On optimal channel configurations for SMR-based brain-computer interfaces

Claudia Sannelli[a], Thorsten Dickhaus[a], Sebastian Halder[c], Eva-Maria Hammer[c], Klaus-Robert Müller[a], Benjamin Blankertz[a,b]

[a]Machine Learning Laboratory, Berlin Institute of Technology, Berlin, Germany
[b]Intelligent Data Analysis Group, Fraunhofer FIRST, Berlin, Germany
[c]Institute of Medical Psychology and Behavioral Neurobiology, Universität Tübingen, Germany

Correspondence: C. Sannelli, Berlin Institute of Technology, Franklinstr., 28/29. 10587 Berlin, Germany. E-mail: claudia.sannelli@tu-berlin.de, phone +49 30 31478624, fax +49 30 31478622

Abstract. One crucial question in the design of EEG-based brain-computer interface (BCI) experiments is the selection of channels. While a setup with few channels is more convenient and requires less preparation time, a dense placement of electrodes provides more detailed information and henceforth could lead to a better classification performance. Here, we investigate this question for a specific setting: a BCI that uses the popular CSP algorithm in order to classify voluntary modulations of sensorimotor rhythms. In a first approach 13 different fixed channel configurations are compared to the full one. The configuration with 48 channels results to be the best one, while configurations with less channels, from 32 to eight, performed not significantly worse than the best configuration in cases of less trials available. In a second approach an optimal channel configuration is obtained by an iterative procedure. As a surprising result, in the second approach a setting with 17 channels centered over the motor areas were selected. Thanks to the acquisition of a large data set recorded from 80 novice subjects using 119 EEG channels, the results of this study can be expected to have a high degree of generalizability.

Keywords: Brain-Computer Interface, Common Spatial Patterns, Channel Selection, EEG, Machine Learning.

1. Introduction

The question of which channel configuration to use is crucial in the design of a BCI system. While the aim to achieve the best possible performance might bias the decision towards the usage of many channels, practical considerations like the ease of preparation and the comfort of the user favor configurations with few channels. Surprisingly, so far there are no large-scale studies which investigate in what way the performance of a system depends on its number of channels. Obviously, there will be no general answer to this question, as it strongly depends on the type of BCI system. Here, we limit the investigation to BCIs which are driven by the modulation of the sensorimotor rhythms (caused by motor imagery). One of the most popular algorithms in such systems is the common spatial pattern (CSP) analysis, see [Blankertz et al., 2008]. In CSP-based systems, a higher number of channels may lead to finer spatial filters, which might be more successful in extracting signals from the discriminative sources. On the other hand, the number of parameters that need to be estimated from calibration data increases quadratically with the number of channels, meaning that overfitting may occur. In summary, it is still a question of debate whether smaller channel configurations could allow to operate BCIs without significant worsening of the performance.

In [Popescu et al., 2007], starting with a full setup of 64 electrodes, an iterative channel removal was used to find the best subject-specific placement of electrodes. In their analysis of data from five subjects, the performance started to decrease after less than 20 channels remained in the setup.

A variant of CSP using an L1-norm was used in [Farquhar et al., 2006] in order to obtain sparse spatial filters, thereby implicitly performing a channel selection. For five subjects individually optimal channel configurations and number of CSP filters were obtained, which utilized between 39 and 4 channels.

In [Lal et al., 2004], a ranking method for electroencephalography (EEG) channels mainly based on Recursive Feature Elimination [Guyon et al., 2002] was proposed and 39 channels were ranked for five good BCI performing subjects. Poor performing subjects were excluded from the analysis. While many

differences in the minimum number of channels were found among subjects, a setting of 17 channels near or close to the motor cortex was hypothesized to be the optimal solution.

In the present study we performed an offline investigation of 119-channel data recorded from 80 subjects in order to find the optimal subject-independent channel configuration, i.e. a setting that yields the best results on average across all subjects. Given the size of the data set, the result of this study can be expected to generalize well to future experiments.

