The Berlin Brain-Computer Interface: Accurate performance from first-session in BCI-naïve subjects

Benjamin Blankertz, Florian Losch, Matthias Krauledat, Guido Dornhege, Gabriel Curio, Klaus-Robert Müller

***Abstract*—The Berlin Brain-Computer Interface (BBCI) project develops a non-invasive BCI system whose key features are (1) the use of well-established motor competences as control paradigms, (2) high-dimensional features from multi-channel EEG and (3) advanced machine learning techniques. Spatio-spectral changes of sensorimotor rhythms are used to discriminate imagined movements (left hand, right hand, foot). A previous feedback study ([1]) with 10 subjects provided preliminary evidence that the BBCI system can be operated at high accuracy for subjects with less than 5 prior BCI exposures. Here, we demonstrate in a group of 14 fully BCI-naïve subjects that 8/14 BCI novices can perform at >84% accuracy in their very first BCI session, and a further 4 subjects >70%. Thus, 12/14 BCI-novices had significant above-chance level performances without any subject training even in the first session, as based on an optimized EEG analysis by advanced machine learning algorithms.**

I. Introduction

Amplitude modulations of sensorimotor rhythms (SMRs) can be voluntarily controlled by most subjects, e.g. by imagining movements. Recently evidence was provided that also patients suffering from amyotrophic lateral sclerosis (ALS) can accomplish SMR modulations ([2]). This ability can be taken as a basis for Brain-Computer Interfaces (BCIs) which are devices that translate the intent of a subject measured from brain signals directly into control commands, e.g. for a computer application or a neuroprosthesis ([3]–[8]). For alternative applications of BCI technology, see [9]–[11]).

One of the challenges in the development of BCI systems is to minimize the amount of subject training that is needed for accurate performance. In this regard the machine learning approach to BCI has been shown to be highly promising ([12, 13]). In our first feedback study ([1]) nine out of ten untrained subjects were able to operate a one-dimensional cursor control feedback with high precision (median $91.7 \pm 5.4\%$ accuracy). Note that the subjects of that study were staff members, some of which had performed feedback with earlier versions of the BBCI system before. Also the number of subjects was limited, so the question was left open whether and how the results would generalize to a broader group of completely untrained subjects.

To date [14] is the only study which investigates BCI feedback accuracy on a larger subject population. At an exposition 99 subjects participated in a BCI experiment where right hand motor imagery was discriminated from feet motor imagery using 2 bipolar EEG channels. One run without and one run with feedback (bar extension) was recorded, each consisting of 40 trials. Feedback was provided either using band power estimation or an adaptive autoregressive model and a linear classifier (trained on the non-feedback run). In the feedback runs, the following results were achieved: 6 % with 90–100 % accuracy, 11.7 % with 80–90 %, 24.9 % with 70–80 %, 45.4 % with 60–70 %, and 12 % below 60 % accuracy. The trial duration was 5 s from appearance of the visual cue to the end of feedback plus 3 s inter-trial break. There was no check for concurrent electromyogram (EMG) activity.

In this paper we report the performance of BCI novices (not from our labs) in their first BBCI feedback session[1] in the framework of a broader study. Here, 12 out of 14 were able to operate a cursor control application accurately (median $86.1 \pm 5.0\%$). A rigorous investigation of EMG signals demonstrates that the success cannot be ascribed to concurrent EMG activity during motor imagery, see Section IV-C.

II. Material

A. Neurophysiology

Macroscopic brain activity during resting wakefulness contains distinct 'idle' rhythms located over various brain areas, e.g. the parietal α-rhythm (8–12 Hz) can be measured over the visual cortex ([15]). The perirolandic sensorimotor cortices show rhythmic macroscopic EEG oscillations (μ-rhythm) ([16,17]), with spectral peak energies of about 9–14 Hz localized predominantly over the postcentral somatosensory cortex and typically phase synchronized components can be found in the beta band ([18]) over the precentral motor cortex. Modulations of the μ-rhythm have been reported for different physiological manipulations, e.g., motor activity, both actual and imagined ([19]–[21]), as well as somatosensory stimulation ([22]). Standard trial averages of μ-rhythm power can reveal attenuation, termed event-related desynchronization (ERD, [23]), or increase (event-related synchronization, ERS).

This work was supported in part by grants of the *Bundesministerium für Bildung und Forschung* (BMBF), FKZ 01IBE01A/B, and by the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. This publication only reflects the authors' views.

BB, MK and KRM are with the Machine Learning Laboratory, Technical University of Berlin, Germany. BB amd KRM are also with Fraunhofer FIRST (IDA), Berlin, Germany. E-mail: blanker@cs.tu-berlin.de.

FL and GC are with the Dept. of Neurology, Campus Benjamin Franklin, Charité University Medicine Berlin, Berlin, Germany.

While the study was conducted, GD was with Fraunhofer FIRST (IDA), Berlin, Germany.

[1]By 'session' we refer to one experimental day. In this study it comprises several calibration (i.e., non-feedback) and feedback runs, see Section II-B.

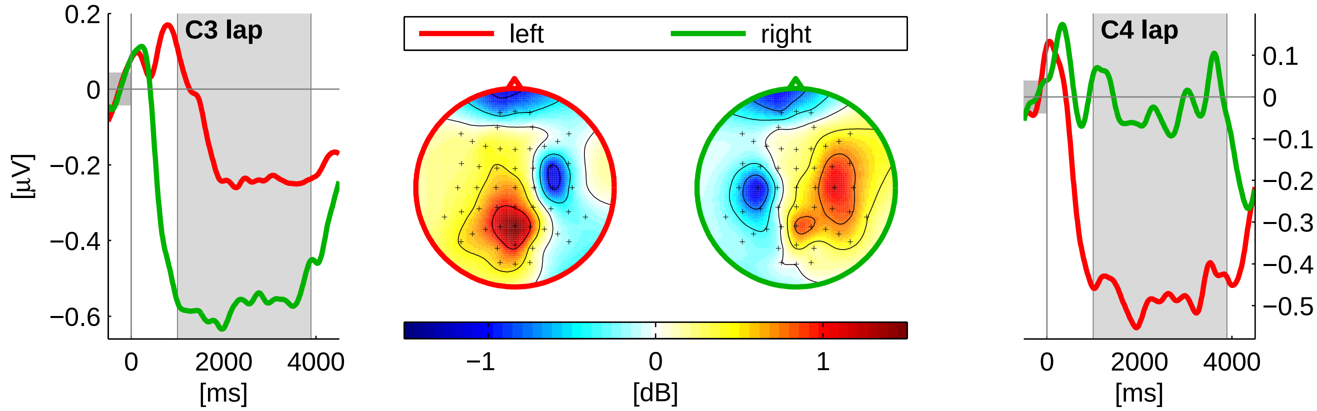


Fig. 1: Event-Related Desynchronization (ERD) during motor imagery of the left and the right hand. Raw EEG signals of one subject have been band-pass filtered in the μ-band. For the time courses, the envelope of the signals has been calculated by Hilbert transform and averaged over segments of -500 to 4500 ms relative to each cue for left resp. right hand motor imagery. ERD curves are shown for Laplace filtered channels at C3 and C4, i.e. over left and right primary motor cortex. The topographical maps of ERD were obtained by calculating the band-power for all (non Laplace filtered) channels in the shaded time interval from 1000 to 4000 ms after stimulus presentation, transforming it to dB, and subtracting the band-power averaged over the whole recording.

