Detecting affective covert user states with passive Brain-Computer Interfaces

Thorsten O. Zander
Team PhyPA
TU Berlin, Germany
tza@mms.tu-berlin.de

Sabine Jatzev
Team PhyPA
TU Berlin, Germany
sja@mms.tu-berlin.de

Abstract

Brain-Computer Interfaces (BCIs) provide insight into ongoing cognitive and affective processes and are commonly used for direct control of human-machine systems [16]. Recently, a different type of BCI has emerged [4, 17], which instead focuses solely on the non-intrusive recognition of mental state elicited by a given primary human-machine interaction. These so-called passive BCIs (pBCIs) do, by their nature, not disturb the primary interaction, and thus allow for enhancement of human-machine systems with relatively low usage cost [12, 18], especially in conjunction with gel-free sensors. Here, we apply pBCIs to detect cognitive processes containing covert user states, which are difficult to access with conventional exogenous measures. We present two variants of a task inspired by an erroneously adapting human-machine system, a scenario important in automated adaptation. In this context, we derive two related, yet complementary, applications of pBCIs. First, we show that pBCIs are capable of detecting a covert user state related to the perception of loss of control over a system. The detection is realized by exploiting non-stationarities induced by the loss of control. Second, we show that pBCIs can be used to detect a covert user state directly correlated to the user’s interpretation of erroneous actions of the machine. We then demonstrate the use of this information to enhance the interaction between the user and the machine, in an experiment outside the laboratory.

1. Introduction

The introduction of methods from statistical machine learning [1] to the field of brain-computer interfacing (BCI) had a deep impact on classification accuracy and it also minimized the effort needed to build up the skill of controlling a BCI system [2]. This enabled other fields to adapt methods from BCI research for their own purposes [18]. A particularly exciting development is the adoption of BCI technology into general Human-Machine Systems (HMS), i.e. for healthy users. In the context of HMS, a BCI constitutes a new communication channel, which enables direct insight into the cognitive and affective user states in response to the environment and technical system (see figure 1). However, not all types of BCIs are equally applicable in this context.

We categorize the methods commonly applied in BCI research into active and reactive. By the term active BCI we denote BCIs which utilize brain activity of direct correlates of intended actions as input. This includes the detection of motor imagery or execution for active control, or the control over slow cortical potentials.

A reactive BCI is still controlled via intended actions. However, in contrast to the active BCI, features are not derived from direct correlates to these actions, but from cognitive reactions to exogenous stimuli, as e.g. in the P300 speller. According to this line of thought, we now define passive BCI [4, 17]. Passive BCIs are not used with the purpose or ability of explicit voluntary control. Instead, they infer cognitive states which are already present within the primary interaction in a human-machine system. Examples are brain states or cognitive events that are automatically and implicitly induced by the primary interaction. Hence, the inclusion of a passive BCI channel to an existing human-machine system does not directly interfere with the primary
Type of BCI | Based on features from | Used for
--- | --- | ---
Active | dedicated and intended action encoding a command focussed on directly generating a BCI detectable signal | direct control.
Reactive | dedicated perception on an exogenous stimulus focussed on indirectly generating a BCI detectable signal | direct control via external stimuli, brain switch
Passive | changes in cognitive states or occurring of cognitive events directly resulting from the current Human-Machine Interaction without the necessity of any additional user effort | implicit interaction, providing information on (affective) covert user states

Table 1. Categorization of BCI systems based on type of the used features and fields of application. The classical definitions of active and reactive BCIs are augmented by the definition of passive BCIs. These benefit from high classification rates and a low usage effort.

Mode of interaction, and the passive BCI forms a secondary communication channel (see Table 1). EEG features based on implicit brain reactions to the environment also seem to be more robust in comparison to features utilized for active BCIs, which are often dependent on intended user actions and more variable in nature. This might be due to the fact that passive BCI features usually depend on automatic processes of cognition which are not as easily modulated by conscious processes. For these reasons, passive BCIs are readily applicable in the general Human-Machine Systems context. Especially, they allow for insights into user states, which are hard to infer from the users behaviour or other exogenous factors. We call these user states 'Covert User States' (CUS) analogous to the term covert attention (defined in [11]). CUS can refer to the user’s interpretation of the current interaction states, which is usually not communicated directly to the technical system, but only conveyed indirectly by reactive user actions in response to this interpretation and task goal. This makes an interpretation and therefore an adequate adaptation for the machine to the user’s need difficult. For instance, the user often expects a specific response of the machine to his behaviour. In adaptive systems this expectation is not always fulfilled, as the adaptation mechanism usually is based on a fixed rule system interpreting the user’s behaviour. Hence, a corrective action by the user is necessary, which may disturb previous goals and strategies. The information of the user’s actual interpretation of the situation, e.g. the CUS "This is wrong!", could fundamentally augment the rule system of the machine and thereby enhance the adaptation performance.

