# BCI detection of deliberately hidden user states – or: Can we detect bluffing in a game?

**Jessika Reissland, Christian Kothe, Sabine Jatzev, Matti Gärtner, Sebastian Welke and Thorsten O. Zander**

*Keywords: passive brain-computer interface (BCI), covert user states (CUS), implicit human-computer interaction (HCI), electroencephalogram (EEG)*

## Abstract

Previous studies suggest that human-machine interaction could be enhanced by providing information about the actual user's state, thus allowing for automated adaption of the system. While some user states might be inferred from the user's behaviors, others like e.g. the perception of a self- or machine-induced error are hard to access externally. But such *covert user states* are reflected in physiological parameters like EEG and can be detected with a passive brain-computer interface (pBCI).

Aim of this study was to expand the complexity of covert user states, which can be detected by a pBCI. An inherently covert and therefore complex state is the deliberate attempt to mislead an opponent in a game (e.g. poker) – namely *bluffing*. We recorded high-density EEG from 6 pairs of subjects while they were playing a bluffing dice game against each other. The game included dedicated states in which the players had to bluff or to give up. When classifying whether a player would bluff or not, we achieved a cross-validated single-trial accuracy of 81.4% (± 6.5%) over all subjects. Based on that, further investigations of the detection of covert user states show promise to enhance the performance of human-machine systems.

## Enhancement of HMS using knowledge about user states

With the proceeding development and increasing deployment of technical systems, the technical part of those systems becomes more and more complex and sophisticated. But every technical system has a user who has to cope with it. Although much research focuses on the technical part of such human-machine systems (HMS), rather little attention has been paid to ensure the user's capability to interact with the system.

User-friendly design of HMS has therefore become an important part of current research. New approaches evolve such as adaptive or interpretative HMS heading for optimal support of the user (Chen & Vertegaal, 2004; Rötting et al., 2009). One possible scenario is the adaptive reallocation of actions between operator and machine according to e.g. fatigue or attention. Or a computer could show the user the information he was looking for if it knew what it was. An application could also exit automatically in case the user notices that he opened it erroneously.

The key information for the design of such context-sensitive systems is knowledge about current mental or cognitive states implicitly generated by the user like arousal, fatigue, workload, or intentions. Furthermore, emotional user states like surprise, satisfaction, or frustration provide interesting information. This could lead to a more human way of interaction between human and machine as well as to a new kind of adaptive and interpretative, and above all, to more holistic HMS (Rötting et al., 2009; Cutrell & Tan, 2008).

## Detection of covert user states by passive BCI

How could these user states be detected? Access would be possible by questionnaires or ratings of the users – but those would suffer from subjectiveness, interruption of the user's interaction with the system, and retrospectivity. Furthermore, users could deliberately hide these user states. Overt measures like behavioral data (e.g. mouse-click behavior, reaction times, or error rates) or psychophysiological measures are more objective without being influenced by the user's memory or social desirability and provide continuous data without interruption of the interaction with the system. Using haptic data (Park et al., 2005) or eye gaze (Asteriadis et al., 2009; Rötting et al., 2009), it has been shown that providing the system with this kind of psychophysiological data leads to improved human-machine interaction (HMI).

Nevertheless, although these measures might correlate to cognitive processes, they can only indirectly give insight into the actual cognitive state of the user (Müller et al., 2008). User states have a covert component, which cannot be observed from the outside. This gives the basis to define those user states of interest as *covert user states*, analogously to *covert attention* (Posner & Cohen, 1984). One psychophysiological measure, which is continuously available and can provide objective real-time information about user states, which are hardly inferable from exogenous factors, is the electroencephalogram (EEG). It has indeed been shown that covert user states like mental effort, workload, fatigue and attention can be identified in EEG (Gevins et al., 2007). Furthermore, Musha et al. (1997) extracted different EEG features for emotions. Very interesting EEG features for HMI design are error potentials (Falkenstein, 2000) or loss of control (Jatzev et al., 2008).

