# Alternative CSP approaches for multimodal distributed BCI data

Stephanie Brandl, Klaus-Robert Müller, *Member, IEEE*, and Wojciech Samek, *Member, IEEE*

***Abstract*—Brain-Computer Interfaces (BCIs) are trained to distinguish between two (or more) mental states, e.g., left and right hand motor imagery, from the recorded brain signals. Common Spatial Patterns (CSP) is a popular method to optimally separate data from two motor imagery tasks under the assumption of an unimodal class distribution. In out of lab environments where users are distracted by additional noise sources this assumption may not hold. This paper systematically investigates BCI performance under such distractions and proposes two novel CSP variants, ensemble CSP and 2-step CSP, which can cope with multimodal class distributions. The proposed algorithms are evaluated using simulations and BCI data of 16 healthy participants performing motor imagery under 6 different types of distraction. Both methods are shown to significantly enhance the performance compared to the standard procedure.**

## I. Introduction

Brain-Computer Interfacing (BCI) serves as a non-muscular communication system between a computer device and a human being [1], [2]. It depends on the user's measured brain signals alone and thus provides a powerful tool for *locked-in* patients who are not able to move, speak, voluntarily blink or control their eye movement [3]. A BCI translates a user's intentions by measuring brain signals e.g. with electroencephalography (EEG) into computer commands and so allows human beings to control a computer device without the use of muscular control or speech. BCI's also find their use in the fields of wheelchair control [4], rehabilitation [5] and mental state monitoring [6].

Combining the field of machine learning with BCI research already reduced calibration time [7], [8] and thus essentially improved BCI efficiency and usability. Novel, more robust, approaches led to important improvements e.g. in artifact classification [9] and feature extraction [10], [11], [12].

Since EEG recordings are highly sensitive to noise, most BCI research has been carried out in very artificial lab environments where users sat still and could entirely focus on the respective task. Those situations do not represent everyday life situations where the user might be distracted by environmental changes, background noise (e.g. street noise, TV, conversations) or just cognitive tasks.

Recent BCI studies started leaving this controlled lab environments and conducted studies with end-users [13], [14], [15], [16]. Other studies were carried out with participants walking indoors [17], outdoors [18] or on a treadmill [19] while controlling a spelling device or also speaking while carrying out motor imagery tasks [20].

With the idea to contribute to this recent out of lab research, we recorded a motor imagery-based BCI study where we simulated multiple real-world scenarios by adding secondary tasks to the primary motor imagery task [21]. This design allowed us to systematically investigate different distraction scenarios.

However, for several participants, common classification methods did not lead to significant BCI control. We therefore propose novel approaches based on ensemble methods and multiple-step classification which tackle the problems of changing environmental situations and are put into perspective with the recorded study.

This paper is organized as follows: We summarize CSP, ensemble methods and multiple-step classifcation in Section 2 and explain our simulation study. In Section 3, we briefly present the recorded BCI study (a more detailed description can be found in [21]), before we evaluate our results in Section 4 and conclude the paper in Section 5.

## II. Methods

### A. Common Spatial Patterns

Common Spatial Patterns (CSP) is a well established spatial filtering method in motor imagery-based BCIs [22] [23]. It detects synchronization and desynchronization processes and computes discriminative spatial filters by maximizing the variance of one class (e.g. left hand imagination) while minimizing the variance of the other class (e.g. right hand imagination), which can be solved by a generalized eigenvalue problem

$$\Sigma_1 w = \lambda \Sigma_2 w.$$

The obtained spatial filters $W = [w_1, w_2, \ldots, w_D]$ will be sorted by $\alpha_i = \max\{\lambda_i, \frac{1}{\lambda_i}\}$ according to their contributing discriminative quality such that $\alpha_1 \geq \ldots \geq \alpha_D$.

### B. Ensemble Methods

Ensemble Methods are a basic concept and widely used in the field of machine learning. They combine multiple classi-

S. Brandl (stephanie.brandl@tu-berlin.de), is with the Berlin Institute of Technology, Marchstr. 23, 10587 Berlin, Germany.

K.-R. Müller (klaus-robert.mueller@tu-berlin.de) is with the Berlin Institute of Technology, Marchstr. 23, 10587 Berlin, Germany and with the Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea.

