Spatial Filtering for Robust Myoelectric Control

Janne Mathias Hahne*, Bernhard Graimann, and Klaus-Robert Müller

***Abstract*—Pattern recognition techniques have been applied to extract information from electromyographic (EMG) signals that can be used to control electrical powered hand prostheses. In this paper, optimized spatial filters that enhance separation properties of EMG signals are investigated. In particular, different multiclass extensions of the common spatial patterns algorithm are applied to high-density surface EMG signals acquired from the forearms of ten healthy subjects. Visualization of the obtained filter coefficients provides insight into the physiology of the muscles related to the performed contractions. The CSP methods are compared with a commonly used pattern recognition approach in a six-class classification task. Cross-validation results show a significant improvement in performance and a higher robustness against noise than commonly used pattern recognition methods.**

***Index Terms*—Common spatial pattern (csp), hand prostheses, myoelectric control, prosthetic control, prosthetics, spatial filters.**

I. Introduction

MUSCLE activity measured noninvasively by surface electromyography (sEMG), can be used to control electrical powered hand prostheses.

With conventional, commercially available myoprostheses, only 1 degree of freedom (DOF) can be controlled directly. This is achieved by a very simple system with only two bipolar electrodes that are usually placed on the flexors and extensors of the residual limb for individuals with a transradial amputation.

A contraction of the flexors causes the prosthesis to close, of the extensors to open. To control another DOF like rotation of the wrist, the user has to switch the active DOF by a cocontraction of both muscle groups. This is slow and cumbersome and limits the benefit of prostheses with more active DOF [1].

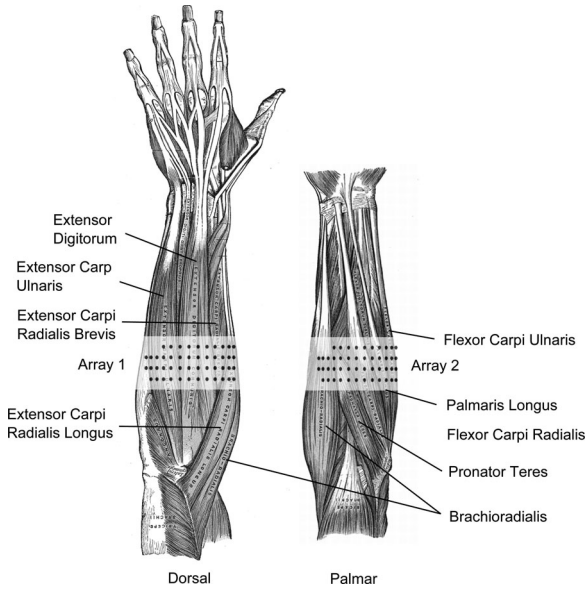

Fig. 1. Anatomy of the forearm and approximated electrode position (modified from [8]).

To overcome these limitations, more electrodes and multivariate multiclass pattern recognition methods have been applied in the last decades with the aim to directly control more functions more intuitively and without mode-switching. [2]. Hudgins and Englehart developed different time-domain, frequency-domain, and wavelet-based features and applied them in combination with different linear and nonlinear classifiers [3].

However, so far those methods are not reliable enough under real world conditions to be used for clinical application. For example, displacement of the electrodes [4] or changing the position of the arm [5] can degrade the performance significantly.

Huang *et al*. [6] showed that fixed local spatial filters like double differential or Laplacian on high dimensional EMG signals can improve classification accuracy on subjects whose arm nerves were selectively transferred to other body regions in a targeted muscle reinnervation surgery.

In this paper, we investigate spatial filters whose coefficients are adapted to each subject using labeled training data. The method common spatial patterns (CSP) has been shown to be very powerful for binary classification problems of electroencephalography (EEG) signals (see [7] for a review). We discuss several multiclass extensions and apply them to high-dimensional EMG signals of healthy subjects. The resulting filters are accessible for physiological interpretation. In a classification task with six different limb movements, the algorithms are compared to a classical pattern recognition approach.

Manuscript received November 15, 2011; revised January 18, 2012; accepted January 28, 2012. Date of publication February 23, 2012; date of current version April 20, 2012. This work was supported by the Marie Currie Industry-Academia Partnerships and Pathways (IAPP) Grant "AMYO," under Project 251555 and 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. *Asterisk indicates corresponding author.*

*J. M. Hahne is with the Machine Learning Laboratory, Berlin Institute of Technology, D-10587 Berlin, Germany, and with the Strategic Technology Management, Otto Bock Healthcare GmbH, Duderstadt D-37115, Germany (e-mail: janne.hahne@tu-berlin.de).

B. Graimann is with the Strategic Technology Management, Otto Bock Healthcare GmbH, Duderstadt D-37115, Germany, and also with the Laboratory of Brain-Computer Interfaces, Institute for Knowledge Discovery, Graz University of Technology, A-8010 Graz, Austria (e-mail: bernhard.graiman@ottobock.de).

