Subject independent mental state classification in single trials

Siamac Fazli[*,b], Florin Popescu[a], Márton Danóczy[b], Benjamin Blankertz[b], Klaus- Robert Müller[b], Cristian Grozea[a]

[a] *Fraunhofer First, Kekuléstr. 7, 12489 Berlin, Germany*
[b] *Technical University, Franklinstr. 28/29, 10587 Berlin, Germany*

Abstract

Current state of the art in Brain Computer Interfacing (BCI) involves tuning classifiers to subject-specific training data acquired from calibration sessions prior to functional BCI use. Using a large database of EEG recordings from 45 subjects, who took part in movement imagination task experiments, we construct an ensemble of classifiers derived from subject-specific temporal and spatial filters. The ensemble is then sparsified using quadratic regression with ℓ_1 regularization such that the final classifier generalizes reliably to data of subjects not included in the ensemble. Our offline results indicate that BCI-naïve users could start real-time BCI use without any prior calibration at only very limited loss of performance.

Key words: machine learning, BCI, zero-training

1. Introduction

Despite the greatly reduced calibration and subject-training time of BCI systems in the last decade, one of the primary goals of current BCI research is to deliver a ready-to-use system, requiring minimal effort to set up [8, 22, 3, 7]. Initial BCI systems were based on operant conditioning and required months of training on the subject side before it was possible to use them [1, 11]. Modern BCI systems require the subject to record a calibration session on which subject-specific, optimally discriminating time and spatial

*corresponding author
Email address: `fazli@cs.tu-berlin.de` (Cristian Grozea)

filters are learned, operating on classes such as EEG recordings of movement imaginations of the extremities [8, 10]. An algorithm was recently proposed which allows expert BCI users to start feedback sessions without any prior calibration session, based on a clustering approach and reuse of their own old training data to infer general patterns of spatial filters [15]. That study uses data of the same subject from earlier sessions to infer a classifier for a new session (session-to-session transfer). A further online-study has very recently confirmed the previous findings [17]. In the present paper we attempt to find a one-fits-all classifier that enables BCI experts as well as novel subjects to start a feedback session immediately.

Ensemble methods construct finite collections of individually weak classifiers from potentially very large ensembles [25, 21, 18]. Some recent use of ensemble methods as classification engines for EEG data have been reported in [28, 23, 12]. The main contribution of this paper is not the use of an ensemble method for BCI, rather it is the first successful attempt to achieve a *subject-independent zero-training BCI*. In our particular case, we operated on the general hypothesis that a very sparse set of features (spatial patterns and frequency bands of cortical EEG) are relevant to movement imagination detection[2]. We therefore chose quadratic regression coupled with an ℓ_1 norm penalty on the weights, which is known to allow a high level of sparsification [29, 26, 20], and concentrated on a judicious choice of penalty hyper-parameter, which would be reliable in terms of generalization error. In this paper we call this regression ℓ_1 regularized regression. To this end 'novel subject' experiments were simulated by removing each particular subject entirely from the training data and testing on that subject. We further confirmed the efficacy of our sparsified ensemble classifier by applying it on a hold out data set, consisting of 53 datasets from 29 subjects. None of these subjects were part of the original dataset.

The paper first provides an overview of the data used, then the ensemble learning algorithm is outlined, consisting of the procedure for building the filters, the classifiers and the gating function. Subsequently we provide results on all subjects and then discuss good subjects. Interestingly we are able to successfully classify trials of novel subjects with zero training suffering only a small loss in performance. Finally we put our results into perspective.

2. Available Data and Experiments

We used 90 BCI datasets (sessions), ranging between 70 to 600 trials each, from 45 individual subjects. Each trial consists of one of two predefined movement imaginations, being left and right hand, i.e. data was chosen such that it relies only on these 2 classes, although originally three classes were cued during the calibration session, being left hand (L), right hand (R) and foot (F). 45 EEG channels, which are in accordance with the 10-20 system, were identified to be common in all sessions considered. The data were recorded while subjects were immobile, seated on a comfortable chair with arm rests. The cues for performing a movement imagination were given by visual stimuli, and occurred every 4.5-6 seconds in random order. Each trial was referenced by a 3 second long time-window starting at 500 msec after the presentation of the cue. Individual experiments consisted of two different training paradigms. The first training paradigm consisted of visual cues in form of a letter. In the second training paradigm the subject was instructed to follow a moving target on the screen. Within this target the edges lit up to indicate the type of movement imagination required. For further details of the exact experimental procedure we would like to refer the reader to [4].

Electromyogram (EMG) on both forearms and the foot were recorded as well as electrooculogram (EOG) to ensure there were no real movements of the arms and that the movements of the eyes were not correlated to the required mental tasks.

Since poorly BCI performing subjects were seldom invited to perform further experiments, there is a considerable bias within our dataset towards subjects for whom BCI generally works well. Since we are interested in achieving real zero-training for the whole range of subjects and not only for expert users, we compensated by always considering the subject mean instead of the session mean, when comparing error-rates. Please refer to Table 1 for the distribution of sessions per subject.

3. Generation of the Ensemble

The ensemble consists of a large set of subject-dependent common spatial pattern filters (CSP) and their matching classifiers (LDA). Each dataset is first preprocessed by nine predefined temporal filters (i.e. band-pass filters) in parallel (see upper panel of Figure 1). A corresponding spatial filter and linear classifier is obtained for every dataset and temporal filter. Each resulting

number of datasets/subject	1	2	3	5	8	9	13
occurrence	32	5	3	2	1	1	1
percentage [%]	39.0	11.0	11.6	12.5	6.8	8.0	11.1

Table 1: The first row gives the numbers of experiments that exist for a single subject, while the second row shows how often this occurs, i.e. 32 subjects carried out only a single experiment each, 5 subjects carried out 2 experiments each, and so on. The third row shows the percentage of trials that fall into each category.