W. Samek (wojciech.samek@hhi.fraunhofer.de) is with the Fraunhofer HHI, Einsteinufer 37, 10587 Berlin, Germany.

*This work was supported by the Federal Ministry of Education and Research (BMBF) under the project Adaptive BCI (FKZ 01GQ1115), by the DFG, by the World Class University Program through the National Research Foundation of Korea funded by the Ministry of Education, Science, and Technology, under Grant R31-10008 and by the Brain Korea 21 Plus Program through the National Research Foundation of Korea funded by the Ministry of Education.

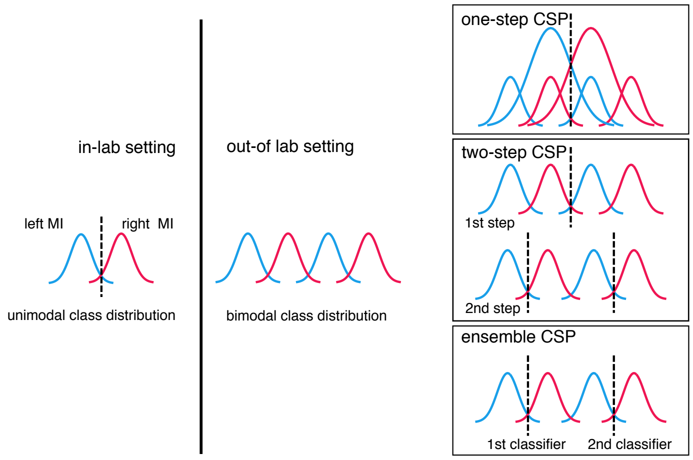


Fig. 1: The two new approaches compared to one-step CSP

fiers to improve accuracy and robustness. Therefore, individual classifiers need to reach higher accuracy than chance level and their errors must be independent or at least uncorrelated [24]. Bagging computes several classifiers and determines the prediction by a majority voting and thus cancels out variance and bias in data and unstable methods [25].

Since original CSP only considers two Gaussian-distributed classes, we face major complications when classifying data that does not arise from the same two Gaussian distribution. An ensemble version of CSP could profit from the diversity of different classifiers in case of multimodal data that is still separable in e.g. *left* vs. *right* (see Figure 1).

### C. Multiple-Step Classification

Assuming different distraction scenarios in real-world situations can lead to immense feature shifting when applying only a single classifier. While data from different scenarios might each be separable (e.g. in *left* and *right*) the whole dataset usually would not. Applying multiple steps in the classification process could separate the data into different *distractions* while in the next step, data is separated into *left* and *right* by a group-based classifier (see Figure 1).

Similar approaches have been applied in BCI research [26], [27], [28] and reviewed in [29].

### D. Simulations

To investigate behaviour and limits of ensemble CSP and multiple-step CSP, we simulated EEG time series with binary decision tasks for three different groups of EEG data (*clean, little noisy, very noisy*). They represent shifts in data distribution as one might discover in presence of artifacts, cognitive distractions or environmental changes.

For *clean*, we generated 2 diagonal covariance matrices (one for each class, e.g. *left* vs. *right*) with uniformly distributed random numbers in the interval $[0, 10]$.

$$\Sigma_{cl} = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} \qquad a_{11}, a_{22} \in [0, 10] \tag{1}$$

For the groups *little noisy* and *very noisy* we added some shifting which varied between 0 and 50 (in steps of 2.5) between *clean* and *little noisy* and double of that shift between *clean* and *very noisy* (0-100, in steps of 5) such that their diagonal covariance matrices were generated by uniformly distributed random numbers from between 0 and 10 up to between 100 and 110. For each of the 21 shifting values, we repeated the simulations 1000 times.

$$\Sigma_{ln} = \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix} \qquad b_{11}, b_{22} \in [0 + s_1, 10 + s_1], \tag{2}$$
$$s_1 = [0 : 2.5 : 50]$$

$$\Sigma_{vn} = \begin{pmatrix} c_{11} & 0 \\ 0 & c_{22} \end{pmatrix} \qquad c_{11}, c_{22} \in [0 + s_2, 10 + s_2], \tag{3}$$
$$s_2 = [0 : 5 : 100]$$

For each group, we generated trials of 2-dimensional Gaussian distributed data with 100 samples (time points), mixed it with a random orthogonal matrix and added Gaussian noise to each trial.

$$X_{cl} = A \cdot S_{cl} + n, \qquad S_{cl} \sim \mathcal{N}(\vec{0}, \Sigma_{cl}) \tag{4}$$
$$n \sim \mathcal{N}(0, \sqrt{2})$$

