

Robust Filter Bank Common Spatial Pattern (RFBCSP) in motor-imagery-based Brain-Computer Interface

Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan

Abstract—The Filter Bank Common Spatial Pattern (FBCSP) algorithm performs autonomous selection of key temporal-spatial discriminative EEG characteristics in motor imagery-based Brain Computer Interfaces (MI-BCI). However, FBCSP is sensitive to outliers because it involves multiple estimations of covariance matrices from EEG measurements. This paper proposes a Robust FBCSP (RFBCSP) algorithm whereby the estimates of the covariance matrices are replaced with the robust Minimum Covariance Determinant (MCD) estimator. The performance of RFBCSP is investigated on a publicly available dataset and compared against FBCSP using 10×10-fold cross-validation accuracies on training data, and session-to-session transfer kappa values on independent test data. The results showed that RFBCSP yielded improvements in certain subjects and slight improvement in overall performance across subjects. Analysis on one subject who improved suggested that outliers were excluded from the robust covariance matrices estimation. These results revealed a promising direction of RFBCSP for robust classifications of EEG measurements in MI-BCI.

I. INTRODUCTION

The Common Spatial Pattern (CSP) algorithm is effective in constructing optimal spatial filters that discriminates two classes of EEG measurements in motor-imagery-based Brain-Computer Interface (MI-BCI) [1], [2]. The Filter Bank Common Spatial Pattern (FBCSP) algorithm was recently proposed to select subject-specific operational frequency band for extracting discriminative CSP features [3]. Although FBCSP is a simple method of selecting appropriate subject-specific band-pass filtering for the CSP algorithm, it performed the best relative to other international submissions in the BCI Competition IV dataset IIa and IIb [4].

The FBCSP algorithm used the classical estimation of multivariate covariance matrices from the EEG measurements for a filter bank of CSP [3]. However, EEG measurements are often contaminated with outliers, such as artifacts or non-standard noise sources [5], that deviate from the usual pattern of the majority of the data [6]. If the EEG measurements are contaminated with even a few extreme outliers, the multivariate covariance estimates typically differs substantially from the estimate without the outliers [6]. Hence, FBCSP is sensitive to outliers in the training data.

This work was supported by the Science and Engineering Research Council of A*STAR (Agency for Science, Technology and Research), Singapore.

The authors are with Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #21-01 Connexis, Singapore 138632. (email: {kkang, zychin, hhzhang, ctguan}@i2r.a-star.edu.sg).

Robust techniques such as channel removal, outlier trial removal, and normalization were first proposed to reduce the influence of outliers in EEG-based BCI in [5]. The study investigated EEG data of 8 subjects and results revealed that outlier trial removal was useful for robust classification of EEG measurements for MI-BCI. However, certain robust technique improved performance of some subjects but deteriorated the performance of others. Hence the study suggested that there is no overall best robust technique and that subject-specific robust technique has to be chosen.

Recently, a robust CSP algorithm was proposed [7] to replace the estimation of the covariance matrices with the robust Minimum Covariance Determinant (MCD) estimator, and the computation of the variance of the CSP projected EEG with the Median Absolute Deviation (MAD). The study investigated EEG data of 5 subjects from the BCI Competition III Dataset IVa using robust CSP and non-robust CSP. The results showed effectiveness of robust CSP on artificially introduced outliers, but results showed slight deterioration of performance compared to non-robust CSP on the data without artificially introduced outliers.

This paper proposes a Robust Filter Bank Common Spatial Pattern (RFBCSP) whereby the estimates of the covariance matrices are replaced with the MCD estimator. Another variant of RFBCSP is also presented that used MCD and replaced the computation of the variance of the CSP projected EEG with MAD. The performances of these two variants of RFBCSP are investigated on the BCI Competition IV dataset IIb [8] using 10×10-fold cross-validations on training data and session-to-session transfer on the test data. The performances are compared with the non-robust FBCSP [3] that out-performed other submissions for this dataset [4].

The remainder of this paper is as follows: Section II and III describe the FBCSP algorithm and the proposed RFBCSP algorithm respectively. Section IV presents the experimental results and section V concludes this paper.

