

The Impact of Loss of Control on Movement BCIs

Boris Reuderink, Mannes Poel, Anton Nijholt

Abstract—Brain-computer interfaces (BCIs) are known to suffer from spontaneous changes in the brain activity. If changes in the mental state of the user are reflected in the brain signals used for control, the behaviour of a BCI is directly influenced by these states. We investigate the influence of a state of loss of control in a variant of Pacman on the performance of BCIs based on motor control. To study the effect a temporal loss of control has on the BCI performance, BCI classifiers were trained on electroencephalography (EEG) recorded during the normal control condition, and the classification performance on segments of EEG from the normal and loss of control condition was compared. Classifiers based on event-related desynchronization (ERD) unexpectedly performed significantly better during the loss of control condition; for the event-related potential (ERP) classifiers there was no significant difference in performance.

Index Terms—brain-computer interfaces, common spatial patterns, electroencephalography, loss of control, mental states, non-stationary signals, event-related desynchronization, lateralized readiness potential.

I. INTRODUCTION

A brain-computer interface (BCI) enables direct communication from the brain to devices, bypassing the traditional pathway of peripheral nerves and muscles. BCIs can help patients that are completely locked-in by paralysis, such as for example patients with amyotrophic lateral sclerosis (ALS), but healthy users can also benefit from a BCI [1], [2]. The traditional approach to BCI is to provide the subject with a control signal that is a fixed function of measurements of neural signals. The user has then to train for weeks or even months to gain adequate, voluntary control over the BCI [3]. An alternative to user training is to use machine learning (ML) methods to automatically construct a (subject specific) model that discriminates between the naturally occurring brain signals associated with a given mental task. These BCIs do require a calibration or training session that is used to automatically construct a subject-specific model of the brain signals before the BCI can be used. A remaining problem is that the basic assumption made by these ML methods that the data in the calibration and application session is independent and identically distributed does often not hold, since the brain signals' statistics are often not stationary but vary over time. This variability in the electroencephalography (EEG) can result in a loss of performance during the application of the BCI [4]–[6].

B. Reuderink is with the Faculty of EEMCS, University of Twente, The Netherlands. e-mail: (b.reuderink@ewi.utwente.nl).

M. Poel is with the Faculty of EEMCS, University of Twente, The Netherlands. e-mail: (mpoel@ewi.utwente.nl).

A. Nijholt is with the Faculty of EEMCS, University of Twente, The Netherlands. e-mail: (nijholt@ewi.utwente.nl).

The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science.

One source of variability in the EEG is related to the changes in the user state. For example, differences in alertness and workload can alter the characteristics of the EEG. This variability might therefore pose a problem for real-world BCI application [5]. In turn, the user state is influenced by the performance of the BCI; for example a non-working BCI could cause frustration, anger or reduced alertness. This interaction between the user state and the BCI performance might form a positive feedback loop, resulting in a BCI that spins out of control. Given the huge drawbacks of a BCI whose accuracy depends on the mental state of a user, understanding the influence of changes in the mental state on the BCI is needed to develop reliable BCIs.

Specifically, the influence of frustration associated with loss of control (LOC) on a BCI is of great interest since it might cause the previously described feedback loop. This influence was investigated in [7], [8]. In these studies, users were instructed to use real movement with their left or right hand to rotate respectively L or R-shaped objects to a target position in order to study the effect of loss of control on the BCI performance. The color of the letter indicated the angle of rotation, and every second the user could press a key to rotate the object in the direction indicated by the shape of the object. After performing a calibration block with cued left/right hand movement and two practise blocks with this so-called RLR paradigm, a LOC was simulated in the third block by occasionally using a wrong angle of rotation in the application. Both an event-related desynchronization (ERD) and an event-related potential (ERP) based classifier were trained on the first block, and applied to the other blocks in an off-line analysis. A significant difference between the training block and the LOC block was found for the distribution of ERD based features, but for ERP features no difference was found. This indicates that there is variability in ERD features related to loss of control.

In this work, we investigate the influence of a mental state associated with loss of control on the performance of a BCI. The difference between our study and [7], [8] is that 1) we use an interleaved block design to control for effects that manifest spontaneously over time, such as increasing fatigue, changing temperature, drying gel on the electrodes etc., 2) we use the same environment for training and evaluating the BCI classifiers to minimize differences not related to LOC, 3) that we use self-reported emotional ratings to validate the effect of loss of control on the mental state and 4) we test and correct for confounding behavioural changes, such as changes in the force, speed or order of the finger movements, and eye movements. In the next sections we describe the methods, results, and discussion followed by conclusions and directions for future research.

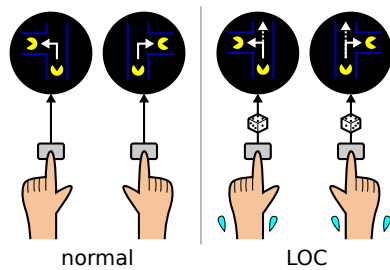


Fig. 1. In the normal condition, the left button rotated the player 90° counterclockwise; the right button rotated the player 90° clockwise. In the LOC condition, 15% of the keyboard input was ignored, and a visual lag was induced (not shown).

II. METHODOLOGY

To study the effect of loss of control (LOC), we designed a Pacman game that periodically reduces the amount of control the user has over his avatar. The EEG was passively recorded during game play, and afterwards the off-line performance of an ERD and an ERP based classifier was used to assess the influence of LOC on the BCI performance. In the rest of this section, we describe the data collection, the preprocessing and classification of the EEG, and the evaluation method in more detail.

A. Data collection

1) *The Pacman Game*: A game was designed to induce a state of LOC, with game play similar to the original Pacman game [9]. The major differences with other Pacman games is that our game periodically tried to induce a state of LOC in the user by responding unreliably to the keyboard commands, which is a proven method for frustration induction [10]–[12]. To simplify (simulated) BCI control, the user input was reduced to a button for the left index finger that turns Pacman 90 degrees counterclockwise, and a button for the right index finger that turns Pacman clockwise.

2) *Experiment design*: The LOC was induced in a randomized interleaved block design with experimental blocks of two minutes. In one third of these two-minute blocks LOC was induced, in the other blocks the game play remained unmodified. The LOC blocks were evenly distributed over the session by building a series from shuffled sequences of three blocks (one LOC and two normal blocks). In the LOC condition, we randomly ignored 15% of the key strokes, resulting in a barely playable game. In addition, the display occasionally lagged in the LOC condition. After each block, the user was asked to rate his mental state in terms of valence (pleasure), arousal and dominance (subjective feelings of control) on a Likert-scale presented under the self-assessment manikin (SAM) [13].

3) *Experimental procedure*: Subjects were asked to read and sign a form of consent, and were subsequently wired with the EEG and physiological sensors. The experimenter briefly explained the game and the self-assessment procedure. The subject was allowed to practise the controls for two minutes before the experiment was started. If users mentioned that the game was unresponsive during the experiment, the experimenter asked them to continue playing and promised to

find the cause later. After 30 minutes, the experimenter stopped the experiment and the users were debriefed.