CSP-LDA couple can be interpreted as a potential basis function. Finding an appropriate weighting for the classifier outputs of these basis functions is of paramount importance for the accurate prediction. We employed different forms of regression, such as ℓ_1 regularized regression, in order to find an optimal but sparse combination weighting for predicting the movement imagination data of unseen subjects[2, 5]. This processing was done by leave-one-subject-out cross-validation, i.e. all sessions of a particular subject were removed, the regression trained on the remaining trials and then applied to this subject's data (see lower panel of Figure 1).

3.1. Temporal Filters

The μ-rhythm (9-14 Hz) and synchronized components in the β-band (16-22 Hz) are macroscopic idle rhythms that prevail over the postcentral somatosensory cortex and precentral motor cortex, when a given subject is at rest. Imaginations of movements as well as actual movements are known to suppress these idle rhythms contralaterally. However, there are not only subject-specific differences of the most discriminative frequency range of the mentioned idle-rhythms, but also session differences thereof.

We identified 9 neurophysiologically relevant temporal filters. Their individual performance over our entire dataset can be seen in Table 2. In all following performance related tables we used the percentage of misclassified trials, or 0-1 loss. As expected from general knowledge of neurophysiology, the μ-band related filters score best across subjects as well as sessions. Note however, that a temporal filter, which on average performs poorly across sessions may still yield high performance on the sessions of a particular subject.

Session means can be unbalanced, since a particular subject may have taken part in multiple sessions, as can be seen from Table 1. We therefore considered subjects means only for our further analysis.

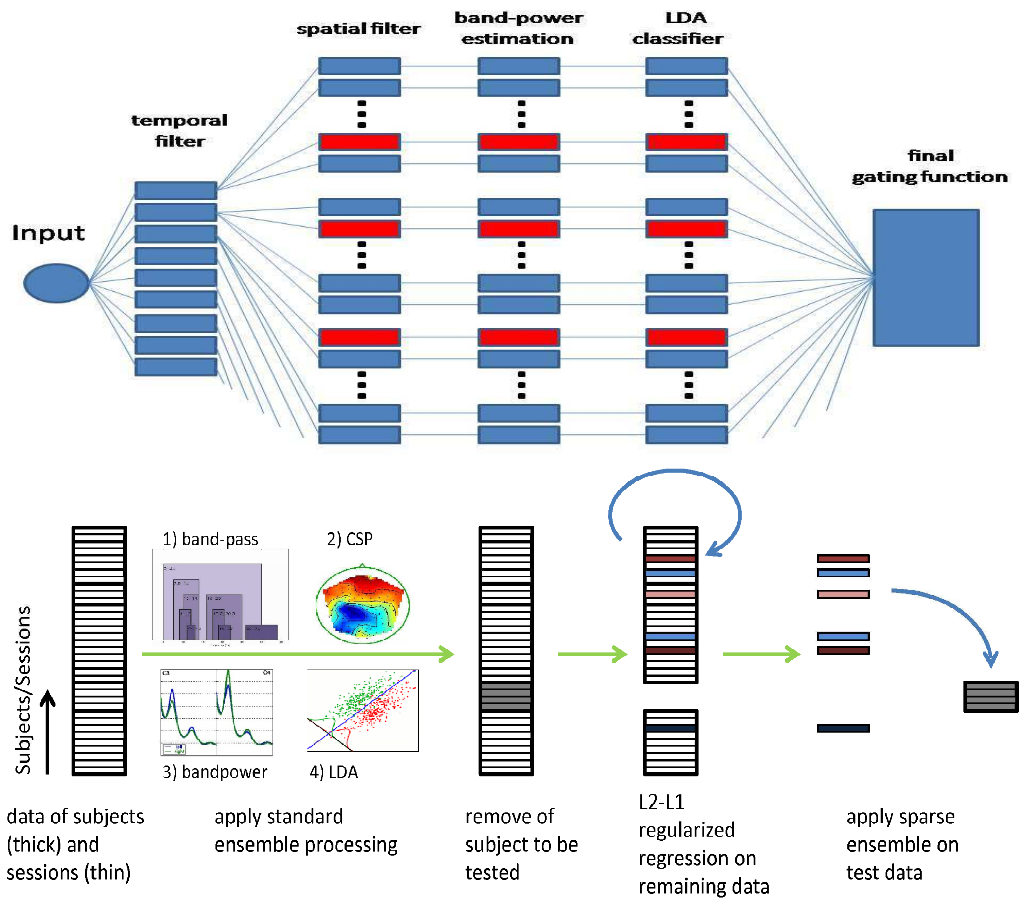


Figure 1: 2 Flowcharts of the ensemble method. The red patches in the top panel illustrate the inactive nodes of the ensemble after sparsification.