II. FILTER BANK COMMON SPATIAL PATTERN

The Filter Bank Common Spatial Pattern (FBCSP) algorithm [3] illustrated in Fig. 1 comprises 4 progressive stages of signal processing and machine learning on the EEG measurements. The CSP projection matrix for each filter band, discriminative CSP features, and classifier model are computed from training data labeled with the respective motor imagery action. These parameters are used to discriminate motor imagery actions from single-trial EEG measurements in the evaluation phase. The following

describe each stage in more detail:

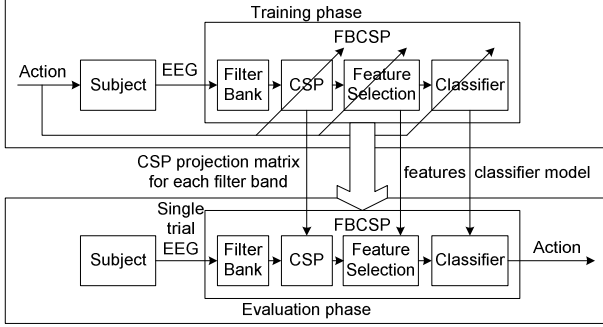


Fig. 1. Architecture of the Filter Bank Common Spatial Pattern (FBCSP) algorithm for the training and evaluation phases.

A. Band-pass filtering

The first stage employs a filter bank that decomposes the EEG into multiple frequency pass bands using causal Chebyshev Type II filter. A total of 9 band-pass filters are used, namely, 4-8 Hz, 8-12 Hz, ..., 36-40 Hz.

B. Spatial filtering

The second stage performs spatial filtering using the CSP algorithm. The CSP algorithm is highly successful in calculating spatial filters for detecting Event Related Desynchronization or Event Related Synchronization (ERD/ERS) in the EEG measurements [1]. Each pair of band-pass and spatial filter in the first and second stage computes the CSP features that are specific to the band-pass frequency range. Spatial filtering is performed using the CSP algorithm by linearly transforming the EEG measurements using

$$\mathbf{Z}_{b,i} = \mathbf{W}_b^T \mathbf{E}_{b,i}, \quad (1)$$

where $\mathbf{E}_{b,i} \in \mathbb{R}^{c \times t}$ denotes the single trial EEG measurement for the i^{th} trial from the b^{th} band-pass filter; $\mathbf{Z}_{b,i} \in \mathbb{R}^{c \times t}$ denotes $\mathbf{E}_{b,i}$ after spatial filtering; $\mathbf{W}_b \in \mathbb{R}^{c \times c}$ denotes the CSP projection matrix for the b^{th} band; c is the number of channels; t is the number of EEG samples per channel; and T denotes transpose operator.

The CSP algorithm computes the transformation matrix \mathbf{W}_b to yield features whose variances are optimal for discriminating 2 classes of EEG measurements [9-12] by solving the eigenvalue decomposition problem

$$\mathbf{\Sigma}_{b,1} \mathbf{W}_b = (\mathbf{\Sigma}_{b,1} + \mathbf{\Sigma}_{b,2}) \mathbf{W}_b \mathbf{D}_b, \quad (2)$$

where $\mathbf{\Sigma}_{b,1}$ and $\mathbf{\Sigma}_{b,2}$ are estimates of the covariance matrices of the b^{th} band-pass filtered EEG measurements of the respective motor imagery action, \mathbf{D}_b is the diagonal matrix that contains the eigenvalues of $\mathbf{\Sigma}_{b,1}$.

Since band-pass EEG measurements have approximately zero mean values, the covariance matrices are estimated by

$$\hat{\mathbf{\Sigma}}_{b,\omega} = \frac{1}{(t \times n_\omega) - 1} \mathbf{E}_{b,\omega} \mathbf{E}_{b,\omega}^T, \quad (3)$$

where $\omega=1, 2$; $\mathbf{E}_{b,\omega} \in \mathbb{R}^{c \times (t \times n_\omega)}$ denotes the concatenated EEG

measurements of all the trials in the training data for the motor imagery action of class ω , t is the number of EEG samples per channel, and n_ω denotes the number of trials in the training data that belongs to class ω .