K.-R. Müller is with the Machine Learning Laboratory, Berlin Institute of Technology, D-10623 Berlin, Germany, with the Bernstein Center for Neurotechnology Berlin, D-10587 Berlin, Germany, and also with the Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea (e-mail: klaus-robert.mueller@tu-berlin.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2012.2188799

II. Methods

A. Experimental Setup

For our study, 96 monopolar EMG signals where recorded from the nondominant forearms of ten healthy subjects, five females and five males. We used "Brain Amp DC" biosignal amplifiers with a sampling frequency of 2500 Hz and two Ag/AgCl electrode arrays with an inter electrode distance of 8 mm (OT-Bioelettronica ELSCH064R3S). The arrays were placed at 1/3 of the distance from elbow to wrist. For a reproducible electrode placement in tangential direction, the location of the ulna was used, which was identified by palpation. The approximate position of the electrodes relative to the forearm muscles is shown in Fig. 1.

The subjects were instructed to perform six different classes of hand movements (hand open, hand close, pronation, supination, wrist flexion, and wrist extension) in five different arm positions (arm down, half-lifted, extended to the front, in front the body, and in front of the face). Visual feedback of the applied force was provided and calculated by the root mean square (200-ms moving window), averaged over all channels and scaled to maximum voluntary contraction (mvc). To adapt the scaling, the subjects had to perform an mvc for each class–position–combination in the beginning of the session.

Each trial was formed by a contraction of 3 s with approximately 40% force. The trials were recorded in several runs, where each run contained each class–position–combination exactly once. In-between two trials was a break of 4 s, in which the subject had to change the arm position. The order of trials within each run was randomized in class and position. When the subject or the experimenter noticed a failure (e.g., wrong or delayed contraction, wrong arm position) in the execution of a trial, it was marked during the experiment and not used in the evaluation of this study. The number of removed trials was very low for all subjects (0 to 0.83% of trials, average $0.41\% \pm 0.28\%$).

After each run the subject could have a break and relax for some minutes. Each subject performed 25 to 35 runs. In this way, 750 to 1050 trials were obtained from each subject. These are enough independent data for a proper statistical evaluation.

For this study, only the static part of the contractions was used, i.e., the beginning and end of each contraction was removed. All signals were filtered with a low pass (500 Hz, fourth-order Butterworth), a high pass (20 Hz, fourth-order Butterworth), and a band stop filter (45–55 Hz, second-order Butterworth) to remove noise, movement artifacts, and power line interferences.

The experiments are in accordance with the declaration of Helsinki and approved by the local ethics commission.

B. Common Spatial Patterns

CSP [7] is a supervised algorithm to obtain linear spatial filters that maximize the variance of one class and at the same time minimize the variance of another class.In this way, the classes become maximally separated by their variances. This method was very successfully applied in classification of motor imaginary EEG data for a brain computer interface (BCI) [9], [10].

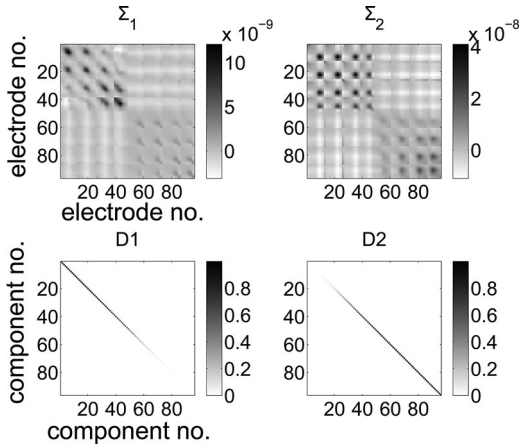

Fig. 2. CSP example: visualization of the covariance matrices $\boldsymbol{\Sigma_1}$, $\boldsymbol{\Sigma_2}$ and the diagonal matrices $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$.

Let $\boldsymbol{X_c}$ denote a $d \times n$ matrix containing d monopolar, temporal filtered, zero mean signals of class c in its rows with n time samples each. We aim to find the coefficients $\boldsymbol{w}$ of the spatial filter $\boldsymbol{y} = \boldsymbol{w^T X}$ whose output $\boldsymbol{y}$ has maximal variance when signals of the corresponding class are applied and minimal for another class. With $\boldsymbol{\Sigma_c} = 1/(n-1) * \boldsymbol{X_c} * \boldsymbol{X_c^T}$ being the empirical covariance matrix of $\boldsymbol{X_c}$, the variance of the filtered signal is $\boldsymbol{w^T \Sigma_c w}$ and the optimization problem can be formulated as follows:

$$\boldsymbol{w} = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w^T \Sigma_1 w}}{\boldsymbol{w^T \Sigma_2 w}}. \tag{1}$$

The solution can be obtained by solving a generalized eigenvalue problem $\boldsymbol{\Sigma_1 w} = \lambda \boldsymbol{\Sigma_2 w}$, i.e., finding the Matrix $\boldsymbol{W}$ that diagonalizes both, $\boldsymbol{\Sigma_1}$ and $\boldsymbol{\Sigma_2}$ (2):

$$\boldsymbol{W \Sigma_1 W^T} = \boldsymbol{D_1} \tag{2a}$$

$$\boldsymbol{W \Sigma_2 W^T} = \boldsymbol{D_2} \tag{2b}$$

$$\boldsymbol{D_1} + \boldsymbol{D_2} = \boldsymbol{I}. \tag{3}$$

This gives d spatial filters as row vectors of $\boldsymbol{W}$. Applying the full filter matrix $\boldsymbol{W}$ to the raw signals would give d output signals $\boldsymbol{Y} = \boldsymbol{W} * \boldsymbol{X}$, so called components. The variances of each component when applying class one would be indicated by the corresponding eigenvalues λ_i^1 on the principal diagonal of $\boldsymbol{D_1}$, for class two of $\boldsymbol{D_2}$. With the constraint (3), the eigenvalues of both classes sum up to one for each component. Choosing the component with the highest λ_i^1 for $\boldsymbol{X_1}$ has automatically the lowest λ_i^2 for $\boldsymbol{X_2}$ and is the solution of (1).