Typically ERD is an indication of cortical activity, while ERS can be observed during cortical idling.

Several EEG-based BCI systems rely on the fact that amplitude modulations of sensorimotor rhythms can be voluntarily controlled by most of the subjects, e.g. by imagining movements as explained above (see [24] for an interesting variation of the paradigm). Fig. 1 shows the time course of the amplitude of the μ-rhythm during left hand and right hand motor imagery and the corresponding topographies.

While some approaches try to achieve the required signal strength by training the subjects ([2,25,26]) an alternative is to calibrate the system to the specific charateristics of each user ([1,27]). For the latter data-driven approaches, calculating subject-specific spatial filters have proven to be useful, cf. Section III-A and [28].

B. Experimental Setup

Fourteen healthy BCI-novices (7m, 7f, age 27.5±3.1) took part in this one-session study. We had no physiological or psychological indicators that these subjects were particularly suited for BCI control. In particular, the subjects did not perform in any motor imagery experiment before. All recordings (calibration and feedback runs) of one subject have been recorded on the same day (one ‘session’).

The subjects were sitting in a comfortable chair with arms lying relaxed on armrests. Brain activity was recorded from the scalp with multi-channel EEG amplifiers (BrainAmp DC by Brain Products, Munich, Germany) using 55 Ag/AgCl electrodes (reference at nasion; manufacturer Electro-Cap International, Inc., Eaton, Ohio) in an extended 10-20 system sampled at 1000 Hz with a band-pass from 0.05 to 200 Hz. Additionally, we recorded EMG from both forearms and the right leg as well as horizontal and vertical electrooculogram (EOG). The EMG channels were exclusively used to control for physical limb movements that could correlate with the task and could be reflected directly (artifacts) or indirectly (afferent signals from muscles and joint receptors) in the EEG channels. An investigation of the potential influence of concurrent EMG activity on the classifier, which should operate on the EEG signals only, is given in Section IV-C.

In the beginning, a short ‘artifact measurement’ was recorded during which the subject performed tasks like eye movements, biting, and relaxing with open or closed eyes. During the ‘calibration measurement’ every 5.25–5.75 seconds one of 3 different letters was displayed in randomized sequence for 3.5 seconds on a screen to indicate which mental task the subject should accomplish during that period. The investigated mental tasks were imagined movements of the left hand (L), the right hand (R), and the right foot (F). For each subject 140 trials per class were recorded within 4 runs with several minutes of break in between. Furthermore, subjects performed 1 or 2 runs of physically executed movements.

After the calibration measurement subjects performed 5 feedback runs of 10 minutes duration (for 3 subjects only 4 runs have been recorded). Here the output of the classifier was translated to the horizontal position of a cursor. See Fig. 2 for a cartoon of the feedback and its timing. One of the two fields on the left and right edge of the screen was highlighted as target at the beginning of a trial. The cursor was initially at the center of the screen and started moving according to the BBCI classifier output 750 ms after the indication of the

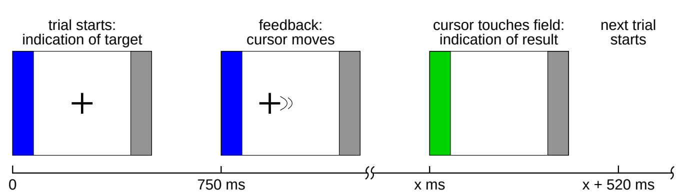


Fig. 2: Course of a feedback trial. The target cue is indicated for 750 ms. Then the cursor starts moving according to the BCI classifier until it touches one of the two fields at the edge of the screen. The touched field is colored green or red according to whether or not its was the correct target. 520 ms later, the next trial starts.

target. The trial ended when the cursor touched one of the two fields. That field was then colored green or red, depending on whether or not it was the correct target. After 520 ms the next target cue was presented (see [1,29] for more details). The number of feedback trials varied according to the subjects' performance between 297 and 1024 (see Table I for the average trial duration of each subject). There was no special motivation (like financial reward) for the subjects to achieve good performance, but most subjects seemed well motivated by their interest in the topic of the study.

III. Methods

A. Subject-specific Spatial Filters

A crucial point in BCI data analysis is the extraction of appropriate spatial filters that optimize the discriminability of multi-channel brain signals based on event-related desynchronization/synchronization (cf. Fig. 1 and Section II-A) of the sensorimotor rhythms. Once these filters have been determined, subsequent processing and classification is relatively straight forward, see Section III-B.

The spatial filters are calculated individually for each subject from the data of a calibration measurement by Common Spatial Pattern (CSP) analysis ([28,30,31]). The objective of the CSP technique is to find spatial filters that maximize variance of signals of one condition and at the same time minimize variance of signals of another condition. Since band-power can be calculated as the variance of band-pass filtered signals, CSP filters can be used to discriminate conditions that are characterized by ERD/ERS effects.

Technically CSP analysis works 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 such that the eigenvalues of Σ_1 and Σ_2 sum to 1. Practically this can be done by calculating the generalized eigenvectors W:

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

Here, the diagonal matrix D contains the (generalized) eigenvalues of Σ_1 and the column vectors of W are the filters for the CSP projections. By this procedure a full decomposition of the sensor space is determined. Best contrast is provided by those filters with high eigenvalues (large variance for condition 1 and small variance for condition 2) and by filters with low eigenvalues (vice versa). Therefore, the common practice in a classification setting is to use several eigenvectors from both ends of the eigenvalue spectrum as features for classification. The CSP filters can also be visualized as scalp maps and chosen according to physiological plausibility. For more details, see the CSP tutorial [28].

B. Features and Classification

The EEG signals of the calibration measurement are band-pass filtered (subject-specific frequency band, see Section III-C) and spatially filtered with the CSP filters determined as described above. From these signals the log-variance is calculated in each trial of the calibration data (interval is selected subject-specfically, typically 750 to 3500 ms relative to the presentation of the visual cue, see Section III-C). This procedure results in a feature vector with dimensionality equal to the number of selected CSP filters, which was between 2 and 6 (most often 4) in this study. To our experience, those features can be well classified by linear methods, so we used linear discriminant analysis (LDA). Nevertheless nonlinear methods (e.g. [32,33]) can potentially improve the results, see also the discussion [34].

For online operation, features are calculated every 40 ms from sliding windows of 500 to 1000 ms width (subject-specific). In the current setup, the CSP filters calculated from the initial calibration measurement are not adapted during online operation. Nevertheless the system allows stable performance even for several hours ([35,36]). The issue of adapting CSP filters during feedback was investigated in [37].