4) *Sensors and recording*: A BioSemi ActiveTwo EEG system was used to record the EEG and physiological signals at a sample rate of 512 Hz. For EEG, 32 Ag/AgCl active electrodes placed at locations of the Extended International 10-20 system were used. To measure the influence of ocular and muscular artifacts, we recorded the EOG (horizontal and vertical pairs) and two pairs of EMG signals over the left and right flexor digitorum profundus (the muscles used to press with the index finger). Additional physiological sensors, such as temperature, respiration, the galvanic skin response and the blood volume pulse were recorded as well, but not used in the present study¹.

B. Preprocessing

The following preprocessing procedure was applied to reduce the influence of noise and artifacts caused by eye movements and muscle tension: first the recording was downsampled to 128 Hz to speed up processing. After downsampling, the data was high-pass filtered using a 4th order Butterworth filter to remove frequencies below 0.2 Hz, and notch-filtered using a 4th order Butterworth filter from 49–51 Hz to remove power line noise. The EEG was then corrected for eye movements using a regression based subtraction method [14]. To prevent noise from spreading to other channels, we performed channel-level preprocessing before we applied the electrooculography (EOG) correction and re-referenced the signals to the common average reference (CAR).

C. Key press classification from EEG

Most motor imagery based BCIs are based on sensory-motor rhythms, specifically the event-related desynchronization (ERD) that occurs during both real and imaginary movement. Because the ERD of real and imagined hand movement is similar [15], we use real movement to train BCIs that predict the movement from the EEG signal, since it provides a clear ground truth and allows for a tighter controlled experiment. In this section we outline the classifiers used for detection of the ERD and the ERP associated with the movements executed to play the game.

1) *ERD features*: The ERD classification was based on the decrease in the Rolandic mu rhythm (8–12 Hz) and Rolandic beta frequencies (peak around 20 Hz) on the contra-lateral motor cortices that occurs when movement is initiated [16]. After preprocessing, we applied a 6th-order Butterworth band-pass filter to extract the frequencies from 8–30 Hz, which includes both the mu and beta rhythms. From the filtered EEG we extracted windows of one second, centered on the moment that key stroke was registered. Visual inspection confirmed that an ERD did indeed occur in this period. For these segments we trained subject-specific spatial filters with the common spatial patterns (CSP) algorithm.

The CSP algorithm [17] finds a matrix \tilde{W} with spatial filters that map the EEG into a new space with basis vectors that have

¹The recordings are available at <http://borisreuderink.nl/perm/affpac/>.

a high variance for the first class and a low variance for the second, and vice versa. Given the number of sensors s , and the number of samples n , W is an $s \times s$ transformation matrix with the following property:

$$\Sigma_{WX_1} = D \quad \text{and} \quad \Sigma_{WX_1} + \Sigma_{WX_2} = I, \quad (1)$$

where D is a diagonal matrix with decreasing values, I is the identity matrix and Σ_{X_i} is the channel covariance matrix of the $s \times n$ EEG measurements matrix X for given class i . Rows of W that correspond to a high value in D have a high variance (power) for the first class and a low variance for the second, and vice versa. Because of this discriminatory property, the $\frac{m}{2}$ first and the $\frac{m}{2}$ last rows were kept to construct the final matrix \tilde{W} with $m = 6$ spatial filters.

After applying the CSP algorithm, we calculated the variance (which corresponded to the band power in the mu and beta band) for each transformed channel, which resulted in m spatial band-power features.

2) *ERP features*: A less frequently used paradigm for classification of EEG related to movement is based on the Bereitschaftspotential (BP), a negative ERP related to movement initiation. The BP consists of an early phase beginning about 2 seconds before the movement onset, and a late phase with a steeper slope 400 ms before the onset [18]. We used the asymmetric distribution of the late BP over the scalp for classification of the laterality of the hand movements, which is known as the lateralized readiness potential (LRP).

For the LRP classification, we used the same preprocessing pipeline as used with the CSP classification up to the band-pass filter. Instead, we applied an (4th-order Butterworth) low-pass filter at 10 Hz, and again extracted windows of one second centered on the moment of registration of keyboard input. These trials were then transformed with a whitening transform P which has the property that the transformed signals are uncorrelated, and have unit variance:

$$\Sigma_{PX} \approx P \Sigma_X P^T = I. \quad (2)$$

With the eigenvalue decomposition $\Sigma_X = U \Lambda U^T$, we find that

$$P = \Lambda^{-\frac{1}{2}} U^T. \quad (3)$$

After whitening with P , we downsample the signal by taking every fourth sample point, resulting in an $s \times \frac{e}{4}$ feature vector where $e = 128$ is the number of samples in a classification window. Despite superficial differences, this method for LRP classification is conceptually similar to the conventional approaches for ERP detection, such as [19], [20], but does not rely on time segment or channel picking.

3) *Classification*: The ERD and LRP features are used to train a final linear support vector machine (SVM) classifier. The SVM's regularization parameter c was selected with a separate cross-validation loop on the two-minute blocks in the training set.

D. Performance measure

BCI classifiers are often evaluated by comparing their respective accuracies. This is problematic, as accuracies (or

equivalently, error rates) are hard to interpret when the prior probabilities of the classes are unequal and/or variable. Furthermore, the statistic does not take the time needed to perform a trial (key press) into account. Due to our short inter-trial intervals (ITIs), lower accuracies are to be expected for our BCIs than the accuracies reported for more traditional BCI environments, where multiple seconds are used to detect an imagined movement. Despite these drawbacks, we provide accuracy measures.

The information transfer rate (ITR) conveniently captures the amount of information a user can communicate through a (noisy) channel with an optimal encoding strategy. It does so by combining the quality of and the time needed for the predictions. As such, the ITR is a better measure to evaluate BCI performance. Note that different formulas to calculate the ITR are used in the BCI literature, for example Wolpaw's definition in [3] is often used. This definition has a number of assumptions that are often violated in practise, most notably the assumption that all classes have the same prior probability. ITR based on mutual information (MI) does not rely on these assumptions, but the labels of the trials still need to be independent of each other for a correct estimate of the ITR. Due to the need of fewer assumptions, we use MI to measure the information contained in the prediction of a single trial. The MI expresses the decrease in uncertainty of a discrete variable \mathcal{Y} (the true labels), given a discrete variable \mathcal{Z} (the predictions of the classifier):

$$I(\mathcal{Z}; \mathcal{Y}) = \sum_{\substack{y \in \mathcal{Y} \\ z \in \mathcal{Z}}} p(z, y) \log_2 \frac{p(z, y)}{p_1(z) p_2(y)} \quad (4)$$

where $p(z, y)$ is the joint probability distribution and $p_1(z)$ and $p_2(y)$ are the marginal probability distribution functions of \mathcal{Z} and \mathcal{Y} . With the base-2 logarithm the reduction in uncertainty is expressed in bits. We use the MI between the classifiers predictions and the ground truth as a second performance measure. The joint and marginal probabilities in (4) are estimated by their relative frequency of occurrence in the confusion matrix.

We calculate the third measure ITR R , in bits per minute, based on the MI (4), and the median² inter-trial interval $\text{med}(\Delta t)$:

$$R = 60 \frac{I}{\text{med}(\Delta t)}. \quad (5)$$

As a fourth, and last performance measure, we use the area under the curve (AUC) of the receiver operating characteristic (ROC) [21] to express the ranking performance of the classifier. The AUC is equal to the probability that a randomly chosen instance of the first class is ranked above an randomly chosen instance of the second class; i.e. an AUC of 0.5 indicates random performance, and an AUC of 0 or 1 indicates perfect ranking ability. Like the MI, the AUC does not assume equal prior probabilities.