In addition, CUS refer to cognitive and affective events or states that might be visible in explicit behaviour but are ambivalent until a direct categorization is possible. Since there is no direct communication between man and machine - with respect to the motivational and emotional response of the user - an adequate adaptation is difficult. Consequently, mental parameters related to these processes are an interesting addition to human-machine systems.

Affective CUS which are possibly detectable within the passive BCI framework range from mental workload, relaxation, surprise, and attention to arousal, frustration and more. Here, we are going to investigate more specific affective CUSs. In the subsequently presented scenarios, we investigate the impact on brain responses of misbehaviour of the machine, and their detectability via passive BCIs. Two different approaches are presented, one focusing on the latent state of loss of control over a system, and the other focusing on the immediate response following a faulty and surprising interaction state.

2. Methods

2.1. Specifications of our BCI system and experimental design

2.1.1. Recording

The EEG system has 32 channels of Ag/AgCl conventional (EasyCap) as well as impedance optimized (ActiCap) electrodes. Signals are amplified by a BrainAmp DC system and recorded by the BrainVision Recorder (BrainProducts). The electrodes are distributed on standard 10/20-based caps with 128 positions. Depending on the type of experiment, they are placed according to the design of the cortex. Additionally, we record electrooculogram (EOG) for controlling feedback-induced correlated eye movements, and electromyogram (EMG) on the relevant limbs, for protocolling correlated movements. Both are bipolarly multiplexed by a BrainAmp (ExG) system and derived with Ag/AgCl electrodes. In order to retain information on exogenous factors, we also record ambient temperature and noise level within the laboratory.

2.1.2. Experimental Conditions

The stimulus presentation in calibration phases before online feedback is designed for providing high control over exogenous and correlating factors besides the one of interest.
This control is relaxed in certain online feedback sessions to allow for a more realistic mode of interaction. A realistic HMS interaction mode is characterized by a distinctive motivational component with regard to the user, whose behaviour is driven by a certain task goal he likes to achieve. In this sense, the experimental paradigm can mimic real world scenarios, where mostly the motivational aspect is modulating the user’s mental states and actions. This can be accomplished by putting the experimental paradigm into a game context, as is the case for the RLR-Game (see section 2.2). This decrement of control over factors might allow for a higher number of artifacts but does decrease the signal to noise ratio. Subjects have been introduced to the main factor of investigation by an instructor. Experimental tasks have been presented in a standardized way on the screen of the Feedback Unit. The course of the experiments contained several breaks for relaxation and recovering of the subjects. Subjects gave information on their overall state and their impressions on different blocks of the experiment by answering questionnaires. All subjects are from age 18 to 45 with German as primary language. The groups of subjects are of mixed and approximately balanced gender. Each subject was paid 20 Euros after completing the experiment.

### 2.1.3 Analyses

**Classification:** For offline analyses, all feature extraction methods, including filtering and resampling, are applied in a strictly causal way. Classifiers are chosen from several linear (LDA, rLDA, SVM) and non-linear (kernel SVM, RDA, GMM) methods. In all analyses presented subsequently, (regularized) LDA was the best performing classifier and was therefore selected. Classification accuracy was estimated by 10x(10×5) [nested] crossvalidation if not otherwise stated. Results from offline analysis are derived from strictly separated training and test blocks. Significance statements are substantiated by standard T-Tests and F-Tests without assumptions on the type of underlying distributions.

