Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring

Klaus-Robert Müller [a,b,*], Michael Tangermann [b], Guido Dornhege [b],
Matthias Krauledat [a], Gabriel Curio [c], Benjamin Blankertz [a,b]

[a] *Technical University Berlin, Str. des 17. Juni 135, 10623 Berlin, Germany*
[b] *Fraunhofer FIRST.IDA, Kekuléstr. 7, 12489 Berlin, Germany*
[c] *Department of Neurology, Campus Benjamin Franklin, Charité University Medicine Berlin, Berlin, Germany*

Abstract

Machine learning methods are an excellent choice for compensating the high variability in EEG when analyzing single-trial data in real-time. This paper briefly reviews preprocessing and classification techniques for efficient EEG-based brain–computer interfacing (BCI) and mental state monitoring applications. More specifically, this paper gives an outline of the Berlin brain–computer interface (BBCI), which can be operated with minimal subject training. Also, spelling with the novel BBCI-based Hex-o-Spell text entry system, which gains communication speeds of 6–8 letters per minute, is discussed. Finally the results of a real-time arousal monitoring experiment are presented.
© 2007 Elsevier B.V. All rights reserved.

Keywords: EEG; Sensorimotor rhythms; α-Rhythm; Single-trial EEG-analysis; Real-time; Machine learning; Mental state monitoring

1. Introduction

Recently, advances in single-trial EEG-analysis have achieved the efficient online differentiation of neuroelectric signals. The present contribution distinguishes between two main application fields of these analysis techniques: (a) brain–computer interfacing and (b) online monitoring of brain states. This paper reviews both along with the machine learning and signal analysis machinery that is necessary for such online EEG processing.

Brain–computer interfaces (BCI) allow for communication that is solely based on brain signals, independent from muscles or peripheral nerves (see Wolpaw et al., 2002; Kübler et al., 2001; Curran and Stokes, 2003; Kübler and Müller, 2007; Dornhege et al., 2007a; Carmena et al., 2003 for a broader overview and background information). The Berlin brain–computer interface (BBCI) is a non-invasive, EEG-based system whose key features are (1) the use of motor imagery for control tasks, (2) advanced

machine learning techniques that automatically extract complex high-dimensional features and classify in a robust manner, and – as a consequence – (3) no need for subject training. The latter characteristic is considered an important contribution since typical BCI systems rely on classical conditioning (cf. Elbert et al., 1980; Rockstroh et al., 1984; Birbaumer et al., 2000) and require extensive subject training of 50–100 h.[1] In contrast, the BBCI approach allows to shift the training effort from the user towards the machine (cf. Section 3). Section 4.1 will demonstrate a BBCI application: the Hex-o-Spell interface for spelling.

Brain–computer interfacing is certainly not the only interesting application when decoding brain activity. General online monitoring of generic brain states beyond voluntarily altered brain activity has in the past been under study, e.g. for the detection of sleep stages, tiredness, arousal, for emotion monitoring and for cognitive workload analysis (Kohlmorgen et al., 2007; Haynes and Rees, 2006; Müller et al., 1995). In Section 4.3 the real-time monitoring of mental states using EEG are discussed

---

* Corresponding author at: Technical University Berlin, Str. des 17. Juni 135, 10623 Berlin, Germany.
   E-mail address: klaus@first.fhg.de (K.-R. Müller).

[1] Consider, that many of these approaches deal with individuals with neurological diseases which affect the brain. This may even lead to slower learning of the subjects or unusual difficulties in the classification of the patient's signals.

and the example of monitoring a subject's arousal to estimate it's concentration ability within an industrial problem setting is briefly outlined.

2. Machine learning for BCI

Since brain data is non-stationary, it offers formidable challenges from the viewpoint of a data analyst. It is characterized by significant trial-to-trial and subject-to-subject variability. Often signals are high-dimensional with only relatively few samples available for fitting models to the data and finally the signal-to-noise ratio is highly unfavorable, In fact, it typically is even ill-defined what signal and, respectively, what noise are (cf. Blankertz et al., 2006c; Dornhege et al., 2007a). Due to this variability, machine learning methods have become the tool of choice for the online analysis of single-trial brain data. In contrast, classical neurophysiological analysis methods apply averaging methods like taking grand averages over trials, subjects and sessions to get rid of various sources of variability. This approach investigates the *average* brain and can answer generic questions of neurophysiological interest, but it is rather blind to the wealth of the dynamics and behavioral variability available only to single-trial analysis methods.

Brain–computer interfacing has been pushing single-trial EEG-analysis to an extreme, as the methods applied in this field need to be performed in real-time without significant processing delays (typically less than 40 ms). In the following the basic machine learning methods for single-trial EEG-analysis are introduced: feature extraction – here an approach using spatial filters is presented – and classification, which is discussed by the example of a framework of regularized linear discriminant analysis (LDA). Furthermore when analyzing high-dimensional data it is not only important to visualize, predict or classify the data with low error, but it is essential that the exploratory data analysis tools allow to *explain* the underlying structure in order to contribute to a better understanding of data. A final paragraph of this section will address this aspect.

2.1. The general machine learning setting

For the training phase of a BCI, $k = 1, \ldots, N$ samples $\mathbf{x}_k$ are measured, where $\mathbf{x}$ is a continuous stream of EEG in the *n*-dimensional sensor space. The subjects are asked to perform individual but fixed mental task, e.g. motor imagery of their left and right hand. This allows to associate the EEG vectors $\mathbf{x}_k$ with the respective class labels for these two mental states $y_k \in \pm 1$. After collecting a sufficient number of trials, the machine learning approach allows to learn the complex unknown mapping f between $\mathbf{x}_k$ and y_k by inferring the typical EEG patterns of left and right motor imagery tasks of a particular subject. While theoretically this abstract mapping f could be learned even from raw EEG data (Blankertz et al., 2002), it has proven to be much more efficient to extract appropriate features from the continuous EEG signal using available physiological a priori knowledge. For the example of motor imagery tasks, known useful discriminative features are the band-power values localized at motor related areas, which can be enhanced in contrast by spatial filtering.[2] Clearly it will depend on the particular paradigmatic setting whether the resulting features can – in a subsequent step – be best classified by linear methods or whether nonlinear kernel-based learning methods (Müller et al., 2001; Schölkopf and Smola, 2002; Vapnik, 1995) are necessary (for a discussion see Müller et al., 2003).