Originally we planned to use the Kullback-Leibler divergence (KLD) as a measure of change in the feature distributions as in [4], but the assumption that the features are

²We use the median instead of the mean because Δt resembles a Poisson distribution.

normally distributed was violated heavily by both our ERD features (even after log-transforming) and our LRP features. This made the estimation of the KLD infeasible due to the need for high-dimensional density estimation.

E. Loss of control analysis

To analyze the influence of user frustration on the performance of a BCI, we trained a BCI on normal blocks, and measured the difference in performance of the same classifier between unseen, normal blocks and unseen blocks from the LOC condition. Because we could not assume that the distribution of the EEG signal would remain stationary, training and evaluating a BCI on samples uniformly spread over the sessions might lead to an overestimation of the performance. In order to have a more reliable measure of how an online BCI would perform, we therefore used a special evaluation scheme where complete experimental blocks were left out for evaluation: the session consisted of series permutations of three experimental blocks; two normal blocks, and a block with LOC simulation. For every three blocks, we added the first normal block to the training set for the BCI classifier. The remaining normal and LOC block were used for evaluation. This way, the training data was spread over time, but we still had independent blocks for evaluation. Note that this evaluation was not symmetrical, since the classifiers were trained on normal blocks but tested on blocks of both the normal and LOC condition.

If there were any difference between the normal and LOC conditions, we expected to find a lower performance on LOC blocks compared to blocks with normal control, since the model was optimized for a different distribution than the observations it was evaluated on had.

F. Confounding factors

To induce mental state change associated with LOC, we intentionally degraded the quality of control. Behavioural changes (e.g. repetitive and force-full keystrokes, and more frequent gazing at the hands) might occur as a result of the method of induction. Therefore, potential differences in BCI performance might be caused solely by these changes in behaviour. In this context, behavioural changes are confounding factors, and need to be corrected for.

However, we cannot discern behavior caused by the method of induction, and behaviour caused by induced changes in the mental state. Correcting for confounding factors could therefore reduce the variation in mental state, and lead to an underestimation of the effect. Therefore, we performed our analysis both with and without correction for confounding behaviour.

Behavioural changes that we identify as confounding factors are the inter-trial interval (ITI), the repetition of key strokes with the same hand, the fraction of key strokes per hand, the force used to press a key, and eye movements. The ITI can be confounding because the EEG is analyzed over a short period of time; key strokes that follow each other quickly could lead to masking of relevant EEG features, or worse, to the leaking of label information from one key stroke to

the next. Repetition of strokes with the same hand might lead to increased performance for the same reason. Force is a confounding factor because force has an influence on the ERD [22]. Artifacts related to eye movements are known to have a profound influence EEG analyses, but these were already removed by the EOG regression method during preprocessing.

To correct for the confounding factors, we used multi-variate frequency matching. Frequency matching involves stratifying the distribution of the confounding variable, and drawing samples such that the number of samples within each stratum is the same per condition [23]. In our case, the multivariate distributions of the previously described confounding variables in the normal and LOC conditions were matched.

These confounding factors were quantified as follows. To quantify the ITI, we used the logarithm of the difference in seconds between consecutive trials. The key stroke patterns were modeled with a discrete bivariate distribution of the label of the current and previous trial. For force, a bivariate distribution of the log-transformed electromyography (EMG) power was used. To calculate EMG power, the procedure outlined in [24] was used: 1) apply a high-pass filter with a cut-off of 30 Hz, 2) apply the Hilbert transform to extract the envelope of the signal and 3) apply a low-pass filter with a cut-off of 40 Hz to smooth the signal.

A multi-dimensional histogram with regularly spaced bins was used to extract strata for frequency matching: 4 bins were used for log ITI, 2×2 bins were used for label patterns, and 5×5 bins were used for log EMG power for index fingers.

G. Statistical tests

Comparisons over subjects were performed using Wilcoxon signed-rank tests, on paired values for both conditions per each subject. This test is a non-parametric alternative to the commonly used paired Student's t-test, which could not be applied because the t-test's assumptions that the measurements are normally distributed and have equal variances do not hold for classification performance [25]. We used a significance level $\alpha = 0.05$ for all tests presented in this article.

The over-subjects analysis tests for group effects. In addition, we performed a more sensitive meta analysis that combines the within-subject p -values to test for individual differences. It combines the p -values of different subjects to reject the combined null hypothesis H_0 , that states that each of the individual null hypotheses is true. The combined alternative, H_A , is that at least one is not true. For this purpose, Fisher's method was suggested in [26] for combining p -values:

$$X^2 = -2 \sum_{i=1}^k \log_e p_i, \quad (6)$$

where p_i is the p -value for subject i . When the null hypotheses are all true, and the p_i 's are independent, X^2 follows a χ^2 distribution with $2k$ degrees of freedom.

III. RESULTS

A. Subjects

Twelve healthy users (age 27 ± 3.9) participated in the experiment. All participants had normal, or corrected to normal

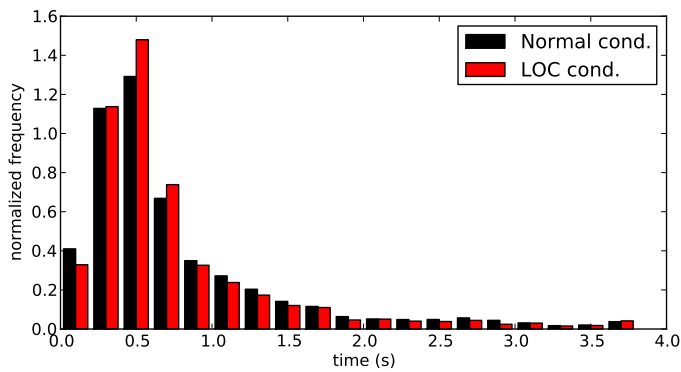


Fig. 2. The histogram for the time between key presses during game play for all subjects. The intervals for the normal condition are displayed in black, the LOC condition is displayed in red. The histogram is dominated by short Δt 's between key presses. The histogram of the LOC condition seems to be slightly more pinched around half a second.

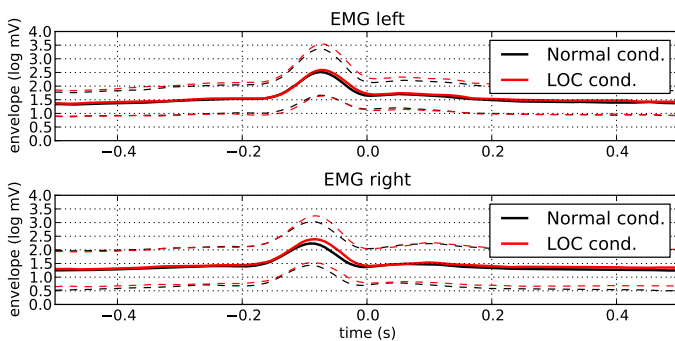


Fig. 3. The mean log EMG power and its standard deviation (dashed lines) is displayed for both the normal and LOC condition, time-locked to the key press at $t = 0$. The top plot shows the EMG power of the left bipolar channel for left index-finger presses, the bottom plot show the EMG power for right hand movement measured with the right bipolar channel.

vision, and reported not to use medication. Only three of our subjects were female, and all subjects were right-handed. Most participants had some video game experience, and four subjects had previous experience with BCIs.