**Feature extractors:** For the extraction of features correlating to finger movements, two methods are used. First, the Common Spatial Patterns for Slow Cortical Potentials (CSPfSCP) algorithm [6]. CSPfSCP aims to find linear combinations (patterns) of EEG channels such that the detection of each trial projected according to these patterns is most discriminative (i.e., differs maximally between the two classes). This version is optimized to detect the deflection of the readiness potential (Bereitschaftspotential). This is an SCP indicated by contralateral low-frequency changes (1-5Hz), in this case localized over motor cortex. A slow negativity can be observed prior to a movement, and the relative strength of this negativity in the channels over the left versus right cortical hemisphere is typically used to infer the laterality of the upcoming movement. And second, for the extraction of spectral features correlating to event related desynchronisations (ERD) we used another version of CSP, Spectrally Weighted CSP (SpecCSP) [15]. SpecCSP iteratively alternates between optimizing spatial and the spectral criteria. This way, the algorithm calculates a set of custom spatial projection together with a set of custom frequency filters. These are generated for discriminating ERD by logarithmic bandpower. For the single trial detection of other event related potentials, in this case the EEG pattern correlating to error responses of the brain, features have been extracted by a derivate of the pattern matching algorithm [3]. It has been extended for detection of several extrema of SCPs within a given epoch. The data is resampled at 100 Hz, epoched relative to the event marker and a FFT band pass filter was applied using a frequency range of 0.1 - 15 Hz. Pattern matching reduces the dimensionality of the EEG data, by partitioning trials into n time windows according to the proposed ERP shape and calculating the mean of each time window and single trial. This results in an n-dimensional feature vector for each of m EEG channels. The EEG data is mapped onto an n*m dimensional feature space, containing the class-specific features of the EEG signal.

**Dependent measures for statistical non-stationarities:**

For detecting non-stationarities in movement-related features, i.e. for executed button presses of the left and right hand, we implemented two methods. Both measures were calculated relative to the training data’s distribution of the initial calibration measurement. In the first one, we explicitly applied a measure of statistical deviation to feature distributions. We measured the Kullback-Leibler divergence (KLD) of the feature distributions for direct observation of non-stationarities [14]. The second method aims at detecting non-stationarities implicitly, by measuring the performance of a movement classifier. We define pseudo online classification rates (POC) for this purpose. POC rates were calculated by offline analysis serving as estimation for online classification results. They were determined as following: A classifier was trained on the initial training block. Then, this classifier was applied to every key press. An average of approx. 100 gradual classifier outputs in a one-second window before each key press was averaged and taken as the classifier’s decision for this key press. The sign of this decision value (by default, left keys, on average, were assigned -1, right keys +1) was remapped according to the key actually pressed, such that correct decisions were assigned positive values and wrong decisions were assigned negative values. The result is a real number for each key that was pressed by the subject. Therefore, positive values would indicate overall correct classifier decisions, while values close to zero or negative would indicate overall wrong decisions.
2.2. The RLR paradigm and its directed restriction, the RLR-Game

For both experiments, variants of the same experimental paradigm have been used, the Rotation-Left-Right paradigm (RLR) [9] and the RLR-Game. The RLR paradigm has been developed to mimic Human-Machine Interaction and to induce different mental user states by manipulating factors of interest (see figure 2). The goal of the experimental task is to rotate a stimulus clockwise or counter-clockwise (by a right or left key press, respectively) until it corresponds to a given target figure. The stimulus is either the letter "L" or "R", indicating the direction of rotation and left or right button press. While the colour of the stimulus is grey, it can not be rotated. However, every 1000 ms it changes into one of three colours, indicating A) the possibility to be rotated by a key press and B) the angle of rotation. If the stimulus lights up in red, the stimulus will rotate by 90 degrees, if it is yellow, by 60 degrees, and if it is green, by 30 degrees, upon key press. Each rotation has to be triggered, which only can be done once per colour change. The subject has to build up an efficient strategy for reaching the target: to rotate the starting stimulus as fast as possible on the target stimulus without rotating too far. The design can be played in two modes: The first was restricted to what we will call ‘Full Control Mode’ and the second, the ‘Reduced Control Mode’ included additional ‘random states’. Random states are different from standard states in that they use a different rotation angle after key press, randomly selected from 90, 60, and 30 degrees. Consequently, in this case, the learned mapping rules do not apply anymore.

A derivate of the RLR paradigm is the RLR-Game, which is restricted to the colours green and red. It has also two stages, the ”correct mode” and the ”error mode” (see figure 3). In the ”error mode”, there will appear error states with a chance of 30%. The new angles in the error states are chosen to be always smaller than the ones from the corresponding standard case. Consequently, the colour red will result in 30° rotation (opposed to 90°), and the colour green will result in a 0° rotation (opposed to 30°). Secondly, the RLR-Game adds a second player, competing to the first one. Their performance is measured and fed back in form of points. A player gets a point when hitting the target earlier than his opponent. Hence the artificially induced machine error has a negative valence for the user, since it decreases his performance and might even lead to frustration.

