Journal of Universal Computer Science, vol. 12, no. 6 (2006), 581-607
submitted: 31/5/06, accepted: 23/6/06, appeared: 28/6/06 © *J.UCS*

The Berlin Brain-Computer Interface: Machine Learning Based Detection of User Specific Brain States

Benjamin Blankertz, Guido Dornhege, Steven Lemm
(Fraunhofer FIRST (IDA), Berlin, Germany
{blanker,dornhege,lemm}@first.fraunhofer.de)

Matthias Krauledat
(Technical University Berlin, Germany
matthias.krauledat@first.fhg.de)

Gabriel Curio
(Dept. of Neurology, Campus Benjamin Franklin, Charité University Medicine Berlin, Berlin, Germany
gabriel.curio@charite.de)

Klaus-Robert Müller
(Fraunhofer FIRST (IDA), Berlin, Germany and University of Potsdam, Germany
klaus@first.fhg.de)

Abstract: We outline the Berlin Brain-Computer Interface (BBCI), a system which enables us to translate brain signals from movements or movement intentions into control commands. The main contribution of the BBCI, which is a non-invasive EEG-based BCI system, is the use of advanced machine learning techniques that allow to adapt to the specific brain signatures of each user with literally no training. In BBCI a calibration session of about 20min is necessary to provide a data basis from which the individualized brain signatures are inferred. This is very much in contrast to conventional BCI approaches that rely on operand conditioning and need extensive subject training of the order 50-100 hours. Our machine learning concept thus allows to achieve high quality feedback already after the very first session. This work reviews a broad range of investigations and experiments that have been performed within the BBCI project. In addition to these general paradigmatic BCI results, this work provides a condensed outline of the underlying machine learning and signal processing techniques that make the BBCI succeed. In the first experimental paradigm we analyze the predictability of limb movement long before the actual movement takes place using only the movement intention measured from the pre-movement (readiness) EEG potentials. The experiments include both off-line studies and an online feedback paradigm. The limits with respect to the spatial resolution of the somatotopy are explored by contrasting brain patterns of movements of left vs. right hand rsp. foot. In a second complementary paradigm voluntary modulations of sensorimotor rhythms caused by motor imagery (left hand vs. right hand vs. foot) are translated into a continuous feedback signal. Here we report results of a recent feedback study with 6 healthy subjects with no or very little experience with BCI control: half of the subjects achieved an information transfer rate above 35 bits per minute (bpm). Furthermore one subject used the BBCI to operate a mental typewriter in free spelling mode. The overall spelling speed was 4.5-8 letters per minute including the time needed for the correction errors.

Key Words: Brain-Computer Interface, Classification, Common Spatial Patterns, EEG, ERD, Event-Related Desynchronization, Feedback, Information Transfer Rate, Readiness Potential, RP, Machine Learning, Single-Trial Analysis
Category: G.1.6, H.1.1, H.1.2, I.2.1, I.2.6, I.5, J.2, J.3, J.7

1 Introduction

A Brain-Computer Interface (BCI) is a man-to-machine communication channel operating solely on brain signatures independent from muscular output, see [Wolpaw et al., 2002, Kübler et al., 2001, Curran and Stokes, 2003, Dornhege et al., 2006c] for a broad overview. The Berlin Brain-Computer Interface (BBCI) is a non-invasive, EEG-based system whose key features are (1) the use of well-established motor competences as control paradigms, (2) high-dimensional features derived from 128-channel EEG, (3) advanced machine learning techniques, and—as a consequence—(4) no need for subject training.

1.1 Why Machine Learning for Brain-Computer Interfacing?

Traditional neurophysiology typically investigates the average brain. As a simple example, an investigation of the neural correlates of motor preparation of index finger movements would involve a number of subjects doing repeatedly such movements. A grand average over all trials and all subjects will then reveal the generic result, a pronounced cortical negativation which is focused in the corresponding (contralateral) motor area. On the other hand comparing intra-subject averages, cf. Fig. 1, shows a huge subject-to-subject variability, i.e., a large amount of variance in the grand average that was not explained. (See Section 2.1 for a detailed description of the experiment.) Now let us go one step further restricting the investigation to one subject. Comparing the session-wise averages in two (motor imagery) tasks between the sessions recorded on different days we encounter again a huge variability (session-to-session variability), cf. Fig. 2. (See Section 3.1 for a detailed description of the experiment.) When it comes to real-time feedback as in brain-computer interfaces we still have to go one step further. The system needs to be able to identify the mental state of a subject based on one single-trial (duration ≤ 1 s) of brain signals. Fig. 3 demonstrates the huge trial-to-trial variance in one subject in one session (the experiment being the same as above). Nevertheless our BBCI system was able to classify all those trials correctly. The tackling of the enormous trial-to-trial variability is a major challenge in BCI research. Given the high subject-to-subject variability it seems reasonable to have a system that adapts to the specific brain signatures of each user. We believe that advanced techniques for machine learning are an essential tool in this endeavor.

This idea contrasts with the operant conditioning variant of BCI, in which the subject learns by neurofeedback to control a specific EEG feature which is hard-wired in

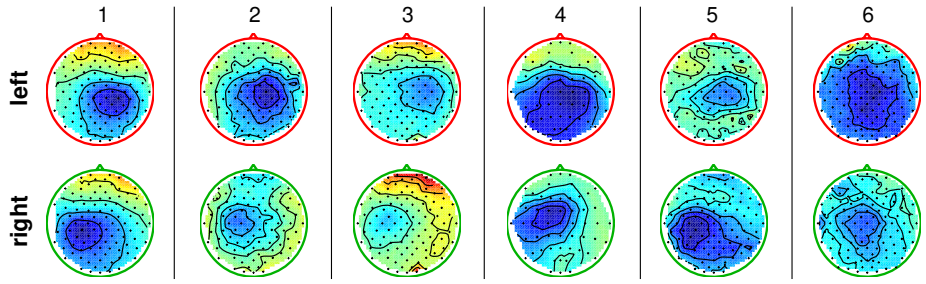

Figure 1: Six subjects performed left vs. right hand index finger tapping. Even though the kind of movement was very much the same in each subject and the task involves a highly overlearned motor competence, the premovement potantial maps (-200 to -100 ms before keypress; blue means negative, red means positive potential) exhibit a great diversity between subjects.

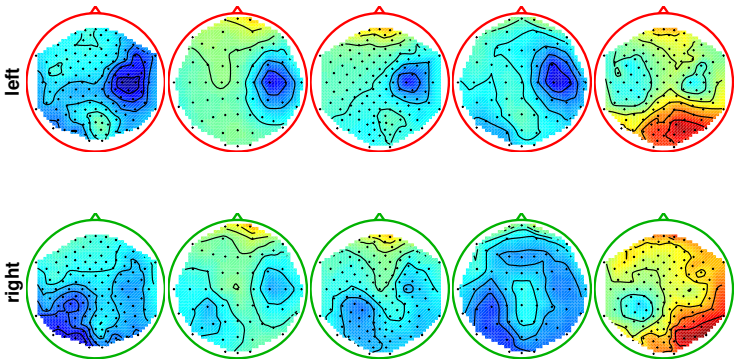

Figure 2: One subject imagined left vs. right hand movements on different days. The maps show spectral power in the alpha frequency band. Even though the maps represent averages across 140 trials each, they exhibit an apparent diversity.

the BCI system, [Elbert et al., 1980, Rockstroh et al., 1984, Birbaumer et al., 2000]. According to the motto 'let the machines learn' our approach minimizes the need for subject training and copes with all kinds of variabilities demonstrated above.

1.2 Overview of this paper

We present two aspects of the BBCI project. The first is based on the discriminability of premovement potentials in voluntary movements. Our initial studies ([Blankertz et al.,

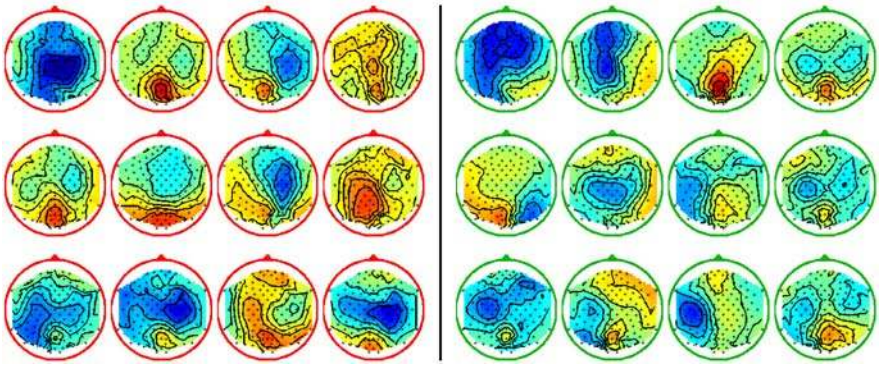

Figure 3: One subject imagined left (red) vs. right (green) hand movements. The topographies show spectral power in the alpha frequency range during single-trials of 3.5 s duration. These patterns exhibit an extreme diversity although recorded from one subject on one day.