An example of the covariance matrices and the diagonal matrices $\boldsymbol{D_1}$ and $\boldsymbol{D_2}$ are shown in Fig. 2. The block structure of the covariance matrices is caused by the row wise positioning of the channels within the electrode arrays. The corresponding eigenvalue spectra are shown in Fig. 4.

The eigenvalues λ_i^1 correspond to activity in component space if the first class is active. It can be seen that for the first components λ are close to one, corresponding to high activity and close to zero for the last components, corresponding to low activity.

Assuming the filters in $\boldsymbol{W}$ are sorted in a way that the eigenvalues in $\boldsymbol{D_1}$ are in descending order, the first and the last components have the best separation properties. However, when dealing with real biosignals, often more components are used to increase the robustness [7].

C. Multiclass CSP

Although the classical CSP approach is limited to two classes, there exist several options of multiclass extension. One way is to combine several binary CSP filters in a one versus rest (OvR) or a one versus one (OvO) combination scheme. Another approach is to extend the optimization problem to a simultaneous diagonalization of more than two covariance matrices [11].

1) One Versus Rest: In the OvR approach each filter is optimized to maximize the variance of the corresponding class and minimize the sum (or average) of the variances of all other classes:

$$\boldsymbol{w_c} = \arg\max_{\boldsymbol{w}} \frac{\boldsymbol{w_c^T \Sigma_c w_c}}{\boldsymbol{w_c^T}\left(\sum_{i\neq c}\boldsymbol{\Sigma_i}\right)\boldsymbol{w_c}}. \tag{4}$$

This can be achieved by solving the generalized eigenvalue problem with $\boldsymbol{\Sigma_1} = \boldsymbol{\Sigma_c}$ and $\boldsymbol{\Sigma_2} = \sum_{i\neq c}\boldsymbol{\Sigma_i}$ for each class c. Since we are interested in the components with high variance for the active class, only the filters with highest eigenvalues in $\boldsymbol{D_1}$ are selected. This is repeated for each class. The features of all selected components are classified by a multiclass classifier.

The number of components is growing linear with the number of classes. Since only the average of the other classes is minimized, there is no guarantee that each individual counter class results in a low variance output. Thus, the filters might be not optimal in terms of separating properties.

2) One Versus One: In the OvO approach the binary CSP is applied for all possible class combinations. The filters are chosen in the same way as in the binary case for each combination. The components have optimal separation properties for the classes belonging to each combination, but the response for unrelated classes is not determined.

The number of filters with that approach grows quadratically with the number of classes. There are $N_{\mathrm{comb}} = (N_c^2 - N_c)/2$ combinations for N_c classes. This can result in a high-dimensional feature space and lead to problems in the classification task. Regularization might be needed if too little training data are available. In this analysis, the features of all components are concatenated into one feature vector and applied to a multiclass classifier.

Another approach is to apply for each class combination a parallel classifier and combine the classification outputs. In a prestudy, this led to worse results than concatenating features of all combinations and is not applied here. Nevertheless, this approach might be interesting to increase the robustness against false activation in a real world scenario (see also [12]).

3) Joint Diagonalization: Another approach is the simultaneous diagonalization of more than two covariance matrices, i.e., to find a filter matrix $\boldsymbol{W}$ that diagonalizes the covariance matrices of all classes: $\boldsymbol{W\Sigma_c W^T} = \boldsymbol{D_c}$, with $\sum_{\boldsymbol{c}}\boldsymbol{D_c} = \boldsymbol{I}$.

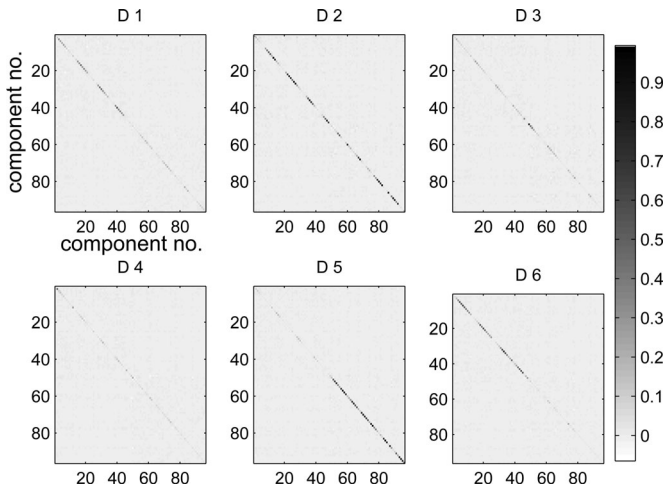

Fig. 3. Example of the joint diagonalization of the covariance matrices of six classes.