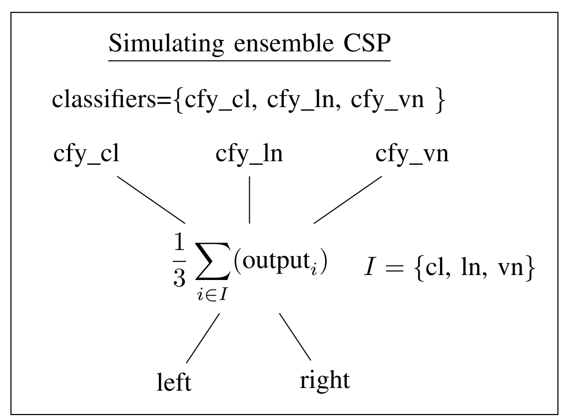


Fig. 2: Simulation Ensemble CSP

*1) Ensemble CSP:* We simulated 480 trials all together and used half of the trials for training and the other half for testing. Since one classifier is trained for each group, we generated 160 trials per group (*clean, little noisy, very noisy*), equally balanced between *left* and *right*.

For each group, we extracted one CSP filter and trained a classifier based on regularized linear discrminant analysis (RLDA) [30], [22], [7]. We then applied all 3 classifiers on all the testing data, calculated the mean of the 3 classifiers' output and compared this averaged output to the real labels, see Figure 2 for an overview.

To compare ensemble CSP with original CSP, we also trained a single RLDA-based classifier on all training trials (one CSP filter) and applied this to the testing data.

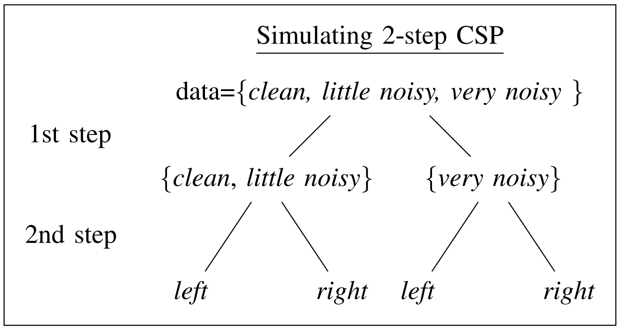


Fig. 3: Simulation 2-step CSP

*2) 2-Step CSP:* We again simulated 480 trials, used half for training and the other half for testing. Here, we assume a scenario where we have one outlier group which highly differs from the rest of the data. Therefore, we generated 320 *clean* trials and 80 of both, *little noisy* and *very noisy*.
We trained one 1st step-classifier to separate {*clean*, *little noisy*} from {*very noisy*} and two *left* vs. *right* 2nd-step classifiers, see Figure 3 for an overview. In case of misclassification during the first step it is possible though that e.g. a *little noisy* trial ends up in the *very noisy* group and the *wrong* classifier is applied in the second step.
We compared this approach to original CSP, where we trained a global one-step classifier on all training data to discriminate between *left* and *right* and applied this classifier to the testing data.
All classifiers were trained based on RLDA and one CSP filter.

## III. EXPERIMENTS

### A. Participants

We recorded EEG data of 16 healthy volunteers (6 female; age: 22-30 years). Most of them had no BCI experience, only 3 had already participated in a BCI study and only 1 of them in a motor imagery experiment. All participants were required to speak and understand German in order to understand the visual and auditory instructions which were given in German. The volunteers were paid for their participation except for 3 participants who are members of the TU Berlin Machine Learning Group.

### B. Distractions

Besides the primary motor imagery task, we added 6 secondary distraction tasks. All distraction tasks are explained briefly in Table I. Those additional tasks lead to e.g. cognitive distractions (*news, numbers*), muscular artifacts (*numbers*) and steady state potentials (*flicker, stimulation*). For a more detailed description, we refer to [21].

### C. Experimental Setup

We recorded with 63 wet Ag/AgCl electrodes placed according to the 10-20 system [31] at symmetrical positions on a Fast'n Easy Cap (Easy Cap GmbH) with reference to the nose. Signals were amplified with two 32-channel amplifiers (Brain Products) and sampled at 1000 Hz.