## Passive BCI

Classic EEG analysis methods apply averaging methods to eliminate various sources of variability (Luck, 2005). This approach is not applicable in HMS because the feedback has to be accessible in real-time and on a single-trial basis. Researchers in the field of brain-computer interfaces (BCI) work on suitable methods to solve this issue (Müller et al., 2008; Zander et al., 2008). The traditional BCI approach uses either *active BCI*, which utilizes the brain activity of direct correlates of dedicated actions encoding a command as input, or *reactive BCI*, which focuses on brain signals elicited by the perception of exogenous stimuli (Zander et al., 2008). But both, active and reactive BCI, have drawbacks with respect to our approach: not only would the user have to learn how to control their brainwaves in order to achieve best results, but also would this method not be applicable for a naturalistic, continuous interaction.

Therefore, our method of choice is the *passive BCI* (pBCI; Zander et al., 2008; Cutrell & Tan, 2008), which is based on neurocognitive events or states. These are automatically induced while interacting with the system, so it can be operationalized without any additional effort by the user.

## Complex covert user states

In the line of pBCI, previous studies showed that real-time classification of covert user states is indeed possible on single-trial basis. Grimes et al. (2008) achieved really good classification results of working memory and cognitive workload using EEG data. Lee and Tan (2006) could convincingly classify which task a user was working on based on EEG. A direct enhancement of an HMS has been shown by Zander et al. (2008): error potentials were detected with a pBCI. This information was fed back to the system and resulted in an automatic error correction, which improved the performance of the HMS.

Based on these results demonstrating that the detection of covert user states can be used to improve HMI, the present study focuses on the question how far this detection of covert user states can go. Which grade of complexity is still detectable by a pBCI on a single-trial basis?

In everyday interactions among humans, actors merely have explicit information about inner states of each other. Nevertheless, understanding intentions of others is an important ability that involves representing the mental states of others in one's own mind as postulated by the "theory of mind" (Premack and Woodruff, 1978). Such information is reflected in covert mental states. These states might also be relevant for a more intuitive HMI (Asteriadis et al., 2009; Moldt & von Scheve, 2002), but are almost completely inaccessible to the machine. Using a pBCI this information could be retrieved. As an example of a complex covert state we chose to investigate an interesting social concept in which the actor's state is inherently covert: *bluffing* - which we define as the deliberate attempt to mislead an opponent in a game. Bluffing is a complex covert state involving many cognitive processes like risk calculation, mentalizing, decision-making, anticipation and preparation of response, and self-monitoring.

A successful classification of bluffing in a game in real-time would provide evidence that even complex covert user states like social intentions can be accessed with the help of pBCI and therefore operationalized as feedback in an HMS. This would bring HMI an exciting step forward towards more intuitivity and naturality and would also lead to promising future perspectives for the design of HMS.

### Expected EEG features for bluffing

To determine heuristically relevant EEG features for the classification, a contingent negative variation (CNV) that appeared more prominent in deliberately deceptive responses in lie detection (Fang et al., 2003) seemed most promising. The CNV occurs prior to a motor response, electrodes over motor areas are therefore likely to be relevant for classification. CNV modulation has also been shown to correlate with increased activation of the dorsal anterior cingulate cortex in decision-making (Walton et al., 2004), so we also expect relevant EEG activity on frontal electrodes.

## Methods

### Experimental Task and Set-up

For the experiment 12 paid subjects were invited. Pairs of them played a game against each other, while we recorded high-density EEG (Brainproducts‘ ActiCap) from each player.

The most characteristic bluffing game, namely poker, faces the problem that it does not objectively allow to discriminate when a player is bluffing or not. So it might well happen that somebody really thinks he is going to win with his hand while another player would bluff with the same hand. Therefore, we adapted a German drinking game called "Mäxchen", which included states dedicated to the decision of the player whether to bluff or to quit.