B. Self assessments

To verify the induction of changes in the mental state, we analysed the self reported emotional ratings of the SAM. Most subjects rated the LOC condition more negatively than the normal condition, over subjects this difference was significant ($T=3$, $p<0.01$). While we expected to find a trend towards more arousal in the LOC condition, there was no significant difference ($T=23$, $p=0.26$). The dominance dimension, which measures the amount of dominance, or control they have on their environment, indicate that people seemed to be significantly ($T=3.5$, $p<0.01$) more in control in the normal condition.

C. Confounding behavioural differences

In this section we describe the analysis of the characteristics of the user's behaviour, as it might have a confounding influence on the BCI performance. Both differences in the ITI, and the pattern of consecutive key strokes can indicate

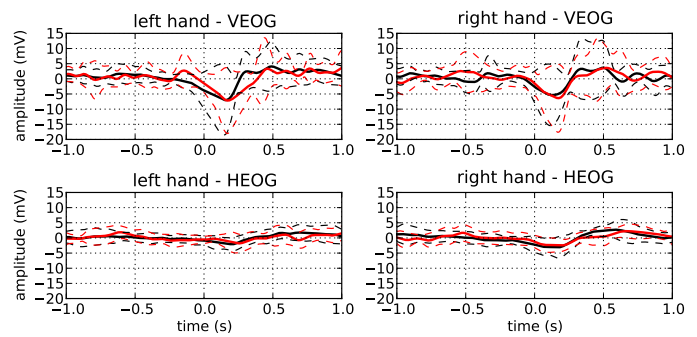


Fig. 4. The vertical eye movement (first row) show downward eye-gaze just after a key stroke at $t = 0$. The horizontal bipolar EOG channel is rather uneventful, only for right hand movement (last column) does there appear to be a delayed reaction to the key press, first negative (looking left) then positive (looking right). The dashed lines indicate the standard deviation. There is no significant difference (16 point Bonferroni corrected Wilcoxon signed-rank test over subjects) between the normal (black) and LOC condition (red).

a confounding behavioural change. The per-subject statistics for these confounding factors are presented in Table I. For the log ITI, we see an insignificant tendency to shorter intervals between key presses in the LOC condition. The probability that a key press is made with the same hand is significantly higher in the LOC condition. This can be caused by increased repetition, by increased imbalance of the class ratios, or a combination thereof. Nevertheless, it indicates a significant behavioural change.

The temporal development of the EMG signal is displayed in Fig. 3. An increase in the EMG power is visible just before the stroke is registered, and a much weaker increase is visible when the key is released. Most of the activity is registered in the interval $[-0.2, 0]$ s relative to the registration of the key press. We used the maximum EMG power in this interval to estimate the force used to press a key (see Table I). Movements with the right index finger produce significantly more EMG power in the LOC condition.

Although we remove (most) of the influence of the EOG signal from the EEG, it is interesting to look at the user's gaze and blink behaviour during a key press (Fig. 4). We can see that users tend to look at their hands 200 ms after a key press, which is most visible in the vertical EOG, and at 300 ms, the variability of the vertical EOG signal seems to increase. This might be caused by eye-blinks, or an adjustment to the new movement direction of the avatar in the game.

In summary, our behaviour analysis has shown that the normal and LOC-conditions are very similar in the timing, the predictability of the key strokes, the amount of force used to press the keys, and in eye movements. However, there is a small but significant increase in repetition of the same movement, and a small significant increase in the force used with the right hand. After balancing the confounding variables and their interactions, on average 25% of the original trials were removed.

D. Impact of LOC on the BCI

To investigate the influence of LOC on the BCI performance, we trained the ERD based and the ERP-based classifiers on blocks from the normal condition, and compared the

TABLE I

STATISTICS OF CONFOUNDING VARIABLES. THE REPETITIVENESS (SECOND COLUMN) AND LOG-EMG POWER (LAST COLUMN) DIFFER SIGNIFICANTLY BETWEEN THE CONDITIONS. THE EMG POWER IS THE MAXIMUM POWER IN THE INTERVAL $[-0.2, 0]$ S. THE START AND END OF THE ARROW SIGNIFY THE MEAN VALUE IN THE NORMAL AND LOC CONDITION RESPECTIVELY.

	$\log \Delta t$	$p(y_t = y_{t-1})$	log pow. EMG L	log pow. EMG R
S0	-0.48 \rightarrow -0.36	0.51 \rightarrow 0.59	1.31 \rightarrow 1.21	1.08 \rightarrow 1.15
S1	-0.51 \rightarrow -0.53	0.54 \rightarrow 0.59	2.70 \rightarrow 2.60	2.04 \rightarrow 2.17
S2	-0.54 \rightarrow -0.59	0.50 \rightarrow 0.52	2.65 \rightarrow 2.58	2.51 \rightarrow 2.46
S3	-0.47 \rightarrow -0.61	0.56 \rightarrow 0.54	1.99 \rightarrow 1.88	2.15 \rightarrow 2.31
S4	-0.46 \rightarrow -0.55	0.53 \rightarrow 0.58	3.14 \rightarrow 3.35	1.76 \rightarrow 1.86
S5	-0.46 \rightarrow -0.44	0.56 \rightarrow 0.60	2.70 \rightarrow 2.56	4.06 \rightarrow 3.91
S6	-0.46 \rightarrow -0.48	0.55 \rightarrow 0.55	2.11 \rightarrow 2.55	2.14 \rightarrow 2.47
S7	-0.40 \rightarrow -0.55	0.58 \rightarrow 0.56	1.70 \rightarrow 1.71	2.13 \rightarrow 2.14
S8	-0.37 \rightarrow -0.44	0.47 \rightarrow 0.56	2.89 \rightarrow 2.95	2.94 \rightarrow 3.02
S9	-0.58 \rightarrow -0.55	0.45 \rightarrow 0.54	3.02 \rightarrow 3.08	2.01 \rightarrow 2.06
S10	-0.58 \rightarrow -0.44	0.47 \rightarrow 0.53	2.43 \rightarrow 2.42	2.52 \rightarrow 2.55
S11	-0.45 \rightarrow -0.42	0.52 \rightarrow 0.58	3.01 \rightarrow 2.86	2.97 \rightarrow 3.13
mean	-0.48 \rightarrow -0.50	0.52 \rightarrow 0.56	2.47 \rightarrow 2.48	2.36 \rightarrow 2.44
Wilc.	T=32, p=0.583	T=6, p=0.010	T=31, p=0.530	T=12, p=0.034

performance on unseen normal blocks with the performance on blocks from the LOC condition. Please refer to Section II-E for more information on this procedure.

The performance of the CSP based features classifier on the normal and LOC blocks without correction for confounds is displayed in Table II. The single trial detection accuracy may seem rather low (60%), but this is similar to the accuracies obtained in other studies that use short ITIs, such as [7]. This was also reflected in the mean ITR of 5.5 bits per minute, which is comparable to the ITRs obtained by naive users with motor-imagery based ERD BCIs. Despite this low recognition rate, the ERD BCIs performance did significantly *increase* in the LOC condition for the AUC and MI measures.

When correction for confounding factors was performed, the results were different (Table III); the over-subject differences disappeared, but there were more significant within-subject differences in sometimes opposing directions. Combined with Fisher's method, the one-sided p -values for a within-subject increase in performance was significant for both AUC and ITR. This indicates that at least one individual increase in performance was significant at the $\alpha = 0.05$ level.