2003]) show that high information transfer rates can be obtained from single-trial classification of fast-paced motor commands. Additional investigations – however beyond the scope of this paper – point out ways of improving bit rates further, e.g., by extending the class of detectable movement related brain signals to the ones encountered, e.g. when moving single fingers within one hand.

In a second step we established a BCI system based on motor imagery. A recent feedback study ([Blankertz et al., 2006]) demonstrated the power of the BBCI approach for 6 healthy subjects with no or very little experience of BCI control : 3 subjects could achieve an information transfer rate above 35 bits per minute (bpm), and further two subjects above 24 and 15 bpm, while one subject had no BCI control. These results indicate that higher transfer rates can be achieved when comparing to classical conditioning approaches, even though our subjects were untrained. We would like to reiterate that the BBCI approach is non-invasive. In Section 2 we present single-trial investigations of premovement potentials including online feedback (2.3). In Section 3 we present our BBCI feedback system based on motor imagery and the results of a systematic feedback study (3.3). Section 3.4 gives evidence that the control is solely dependent on central nervous system activity. Section 4 uses machine learning to not only classify and predict but even to explain the underlying structure of the EEG data. In Section 5 we point out lines of further improvement before the concluding discussion 6.

2 Premovement Potentials in Executed Movements

In our first paradigm we studied the pre-movement potentials in overlearned movements, like typewriting on a computer keyboard. Our aim here was to build a classifier based on the Bereitschaftspotential (BP, or readiness potential) which is capable of detecting movement intentions and predicting the type of intended movement (e.g. left vs. right hand) before EMG onset. The basic rationale behind letting healthy subjects actually perform the movements in contrast to movement imagination is that the latter poses a dual task (motor command preparation plus vetoing the actual movement). This suggests that movement imagination by healthy subjects might not guarantee an appropriate correspondence to paralyzed patients as the latter will emit the motor command without veto (but see [Kübler et al., 2005] for a study showing that ALS patients can indeed use modulations of sensorimotor rhythms for BCI control). In order to allow a safe transfer of the results in our setting to paralyzed patients it is essential to make predictions about imminent movements prior to any EMG activity to exclude a possible confound with afferent feedback from muscle and joint receptors contingent upon an executed movement. Being able to predict movements in real time before the EMG activity starts, opens interesting perspectives for assistance of action control in time-critical behavioral contexts, an idea further pursued in [Krauledat et al., 2004].

2.1 Left vs. Right Hand Finger Movements

Our goal is to predict in single-trials the laterality of imminent left vs. right finger movements at a time point prior to the start of EMG activity. The specific feature that we use is the readiness potential (LR, or Bereitschaftspotential), which is a transient postsynaptic response of main pyramidal peri-central neurons, see [Kornhuber and Deecke, 1965]. It leads to a pronounced cortical negativation which is focused in the corresponding motor area, i.e., contralateral to the performing limb reflecting movement preparation, see Fig. 4. Neurophysiologically, the RP is well investigated and described, cf. [Kornhuber and Deecke, 1965, Lang et al., 1989, Cui et al., 1999]. New questions that arise in this context are (a) can the lateralization be discriminated on a single-trial basis, and (b) does the refractory behavior allow to observe the RP also in fast motor sequences? Our investigations provided positive answers to both questions.

In a series of experiments healthy volunteers performed self-paced finger-movements on a computer keyboard with approximate tap-rates of 30, 45, 60 and 120 taps per minute (tpm). EEG was recorded from 128 Ag/AgCl scalp electrodes (except for some experiments summarized in Fig. 5 which were recorded with 32 channels). To relate the prediction accuracy with the timing of EMG activity we recorded electromyogram (EMG) from *M. flexor digitorum communis* from both sides. Also electrooculo-

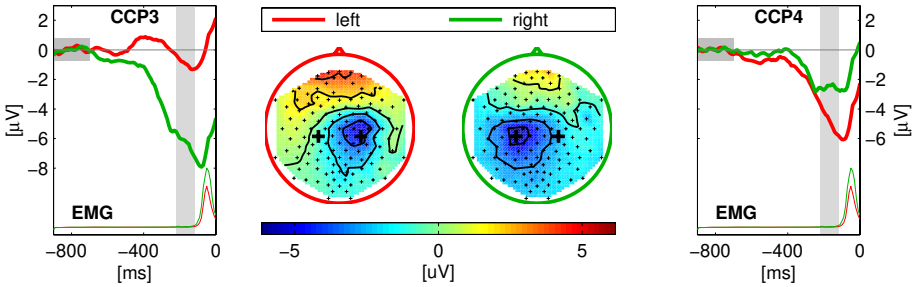

Figure 4: Response *averaged* event-related potentials (ERPs) of one right-handed subject in a left vs. right hand finger tapping experiment (N =275 resp. 283 trials per class). Finger movements were executed self-paced, i.e., without any external cue, in an approximate inter-trial interval of 2 seconds. The two scalp plots show a topographical mapping of scalp potentials averaged within the interval -220 to -120 ms relative to keypress (time interval shaded in the ERP plots). Larger crosses indicate the position of the electrodes CCP3 and CCP4 for which the time course of the ERPs is shown in the subplots at both sides. For comparison time courses of EMG activity for left and right finger movements are added. EMG activity starts after -120 ms and reaches a peak of 70 μV at -50 ms. The readiness potential is clearly visible, a predominantly contralateral negativation starting about 600 ms before movement and raising approximately until EMG onset.

gram (EOG) was recorded to control for the influence of eye movements, cf. Fig. 8. No trials have been discarded from analysis.

The first step towards RP-based feedback is evaluating the predictability of the laterality of upcoming movements. We determined the time point of EMG onset by inspecting classification performance based on EMG-signals (like in Fig. 8) and used it as end point of the windows from which features for the EEG-based classification analysis were extracted. For the data set shown in Fig. 4 the chosen time point is -120 ms which is in coincidence with the onset seen in averaged EMG activity. The choice of the relative position of the classification window with respect to the keypress makes sure that the prediction does not rely on brain signals from afferent nerves. The extraction of the RP features and the classification techniques are described in section 2.2. The result of EEG-based classification for all subjects is shown in Fig. 5 where the cross-validation performance is quantified in bits per minute (according to Shannon's formula) in order to trade-off accuracy vs. decision speed. A discussion of the possible influence of non-central nervous system activity on the classfication can be found in the next section 2.3,

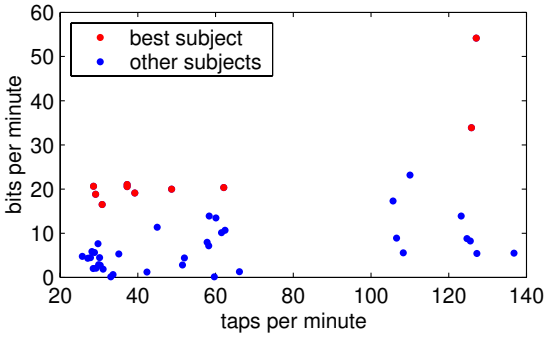

Figure 5: Tapping rates [taps per minute] vs. information transfer rate as calculated by Shannon's formula from the cross-validation error for different subjects peforming self-paced tapping at different average tapping rates with fingers of the left and the right hand. The results of the best subject (marked by red dots) were confirmed in several experiments.

especially in Fig. 8.

The results indicate that the refractory period of the RP is short enough to effectively discriminate pre-movement potentials in finger movement sequences as fast as 2 taps per second. On the other hand it turned out that the performance of RP-based premovement potential detection in a self-paced paradigm is highly subject-specific. Further investigations have studied event-related desynchronization (ERD) effects in the μ and β frequency range, cf. [Pfurtscheller and da Silva, 1999], compare systematically the discriminability of different features and combined RP+ERD features, cf. [Dornhege et al., 2004], and search for modifications in the experimental setup in order to gain high performance for a broader range of subjects.