2. Material and Methods

2.1 Experimental Setup

In this study, 80 data sets from 80 BCI naive subjects (39m, 41f; age 29.9±11.5y; 4 left-handed) were investigated. Each data set was acquired during a single BCI session with a classical motor imagery paradigm.

During the calibration measurement, in each trial a visual stimulus indicated to the subject which type of movement s/he should imagine. The visual stimulus was an arrow directed to the left, to the right or to the bottom, corresponding to left hand, right hand and foot movement imagination, giving rise to three classes and three possible binary class combinations. Calibration data sets consist of three concatenated runs, each run with 25 trials per class, resulting in 75 trials per class.

Automatic variance based artifact rejection was applied on the calibration data to reject trials and channels corresponding to amplitude abnormalities in the raw electroencephalogram (EEG). For each binary class combination, a semi-automatic procedure selected a specific frequency band and time interval, in which the two classes were best discriminable and CSPs were trained on the corresponding filtered and segmented EEG. Afterwards, CSPs filters were chosen by a heuristic and the log-variance of the CSP filtered EEG signals (in the following called CSP features) was used to train a Linear Discriminant Analysis (LDA) [Friedman, 1989] resulting in a generalized calibration performance. The best class combination was then chosen depending on the best classification performance and used for the feedback session, see [Blankertz et al., 2008] for more details on this procedure.

During the feedback measurement, the previously trained CSP filters and LDA were used to respectively filter the raw data and classify the CSP features, and the classifier output was visualized at the same time as the stimulus presentation. Feedback data consists of three concatenated runs, each run containing 50 trials per class. The first 20 trials (ten per class) were used to adapt the bias of the LDA classifier, and the feedback performance was calculated on the remaining 80 trials, resulting in 240 trials (120 per class) per session.

The stimulus duration was five seconds, while the inter-stimulus interval (ISI) was nine seconds. EEG was recorded using using 119 Ag/AgCl electrodes at positions according an extended international 10-20 system and a sample frequency of 1000 Hz. Electromyogram (EMG) and electrooculogram (EOG) have been also recorded in order to assure that no muscle activity is present during the mental task and that no eye movements could influence the classification.

2.2 Classification performance

In order to evaluate various channel configurations, offline classification performance was determined according to the procedures described in Sections 2.3 to 2.5. The calibration data were used as training set for CSP analysis, and the feedback data was used as test set. For each channel configuration, the parameters of the procedure described in Section 2.1 (time interval, frequency band, set of CSP filters) were selected completely automatically. Different from online feedback, we did not try to mimic continuous cursor movement in this offline analysis. Rather, complete trials extracted from feedback data have been classified. The classification error was then calculated as the area over the Receive Operative Characteristic (ROC) constructed using the output of the LDA classifier.

2.3 Testing predefined channel configurations

Thirteen different channel configurations were investigated. Twelve of them are summarized in Fig. 1. The configuration *'most'*, not shown in Fig. 1, contains all channels except for the very frontal,

temporal and occipital ones. For each channel configuration, the classification error was calculated for all 80 subjects as described in Section 2.2.

Frequency band and time interval for the CSP analysis were chosen in two different modalities. In the first one called *'broad'*, a fixed frequency band of 8-35 Hz and time interval of 750-3750 ms were used. In the second one called *'auto'*, frequency band and time interval were automatically chosen by the heuristic as during the experiment.

Wilcoxon signed rank tests for equality of medians [Gibbons, 1985; Hollander and Wolfe, 1973] were then applied on the test error data for each classification modality to find out the 'best set' in terms of calibration accuracy. In particular, paired tests for comparison of the test error percentages of each configuration with that of the full channel configuration called *'all'* resulted in 13 p-values.

Among the n (say) configurations yielding a p-value < 0.05 and with a median test error smaller than the 'all' set, the configuration with the smallest p-value was selected as the *'best set'*. The test errors of the other *n-1* configurations were again compared to the test error of the 'best one' and those sets corresponding to p-values > 0.1 are listed as *'sets comparable to the best'*, since they result in a performance not significantly worse than the *'best set'*.