The spatial distribution of the movement related ERD is shown in Fig. 5. Subjects S0, S1, S5, S9 and S10 do display the prototypical ERD on the motor cortices. Remarkably, these activations are more pronounced in the LOC condition (second row), which supports the observed increase in performance. Note that the CSP classification is based on covariance of the EEG channels, while in this figure only the variance is shown.

In contrast to the ERD based classifiers, the LRP classifiers had a constant high performance with a minimum ITR of 11.6 bits per minute. Furthermore, they did not seem to behave differently in the LOC condition, not with (Table V) and not without correction for confounding factors (Table IV). Visual inspection of the classifier's weights confirmed that the most discriminative features were located on the motor cortices, i.e. the BCI was based on brain activity. The increase in performance for S8 was probably a false positive, since the combination over subjects with Fisher's method was not significant.

IV. DISCUSSION

We successfully induced a negative mental state with lack of feeling of control in the LOC condition, as shown by the significant differences in the SAM ratings. The performance of our ERD classifier improved significantly in the LOC condition. However, we did find small differences in the statistics of known confounding behaviour. After correcting for these factors, there were still differences in the performance of the CSP-based BCI. This difference in performance was probably caused by a change in feature distribution associated with the change in the mental state of the subject.

The correction for confounding statistics could have reduced the contrast between the experimental conditions, since changes in behaviour and emotional state are interrelated and the behavioural changes were removed. This might explain why the over the over-subjects effect disappeared. After correction, more within-subject differences became significant, sometimes in opposing directions. While correcting for confounding factors was necessary to exclude the possibility that the effects depend solely on the LOC induction method, the results without correction are more ecologically valid, since increased effort and repeated attempts to perform the same action are to be expected when the BCI fails.

It is surprising that the same ERD classifier predicts the labels of data from a different distribution significantly better than it predicts labels of the data with the distribution it was trained on. The significant differences in the AUC values indicates that the improvement in the LOC condition cannot be simply explained by a bias shift, since that would not account for improved ranking of trials. This suggests that the ERD feature distributions of the two classes move away from the classifier's hyperplane in the LOC condition.

Our findings for the ERD classifiers contradict the findings presented in [7], where a *decrease* in performance for the LOC condition was reported. On ERP classification the studies do agree; [7] reports no significant difference, and also in our study no significant effect was found for LRP classification. The differences between our study and [7] for ERD classification could be caused by several differences in the experimental

TABLE II

THE INFLUENCE OF LOC ON A CSP CLASSIFIER IS SHOWN BELOW, WITHOUT CORRECTION FOR CONFOUNDING FACTORS. THE START AND END OF THE ARROW INDICATE THE MEDIAN PERFORMANCE FOR THE NORMAL AND LOC CONDITION RESPECTIVELY. THE p -VALUE OF A MANN-WHITNEY U TEST ON THE PER-BLOCK PERFORMANCE IS DISPLAYED ABOVE THE ARROW. THE ROW DENOTED WITH "WILC." SIGNIFIES THE OVER-SUBJECT COMPARISON WITH THE WILCOXON SIGNED RANK TEST. THE ROW DENOTED BY "FISH." PRESENTS THE RESULTS OF COMBINING ONE-SIDED p -VALUES FOR AN INCREASE IN PERFORMANCE.

	Accuracy	AUC	MI	ITR
S0	0.679 $\xrightarrow{p=0.35}$ 0.736	0.759 $\xrightarrow{p=0.35}$ 0.843	0.110 $\xrightarrow{p=0.35}$ 0.193	11.680 $\xrightarrow{p=0.48}$ 13.763
S1	0.599 $\xrightarrow{p=0.48}$ 0.618	0.616 $\xrightarrow{p=0.64}$ 0.652	0.021 $\xrightarrow{p=0.64}$ 0.041	2.601 $\xrightarrow{p=0.64}$ 4.525
S2	0.619 $\xrightarrow{p=0.92}$ 0.555	0.543 $\xrightarrow{p=0.62}$ 0.578	0.001 $\xrightarrow{p=0.77}$ 0.000	0.088 $\xrightarrow{p=0.92}$ 0.025
S3	0.474 $\xrightarrow{p=0.13}$ 0.519	0.466 $\xrightarrow{p=0.05}$ 0.518	0.004 $\xrightarrow{p=0.62}$ 0.003	0.452 $\xrightarrow{p=0.77}$ 0.338
S4	0.519 $\xrightarrow{p=0.92}$ 0.507	0.549 $\xrightarrow{p=0.92}$ 0.526	0.002 $\xrightarrow{p=0.77}$ 0.003	0.224 $\xrightarrow{p=0.77}$ 0.351
S5	0.741 $\xrightarrow{p=0.92}$ 0.750	0.828 $\xrightarrow{p=0.92}$ 0.832	0.167 $\xrightarrow{p=0.92}$ 0.188	15.616 $\xrightarrow{p=0.92}$ 20.297
S6	0.538 $\xrightarrow{p=0.65}$ 0.529	0.543 $\xrightarrow{p=0.86}$ 0.522	0.003 $\xrightarrow{p=0.59}$ 0.006	0.302 $\xrightarrow{p=0.59}$ 0.664
S7	0.544 $\xrightarrow{p=0.10}$ 0.600	0.565 $\xrightarrow{p=0.27}$ 0.657	0.005 $\xrightarrow{p=0.19}$ 0.027	0.532 $\xrightarrow{p=0.08}$ 3.082
S8	0.542 $\xrightarrow{p=0.34}$ 0.581	0.556 $\xrightarrow{p=0.34}$ 0.589	0.004 $\xrightarrow{p=0.48}$ 0.010	0.366 $\xrightarrow{p=0.48}$ 1.035
S9	0.735 $\xrightarrow{p=0.15}$ 0.768	0.797 $\xrightarrow{p=0.15}$ 0.839	0.160 $\xrightarrow{p=0.15}$ 0.218	18.879 $\xrightarrow{p=0.15}$ 24.494
S10	0.612 $\xrightarrow{p=0.35}$ 0.582	0.670 $\xrightarrow{p=0.48}$ 0.651	0.046 $\xrightarrow{p=0.35}$ 0.018	5.611 $\xrightarrow{p=0.35}$ 2.108
S11	0.608 $\xrightarrow{p=0.34}$ 0.553	0.594 $\xrightarrow{p=0.95}$ 0.599	0.020 $\xrightarrow{p=0.95}$ 0.023	2.371 $\xrightarrow{p=0.64}$ 2.059
mean	0.601 \rightarrow 0.608	0.624 \rightarrow 0.651	0.045 \rightarrow 0.061	4.893 \rightarrow 6.062
Wilc.	T=31.0, p=0.530	T=12.0, p=0.034	T=14.0, p=0.050	T=17.0, p=0.084
Fish.	p=0.123	p=0.078	p=0.259	p=0.222

TABLE III

THE INFLUENCE OF LOC ON A CSP CLASSIFIER IS SHOWN BELOW, WITH CORRECTION FOR CONFOUNDING FACTORS ENABLED. PLEASE REFER TO TABLE II FOR AN EXPLANATION.