With the help of modern machine learning methods, the BBCI system needs only between 50 and 150 trials, i.e., between 7 and 20 min (depending on the signal-to-noise ratio of the subject's EEG) of calibration recordings for constructing a high performance BCI (Blankertz et al., 2007a). Note that some subjects need no calibration measurement at all, if data from earlier BCI sessions is available (Krauledat et al., 2007). With the BBCI even BCI naive subjects can successfully communicate within the same morning. Our approach follows the motto *let the machines learn* and has no need for explicit subject training. However, note that (as in all other groups worldwide) about one third of the BCI users do not achieve BCI communication at all.

2.2. Spatial filtering with common spatial patterns

A crucial point in the data processing is to extract appropriate spatio-temporal filters that can serve as feature extractors for the subsequent steps. Ideally they optimize the discriminability of the multi-channel brain signals based on some chosen physiological paradigm, for instance event-related (de-)synchronization (ERD/ERS) effects of the (sensory-) motor rhythms (Pfurtscheller and Lopes da Silva, 1999). The filters that have been learned from the training data set are used to project the continuous EEG to a lower dimensional informative subspace. Then the final features are constructed as the log variances from those projections.

A very successful method to determine spatio-temporal filters for a BCI system is the so-called common spatial pattern (CSP) algorithm, which has recently been extended and improved in various directions (Lemm et al., 2005; Dornhege et al., in 2007b, 2006; Tomioka et al., 2006, 2007). Note that for simplicity only the basic CSP version (Fukumizu, 1996) that was originally introduced to BCI by Ramoser et al. (2000) is discussed. The objective of the CSP technique is to find spatial filters that maximize the variance of signals of one condition and at the same time minimize the variance of signals of another condition (see Fig. 1). Since the variance of band-pass filtered signals is equal to the band power, CSP filters can be used to discriminate conditions that are characterized by ERD/ERS effects.

Technically CSP filters are constructed as follows. Let Σ_1 and Σ_2 be estimates of the covariance matrices of the band-pass filtered EEG signals under two conditions (for the extension to more conditions see Dornhege et al., 2004). These two matrices are simultaneously diagonalized in a way that the eigenvalues of Σ_1 and Σ_2 sum to 1. Practically this can be done by calculating

---

[2] But even if no prior knowledge is available for a new mental task, suitable discriminative features can be found by machine learning algorithms.
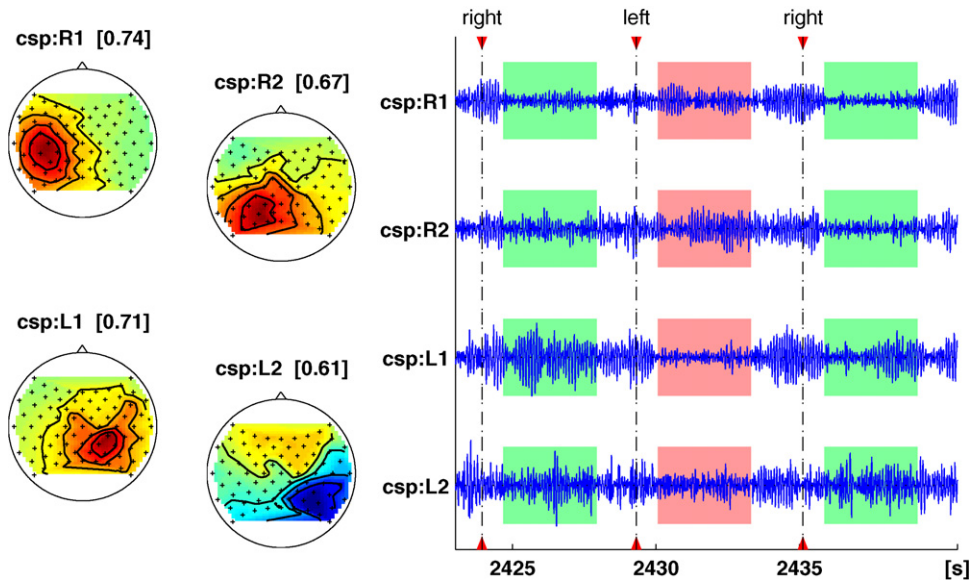
Fig. 1. Two spatial filters for right hand imagery (R1 and R2) and two filters for left hand imagery (L1 and L2) result in four CSP-channels are shown on the right side. The variance of the filtered EEG data changes depending on the task cue, which is indicated by the dashed lines.

the generalized eigenvectors V:

$$\Sigma_1 V = (\Sigma_1 + \Sigma_2)VD. \tag{1}$$

Then the diagonal matrix D contains the eigenvalues of Σ_1 and the column vectors of V are the filters of the common spatial patterns. In theory the best contrast is provided by filters with high eigenvalues (large variance for condition 1 and small variance for condition 2) and by filters with low eigenvalues (vice versa). In practice it is useful to inspect the resulting filters visually in order to detect miss-estimates due to outliers. Further details about the processing methods and the selection of parameters can be found in Blankertz et al. (2005); Fukunaga (1990); Lemm et al. (2005); Dornhege et al. (2007b).