C. Selection of Subject-Specific Hyperparameters

While many parameters of the processing method are estimated automatically like the filter matrix and the weighting of the linear classifier, there are several hyperparameters which were selected semi-automatically like the frequency band and the selection of the CSP filters. To this end, class-wise averaged plots of the spectra, of the ERD curves and of the respective squared bi-serial correlation coefficient (r^2-value) were investigated. The r^2-coefficient reflects how much of the variance in the distribution of all samples is explained by the class affiliation. A heuristic based on the r^2-values suggested parameters to the experimenter. Additionally, the cross-validation error was used as an indicator for good parameter values. Further details about the processing methods and the selection of parameters can be found in [1,38]. A heuristic procedure for fully automatic selection of all neccessary parameters for CSP is proposed and evaluated in [28].

IV. Results

A. Neurophysiological Outcome

The neurophysiological properties of the EEG of all successful subjects are shown in Fig. 3. Only the two imagery conditions that have been used for feedback are displayed. The frequency band that was chosen for online feedback is shaded gray, see Section III-C. Here we discuss the results for two representative subjects (*ct* and *cu*) in more detail. Cross-validation results suggested to use only the alpha band for subject *cu* while using a broader band encompassing alpha and beta range for subject *ct*. The topographies of the reference condition (top scalp map row) look quite similar for most subjects with a dominating occipital/parietal α rhythm, which is only absent in subjects *cn* and *co*. For the motor imagery conditions we essentially expect two effects: regularly, an ERD over the sensorimotor area corresponding to the limb for which motor imagery was performed ([23]), and, potentially, an ERS over flanking sensorimotor areas, possibly reflecting an 'surround inhibition' enhancing focal cortical activation ([39,40]). Subject *ct* shows clearly the contralateral ERD during hand

See large Figures on last two pages.

Fig. 3: The first row displays the averaged spectra of the two motor imagery tasks (red: left hand, green: right hand; blue: right foot) in the calibration measurement that have been used to train the classifier (subject *co* performed physical movements, see Section IV-A). The r^2-values of the difference between those conditions are color coded and the frequency band that has been chosen is shaded gray. The second row shows the average amplitude envelope of that frequency band with 0 being the time point of stimulus presentation in the calibration measurement. Spectra and amplitude envelopes are both shown for the Laplace filtered channel that is indicated in the top of the spectra subplot. The top scalp maps (row 3) show the log power within the chosen frequency band averaged over the whole calibration measurement. The fourth and fifth row display the log band power difference topographies of the particular motor imagery tasks (indicated by L, R, or F, respectively), from which the global average (shown in row 3) is subtracted. The bottom row (6) displays the r^2-values of the difference (row 4 minus row 5) between the individually chosen motor imagery tasks as scalp map.

motor imagery. An ipsilateral ERS was not observed. For subject *cu* left hand motor imagery results in a contralateral ERD accompanied by a weaker ipsilateral ERD. During foot imagery no ERD over the foot area was observed (the same applies for the other subjects that used foot imagery), but a strong ERS over both hand areas.

For subject *cn* β-ERS at central scalp position was observed during foot imagery, while the same region showed an ERD during left hand motor imagery. This is not the typical case, see Section II-A, but ERS of β oscillations over the foot representation area was also reported, e.g., in [41] where this effect was observed only after several months of BCI training.

For subject *co* no sensorimotor rhythm (SMR) was visible in the calibration measurement, i.e. the spectra from Laplacian filtered channels over sensorimotor cortex did not show any peak, but followed the $1/f$ noise shape. We then recorded in the same setup real movements (but acquiring only 35 trials per class). Here a SMR with the expected desynchronization was observed. We were able to train a classifier on the real movement recording and it could be successfully used to provide feedback when the subject *imagined* movements (results reported in Table I). This interesting relation of imagined and real movements goes beyond the scope of this study and will be the topic of a future publication.

Also for subject *cq* the spectra from Laplacian filtered channels over sensorimotor cortex did not show any peak. But here the same applied to the measurement with real movements. Deeper analysis of such cases and the development of strategies to provide BCI control also for subjects of this category is subject of present resarch.

B. Feedback Performance

For subject *cq*, no distinguishable classes were identified. The other 13 subjects performed feedback: 1 near chance level, 3 with 70-80%, 6 with 80-90% and 3 with 90-100% hits, see Table I. The results of all feedbacks runs are shown in the left plot of Fig. 4.

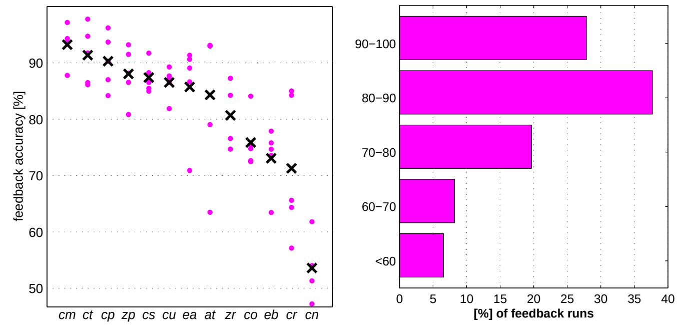


Fig. 4: *Left:* Feedback accuracy of all runs (magenta dots) and intra-subject averages (black crosses). *Right:* Histogram of accuracies obtained in BBCI-controlled cursor movement task in all feedback runs of the study. Note, that the duration of trials (from cue presentation to end of feedback) was variable in our study, see Table I.

TABLE I: Performance results for all 14 subjects of the study. The first column shows the subject code and the second column a two letter code which indicates the classes which have been used for feedback (L: left hand, R: right hand, F: right foot). The third column shows the average accuracy during the feedback ± the standard error of intra-run averages. The average duration ± standard deviation of the feedback trials is provided in the fourth column (duration from cue presentation to target hit, inter-cue interval is 520 ms longer, see Section II-B). Subject order is sorted according to feedback accuracy. The last three columns investigate the influence of concurrent EMG activity. Column 5 reports the results of EMG-based classification; columns 6 and 7 compare the EEG-based feedback accuracy in the subset of trials that have been classified correctly (column 6) or incorrectly (column 7) based on EMG data.

sbj.	**cls.**	**acc. fb [%]**	**dur. [s]**	EMG acc	acc in EMG+	acc in EMG-
cm	LR	93.2 ± 3.9	3.5 ± 2.7	59.0	93.0	94.1
ct	LR	91.4 ± 5.1	2.7 ± 1.5	55.2	90.4	93.2
cp	LF	90.3 ± 4.9	3.1 ± 1.4	87.7	90.2	92.7
zp	LR	88.0 ± 4.8	3.6 ± 2.1	51.0	81.4	79.3
cs	LR	87.4 ± 2.7	3.9 ± 2.3	61.5	87.0	87.7
cu	LF	86.5 ± 2.8	3.3 ± 2.7	53.0	84.9	88.2
ea	FR	85.7 ± 8.5	3.8 ± 2.2	67.7	84.1	86.9
at	LF	84.3 ± 13.1	10.0 ± 8.3	43.1	75.3	76.7
zr	LF	80.7 ± 6.0	3.1 ± 1.9	51.4	72.8	78.7
co	LF	75.9 ± 4.8	4.6 ± 3.1	69.1	75.5	76.3
eb	LF	73.1 ± 5.6	5.9 ± 4.8	69.8	71.5	77.3
cr	LR	71.3 ± 12.6	4.9 ± 3.7	50.5	72.0	70.3
cn	LF	53.6 ± 6.1	3.9 ± 2.4	64.5	60.1	42.5
cq	—	–	–			