2.2 Preprocessing and Classification

The following feature extraction method is specifically tailored to extract information from the readiness potential. It extracts the low frequency content with an emphasis on the late part of the signal, where the information content can be expected to be largest in pre-movement trials. Starting points are epochs of 128 samples (i.e. 1280 ms) of raw EEG data as depicted in Fig. 6 (a) for one channel. To emphasize the late signal content, the signal is convoluted with one-sided cosine window (Fig. 6 (b))

$$w(n) := 1 - \cos(n\pi/128) \quad \text{for } n = 0, \ldots, 127,$$

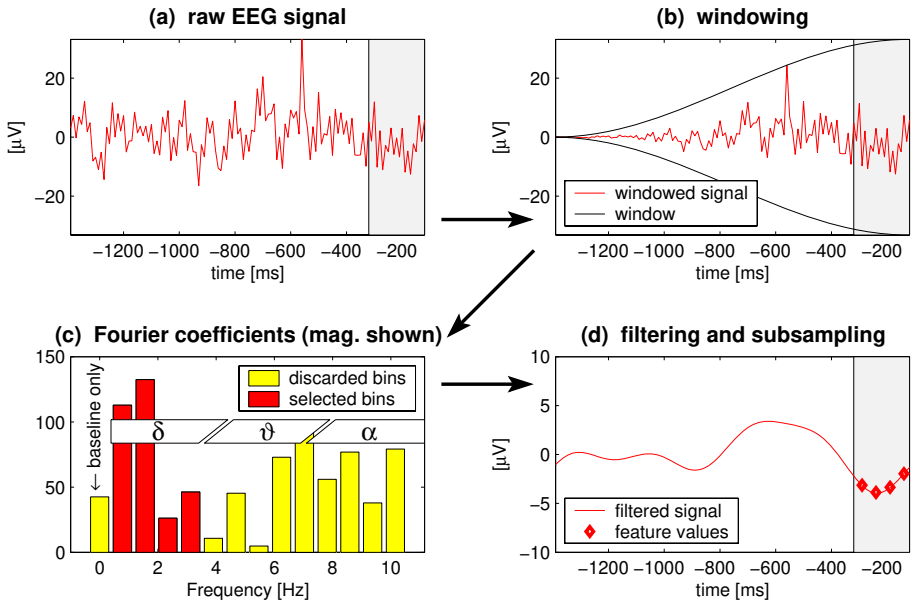

Figure 6: This example shows the feature calculation in one channel of a pre-movement trial $[-1400 - 120]$ ms with keypress at $t = 0$ ms. The pass-band for the FT filtering is 0.4–3.5 Hz and the subsampling rate is 20 Hz. Features are extracted only from the last 200 ms (shaded) where most information on the upcoming movement is expected.

before applying a Fourier transform (FT) filtering technique: from the complex-valued FT coefficients all are discarded but the ones in the pass-band (including the negative frequencies, which are not shown), (Fig. 6 (c)). Transforming the selected bins back into the time domain gives the smoothed signal of which the last 200 ms are subsampled at 20 Hz resulting in 4 feature components per channel (Fig. 6 (d)). The full (RP-) feature vector is the concatenation of those values from all channels for the given time window. For online operation those features are calculated every 40 ms from sliding windows.

Due to our observation that RP-features under particular movement conditions are normally distributed with equal covariance matrices ([Blankertz et al., 2003]), the classification problem meets the assumption of being optimally separated by a linear hyperplane. The data processing described above preserves gaussianity, hence we classify with regularized linear discriminant analysis (RLDA, see [Friedman, 1989]). Regularization is needed to avoid overfitting since we are dealing with a high-dimensional

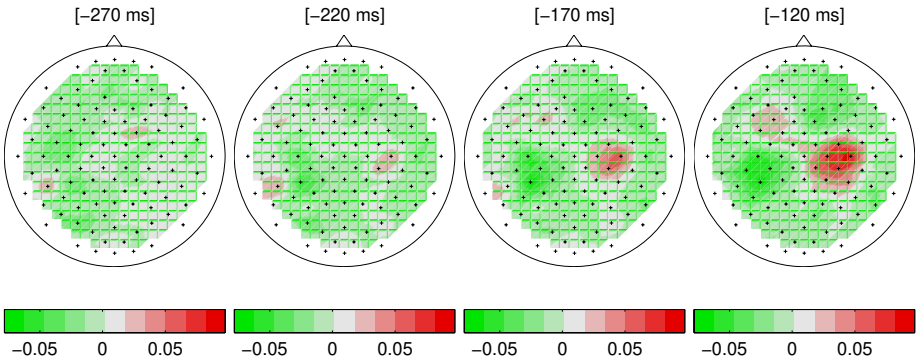

Figure 7: Projection vectors of linear classifiers are instructive. In the case RP-features (see text) they correspond to a temporal sequence of scalp topographies. The weights on both hemispheres get the opposite sign (red vs. green) reflecting the lateralized nature of the readiness potential and the magnitude of the weights is increasing with time reflecting the greater confidence of the potential distribution shortly before keypress.

dataset with only few samples available. Without regularization the estimates of covariance matrices may be inacccurate leading to degraded classification results. The choice of linear classifiers has the advantage that the learned projection vector of the classifier can be visualized and neurophysiologically validated. In the case of RP-features as calculated above the projection vector corresponds to a temporal sequence of scalp topographies, see Fig. 7. See Section 4 for more advanced classification techniques and a discussion of how these techniques can help to explain underlying structures of the analyzed data.

2.3 RP-based feedback in asynchronous mode

The general setting is the following. An experimental session starts with a short period during which the subject performs self-paced finger movements. This session is called calibration session, and the data is used to train a classifier which is then used to make instantaneous predictions on whether the subject intends a hand movement and what its laterality will be.

Although the results of the preceding section demonstrate that an effective *discrimination* of left vs. right hand finger movements is possible well before keypress, it remains a challenge to build a system that predicts movement intentions from ongoing EEG. One point that made the previous classification task easier was that the single

trials were taken from intervals in fixed time relation to the keypress. For the implementation of a useful continuous feedback in an asynchronous mode (meaning without externally controlled timing) we need two more things: (1) the classifier must work reasonably well not only for one exact time point but for a broader interval of time, and (2) the system needs to detect the build up of movement intentions such that it can trigger BCI commands without externally controlled timing.

With respect to the first issue we found that a quite simple strategy (*jittering*) leads to satisfying results: instead of taking only one window as training samples ones extracts several with some time jitter between them. More specifically we extracted two samples per keypress of the calibration measurement, one from a window ending at 150 ms the other at 50 ms before keypress. This method makes the resulting classifier somewhat invariant to time shifts of the samples to be classified, i.e., better suited for the online application to sliding windows. Using more than two samples per keypress event did not improve classification performance further. Extracting samples from windows ending at 50 ms before keypress may seem critical since EMG activity starts at about 120 ms before keypress. But what matters is that the trained classifier is able to make predictions before EMG activity starts no matter what signals it was trained on. This can be seen in Fig. 8 in which EEG-, EMG- and EOG-based classification is compared in relation to the time point of classification. The left plot shows a leave-one-out validation of the calibration measurement, while the right plot shows the accuracy of a classifer trained on the calibration measurement applied to signals of the feedback session, both using jittered training.

To implement the *detection* of upcoming movements we train a second classifier as outlined in [Blankertz et al., 2002]. Technically, the detector of movement intentions was implemented as a classifier that distinguishes between motor preparation intervals (for left and right taps) and 'rest' intervals that were extracted from intervals between movements. To study the interplay of the two classifiers we pursued exploratory feedback experiments with one subject, selected for his good offline results. Fig. 9 shows a statistical evaluation of the two classifiers when applied in sliding windows to the continuous EEG.

The movement discriminator in the left plot of Fig. 9 shows a pronounced separation during the movement (preparation and execution) period. In other regions there is a considereable overlap. From this plot it becomes evident that the left/right classifier alone does not distinguish reliably between *movement intention* and *rest condition* by the magnitude of its output, which explains the need for a movement detector. The elevation for the left class is a little less pronounced (e.g., the median is -1 at $t = 0$ ms compared to 1.25 for right events). The movement intention detector in the right plot of Fig. 9 brings up the movement phase while giving (mainly) negative output to the post movement period.

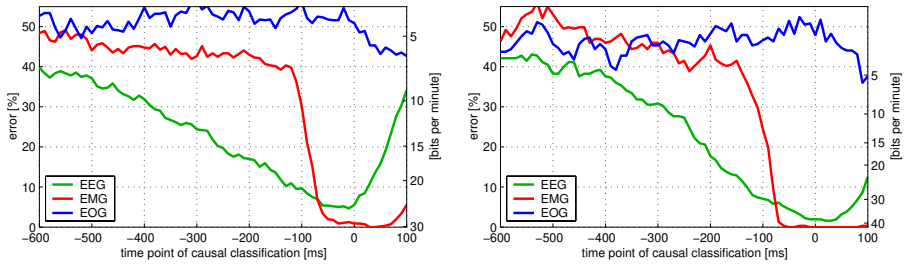

Figure 8: Comparison of EEG, EMG and EOG based classification with respect to the endpoint of the classification interval with $t = 0$ ms being the time point of keypress. For the left plot classifiers were trained in a leave-one-out fashion and applied to a window sliding over the respective left out trials on data of the calibration measurement. For the right plot a classifier (for each type of signal) was trained on data of the calibration measurement and applied to a window sliding over all trials of a feedback session. Note that the scale of the information transfer rate [bits per minute] on the right is different due to a higher average tapping speed in the feedback session.