Two subjects played against each other, starting with an account of 20 points, inconspicuously rolling two dice indicating a two-digit number in alternation. The player in turn needed a higher number than his predecessor, otherwise, he had to bluff or to quit. However, bluffing was more risky as its detection by the opponent cost 2 points, while quitting only cost 1 point. If a bluff was wrongly accused, the accuser lose 2 points. A game was won when the opponent had lost all 20 points. Each pair of subjects played eight games. For motivation, subjects got extra monetary bonus for each won game.

The players were sitting at a table with a computer screen per player, facing each other, and rolling virtual dice on the computer. The screen was set low enough for them to see each other, as this is crucial for the social interaction component of bluffing. A typical trial had the following time line: player A pressed a button to roll his dice, the result appeared on screen 2 seconds later. He had to wait for an auditory go-signal to announce his true or alleged number or to quit. Responses were given verbally, synchronously with a button press to cue the time of response. After player A's response player B had to decide whether to accuse A of having bluffed or to accept the number and to roll the dice in his turn.

## Data Analysis and Classification

The goal of our method was to classify trials into the categories *bluffing* or *not bluffing* based on ongoing EEG. For this, we only used trials in which a player announced a different number than he rolled and contrasted bluffs versus quits. The alternative was to contrast bluffs versus non-bluffs in general, but that comparison would have been confounded with whether the situation was risky or not. Hence, we could not reliably tell whether our method captured bluffing itself or just winning or not winning. Of course, this reduced the amount of trials for the classification remarkably, which was a particular challenge for the training of the pBCI classifier: on average, in each game one player rolled the dice 203.5 times; from that, he told 123.2 times the truth, bluffed 44.3 times and quitted 35 times. One subject had to be excluded from analysis due to a number of bluff/quit trials far below average.

For training of the classifier we used an extension of the pattern matching method proposed by Blankertz et al. (2002). We preprocessed the data by sampling it down to 100 Hz and by using a bandpass filter from 0.1 to 8 Hz. Previous findings (Fang et al., 2003) suggested relevant EEG features in the time window around the button press when announcing the number or when quitting. Therefore we defined four features by averaging the EEG data in four concatenated 50 ms time windows around the button press. The features were separately calculated per subject for each of the 68 EEG channels. Due to a subject-specific variance of the CNV the set of windows was chosen with a jitter of 75 ms around the defined time windows (see figure 1). In order to show that the pBCI is capable of discriminating the examined states on single-trial and real-time basis, we used a cross-validated offline-analysis (Bishop, 2006). With this method we estimated how a classifier, for which we chose a regularized linear disriminant analysis (LDA), would perform in an online application. To cope with the small number of trials we estimated a regularization parameter by a 10x10 nested crossvalidation.