3.2. Spatial Filters and Classifiers

CSP is a very widely used algorithm for calculating spatial filters, used for detecting event-related (de-)synchronization (ERD/ERS), and is considered to be the gold-standard of ERD-based BCI systems [14, 24]. The CSP algorithm maximizes the variance of right hand trials, while simultaneously minimizing the variance for left hand trials. Given the two covariance matrices Σ_1 and Σ_2, of size *channels* x *concatenated timepoints*, the CSP algorithm returns the matrices W and D. W is a matrix of projections, where the i-th row has a relative variance of d_i for trials of class 1 and a relative variance

	subject mean		session mean	
frequency [Hz]	loss±std [%]	best [%]	loss±std [%]	best [%]
$7.5 - 14$	$\mathbf{30.0 \pm 8.0}$	**20.0**	$\mathbf{23.3 \pm 7.8}$	17.6
$11 - 13$	33.1 ± 8.4	15.6	27.3 ± 8.2	17.6
$10 - 14$	30.8 ± 8.3	15.6	23.4 ± 7.8	**20.9**
$9 - 12$	32.3 ± 7.9	11.1	25.4 ± 7.7	17.6
$19 - 22$	42.8 ± 8.8	2.2	39.2 ± 8.9	1.1
$16 - 22$	39.2 ± 8.4	8.9	34.5 ± 8.5	7.7
$26 - 34$	45.7 ± 8.7	6.6	44.9 ± 9.3	3.3
$17.5 - 20.5$	43.3 ± 8.5	2.2	39.3 ± 8.9	2.2
$7 - 30$	30.4 ± 8.4	17.8	24.1 ± 8.5	12.1
auto-band	$\mathbf{28.9 \pm 7.4}$	–	$\mathbf{22.9 \pm 7.5}$	–

Table 2: Mean performance of all temporal filters, included for the ensemble. For comparison we included an automatic heuristic that searches for an optimal, subject-dependent temporal filter, which has been proven to yield favorable results [8]. The first column shows, which frequency band was used. The second and fourth columns represent the mean of the cross-validated loss over each subject and session, respectively. The third and fifth columns show how often the given frequency band performed best over subjects and sessions, respectively.

of $1 - d_i$ for trials of class 2. D is a diagonal matrix with entries $d_i \in [0, 1]$, with length n, the number of channels:

$$W\Sigma_1 W^T = D \qquad \text{and} \qquad W\Sigma_2 W^T = I - D \tag{1}$$

Best discrimination is provided by filters with very high or very low eigenvalues, we therefore chose to only include projections with the highest 2 and corresponding lowest 2 eigenvalues for our analysis. For a thorough review of the CSP algorithm we would like to refer the reader to [8].

We used a linear classifier, since non-linear classification methods on these features could not be observed to have any statistically significant gain for the given experimental setup [19]. We therefore chose Linear Discriminant Analysis (LDA) [6]. Each time filtered session corresponds to a CSP set and to a matched LDA.

3.3. Final gating function

The final gating function is responsible for combining the outputs of the individual ensemble members to a single one and this can be realized in many ways. For a number of ensemble methods the mean has proven to be a surprisingly good choice [21]. As a first try and baseline for our ensemble we simply averaged all outputs of our individual classifiers. This result is given as *mean* in Table 3. Furthermore we employed various forms of regression to the outputs of the ensemble. The implementation of these regressions were performed using `CVX`, a package for specifying and solving convex programs [13]. We coupled an ℓ_2 loss with an ℓ_1 penalty term on a linear voting scheme ensemble.

$$\begin{aligned}\underset{w_{ij}^{(k)}}{\operatorname{argmin}} & \sum_{x \in X \setminus X_k} (h_k(x) - y(x))^2 \\ & + \alpha \sqrt{\sum_{i=1}^{B} \sum_{j \in S \setminus S_k} \sum_{x \in X \setminus X_k} c_{ij}(x)^2} \left(\sum_{i=1}^{B} \sum_{j \in S \setminus S_k} |w_{ij}^{(k)}| + |b| \right)\end{aligned} \tag{2}$$

$$h_k(x) = \sum_{i=1}^{B} \sum_{j \in S \setminus S_k} w_{ij}^{(k)} c_{ij}(x) - b \tag{3}$$

where $c_{ij}(x) \in [-\infty; \infty]$ is the continuous classifier output, before thresholding, obtained from the session j by applying the bandpass filter i, B is the number of frequency bands, S the complete set of sessions, X the complete data set, S_k the set of sessions of subject k, X_k the dataset for subject k, $y(x)$ is the class label of trial x and w_{ij}^{k} in equation (3) are the weights given to the LDA outputs.

The hyperparameter α in equation (2) was varied on a logarithmic scale and multiplied by a dataset scaling factor which accounted for fluctuations in voting population distribution and size for each subject. The dataset scaling factor is computed using $c_{ij}(x)$, for all $x \in X \setminus X_k$. For computational efficiency reasons the hyperparameter was tuned on a small random subset of subjects whose labels are to be predicted from data obtained from other subjects such that the resulting test/train error ratio was minimal, which in turn affected the choice (leave in/out) of classifiers among the 90x9 candidates. The ℓ_1 regularized regression with this choice of α was then applied to all subjects, with results (in terms of feature sparsification) shown in Figure 3. The final sparsified ensemble used classifiers which were chosen for prediction of performance in at least 20% of subjects.