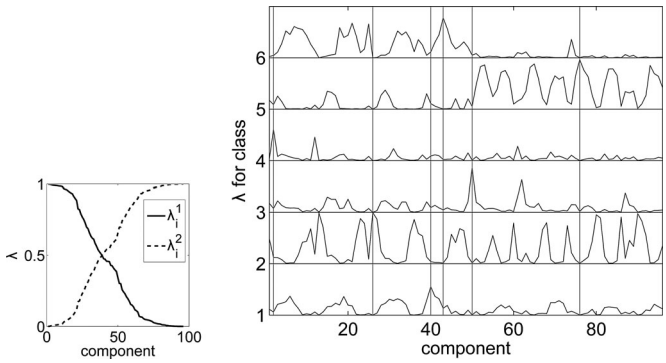

Fig. 4. Example of eigenvalue spectra for binary CSP (l.) and joint diagonalization (r.).

Contrary to the two class case the solution can be only approximated. We use the FFDIAG algorithm described in [13].

The approximated diagonal matrices of the six class problem are shown for a representative subject in Fig. 3. The standard error of the off-diagonal elements was for all subjects smaller than 5×10^{-3}, indicating a successful approximation.

The elements of the principal diagonals of D_c form the eigenvalue spectrum for class c. As in the binary case the eigenvalue λ_i^c indicates the variance (power) of component i for signals of class c.

Selecting good components for classification is not as straight forward as in the binary case, where the component with highest eigenvalue for one class has automatically the lowest eigenvalue for the other class as seen in Fig. 4. Dornhege *et al*. [11] investigated different strategies in a BCI context. In our study, the best results were obtained by adding for each class the component with highest eigenvalues λ_i^c. If one component would be selected for two classes, for the class with lower λ_i^c the component with the next highest eigenvalue λ_j^c is selected.

The eigenvalue spectra for an representative example are shown in Fig. 4(r). If λ_i^c is close to one for a class and very low for all other classes this component has good separation properties. In the given example, this is the case for some of

the classes (e.g., component 75 for class 5). In other cases (e.g., class one), the highest λ_i^c is much lower than one and the corresponding component will capture power also for other classes. The first components that are selected for each class are marked by the vertical lines.

D. Feature Extraction

For all methods, the features were extracted from blocks of 200 ms, overlapping each other by 150 ms. This scheme was chosen in order to fulfill the time constraints that would arise in a real-time application [14]. Longer window sizes would increase the classification accuracy, but cause longer delays that reduces the usability of an online control system. The choice of 200-ms window length is a good compromise in this tradeoff.

To capture the discriminative information of the CSP-components variance-sensitive features have to be applied. We used the logarithm of the block-wise calculated variance. This has been shown to work well in combination with linear discriminant analysis on EEG signals [7]. The logarithm causes a compression of the value range, which results in a feature distribution that is closer to Gaussian.

E. Bipolar Derivation and Hudgins Feature Set

In order to compare the performance of the proposed algorithms, a commonly used method was implemented to obtain a reference performance. A set of bipolar channels was obtained by subtracting couples of monopolar channels from each other, aligned with the direction of the muscle fibers. This can be seen as a fixed, a nonsupervised spatial filter. From the bipolar signals the Hudgins feature set (mean absolute value, zero crossing, slope sign changes, and wavelength) was extracted in the same block-wise scheme as for the other methods. Englehart and Hudgins found out that for steady-state EMG signals in a real-time processing scheme those rather simple features outperform many other, more complex methods [3]. The Hudgins feature set in combination with bipolar signals was used in many other studies and can be seen as the state of the art for myoelectric control.

F. Classification

For both, classification of CSP features in BCI [7] and of the Hudgins feature set in myoelectric control [3], linear discriminant analysis after Fisher has been proposed. This is the optimal Bayes classifier for data that is Gaussian distributed and has the same covariances for all classes in feature space.

Since the dimensionality of the feature space can become quite large depending on the number of selected components and the type of multiclass extension, a shrinkage regularized version of LDA was applied (see [15]). The covariance matrices (in feature space) were regularized toward their values on the primary diagonal, i.e., their variances. The regularization parameter was automatically selected solely based on the training set [16].

All arm positions were included in training and test sets. For extensive investigations on training strategies for position robust control with reduced amount of training data see [5].

For a practical application, e.g., for the control of hand prostheses, it is not suitable to have too many channels. The power consumption and the production costs would be too high for a clinical application. Therefore, we investigated the classification performance with a reduced number of 22 channels in two rows (distance 16 mm) of 11 equally spaced channels around the forearm. All classification results were obtained from the reduced channel set.

In case of bipolar derivation, the couples were formed by opposite channels from both rows. This results in bipolar channels with an inter electrode distance of 16 mm and orientation approximately longitudinal to the muscle fibers.

G. Noise Investigations

Due to the low electrode-to-skin impedances of the wet Ag/AgCl electrodes and the high-quality biosignal amplifiers (very high input impedance and common mode rejection ratio), the noise level of the signals is rather low. For a practical application in a prostheses control, dry electrodes are needed. This can increase the noise level depending on the skin conditions.

Therefore, in a second analysis, the robustness of all methods to an increased noise level is investigated. To all monopolar signals (training and test data) uncorrelated white Gaussian noise was added. The noise was filtered in the same way as the EMG, so only the frequency components in the pass bands affected the signals. Five different noise-levels with an postfilter-amplitude of 1–100 μV_{rms} are investigated.