We recorded 7 runs with 72 motor imagery trials each (36 left, 36 right). Recording one run took about 10 minutes, one trial lasted 4.5 seconds. The first run served as a calibration phase where no secondary task was added. In each of the following runs, we included 12 trials (6 left, 6 right) of each secondary task (see Table I). This means, we recorded 72 trials of each secondary task, except for *clean*, where we also recorded 72 trials during the calibration phase. Each task was equally balanced between left and right hand motor imagination.
After the calibration phase, we extracted Laplacian filters [32] of the electrodes C3 and C4 and trained an RLDA-based classifier in broad band (9-13Hz, 18-26Hz) which was applied in the online classification during run 2-7.

### D. Data Analysis

After downsampling data to 100Hz, we selected an individual frequency band (in the maximum range of 5 and 35Hz) and time interval for each participant according to [22].
With three CSP filters per class, we trained an RLDA-based classifier on the calibration data (no secondary distraction tasks) and tested on the remaining data (with distraction tasks). Average classification rates are displayed in Table II. Each row represents one participant and the corresponding averaged classification accuracy over the whole experiment as well as the average classification rates for each secondary task. The ones highlighted in bold represent the participant's best task and the ones in red the participant's weakest. We categorized the participants according to their classification outcome in 3 groups. The first group achieved significant BCI control in all secondary tasks (threshold of $61.11\%$). The second group reached significant BCI control in the overall experiment (threshold of $54.17\%$) and the third group did not reach significant BCI control. Those thresholds were calculated by applying a binomial test ($\alpha = 0.05$). Classification rates vary clearly between the different tasks, especially the *numbers* task shows a major decrease in classification accuracy. Since only *clean* data was used in the training phase, we can assume major feature shifts between e.g. *numbers* and calibration. To support this assumption, we also classified *clean* against each of the other secondary tasks and detected major feature shifts especially between *clean* and *numbers* [21].
To overcome those feature shifts we further applied ensemble CSP where we trained 6 classifiers (with 3 filters per class each), one for each secondary task and averaged over the output of all classifiers to compute classification accuracy.
For the 2-step approach we first trained a classifier with one filter per class to separate *numbers* from the rest of the data. For the second step, we trained one classifier on *numbers* and one on *not-numbers* (both with 3 filters per class) to discriminate between left and right hand motor imagery.

## IV. EVALUATION

### A. Simulations

Results of both simulations are displayed in Figure 4. On the x-axis, we plotted the symmetric Kullback-Leibler divergence [33] as a distance measure between the average covariance

TABLE I: 6 different secondary tasks which were added to the primary motor imagery task

| **Name** | **Distraction Task** | **Motivation** | **Real-World Scenario** |
|---|---|---|---|
| **Clean** | without distraction | control task | |
| **Eyes** | closing eyes | overlay of $\alpha$ and $\mu$ rhythms | getting tired, relaxing |
| **News** | listening to a public newscast | cognitive distraction<br>activation of auditory cortex | noisy environments (TV, music, street noise) |
| **Numbers** | searching the room for a certain letter-number combination | cognitive distraction<br>additional muscular artifacts | cognitive distractions in everyday life |
| **Flicker** | watching a flickering video (10Hz) | steady state visually evoked potential (SSVEP) | watching TV, using a computer |
| **Stimulation** | vibro-tactile stimulation | steady state vibration somatosensory evoked potential (SSVESP) | |