2.4 Evaluation on small training sets

The analysis described in Section 2.3 was repeated using just *n = 50, 40, 30* trials from the calibration data in three different ways: 1) just the first *n* trials, 2) the last *n* trials and 3) *n* trials linearly
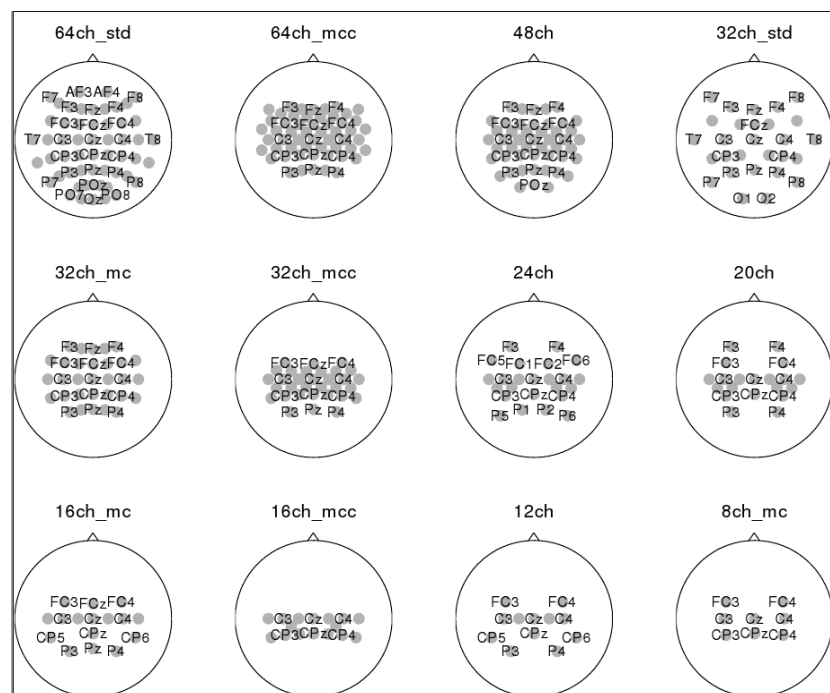
equally spaced in the time. Results for these three modes were averaged.

Figure 1.   *Predefined channel configurations. Some channel labels are omitted for a better visualization.*

2.5 Iterative channel seletion

An iterative channel selection procedure was applied, which allows to determine an optimal channel configuration without being confined to a predefined list of configurations. The method is based on statistical tests to see whether each single channel contributes to the classification of the feedback data significantly or not.

The procedure considers a set of *'selected'* channels and a *'pool'* containing all other channels. The algorithm alternates *'intern cycles'* and *'extern cycles'*. Within an *'intern cycle'* one or more channels belonging to the *'selected'* set are removed from it and added to the *'pool'* because their removal does not yield any significant increase of test error. Within an *'extern cycles'* one or more channel belonging to the *'pool'* are added to the *'selected'* set and removed from the *'pool'* due to a significant test error improvement.

The procedure starts with an *'intern cycle'* using the channel configuration called '32ch_mcc', which contains 32 channels distributed in the very central motor area. In the first iteration, the test error is

calculated for all 80 subjects using the *'selected'* 32 channel configuration and for the 32 *'candidate'* configurations consisting of 31 channels obtained by leaving out one channel from '32ch_mcc'. Testing the difference between the median of the *'selected'* configuration and the median of each *'candidate'* configuration by Wilcoxon signed rank test, 32 p-values are obtained. Channels corresponding to p-values > 0.1 were considered not relevant for the classification because their removal does not cause significant difference in the performance. An additional condition in order to remove a channel was that its elimination leads to an improvement of the median test error. Therefore, if existing, the *'candidate'* configuration with the highest p-value > 0.1 and with median test error smaller than the median test error of the *'selected'* one, becomes the *'selected'* configuration, and the algorithm stays in the *'intern cycle'* modality. If no *'candidate'* configuration satisfies the two conditions on the p-value and the median test error, the procedure enters the *'extern cycle'* modality.