C. Independence of BCI Control from EMG Activity

It is in principle possible to voluntarily modulate sensorimotor rhythms without concurrent EMG activity ([42]), nevertheless EMG contributions always need to be thoroughly investigated. While the experimenter checked for EMG activity throughout the whole measurement some subjects of this study did not manage to completely refrain from small EMG activations in some trials. We investigated the influence on the (EEG-based) feedback performance in the following way. The 3 EMG channels were high-pass filtered at 20 Hz and segmented in 500 to 3000 ms epochs relative to cue presentation. Then features were calculated as log variance in 5 subwindows of 500 ms of each trial and classified by linear discriminant analysis (LDA) in a leave-one-out fashion, such that each feedback trial obtained a label 'EMG+' or 'EMG-' according to correct or incorrect classification. Note that 'EMG+' trials do not necessarily contain EMG activity. In the absence of EMG activity classification would perform at chance, i.e. about 50 % of the trials would be 'EMG+'. This was the case for most of the subjects. However, for some subjects there were substantially more 'EMG+' trials indicating that a portion of those trials contains EMG activity. We then compared the accuracy of the EEG-based feedback in the two subgroups of 'EMG+' and 'EMG-' trials, see the last two columns in Table I. If muscle activity has a positive bias then the first quantity will be larger than the second one. It is interesting to see that in most cases where these quantities differ, higher accuracy was achieved in 'EMG-' trials. This suggests that the causal relationship between BCI performance and EMG artifacts might as well be the other way around: when for some subjects the EEG-based feedback does not work reliably, they subconciously used muscle activity in their effort to produce the intended feedback. This activity, however, has apparently no positive influence on the feedback signal. The only subject for whom a positive dependence of EMG activity and feedback accuracy was found (subject *cn*) was the worst perfoming one. In order not to bias our performance statistics we decided to keep this subject in the result table.

V. Discussion

A. What makes the difference?

For a variety of reasons it is difficult in BCI research to pinpoint the reasons, why a particular study leads to good results. In order to approach this question in some aspects, we did an offline analysis of our data using alternative processing techniques. From the vast amount of possible alternatives, we chose to investigate those parameters of data analysis in which our system differs from the one used in [14] (see Section I). That study provides an interesting counterpart to ours since it requires much less resources (number of electrodes, preparation time). The achieved feedback accuracy is remarkable for the quick setup, but considerably worse compared to ours. Due to a number of reasons, those two studies are not directly comparable (the latter study was conducted at a public exposition with additional noise sources as well as potentially psychological pressure on the subjects; a much lower number of trials was recorded, etc.). Still it can be used as a motivation to investigate how the (offline) performance in our data sets varies, when the complexity of the system is reduced. We limit the investigation to three factors: choice of motor imagery classes, selection of frequency band and spatial filtering.

- **CLASSES**. In [14] the pair of motor imagery conditions was fixed to be right hand vs. foot imagery. In our study the calibration measurement additionally encompassed left hand imagery. The pair giving the best cross-validation error on the calibration data was chosen for feedback. For factor CLASSES we investigate *right-foot* and *best pair*.
- **FREQBAND**. In [14] band-power was calculated in the two frequency bands 8–10 Hz and 16–20 Hz, while in our study one frequency band was chosen individually for each subject. For the factor FREQBAND we investigate *fixed* and *individual*.
- **SPATFILT**. In [14] two bipolar channels corresponding to FC3-CP3 and FCz-CPz have been used. In our study subject-specific spatial filters were optimized by CSP analysis. For factor SPATFILT we investigate *bip*[2] and *csp*.

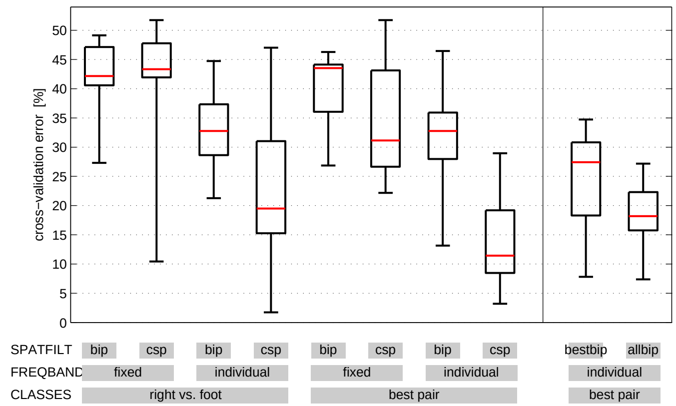


Fig. 5: Comparison of different preprocessing/classification methods. Three different factors have been investigated regarding the impact of their specificity on classification performance. For each factor (SPATFILT, FREQBAND, CLASSES) one fixed value and one subject-adapted adjustment was contrasted. See Section V-A for details. Cross-validation results were obtained from the calibration measurements (we excluded subjects *co* (too little number of trials) and *cq* (signals no discriminable with any of the applied methods) and are displayed as boxplots. Each box ranges from the 25- to the 75-percentile with the median marked in the center. The whiskers extend to the minimum resp. the maximum.

The results of the investigation are presented in Fig. 5 (the two rightmost bars are explained below). While in the settings with fixed frequency bands (FREQBAND=*fixed*), CSP-based classification performs similar to SPATFILT=*bip*, the selection of subject-specific frequency bands in particular in connection with CSP drastically improves performance. Furthermore, the individual selection of two motor imagery classes out of three had a positive effect. Note that only for one subject, combination *foot* vs. *right* (which was the fixed choice in [14]) was selected (see Table I). The performance for FREQBAND=*fixed* is generally quite bad. As can been seen in Fig. 3, the most discriminative frequency band for most subjects in our study falls between 10 and 15 Hz, a range which is left out in the fixed setting of [14] that has also been used here for comparison.

Since the number of channels is an essential factor for the preparation time (but see Section V-D) we investigated the factor SPATFILT further. The question was, whether the essential cause for the lower performance of *bip* was the lower number of channels or the lack of subject-specific positioning. Therefore we studied the following variants for factor SPATFILT:

[2]For CLASSES=*best pair*, the selection of electrode positions in SPATFILT=*bip* was chosen accordingly, e.g., FC4-CP4 was used if class *left* was involved.

- In *allbip* all bipolar channels (of neighboring electrodes in anterior to posterior direction like CP3-P3 or F4-FC4) have been used for classification.
- In *bestbip* the two most effective channels among all bipolar channels have been selected. The selection was based on the Fisher score ([38]). For validation purpose, channel selection was performed within cross-validation on each training set.

The rationale of *bestbip* is that after a first session with the full channel setup, only the two selected bipolar channels might be sufficient in future sessions. The stability of this choice across sessions would have to be investigated, though.

The results (for the most favorable settings of FREQBAND and CLASSES) are shown as the two rightmost boxes in Fig. 5. For method *bestbip* the median error is about 5% lower than for *bip*. Using more channels in *allbip* lead to a substantial improvement of performance. The error dropped by nearly 10%. Note that this error is still 5% above *csp* in the same setting.

