20th IRANIAN STUDENT CONFERENCE ON ELECTRICAL ENGINEERING

Improving Classification of Multi-class Motor Imagery by Statistical Feature Selection

Mohammad Dehghan Manshadi
Master student, School of Automotive Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran
Mohammad_dehghan@auto.iust.ac.ir

Abdollah Amirkhani
Assistant professor, School of Automotive Engineering, Iran University of Science and Technology, Tehran 16846-13114, Iran
amirkhani@iust.ac.ir

Abstract
Brain-computer interface (BCI) is a novel technology that is assisting not only disabled people but also healthy people to control an external device by using motor imagery (MI). Although much work has been done in BCI system, achieving ideal accuracy has not been achieved due to the difficulty of pattern recognition of EEG signals. BCI systems are made up of various components that perform preprocessing, feature extraction, and decision making. Common spatial pattern (CSP) is an effective algorithm which is extensively used in extracting feature of EEG motor imagery task. In this article, the CSP algorithm has extended to multi-class classification by one-versus-one (OVO) and one-versus-rest (OVR) methods. To improve classifier in terms of accuracy and less complexity, Fisher algorithm has been used. The average accuracy 73.41 ± 1.62 has been achieved on BCI Competition IV-IIa dataset. The experimental results show that the Fisher algorithm in reducing complexity and increasing the accuracy of classifier has been effective.

Keywords: Brain computer interface, Common spatial pattern, Electroencephalography, Feature selection, Motor imagery task

Introduction

Brain-computer interface technology establishes a way of communication between the human brain and other external equipment so that one can control external equipment by analyzing brain signals. BCI divided into invasive and noninvasive category which in invasive one, signals are achieved by electrodes on the skin of brain and the surgery is needed but the other one, signals are achieved by electrodes on the surface of the scalp so there is no need to surgery (Baig et al, 2017).

So many people around the world lose their ability not only to move their body parts but also their ability to speak due to accidents, aging or congenital. Some people because of sickness like amyotrophic lateral sclerosis (ALS), locked-in syndrome, Lou Gehrig's disease and high spinal cord injury cannot control their muscle to do even simple routine things (Dong et al, 2017). Not only BCI can play a prominent role in applications like medical, robotic, game, etc. but also they can help people to control their environment without using their muscles.

BCIs based on motor imagery task because of simplicity, flexibility, noninvasive, good spatial accuracy and low equipment cost have been noticed a lot (Luo et al, 2020). In motor imagery task, asked people to imagine a task without any Physical or verbal expression then their brain signals achieved by a computer to recognize what imagined task is.

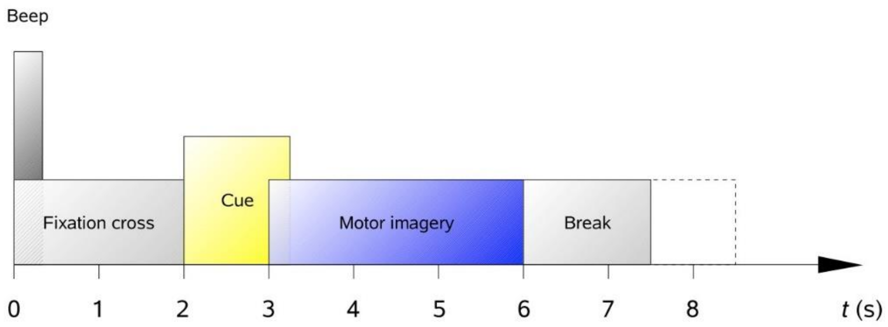
There are many challenges across the BCI based on motor imagery task which the most important one is the quality of EEG signal because they are contaminated easily by noise and different artifacts like muscle movement, blinking, heart rate, magnetic field of the environment, etc (He et al, 2018). Consequently, spatial and temporal filter are important to increase the signal-to-noise ratio of EEG signals (Blankertz et al, 2007). CSP has had good performances in extracting features from EEG based on motor imagery tasks (Dong et al, 2017 Belhadj et al, 2015 Ma et al, 2016 Zhang et al, 2015 Mishuhina and Jiang, 2018) and they are better in binary class motor imagery tasks. The performance of the CSP algorithm is affected by the number of train data and sensitive to noise so the regularized CSP (R-CSP) algorithm was introduced to overcome the problem of small training data set and noise (Lotte and Guan et al, 2010). Another issue in CSP is the choice of the optimal frequency band. The filter bank CSP (FBCSP) algorithm, which won the BCI IV Competition, was presented for this purpose. Many kinds of research have been done on the selection of the optimal frequency band.