In each iteration in the *'extern cycle'*, the test error is calculated for all 80 subjects and for all possible *N 'candidate'* sets formed by the *'selected'* set plus one channel coming from the *'pool'*, where *N* is the number of channels in the *'pool'*. The paired comparison between the median test error obtained used the *'selected'* set and the median test error of each *'candidate'* set results in *N* p-values. Channels corresponding to p-values < 0.05 were considered relevant for the classification because their inclusion causes significant difference in the performance. An additional condition in order to add a channel, was that its inclusion leads to an improvement of the median test error. Therefore, if existing, the *'candidate'* configuration with the smallest p-value < 0.05 and with median test error smaller than the median test error of the *'selected'* one, becomes the *'selected'* configuration, and the algorithm stays in the *'intern cycle'* modality. If no *'candidate'* configuration satisfies the two conditions on the p-value and the median test error, the procedure enters the *'intern cycle'* modality again. The procedure terminates when no further channels can be removed from or added to the *'selected'* set.

Frequency band and time interval for training CSP were the same as selected in the experiment.

***Table 1.*** *Results obtained using all subjects.*

FILT	BROAD			AUTO		
TRIALS	**Comp. to the best**	**Best**	**Err.[%]**	**Comp. to the best**	**Best**	**Err.[%]**
ALL		48ch	15.20		48ch	11.75
MEAN50	48ch, 20ch, 32ch_mcc, 32ch_mc, 64ch_mcc, 16ch_mc	24ch	20.55	32ch_mcc, 20ch, 32ch_mc, 12ch, 16ch_mc, 24ch	48ch	24.81
MEAN40	48ch, 24ch	32ch_mc	21.60	32ch_mcc, 16ch_mc, 32ch_mc, 20ch, 12ch	48ch	26.07
MEAN30	32ch_mcc, 32ch_mc, 20ch, 16ch_mc	24ch	24.31	16ch_mc, 20ch, 12ch, 8ch_mc, 32ch_mc	32ch_mcc	29.25

3. Results

3.1 Best channel setup of fixed configurations

The results of the significance tests run on the classification performance of the predefined 13 channel configurations are summarized in Table 1. The first row indicates the type of frequency and time filtering applied, called *'broad'* and *'auto'*. The fields *'All'*, 'S*um50'*, *'Sum40'*, *'Sum30'* in the first column mean that the results shown were obtained using all trials from the calibration data and the mean value of the test error resulting using the three modalities with 50, 40 and 30 trials respectively explained in Section 2.4. In the column *'Comp. to the best'*, the sets are listed that, if existing, performed not significantly worse than the best set, visualized in the column *'Best'*. Finally, the column *'Err.'* visualized the median test error of the *'Best'*.

In the more detailed analysis for subgroups of good (>88% accuracy in original feedback), fair (70%-88%) and bad (<70%) performing subjects (not included in Table 1 because of lack of space), it could be observed that in better performing subjects configurations with more channels were preferred.

Channel selected by iteration starting from 32mcc
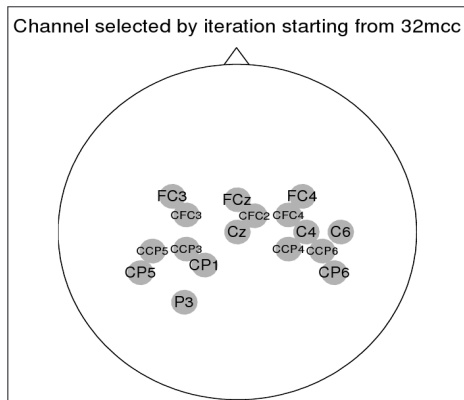
'32mcc' as initial channel configuration.