Note that while the group size of 14 subjects in the present study is large enough to draw conclusions, future studies with the BBCI system will ultimately need to confirm them for a larger cohort. See [43] for an investigation of relevant features for BCI control in an offline study with 34 subjects performing motor imagery.

B. Effect of Learning During the First Session

One important aspect of a BCI setup is how the feedback accuracy develops over time, during one session or over a sequence of sessions. Generally, an improvement of accuracy over time is taken as a positive result, since it shows that the subject is learning and the system can profit from it. (In adaptive BCI systems, e.g. [26], improvement can also be attributed to the 'learning' of the system). We would like to add another view on the topic. Let us assume that tasks like motor imagery can be performed quite effectively by most subjects. Then a flexible BCI system which successfully detects the subject-specific natural signals might not leave much room for improvement. On the other hand, in a less flexible system which is fixed to the 'average' neurophysiology, the initial accuracy may be low but increase over time due to the subjects' ability to learn to produce brain signals as expected by the system.

Here we investigate the effect of learning during the first session. For this purpose we have split the feedback data of each subject chronologically into 12 intervals and evaluated the classification accuracy within each part (about 55 trials). In order to check for training effects during the calibration runs (without feedback), the calibration data has been split into 6 intervals and classification accuracy was determined within each part (about 43 trials) using the same classifier, and the same temporal and spatial filters that were used during feedback. Since filters and classifier were determined using the calibration data, the results on the calibration data cannot be seen as a general estimate of classifyability (overfitting). Still it is an appropriate way to investigate the *relative* evolution of discriminability over time.

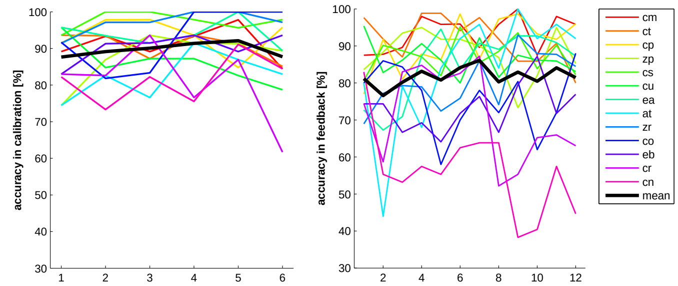


Fig. 6: *Left:* Offline classification accuracy of the feedback classifier in 6 chronological parts of the calibration data. Note that the feedback classifier was trained on the calibration data, so classification may be subject to overfitting. Nevertheless, the relative development of classifyability over time can be seen. *Right:* Online classification accuracy during BBCI feedback evaluated in 12 chronological parts of the feedback data. In both cases, no significant trend in accuracy over time was found.

The results in Fig. 6 show a considerable variation of accuracy over time. While the across-subject average shows a slight positive trend in accuracy (correlation coefficient r= 0.1 in both cases), it is not significant (p= 0.4 for calibration, and p= 0.2 for feedback data).

The results presented in this section substantiate that the good feedback results cannot be attributed to mental training during the relatively long calibration measurement or due to learning during the feedback runs. The feedback results rather reflect the ability of our system to detect the natural subject-specific brain patterns of motor imagery. Obviously the present one-session study is not suitable to discuss the issue of learning in general.

C. Spatial Filtering Prior to CSP

Due to volume conduction effects, raw scalp EEG is associated with a large spatial scale ([44]). Spatial filters, like those determined by CSP analysis, are used to access signals from well localized sources. One concern might be that this smearing leads to suboptimal CSP filters. So a reasonable question is whether additional spatial filtering (like Laplace) should be applied *prior* to CSP analysis. Mathematical details of the following discussion can be found in [28]. If the signals are spatially filtered with an invertible matrix B before CSP is applied, exactly the same features are obtained as without applying B. This is the case if, e.g., principal component analysis (PCA) or independent component analysis (ICA) is applied without discarding components. If the signals are spatially filtered with a noninvertible matrix B, the discriminative value of the obtained CSP filters as measured by Eigenvalues can only decrease. Nevertheless it might be the case that the generalization performance improves, e.g. if B eliminates artifacts to which CSP is susceptible.

For the present study, we did not use spatial filters prior the CSP analysis. In an offline analysis, we evaluated the impact of Laplace filtering before applying CSP and no systematic difference was found: In 7/12 subjects (subjects *co* and *cq* were left out as in Fig. 5) plain CSP obtained a lower error compared to prior Laplce filtering, and the mean error differed non significantly. The median error was 10% for plain CSP and 10.8% with prior Laplace filtering. Still, the result may be different for other data sets.

D. Facilitation of preparation

The proposed system reduces the effort for BCI control in one aspect but increased it in another. While the machine learning methods successfully reduced the need for subject training, the demands for electrode preparation have risen due to the large number of used electrodes. Furthermore a calibration measurement in which the subject does not get feedback needs to be recorded.

With respect to the cumbersome electrode preparation great advancements could be achieved in the meantime. In [45] we present a novel dry EEG recording technology which does not need preparation with conductive gel. In the reported study with good BCI subjects, feedback performance was comparable to the approach with conventional EEG caps for most subjects. Note that this system only uses 6 electrodes and can thus be miniaturized to run with a tiny EEG amplifier and a pocket PC.

We recorded 140 trials per class in the calibration measurement to be on the safe side and to have an elaborate database for offline analysis. The minimal demand for online feedback is much lower. Fig. 7 shows how the validation error on the calibration data depends on the number of training samples. For this evaluation we did not use the parameters (like frequency band) that have been used for feedback, since these parameters were chosen in knowledge of the whole calibration data. Instead we used an automatic procedure to perform the selection ([28]) which is a heuristic based the bi-serial correlation coefficient. From the figure we may conclude that around 40 samples per classes are sufficient for good performance. However, note that here we drew the training set randomly from the whole data set. A shorter calibration measurement might contain less variation in background brain activity which can result in a classifier that is more prone to nonstationarities in ongoing activity. A possible remedy was proposed in [46] where a specific measurement captures possible non task related variations and is used to enforce invariance properties of the classifier.

A further strategy to reduce the training time was developed in [47]. For users who participated in several calibration sessions, the proposed approach allows to calculate a classifier setup (CSP filters + LDA) on these past sessions. It was demonstrated to give stable performance in follow-up feedback session without the need to record new calibration data.