Application of deep learning (DL) methods in BCI for classification of EEG recordings have confronted two main challenges: lack of large data set and privacy concern. Due to gathering large dataset are time-consuming and expensive, there are multiple small EEG-BCI dataset around the world. On the other hand, because there is potential abuse of EEG data which may lead to privacy violation, organizations should have explicit participant approval to exchange data. To tackle these problems, for example, in (Ju et al, 2020) has used federated learning and has reached state-of-art results.

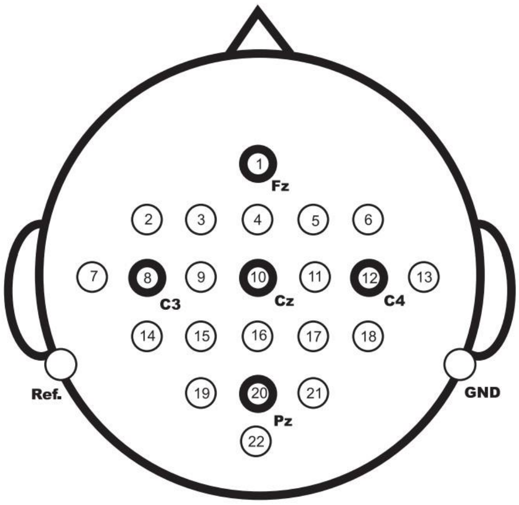
Section Dataset describes important database information. Section Methods details all stages of the experiments. Section Experimental Results presents the results of the classification and discusses the advantages and disadvantages of the proposed method. Finally, Section Conclusion draws conclusions.

Dataset

The database used in this article is related to BCI Competition IV-IIa, which includes 4 motor imagery tasks, left hand movement, right hand movement, foot movement and tongue movement. The number of channels or electrodes is 25 which 22 is related to EEG signals and the other 3 channels are related to EOG signal. The sampling rate is 250 HZ. The schedule of motor imagery tasks performed is shown in Fig. 1. EEG data were recorded in two sessions on different days to consider the nature of the instability of the EEG data. Each session consists of 6 run with a short break between each run. In each run, for each motor imagery task, 12 trials are recorded, in total, 48 trials are recorded, so in each session, 288 trials are recorded (72 trials for each mental activity). See (Brunner et al, 2008) for more information about the database.



(a)



(b)

Fig. 1. Database information ((a) motor imagery task schedule and (b) a view of the electrodes location [11])

Methods

Preprocessing

The neurophysiological basis for motor imagery tasks is that in each motor imagery task, the frequency band power of the EEG sensory-motor rhythms, i.e. Mu rhythm (8 HZ to 13 HZ) and Beta rhythm (18 HZ to 30 HZ) decreases or increases, which are called event-related desynchronization (ERD) / event-related synchronization (ERS) (Pfurtscheller et al, 1999 Duan et al, 2017). As a result, we extract the Mu and Beta rhythms, that is, the frequency between 8 HZ to 30 HZ, with the help of a 5th order Butterworth bandpass filter, so that only the band information related to motor imagery remains in the

EEG signal. According to (Lotte and Guan et al, 2010), we use the information of each trial from 0.5 to 2.5 s after the cue instructing the subject to perform motor imagery task as done by the winner of BCI Competition IV, data set IIa.

The 6-fold cross-validation method has been used to evaluate the results and prevent over-fitting. In 6 fold, the data is randomly divided into 6 fold (9 trials for each layer). The cross-validation process is repeated 6 times so that all 6 fold are tested exactly once.