2.3. Classification and explanation by machine learning techniques

For log band-power features in CSP-filtered channels the classification task is feasible very well by linear discriminant analysis in our experience. If the dimensionality of the features is higher, regularization is advisable, see Blankertz et al. (2006c, 2002) for an example in the context of BCI. Features that have more complex distributions typically require nonlinear methods (Müller et al., 2003, 2001). Here we would like to stress further aspects of machine learning techniques for the analysis of single-trial EEG.

First of all, the use of state-of-the-art learning machines enables us to achieve high decision accuracies for BCI (e.g. Blankertz et al., 2002, 2003; Dornhege et al., 2004). When no sufficient prior knowledge is available for a task, then feature selection techniques are required. Particularly those methods are attractive that reveal interpretable results (Lal et al., 2004). Recently, so-called mathematical programming methods like the linear programming machines (LPMs) (Blankertz et al., 2006c;

Bennett and Mangasarian, 1992; Vapnik, 1995; Müller et al., 2001) have been proposed which fulfill this condition.

Applied to EEG measurements from a motor imagery paradigm like the one presented in Section 3.1, the LPM selects less than 4% of the feature dimensions. The chosen features result in very accurate classification of left vs. right hand imagery signals and generalize well to new data. The outcome of the algorithm coincides nicely with what is expected from neurophysiology, i.e., high loadings for electrodes close to sensorimotor cortices in the left and right hemisphere with a strong focus at 12 Hz, i.e., the frequency range of the sensorimotor μ-rhythm, cf. Fig. 2. Note that the feature selection is an integrative part of the learning process and is automatically adapted to a subject, electrode placement, etc.

For the above paradigm a clear physiological expectation exists and the mathematical programming method could match perfectly with this expectation. More interesting and realistic is an exploratory scenario, where a new paradigm is tested that possibly generates unexpected neurophysiological signatures. The learning machine could automatically generate a hypothesis about the underlying task-relevant brain processes. This can serve to adapt and explore the experimental paradigm, such that in principle a better understanding of the brain processes can be inferred. In this sense a machine learning method can offer explanation, which is of great use in the exploration loop for testing new paradigms.

3. BCI control based on movement imagery without subject training

So far the BBCI has mainly studied two paradigms: (a) the discriminability of pre-movement potentials in self-paced executed movements (Blankertz et al., 2003, 2006a,c), where it can be shown that high information transfer rates can be obtained from single-trial classification of fast-paced motor commands

and (b) motor imagery (Blankertz et al., 2007a). Both paradigms do not require subject training. Due to space limitations only the results of the second paradigm are reviewed.

3.1. Experimental setup

Six subjects who all had no or very little experience with BCI feedback took part in this BCI study. To capture brain signals related with motor imagery tasks, EEG was recorded from 118 electrodes mounted on the scalp. To investigate the influence from non-central nervous system activity, electrooculogram (EOG) and electromyogram (EMG) signals were additionally recorded but not used to generate the feedback signal.

In the cued calibration measurement the subjects performed motor imagery tasks of the left hand, the right hand or the right foot. The imagery was sustained after a cue, and cues appeared at intervals of 6–7 s. The subjects did not receive any feedback. After analyis, those two of the three conditions whose signals allowed for the best discrimination were used subsequently used in the three two-class feedback applications described below. Note that the whole experiment which comprises a calibration recording plus all feedback applications was done within one morning.

In the first feedback application (*position controlled cursor*, PCC), the output of the classifier was directly translated to the horizontal position of a cursor. There were two fields on the left resp. right edge of the screen, one of which was highlighted as a target at the beginning of each trial. The cursor started in a deactivated mode (in which it could move into but not activate the target field) and became activated after the user has held the cursor in a central position for 500 ms. The trial ended when the activated cursor touched one of the two fields. That field was then colored green or red, depending on whether it had been the target field that was hit or not. The cursor was then deactivated and the next target was highlighted.

The second feedback application (*rate controlled cursor*, RCC) was very similar, but the control of the cursor was rel-
ative to the actual position, i.e., at each update step a fraction of the classifier output was added to the current cursor position. Each trial started by setting the cursor to the middle of the screen and releasing it after 750 ms.

The third feedback application (*basket game*) is similar to applications in McFarland et al. (2003) and Krausz et al. (2003) and was operated in a synchronous BCI mode. A ball was falling down from the top of the screen at constant speed. Its horizontal position was controlled by the classifier output. At the bottom of the screen there were three horizontally aligned possible target fields. The outer ones had half the width of the middle field to account for the fact that outer positions are easier to hit.

3.2. Results

To compare the results of the different feedback scenarios, the information transfer rate (ITR, Wolpaw et al., 2002) measured in bits per minute (bpm) is used (cf. Table 1). The BBCI can be

Table 1
The first two columns compare the accuracy as calculated by cross-validation on the calibration data with the accuracy obtained online in the feedback application RCC (see text)

Subject no.	Acc (%)		PCC		RCC		Basket	
	Cal.	Fb.	Ø	Peak	Ø	Peak	Ø	Peak
1	95.4	80.5	7.1	15.1	5.9	11.0	2.6	5.5
3	98.0	98.0	12.7	20.3	24.4	35.4	9.6	16.1
4	78.2	88.5	8.9	15.5	17.4	37.1	6.6	9.7
5	78.1	90.5	7.9	13.1	9.0	24.5	6.0	8.8
6	97.6	95.0	13.4	21.1	22.6	31.5	16.4	35.0
Ø	89.5	90.5	10.0	17.0	15.9	27.9	8.2	15.0