These two classifiers were used for an exploratory feedback in which a cross was moving in two dimensions, see left plot of Fig. 10. The position on the x-axis was controlled by the left/right movement discriminator and the vertical position was determined by the movement intention detector. Obviously this is *not* an independent control of two dimensions. Rather the cursor was expected to stay in a middle of the lower half during rest and it should move to the upper left or right field when a movement of the left resp. right hand was prepared. The red and green colored fields are the decision areas which only have a symbolic meaning in this application, because no further actions are triggered. In a case study with one subject the expected behavior was indeed found. Although the full flavor of the feeback can only be experienced by watching it, we tried to demonstrate its dynamics by showing the traces of the first 100 trials of the feedback in the right plot of Fig. 10. Each trace displays an interval of the feedback signal -160 to -80 ms relative to keypress. The last 40 ms are intensified and the end point of each trace is marked by a dot.

3 BCI Control based on Imagined Movements

The RP feature presented in the previous section allows an early distinction between motor related mental activities since it reflects movement intent. But even in repetitive movements the discrimination decays already after about 1 second, cf. [Dornhege, 2006]. Accordingly we take an alternative approach for the design of proportional BCI-

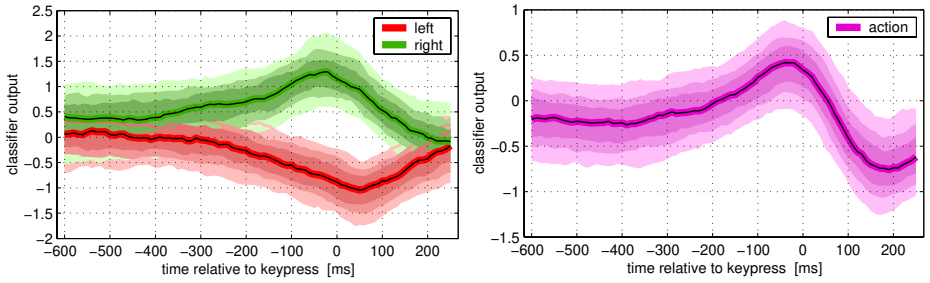

Figure 9: Classifiers were trained in a leave-one-out fashion and applied to windows sliding over unseen epochs yielding traces of graded classifier outputs. The tubes show the 10, 20, 30 resp. 90, 80, and 70 percentile values of those traces. On the left the result is shown of the left vs. right classifier with tubes calculated separately for left and right finger tapping. The subplot on the right shows the result for the movement detection classifier.

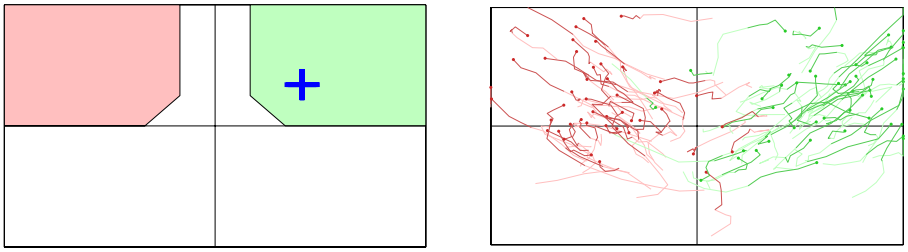

Figure 10: *Left panel:* In a BCI feedback experiment a cursor was controlled by two classifiers. The output of a classifier trained to discriminate left vs. right hand finger movements determined the x-coordinate, while a classifier trained to detect upcoming finger movements determined the y-coordinate. Accordingly the cursor should stay in the lower center area when the subject is at rest while approaching one of the target fields upon movement intentions. This behavior was indeed achieved as can be seen in the *right panel:* Traces of feedback control. Each trace displays an interval of the feedback signal -160 to -80 ms relative to keypress. The last 40 ms are intensified and the end point of each trace is marked by a dot. Traces are colored red or green for subsequent left resp. right hand finger taps.

control, like continuous cursor control. Here we focus on modulations of sensorimotor rhythms evoked by imagined movements. The neurophysiological feature that is exploited here is the event-related desynchronization (ERD): When a subject is at rest (sensori-) motor cortices exhibit a so-called idling rhythm typically with a fundamental frequency at about 12 Hz and a harmonic at 24 Hz. During motor preparation, imagination or execution this rhythm is attenuated or even total blocked in the area of the cortex that corresponds to the respective limb, an effect termed ERD, cf. [Pfurtscheller and Lopes da Silva, 1999]. The (sensori-) motor area of the left hand is in the center of the right hemisphere, the area of the right hand on the left hemisphere and the area of the foot is in the middle of the vertex. The opposite effect of enhancment of brain rhythms is called event-related synchronization and can, e.g., by observed after movement offset.

3.1 Experimental Setup

We designed a setup for a feedback study with 6 subjects who all had no or very little experience with BCI feedback. Brain signals were measured from 118 electrodes mounted on the scalp. To exclude the possibility of influence from non central nervous system activity, EOG and EMG were recorded additionally, see Section 3.4. Those channels were not used to generate the feedback signal.

Each experiment began with a calibration measurement (also called training session but note that this refers to *machine* training) in which labeled trials of EEG data during motor imagery were gathered. This data is used by signal processing and machine learning techniques to estimate parameters of a brain-signal to control-signal translation algorithm. The learning machine can then be applied online to continuously decode incoming signals for producing an instantaneous feedback control signal.

In the training sessions visual stimuli indicated for 3.5s which of the following 3 motor imageries the subject should perform: (L) left hand, (R) right hand, or (F) right foot. The presentation of target cues was interrupted by periods of random length, 1.75 to 2.25s, in which the subject could relax.

Then the experimenter investigated the data to adjust subject-specific parameters of the data processing methods and identified the two classes that gave best discrimination. See Fig. 11 for band-energy mappings of 5 successful subjects and r^2 maps showing that discriminative activity is found over (sensori-) motor cortices only. When the discrimination was satisfactory, a binary classifier was trained and three different kinds of feedback applications followed. This was the case for 5 of 6 subjects who typically performed 8 runs of 25 trials each for each type of feedback applications

During preliminary feedback experiments we realized that the initial classifier was often performing suboptimal, such that the bias and scaling of the linear classifier had to be adjusted. Later investigations have shown that this adaption is needed to account for

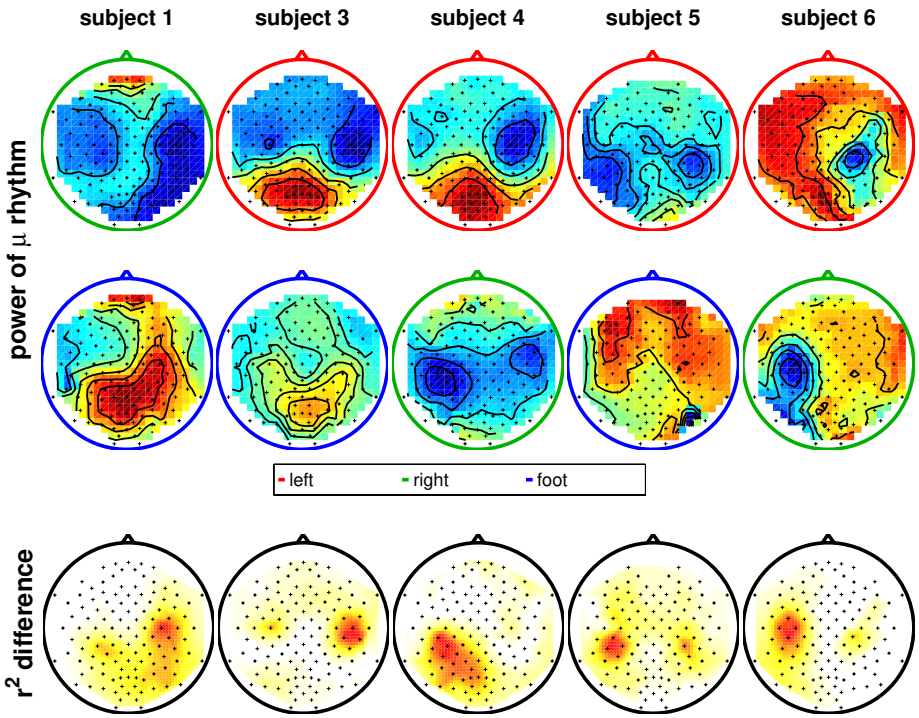

Figure 11: The upper two rows show a topographic display of the energy in the specific frequency band that was used for feedback (as baseline the energy in the inter-stimuli intervals was subtracted). Darker shades indicate lower energy resp. ERD. From the calibration measurement with three types of motor imagery, two were selected for feedback. Energy plots are shown only of those two selected conditions. The type of motor imagery is indicated by the color of the scalp outline. The lower row shows the r^2 differences between the band energy values of the two classes demonstrating that distictive information found over from (sensori-) motor cortices.

the different experimental condition of the (exciting) feedback situation as compared to the calibration session (see e.g. [Shenoy et al., 2006]).