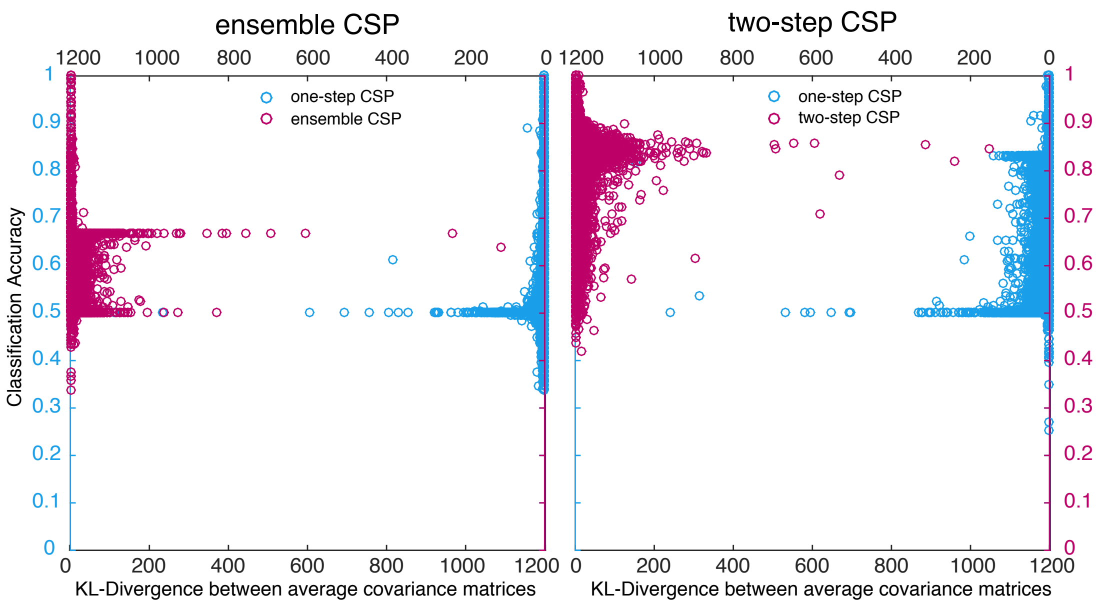


Fig. 4: Results of a 2-step CSP simulation ({clean, little noisy} vs. {very noisy}) with 21000 data points. The x-axis shows the KL-divergence between average covariance matrices and the y-axis the classification accuracy.

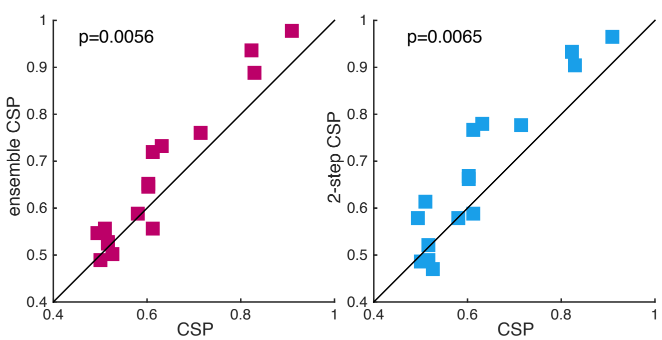


Fig. 5: The two new approaches compared to CSP trained on *clean*

matrices.

Ensemble CSP outperforms one-step CSP in 77.02% of all repetitions. Classification results for both methods are not very high though. One-step CSP achieves an average classification accuracy of 53%, while ensemble CSP reaches an average of 62.29%. The high data contamination (only one third is considered *noise-free*) might be the main reason for this. For both methods we can identify a peak at 50% accuracy where the classifier did not find any discriminative function. For ensemble CSP there is a second peak at around 66%. Misclassifying one third of the data could indicate that both outlier groups were not separable at all and therefore only reached accuracies around chance level. Those results show that ensemble CSP indeed classifies significantly better than one-step CSP but also is not able to overcome serious data contamination. To prove significance, we applied a right-tailed Wilcoxon signed rank test ($\alpha = 0.05$) [34].

Two-step CSP clearly outperforms one-step CSP especially for higher distances. In 93.02% of all cases, 2-step CSP achieves higher classification accuracy than one-step CSP. On average, one-step CSP reaches 62.12% classification accuracy, 2-step CSP even 83.04%. For 2-step CSP there is a clear peak at

TABLE II: Mean classification accuracies for **one-step CSP**. One row represents one participant. For each participant, the conditions with highest (bold) and lowest (red) performance rates are highlighted.