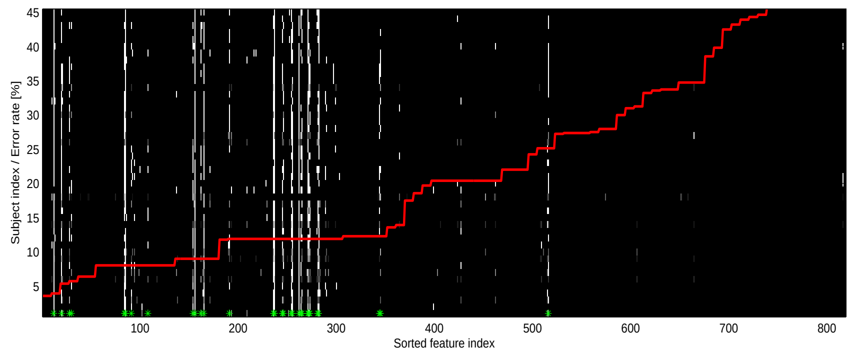
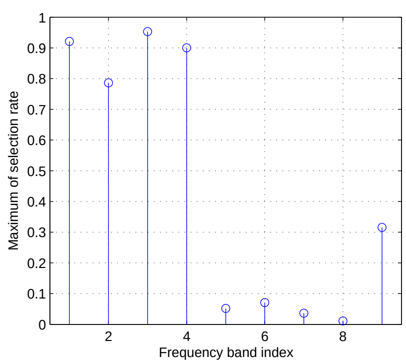


Figure 2: Feature selection during cross-validation: white dashes mark the features kept after regularization for the prediction of the data of each subject. The numbers on the vertical axis represent the subject index as well as the Error Rate (%). The red line depicts the baseline error of individual subjects (classical auto-band CSP). Features as well as baseline errors are sorted by the magnitude of the subject error. Green dots denote the features selected for prediction in more than 20% of the subjects. Note that some of the features are useful in predicting the data of most other subjects, while some are rarely or never used.



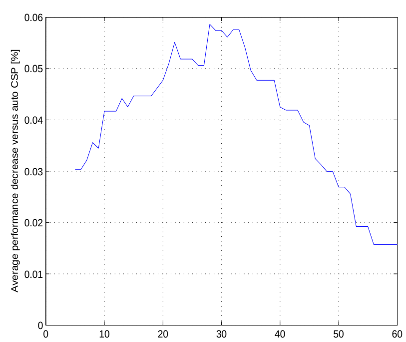


Figure 3: On the left: The maximum rate of selection of all features corresponding to each frequency band over subjects. The individual frequencies can be seen in Table 2. On the right: The average increased loss of the ensemble with respect to the classical auto-band CSP performance of the individual subjects.

3.4. Validation

The subject-specific CSP-based classification methods with broad-band and automatically, subject-dependent tuned temporal filters (termed reference methods) are validated by an 8-fold cross-validation, splitting the data chronologically. The chronological splitting for cross-validation is a common practice in EEG classification, since the non-stationarity of the data is thus preserved.
To validate the quality of the ensemble learning we employed a leave-one-subject out cross-validation (LOSO-CV) procedure. For predicting the labels of a particular subject we only use data from other subjects. We furthermore test our obtained general classifier on a hold-out set of subjects, not included in the ensemble set.

4. Results

Overall performance of the reference methods, other baseline methods and of the ensemble method is presented in Table 3. Reference method performances of subject-specific CSP-based classification are presented with both broad-band and heuristically tuned frequency bands (auto-band). Furthermore we considered much simpler (zero-training) methods as a control. Laplacian stands for the power difference in two Laplace filtered channels (C3 vs. C4) and simple band-power stands for the power difference of the same two channels without any spatial filtering. For the simple zero-training methods we chose a broad-band filter of $7 - 30$Hz, since it is the least restrictive and scored one of the best performances on a subject level, as can be seen from Table 2.
The bias b in equation (3) can be tuned broadly for all sessions or corrected individually by session, and implemented for online experiments in multiple ways [17, 27, 16]. In our case we chose to adapt b without label information, but operating under the assumption that class frequency is balanced. We therefore simply subtracted the mean over all trials of a given session. This procedure is termed (bc) in Table 3.
In Figure 5, we show overall classification results on training data and hold-out data highlighting relative performance of ensemble classification versus the auto-band reference method (zero-training versus subject-dependent training) over both subjects and sessions.

ensemble zero-training		
method	multi-band	broad-band
mean	36.0	36.5
least square regression	35.3	37.8
ℓ_1 regularized regression	33.3	37.3
mean (bc)	31.0	32.1
least square regression (bc)	32.3	32.7
ℓ_1 regularized regression (bc)	**30.1**	**31.8**
baselines – with subject-dependent training		
classical auto-band CSP	**28.9**	
broad-band CSP	30.4	
baselines – without subject-dependent training		
Laplacian	40.7	
simple band-power	40.5	
Laplacian (bc)	36.3	
simple band-power (bc)	38.5	

Table 3: Final results of the ensemble classifiers and two baseline classifiers. mean stands for simply averaging the ensemble outputs, least square regression for learning the linear combination weights through least-squares regression, ℓ_1 regularized regression as according to equations (2) and (3), (bc) stands for bias corrected outputs.

4.1. Ensemble with good subjects only, tested on all subjects

In order to find out whether a lower number of subjects who generally perform well with BCI would be sufficient for the given task of classifying data from unseen subjects, we chose to only include datasets with a auto-band reference method cross-validation error of less than θ. Where $\theta = 0.1, 0.2, ..., 0.5$. The results of the LOSO-CV can be seen in Figure 4.
Two versions were implemented. The three top panels show the results, if solely one broad-band temporal filter is used to generate the ensemble, while the lower three panels describe the results of the full ensemble with all 9 temporal filters present. For both versions the standard least-squares regression (LSR) overfits on the given training data. Given the multi-band version of the ensemble, this effect becomes even more apparent. The ℓ_1 regularized regression performs favorably, as can be seen on the middle panels. The rightmost panel shows the number of features chosen by the ℓ_1 regularized regression, given all basis classifiers. For the single-band version the average chosen number of features rises, even if only classifiers are added, which do not perform very well on the auto-band reference method cross-validation score. For the multi-band approach we can observe a saturation of features, that occurs, when datasets with more than 30% cross-validation error are added to the ensemble. It is therefore not sufficient to operate the ensemble with a single temporal filter or the data of well performing subjects only, but necessary to maintain the full set of features in order to obtain optimal results.