Columns three to eight report the information transfer rates (ITR) measured in bits per minute obtained in all three feedback applications PCC, RCC and basket. For each application the first of the two columns reports the average ITR of all runs (of 25 trials each), while the second column reports the peak ITR of all runs. Subject 2 is not reported in this table as he or she did not achieve sufficient BCI control to run the feedback applications (64.6% accuracy in the calibration data).
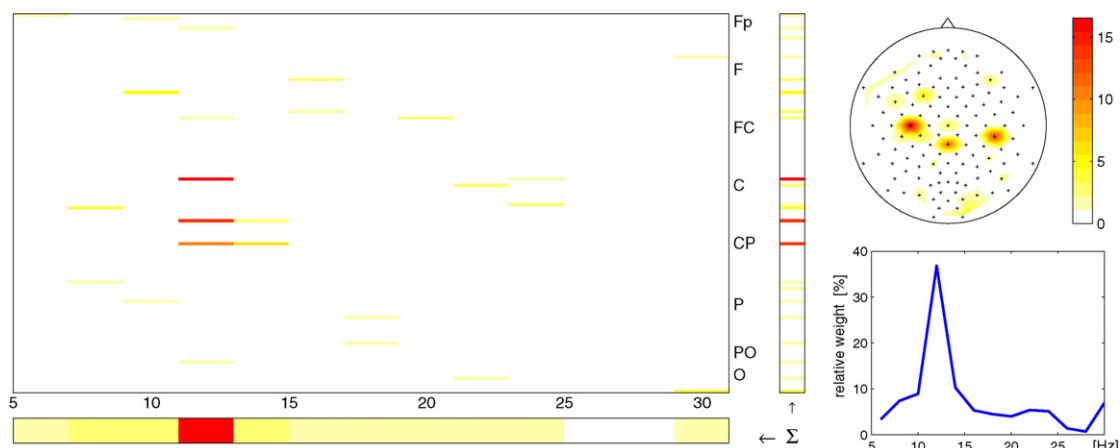

Fig. 2. This figure shows the absolute weight vector (displayed as a channel × frequency matrix) of a sparse LPM classifier. It was trained to discriminate left vs. right hand motor imagery. The bar on the bottom shows the sum across all channels and is displayed also in the lower right plot. The focus in the frequency range lies on the α-band (here 11–14 Hz). The bar on the right side of the matrix shows the sum across all frequency bands and is displayed as a scalp topography in the upper right plot. Note that less than 4% of the features were assigned non-zero weights.

operated at a high decision speed. For the five out of six subjects that showed discriminable brain signals during the calibration recording, this high speed was enforced by using a setting where the subjects were competing to be the winner in one discipline. In the position control the average trial length was 3 s, in rate control 2.5 s. In the basket feedback the trial length is constant (synchronous protocol) but was individually selected for each subject, ranging from 2.1 s to 3 s. The fastest subject was no. 4 which performed at an average speed of one decision every 1.7 s. The most reliable performance was achieved by subject 3: only 2% of the total 200 trials in the rate controlled cursor were misclassified at an average speed of one decision per 2.1 s.

Although EOG artifacts and a small amount of concurrent EMG activity were present in some trials of motor imagery, they were found to have no influence on the classification accuracy. For details see Blankertz et al. (2007a).

4. Applications of BBCI

The machine learning tools that have been developed for the BBCI system enable us to analyze EEG signals in real-time and on a single-trial basis. As a prerequisite the algorithms generally have to be calibrated based on examples of the specific brain patterns of an individual. In the following, two applications of single-trial analysis are presented. First, *Hex-o-Spell*, a text entry system for communicating, which is a classical BCI feedback application. Second, the online monitoring of arousal that reflects the concentration ability of subjects. In the second application, the BBCI provides additional information about the user, that would not be accessible in real-time otherwise. This additional information could in principal serve to improve human-computer interfaces (HMI): the characteristics of the HMI can adapt to the user based on the results of the monitoring estimate (Krepki et al., 2007).

4.1. Text entry with the BBCI: Hex-o-Spell

The challenge in designing a mental text entry system is to map a small number of BCI control states (typically two) to a much high number of symbols (26 letters plus punctuation

marks, blanks and *erase*) while accounting for the low signal-to-noise ratio in the control signal. The fluid interaction in the BBCI system was enabled by applying a combination of probabilistic data and dynamic systems theory and is based on work on mobile interfaces (Williamson and Murray-Smith, 2005). For details see Blankertz et al. (2007b).

The initial configuration of the visual user interface is shown in the leftmost plot in Fig. 3. Six hexagonal fields surround a circle and in each of them five symbols are arranged. An arrow in the center of the circle allows for the selection of a neighboring field of symbols (in a first step) or a neighboring symbol (in a second step). By imagining a right hand movement the arrow turns clockwise. An imagined left hand movement stops the rotation and the arrow starts extending. If this imagination persists, the arrow touches the neighboring hexagon and thereby selects it. Then all other hexagons are cleared and the five symbols of the selected hexagon are moved to individual hexagons as shown in the rightmost screenshot in Fig. 3. For the following second step the arrow is reset to its minimal length while maintaining its last direction. Now the same procedure (rotation and extension) is repeated to finally select a symbol. Note that there are only 5 symbols to choose from in the second step as shown in the rightmost screenshot in Fig. 3. Choosing the empty hexagon makes the application return to the first step without selecting a symbol. This transition allows to revoke the last choice. Misspelt characters can be erased by selecting the backspace symbol. The *turning speed* and the *growing speed* of the arrow are parameters that can be adapted to the user.