| CSP | ∅ | **clean** | **eyes** | **news** | **num** | **flicker** | **stim** |
|---|---|---|---|---|---|---|---|
| od | 90.97 | **95.83** | **95.83** | 93.06 | 72.22 | **95.83** | 93.06 |
| obx | 82.87 | 88.89 | 87.50 | 81.94 | 70.83 | **91.67** | 76.39 |
| nko | 82.13 | 93.06 | 83.33 | 80.56 | 62.50 | **94.44** | 78.87 |
| njz | 71.30 | **83.33** | 81.94 | 75.00 | 45.83 | 77.78 | 63.89 |
| nkq | 63.26 | **73.61** | 59.72 | 61.97 | 47.89 | 66.67 | 69.44 |
| nkt | 61.34 | 66.67 | 62.50 | 66.67 | 51.39 | **70.83** | 50.00 |
| nkr | 61.11 | 63.89 | 61.11 | 62.50 | 51.39 | **65.28** | 62.50 |
| njy | 60.42 | 62.50 | 54.17 | 65.28 | 50.00 | **69.44** | 61.11 |
| nkm | 60.42 | **68.06** | 52.78 | 65.28 | 56.94 | 55.56 | 63.89 |
| nkn | 58.00 | 62.50 | 52.78 | 61.11 | 49.30 | **65.28** | 56.94 |
| nkl | 52.55 | 45.83 | 48.61 | 54.17 | 54.17 | **61.11** | 51.39 |
| nkp | 51.62 | 51.39 | **55.56** | 50.00 | 50.00 | 52.78 | 50.00 |
| nku | 51.62 | **61.11** | 52.78 | 47.22 | 50.00 | 48.61 | 50.00 |
| ma4 | 51.16 | 56.34 | **58.33** | 48.61 | 49.30 | 41.67 | 52.78 |
| nkk | 50.00 | 48.61 | **55.56** | 43.06 | 51.39 | 51.39 | 50.00 |
| nks | 49.42 | 47.22 | 47.14 | 45.83 | 47.89 | **54.17** | **54.17** |
| ∅ | 62.39 | **66.68** | 63.10 | 62.64 | 53.81 | 66.41 | 61.53 |

TABLE III: Mean classification accuracies for **ensemble CSP**. The results which improved compared to Table II are highlighted in (blue).

| od | obx | nko | njz | nkq | nkt |
|---|---|---|---|---|---|
| 97.92 | 88.89 | 93.66 | 75.93 | 55.56 | 71.99 |
| nkr | njy | nkm | nkn | nkl | nkp |
| 73.24 | 64.58 | 65.28 | 58.92 | 50.23 | 52.78 |
| nku | ma4 | nkk | nks | ∅ | |
| 52.55 | 55.63 | 49.07 | 54.69 | 66.31 | |

around 85% where, again, separation of the outlier groups might have failed. After applying a Wilcoxon signed rank test ($\alpha = 0.05$), we can also state significant improvement for 2-step CSP compared to one-step CSP.
The differences in performance between ensemble CSP and 2-step CSP can be explained by the different simulation setting. For ensemble CSP, we simulated the same amount of trials for all 3 groups (*clean, little noisy, very noisy*). In the case of 2-step CSP, where we assume one particular outlier group, we have twice as many *clean* trials as we have contaminated trials. This makes it easier to achieve higher classification accuracies compared to the ensemble CSP setting.

### B. Study

Results of ensemble CSP and 2-step CSP applied to our recorded BCI data can be found in Table III and IV. Comparing the result of one-step CSP in Table II with the ensemble results in Table III shows that we could improve the overall performance in 13 out of 16 participants (improvements are marked in blue). It is noteworthy that the 3 participants who achieve lower classification rate in the ensemble approach, only achieved between 50 and 61.11% accuracy with one-step CSP.

TABLE IV: Mean classification accuracies for **2-step CSP**. The results which improved compared to Table II are highlighted in (blue).

| | | 1st step | 2nd step | |
|---|---|---|---|---|
| | **overall** | **cond** | **not numbers** | **numbers** |
| od | 96.53 | 100.00 | 99.17 | 83.33 |
| obx | 90.28 | 99.31 | 91.92 | 82.19 |
| nko | 93.19 | 96.71 | 93.82 | 90.00 |
| njz | 77.55 | 97.45 | 78.71 | 72.00 |
| nkq | 77.93 | 99.53 | 80.28 | 66.20 |
| nkt | 76.85 | 99.77 | 79.11 | 65.75 |
| nkr | 58.80 | 99.07 | 57.26 | 66.22 |
| njy | 66.20 | 96.53 | 70.54 | 46.84 |
| nkm | 66.90 | 86.34 | 70.66 | 50.62 |
| nkn | 57.75 | 96.95 | 59.08 | 51.90 |
| nkl | 46.99 | 99.31 | 47.90 | 42.67 |
| nkp | 49.07 | 95.37 | 48.56 | 51.19 |
| nku | 52.08 | 98.84 | 52.65 | 49.32 |
| ma4 | 61.27 | 98.83 | 60.45 | 65.28 |
| nkk | 48.61 | 94.68 | 48.12 | 50.57 |
| nks | 57.75 | 98.83 | 58.43 | 54.29 |
| ∅ | 67.36 | 97.34 | 68.54 | 61.77 |

Comparing 2-step CSP with one-step CSP yields similar results. We could improve accuracies for 11 out of 16 participants and the five participants with lower accuracies in the 2-step approach also only achieved between 50 and 61.11% accuracy with one-step CSP.
A comparison of both methods with the original CSP approach is displayed in Figure 5. Each square represents one participant, p-values of statistical testing (one-sided Wilcoxon signed rank test, $\alpha = 0.05$) are displayed in the upper left.