H. Cross Validation

To estimate the performance of machine learning methods, the data have to be split into training and test data. All parameters, including the classifier model, possible regularization parameters, and the coefficients of the spatial filters have to be determined only based on the training data [17].

The performance obtained in this way is a stochastic value that is strongly influenced by the choice of the training and test dataset. To decrease the variance of this value, a k-fold cross validation can be applied. The data are split into k parts. The test set is formed by one part, the training set by the other $k-1$ parts. This is repeated k times so that each part forms once the test set. Averaging the k results gives a more stable performance approximation.

It is important that training and test set are not only disjoint, but also independent of each other. If a subject would for example perform the same type of contraction several times after each other, the trials would be less independent, as if other contraction types are performed in between [18].

In this study fivefold cross validation was used. The parts were formed by entire runs. Due to the randomization of trials and the pauses in between two runs the dependencies between trials of the training and test parts parts are reduced to a minimum. Each part consisted of five to seven runs.

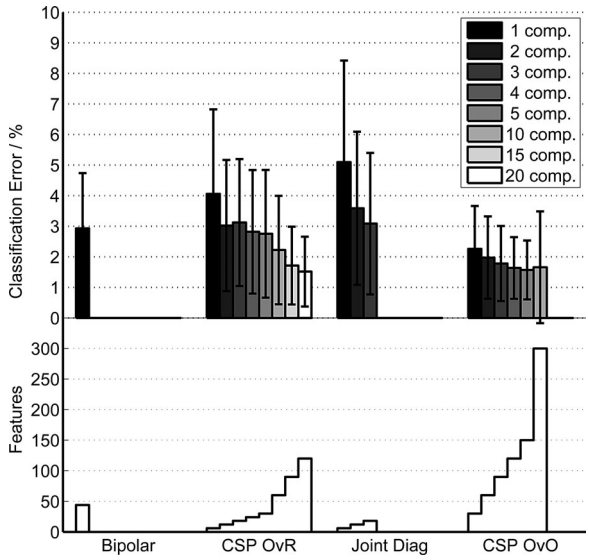

Fig. 5. Classification error from cross validation, mean, and standard deviation over all subjects (u.) and dimensionality of feature space (d.). "comp." referes to components.

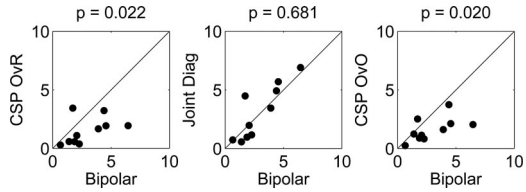

Fig. 6. Comparison of the classification errors with the reference method *bipolar* for each subject and *p*-values for *t*-test.

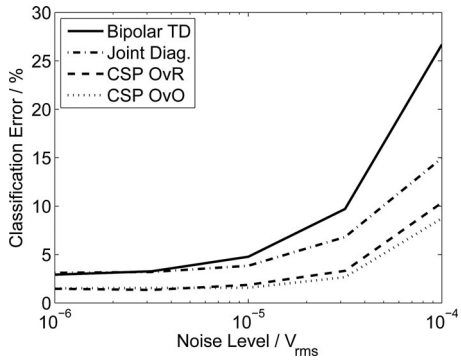

Fig. 7. Classification error with added noise.

III. Results

In the following section, we show classification results for the proposed methods and compare them with a state-of-the-art approach described in Section II-E. An interpretation of the CSP-filters is done in Section IV.

A. Cross-Validation Results

To compare the performance of the investigated methods, the mean classification error is shown in Fig. 5. The error bars represent the subject-to-subject standard deviation. For all methods except *bipolar*, results are presented with different numbers of components. The gray level indicates the number of components per class for *CSP OvR* and *Joint Diagonalization* or per class in each combination for *CSP OvO*. The resulting total number of components is indicated by the curve under the bars, which shows the dimensionality of the feature space.

It is seen that choosing more components within each of the proposed CSP extensions improves the performance. Even if the total number of components becomes larger than the number of raw signals, the performance can be further increased.

If a sufficient number of components is selected *CSP OvR* and *CSP OvO* perform better than the reference method *Bipolar*. They have approximately the same error rates for the same number of total components. For the highest number of investigated components at *CSP OvO*, the error is slightly increased. This might be caused by overfitting, due to the high-dimensional feature space and limited amount of available training data.

For *Joint Diagonalization*, it is not possible to select more components than raw signals. With the highest choice of three components per class it does not perform better than the reference method.

In Fig. 6, all CSP methods are directly compared with the reference method *bipolar* for each subject individually. For each method only a fixed number of components is shown. The total number of components was 18 in case of *Joint Diagonalization* (three components times six classes), 120 for *CSP OvR* (20 components times six combinations) and 120 for *CSP OvO* (eight components times 15 combinations). Above each plot the *t*-test significance value is shown for that comparison.

CSP OvR and *CSP OvO* perform better than the reference method for all but one subject. The difference is significant in both cases, indicated by the low *p*-values of 0.022 and 0.020. For *Joint Diagonalization* there is no significant improvement. Between *CSP OvR* and *CSP OvO* there is no significant difference in performance ($p = 0.418$).

