Ensembles of temporal filters enhance classification performance for ERD-based BCI systems

S. Fazli[1], C. Grozea[1], M. Danóczy[1], B. Blankertz[2], K.- R. Müller[2], F. Popescu[1]

[1]Fraunhofer First, Kekuléstr. 7, 12489 Berlin, Germany
[2]Technical University, Franklinstr. 28/29, 10587 Berlin, Germany

fazli@first.fhg.de

Abstract

Finding suitable subject-dependent temporal and spatial filters is of paramount importance for achieving high information transfer rates in BCI systems, depending on Event-Related (De-)Synchronization (ERD/ERS). The temporal filter can be chosen manually, or automatically by use of a heuristic. We employ a multi-classifier system (MCS), based on a predefined filter-bank of temporal filters and apply it to 91 datasets, comprising 2-class experiments from 45 subjects. Our results indicate that this approach is a viable alternative to existing methods and can completely automate the temporal filter choice while promising higher performance than either a broad, subject-independent frequency band choice or the current heuristic.

1 Introduction

Classical BCI systems relied on subject-training, i.e. the subject was equipped with a BCI system and had to adapt to the system in order to learn how to use it, a method termed *operant conditioning* [1, 2]. Reducing the subject-training time has been and is still one of the goals of the BCI community, for the achievement of which Machine learning has proved to be a suitable method [3, 4]. Very recently, a technique emerged, where reusage of subject-specific training data of previous sessions is utilized, not only enabling expert BCI subjects to engage in feedback sessions without the need of a calibration session, but also yielding highly competitive accuracies [5]. However, for naïve users, ERD-based BCI systems such as the Berlin Brain Computer Interface (BBCI) require a calibration phase, where the subject is instructed to imagine movements of their respective limbs a number of times. The types of imaginations, called classes, are chosen with respect to their topography on the motor-cortex. This calibration data is then examined and machine learning methods are employed to build subject-dependent spatial and temporal filters as well as a classifier, in order to be able to discriminate the classes as accurately as possible in a real-time feedback paradigm.

Instead of using a single temporal filter that can be identified heuristically or adjusted manually via an expert user from the calibration data, we identified multiple temporal filters which have proven to be effective over a large set of subjects. As can be seen in Figure 1, the calibration data are filtered by one of the predefined temporal filters. Subsequently, corresponding spatial filters and classifiers are calculated for each of the given temporal filters. The individual outputs thus obtained are then combined by a final gating function to yield the resulting output. The final gating function may be a simple voting scheme, but can also be designed in a more complex fashion.

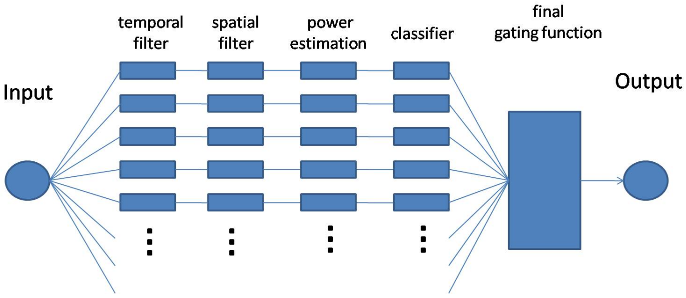

Figure 1: The flowchart shows the general setup of the method. Data is run through predefined temporal filters, and subsequently filtered by a spatial filter, which was obtained by the training set and classified, once the spectral power has been estimated. A gating function combines the outputs from the individual classifiers to a single output.

number of datasets/subject	1	2	3	5	8	9	13
occurance	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 times a particular subject carried out only a single experiment, 5 subjects carried out 2 experiments, and so on. The third row show the percentage of all trials that fall into each category (not every experiment contains the same number of trials).

2 Methods

2.1 Datasets used

We used 91 experiments of 45 individual subjects, all of which have been recorded at the IDA group. For each dataset the number of trials ranged from 70 to 600 trials and in total $\approx$ 22000 trials were examined. Each trial was referenced by 500-3500 msec, relative to stimulus onset. We considered only datasets with 2 classes of movement imaginations: left hand versus right hand. 45 channels were identified to be present in all datasets and only these were used for reasons of consistency. The channels we used are:'F5', 'F3', 'F1', 'Fz', 'F2', 'F6', 'FC5', 'FC3', 'FC1', 'FCz', 'FC2', 'FC4', 'FC6', 'T7', 'C5', 'C3', 'C1', 'Cz', 'C2', 'C4', 'C6', 'T8', 'TP7', 'CP5', 'CP3', 'CP1', 'CPz', 'CP2', 'CP4', 'CP6', 'TP8', 'P5', 'P3', 'P1', 'Pz', 'P2', 'P4', 'P6', 'P8', 'PO3', 'POz', 'PO4', 'O1', 'Oz' and 'O2'. Individual experiments consisted of different training paradigms. While there were minor differences for the various paradigms, generally speaking visual cues were presented to the subject, and she was intructed to perform the cued imagination upon appearance. For further details we refer the reader to [6]

We would like to draw the readers attention to the fact that the dataset, considered here is biased towards subjects, for which BCI generally works, since the subjects who were chosen to take part in more than one of experiment were generally not BCI-illiterates. Please refer to Table 1 to see how many experiments per subject are present in the data we used.