	Accuracy	AUC	MI	ITR
S0	0.678 $\xrightarrow{p=0.64}$ 0.667	0.718 $\xrightarrow{p=0.35}$ 0.785	0.094 $\xrightarrow{p=0.82}$ 0.056	7.370 $\xrightarrow{p=0.95}$ 4.896
S1	0.543 $\xrightarrow{p=0.05}$ 0.646	0.572 $\xrightarrow{p=0.09}$ 0.686	0.007 $\xrightarrow{p=0.05}$ 0.063	0.734 $\xrightarrow{p=0.09}$ 6.751
S2	0.640 $\xrightarrow{p=1.00}$ 0.626	0.558 $\xrightarrow{p=0.92}$ 0.548	0.009 $\xrightarrow{p=0.19}$ 0.002	0.825 $\xrightarrow{p=0.27}$ 0.165
S3	0.541 $\xrightarrow{p=0.62}$ 0.553	0.483 $\xrightarrow{p=0.62}$ 0.514	0.007 $\xrightarrow{p=0.77}$ 0.005	0.482 $\xrightarrow{p=0.62}$ 0.450
S4	0.569 $\xrightarrow{p=0.19}$ 0.532	0.572 $\xrightarrow{p=0.77}$ 0.536	0.013 $\xrightarrow{p=0.92}$ 0.008	1.416 $\xrightarrow{p=0.92}$ 0.769
S5	0.753 $\xrightarrow{p=0.77}$ 0.743	0.857 $\xrightarrow{p=0.62}$ 0.821	0.198 $\xrightarrow{p=0.77}$ 0.180	14.086 $\xrightarrow{p=0.77}$ 15.423
S6	0.532 $\xrightarrow{p=0.21}$ 0.558	0.498 $\xrightarrow{p=0.01}$ 0.579	0.004 $\xrightarrow{p=0.10}$ 0.011	0.308 $\xrightarrow{p=0.06}$ 0.874
S7	0.549 $\xrightarrow{p=0.37}$ 0.577	0.613 $\xrightarrow{p=0.92}$ 0.623	0.010 $\xrightarrow{p=0.49}$ 0.015	0.964 $\xrightarrow{p=0.37}$ 1.534
S8	0.543 $\xrightarrow{p=0.48}$ 0.605	0.549 $\xrightarrow{p=0.23}$ 0.583	0.002 $\xrightarrow{p=0.02}$ 0.017	0.166 $\xrightarrow{p=0.02}$ 1.394
S9	0.687 $\xrightarrow{p=0.23}$ 0.740	0.783 $\xrightarrow{p=0.23}$ 0.833	0.086 $\xrightarrow{p=0.15}$ 0.162	8.770 $\xrightarrow{p=0.15}$ 16.045
S10	0.658 $\xrightarrow{p=0.05}$ 0.617	0.676 $\xrightarrow{p=0.82}$ 0.673	0.054 $\xrightarrow{p=0.24}$ 0.035	4.888 $\xrightarrow{p=0.35}$ 3.147
S11	0.585 $\xrightarrow{p=0.23}$ 0.551	0.615 $\xrightarrow{p=0.01}$ 0.544	0.021 $\xrightarrow{p=0.23}$ 0.005	1.329 $\xrightarrow{p=0.15}$ 0.366
mean	0.606 \rightarrow 0.618	0.625 \rightarrow 0.644	0.042 \rightarrow 0.047	3.445 \rightarrow 4.318
Wilc.	T=31.0, p=0.530	T=27.0, p=0.347	T=35.0, p=0.754	T=35.0, p=0.754
Fish.	p=0.237	p=0.050	p=0.063	p=0.047

setup: a) our ITI is about half the ITI used in [7], b) we did gather training and evaluation data in the same environment with the same user tasks, c) our subjects were not informed that the system would respond incorrectly, and d) unlike [7], we show that the performance differences between the two conditions were not (fully) caused by behavioural differences. But both our study and [7] do agree that LOC has a profound influence on the ERD features.

V. CONCLUSIONS AND FUTURE WORK

In this experiment, we simulated a non-responding brain-computer interface (BCI) controller to study whether asso-

ciated changes in the users mental state have an influence on the BCI performance. The self-reported emotional ratings confirmed that the loss of control (LOC) condition induced a more negative, and less dominant mental state. These different mental states were accompanied by minor behavioural changes for which we corrected. Contrary to our expectations, we observed a significant performance *increase* during the LOC condition for the event-related desynchronization (ERD) based BCIs. For the event-related potential (ERP) based BCIs, we found no change in performance. The image that we sketched in the introduction of a BCI spiralling completely out of control appears to be an illusion. However, the difference

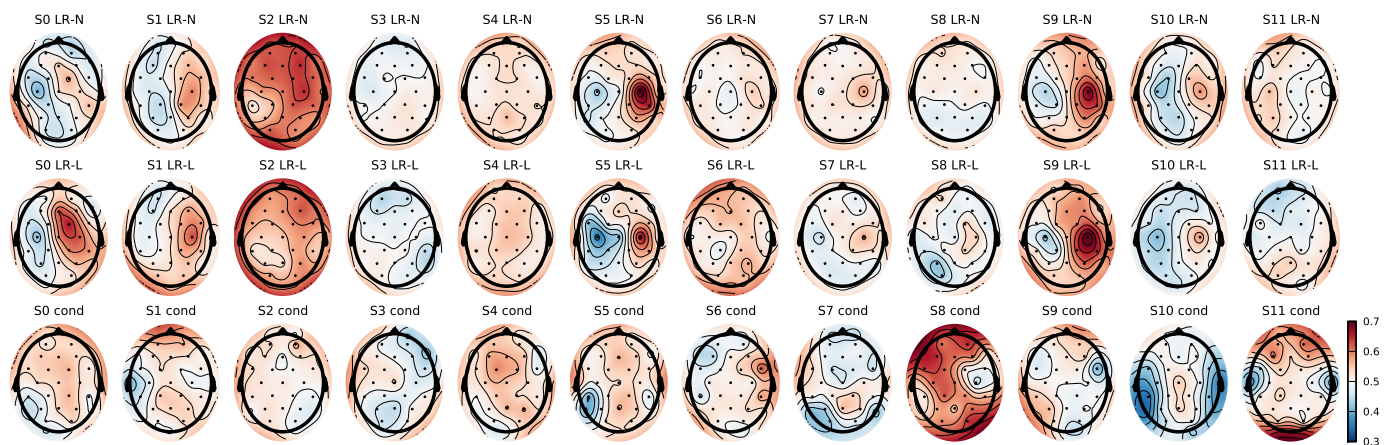


Fig. 5. These scalp plots display the difference between left and right hand movement in the normal (first row) and LOC condition (second row), and the difference between the normal and LOC in the last row. The color encodes the AUC-ROC ranking performance of the 8–30 Hz band power at the specified location; red indicates a positive rank correlation with the target class (right hand for the first two rows, or LOC in the last), blue a negative correlation. The conditions were corrected for confounding factors with frequency matching. Most subjects display a more pronounced spatial activation in the LOC condition.

TABLE IV

THE INFLUENCE OF LOC ON A ERP CLASSIFIER IS SHOWN BELOW, WITHOUT CORRECTION FOR CONFOUNDING FACTORS. PLEASE REFER TO TABLE II FOR AN EXPLANATION.