B. Noise Investigations

Fig. 7 shows the degradation of classification error depending on the level of added noise. The number of components is the same as for the significance tests (totally 120 for *CSP OvO* and *CSP OvR*, 18 for *Joint Diagonalization*).

For all methods, the classification error is monotonically increasing with increasing noise level. The drop in performance is strongest in the case of bipolar derivation. *CSP OvR* and *CSP OvO* that performed already best without additional noise show also the lowest error under noisy conditions.

IV. Discussion

A. Interpretation of Spatial Filters

To understand the properties of the spatial CSP filters, we visualize the corresponding patterns for one representative subject. Patterns are columns of the inverse filter matrix $\boldsymbol{A} = \boldsymbol{W}^{-1}$ and are easier to interpret as the filter coefficients. The *i*th pattern can be seen as that source signal distribution to the sensors that produces activity in the *i*th CSP component. In the graphical representations (see Figs. 8, 9, and 10, colors only in online

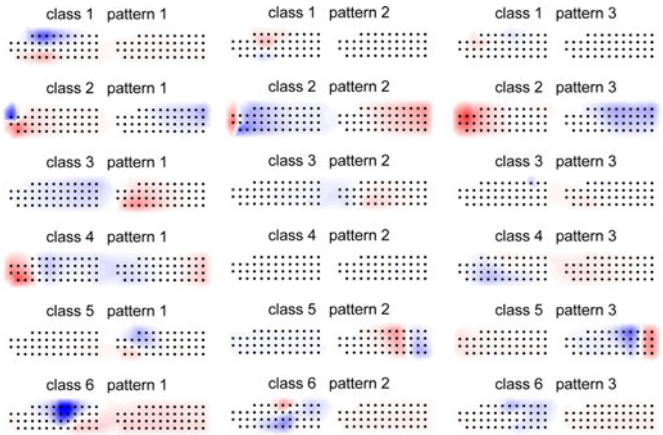

Fig. 8. First three patterns for each class of CSP OvR.

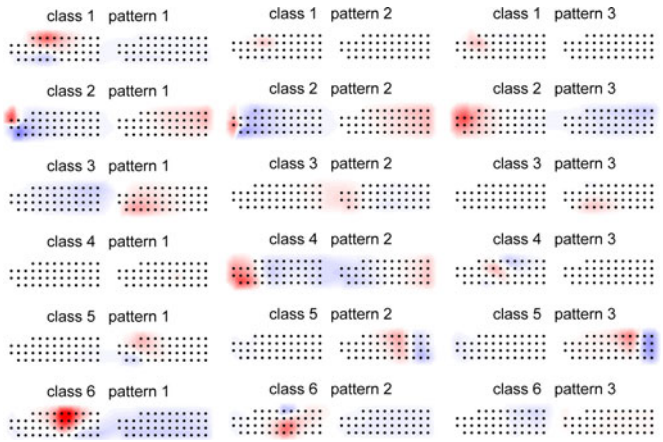

Fig. 9. First three patterns for each class of CSP joint diagonalization.

version available) blue corresponds to negative entries of A, red to positive entries, and white to values close to zero. Of particular interest are the first and last patterns, since they show the sources of the most discriminative components.

We compare the CSP patterns with a graphical representation of the signal variances (see Fig. 11) that show the distribution of EMG activity over the arrays. Each dot corresponds to one sensor and the variances in between are estimated by cubic interpolation. Since the range of variances is quite different for different classes, each plot is scaled individually. White corresponds to no activity and black to the variance of the strongest channel.

B. Patterns of CSP OvR and Joint Diagonalization

The patterns of *joint diagonalization* (see Fig. 9) and *CSP OvR* (see Fig. 8) are very similar. In some cases, the signs are exchanged, but this has no influence on the variance of the corresponding components. The reason for that similarity can be found in the optimization criteria of these two approaches, which are actually not so different. Even if *CSP OvR* does not diagonalize each of the individual covariance matrices of the "rest classes," in both methods the variance for one class is maximized and the sum of the other minimized.

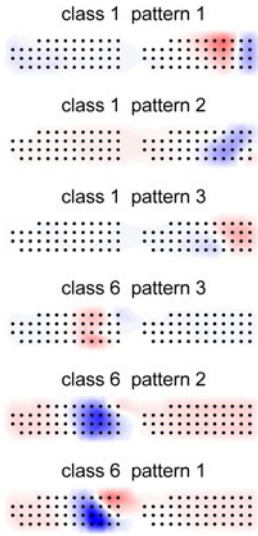

Fig. 10. Example of CSP patterns of the OvO extension for the class combination *hand open* and *wrist extension*.