3. Experimental Scenarios

In this section we are going to present two studies. In both the factor of investigation is the utilization of a specific CUS. The first study is based on the RLR design and handles the CUS of perceiving loss of control within human-
Figure 3. The RLR-Game and its two modes, correct and error. In error mode the system will react on a chance of 30% with a decreased rotation angle. A correct classification will then switch back to the standard rotation angle. On a false positive classification the angle would be correct and then, due to the classifier's mistake, be decreased.

machine interaction. In the second study we investigate the user’s interpretation of automated adaptation within the RLR-Game.

3.1. Defining a pBCI for detecting the perceived loss of control within Human-Machine Interaction

3.1.1 Motivation

When conventional active BCI applications are transferred from the laboratory to interactive scenarios, influence of most interfering factors is lost, and such interference can lead to prolonged drops in performance. In this context, pBCIs may give insight into the underlying mental or affective user states, and related overall system states. In general, most HMSs are lacking information about the user’s capability of handling the technical system, or whether the user is overwhelmed with the current task. Therefore, the machine is unable to adapt to the users needs and cannot supply the necessary support to avoid interaction mistakes.

3.1.2 Factor of investigation

An affective parameter that may underlie all of these situations is the loss of control (LoC), which makes conventional active BCI applications a good candidate for the investigation of the LoC. Moreover, a sufficiently universal method in the pBCI framework for detecting the LoC state may lend itself well to a much broader range of applications. This applies especially to those in which the primary interaction is performed by manual actions. As previously mentioned, accessing this state is difficult using conventional HMS channels, as LoC falls into the category of covert user states.

3.1.3 Approach

In this explorative scenario, we investigate the feasibility of designing a pBCI to detect the CUS of the loss of control over a task. It has been assumed that in the restricted context of active BCIs, e.g. control via imagined movements, the LoC manifests itself in non-stationarities in the underlying features. This leads to deviations from the feature statistics during the BCI calibration phase, with the consequence of degraded system performance [5,14]. Theoretically, if present, this statistical behaviour can be detected passively, using the same features as a basis. However, to also allow for operation in more general HMS cases, features must be derived from executed movements, such as typing. Assuming that typing produces features that are similar to those occurring during imagined movements in conventional BCIs, we can reapply standard techniques for movement-related features to our new situation in a passive way. Standard feature extractors for imagined and executed movements are compared in their sensitivity with respect to the LoC, and thus in their predictive performance for use in a passive LoC detector.

3.1.4 Experimental Design

By utilizing the RLR Paradigm we have been able to artificially induce phases of reduced user control (phase BUc, see figure 5) by permuting the mapping between colours and angles of rotation. The learned rule system would not apply any more and therefore the user is confronted with an unexpected behaviour of the technical system, experiencing a loss of control of the task. In these experiments 22 subjects participated. We tracked features representing the primary mode of interaction, pressing a key, in the EEG data. Details on this study can be found in [9].
3.1.5 Features and analyses

Loss of control is accessed by analysing feature modulations of EEG patterns correlated to the executed hand movements. The EEG features that allow left and right hand movements to be discriminated fall into two categories: Slow Cortical Potentials (SCPs) and Event-Related Desynchronization (ERD) features. We have chosen features from both categories which have been extracted by Common Spatial Patterns for SCP (CSPfSCP) and Spectrally Weighted CSP (as described in section 2.0.3) for ERD from 200 ms of data prior to the button press. For the detection of LoC we have calculated the KLD on a moving window, containing the data of 10 button presses compared to the data from the initial training phase. Also, we estimated the POC for each buttonpress.

3.2. Applicability of a pBCI for enhancement of efficiency in HMS

3.2.1 Motivation

Errors in communication are highly relevant factors regarding the efficiency of HMS. Especially with regard to automated adaptation of the machine to the interaction mode of the user [10]. The machine tries to adapt to the behaviour and needs of the user. The currently used approaches are based on inferences of the user’s actions or machine inputs. But a precise adaptation on this restricted information is hard to accomplish, because the mental states of interest are mostly CUSs. A wrong automation decision induces effects of surprise and frustration and in this respect, adaptation reduces the performance and the safety in HMS [13]. Additionally it triggers a correction action which enforces a shift in the intention focus of the user. According to this it reduces the overall acceptance of the adaptation and of the whole system. The goal of this scenario was to investigate a pBCI that is capable of communicating these error-related brain responses of the user to the technical system.

3.2.2 Factor of Investigation

The RLR-Game mimics the interaction in an HMS and allows for modelling an unexpected and negative effect, the error states. While this game is based on common interaction channels, we have added a secondary and passive BCI channel capable of automatically correcting the effects of reduced angles in the error states (see figure 3). This correction is triggered by an event-related potential reflecting the mental processing of an error trial. If it is correctly detected by the pBCI during an error trial, the rotation angle was set to the correct mapping. In case of a false positive the angle was reduced to that of a corresponding error state. Hence, each correct detection of an error brain response speeds the player up and a false detection slows him down. Therefore, if the classifier works properly, it will enhance the performance of the player and it will reduce it otherwise. See figure 4 for details.