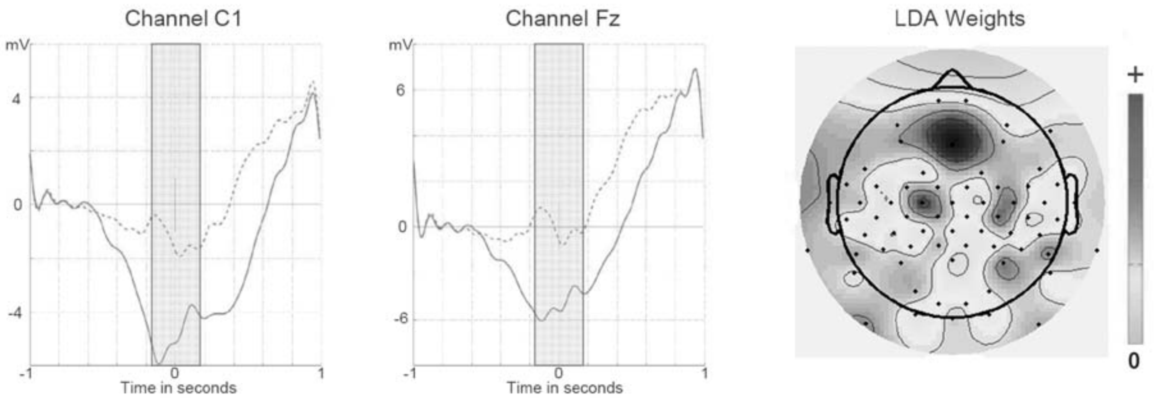


Fig. 1: Left: The grand average of the event related potentials (ERPs) of the bluffing (solid) and the quit (dashed) condition. The marked areas indicate the timeframe from which the features could be extracted. Right: The relevance of electrodes for classification is displayed by topography of the LDA weights. The black dots indicate the 68 EEG channel locations.

## Results

The nested crossvalidation estimated a performance of the pBCI of 81.4% (±6.5%) accuracy, with 94% as maximum and 72.4% as minimum. Averaging the window sets chosen by the classifier resulted in a timeframe from -80 to 120 ms relative to the button press, which matched the assumption that the prospected features could be found around the button press. As we faced the problem of a restricted number of trials, we could indeed observe a correlation of the number of trials in both classes to the individual classification accuracy: if the number of trials was low, the accuracy was low as well. As expected, the topography of the LDA weights (see figure 1) showed that relevant electrodes for classification were selected over motor cortex (C1 and C4) and anterior cingulate cortex (Fz). Some peripheral electrodes were also weighted, pointing to non-brain influences.

## Discussion and Conclusion

In this study, the classification accuracies we achieved suggest that we can reliably identify whether a person would bluff or not based on EEG signals. Since the hidden bluffing event served as an example for complex covert user states we could indeed show that such states are generally determinable with the method of pBCI.

However, based on current analyses, we cannot be certain that the phenomena providing the classification power is entirely generated by neuronal activity. EEG signals are also sensible to voltage changes evoked by eye movements or muscular activities. Such physiological responses might be involuntarily coupled with the bluffing event, e.g. if the player might show a specific eye movement pattern or neck muscle tension while bluffing in contrast to not bluffing. External psychophysiological data like eye movements can already give insights to cognitive processes to a certain extent (Rayner, 1977). If such artifacts are highly correlated with the covert states, which are to be identified, they can be exploited to improve the classification accuracy (Lee and Tan, 2006). Analysis tools like independent component analysis (ICA) and source localization (Makeig et al., 1996) provide more insight into neural processes represented in the EEG. For this study, first preliminary results using ICA indeed suggest that classification accuracies increase when using only EEG data assumed to be non-artifactual neural activity. EEG data based on pure neural processes might be more stable for user state classification because of a high subject specific variance and situational impacts of the environment. More research is needed at this point.

It might also be possible to calculate the bluffing behavior of a player based on probabilistic game scores. Cost-benefit trade offs have to be taken into account on an individual basis to determine which method is suitable to be integrated in an HMS. However, the aim of this study was to make a first methodological move towards pBCI classification of complex covert user states. As we could indeed show this, further investigations are needed to elicit which other complex covert user states can be identified and classified from the EEG patterns. Considerations how to include this method into user interfaces to enhance HMS have to be further explored.

## References

Asteriadis, S., Tzouveli, P., Karpouzis, K., Kollias, S. (2009). Estimation of behavioral user state based on eye gaze and head pose – application in an e-learning environment. *Multimedia Tools and Applications, 41*, 469-493.

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning.* New York: Springer Verlag.

Blankertz, B., Schaefer, C., Dornhege, G., Curio, G. (2002). Single trial detection of EEG error potentials: A tool for increasing BCI transmission rates. Artificial Neural Networks, ICANN 2002.

Chen, D., Vertegaal, R. (2004). Using mental load for managing interruptions in physiologically attentive user interfaces. *Extended Abstracts of SIGCHI 2004 Conference on Human Factors in Computing Systems*, 1513-1516.