VI. Conclusion

The presented study provides evidence that most subjects can operate the SMR-based BBCI system at high accuracy in their first feedback session. In contrast, in [14] more than 57 % of the subjects achieved accuracies below 70 % during feedback (see Section I), which is the level criterion that is commonly assumed to be the threshold which needs to be surpassed in order to be able to control feedback applications

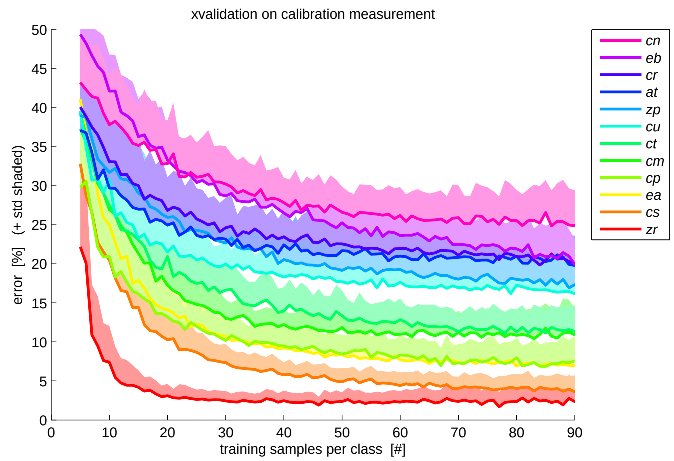


Fig. 7: Dependence of the calibration error on the number of samples used for training. The number of training samples per class was varied from 5 to 90. For each value, 100 subsets of training data have randomly been drawn from the calibration data. CSP and LDA have been trained and then applied to the remaining test data. The curves show the average over the 100 errors values for each subject. The shaded area extends to error plus standard deviation. Note that for subject *co* data from real movements have been used for training. Since only 35 samples per class have been recorded this data set is left out in this figure.

like mental typewriters. Note that in [14] a very reduced system with only 2 bipolar channels was used as opposed to our system with 55 sensors.

The feature of requiring no user training makes the BBCI particularly attractive for BCI research studies. Furthermore it shows that our system is a good candidate for clinical use, because prolonged training is burdensome for paralyzed patients as well as costly. It has been demonstrated that ALS patients can indeed operate a BCI by the voluntary control of sensorimotor rhythms ([2]). Nevertheless it has still to be shown that the presented approach with minimal training works for patients as well.

Future studies will investigate whether subjects for whom the SMR-based approach fails could succeed in settings using different neurophysiological features. Furthermore it is of high interest to see whether those unsuccessful subjects are able to learn BCI control after training.

References

[1] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, “The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects,” *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007. [Online]. Available: http://dx.doi.org/10.1016/j.neuroimage.2007.01.051

[2] A. Kübler, F. Nijboer, J. Mellinger, T. M. Vaughan, H. Pawelzik, G. Schalk, D. J. McFarland, N. Birbaumer, and J. R. Wolpaw, “Patients with ALS can use sensorimotor rhythms to operate a brain-computer interface,” *Neurology*, vol. 64, no. 10, pp. 1775–1777, 2005.

[3] G. Dornhege, J. del R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, Eds., *Toward Brain-Computer Interfacing*. Cambridge, MA: MIT Press, 2007.

[4] B. Allison, E. Wolpaw, and W. J.R., “Brain-computer interface systems: progress and prospects,” *Expert Rev Med Devices*, vol. 4, no. 4, pp. 463–474, 2007.

[5] N. Birbaumer, C. Weber, C. Neuper, E. Buch, K. Haapen, and L. Cohen, “Physiological regulation of thinking: brain-computer interface (BCI) research,” *Prog Brain Res*, vol. 159, pp. 369–391, 2006.

[6] G. Pfurtscheller, C. Neuper, and N. Birbaumer, "Human Brain-Computer Interface," in *Motor Cortex in Voluntary Movements*, A. Riehle and E. Vaadia, Eds. New York: CRC Press, 2005, ch. 14, pp. 367–401.

[7] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin Neurophysiol*, vol. 113, no. 6, pp. 767–791, 2002.

[8] A. Kübler, B. Kotchoubey, J. Kaiser, J. Wolpaw, and N. Birbaumer, "Brain-computer communication: Unlocking the locked in," *Psychol Bull*, vol. 127, no. 3, pp. 358–375, 2001.

[9] J. Kohlmorgen, G. Dornhege, M. Braun, B. Blankertz, K.-R. Müller, G. Curio, K. Hagemann, A. Bruns, M. Schrauf, and W. Kincses, "Improving human performance in a real operating environment through real-time mental workload detection," in *Toward Brain-Computer Interfacing*, G. Dornhege, J. del R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, Eds. Cambridge, MA: MIT press, 2007, pp. 409–422.

[10] K.-R. Müller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, and B. Blankertz, "Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring," *J Neurosci Methods*, vol. 167, no. 1, pp. 82–90, 2008. [Online]. Available: http://dx.doi.org/10.1016/j.jneumeth.2007.09.022

[11] A. Gerson, L. Parra, and P. Sajda, "Cortically coupled computer vision for rapid image search," *IEEE Trans Neural Sys Rehab Eng*, vol. 14, no. 2, pp. 174–179, 2006.

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

[13] K.-R. Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz, "Machine learning techniques for brain-computer interfaces," *Biomed Tech*, vol. 49, no. 1, pp. 11–22, 2004.

[14] C. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How many people are able to operate an EEG-based Brain-Computer Interface (BCI)?" *IEEE Trans Neural Sys Rehab Eng*, vol. 11, no. 2, pp. 145–147, 2003.

[15] H. Berger, "Über das Elektroenkephalogramm des Menschen," *Archiv für Psychiatrie und Nervenkrankheiten*, vol. 99, no. 6, pp. 555–574, 1933.

[16] H. Jasper and H. Andrews, "Normal differentiation of occipital and precentral regions in man," *Arch. Neurol. Psychiat. (Chicago)*, vol. 39, pp. 96–115, 1938.

[17] R. Hari and R. Salmelin, "Human cortical oscillations: a neuromagnetic view through the skull," *Trends in Neuroscience*, vol. 20, pp. 44–9, 1997.

[18] V. V. Nikulin and T. Brismar, "Phase synchronization between alpha and beta oscillations in the human electroencephalogram," *Neuroscience*, vol. 137, pp. 647–657, 2006.

[19] H. Jasper and W. Penfield, "Electrocorticograms in man: Effect of voluntary movement upon the electrical activity of the precentral gyrus," *Arch. Psychiatrie Zeitschrift Neurol.*, vol. 183, pp. 163–74, 1949.

[20] G. Pfurtscheller and A. Aranbibar, "Evaluation of event-related desynchronization preceding and following voluntary self-paced movement," *Electroencephalogr Clin Neurophysiol*, vol. 46, no. 2, pp. 138–46, 1979.

[21] A. Schnitzler, S. Salenius, R. Salmelin, V. Jousmäki, and R. Hari, "Involvement of primary motor cortex in motor imagery: a neuromagnetic study," *NeuroImage*, vol. 6, no. 3, pp. 201–208, 1997.

[22] V. Nikouline, K. Linkenkaer-Hansen, W. H., M. Kesäniemi, E. Antonova, R. Ilmoniemi, and J. Huttunen, "Dynamics of mu-rhythm suppression caused by median nerve stimulation: a magnetoencephalographic study in human subjects," *Neurosci Lett*, vol. 294, no. 3, pp. 163–166, 2000.

[23] G. Pfurtscheller and F. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin Neurophysiol*, vol. 110, no. 11, pp. 1842–1857, Nov 1999.