3.2.3 Experimental Design

For keeping the environment as realistic as possible, we have chosen the Open House of the TU Berlin (LNdW 2007) as the setting. Four times two different players from the audience played the RLR game against each other (see figure 3) for a visualisation of the rule system). Each pair played three sessions of 50 trials. The first was for user training, without error states. In the second session we introduced the error trials. The automatic adaptation has been applied in the last session, only for one player.
3.2.4 Features and application

For detecting brain response relative to an erroneous rotation, we epoched the data from 0 to 800 ms relative to the stimulus rotation. On these epochs we applied the pattern matching method with 8 windows of 50 ms length starting at 300 ms after the event. This is resulting in a 256 dimensional feature space. Based on the features extracted from the data of the second RLR-Game session a classifier has been trained. In the third session this classifier was applied. On each trial a classification was requested directly after the stimulus rotation.

4. Results

The results of the LoC study (Figures 5 and 6) show that for the phases with full control (A1, A2, Ba1, Ba2) the variance of the averaged Kullback-Leibler divergence (KLD) is bounded for ERD features. In contrast, the phases of reduced control (BUc) reveal a significant (p<0.05) increase of the KLD, for ERD-based features. The POC drops down in phase Bm, which correlates significantly (p<0.05) to the course of the KLD. Hence, the KLD of this feature category is a measure strongly related to the perception of control by the user. Contrary there are no significant changes in the features extracted for slow cortical potentials.

Figure 7 shows the results from the sessions from the open house of the TU Berlin 2007, investigating the pBCI based on error potentials. During the third session one player was supported by the pBCI, correcting erroneous states of the machine by detecting brain activity induced by the respective machine error states. While the points have been equally distributed between session 1 and 2, the performance of all pBCI supported players has been increased significantly. The classifier hat an accuracy of 81.2% with error ratios equally distributed over the two classes.

5. Discussion

The results of the LoC scenario show that we have found a possible BCI measure for an affective mental parameter, which is sensitive with respect to loss of control in an executed movement task, relying on oscillatory features. Features based on the readiness potential, however, show no sensitivity. This indicates that the corresponding mental parameter could be detected passively, using SpecCSP and KLD as building blocks, and a technical system could be supported by it. As the loss of control is an important CUS to be transferred to the technical system in order to enable an adaptation of the system to the user’s needs, this is an important route for further investigation. Especially, the online applicability and specificity of the inference have to be further investigated.

Erroneous system behaviour results in frustration of the user and a deteriorated man-machine interaction. The investigated pBCI online detection of brain responses to machine errors clearly allows for an enhancement of the human-machine interaction. Currently, a further study is being undertaken to validate the pBCI based on error responses, by investigating EEG patterns correlating to different error cat-
6. Conclusion

Here we gave examples of two types of pBCIs. One establishing an information flow from the human brain to the HMS, reflecting CUS correlated to current modes of interaction. The other one extracting the actual interpretation of dedicated system states from the users cognition. Both can be applied in the context of BCI for enhancing classification accuracy. First, for automated adaptation of BCI classifiers, and second, for correcting errors in Human-Machine Interaction as proposed in [8]. In the more general context, our results show that pBCIs are suitable for an application in the field of HMS, providing information about the mental user state, which can only hardly be inferred by typical information channels in HMS. Our experiences with pBCIs show that these enable new channels of information within the interaction between man and machine. Next to an increased efficiency of work, automation technology has caused additional difficulties in HMS. This leads to errors and safety risks, mainly due to maladapted man-machine communication. pBCIs enable a direct access into CUSs, which is not currently accomplished by any other HMS method. Anyhow this is an important precondition for an optimal adaptation of automated agents, to make the man-machine interaction more efficient and less prone to errors. Here we were able to show that pBCIs are capable of detecting brain activity in response to machine errors and thereby enhancing automation adaptation. Additionally it seems to be fruitful to exchange experiences between the fields of HMS and BCI research, which will hopefully be done extensively in the near future. These studies could be a starting point for a whole series of new approaches. Currently we are investigating pBCIs for detection of mental workload, cognitive interpretation of the perception of human movements [8] and information on driver intentions. Please see www.phyPA.org for details.

References