4.2. Performance on good subjects

BCI performance generally varies to a high degree from subject to subject. For subjects, where discriminability of mental tasks is high, subject-dependent tuning of temporal and spatial filters is of paramount importance to ensure optimal performance [10]. To examine how well our presented method performs on these subjects, we chose to generate a subset of subjects, whose average loss was less than 20% in any of the following methods: classical auto-band CSP, Laplacian, simple bandpower and ℓ_1 regularized regression. Not surprisingly classical auto-band CSP still performs best as compared to all three other blind methods, since the tuning of the temporal and spatial filters is done in a subject-dependent manner. However, the proposed method outperforms simpler subject-independent methods by far. The results of all four methods are given in Table 4.

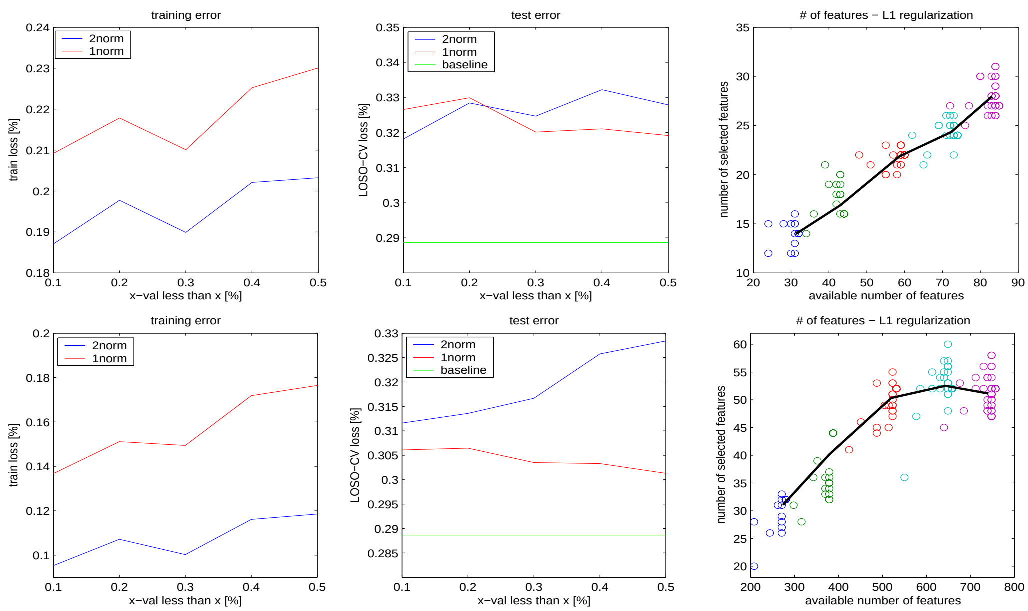


Figure 4: The two left most panels show the training and test errors for a simple least-squares regression and for equations (2) and (3), denoted as 2norm and 1norm, respectively. We chose to stepwise increase the number of datasets, included in the ensemble, according to their auto-band reference method cross-validation loss (x-axis). The rightmost panel shows the number of features, that were selected by the ℓ_1 regularized regression. The individual colors code for the stepwise increases of datasets available for the regression. The black line denotes the subject mean of the selected number of features.

	ℓ_1	CSP	lap	bp
# <20% loss	11	14	6	2
25%-tile	10.5	6.4	18.6	22.8
median	12.7	8.7	22.6	27.1
75%-tile	18.2	13.7	28.2	32.1

Table 4: The first row shows how many subjects scored a loss of less than 20% for the given method. Row 2-4 show median and percentile errors of the union of subjects scoring less than 20% in any of the given methods. ℓ_1 stands for ℓ_1 regularized regression.

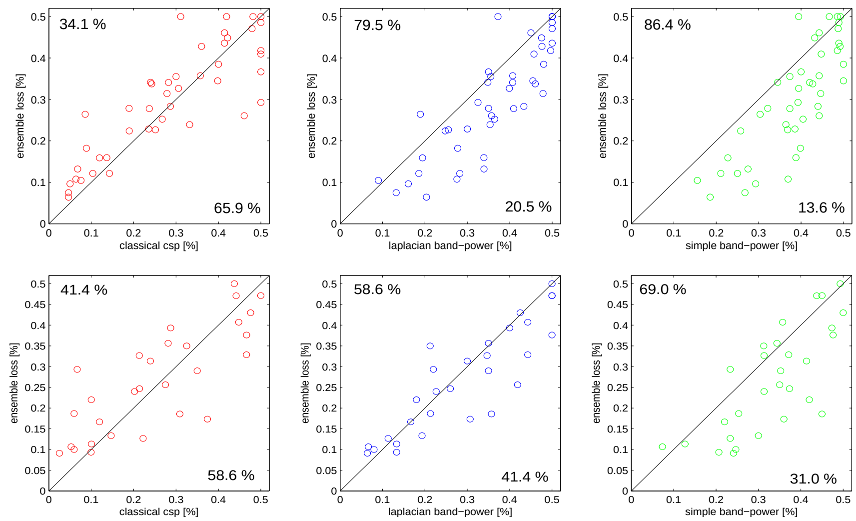


Figure 5: Results of the LOSO-CV ensemble loss versus various baselines (top panels) and the ensemble loss of the hold-out set (bottom panels). Left panel: versus classical auto-band CSP, cross-validated with subject-dependent training. Middle panel: versus Laplacian filtered band-power. Right panel: versus simple band-power. The percentages show how often a given method performed advantageously.