[24] V. V. Nikulin, F. U. Hohlefeld, A. M. Jacobs, and G. Curio, "Quasi-movements: A novel motor-cognitive phenomenon," *Neuropsychologia*, 2008, in press. [Online]. Available: http://dx.doi.org/10.1016/j.neuropsychologia.2007.10.008

[25] J. R. Wolpaw and D. J. McFarland, "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans," *Proc Natl Acad Sci USA*, vol. 101, no. 51, pp. 17 849–17 854, 2004.

[26] C. Vidaurre, A. Schlögl, R. Cabeza, R. Scherer, and G. Pfurtscheller, "A fully on-line adaptive BCI," *IEEE Trans Biomed Eng*, vol. 53, no. 6, pp. 1214–1219, 2006.

[27] C. Guger, H. Ramoser, and G. Pfurtscheller, "Real-time EEG analysis with subject-specific spatial patterns for a Brain Computer Interface (BCI)," *IEEE Trans Neural Sys Rehab Eng*, vol. 8, no. 4, pp. 447–456, 2000.

[28] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Proc Magazine*, vol. 25, no. 1, pp. 41–56, Jan. 2008. [Online]. Available: http://dx.doi.org/10.1109/MSP.2008.4408441

[29] B. Blankertz, G. Dornhege, S. Lemm, M. Krauledat, G. Curio, and K.-R. Müller, "The Berlin Brain-Computer Interface: Machine learning based detection of user specific brain states," *J Universal Computer Sci*, vol. 12, no. 6, pp. 581–607, 2006.

[30] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. San Diego: Academic Press, 1990.

[31] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans Rehab Eng*, vol. 8, no. 4, pp. 441–446, 2000.

[32] J. H. Friedman, "Regularized discriminant analysis," *J Amer Statist Assoc*, vol. 84, no. 405, pp. 165–175, 1989.

[33] K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Schölkopf, "An introduction to kernel-based learning algorithms," *IEEE Neural Networks*, vol. 12, no. 2, pp. 181–201, May 2001.

[34] K.-R. Müller, C. W. Anderson, and G. E. Birch, "Linear and non-linear methods for brain-computer interfaces," *IEEE Trans Neural Sys Rehab Eng*, vol. 11, no. 2, pp. 165–169, 2003.

[35] K.-R. Müller and B. Blankertz, "Toward noninvasive brain-computer interfaces," *IEEE Signal Proc Magazine*, vol. 23, no. 5, pp. 125–128, September 2006.

[36] B. Blankertz, M. Krauledat, G. Dornhege, J. Williamson, R. Murray-Smith, and K.-R. Müller, "A note on brain actuated spelling with the Berlin Brain-Computer Interface," in *Universal Access in HCI, Part II, HCII 2007*, ser. LNCS, C. Stephanidis, Ed., vol. 4555. Berlin Heidelberg: Springer, 2007, pp. 759–768.

[37] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *J Neural Eng*, vol. 3, no. 1, pp. R13–R23, 2006. [Online]. Available: http://dx.doi.org/10.1088/1741-2560/3/1/R02

[38] G. Dornhege, M. Krauledat, K.-R. Müller, and B. Blankertz, "General signal processing and machine learning tools for BCI," in *Toward Brain-Computer Interfacing*, G. Dornhege, J. del R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, Eds. Cambridge, MA: MIT Press, 2007, pp. 207–233.

[39] C. Neuper and G. Pfurtscheller, "Evidence for distinct beta resonance frequencies in human EEG related to specific sensorimotor cortical areas," *Clin Neurophysiol*, vol. 112, pp. 2084–2097, 2001.

[40] G. Pfurtscheller, C. Brunner, A. Schlögl, and F. L. da Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," *NeuroImage*, vol. 31, no. 1, pp. 153–159, 2006.

[41] G. Pfurtscheller, G. R. Müller, J. Pfurtscheller, H. J. Gerner, and R. Rupp, "'thought' - control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia," *Neurosci Lett*, vol. 351, pp. 33–36, 2003.

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

[43] C. Neuper, G. R. Müller-Putz, R. Scherer, and G. Pfurtscheller, "Motor imagery and eeg-based control of spelling devices and neuroprostheses," *Prog Brain Res*, vol. 159, pp. 393–409, 2006.

[44] P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch, "EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales," *Electroencephalogr Clin Neurophysiol*, vol. 103, no. 5, pp. 499–515, 1997.

[45] F. Popescu, S. Fazli, Y. Badower, B. Blankertz, and K.-R. Müller, "Single trial classification of motor imagination using 6 dry EEG electrodes," *PLoS ONE*, vol. 2, no. 7, 2007. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0000637

[46] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K.-R. Müller, "Invariant common spatial patterns: Alleviating non-stationarities in brain-computer interfacing," in *Advances in Neural Information Processing Systems 20*. Cambridge, MA: MIT Press, 2008, accepted.

[47] M. Krauledat, M. Schröder, B. Blankertz, and K.-R. Müller, "Reducing calibration time for brain-computer interfaces: A clustering approach," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007, pp. 753–760.