The spatial filtered signal \mathbf{Z}_b in equation (1) using \mathbf{W}_b from equation (2) thus maximizes the differences in the variance of the 2 classes of band-pass filtered EEG measurements. The m pairs of CSP features for the b^{th} band-pass filtered EEG measurements is then given by

$$\mathbf{cf}_b = \log \left(\text{diag} \left(\bar{\mathbf{W}}_b^T \mathbf{E}_b \mathbf{E}_b^T \bar{\mathbf{W}}_b \right) / \text{tr} \left[\bar{\mathbf{W}}_b^T \mathbf{E}_b \mathbf{E}_b^T \bar{\mathbf{W}}_b \right] \right), \quad (4)$$

where $\mathbf{cf}_b \in \mathbb{R}^{2m}$, $\bar{\mathbf{W}}_b$ represents the first m and the last m columns of \mathbf{W}_b ; $\text{diag}(\cdot)$ returns the diagonal elements of the square matrix; $\text{tr}[\cdot]$ returns the sum of the diagonal elements in the square matrix.

The FBCSP feature vector for the i^{th} trial is then formed as follows

$$\mathbf{x}_i = [\mathbf{cf}_1, \mathbf{cf}_2, \dots, \mathbf{cf}_9], \quad (5)$$

where $\mathbf{x}_i \in \mathbb{R}^{1 \times (9 \times 2m)}$, $i=1, 2, \dots, n$; n denotes the total number of trials in the training data.

The FBCSP feature matrix from training data is then

$$\mathbf{X} = [\mathbf{x}_1^T \quad \mathbf{x}_2^T \quad \dots \quad \mathbf{x}_n^T]^T, \quad (6)$$

where $\mathbf{X} \in \mathbb{R}^{n \times (9 \times 2m)}$.

C. Feature selection

The third stage employs a feature selection algorithm, namely the Mutual Information-based Best Individual Feature (MIBIF) algorithm [13], to select discriminative CSP features from \mathbf{X} for the subject's task.

Given a set of features $\mathbf{F} = [\mathbf{f}_1^T, \mathbf{f}_2^T, \dots, \mathbf{f}_{9 \times 2m}^T]^T = \mathbf{X}$ such that \mathbf{X} is from equation (6), $\mathbf{f}_j^T \in \mathbb{R}^{1 \times 1}$ is the j^{th} column vector of \mathbf{X} ; MIBIF selects k best features that results in the highest estimate of mutual information with the class labels. Based on the study in [3], $k=4$ is used. Since the CSP features are paired, the corresponding CSP features that come in pairs with the selected k features are also included as well. The feature selected training data is denoted as $\bar{\mathbf{X}} \in \mathbb{R}^{n \times d}$ where d ranges from 4 to 8.

D. Classification

The fourth stage employs a classification algorithm, namely the Naïve Bayesian Parzen Window (NBPW) classifier [13], to model and classify the selected CSP features. The classification rule of NBPW is given as

$$\omega = \arg \max_{\omega=1,2} p(\omega | \mathbf{x}), \quad (7)$$

where $p(\omega | \mathbf{x})$ denotes posterior probability of the class being $\omega=1, 2$, given the random trial $\mathbf{x} = [x_1, x_2, \dots, x_d]$ and d denotes the number of selected features from the third stage.

For further details on the FBCSP algorithm, the reader is referred to [3].

III. ROBUST FILTER BANK COMMON SPATIAL PATTERN

Many multivariate datasets contain outliers that deviate from the patterns of a majority of the data [5], [6]. Thus the use of the classical multivariate method to estimate the covariance matrices of EEG measurements in the CSP algorithm is sensitive to outliers [7]. Robust statistics, on the other hand, provides alternatives to the classical statistical estimates that are not affected by outliers [14]. A measure of robustness of an estimator is the breakdown value, which states the smallest amount of outlier contamination that can have arbitrarily large effect on the estimator [6].

A. Minimum Covariance Determinant estimate

The non-robust CSP algorithm has an inherent breakdown value of 0 [7]. Thus the robust CSP algorithm was proposed [15] to replace the classical multivariate estimate with the Minimum Covariance Determinant (MCD) estimator.