4.3. Focusing on a particular subject

In order to exemplify how the ensemble works in detail we will focus on a particular subject. We chose to use the subject with the lowest auto-band reference method cross-validation error (10%). Given the non-linearity in the band-power estimation (see Figure 1) it is impossible to picture the resulting ensemble spatial filter exactly. However, by averaging the chosen CSP filters with the weightings, obtained by the ensemble and multiplying them by their LDA classifier weight, we get an approximation:

$$P_{ENS} = \sum_{i=1}^{B} \sum_{j \in S \setminus S_k} w_{ij} W_{ij} C_{ij} \tag{4}$$

where w_{ij} is the weight matrix, resulting from the ℓ_1 regularized regression, given in equations (2) and (3), W_{ij} the CSP filter, corresponding to temporal filter i and subject j and C_{ij} the LDA weights (B in Figure 6). For the case of classical auto-band CSP this simply reduces to $P_{CSP} = WC$ (A in Figure 6). Another way to exemplify the ensemble performance is to refer to a transfer function. By injecting a sinusoid with a frequency within the corresponding band-pass filter into a given channel and processing it by the four CSP filters, estimating the bandpower of the resulting signal and finally combining the four outputs by the LDA classifier, we obtain a response for the particular channel, where the sinusoid was injected. Repeating this procedure for each channel results in a response matrix. This procedure can be applied for a single CSP/LDA pair, however we may also repeat the given method for as many times as features were chosen for a given subject by the ensemble and hence obtain an accurate description of how the ensemble processes the given EEG data. The resulting response matrices are displayed in panel C of Figure 6. While the subject-specific pattern (classical) looks less focused and more diverse the general pattern matches the one obtained by the ensemble. A third way of visualizing how the ensemble works, we show the primary projections of the CSP filters that were given the 6 highest weights by the ensemble on the left panel F and the distribution of all weights in panel D. The spatial positions of highest channel weightings differ slightly for each of the CSP filters given, however the maxima of the projection matrices are mostly positioned around the primary motor cortex. In panel G the outputs of all basis classifiers are applied to each trial of one subject. The top row (broad) gives the label, the second row (broad) gives the output of the classical auto-band CSP, and each of the following rows (thin) gives the outputs of the individual classifiers of other subjects. The individual classifier outputs are sorted by their correlation coefficient with respect to the class labels. The trials (columns) are sorted by true labels with primary key and by mean ensemble output as a secondary key. The row at the bottom gives the sign of the average ensemble output.

4.4. Validation on a hold-out set

A hold-out set of 53 sessions, comprised of 29 subjects, of which none were included in the dataset for deriving the general classifier, was used as means of validation, to ensure our learning algorithm does not overfit. In other words, these 53 sessions of 29 subjects are not a subset of the original dataset, but consist of different subjects. For our best performing model, produced by ℓ_1 regularized regression, with bias removal, the average subject performance was 26.7%, and the average session performance was 28.3%. For comparison, the standard processing of cross-validating CSP with optimal frequency bands led to 25.4% and 28.5% for subject and session means, respectively. For the zero-training methods of Laplace filtered channels we