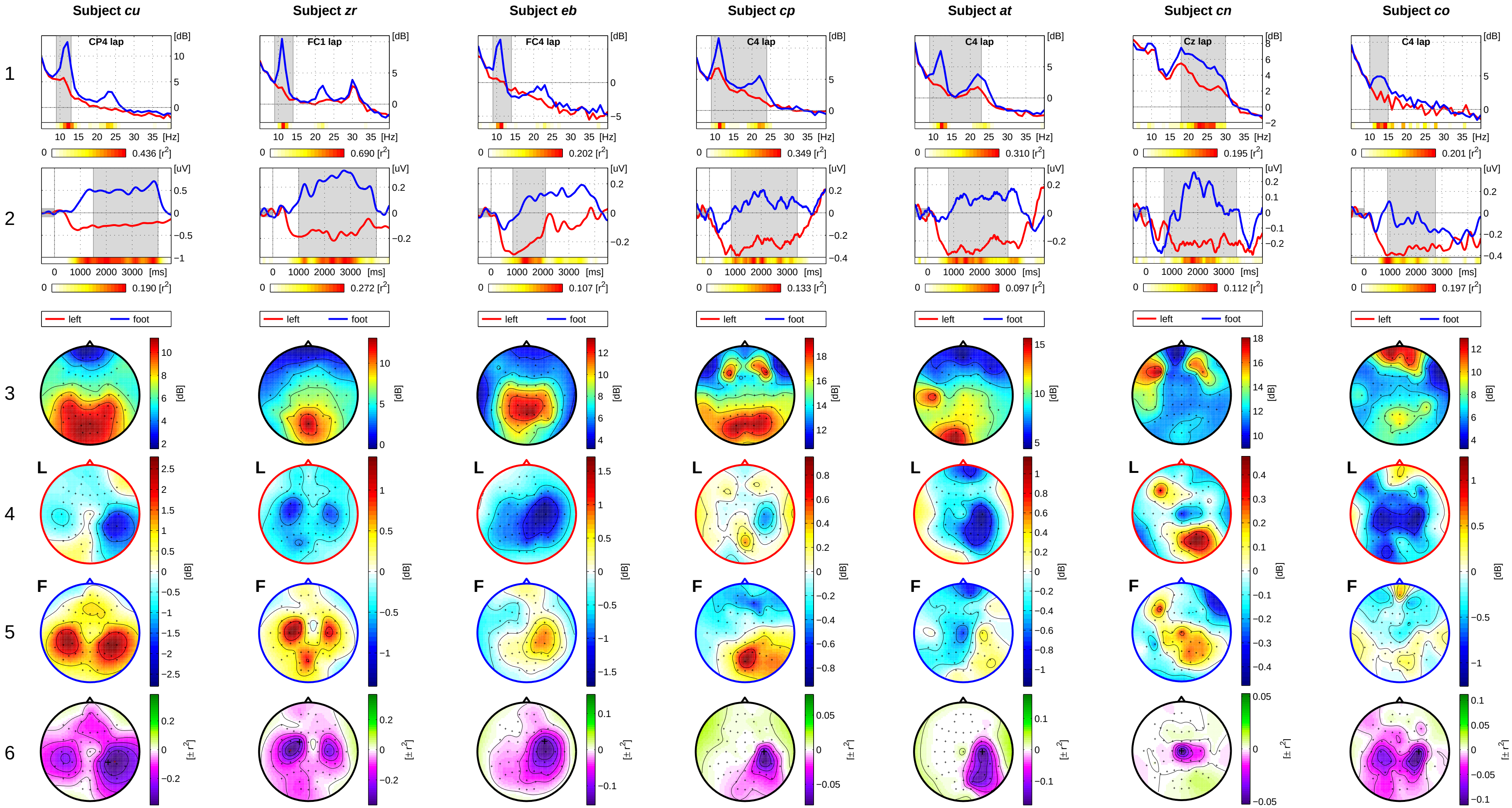
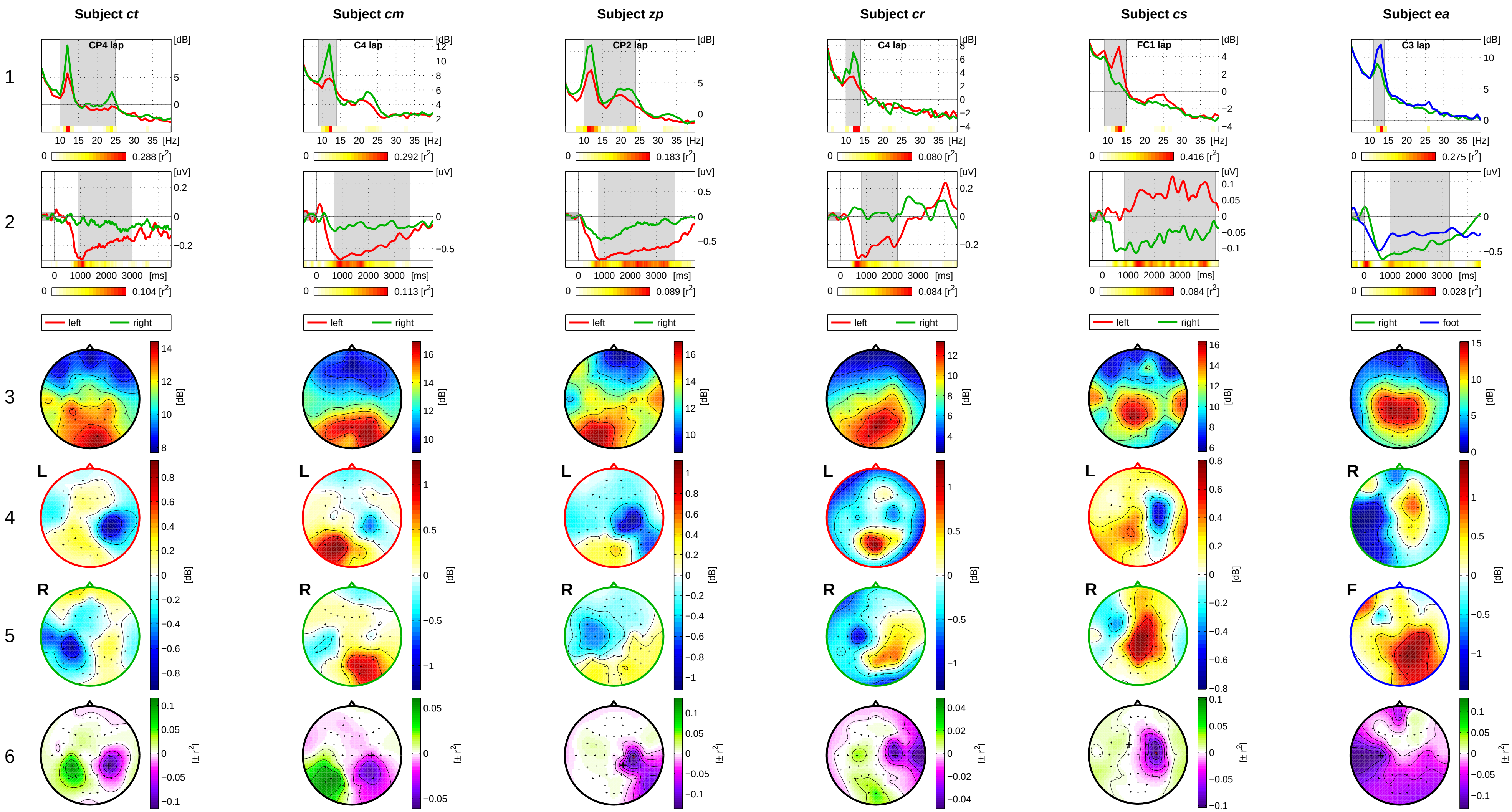


Fig. 3: The first row displays the averaged spectra of the two motor imagery tasks (red: left hand, green: right hand; blue: right foot) in the calibration measurement that have been used to train the classifier (subject *co* performed physical movements, see Section IV-A. The r^2-values of the difference between those conditions are color coded and the frequency band that has been chosen is shaded gray. The second row shows the average amplitude envelope of that frequency band with 0 being the time point of stimulus presentation in the calibration measurement. Spectra and amplitude envelopes are both shown for the Laplace filtered channel that is indicated in the top of the spectra subplot. The top scalp maps (row 3) show the log power within the chosen frequency band averaged over the whole calibration measurement. The fourth and fifth row display the log band power difference topographies of the particular motor imagery tasks (indicated by L, R, or F, respectively), from which the global average (shown in row 3) is subtracted. The bottom row (6) displays the r^2-values of the difference (row 4 minus row 5) between the individually chosen motor imagery tasks as scalp map.


Subject ct
Subject cm
Subject zp
Subject cr
Subject cs
Subject ea
1
2
3
4
5
6
CP4 lap
C4 lap
CP2 lap
C4 lap
FC1 lap
C3 lap
[dB]
[Hz]
0.288 [r²]
0.292 [r²]
0.183 [r²]
0.080 [r²]
0.416 [r²]
0.275 [r²]
[uV]
[ms]
0.104 [r²]
0.113 [r²]
0.089 [r²]
0.084 [r²]
0.084 [r²]
0.028 [r²]
left
right
foot
L
R
F
[± r²]