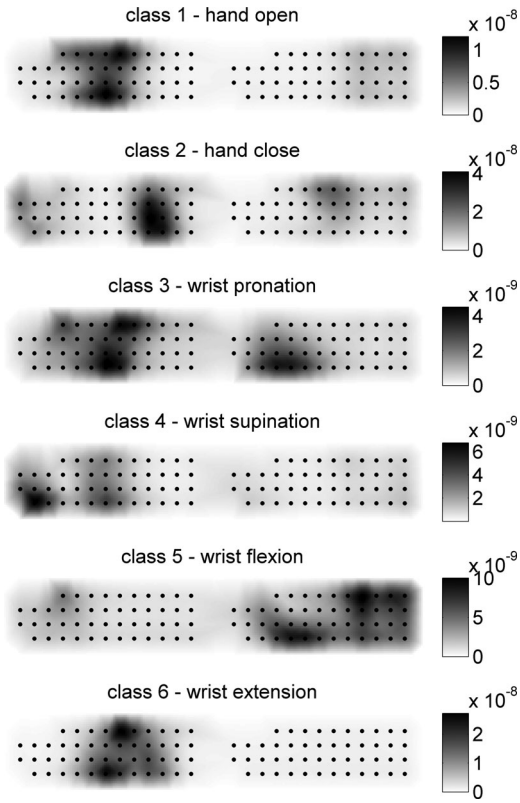

Fig. 11. Variance patterns of the raw, temporal filtered signals.

For *hand open* and *wrist flexion* the CSP patterns focus in the same regions as the variance patterns. For *pronation*, *supination* and *wrist extension* the CSP-patterns focus partly in the regions with high variance, but additional further discriminative regions are found that are not visible in the variance patterns. For *hand close* the best discriminative regions are found in areas completely different form those with highest signal power. This

shows that the regions with strongest muscle activity are not necessarily the regions with most discriminative information.

C. Patterns of CSP OvO

Due to the high number of combinations in the *OvO* extension it would go beyond the scope of this article to interpret all patterns. As an interesting example, we picked out a combination of two classes that corresponds to similar muscle activation, namely *hand open* and *wrist extension*. This is seen in the variance patterns (see Fig. 11 class one and six). The active regions are very similar for those classes, partly overlapping each other.

In Fig. 10, the first and last three patterns of the CSP filter that separate the chosen classes are shown. The lower three patterns correspond to the filters with high output for the class *wrist extension* and focus in a region with high EMG activity for that class. Careful comparison of the CSP patterns with the variance patterns shows that the focus of the CSP patterns is on the very right site of the region with high variance for that class. In this region, the other class *hand open* is less active, since its variance pattern is located a bit more to the left.

The upper three patterns correspond to the class *hand open*. They focus in a region on the second array, far away from the region of maximal variance. The region is rather in a location, where *hand open* has low, but still present activity, while the other class has no activity at all.

The fact that the region of best discriminative information is not necessarily in regions of high EMG-activity is consistent with our findings on heuristic channel selections [19]. The CSP patterns could become an equally useful tool for channel selection; however, this interesting aspect goes beyond the scope of this paper.

D. Discussion and Impact of the Results

The results of our analysis show the offline performance for static contractions, which cannot be put on a level with the performance in a real world application. But by including different arm positions and adding artificial noise, we included two important factors that can degrade the performance in a real world application. This does not render online-tests under out-of-lab conditions redundant, but it brings the offline analysis closer to the real-world scenario.

It was shown that with a sufficient high number of CSP-components the classification error can be almost halved compared to the baseline method. Under noisy conditions this significant advantage is even larger. Those results indicate that our proposed methods can significantly increase the reliability of a myoelectric control system.

The fact that the error was further reduced if the number of components exceeds the number of EMG channels might lead to the suspicion that the additional components do not contain additional information, but instead render a nonlinearly separable classification problem linear separable. However, tests with nonlinear classifiers like quadratic discriminant analysis and k-nearest-neighbors (with optimal choice of k) that are not shown here led to similar results as LDA.

In further steps online tests with appropriate performance metrics will be performed. Scheme and Englehart showed in [20] that results from healthy subjects can be generalized to persons with amputation. In their study with ten healthy and seven transradial amputated subjects the ranking was consistent across the two groups when different classifiers or different features were investigated. We expect a similar consistency for our methods, this, however, will need to be evaluated in a further dedicated study. A specific user training might additionally improve the reliability of the system. This will be of particular importance for amputated subjects that did not use the related muscles for many years.

V. Conclusion

We presented several multiclass extensions of the CSP algorithm and applied them to EMG signals in a myoelectric control task.

It was shown that the OvO and the OvR extension of CSP can significantly improve the classification performance in comparison with the commonly used bipolar derivations and the Hudgins feature set. The CSP methods are also much less affected by sensor noise, indicating that it could help to improve the reliability and robustness of a myoelectric control in a real world application.

The patterns of the discriminating CSP filters provide valuable information about the underlying physiological processes and the related muscles. In opposition to the variance patterns, which provide only information about the muscle activation, the CSP patterns emphasize the locations that provide most information to distinguish in between different contractions. The most discriminative areas can be in locations with rather low signal power.

In further steps, the method will be tested in online experiments and with amputees. Also the robustness against shifts of the electrodes, variations in force for proportional control, and the behavior in nonstatic contractions are subject of further investigations.

Acknowledgment

The authors would like to thank the members of the Berlin BCI and IDA group, who supported this study with many constructive discussions and good advice. Especially C. Sannelli for her assistance on the interpretation of the CSP patterns and S. Dähne for correcting this manuscript.

References

[1] P. Parker, K. Englehart, and B. Hudgins, "Myoelectric signal processing for control of powered limb prostheses," *J. Electromyography Kinesiol.*, vol. 16, no. 6, pp. 541–548, Dec. 2006.

[2] M. A. Oskoei and H. Hu, "Myoelectric control systems—A survey," *Biomed. Signal Process. Control*, vol. 2, no. 4, pp. 275–294, Oct. 2007.