2.2 Selection of a frequency band

2.2.1 heuristic

As has been shown in [7] a heuristic can be very useful in detecting the most discriminant frequency range for a given subject. In short the method's steps are the following:

1. Use Laplacian or bipolar channels, from motor cortex related electrodes

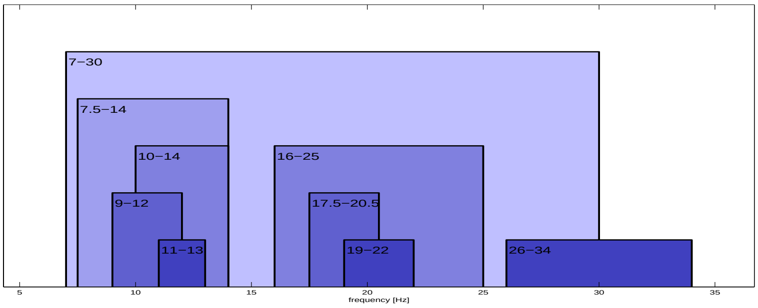

Figure 2: The Figure shows all temporal filters, used in the ensemble. Note that the heights of the patches are not related to the orders or the magnitude responses of the filters. All filters are order 5 butterworth filters.

frequency [Hz]	loss [%]	best performance [%]
7.5 – 14	**12.4**	16.5
11 – 13	19.6	9.9
10 – 14	12.8	**30.8**
9 – 12	18.6	11.0
19 – 22	42.9	1.1
16 – 25	31.8	6.6
26 – 34	46.4	2.2
17.5 – 20.5	41.4	2.2
7 – 30	14.8	19.8

Table 2: The Table summarizes the median loss of each temporal filter, we chose to include for the ensemble over all subjects. The first column gives the actual frequency band, the second column represents the median, x-validated loss over all datasets and the third column shows for how many datasets the given filter performed best. Note that for seemingly unsuitable filters, some datasets score their best validation loss.

2. For each trial, channel and frequency in the range from 7 to 35 Hz, calculate the log-bandpower
3. Calculate the correlation coefficient between the log-bandpowers and their true labels
4. Find the frequency with the highest correlation coefficient and broaden the band step-wise in both diretions, until the next frequency bin is smaller than 5% of the peak

2.2.2 Filter bank

From neurophysiology we know, that the μ-rhythm (9-14 Hz) and synchronized components in the beta band (16-22 Hz) are macroscopic idle-rhythms, that occur when a subject is at rest and are located over the postcentral somatosensory cortex and precentral motor cortex, respectively. Imaginations of movements as well as actual movements are known to suppress this idle rhythm contralaterally. However the motoric μ-rhythm as well as the beta-rhythm can have a slightly different frequency ranges for individual subjects. From empirical considerations we identified 9 different band-pass filters that have proven to discriminate best over a large range of subjects. Figure 2 shows the filters we used to generate the ensemble, table 2 summarizes how well each of the chosen filters performs on average. Also in the same table it can be seen how often each one of the proposed filters acts best on all experiments, percentagewise.

2.3 Validation

Every dataset was split into two chronological halves, i.e. each method was trained on the first half and then validated on the second half.

2.4 final gating function

The final gating function can be realized in different ways, for a intuitive review of the most common methods, we would like to refer the reader to [8]. Given the classifier outputs for a single trial $X \in R^{d \times t}$, one approach is to average out the resulting outputs of the individual classifiers $\hat{y}_m = \sum_{j=1}^{d} X$. However, we may also choose to let only the classifier with highest absolute value take the decision or let the majority decide or the median of the individual outputs. Table 3 shows the validation results of these combination rules.

3 Results

Each LDA output for a given trial indicates how far the feature is from the hyperplane. This can be interpreted as how confident a classifier is. In this sense the weighting of the individual classifiers is already optimal. It is therefore not surprising that the ensemble mean yields the best results, as can be seen from Table 3. For good subjects the heuristic performs very well, while for subjects, where the discriminativities are not so well detectable, a broadband CSP perfoms favourably.

	CSP		Ensemble			
	broad	auto	mean	max	maj	med
25%-tile	7.2	4.1	**3.6**	6.8	24.1	5.1
median	14.8	15.5	**11.2**	17.3	43.8	11.3
75%-tile	31.7	36.7	**30.7**	32.7	64.9	31.4

Table 3: The results given above were calculated for each subject individually and then averaged. The left part of the table shows the loss of the baseline, using CSP and a broad band-pass filter ($7 - 30$ Hz) and the automatic heuristic. On the right hand side of the table the results of the ensemble are presented. "mean" uses the mean of all classifiers, "max" uses only the classifier output with the highest absolute value, "maj" stands for majority voting and "med" for median voting

4 Discussion

The principal aim of this work is to make classifier tuning as automatic and fast as possible. Certainly, the mean ensemble method obviates the need to find any extra parameter estimation for the weighting of the individual classifier outputs.