The FBCSP algorithm, which uses the classical multivariate estimate in equation (3), also has an inherent breakdown value of 0. Therefore, the Robust Filter Bank Common Spatial Pattern (RFBCSP) is proposed to replace equation (3) with the MCD estimator

$$\hat{\Sigma}_{b,\omega} = \frac{1}{(\alpha \times t \times n_\omega) - 1} \hat{\mathbf{E}}_{b,\omega} \hat{\mathbf{E}}_{b,\omega}^T, \quad (8)$$

where

$$\hat{\mathbf{E}}_{b,\omega} = \arg \min_{\hat{\mathcal{E}} \subset \mathcal{E}} |\hat{\Sigma}_{b,\omega}|; \quad (9)$$

$(1-\alpha)$ is the fraction of outliers to resist, $\alpha=[0.5, 1]$; \mathcal{E} is the set of $t \times n_\omega$ c -dimensional elements of $\mathbf{E}_{b,\omega} \in \mathbb{R}^{c \times (t \times n_\omega)}$; $\hat{\mathcal{E}}$ is the subset of \mathcal{E} containing $\alpha \times t \times n_\omega$ c -dimensional elements of $\hat{\mathbf{E}}_{b,\omega} \in \mathbb{R}^{c \times (\alpha \times t \times n_\omega)}$; and $|\cdot|$ denotes determinant.

The MCD estimator in equation (9) thus computes a defined fraction α of the data such that the determinant of the estimate of the covariance matrix is minimized. The issues of the MCD estimator are that it depends on the initial estimates and it is iterative [15]. The FASTMCD algorithm resolves these issues by drawing multiple random subsets of the data and iteratively approximates towards a lower determinant [15]. The implementation of FASTMCD is available as the MATLAB function “mcdcov” in the LIBRA Toolbox from [16].

B. Median Absolute Deviations estimate

The non-robust CSP algorithm uses the classical variance estimate of the spatial filtered signal. Hence, the robust CSP algorithm proposes replacing the classical variance estimate with the Median Absolute Deviation (MAD) estimate [7].

Similarly, in the proposed RFBCSP algorithm, the use of the classical variance estimate of the spatial filter signal in the equation (4) can be replaced with the Median Absolute Deviation (MAD) estimate [14] given by

$$\mathbf{cf}_b = \left(1.4826 \times \text{med} \left(\bar{\mathbf{W}}_b^T \mathbf{E}_b - \text{med} \left(\bar{\mathbf{W}}_b^T \mathbf{E}_b \right) \right) \right)^2, \quad (10)$$

where $\text{med}(\mathbf{A})$ gives the median of the rows of \mathbf{A} .

IV. EXPERIMENTAL RESULTS

This section evaluates the performance of the proposed RFBCSP algorithm on the BCI Competition IV dataset IIB [8]. The dataset consists of 9 subjects whereby the training data for each subject comprises 3 sessions of single-trial EEG data from 3 bipolar recordings (C3, Cz and C4) while the subject performed 2-class hand motor imagery. The first 2 sessions consist of 240 single trials without visual feedback and the third session consist of 160 single trials with visual feedback to the subject. The selections of the training sessions to be used as training data for each subject are based on the winners of the BCI Competition IV dataset IIB [4]. The evaluation data consists of 2 sessions of single-trial EEG data, totaling 320 trials.

The running classification performance of the proposed RFBCSP with MCD (denoted MCD), and the variant of RFBCSP with MCD and MAD (denoted MCDMAD) are compared against the non-robust FBCSP on the training data using 10×10-fold cross-validations. For running classification, the same EEG time segment of 0.5 to 2.5s relative to the visual cue is used for training and validation [17]. The MCD and MCDMAD are configured with the default $\alpha=0.75$ implemented in the “mcdcov” function from the LIBRA toolbox [16]. The results presented in Table I showed that MCD yielded better accuracies for subjects 2, 7 and 9 (shaded in gray). The results also showed that MCD yielded slightly better overall accuracy than non-robust FBCSP (79.21% versus 79.20, $p=0.97$) and MCDMAD yielded poorer accuracy (78.06%, $p=0.20$), but both are not statistically significant using paired-sample t -test.