[3] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 7, pp. 848–854, Jul. 2003.

[4] L. Hargrove, K. Englehart, and B. Hudgins, "A training strategy to reduce classification degradation due to electrode displacements in pattern recognition based myoelectric control," *Biomed. Signal Process. Control*, vol. 3, no. 2, pp. 175–180, Apr. 2008.

[5] A. Fougner, E. Scheme, A. Chan, K. Englehart, and O. Stavdahl, "Resolving the limb position effect in myoelectric pattern recognition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 9, no. 6, pp. 644–651, Dec. 2011.
[6] H. Huang, P. Zhou, G. Li, and T. Kuiken, "Spatial filtering improves EMG classification accuracy following targeted muscle reinnervation," *Ann. Biomed. Eng.*, vol. 37, no. 9, pp. 1849–1857, Sep. 2009.
[7] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.
[8] H. Gray, *Anatomy of the Human Body*. Philadelphia, PA: Lea & Febiger, 1918, ch. 4, pp. 1821–1865.
[9] G. Dornhege, J. del R. Millán, T. Hinterberger, D. J. McFarland, and K.-R. Müller, *Toward Brain-Computer Interfacing*. Cambridge, MA: MIT Press, 2007, pp. 207–233.
[10] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive berlin brain-computer interface," *NeuroImage*, vol. 37, no. 2, pp. 539–550, Aug. 2007.
[11] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.
[12] E. Scheme, K. Englehart, and B. Hudgins, "Selective classification for improved robustness of myoelectric control under non-ideal conditions," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 6, pp. 1698–1705, Jun. 2011.
[13] A. Ziehe, P. Laskov, G. Nolte, and K.-R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 777–800, Jul. 2004.
[14] L. H. Smith, L. J. Hargrove, B. A. Lock, and T. A. Kuiken, "Determining the optimal window length for pattern recognition-based myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 2, pp. 186–192, Apr. 2011.
[15] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K.-R. Müller, "Single-trial analysis and classification of ERP components—A tutorial," *NeuroImage*, vol. 56, no. 2, pp. 814–825, May 2011.
[16] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.
[17] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*, corrected ed. New York: Springer-Verlag, Jul. 2003, ch. 7, pp. 241–245.
[18] 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, May 2011.
[19] J. Hahne, M. Herrmann, D. Hofmann, M. Paulus, and B. Graimann, "Electrode configuration for multifunctional myoprostheses," presented at the 18th Congress of the International Society of Electrophysiology and Kinesiology, Aalborg, Denmark, Jun. 16–19, 2010.
[20] E. Scheme and K. Englehart, "Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use," *J. Rehabil. Res. Development*, vol. 48, no. 6, pp. 643–659, 2011.

Janne Mathias Hahne received the Diploma degree in electrical engineering from the Berlin Institute of Technology (TU-Berlin), Germany, in 2008, where since 2010, he has been working toward the Ph.D. degree in the Machine Learning Laboratory.

In 2009, he became member of the Department of Strategic Technology Management at Otto Bock HC, Duderstadt, Germany, where he was involved in control systems for myoelectric hand prostheses. His research interests include myoelectric control, biomedical signal processing, and machine learning.

Bernhard Graimann received the Ph.D. degree in biomedical engineering from the Graz University of Technology (TU Graz), Graz, Austria, in 2002.

He was a Postdoctoral Researcher at the Laboratory of Brain-Computer Interfaces, TU Graz, and at Institute of Automation, University of Bremen in biosignal processing, pattern recognition, and machine learning with applications in brain-computer communication and rehabilitation robotics. Since 2006, he has been a Lecturer at the Institute of Knowledge Discovery, TU Graz. In 2008, he became the Scientific Coordinator for Neurotechnology at Otto Bock HealthCare GmbH, Duderstadt, Germany. In this position, his main responsibility is the coordination of research projects in the field of neurotechnology with national and international academic and industrial partners.

Klaus-Robert Müller received Graduate and Master's degrees in physics during session 1984–1989, and the Ph.D. degree in computer science in 1992, all from the Karlsruhe Institute of Technology (TU Karlsruhe), Karlsruhe, Germany.

Since 2006, he has been a Professor of computer science at the Berlin Institute of Technology (TU Berlin), Berlin, Germany; at the same time he is the Director of the Bernstein Focus on Neurotechnology Berlin, Germany. From 1992 to 1994, he was a Postdoctoral Researcher at GMD FIRST (later Fraunhofer FIRST), Berlin, Germany, from 1994 to 1995, he was an European Community STP Research Fellow at the University of Tokyo, Tokyo, Japan. From 1995 to 2008, he built up the Intelligent Data Analysis Group at GMD FIRST and directed it. From 1999 to 2006, he was a Professor for computer science at the University of Potsdam. His research interests include intelligent data analysis, machine learning, statistical signal processing, and statistical learning theory with the application foci computational finance, computational chemistry, computational neuroscience, and genomic data analysis. Since 2000, one of his main scientific interests is to study the interface between brain and machine: noninvasive EEG-based brain computer interfacing.

Dr. Müller received the Olympus Prize by the German Pattern Recognition Society, DAGM, in 1999, and the SEL Alcatel Communication Award in 2006 .