The motivation of the ensemble of temporal filters was, that for small numbers of training trials, or for subjects, where the detection of the correct frequency band is difficult, it is certainly possible that the heuristic fails. By using the ensemble we introduce prior information from neurophysiology and BCI classifier calibration experience and let the ensemble of classifiers decide which band scores the highest confidence at minimal computational cost.

It would be unrealistic to claim that the data presented here can be seen as an unbiased sample of society, as only successful BCI subjects are likely to participate in more than one experiment. However, since most of the BCI community is interested in well performing subjects, the results presented here should be of interest. Furthermore, when possible we look at individual subject performance as well as experiment performance, as to reduce this bias as much as possible.

While this study is solely based on offline results, it can be of significant value only if designed in a

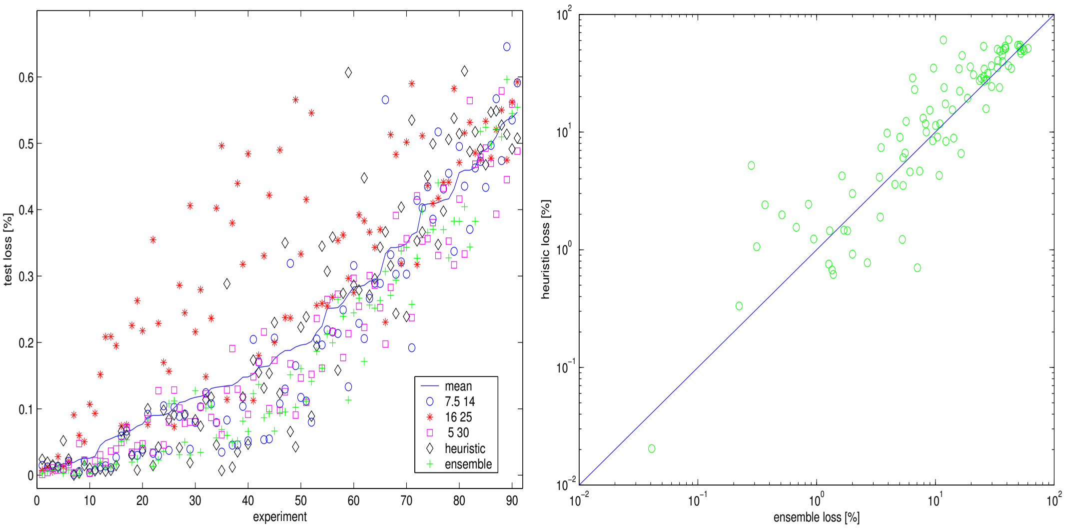

Figure 3: The Figure on the left shows the resulting loss of 4 different frequency bands, data is sorted by the mean performance of all bands. The Figure on the right shows the test loss for each individual experiment for the best ensemble method, versus the classical procedure, with the automatic heuristic

meaningful manner. The implementation of the presented method is computationally manageable and in fact we plan to validate the presented results by online experiments in novel subjects.

5 Conclusion

Ensemble methods have only recently been applied to BCI related data [9, 10] with promising results, but to our knowledge not using the filter-bank approach presented here. We show that our approach of parallelizing the decoding of ERD-based EEG signals has a beneficiary effect on the classification performance. Furthermore, we see that the combination of the individual classifier outputs can be realized in a very simple and effective manner. Using ensemble methods has an increased but manageable computational cost, while it has the inherent advantage over optimization of the correct band pass filter that it is less prone to overfitting.
It remains to be seen, if by this method, the resulting architecture is more robust to non-stationarities, which may occur over long feedback sessions. Whatever the nonstationarities, it is less likely that a particular subject changes his frequency band corresponding to motor imagination over time. Furthermore, this could be easily tested by applying the presented method to datasets where non-stationarities are known to exist or by putting the method into practice in a feedback environment.
In the more likely case that slightly different cortical regions are active and encode information with different frequencies, the approach we present here ensures that these type of effects are taken into account and can be tracked in the nonstationary case.

References

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

[2] T. Elbert, B. Rockstroh, W. Lutzenberger, and N. Birbaumer. Biofeedback of slow cortial potentials. i. *Electroencephalogr. Clin. Neurophysiol.*, 48:293–301, 1980.

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

[4] K.-R. Müller, M. Krauledat, G. Dornhege, G. Curio, and B. Blankertz. Machine learning techiques for brain-computer interfaces. *Biomed Tech*, 49(1):11–22, 2004.

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

[6] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K.-R. Müller. Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. In *Advances in Neural Inf. Proc. Systems (NIPS 08)*, volume 20, 2008.

[7] 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*, 25(1):41–56, January 2008.

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

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

[10] A. Rakotomamonjy and V. Guigue. Bci competition iii: dataset ii- ensemble of svms for bci p300 speller. *IEEE Trans Biomed Eng*, 55(3):1147–1154, March 2008.