Common Spatial Pattern

Spatial filters are an important part of detecting neural changes in the motor cortex, which the most successful way to differentiate such changes is by common spatial patterns (CSP) (Gonzalez et al, 2018). CSP designs subject-based spatial filters, which linearly combine different channels to maximize the variance of filtered signals in one class and minimize in another class (Jiang et al, 2020). This paper uses traditional CSP and uses one-versus-rest (OVR) and one-versus-one (OVO) strategies (Dong et al, 2017) to extend it to multi-classes.

Suppose $X_{c,i}$ $(i \in 1,2,\dots,n\ ,\ c \in 1,2,\dots,k)$ represents the i-th trail of class c (n is equal to the number of trials in each class and k is equal to the number of classes). Each trial $X \in R^{M\times N}$, depends on the number of channels and the sampling rate that M indicates the number of channels and N indicates the number of samples in each time period.

Traditional CSP is formulated as follows:

$$\mathrm{W} = \mathrm{argmax}_{\mathrm{W}} \frac{\mathrm{W}^{\mathrm{T}} \mathrm{C}_1 \mathrm{W}}{\mathrm{W}^{\mathrm{T}} \mathrm{C}_2 \mathrm{W}} \tag{1}$$

CSP look for spatial filters $W \in R^{M' \times N}$ that W are the spatial filters and C_k $(k = 1,2)$ is the mean of the normalized covariance matrix related to the k-th class calculated as follows:

$$C_k = \frac{1}{n} \sum_{i=1}^{n} \frac{X_{k,i} X_{k,i}^T}{trace(X_{k,i} X_{k,i}^T)} \tag{2}$$

n is the number of trial related to class k. The composite covariance matrix is defined as follows:

$$C_c = c_1 + c_2 \tag{3}$$

Its SVD is as follows:

$$C_c = U_c \Lambda_c U_c^T \tag{4}$$

U_c is an eigenvector matrix, each column forms an eigenvector. Eigenvectors, each corresponding to one of the eigenvalues in the diagonal matrix Λ_c (eigenvalue matrix). Also, $U_c$is a unit matrix. so $U_c U_c^T = I$. whitening transformation is equal to:

$$P = \Lambda_c^{-\frac{1}{2}} U_c^T \tag{5}$$

Because:

$$P C_c P^T = \sqrt{\Lambda_c^{-1}} U_c^T C_c U_c \sqrt{\Lambda_c^{-1}}^T$$

$$= \sqrt{\Lambda_c^{-1}} U_c^T [U_c \Lambda_c U_c^T] U_c \sqrt{\Lambda_c^{-1}}^T$$
$$= I$$

It means, this transformation makes the composite covariance matrix diagonal (identical matrix). By applying the whitening transformation to C_1 and C_2we have:

$$S_1 = PC_1P^T \quad (6)$$

$$S_2 = PC_2P^T \quad (7)$$

SVD of S_1 and S_2 :

$$S_1 = B\Lambda_1B^T \quad (8)$$

$$S_2 = B\Lambda_2B^T \quad (9)$$

That the sum of Λ_2 and Λ_1 is a unit matrix: $\Lambda_1 + \Lambda_2 = I$, which means that the largest eigenvalue in S_1 corresponds to the smallest eigenvalue in S_2. Eigenvectors B are used to classify the two classes. Spatial filters matrix:

$$W = B^TP \quad (10)$$

As a result, the trials are mapped as follows:

$$Y_i = WX_i \quad (11)$$

The reduced features are obtained by mapping the trial X_i to Y_i corresponding to m from the smallest and largest Λ_1 (or Λ_2). There is no definite evidence that performance improved with increasing m (Kang et al, 2009) but for example in (Ramoser et al, 2000 Müller-Gerking et al, 1999), it is suggested that $m = 2$ or $m = 3$ have good experimental performance. In this article, we have considered $m = 3$. In summary, the procedure for calculating spatial filters $W \in R^{M' \times N}$ is as follows:

1) Calculate $C_c\ in\ (3)\ and\ its\ SVD\ in\ (4)$

2) Calculate whitening matrix P in (5)

3) Calculate S_1 in (6) and its SVD in (8)

4) Arrange the elements of diameter $\Lambda_1$in descending order and arrange the columns of the matrix B accordingly and finally create the matrix B from the m first column and $m$the last column of the matrix B

5) Calculate $W = B^TP$

The feature vector f_i for the $X_i$trial is calculated as follows:

$$[f_i] = \log\left(\frac{var(y_i)}{\sum_{i=1}^{v} ar(y_i)}\right) \quad (12)$$

Finally, the combined feature vector resulting from the OVO and OVR strategies is as follows:

$$f = [f_{12}, f_{13}, f_{14}, f_{23}, f_{24}, f_{34}, f_1, f_2, f_3, f_4] \tag{13}$$

which $f_{12}, f_{13}, f_{14}, f_{23}, f_{24}, f_{34}$ are feature vectors corresponding to the strategy OVO and f_1, f_2, f_3, f_4 are the feature vectors corresponding to the OVR strategy.

Feature Selection

One of the most important reasons for feature selection is the peaking phenomenon, which is indicated in Fig. 2. The peaking phenomenon alone demonstrates the importance of using the right amount (neither too much nor too little) of features. Feature selection can reduce computational complexity (both in terms of execution load and storage).

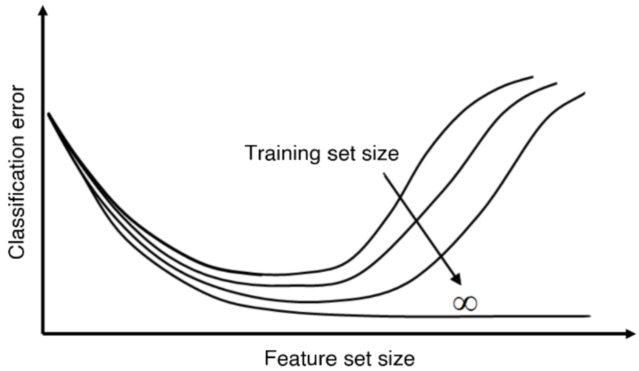


Fig. 2. Picking phenomenon as the number of features increases from one point onwards, the classification error increases.

Different types of feature selection methods can be divided into 3 general categories: filter, wrapper and hybrid methods, each with its own advantages and disadvantages. Filter methods do not use classifier during feature selection, which can be both an advantage and a disadvantage of these methods. The advantage of this approach is that the execution time of the algorithm is very low and in a very short time without using the model, features are scored then a number of features are selected depending on the needs. The disadvantage of this approach is that because the classifier don’t participate in feature selection process, the appropriate features for the classifier may not be selected. Unlike filter methods, wrapper methods use the classifier during feature selection so considering the relationship between features. These methods try to choose the best combination of available features. Wrapper methods are time consuming because they keep the model informed of feature selection, and may be costly when the number of features increases. Hybrid methods try to combine the two methods of scalar and vector, to perform the problem of feature selection faster than the wrapper method and more efficient.

The fisher discriminant ratio (FDR) method is one of the filter methods and based on the two parameters of mean and variance of a feature, determines its importance for classifying data of two or more classes. In this method, according to the relation:

$$F_{ratio} = \frac{variance\ of\ means\ between\ the\ classes}{average\ variance\ within\ the\ classes}$$

So:

$$F_{ratio} = \frac{\frac{1}{k}\sum_{j=1}^{k}\left(\mu_j - \bar{\mu}\right)^2}{\frac{1}{k}\sum_{j=1}^{k}\frac{1}{n_j}\sum_{i=1}^{n_j}\left(x_{ij} - \mu_j\right)^2} \qquad (14)$$

The value obtained determines the importance of the feature (k is the number of classes, μ_j is the average of class j-th, $\bar{\mu}$ is the average of the total data and n_j is the number of instances of the class j). For each feature, the value F_{ratio} is calculated and based on it features are ranked. After ranking the features, determining the number of properties is a hyperparameter. Fig. 3 shows the results of classifying the participant 1 based on the number of selected features, and by experimenting with other participants, we have considered the number of features to be 10. Our overall framework has been demonstrated in Fig. 4.