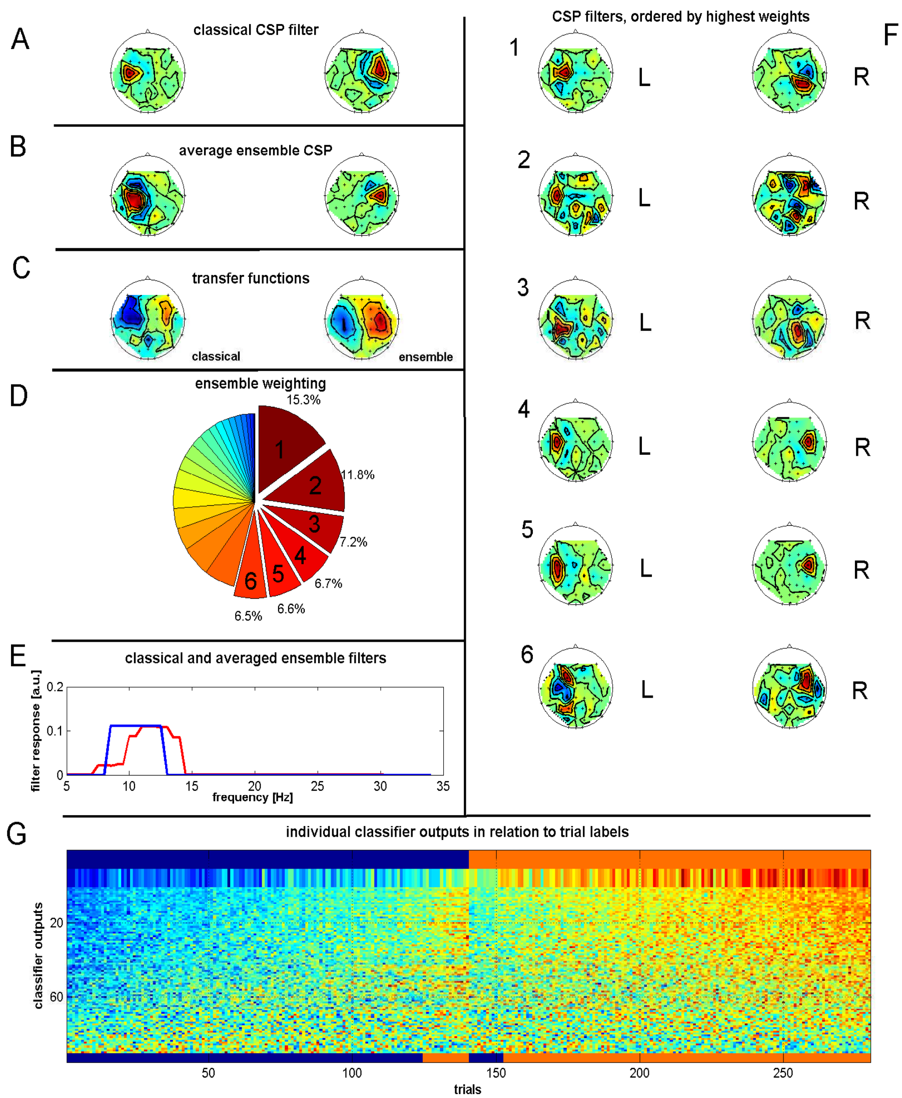


Figure 6: A: primary projections for classical auto-band CSP. B: linearly averaged CSP's from the ensemble. C: transfer function for classical auto-band and ensemble CSP's. D: weightings of 28 ensemble members, the six highest components are shown in F. E: linear average ensemble temporal filter (red), heuristic (blue). F: primary projections of the 6 ensemble members that received highest weights. G: Broad-band version of the ensemble for a single subject.

obtained 29.3% and 30.0% for subject and session means and for simple bandpower 33.5% and 35.7% of the trials were misclassified. The detailed results for each individual subject may be obtained from the lower part of Figure 5. Our hold-out set therefore confirms the quality of findings we previously achieved by the leave-one-subject-out cross-validation, as can be seen in Table3.

5. Discussion

The practical significance of our result is clear: BCI use, by dispensing with calibration sessions, would become more widely accessible to those users who are inherently proficient at it (i.e. are BCI 'literate'). We have taken great care in this work to exclude data from a given subject in predicting his/her performance, and kept a reasonable number of subjects out of our analysis entirely as a validation set, on which prediction was successful. In contrast with previous work on ensemble approaches to BCI classification based on simple majority voting and Adaboost [30, 9] utilizing only a limited dataset, we have been aided greatly by the laborious accumulation of BCI experiments over the years. This has enabled us to choose a very sparse set of voting classifiers which performed as well as standard, state-of-the-art subject calibrated methods. ℓ_1 regularized regression in this case performed better than other methods (such as majority voting) which we have also tested. Note that, interestingly, the chosen features (see Figure 2), do not exclusively come from the best performing subjects, in fact an average performer was also selected. However most white dashes are present in the left half, i.e. most subjects with high auto-band reference method performance were selected. Interestingly some subjects with very high BCI performance are not selected at all, while others generalize well in the sense that they are able to predict other subject's data.

No single frequency band dominated classification accuracy – see Figure 3. Therefore, the regularization must have selected diverse features. Nevertheless, as can be seen in panel G of Figure 6 there is significant correlation between classifiers in the ensemble. Our approach of finding a sparse solution reduces the dimensionality of the chosen features significantly, For very good subjects the ensemble does not always perform as well as auto-band reference method classifiers. We show on the right side of Figure 3 that the average error of the ensemble never exceeds an increased loss of more than

6%. However this increased loss for very able subjects will not prevent them from performing successfully in BCI.

The sparsification of classifiers, in this case, also leads to potential insight into neurophysiological processes. It identifies relevant cortical locations and frequency bands of neuronal population activity which are in agreement with general neuroscientific knowledge. While this work concentrated on zero training classification and not brain activity interpretation, a much closer look is warranted. Movement imagination detection is not only determined by the cortical representation of the limb whose control is being imagined (in this case the arm) but also by differentially located cortical regions involved in movement planning (frontal), execution (fronto-parietal) and sensory feedback (occipito-parietal). Patterns relevant to BCI detection appear in all these areas and while dominant discriminant frequencies are in the α range, higher frequencies appear in our ensemble, albeit in combination with less focused patterns.

In this study we show that by taking advantage of a large dataset from a high number of subjects and by applying the appropriate machine learning tools, it is possible to classify data from unseen subjects with similar accuracies as compared to techniques which are dependent on subject specific data.

6. Conclusions and Outlook

We have shown that a subject-independent zero training procedure is possible for ERD-based BCI. A general classifier that allows instantaneous feedback sessions for BCI experts as well as for naïve BCI subjects has been produced.

Further online studies will be needed to add more experimental evidence in support of our findings. However, a preliminary online experiment was carried out in which the broad-band version of the classifier was tested on a single subject, which was not included in the ensemble, but known to be a BCI literate. Two feedback sessions of 40 trials each were carried out: one without prior calibration and one with. EEG data and a video were recorded. The correct class was hit 38 out of 40 times in both sessions , i.e. an accuracy of 95% was achieved. This gives a first anecdotal evidence that our novel ensemble method will be transferable to online scenarios. Imagining movements over a longer time-period, without feedback, can present difficulties for some subjects. It may also lead to non-stationarities and signal differences to the following feedback session, since the visual feedback may induce conflicting

neural mechanisms. Starting a feedback session straightaway may help to alleviate those effects and lead to more robust BCI.
We furthermore consider to extend the presented static framework to an adaptive one, where the weights of the individual classifiers can adapt, via a supervised and even a reinforcement learning scheme. We plan to adopt the ensemble approach in combination with a recently developed EEG cap having dry electrodes [22] and thus being able to reduce the required preparation time for setting up a running BCI system to essentially zero.