## V. Conclusion

Everyday life situations bear much more complexity than controlled lab environments. Adjusting classification and feature extraction methods are therefore crucial when bringing BCI research out of the lab.
In this paper we proposed two new methods, ensemble CSP and 2-step CSP. They tackle the problem of multimodal data distribution and major feature shifts. Both perform significantly better than original CSP in simulation scenarios and artifact contaminated BCI data. However, they both still have major difficulties when it comes to seriously contaminted data as we have seen in the simulated and real scenarios. Ensemble CSP needs diverse and accurate classifiers to improve its accuracy with respect to original CSP, if individual classifiers are not accurate, combining them will not yield the desired result. For the first step of 2-step CSP it is important that data can be separated in different groups. If that step fails, the error propagates itself and the method would not work properly.
Both methods mean significant improvement when it comes

to multimodal distributed BCI data. Future work could focus on a more robust feature extraction or classification method so that even noisy data can be classified correctly. Deep neural networks [35] and advanced data fusion techniques [36], [37] may help to tackle this problem.

## References


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

[2] B. Graimann, B. Z. Allison, and G. Pfurtscheller, *Brain-computer interfaces: Revolutionizing human-computer interaction*. Springer, 2010.

[3] 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*, vol. 398, no. 6725, pp. 297–298, 1999.

[4] T. Kaufmann, A. Herweg, and A. Kübler, "Toward brain-computer interface based wheelchair control utilizing tactually-evoked event-related potentials," *Journal of neuroengineering and rehabilitation*, vol. 11, no. 1, p. 7, 2014.

[5] J. J. Daly and J. R. Wolpaw, "Brain–computer interfaces in neurological rehabilitation," *The Lancet Neurology*, vol. 7, no. 11, pp. 1032–1043, 2008.

[6] 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," *Journal of neuroscience methods*, vol. 167, no. 1, pp. 82–90, 2008.

[7] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, "Introduction to machine learning for brain imaging," *Neuroimage*, vol. 56, no. 2, pp. 387–399, 2011.

[8] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for brain-computer interfacing," *PLoS ONE*, vol. 3, no. 8, p. e2967, 2008.

[9] I. Winkler, S. Brandl, F. Horn, E. Waldburger, C. Allefeld, and M. Tangermann, "Robust artifactual independent component classification for bci practitioners," *Journal of neural engineering*, vol. 11, no. 3, p. 035013, 2014.

[10] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve bci designs: unified theory and new algorithms," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 2, pp. 355–362, 2011.

[11] W. Samek, M. Kawanabe, and C. Vidaurre, "Group-wise stationary subspace analysis - a novel method for studying non-stationarities," in *Proc. of the 5th Int. BCI Conference*. Verlag der TU Graz, 2011, pp. 16–20.

[12] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Reviews in Biomedical Engineering*, vol. 7, pp. 50–72, 2014.

[13] C. Neuper, G. Müller, A. Kübler, N. Birbaumer, and G. Pfurtscheller, "Clinical application of an eeg-based brain–computer interface: a case study in a patient with severe motor impairment," *Clinical neurophysiology*, vol. 114, no. 3, pp. 399–409, 2003.

[14] K. K. Ang, C. Guan, K. S. G. Chua, B. T. Ang, C. W. K. Kuah, C. Wang, K. S. Phua, Z. Y. Chin, and H. Zhang, "A large clinical study on the ability of stroke patients to use an eeg-based motor imagery brain-computer interface," *Clinical EEG and Neuroscience*, vol. 42, no. 4, pp. 253–258, 2011.

[15] R. Leeb, S. Perdikis, L. Tonin, A. Biasiucci, M. Tavella, M. Creatura, A. Molina, A. Al-Khodairy, T. Carlson, and J. dR Millán, "Transferring brain–computer interfaces beyond the laboratory: successful application control for motor-disabled users," *Artificial intelligence in medicine*, vol. 59, no. 2, pp. 121–132, 2013.