In the first feedback application ('position controlled cursor'), the output of the classifier was directly translated to a horizontal position of a cursor. Out of two target fields at both sides, one was highlighted at the beginning of a trial. The cursor started in a deactivated mode (in which it could move but not trigger a target field) and became activated after the user has held the cursor in a central position for 500 ms. The trial

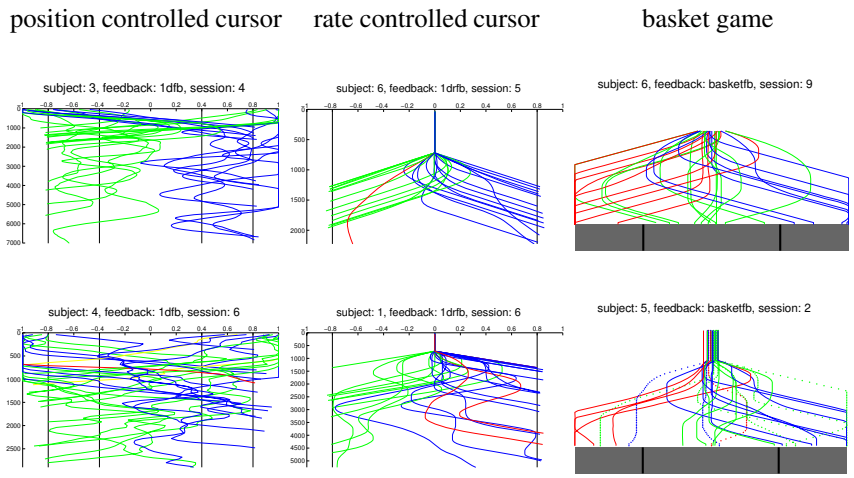

Figure 12: This plots shows the single trial feedback traces from two runs for each type of feedback with time on the vertical axis and BCI control on the horizontal axis. Data sets were chosen from all 5 subjects. For the cursor control, green and blue code correct trials with left resp. right target while erroneous trials are shown in red. For the basket game feedback erroneous trials are coded by dotted lines. The upper row shows examples of very good performance, while performance in the lower row was a bit degraded.

ended when the activated cursor touched a target field which was then colored green or red, depending on whether it was the correct target or not. The cursor was deactivated and the next target appeared.

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

The last feedback application ('basket game') operated in a synchronous mode and is similar to what is used in Graz, cf. [Krausz et al., 2003]. A ball fell at constant speed while its horizontal position was controlled by the classifier output. At the bottom of the screen there were three target fields, the outer having half the width of the middle fields to account for the fact that outer positions were easier to hit.

Fig. 12 shows the trajetories of the selected feedback runs for each of the three types of feedback. The upper row shows examples of very good performance, while performance in the lower row was a bit degraded.

3.2 Processing and Classification

A crucial point in the data processing is to extract appropriate spatial filters that optimize the discriminability of multi-channel brain signals based on ERD/ERS effects of the (sensori-) motor rhythms. Once these filters have been determined, features are extracted as the log of the variance in those surrogate channels. In our experience those features can best be classified by linear methods; we use linear discriminant analysis (LDA). For online operation, features are calculated every 40 ms from sliding windows of 250 to 1000 ms (subject-specific). The spatial filters are calculated individually for each subject on the data of the calibration measurement by Common Spatial Pattern (CSP) analysis, see [Fukunaga, 1990, Lemm et al., 2005, Dornhege et al., 2006d]. The objective of the CSP technique is to find spatial filters that the maximize variance of signals of one condition and at the same time minimize variance of singals of another condition. Since variance of band-pass filtered signals is equal to band-power, CSP filters can be used to discriminate conditions that a characterized by ERD/ERS effects.

Technically CSP analysis goes as follows. Let Σ_1 and Σ_2 be estimates of the covariance matrices of the band-pass filtered EEG signals under the two conditions. 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 caluclating the generalized eigenvectors V:

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

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. 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).

Further details about the processing methods and the selection of parameters can be found in [Blankertz et al., 2005].

3.3 Results

To compare the results of different feedback sessions we use the information transfer rate (ITR, [Wolpaw et al., 2002]) measured in bits per minute (bpm). We calculated this measure for each run according to the following formula:

$$\mathrm{ITR} = \frac{\#\text{ of decisions}}{\text{duration in minutes}} \cdot \left(p \log_2(p) + (1-p) \log(\frac{1-p}{N-1}) + \log_2(N) \right) \tag{2}$$

where p is the accuracy in decisions between N classes ($N = 2$ for cursor control and $N = 3$ for the basket game). Note that the *duration in minutes* refers to the total duration of the run including all inter-trial intervals. In contrast to error rates or ROC curves the ITR takes different duration of trials and different number of classes into account. The

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 'rate controlled cursor'. Columns three to eight report the information transfer rates (ITR) measured in bits per minute as obtained by Shannon's formula, cf. (2). For each feedback application the first column reports the average ITR of all runs (of 25 trials each), while the second column reports the peak ITR of all runs. Subject 2 did not achieve BCI control (64.6% accuracy in the calibration data).

	acc [%]		cursor pos. ctrl		cursor rate ctrl		basket	
	cal.	fb.	overall	peak	overall	peak	overall	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
∅	89.5	90.5	10.0	17.0	15.9	27.9	8.2	15.0

ITR of a random classifier is 0. Table 1 summarizes the information transfer rates that were obtained by the 5 subjects in the three feedback sessions. Highest ITRs were obtained in the 'rate controlled cursor' scenario which has a asynchronous protocol.

One point that is to our knowledge special about the BBCI is that it can be operated at a high decision speed, not only theoretically, but also in practice. In the absolute cursor control the average trial length was 3 seconds, in rate controlled cursor 2.5 seconds. In the basket feedback the trial length is constant (synchronous protocol) but was individually selected for each subject, ranging from 2.1 to 3s. The fastest subject was no. 4 which performed at an average speed of one decision every 1.7s. 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 2.1s. Note that in our notion a trial is ranging from one target presentation to the next including the 'non-control' period during which the selected field was highlighted.

In a later experiment subject 3 operated a mental typewriter based the second feedback application. The alphabet (including a space and a deletion symbol) was split into two parts and those groups of characters were placed on the left resp. right side of the screen. The user selects one subgroup by moving the cursor to the respective side and the process is iterated until a 'group' of one character is selected. The splitting was done alphabetically based on the probabilities of the German alphabet, but no elaborated lan-

guage model was used. In a free spelling mode subject 3 spelled 3 german sentences with a total of 135 characters in 30 minutes, which is a typing speed of 4.5 letters per minutes. Note that all erros have been corrected by using the deletion symbol. For details, see [Dornhege, 2006]. Note that with a novel mental typewriter that is based on principles of human-computer interaction the same subject achieved a typing speed of more than 7 letters per minute, cf. [Müller and Blankertz, 2006].

3.4 Investigating the Dependency of BCI Control

The fact that it is in principle possible to voluntarily modulate motorsensory rhythms without concurrent EMG activity was studied in [Vaughan et al., 1998]. Nevertheless it has to be checked for every BCI experiment involving healthy subjects. For this reason we always record EMG signals even though they are not used in the online system. On one hand we investigated classwise averaged spectra, their statistical significant differences and the scalp distributions and time courses of the power of the μ and β rhythm. The results substantiated that differences of the motor imagery classes indeed were located in sensorimotor cortices and had the typical time courses (except for subject 2 in whom no consistent differences were found), cf. Fig. 11. On the other hand we compared how much variance of the classifier output and how much variance of the EMG signals can be explained by the target class. Much in the spirit of [Vaughan et al., 1998] we made the following analysis using the squared bi-serial correlation coefficient r^2. The r^2-value was calculated for the classifier output and for the band-pass filtered and rectified EMG signals of the feedback sessions. Then the maximum of those time series was determined resulting in one r^2-value per subject and feedback session for EMG resp. for the BCI classifier signal. The r^2 for EMG was in the range 0.01 to 0.08 (mean 0.04 ± 0.03) which is very low compared to the r^2 for the BCI classifier signal which was in the range 0.36 to 0.79 (mean 0.52 ± 0.15).

The fact that the BBCI works without being dependent on eye movements or visual input was additionally verified by letting two subjects control the BBCI with closed eyes which resulted in a comparable performance as in the closed loop feedback.