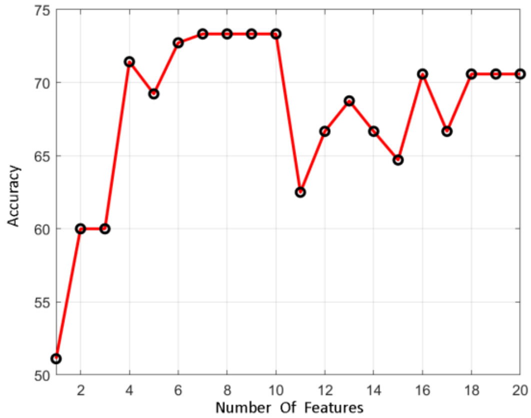


Fig. 3. Accuracy of classification based on the number of features selected in participant 1

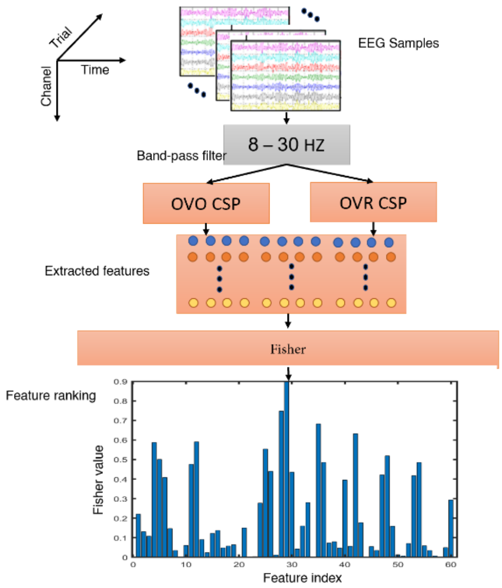


Fig. 4. Overall framework

Experimental Results

Experiments were performed using OVO-OVR CSP. In this article, we have used the linear discriminant analysis (LDA) as a classifier and to evaluate the results, we have used a 6-fold cross-validation with 50 run to get statistically meaningful results. Fig. 5 shows the results of the classification that the best accuracy for participant 9 is 94.58 ± 1.7 (mean ± standard deviation) and the overall accuracy among all participants is 73.41 ± 1.62. Two-way ANOVA was used to evaluate the accuracy obtained for 9 participants, which showed a significant difference ($F_{8,49} = 830.66$ $P = 1.7 \times 10^{-240}$). The results obtained in this article are compared with other articles which used the same method but without feature selection (see Table I). In (Dong et al, 2017), the number of features extracted is equal to 20, which we have reduced to 10, as well as the average total accuracy and accuracy obtained in 6 of the participants has increased. In (He et al, 2018), the EEG data were preprocessed by EEGLAB, although the total accuracy obtained are almost the same but the number of features extracted from 60 decreased to 10.

Table 1. A comparison between the results of this paper and other related work based on the accuracy index

Algorithm	Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Sub 8	Sub 9	Mean ± std
OVR-OVO CSP + HSVM (Dong et al, 2017)	68	72	82.1	45	40	38	76	78	74	64.4 ± 16.7
CSP + LDA (He et al, 2018)	82	61	96	63	53	60.5	68	95	88	74
Our work	75.87	65.69	86.99	56.38	59.56	66.33	66.98	88.34	94.58	73.41 ± 1.62