[16] J. Höhne, E. Holz, P. Staiger-Sälzer, K.-R. Müller, A. Kübler, and M. Tangermann, "Motor imagery for severely motor-impaired patients: evidence for brain-computer interfacing as superior control solution," *PLOS ONE*, vol. 9, no. 8, p. e104854, 2014.

[17] F. Lotte, J. Fujisawa, H. Touyama, R. Ito, M. Hirose, and A. Lécuyer, "Towards ambulatory brain-computer interfaces: A pilot study with p300 signals," in *Proc. of the Int. Conf. on Advances in Computer Enterntainment Technology*, 2009, pp. 336–339.

[18] M. De Vos, K. Gandras, and S. Debener, "Towards a truly mobile auditory brain–computer interface: exploring the p300 to take away," *International journal of psychophysiology*, vol. 91, no. 1, pp. 46–53, 2014.

[19] T. Castermans, M. Duvinage, M. Petieau, T. Hoellinger, C. Saedeleer, K. Seetharaman, A. Bengoetxea, G. Cheron, and T. Dutoit, "Optimizing the performances of a p300-based brain–computer interface in ambulatory conditions," *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, vol. 1, no. 4, pp. 566–577, 2011.

[20] H. Gürkök, M. Poel, and J. Zwiers, "Classifying motor imagery in presence of speech," in *Proc. of Int. Joint Conference on Neural Networks (IJCNN)*. IEEE, 2010, pp. 1–8.

[21] S. Brandl, J. Höhne, K.-R. Müller, and W. Samek, "Bringing bci into everyday life: Motor imagery in a pseudo realistic environment," in *Proc. of the Int. IEEE/EMBS Neural Engineering Conference (NER)*, 2015, pp. 224–227.

[22] 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, 2008.

[23] 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, 1998.

[24] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple classifier systems*. Springer, 2000, pp. 1–15.

[25] E. Bauer and R. Kohavi, "An empirical comparison of voting classification algorithms: Bagging, boosting, and variants," *Machine learning*, vol. 36, no. 1-2, pp. 105–139, 1999.

[26] H. Lee and S. Cho, "Pca+ hmm+ svm for eeg pattern classification," in *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, vol. 1. IEEE, 2003, pp. 541–544.

[27] P. S. Hammon and V. R. De Sa, "Preprocessing and meta-classification for brain-computer interfaces," *Biomedical Engineering, IEEE Transactions on*, vol. 54, no. 3, pp. 518–525, 2007.

[28] N. H. Huy, S. Frenzel, and C. Bandt, "Two-step linear discriminant analysis for classification of eeg data," in *Data Analysis, Machine Learning and Knowledge Discovery*. Springer, 2014, pp. 51–59.

[29] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for eeg-based brain–computer interfaces," *Journal of neural engineering*, vol. 4, no. 2, p. R1, 2007.

[30] J. H. Friedman, "Regularized discriminant analysis," *Journal of the American statistical association*, vol. 84, no. 405, pp. 165–175, 1989.

[31] H. Jasper, "The ten twenty electrode system of the international federation," *EEG Clin. Neurophysiol.*, vol. 10, pp. 371–375, 1958.

[32] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for eeg-based communication," *Electroencephalography and clinical Neurophysiology*, vol. 103, no. 3, pp. 386–394, 1997.

[33] D. H. Johnson, S. Sinanovic *et al.*, "Symmetrizing the kullback-leibler distance," *IEEE Transactions on Information Theory*, vol. 1, no. 1, pp. 1–10, 2001.

[34] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[35] I. Sturm, S. Bach, W. Samek, and K.-R. Müller, "Interpretable deep neural networks for single-trial eeg classification," *arXiv:1604.08201*, 2016.

[36] S. Dähne, F. Bießmann, W. Samek, S. Haufe, D. Goltz, C. Gundlach, A. Villringer, S. Fazli, and K.-R. Müller, "Multivariate machine learning methods for fusing multimodal functional neuroimaging data," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1507–1530, 2015.

[37] S. Fazli, S. Dähne, W. Samek, F. Bießmann, and K.-R. Müller, "Learning from more than one data source: data fusion techniques for sensorimotor rhythm-based brain-computer interfaces," *Proceedings of the IEEE*, vol. 103, no. 6, pp. 891–906, 2015.