Hex-o-Spell incorporates a language model that determines the probability for all symbols conditioned on the letters written so far. Thus, more probable symbols can be reached faster. However, the grouping of the symbols to the six hexagons is fixed (e.g. letters 'A'–'E' are always in the topmost hexagon, see Fig. 3) in order to avoid confusion of the user. Only the arrangement of the symbols within one hexagon is controlled by the language model. Therefore, the symbol probability is matched with the rank of the symbol position which reflects how easy that position can be reached. The symbol that is in forward direction can be reached most easily, since in this case the user can simply maintain the 'go straight' command. To reach the position next
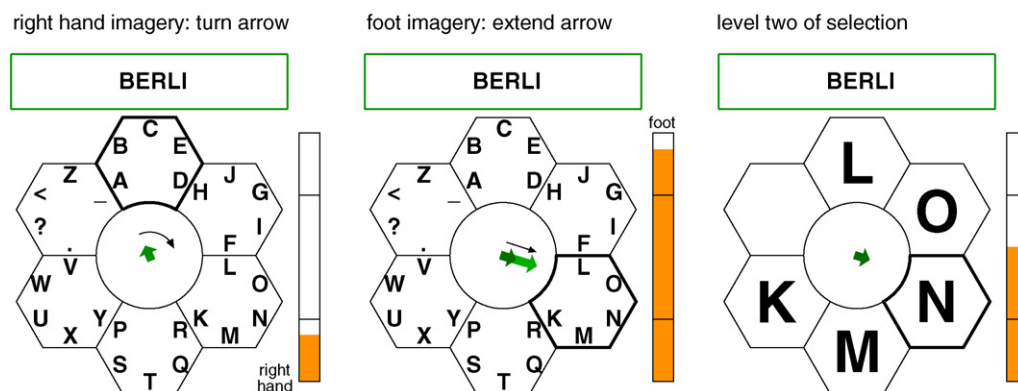

Fig. 3. The mental text entry system *Hex-o-Spell*. The two states classified by the BBCI system (bar on the right in each screenshot) control the turning and growing of the central green arrow (see also text). All letters and a number of meta symbols can be chosen in a two step procedure. If the classifier output is undecided the orange bar is between the thresholds. In this case the arrow maintains its direction and its length diminishes continuously to a minimum.

to it, switching to the turning state and a clockwise turn of 60° is required, cf. Fig. 3. In general, after one symbol has been selected the arrow will be set to point to the hexagon which according to the language model contains the most probable next letter.

4.2. Results

On 2 days in the course of the CeBIT fair 2006 in Hannover, Germany, live demonstrations were given with two subjects that simultaneously used the BBCI system. These demonstrations turned out to be BBCI robustness tests *par excellence*. All over the fair pavilion, noise sources of different kinds (electric, acoustic, ...) were potentially jeopardizing the performance. Low air humidity affected the EEG electrode gel, last but not least, the subjects were under psychological pressure to perform well, for instance in front of several running TV cameras or in the presence of the German minister of research. The preparation of the experiments started at 9:15 a.m. and the live performance began at 11 a.m. The two subjects played a simple computer game *Brain-Pong* against each other or wrote messages with Hex-o-Spell. Except for a few very short breaks and one longer lunch break, the subjects continued until 5 p.m. without any degradation of the classification performance over time. This stability was demonstrated by an impressive typing speed of between 2.3 and 5 char/min for one subject and between 4.6 and 7.6 char/min for the other subject. The speed was measured only for error-free, completed phrases, i.e., all typing errors had to be corrected by using the backspace command.

For a BCI-driven text entry system that did not operate on evoked potentials these numbers are a world class spelling speed, especially when the difficult environment is taken into control. Note that the the subjects had used the BBCI text entry interface only twice before the demonstration.

4.3. Mental state monitoring of users

When aiming to optimize the design of user interfaces or the work flow of manufacturing environments, the mental state of a user during the execution of a task can provide interesting information. Examples of these mental states are the levels of arousal, fatigue or workload. However, questionnaires are of limited use for precisely assessing this information as the delivered answers are often distorted by subjectiveness. Questionnaires cannot determine the quantities of interest in real-time (during the execution of the task) but only in retrospective. Even the monitoring of eye blinks or eye movements only allows for an indirect access to the user's mental state. In contrast to this, the monitoring of a user's errors is a more direct measure but detects critical changes of the user state post hoc only.

We propose to evaluate the use of electroencephalogram (EEG) signals for arousal monitoring. The experimental setting simulates a security surveillance system where the sustained concentration ability of the user is crucial. For the data analysis methods developed in the context of Brain–computer interfaces are applied. With this approach the signals of interest can be isolated and in principal also evaluated in real-time and on a single-trial basis.

4.4. Experimental setup for arousal monitoring

For this pilot study a subject was seated approx. 1 m in front of a computer screen that displayed different stimuli in a forced choice setting. It was asked to respond quickly to stimuli by pressing keys of a keyboard with either the left or right index finger; recording was done with a 128 channel EEG at 100 Hz. The subject had to rate several hundred X-ray images of luggage objects as either dangerous or harmless by a key press after each presentation. The experiment was designed as an oddball paradigm where the number of the harmless objects was much larger than that of the dangerous objects. The terms standard and deviant will subsequently be used for the two conditions. One trial was usually performed within 0.5 s after the cue presentation.

The subject was asked to perform 10 blocks of 200 trials each. Due to the mono-tonous nature of the task and the long duration of the experiment, the subject was expected to show a fading level of arousal which results in worse concentration and the generation of more and more erroneous decisions during later blocks.

For the off-line analysis of the collected EEG signals, the following steps were applied. After exclusion of channels with bad impedances a spatial Laplace filter was applied and the band power features from 8 to 13 Hz were computed on 2 s windows. The resulting band power values of all channels were concatenated into a final vector. As the subject's correct and erroneous decisions were known, a supervised LDA classifier was trained on the data. The classification error of this procedure was estimated by a cross-validation scheme that left out a whole block of 200 trials during each fold for testing. As the number of folds was determined by the number of experimental blocks it varied slightly from subject to subject.