4 Explaining Underlying Structures by Machine Learning Techniques

When analyzing high dimensional data it is not only important to visualize, predict or classify with low error, but it is essential that exploratory data analysis tools allow to *explain* the underlying structure in order to contribute to a better *understanding* of data.

The ability to generate excellent *and* interpretable results is a long standing problem for general nonlinear methods. Decision trees such as CART [Breiman et al., 1984] or MARS [Friedman, 1991] or more recent tree-algorithms offer a first reasonable compromise. Here the classification can be translated into a set of rules, which, however,

due to their large number, might still not be overly illuminating. Using vanilla neural network approaches (e.g. [Bishop, 1995]), the hidden units nonlinearly combine the measured data and thus interpretation also becomes hard in terms of input variables. Methods like input pruning can alleviate this problem to some extent since such pruning algorithms give rise to solutions which are sparse in the number of input variables ([Bishop, 1995]).

In contrast, linear methods reveal explanatory power. (See e.g. [Müller et al., 2003] for a discussion of linear vs. nonlinear classification.) In the linear separating hyperplane formulation ($w^\top x + b = 0$), the estimated label $\{1, -1\}$ of an input vector $x \in \mathbb{R}^N$ is $\hat{y} = \text{sign}(w^\top x + b)$. If no a-priori knowledge on the probability distribution of the data is available, a typical objective is to minimize a combination of the empirical risk function and a regularization term that restrains the algorithm from overfitting to the training set $\{(x_k, y_k) \,|\, k = 1, \ldots, K\}$. Taking a soft margin loss function yields the empirical risk function $\sum_{k=1}^{K} \max(0, 1 - y_k (w^\top x_k + b))$. In most approaches of this type there is a hyper-parameter that determines the trade-off between risk and regularization, which has to be chosen by model selection on the training set, e.g. similar to [Rätsch et al., 2001, Müller et al., 2001]. When classifying with a linear hyperplane classifier w we can easily rank and thus quantify the contribution of every dimension x_i to the classifier's decision for a given data point x. For instance, the more positive the contribution to the scalar product $w^\top x$ for a certain dimension i, the higher its relative contribution to the overall decision to label the point x as positive.

Nonlinear kernel-based methods (e.g. [Vapnik, 1995, Müller et al., 2001, Schölkopf et al., 1998, Rosipal and Trejo, 2001]) could in principle assess such individual contributions in some appropriate high dimensional feature space $\mathscr{F}$, but when projecting back to the original input space, single feature space dimensions typically correspond to an uninterpretable nonlinear mix of input variables ([Schölkopf et al., 1999]), most similar to the above discussed neural network scenario.

Thus, an ideal strategy should simultaneously construct a good classifier *and* select features for explaining. This integrative approach is in contrast to common sequential methods that first use dimensionality reduction such as PCA and later on build a classifier on the reduced space. Recently, mathematical programming machines especially linear programs (rsp. sparse regularized fisher's discriminant, [Mika et al., 2001, Mika et al., 2000]) have become popular since they can exactly fulfill this integrative task, see e.g. [Blankertz et al., 2002] for a first application of this technique in the context of BCI data analysis.

Let us introduce some mathematical programming machines that are based on the above linear classifier (w, b). The mathematical programming formulation of Regular-

ized Fisher Discriminant (RFD) is [Mika et al., 2001]:

$$\min_{w,b,\xi} \quad 1/2\|w\|_2^2 + C/K\|\xi\|_2^2 \quad \text{subject to} \tag{3}$$

$$y_k(w^\top x_k + b) = 1 - \xi_k \text{ and } \xi_k \geqslant 0, \text{ for } k = 1,\ldots,K \tag{4}$$

where $\|\cdot\|_2$ denotes the ℓ_2-norm ($\|w\|_2^2 = w^\top w$) and C is a model parameter. The constraint $y_k(w^\top x_k + b) = 1 - \xi_k$ ensures that the class means are projected to the corresponding class labels, i.e., 1 and -1. Minimizing the length of w maximizes the margin between the projected class means relative to the intra class variance. A reformulation of eqs. (4) and (4) allows to consider some interesting variants, e.g., Sparse Fisher Discriminant (SFD), which uses the ℓ_1-norm ($\|w\|_1 = \Sigma|w_n|$) on the regularizer, i.e., the goal function is $\|w\|_1 + C/K\|\xi\|_2^2$. This choice favours solutions with sparse vectors w, such that it automatically also yields integrated feature selection in input space while providing excellent classification. Our implementation of SFD uses the *cplex* optimizer [ILOG, 1999].

A further mathematical programming technique are Linear programming machines (LPMs) that are well known for their sparseness [Bennett and Mangasarian, 1992, Vapnik, 1995, Müller et al., 2001]

$$\min_{w,b,\xi} \; 1/2\|w\|_1 + C/K\|\xi\|_1 \quad \text{subject to}$$

$$y_k(w^\top x_k + b) \geqslant 1 - \xi_k, \quad \text{and} \quad \xi_k \geqslant 0 \text{ for } k = 1,\ldots,K.$$

Note that an interesting class of sparse mathematical programming machines are implemented by minimizing partial least squares problems; also here sparsity and accuracy can be provided in an integrative manner (e.g. [Rosipal and Trejo, 2001]).

When applied to one of our EEG measurements from an imagined movement experiment as presented in Section 3.1, the LPM selects less than 4% of the feature dimensions that allow for a left vs. right classification with good generalization. The outcome of the algorithm coincides nicely with what we would expect 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 μ-rhythm, cf. Fig. 13. Note that the feature selection is an integrative part of the learning process and is automatically adapted to subject, electrode placement, etc.

Thus, the use of state-of-the-art learning machines enables us not only to achieve high decision accuracies for BCI (e.g. [Blankertz et al., 2002, Blankertz et al., 2003, Dornhege et al., 2004]), but also, as a by-product of the classification, the few most prominent features that are found match nicely with neurophysiological intuition: the most salient information can be gained in the frequency range of sensorimotor rhythms with a focus over motor cortices, cf. Fig. 13. For the above paradigm it was clear what to

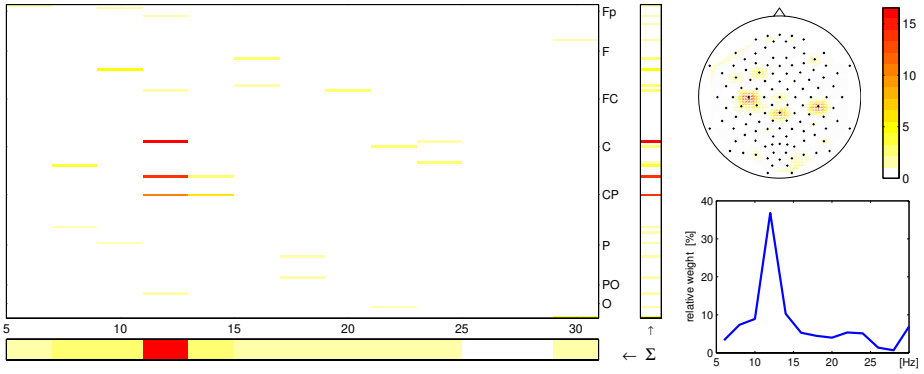

Figure 13: This figure shows the weight vector (display as a channel $\times$ frequency matrix) of a sparse classifier (absolute values) that was trained to discriminate left .vs right hand motor imagery. The bar on the bottom shows the sum across all channels and is displayed also in the lower right 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 show the sum across all frequency bands and is displayed as scalp topography in the upper right plot. Note that less than 4% of the features were assigned non-zero weights.

expect from a physiological point of view and the mathematical programming method could match perfectly with neurophysiological intuition. More interesting and realistic is an exploratory scenario, where (1) a new paradigm is tested that could generate also unexpected neurophysiological signatures, (2) a hypothesis about underlying task relevant brain processes is generated automatically by the learning machine, and (3) finally the paradigm is adapted, so in principle a better understanding of the brain processes could be inferred. In this sense a machine learning method offering explanation, can be of great use in the semi-automatic exploration loop for testing new paradigms.

5 Lines of Further Improvement

5.1 CSSP: CSP with simultaneous spectral optimization

One drawback of the classical CSP algorithm from eqn. 1 is that its performance strongly depends on the appropriate choice of the band-pass filter that needs to be applied to the EEG data in advance. Although [Müller-Gerking et al., 1999] found evidence that a broad band filter is the best general choice, a subject-specific fine tuned filter that is adapted to the individual spectral properties can enhance the results.

The idea of our Common Spatio-Spectral Pattern (CSSP) algorithm ([Lemm et al., 2005]) is to simultaneously optimize the spatial filter in conjunction with very simple frequency filters (one tapped delay FIR filter) for each channel.