Cutrell E., Tan, D. (2008) BCI for passive input in HCI. *Proceedings of the CHI-Conference*, Florence, Italy, 2008.

Falkenstein, M., Hoormann, J., Christ, S., Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology, 51*, 87-107.

Fang, F., Liu, Y., Shen, Z. (2003). Lie detection with contingent negative variation. *International Journal of Psychophysiology, 50*, 247-255.

Gevins, A., Smith, M.E. (2007). Electroencephalography (EEG) in Neuroergonomics. In R. Parasuraman & M. Rizzo (eds.), *Neuroergonomics – The Brain at Work*. New York: Oxford University Press.

Grimes, D., Tan., D.S., Hudson, S.E., Shenoy, P., Rao, R.P. (2008). Feasibility and pragmatics of classifying working memory load with an electroencephalograph. *Proceedings of the 20th annual SIGCHI conference of Human factors in Computing Systems*, Florence.

Jatzev, S., Zander, T.O., DeFilippis, M., Kothe, C., Welke, S., Rötting, M. (2008). Examining causes for non-stationarities: The loss of controllability is a factor which induces non-stationarities. *Proceedings of the 4th Int. BCI Workshop & Training Course, Graz University of Technology*, Graz, Austria: Publishing House.

Lee, J.C., Tan, D.S. (2006). Using a low-cost electroencephalograph for task classification in HCI research. *Proceedings of the 19th annual ACM symposium on User interface software and technology*, Montreux, Switzerland.

Luck, S.J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: The MIT Press.

Makeig, S., Bell, A.J., Jung, T., Sejnowski, T.J. (1996). Independent Component Analysis of Electroencephalographic data. *Advances in Neural Information Processing Systems 8*, MIT Press, Cambridge MA.

Moldt, D., von Scheve, C. (2002). Attribution and Adaptation: The Case of Social Norms and Emotion in Human-Agent Interaction. *Proceedings of "The Philosophy and Design of Socially Adept Technologies"*, workshop held in conjunction with CHI'02, April 20th, Minneapolis/Minnesota, USA. National Research Council Canada.

Müller, K., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods, 167*, 82-90.

Musha, T., Terasaki, Y., Haque, H.A., Ivamitsky, G.A. (1997). Feature extraction from EEG associated with emotions. *Artificial Life and Robots, 1(1)*.

Park, N., Zhu, W., Jung, Y., McLaughlin, M., Jin, S., (2005). Utility of haptic data in recognition of user state. *Proceedings of HCI International 11*. Lawrence Erlbaum Associates.

Posner, M.I., Cohen, Y. (1984). Components of Visual Orienting. In H. Bouma and D.G. Bouwhuis (eds.), *Attention and Performance, 10*, 531–556. Hilldale, N.J.: Erlbaum.

Premack, D., Woodruff, G. (1978). Does the chimpanzee have a "theory of mind"? *Behavioural and Brain Sciences, 4*, 515-526.

Rayner, K. (1977). Visual attention in reading: Eye movements reflect cognitive processes. *Memory and Cognition, 5(4)*, 443-448.

Rötting, M., Zander, T., Trösterer, S., Dzaack, J. (2009). Implicit interaction in multimodal human-machine systems. In C. Schlick (ed.), *Methods and Tools of Industrial Engineering and Ergonomics*, Berlin: Springer.

Walton, M.E., Devlin, J.T., Rushworth, M.F. (2004). Interactions between decision making and performance monitoring within prefrontal cortex. *Nature Neuroscience, 7*(11), 1259-65.

Zander, T.O., Kothe, C., Welke, S., Rötting, M. (2008). Enhancing human-machine systems with secondary input from passive brain-computer interfaces. *Proceedings of the 4th Int. BCI Workshop & Training Course, Graz University of Technology*, Graz, Austria: Publishing House.