4.5. Results

The erroneous decisions taken by a subject were recorded and smoothed in order to form a measure for the arousal. This measure is further referred to as *error index* and reflects the ability of the subject to concentrate and fulfill the security task. To enhance the contrast of the discrimination analysis, two thresholds were introduced for the error index and set after visual inspection. Extreme trials outside these thresholds defined two sets of trials with a rather high rsp. a low value. The EEG data of the trials were labeled as *sufficiently concentrated* or *insufficiently concentrated* depending on these thresholds for later analysis. Fig. 4 shows the error index. The subject did perform nearly error-free during the first blocks but then showed increasing errors beginning with block 4. However, as the blocks were separated by short breaks, the subject could regain arousal at the beginning of each new block at least for a small number of trials. The trials of high and low error index formed the training data for teaching a classifier to discriminate mental states of insufficient arousal based on single-trial EEG data.

A so-called concentration insufficiency index (CII) of a block was generated by an LDA classifier that had been trained off-line on the labeled training data of the remaining blocks. The
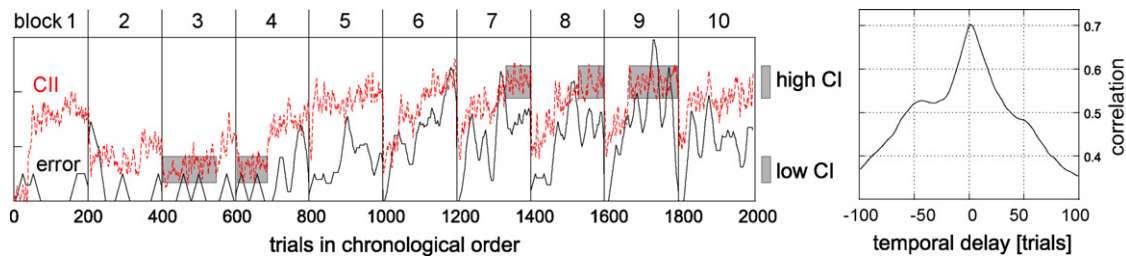
Fig. 4. Left plot: Comparison of the concentration insufficiency index (CII, dotted curve) and the error index for subject 1. The error index (the true performed errors smoothed over time) reflects the inverse of the arousal of the subject. Right plot: Correlation coefficient between the CII (returned by the classifier) and the true performance for subject 1 for different time shifts. Highest correlation is around a zero time shift as expected. Please remark that the CII has an increased correlation with the error even before the error appears.

classifier output (CII) of each trial is plotted in Fig. 4 together with the corresponding error index. It can be observed that the calculated CII mirrors the error index for most blocks. More precisely the CII mimics the error increase inside each block and in blocks 3 and 4 it can anticipate the increase of later blocks. For those later blocks the CII reveals that the subject could not recover its full arousal during the breaks. Instead it shows a short-time arousal for the time immediately after a break, but the CII accumulates over time.

The correlation coefficient of both time series with varying temporal delay is shown in the right plot in Fig. 4. The CII inferred by the classifier and the errors that the subject had actually produced correlate strongly. Furthermore the correlation is high even for predictions that are up to 50 trials ahead in the future.

4.6. Physiological analysis

The trials of subject one are grouped into different CII levels (low/medium/high) as depicted in Fig. 4. To enlarge the contrast for the following analysis, only the extreme CII levels are further considered. Calculating the EEG spectrum for the groups of high and low CII reveals that a main difference between these groups expresses in the frequency domain. Fig. 5 shows the spectra for a number of EEG channels together with signed r^2-values for low CII trials against high CII trials. It shows that the difference between low and high CII levels is reflected best by the differences in the 10 Hz range.

Of course it is also of interest which channels have contributed to the class discriminability by means of LDA in the CII detection system. The scalp plot in Fig. 5 shows the map of signed r^2-values for the discrimination of low and high CII trials (see marked trials in Fig. 4) in the α-band. (Left-)occipital channels account strongest for this difference, which nicely reflects that changes of α-activity depend on visual processes.

5. Concluding discussion

Analyzing EEG signals robustly and in real-time, despite their high variability, and the obviously noisy signal characteristics, is a major challenge. Recently modern machine learning and adaptive signal processing techniques have been able to successfully contribute to this exciting field. In particular, it has become possible to analyze EEG on a single-trial basis for two fields of applications: (a) new insights into general mental state monitoring can be gained and (b) brain–computer interfacing becomes feasible without the need for subject training. Our contribution presents machine learning as a key technology to access these
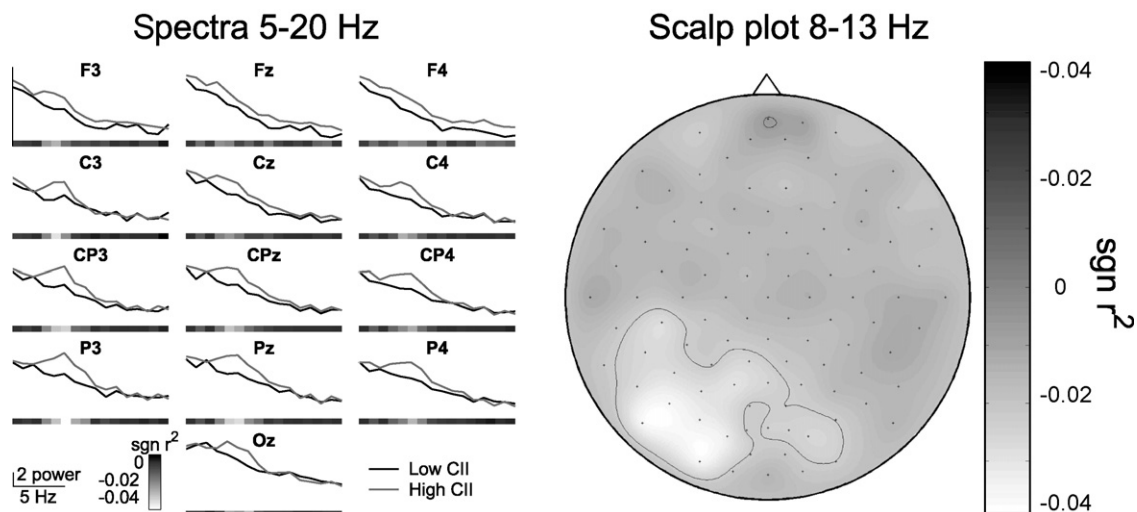