References

[1] Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., Flor, H., 1999. A spelling device for the paralysed. Nature 398, 297–298.

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

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

[4] Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., Kunzmann, V., Losch, F., Curio, G., 2006. The Berlin Brain-Computer Interface: EEG-based communication without subject training. IEEE Trans Neural Syst Rehabil Eng 14, 147–152.

[5] Blankertz, B., Dornhege, G., Lemm, S., Krauledat, M., Curio, G., Müller, K.-R., 2006. The Berlin Brain-Computer Interface: Machine learning based detection of user specific brain states. Journal of Universal Computer Science 12, 2006.

[6] Blankertz, B., Dornhege, G., Schäfer, C., Krepki, R., Kohlmorgen, J., Müller, K.-R., Kunzmann, V., Losch, F., 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.

[7] Blankertz, B., Losch, F., Krauledat, M., Dornhege, G., Curio, G., Müller, K.-R., 2008. The Berlin Brain–Computer Interface: accurate performance from first-session in BCI-naïve subjects. IEEE Trans Biomed Eng 55, 2452–2462.

[8] Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., Müller, K.-R., 2008. Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Proc Magazine 25 (1), 41–56.

[9] Boostani, R., Moradi, M. H., 2004. A new approach in the BCI research based on fractal dimension as feature and adaboost as classifier. J. Neural Eng. 1, 212–217.

[10] Dornhege, G., Millán, J. R., Hinterberger, T., McFarland, D., Müller, K.-R. (Eds.), 2007. Toward Brain-Computer Interfacing. Cambridge, MA: MIT Press.

[11] Elbert, T., Rockstroh, B., Lutzenberger, W., Birbaumer, N., 1980. Biofeedback of slow cortial potentials. I. Electroencephalogr. Clin. Neurophysiol. 48, 293–301.

[12] Fatourechi, M., Ward, R. K., Birch, G. E., 2008. A self-paced brain-computer interface system with a low false positive rate. J. Neural Eng 5, 9–23.

[13] Grant, M., Boyd, S., 2008. CVX: Matlab software for disciplined convex programming (web page and software). http://stanford.edu/ boyd/cvx.

[14] Koles, Z., Soong, A. C. K., 1998. EEG source localization: implementing the spatio-temporal decomposition approach. Electroencephalogr. Clin Neurophysiol 107, 343–352.

[15] Krauledat, M., Schröder, M., Blankertz, B., Müller, K.-R., 2007. Reducing calibration time for brain-computer interfaces: A clustering approach. In: Schölkopf, B., Platt, J., Hoffman, T. (Eds.), Advances in Neural Inf. Proc. Systems (NIPS 07). Vol. 19. pp. 753–760.

[16] Krauledat, M., Shenoy, P., Blankertz, B., Rao, R., Müller, K.-R., 2007. Adaptation in csp-based BCI systems. In: Toward Brain-Computer Interfacing. MIT Press, pp. 305–309.

[17] Krauledat, M., Tangermann, M., Blankertz, B., Müller, K.-R., 2008. Towards zero training for brain-computer interfacing. PLoS ONE 3, e2967.

[18] Meir, R., Rätsch, G., 2003. An introduction to boosting and leveraging. Springer-Verlag New York, Inc., New York, NY, USA, pp. 118–183.

[19] Müller, K.-R., Anderson, C. W., Birch, G. E., 2003. Linear and nonlinear methods for brain-computer interfaces. Neural Systems and Rehabilitation Engineering, IEEE Transactions on 11 (2), 165–169.

[20] Müller, K.-R., Mika, S., Rtsch, G., Tsuda, S., Schlkopf, B., 2001. An introduction to kernel-based learning algorithms. IEEE Transactions on Neural Networks 12 (2), 181–202.

[21] Polikar, R., 2006. Ensemble based systems in decision making. IEEE Circuits and Systems Magazine 6 (3), 21–45.

[22] Popescu, F., Fazli, S., Badower, Y., Blankertz, B., Müller, K.-R., 2007. Single trial classification of motor imagination using 6 dry EEG electrodes. PLoS ONE 2, e637.

[23] Rakotomamonjy, A., Guigue, V., 2008. BCI competition III: dataset II - ensemble of svms for bci P300 speller. IEEE Trans Biomed Eng 55 (3), 1147–1154.

[24] Ramoser, H., Müller-Gerkin, J., Pfurtscheller, G., 2000. Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans. Rehab. Eng 8 (4), 441–446.

[25] Rätsch, G., Demiriz, A., Bennett, K., 2002. Sparse regression ensembles in infinite and finite hypothesis spaces. Mach. Learn. 48 (1-3), 189–218.

[26] Rätsch, G., Mika, S., Warmuth, M. K., 2001. On the convergence of leveraging. In: Dietterich, T. G., Becker, S., Ghahramani, Z. (Eds.), NIPS. MIT Press, pp. 487–494.

[27] Shenoy, P., Krauledat, M., Blankertz, B., Rao, R., Müller, K.-R., 2006. Towards adaptive classification for BCI. Journal of Neural Engineering 3 (1), R13–R23.

[28] Sun, S., Zhang, C., Zhang, D., 2007. An experimental evaluation of ensemble methods for EEG signal classification. Pattern Recognition Letters 28 (15), 2157–2163.

[29] Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B 58 (1), 267–288.

[30] Wang, S., Lin, Z., Zhang, C., 2005. Network boosting for BCI applications. Book Series Lecture Notes in Computer Science 3735, 386–388.