Therefore CSSP solves the standard CSP problem in state space, given by the concatenation of the original signal x and its off τ ms delayed version x^τ. More intuitively, we are optimizing a spatial filter in an extended spatial domain, were the delayed signals are treated as new channels $(x, x^\tau)^\top$. Consequently this yields spatial projections $\left(w^0, w^\tau\right)$, that correspond to vectors in this extended spatial domain. Any spatial projection in state space can be expressed as a combination of a pure spatial and spectral filter applied to the original data x, as follow:

$$\begin{aligned}\langle (w^0, w^\tau)^\top, (x, x^\tau)^\top \rangle &= \sum_{c=1} w_c^{(0)} x_{c,\cdot} + w_c^{(\tau)} x_{c,\cdot}^\tau \\ &= \sum_{c=1} \gamma_c \left(\frac{w_c^{(0)}}{\gamma_c} x_{c,\cdot} + \frac{w_c^{(\tau)}}{\gamma_c} x_{c,\cdot}^\tau \right), \end{aligned} \tag{5}$$

where $(\gamma_c)_{c=1,\ldots,C}$ defines a pure spatial filter, whereas $(\frac{w_c^{(0)}}{\gamma_c}, \overbrace{0, \ldots, 0}^{\tau-1}, \frac{w_c^{(\tau)}}{\gamma_c})$ defines a Finite Impulse Response (FIR) filter at each electrode c. This decomposition into a spatial and a FIR filter is not unique, but there exists a very intuitive partitioning, i.e.

$$\gamma_c := \frac{\sqrt{{w_c^{(0)}}^2 + {w_c^{(\tau)}}^2}}{\widetilde{\operatorname{sign}}\left(w_c^{(0)}\right)}. \tag{6}$$

The use of the signed norm as spatial filter γ enables us to examine the spatial origin of the projected source signals. In addition it allows us a one-dimensional parameterization of the FIR filter at each electrode

$$\phi_c^{(\tau)} := \operatorname{atan}\left(\frac{w_c^{(0)}}{w_c^{(\tau)}}\right) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right], \tag{7}$$

and consequently the opportunity for easy visualization of the spatial distribution of the utilized spectral information.

Summarizing, to solve the CSP analysis in the state space, allows us to neglect or emphasize specific frequency bands at each electrode position. However, the performance of a classification method using CSSP-based extracted features will explicitly depend on the specific choice of τ which can be accomplished by some validation approach on the calibration data. More complex frequency filters can be found by concatenating more EEG-signals with several delays. However, as pointed out in [Lemm et al., 2005] in typical BCI situations where only small training sets are available, the

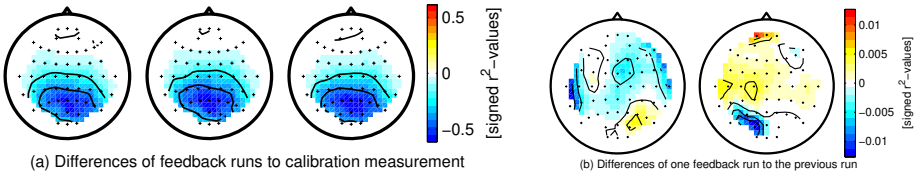

Figure 14: Subfigure (a) shows the differences in band energy from three feedback runs to the data of the calibration measurement as signed r^2-values. The decrease of occipital alpha is most likely due to the increase visual input during BCI feedback. Subfigure (b) shows the difference in band energy of one feedback run (2 and 3) to its predecessor run. The important obervation here is that the r^2-values for the differences between runs is 50 times smaller compared to (a).

choice of a single delay parameter is most effective. Different approaches that implement one global but more complex spectral filter into CSP are currently under investigation [Dornhege, 2006, Dornhege et al., 2006a].

5.2 Investigating the need for adaptvity

Non-stationarities are ubiquitous in EEG signals. The question that is relevant in BCI research is (a) how much of this nonstationarity is reflected in the EEG features, that are used for BCI control, (b) how strongly is the classifer output affected by this change in class distributions, and (c) how this can be remedied. We quantified the shifting of the statistical distributions in particular in view of band energy values and the features one gets from CSP analysis. In contrast to several studies [del R. Millán, 2004, Vidaurre et al., 2004, Wolpaw and McFarland, 2004] that found substantial nonstationarities that need to be accounted for by adaptive classification, our investigations lead to results of somewhat different flavor. Notably the most serious shift of the distributions of band energy features occurred between the initial calibration measurement and online feedback operation of BCI. In contrast the differences during online operation from one run to another were rather inconsequential in most subjects, see Fig. 14. In other subjects those shifts were largely compensated by the CSP filters or the final classifier. The good news with respect the observed shift of distributions is that a simple adaption of classification bias can successfully cure the problem. A thorough description of this study including new techniques for visualization and a systematic comparison of different classification methods coping with shifting distributions can be found in [Shenoy et al., 2006] and forthcoming papers.

6 Discussion and Outlook

The Berlin Brain-Computer Interface makes use of a machine learning approach towards BCI. Working with high dimensional, complex features obtained from 128 channel EEG allows the system a distinguished flexibility for adapting to the specific individual characteristics of each user's brain. This way the BBCI system can provide feedback control even for untrained users typically after a 20 minutes calibration measurement which is then used for the training of the machine learning algorithms.

In one line of investigation we studied the detectability of premovement potentials in healthy subjects. It was shown that high bit rates in single-trial classifications can be achieved by fast-paced motor commands. An analysis of motor potentials during movements with different limbs, e.g. finger II and V within one hand, exposed a possible way of further enhancement. A preliminary study involving patients with traumatic amputations showed that the results can be expected to transfer to phantom movements. Details will be published in a forthcoming paper.

In a second approach we investigate the possibility of establishing BCI control based on motor imagery without subject training. The result from a feedback study with six subjects impressively demonstrates that our system (1) robustly transfers the discrimination of mental states from the calibration to the feedback sessions, (2) allows a very fast switching between mental states, and (3) provides reliable feedback directly after a short calibration measurement and machine training without the need that the subject adapts to the system, all at high information transfer rates, see Table 1.

Recent BBCI activities comprise (a) mental typewriter experiments, with an integrated detector for the error potential, an idea that has be investigated off-line in several studies, cf. [Blankertz et al., 2003, Schalk et al., 2000, Parra et al., 2003, Ferrez and del R. Millán, 2005, Müller and Blankertz, 2006], (b) the online use of combined feature and multi-class paradigms (c) exploration of feature extraction for BCI and (d) real-time analysis of mental workload in subjects engaged in real world cognitive tasks, e.g., in driving situations [Dornhege et al., 2006b].

Our future studies will strive for 2D cursor control and robot arm control, still maintaining our philosophy of minimal subject training.

Acknowledgments

We would like to thank Siamac Fazli, Florin Popescu, Christin Schäfer, Ryota Tomioka, and Andreas Ziehe for fruitful discussions.

This work was supported in part by grants of the *Bundesministerium für Bildung und Forschung* (BMBF), FKZ 01IBE01A/B and 01IGQ0414, by the *Deutsche Forschungsgemeinschaft* (DFG), FOR 375/B1, and by the IST Programme of the European Com-

munity, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors' views.

References

[Bennett and Mangasarian, 1992] Bennett, K. and Mangasarian, O. (1992). Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1:23–34.

[Birbaumer et al., 2000] Birbaumer, N., Kübler, A., Ghanayim, N., Hinterberger, T., Perelmouter, J., Kaiser, J., Iversen, I., Kotchoubey, B., Neumann, N., and Flor, H. (2000). The though translation device (TTD) for completly paralyzed patients. *IEEE Trans. Rehab. Eng.*, 8(2):190–193.

[Bishop, 1995] Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press.

[Blankertz et al., 2002] Blankertz, B., Curio, G., and Müller, K.-R. (2002). Classifying single trial EEG: Towards brain computer interfacing. In Diettrich, T. G., Becker, S., and Ghahramani, Z., editors, *Advances in Neural Inf. Proc. Systems (NIPS 01)*, volume 14, pages 157–164.

[Blankertz et al., 2005] Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., and Curio, G. (2005). The Berlin Brain-Computer Interface: Report from the feedback sessions. Technical Report 1, Fraunhofer FIRST.

[Blankertz et al., 2006] Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., Kunzmann, V., Losch, F., and Curio, G. (2006). The Berlin Brain-Computer Interface: EEG-based communication without subject training. *IEEE Trans. Neural Sys. Rehab. Eng.*, 14(2). in press.

[Blankertz et al., 2003] Blankertz, B., Dornhege, G., Schäfer, C., Krepki, R., Kohlmorgen, J., Müller, K.-R., Kunzmann, V., Losch, F., and Curio, G. (2003). Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis. *IEEE Trans. Neural Sys. Rehab. Eng.*, 11(2):127–131.

[Breiman et al., 1984] Breiman, L., Friedman, J., Olshen, J., and Stone, C. (1984). *Classification and Regression Trees*. Wadsworth.