	Accuracy		AUC		MI		ITR	
S0	0.765	$\xrightarrow{p=0.73}$ 0.747	0.824	$\xrightarrow{p=0.95}$ 0.812	0.223	$\xrightarrow{p=0.64}$ 0.179	24.349	$\xrightarrow{p=0.48}$ 17.575
S1	0.802	$\xrightarrow{p=0.34}$ 0.758	0.866	$\xrightarrow{p=0.64}$ 0.823	0.282	$\xrightarrow{p=0.23}$ 0.195	32.838	$\xrightarrow{p=0.34}$ 22.092
S2	0.726	$\xrightarrow{p=0.27}$ 0.750	0.790	$\xrightarrow{p=0.49}$ 0.809	0.142	$\xrightarrow{p=0.62}$ 0.154	16.269	$\xrightarrow{p=0.37}$ 18.677
S3	0.711	$\xrightarrow{p=0.77}$ 0.724	0.779	$\xrightarrow{p=0.62}$ 0.788	0.133	$\xrightarrow{p=0.77}$ 0.139	11.719	$\xrightarrow{p=0.77}$ 16.195
S4	0.692	$\xrightarrow{p=0.92}$ 0.714	0.767	$\xrightarrow{p=0.37}$ 0.803	0.096	$\xrightarrow{p=0.92}$ 0.110	11.672	$\xrightarrow{p=0.77}$ 11.882
S5	0.720	$\xrightarrow{p=0.77}$ 0.715	0.800	$\xrightarrow{p=0.92}$ 0.790	0.142	$\xrightarrow{p=0.92}$ 0.136	14.584	$\xrightarrow{p=0.62}$ 15.041
S6	0.703	$\xrightarrow{p=0.15}$ 0.726	0.766	$\xrightarrow{p=0.15}$ 0.789	0.116	$\xrightarrow{p=0.10}$ 0.155	13.460	$\xrightarrow{p=0.10}$ 16.812
S7	0.772	$\xrightarrow{p=0.77}$ 0.778	0.833	$\xrightarrow{p=0.62}$ 0.857	0.225	$\xrightarrow{p=0.77}$ 0.234	23.515	$\xrightarrow{p=0.62}$ 26.844
S8	0.702	$\xrightarrow{p=0.64}$ 0.742	0.778	$\xrightarrow{p=0.05}$ 0.830	0.121	$\xrightarrow{p=0.64}$ 0.175	10.529	$\xrightarrow{p=0.23}$ 18.768
S9	0.831	$\xrightarrow{p=0.48}$ 0.785	0.886	$\xrightarrow{p=0.34}$ 0.873	0.328	$\xrightarrow{p=0.48}$ 0.246	38.768	$\xrightarrow{p=0.34}$ 29.488
S10	0.704	$\xrightarrow{p=0.82}$ 0.701	0.790	$\xrightarrow{p=0.95}$ 0.783	0.129	$\xrightarrow{p=0.82}$ 0.120	14.146	$\xrightarrow{p=0.95}$ 14.663
S11	0.835	$\xrightarrow{p=0.81}$ 0.843	0.931	$\xrightarrow{p=0.81}$ 0.921	0.344	$\xrightarrow{p=0.64}$ 0.375	41.528	$\xrightarrow{p=0.81}$ 38.285
mean	0.747 \rightarrow 0.749		0.818 \rightarrow 0.823		0.190 \rightarrow 0.185		21.115 \rightarrow 20.5	
Wilc.	T=32.0, p=0.583		T=30.0, p=0.480		T=36.0, p=0.814		T=37.0, p=0.875	
Fish.	p=0.546		p=0.197		p=0.570		p=0.350	

in performance demonstrates that variabilities in the feature distributions related to LOC do in fact exist, and could be more dire under different circumstances.

For future work, a logical next step would be the investigation of the origin of the increase of performance for ERD classifiers. We suspect that it might be related to a shift in attention from the game context during normal play to the movement of the hands in the LOC condition. Since the strength of the beta band event-related synchronization (ERS) is related to attention in constant isometric force motor tasks [27], an increase of attention on the motor task in the LOC condition could result in more pronounced beta ERD/ERS, and indirectly in better classification results. This would form an interesting hypothesis for a follow-up study. Related is also the study presented in [28] that shows a pronounced beta rebound when the observed movement does not match the movement the user was supposed to execute.

The recordings from our current experiments could also be analyzed for correlates with emotions, as we already have self-reported emotions for every two-minute block in the experiment. The first steps in this direction have been taken in [29]. The recognition of emotions from electroencephalography (EEG) would be immensely valuable to both locked-in patients — who would otherwise have to verbalise their mood using other means, such as the P300 speller — and healthy users.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the BrainGain Smart Mix Programme of the Netherlands Ministry of Economic Affairs and the Netherlands Ministry of Education, Culture and Science.

TABLE V

THE INFLUENCE OF LOC ON A ERP CLASSIFIER IS SHOWN BELOW, WITH CORRECTION FOR CONFOUNDING FACTORS ENABLED. PLEASE REFER TO TABLE II FOR AN EXPLANATION.

	Accuracy	AUC	MI	ITR
S0	0.710 $\xrightarrow{p=0.35}$ 0.732	0.800 $\xrightarrow{p=0.64}$ 0.820	0.134 $\xrightarrow{p=0.48}$ 0.164	11.145 $\xrightarrow{p=0.95}$ 9.779
S1	0.810 $\xrightarrow{p=0.15}$ 0.777	0.870 $\xrightarrow{p=0.48}$ 0.854	0.295 $\xrightarrow{p=0.15}$ 0.233	28.990 $\xrightarrow{p=0.23}$ 19.258
S2	0.720 $\xrightarrow{p=0.92}$ 0.746	0.792 $\xrightarrow{p=0.77}$ 0.801	0.121 $\xrightarrow{p=0.62}$ 0.127	12.031 $\xrightarrow{p=0.77}$ 12.806
S3	0.716 $\xrightarrow{p=0.77}$ 0.702	0.763 $\xrightarrow{p=0.62}$ 0.766	0.133 $\xrightarrow{p=0.92}$ 0.134	9.135 $\xrightarrow{p=0.27}$ 11.983
S4	0.696 $\xrightarrow{p=0.62}$ 0.707	0.759 $\xrightarrow{p=0.77}$ 0.787	0.110 $\xrightarrow{p=0.77}$ 0.133	12.860 $\xrightarrow{p=0.37}$ 13.445
S5	0.640 $\xrightarrow{p=0.77}$ 0.649	0.729 $\xrightarrow{p=0.92}$ 0.735	0.094 $\xrightarrow{p=0.27}$ 0.071	7.364 $\xrightarrow{p=0.92}$ 6.853
S6	0.671 $\xrightarrow{p=0.47}$ 0.695	0.769 $\xrightarrow{p=0.86}$ 0.760	0.086 $\xrightarrow{p=0.47}$ 0.124	6.197 $\xrightarrow{p=0.72}$ 7.992
S7	0.739 $\xrightarrow{p=0.92}$ 0.733	0.824 $\xrightarrow{p=0.77}$ 0.811	0.174 $\xrightarrow{p=0.92}$ 0.164	15.149 $\xrightarrow{p=0.92}$ 15.587
S8	0.709 $\xrightarrow{p=0.05}$ 0.799	0.777 $\xrightarrow{p=0.02}$ 0.847	0.106 $\xrightarrow{p=0.05}$ 0.265	9.679 $\xrightarrow{p=0.02}$ 21.099
S9	0.845 $\xrightarrow{p=0.34}$ 0.803	0.925 $\xrightarrow{p=0.23}$ 0.874	0.379 $\xrightarrow{p=0.48}$ 0.278	35.293 $\xrightarrow{p=0.34}$ 29.589
S10	0.671 $\xrightarrow{p=0.82}$ 0.688	0.764 $\xrightarrow{p=0.95}$ 0.765	0.089 $\xrightarrow{p=0.82}$ 0.099	7.261 $\xrightarrow{p=0.35}$ 9.269
S11	0.835 $\xrightarrow{p=0.81}$ 0.822	0.921 $\xrightarrow{p=0.34}$ 0.883	0.350 $\xrightarrow{p=0.81}$ 0.331	23.343 $\xrightarrow{p=0.64}$ 26.231
mean	0.730 \rightarrow 0.738	0.808 \rightarrow 0.809	0.173 \rightarrow 0.177	14.871 \rightarrow 15.3
Wilc.	T=31.0, p=0.530	T=38.0, p=0.937	T=36.0, p=0.814	T=28.0, p=0.388
Fish.	p=0.392	p=0.484	p=0.456	p=0.162