Fig. 5. Left plot: EEG spectra for Laplace-filtered channels and two extreme levels of CII. Gray values code the r^2-values for low CII vs. high CII. Right plot: Scalp plot showing signed r^2-values for 8–13 Hz. The plot reveals discriminability at occipital channels in the α-band.

new fields. We would now like to discuss interesting future directions. Firstly, further gains in the BCI information transfer rate will be achieved by combining different physiological features, by moving from binary to multi-class classification, by adapting the system to non-stationarity and by including the error potential cf. Dornhege et al. (2003, 2004); Dornhege (2006); Sugiyama and Müller (2005). Secondly, as we demonstrated with the mental text entry application *Hex-o-Spell*, it can be foreseen that the incorporation of principles from human–computer interaction research into the desing of BCI feedback applications will boost the usability and the performance of BCIs and brain state monitoring systems. Thirdly, using BCI technology as a tool for computational neuroscience or for rehabilitation engineering is attractive. Gaming and mental state monitoring are only the beginning of a wider *direct* use of information on brain activity for human–machine interaction. BCI output (in addition to the motor output of healthy users) offers a new communication channel for the HMI field that is yet to be explored (Müller and Blankertz, 2006; Krepki et al., 2007). Finally, EEG-based neurotechnology must profit from further improvement of machine learning and signal processing algorithms, but it will require a substantial advancement of sensor technology. EEG caps and amplifiers need to become cheap and easy to use even by non-trained individuals. To be accepted on the market they need to be well designed and gel-free. Once such systems exist, a large number of applications that enhance human–machine interaction will emerg—not only for the disabled user, but also for the healthy.

Acknowledgments

The authors would like to thank their co-authors for granting the use of materials published in Blankertz et al. (2006b).

References

Bennett K, Mangasarian O. Robust linear programming discrimination of two linearly inseparable sets. Optim Methods Software 1992;1:23–34.

Birbaumer N, Kübler A, Ghanayim N, Hinterberger T, Perelmouter J, Kaiser. The though translation device (TTD) for completely paralyzed patients. IEEE Trans Rehab Eng 2000;8(2):190–3.

Blankertz B, Curio G, Müller K-R. Classifying single trial EEG: towards brain–computer interfacing. In: Diettrich TG, Becker S, Ghahramani Z, editors. Advances in Neural Inf. Proc. Systems (NIPS 01), vol. 14; 2002. p. 157–64.

Blankertz B, Dornhege G, Krauledat M, Müller K-R, Curio G. The non-invasive Berlin Brain–Computer Interface: fast acquisition of effective performance in untrained subjects. NeuroImage, 2007;37(2):539–550.

Blankertz B, Dornhege G, Krauledat M, Müller K-R, Curio G. The Berlin brain–computer interface: Report from the feedback sessions. Tech. Rep. 1, Fraunhofer FIRST; 2005.

Blankertz B, Dornhege G, Krauledat M, Müller K-R, Kunzmann V, Losch F. The Berlin brain–computer interface: EEG-based communication without subject training. IEEE Trans Neural Syst Rehab Eng 2006a;14(2):147–52.

Blankertz B, Dornhege G, Krauledat M, Schröder M, Williamson J, Murray-Smith R. The Berlin brain–computer interface presents the novel mental typewriter Hex-o-Spell. In: Proceedings of the 3rd International brain–computer Interface Workshop and Training Course; 2006b. p. 108–9.

Blankertz B, Dornhege G, Lemm S, Krauledat M, Curio G, Müller K-R. The Berlin brain–computer interface: machine learning based detection of user specific brain states. J Univ Comput Sci 2006c;12(6):581–607.

Blankertz B, Dornhege G, Schäfer C, Krepki R, Kohlmorgen J, Müller K-R. Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG-analysis. IEEE Trans Neural Syst Rehab Eng 2003;11(2):127–31.

Blankertz B, Krauledat M, Dornhege G, Williamson J, Murray-Smith R, Müller K-R. A note on brain actuated spelling with the Berlin brain–computer interface. In: Stephanidis C, editor. Universal Access in HCI, Part II, HCII, Vol. 4555 of LNCS. Berlin Heidelberg: Springer; 2007. p. 759–68.

Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, Dimitrov DF. Learning to control a brain–machine interface for reaching and grasping by primates. PLoS Biol 2003:42.

Curran EA, Stokes MJ. Learning to control brain activity: a review of the production and control of EEG components for driving brain–computer interface (BCI) systems. Brain Cogn 2003;51:326–36.

Dornhege G. Increasing information transfer rates for brain–computer interfacing. Ph.D. Thesis. University of Potsdam, 2006.

Dornhege G, Blankertz B, Curio G, Müller K-R. Combining features for BCI. In: Becker S, Thrun S, Obermayer K, editors. Advances in Neural Inf. Proc. Systems (NIPS 02), vol. 15; 2003. p. 1115–22.

Dornhege G, Blankertz B, Curio G, Müller K-R. Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms. IEEE Trans Biomed Eng 2004;51(6):993–1002.

Dornhege G, Blankertz B, Krauledat M, Losch F, Curio G, Müller K-R. Combined optimization of spatial and temporal filters for improving brain–computer interfacing. IEEE Trans Biomed Eng 2006;53(11):2274–81.