[Cui et al., 1999] Cui, R. Q., Huter, D., Lang, W., and Deecke, L. (1999). Neuroimage of voluntary movement: topography of the Bereitschaftspotential, a 64-channel DC current source density study. *Neuroimage*, 9(1):124–134.

[Curran and Stokes, 2003] Curran, E. A. and Stokes, M. J. (2003). Learning to control brain activity: A review of the production and control of EEG components for driving brain-computer interface (BCI) systems. *Brain Cogn.*, 51:326–336.

[del R. Millán, 2004] del R. Millán, J. (2004). On the need for on-line learning in brain-computer interfaces. In *Proceedings of the International Joint Conference on Neural Networks*, Budapest, Hungary. IDIAP-RR 03-30.

[Dornhege, 2006] Dornhege, G. (2006). *Increasing Information Transfer Rates for Brain-Computer Interfacing*. PhD thesis, University of Potsdam.

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

[Dornhege et al., 2006a] Dornhege, G., Blankertz, B., Krauledat, M., Losch, F., Curio, G., and Müller, K.-R. (2006a). Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE Trans. Biomed. Eng.* accepted.

[Dornhege et al., 2006b] Dornhege, G., Braun, M., Kohlmorgen, J., Blankertz, B., Müller, K.-R., Curio, G., Hagemann, K., ns, A. B., Schrauf, M., and Kincses, W. (2006b). 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., and Müller, K.-R., editors, *Towards Brain-Computer Interfacing*. MIT press. accepted.

[Dornhege et al., 2006c] Dornhege, G., del R. Millán, J., Hinterberger, T., McFarland, D., and Müller, K.-R., editors (2006c). *Towards Brain-Computer Interfacing*. MIT Press. in preparation.

[Dornhege et al., 2006d] Dornhege, G., Krauledat, M., Müller, K.-R., and Blankertz, B. (2006d). *Towards Brain-Computer Interfacing*, chapter General signal processing and machine learning tools for BCI. MIT Press. accepted.

[Elbert et al., 1980] Elbert, T., Rockstroh, B., Lutzenberger, W., and Birbaumer, N. (1980). Biofeedback of slow cortical potentials. I. *Electroencephalogr. Clin. Neurophysiol.*, 48:293–301.

[Ferrez and del R. Millán, 2005] Ferrez, P. and del R. Millán, J. (2005). You are wrong! – automatic detection of interaction errors from brain waves. In *19th International Joint Conference on Artificial Intelligence*, pages 1413–1418.

[Friedman, 1991] Friedman, J. (1991). Multivariate adaptive regression splines. *Annals of Statistics*, 19(1):1–141.

[Friedman, 1989] Friedman, J. H. (1989). Regularized discriminant analysis. *J. Amer. Statist. Assoc.*, 84(405):165–175.

[Fukunaga, 1990] Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, 2nd edition.

[ILOG, 1999] ILOG (1999). Ilog solver, ilog cplex 6.5 reference manual. www.ilog.com.

[Kornhuber and Deecke, 1965] Kornhuber, H. H. and Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflügers Arch.*, 284:1–17.

[Krauledat et al., 2004] Krauledat, M., Dornhege, G., Blankertz, B., Curio, G., and Müller, K.-R. (2004). The Berlin brain-computer interface for rapid response. *Biomed. Tech.*, 49(1):61–62.

[Krausz et al., 2003] Krausz, G., Scherer, R., Korisek, G., and Pfurtscheller, G. (2003). Critical decision-speed and information transfer in the "Graz Brain-Computer Interface". *Appl. Psychophysiol. Biofeedback*, 28(3):233–240.

[Kübler et al., 2001] Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J., and Birbaumer, N. (2001). Brain-computer communication: Unlocking the locked in. *Psychol. Bull.*, 127(3):358–375.

[Kübler et al., 2005] Kübler, A., Nijboer, F., Mellinger, J., Vaughan, T. M., Pawelzik, H., Schalk, G., McFarland, D. J., Birbaumer, N., and Wolpaw, J. R. (2005). Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface. *Neurology*, 64(10):1775–1777.

[Lang et al., 1989] Lang, W., Lang, M., Uhl, F., Koska, C., Kornhuber, A., and Deecke, L. (1989). Negative cortical DC shifts preceding and accompanying simultaneous and sequential movements. *Exp. Brain Res.*, 74(1):99–104.

[Lemm et al., 2005] Lemm, S., Blankertz, B., Curio, G., and Müller, K.-R. (2005). Spatio-spectral filters for improved classification of single trial EEG. *IEEE Trans. Biomed. Eng.*, 52(9):1541–1548.

[Mika et al., 2001] Mika, S., Rätsch, G., and Müller, K.-R. (2001). A mathematical programming approach to the Kernel Fisher algorithm. In Leen, T., Dietterich, T., and Tresp, V., editors, *Advances in Neural Information Processing Systems*, volume 13, pages 591–597. MIT Press.

[Mika et al., 2000] Mika, S., Rätsch, G., Weston, J., Schölkopf, B., Smola, A., and Müller, K.-R. (2000). Learning discriminative and invariant nonlinear features. unpublished manuscript.

[Müller et al., 2003] Müller, K.-R., Anderson, C. W., and Birch, G. E. (2003). Linear and nonlinear methods for brain-computer interfaces. *IEEE Trans. Neural Sys. Rehab. Eng.*, 11(2):165–169.

[Müller and Blankertz, 2006] Müller, K.-R. and Blankertz, B. (2006). Toward non-invasive brain-computer interfaces. *IEEE Signal Proc. Magazine*. accepted.

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

[Müller-Gerking et al., 1999] Müller-Gerking, J., Pfurtscheller, G., and Flyvbjerg, H. (1999). Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clin. Neurophysiol.*, 110:787–798.

[Parra et al., 2003] Parra, L., Spence, C., Gerson, A., and Sajda, P. (2003). Response error correction - a demonstration of improved human-machine performance using real-time EEG monitoring. *IEEE Trans. Neural Sys. Rehab. Eng.*, 11(2):173–177.

[Pfurtscheller and da Silva, 1999] Pfurtscheller, G. and da Silva, F. H. L. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.*, 110(11):1842–1857.

[Pfurtscheller and Lopes da Silva, 1999] Pfurtscheller, G. and Lopes da Silva, F. (1999). Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clin. Neurophysiol.*, 110(11):1842–1857.

[Rätsch et al., 2001] Rätsch, G., Onoda, T., and Müller, K.-R. (2001). Soft margins for AdaBoost. *Machine Learning*, 42(3):287–320.

[Rockstroh et al., 1984] Rockstroh, B., Birbaumer, N., Elbert, T., and Lutzenberger, W. (1984). Operant control of EEG and event-related and slow brain potentials. *Biofeedback and Self-Regulation*, 9(2):139–160.

[Rosipal and Trejo, 2001] Rosipal, R. and Trejo, L. (2001). Kernel partial least squares regression in reproducing kernel hilbert space. *J. Machine Learning Res.*, 2:97–123.

[Schalk et al., 2000] Schalk, G., Wolpaw, J. R., McFarland, D. J., and Pfurtscheller, G. (2000). EEG-based communication: presence of an error potential. *Clin. Neurophysiol.*, 111:2138–2144.

[Schölkopf et al., 1999] Schölkopf, B., Mika, S., Burges, C., Knirsch, P., Müller, K.-R., Rätsch, G., and Smola, A. (1999). Input space vs. feature space in kernel-based methods. *IEEE Transactions on Neural Networks*, 10(5):1000–1017.

[Schölkopf et al., 1998] Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10:1299–1319.

[Shenoy et al., 2006] Shenoy, P., Krauledat, M., Blankertz, B., Rao, R. P. N., and Müller, K.-R. (2006). Towards adaptive classification for bci. *J. Neural Eng.*, 3:R13–R23.

[Vapnik, 1995] Vapnik, V. (1995). *The nature of statistical learning theory*. Springer Verlag, New York.

[Vaughan et al., 1998] Vaughan, T. M., Miner, L. A., McFarland, D. J., and Wolpaw, J. R. (1998). EEG-based communication: analysis of concurrent EMG activity. *Electroencephalogr. Clin. Neurophysiol.*, 107:428–433.

[Vidaurre et al., 2004] Vidaurre, C., Schlögl, A., Cabeza, R., and Pfurtscheller, G. (2004). About adaptive classifiers for brain computer interfaces. *Biomed. Tech.*, 49(1):85–86.

[Wolpaw et al., 2002] Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., and Vaughan, T. M. (2002). Brain-computer interfaces for communication and control. *Clin. Neurophysiol.*, 113:767–791.

[Wolpaw and McFarland, 2004] Wolpaw, J. R. and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proc. Natl. Acad. Sci. USA*, 101(51):17849–17854.