REFERENCES

- [1] A. Nijholt, D. Plass-Oude Bos, and B. Reuderink, "Turning shortcomings into challenges: Brain-computer interfaces for games," *Entertainment Computing*, vol. 1, no. 2, pp. 85–94, October 2009.
- [2] B. Reuderink, "Games and brain-computer interfaces: The state of the art," Human Media Interaction, Faculty of EEMCS, University of Twente, Tech. Rep. TR-CTIT-08-81, September 2008.
- [3] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.
- [4] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *Journal of Neural Engineering*, vol. 3, no. 1, pp. R13–R23, 2006.
- [5] B. Blankertz, M. Kawanabe, R. Tomioka, F. U. Hohlefeld, V. Nikullin, and K.-R. Müller, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 20, 2007, pp. 113–120.
- [6] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for brain-computer interfacing," *PLoS ONE*, vol. 3, p. e2967, 2008.
- [7] S. Jatzew, T. O. Zander, M. D. Filippis, C. Kothe, S. Welke, and M. Rötting, "Examining causes for nonstationarities: the loss of controllability is a factor which induces nonstationarities," in *Proceedings of the 4th International BCI Workshop & Training Course*. Verlag der Technischen Universität Graz, 2008, pp. 138–143.
- [8] T. O. Zander, C. Kothe, S. Welke, and M. Rötting, "Utilizing secondary input from passive brain-computer interfaces for enhancing human-machine interaction," in *Lecture Notes in Computer Science*, vol. 5638/2009, 2009, pp. 759–771.
- [9] B. Reuderink, A. Nijholt, and M. Poel, "Affective Pacman: A frustrating game for brain-computer interface experiments," in *Proceedings of the 3rd International Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2009)*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 9, May 2009, pp. 221–227.
- [10] J. Scheirer, R. Fernandez, J. Klein, and R. W. Picard, "Frustrating the user on purpose: A step toward building an affective computer," *Interacting with Computers*, vol. 14, pp. 93–118, February 2002.
- [11] J. Klein, Y. Moon, and R. W. Picard, "This computer responds to user frustration: Theory, design, and results," *Interacting with Computers*, vol. 14, pp. 119–140, 2002.
- [12] H. Diener and K. Oertel, "Experimental approach to affective interaction in games," in *Edutainment*, vol. 3942/2006, 2006, pp. 507–518.
- [13] M. M. Bradley and P. J. Lang, "Measuring emotion: The self-assessment manikin and the semantic differential," *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 25, no. 1, pp. 49–59, March 1994.
- [14] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical Neurophysiology*, vol. 118, pp. 98–104, 2007.
- [15] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topography*, vol. 12, no. 3, pp. 117–186, October 2000.
- [16] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, pp. 1842–1857, 1999.
- [17] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 6, pp. 440–447, December 1991.
- [18] H. Shibasaki and M. Hallett, "What is the Bereitschaftspotential?" *Clinical Neurophysiology*, vol. 117, no. 11, pp. 2341–2356, 2006.
- [19] B. Blankertz, G. Dornhege, C. Schäfer, R. Krepki, J. Kohlmorgen, K.-R. Müller, V. Kunzmann, F. Losch, and G. Curi, "Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 11, no. 2, pp. 127–131, 2003.
- [20] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components — a tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, May 2011.
- [21] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2005.
- [22] A. Stančák Jr., A. Riml, and G. Pfurtscheller, "The effects of external load on movement-related changes of the sensorimotor EEG rhythms," *Electroencephalography and Clinical Neurophysiology*, vol. 102, pp. 495–504, 1997.
- [23] S. Anderson, A. Auquier, W. W. Hauck, D. Oakes, W. Vandaele, and H. I. Weisberg, *Statistical Methods for Comparative Studies: Techniques for Bias Reduction*, ser. Wiley series in probability and mathematical statistics. New York: John Wiley & Sons, Inc., 1980.
- [24] A. L. Hof, "EMG and muscle force: An introduction," *Human Movement Science*, vol. 3, no. 1–2, pp. 119–153, 1984.
- [25] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, pp. 1–30, December 2006.
- [26] T. M. Loughin, "A systematic comparison of methods for combining p -values from independent tests," *Computational Statistics & Data Analysis*, vol. 47, pp. 467–485, 2004.
- [27] R. Kristeva-Feige, C. Fritsch, J. Timmer, and C.-H. Lücking, "Effects of attention and precision of exerted force on beta range EEG-EMG

synchronization during a maintained motor contraction task,” *Clinical Neurophysiology*, vol. 113, pp. 123–131, 2002.

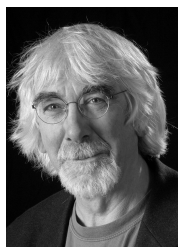
- [28] T. Koelewijn, H. T. van Schie, H. Bekkering, R. Oostenveld, and O. Jensen, “Motor-cortical beta oscillations are modulated by correctness of observed action,” *NeuroImage*, vol. 40, pp. 767–775, 2008.
- [29] B. Reuderink, C. Mühl, and M. Poel, “Valence, arousal and dominance in the EEG during game play,” *International Journal of Autonomous and Adaptive Communication Systems*, 2012, to appear.



Boris Reuderink obtained his master’s degree at the University of Twente in 2007, after working on different machine learning problems. He combines his big interests — brains and intelligence — in his PhD position at the University of Twente, in which he focuses on making brain-computer interfaces work in real-world settings for healthy users.



Mannes Poel is assistant professor in the Human Media Interaction group at the University of Twente. His main research involves applied machine learning for vision based detection and interpretation of human behavior and the analysis and classification EEG based brain signals.



Anton Nijholt received an MSc in Mathematics and Computer Science (1974) at Delft University of Technology and a PhD in Computer Science at the Vrije Universiteit Amsterdam. He has held positions at various universities in Canada, Belgium and the Netherlands. Since 1989, he has been a Full Professor of Computer Science at the University of Twente (Enschede, The Netherlands). He is now the Head of the Human Media Interaction research group at the Department of Computer Science of the University of Twente. His recent interests include entertainment

computing, multimodal interaction, and brain-computer interfacing.