Dornhege G, del R Millán J, Hinterberger T, McFarland D, Müller K-R, editors. Towards brain-computer interfacing. MIT Press, Cambridge, MA, 2007.

Dornhege G, Krauledat M, Müller K-R, Blankertz B. General signal processing and machine learning tools for BCI, In: Towards brain–computer interfacing. MIT Press, Cambridge, MA, 2007, pp. 207–233.

Elbert T, Rockstroh B, Lutzenberger W, Birbaumer N. Biofeedback of slow cortical potentials.I. Electroencephalogr Clin Neurophysiol 1980;48:293–301.

Fukumizu K. Active learning in multilayer percetrons. Advances in neural information processings systems, vol. 8. MIT Press; 1996, 295–302.

Fukunaga K. Introduction to statistical pattern recognition. 2nd ed San Diego: Academic Press; 1990.

Haynes J, Rees G. Decoding mental states from brain activity in humans. Nat Rev Neurosci 2006;7:523–34.

Kohlmorgen J, Dornhege G, Braun M, Blankertz B, Müller K-R, Curio G, Hagemann, K., Bruns, A., Schrauf, M., Kincses, W., 2007. Improving human performance in a real operating environment through real-time mental workload detection. In: Dornhege G, del R Millán J, Hinterberger T, McFarland D, Müller K-R, editors, Towards brain–computer interfacing. MIT Press, Cambridge, MA, pp. 409–422.

Krauledat M, Schröder M, Blankertz B, Müller K-R 2007. Reducing calibration time for brain–computer interfaces: a clustering approach. In: Schökopf, B., Platt, J., Hoffman, T. (Eds.), Advances in Neural Inf. Proc. Systems 19. MIT Press, Cambridge, MA, pp. 753–760.

Krausz G, Scherer R, Korisek G, Pfurtscheller G. Critical decision-speed and information transfer in the "Graz brain–computer interface". Appl Psychophysiol Biofeedback 2003;28(3):233–40.

Krepki R, Curio G, Blankertz B, Müller K-R. Berlin brain–computer interface—the hci communication channel for discovery. Int J Human Comput Studies 2007;65:460–77 [special Issue on Ambient Intelligence].

Kübler A, Kotchoubey B, Kaiser J, Wolpaw J, Birbaumer N. Brain–computer communication: Unlocking the locked in. Psychol Bull 2001;127(3):358–75.

Kübler A, Müller K-R, 2007. An introduction to brain–computer interfacing. In: Dornhege G, del R Millán J, Hinterberger T, McFarland D, Müller K-R, editors. Towards brain–computer interfacing. MIT Press, Cambridge, MA, pp. 1–25.

Lal T, Schröder M, Hinterberger T, Weston J, Bogdan M, Birbaumer N. Support vector channel selection in BCI. IEEE Trans Biomed Eng 2004;51(6):1003–10.

Lemm S, Blankertz B, Curio G, Müller K-R. Spatio-spectral filters for improved classification of single trial EEG. IEEE Trans Biomed Eng 2005;52(9):1541–8.

McFarland DJ, Sarnacki WA, Wolpaw JR. Brain–computer interface (BCI) operation: optimizing information transfer rates. Biol Psychol 2003;63:237–51.

Müller K-R, Anderson CW, Birch GE. Linear and non-linear methods for brain–computer interfaces. IEEE Trans Neural Syst Rehab Eng 2003;11(2):165–9.

Müller K-R, Blankertz B. Toward noninvasive brain–computer interfaces. IEEE Signal Proc Mag 2006;23(5):125–8.

Müller K-R, Kohlmorgen J, Rittweger J, Pawelzik K. Analysing physiological data from the wake-sleep state transition with competing predictors. In: NOLTA '95: Las Vegas Symposium on Nonlinear Theory and its Applications; 1995. p. 223–6.

Müller K-R, Mika S, Rätsch G, Tsuda K, Schölkopf B. An introduction to kernel-based learning algorithms. IEEE Neural Networks 2001;12(2):181–201.

Pfurtscheller G, Lopes da Silva F. Event-related EEG/MEG synchronization and desynchronization: basic principles. Clin Neurophysiol 1999;110(11):1842–57.

Ramoser H, Müller-Gerking J, Pfurtscheller G. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehab Eng 2000;8(4):441–6.

Rockstroh B, Birbaumer N, Elbert T, Lutzenberger W. Operant control of EEG and event-related and slow brain potentials. Biofeedback Self-Regulat 1984;9(2):139–60.

Schölkopf B, Smola A. Learning with kernels. Cambridge, MA: MIT Press; 2002.

Sugiyama S, Müller K-R. Input-dependent estimation of generalization error under covariate shift. Stat Dec 2005;23(4):249–79.

Tomioka R, Aihara K, Müller K-R. Logistic regression for single trial EEG classification. In: Schökopf, B., Platt, J., Hoffman, T. (Eds.), Advances in Neural Inf. Proc. Systems 19. MIT Press, Cambridge, MA, pp. 1377–1384.

Tomioka R, Hill J, Blankertz B, Aihara K. Adapting spatial filtering methods for nonstationary bcis. In: Proceedings of 2006 Workshop on Information-Based Induction Sciences (IBIS2006); 2006. p. 65–70.

Vapnik V. The nature of statistical learning theory. New York: Springer Verlag; 1995.

Williamson J, Murray-Smith R. Dynamics and probabilistic text entry. In: Murray-Smith R, Shorten R, editors. Proceedings of the Hamilton Summer School on Switching and Learning in Feedback systems, vol. 3355 of Lecture Notes in Computing Science; 2005. p. 333–42.

Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain–computer interfaces for communication and control. Clin Neurophysiol 2002;113:767–91.