3.2 Best channel setup by iterative channel selection

The iterative channel selection described in the Section 2.5 chose 17 channels concentrated in the motor area as visualized in Fig. 2. The presence of more channels on the right hemisphere could be due to the fact that the class *'left'* and especially the class combination *'left-foot'*, *was* more often selected during the experiment than the class *'right'* and the combinations *'left-right'* and *'right-foot'*. The median test error using this configuration was *12.79 %*, resulting very competitive to the errors reported with the *'48ch'* set in Section 3.1.

Figure 2.   *Channels selected by iteration using*

4. Discussion and Conclusion

In a detailed study with 80 subjects we investigated how many channels are sufficient for a stable classification of BCI using the motor imagery paradigm.

In the first approach the general tendency was that configurations with more channels were preferred 1) with bigger training sets, 2) for the *'auto'* condition in comparison to the *'broad'* one and 3) for better performing subjects. These facts are presumably due to overfitting effects of CSP [Blankertz et al., 2008] if too many channels are used just with a small number of trials (for points 1 and 2) or with less quality information present in the data of bad performing subjects (for point 3).

In general, the configuration with 48 channels resulted to be the best one. This result contradicts previous findings [Lal et al., 2004; Farquhar et al., 2006; Popescu et al., 2007] where subject depending channel selections showed that much less channels are necessary for a robust classification, even if using good performing subjects and 400 trials [Lal et al., 2004; Farquhar et al., 2006].

On the contrary, the second iterative approach presented in this paper selected 17 channels as the best configuration, agreeing with the previous studies.

It can be then hypothesized that the iterative procedure, as well as the algorithms used in [Lal et al., 2004; Farquhar et al., 2006; Popescu et al., 2007] are probably more efficient against the high correlation present in EEG signals, since they consider either a sparse solutions [Farquhar et al., 2006] or a recursive one, which considers each channel separately.

While the resulting set with 17 channels may seem as expected as motor imagery necessarily focuses on motor activity, it is actually far from obvious: a more fine grained resolution with 119 channels may have yielded more stable CSP filters and likewise more robust and individually adapted classifiers. A minimal number of channels always holds the risk that outliers and artifacts of any kind may spoil statistical estimation: this trade-off between feature sparsity and robustness is a general statistical dilemma. Interestingly, the statistical testing procedure outlined here showed that a sparse subspace is optimal and noise robust. Thus, the use of a well-estimated low dimensional informative subspace not only regularizes against noise and artifacts but also makes optimal use of the complexity of the classifier and focuses the resources of the estimator onto the proper subspace.

The present work goes well beyond existing studies due to its rigor in the statistical analysis and the large number of subjects all recorded in a unified setting.

In the future we will use the channel subset established in this paper as an excellent starting point for subject independent adaptive training procedures [Vidaurre and Schlögl, 2008]. While a standard or even underperforming BCI user can use the channel subset as such, excellent BCI may further gain additional communication performance by adaptively increasing the channel set beyond the standard one.

References

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

Farquhar J, Hill NJ, Lal TN, Schölkopf B. Regularised CSP for sensor selection in BCI. Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course Graz, 14-15, 2006.

Friedman JH. Regularized Discriminant Analysis. *Journal of the American Statistical Association.* 84(405): 165-175,1989.

Gibbons JD. Nonparametric Statistical Inference. Dekker M. Editor, 2nd Edition. New York, 1985.

Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Machine Learning,* 46(1-3): 389-422, 2002.

Hollander M, Wolfe DA. Nonparametric Statistical Methods. Wiley Editor. New York, 1973.

Lal TN, Schröder M, Hinterberger T, Weston J, Bodgan M, Birbaumer N, Schölkopf B. Support vector channel selection in BCI. *IEEE Transactions on Biomedical Engineering,* 51(6): 1003-1010, 2004.

Popescu F, Fazli S, Badower Y, Blankertz B, Müller K-R. Single trial classification of motor imagination using 6 dry EEG electrodes. *PLoS ONE.* 2(7): e637. doi:10.1371/journal.pone.0000637, 2007.

Vidaurre C, Schlögl A. Comparison of adaptive features with liner discriminant classifier for Brain Computer Interfaces. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008. to appear.