TABLE I
RESULTS IN VALIDATION ACCURACIES OF RUNNING CLASSIFICATION USING FBCSP, RFBCSP WITH MCD, AND RFBCSP WITH MCDMAD

Method	Subjects									AVG
	1	2	3	4	5	6	7	8	9	
FBCSP	77.32	56.73	61.05	98.63	85.88	81.81	85.88	87.06	78.44	79.20
	± 0.51	± 1.15	± 1.28	± 0.26	± 0.60	± 1.54	± 2.54	± 0.42	± 0.53	± 0.98
MCD	77.18	58.13	59.85	98.56	85.81	79.94	86.88	87.44	79.13	79.21
	± 0.39	± 1.57	± 1.34	± 0.30	± 1.10	± 1.57	± 2.19	± 0.46	± 0.84	± 1.09
MCDMAD	78.32	53.60	57.93	97.50	80.63	78.50	89.00	87.81	79.25	78.06
	± 0.92	± 1.73	± 1.46	± 0.00	± 1.21	± 1.75	± 0.44	± 0.53	± 0.87	± 0.99

The session-to-session transfers of MCD, MCDMAD are then evaluated and compared with the non-robust FBCSP on the evaluation data in terms of kappa values. The kappa value is computed using the BIOSIG toolbox [18]. For static classification, the EEG time segment of 0.5 to 2.5s relative to the visual cue is used for training, then the entire time segment from a single trial is used for evaluation [17]. The results presented in Table II showed that MCD yielded better kappa values for subjects 1, 3, 5 and 6 (shaded in gray). The results also showed that MCD yielded slightly better overall kappa value compared to the non-robust FBCSP (0.606 versus 0.585, $p=0.11$) and MCDMAD yielded poorer overall kappa value (0.568, $p=0.06$), but both are not statistically significant using paired-sample t -test. It is noted that the overall kappa value of RFBCSP with MCD is relatively

better than the winner of this dataset [4].

TABLE II
EXPERIMENTAL RESULTS IN KAPPA VALUES OF STATIC CLASSIFICATION
USING FBCSP, RFBCSP WITH MCD, AND RFBCSP WITH MCDMAD

Method	Subjects									AVG
	1	2	3	4	5	6	7	8	9	
FBCSP	0.356	0.171	0.169	0.963	0.850	0.594	0.556	0.856	0.750	0.585
MCD	0.363	0.171	0.256	0.956	0.869	0.669	0.563	0.856	0.750	0.606
MCDMAD	0.319	0.143	0.150	0.931	0.775	0.569	0.594	0.888	0.744	0.568

Since the results in Table II showed that the kappa value of subject 3 is significantly improved from 0.169 for non-robust FBCSP to 0.256 for MCD, a further analysis is performed to investigate the band-pass filtered EEG measurements that are selected by RFBCSP but excluded by MCD. Fig. 2(a) shows the plot of a single trial of EEG measurements from the training data of subject 3 that are band-pass filtered from 4-8 Hz without outliers identified by MCD. Fig. 2(b) shows the plot of a single trial of EEG measurements with more than half of the time-samples identified as outliers by MCD. The plot of the latter shows a relatively high amplitude EEG data that are band-passed using 4-8 Hz, which could suggest the presence of artifacts.

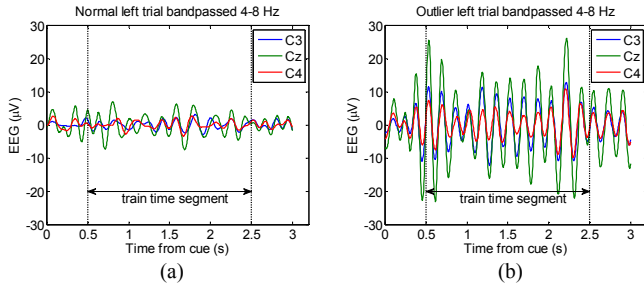


Fig. 2. (a) Plot of a signal trial from the training data of subject 3 without outliers identified by MCD, (b) plot of a trial from the training data of subject 3 with more than half time samples identified as outliers by MCD.

V. CONCLUSIONS

The Filter Bank Common Spatial Pattern (FBCSP) algorithm is sensitive to outliers because it involves multiple estimations of covariance matrices from the EEG measurements. This paper proposes a Robust FBCSP (RFBCSP) algorithm whereby the estimates of the covariance matrices are replaced with the robust Minimum Covariance Determinant (MCD). Another variant of RFBCSP that used MCD and replaced the variance estimate of the spatial filter signal with the Median Absolute Deviation (MAD) is also presented. The performances of RFBCSP with MCD, and RFBCSP with MCD and MAD are investigated on the BCI Competition IV dataset IIB and compared against the non-robust FBCSP using running and static classification.

Experimental results showed that RFBCSP with MCD yielded improvements in certain subjects. The results also showed that RFBCSP with MCD yielded slight overall improvements in both static and running classification, even though the improvements are not statistically significant. However, the results showed that RFBCSP with MCD and MAD yielded deterioration. These results suggest a

promising direction of replacing the classical multivariate covariance estimates with the robust MCD, but not to replace the variance estimate of the spatial filter signal with MAD. Further analysis on one of the subjects with improved performance suggested that outliers were excluded from the estimation of the covariance matrices. Thus the results in this study revealed a promising direction in RFBCSP on robust classifications of EEG measurements for MI-BCI.

ACKNOWLEDGMENT

The authors would like to thank the organizers [4] and dataset providers of the BCI Competition IV dataset IIB [8].

REFERENCES

- [1] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539-550, Aug. 2007.
- [2] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," *Proc. IEEE*, vol. 89, no. 7, pp. 1123-1134, 2001.
- [3] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," in *Proc. IJCNN'08*, 2008, pp. 2391-2398.
- [4] B. Blankertz, "BCI Competition IV", Fraunhofer FIRST (IDA), http://ida.fraunhofer.de/projects/bci/competition_iv/, 2008.
- [5] M. Krauledat, G. Dornhege, B. Blankertz, and K.-R. Müller, "Robustifying EEG data analysis by removing outliers," *Chaos Complex. Lett.*, vol. 2, no. 3, pp. 259-274, 2007.
- [6] M. Hubert, P. J. Rousseeuw, and S. Van Aelst, "High-Breakdown Robust Multivariate Methods," *Stat. Sci.*, vol. 23, no. 1, pp. 92-119, 2008.
- [7] X. Yong, R. K. Ward, and G. E. Birch, "Robust Common Spatial Patterns for EEG Signal Preprocessing," in *Proc. EMBC'08*, 2008, pp. 2087-2090.
- [8] R. Leeb, F. Lee, C. Keinrath, R. Scherer, H. Bischof, and G. Pfurtscheller, "Brain-Computer Communication: Motivation, Aim, and Impact of Exploring a Virtual Apartment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 15, no. 4, pp. 473-482, Dec. 2007.
- [9] B. Blankertz, F. Losch, M. Krauledat, G. Dornhege, G. Curio, and K. R. Müller, "The Berlin Brain-Computer Interface: Accurate Performance From First-Session in BCI-Naive Subjects," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 10, pp. 2452-2462, 2008.
- [10] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441-446, Dec. 2000.
- [11] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, no. 5, pp. 787-798, 1999/5/1 1999.
- [12] K. Fukunaga, "Introduction to Statistical Pattern Recognition," 2nd ed. New York: Academic Press, 1990.
- [13] K. K. Ang and C. Quek, "Rough Set-based Neuro-Fuzzy System," in *Proc. IJCNN'06*, 2006, pp. 742-749.
- [14] P. J. Huber and E. M. Ronchetti, *Robust Statistics*, 2nd ed. U.S.: John Wiley and Sons, 2009.
- [15] P. J. Rousseeuw and K. V. Driessen, "A Fast algorithm for the Minimum Covariance Determinant Estimator," *Technometrics*, vol. 41, no. 3, pp. 212-223, 1999.
- [16] M. Hubert, "LIBRA: A Matlab Library for Robust Analysis", Katholieke Universiteit Leuven, <http://wis.kuleuven.be/stat/robust/LIBRA.html>, 2009.
- [17] M. Naeem, C. Brunner, R. Leeb, B. Graimann, and G. Pfurtscheller, "Seperability of four-class motor imagery data using independent components analysis," *J. Neural Eng.*, vol. 3, no. 3, p. 208, 2006.
- [18] A. Schlögl, "BIOSIG - an open source software library for biomedical signal processing", Graz University of Technology, <http://biosig.sourceforge.net/>, 2009.