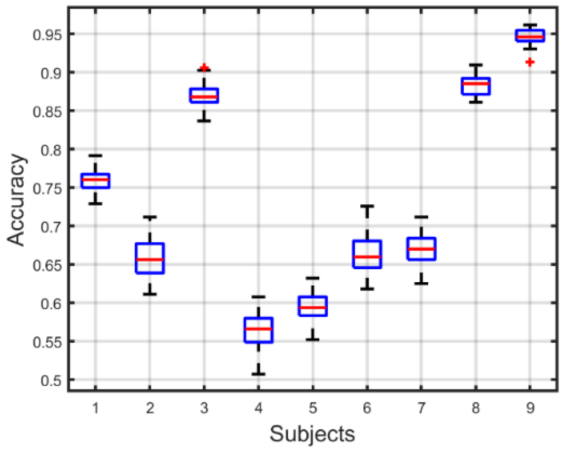


Fig. 5. The box plot of the accuracy obtained for each database participant

Conclusion

CSP is one of the most popular spatial filtering methods in brain-computer interfaces based on EEG data, especially in motor imagery task applications. CSP is used to increase the signal-to-noise ratio of EEG before giving to the classifier. In this paper described how to apply CSP to EEG data, extend it to multi-classes using the OVO and OVR methods, and reduce feature dimensions with the Fisher algorithm. The results of experiments on motor imagery data showed that the Fisher algorithm improved the model in terms of accuracy and reduced the complexity of the model.

References

M. Z. Baig, N. Aslam, H. P. H. Shum, and L. Zhang, "Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG," *Expert Syst. Appl.*, vol. 90, pp. 184–195, 2017.

E. Dong, C. Li, L. Li, S. Du, A. N. Belkacem, and C. Chen, "Classification of multi-class motor imagery with a novel hierarchical SVM algorithm for brain--computer interfaces," *Med. Biol. Eng. Comput.*, vol. 55, no. 10, pp. 1809–1818, 2017.

J. Luo, X. Gao, X. Zhu, B. Wang, N. Lu, and J. Wang, "Motor Imagery EEG Classification Based on Ensemble Support Vector Learning," *Comput. Methods Programs Biomed.*, p. 105464, 2020.

H. He and D. Wu, "Spatial filtering for brain computer interfaces: A comparison between the common spatial pattern and its variant," in *2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*, 2018, pp. 1–6.

B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, 2007.

S. A. Belhadj, N. Benmoussat, and M. Della Krachai, "CSP features extraction and FLDA classification of EEG-based motor imagery for Brain-Computer Interaction," in *2015 4th International Conference on Electrical Engineering (ICEE)*, 2015, pp. 1–6.

Y. Ma, X. Ding, Q. She, Z. Luo, T. Potter, and Y. Zhang, "Classification of motor imagery EEG signals with support vector machines and particle swarm optimization," *Comput. Math. Methods Med.*, vol. 2016, 2016.

Y. Zhang, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain--computer interface," *J. Neurosci. Methods*, vol. 255, pp. 85–91, 2015.

V. Mishuhina and X. Jiang, "Feature weighting and regularization of common spatial patterns in EEG-based motor imagery BCI," *IEEE Signal Process. Lett.*, vol. 25, no. 6, pp. 783–787, 2018.

Ju, Ce and Gao, Dashan and Mane, Ravikiran and Tan, Ben and Liu, Yang and Guan, C. (2020). Federated Transfer Learning for EEG Signal Classification. *ArXiv Preprint ArXiv:2004.12321*.

F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, 2010.

C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI Competition 2008--Graz data set A," *Inst. Knowl. Discov. (Laboratory Brain-Computer Interfaces), Graz Univ. Technol.*, vol. 16, 2008.

G. Pfurtscheller and F. H. L. Da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

L. Duan, Z. Hongxin, M. S. Khan, and M. Fang, "Recognition of motor imagery tasks for BCI using CSP and chaotic PSO twin SVM," *J. China Univ. Posts Telecommun.*, vol. 24, no. 3, pp. 83–90, 2017.

C. D. V. Gonzalez, J. H. S. Azuela, E. R. Espino, and V. H. P. Ponce, "Classification of motor imagery EEG signals with CSP filtering through neural networks models," in *Mexican International Conference on Artificial Intelligence*, 2018, pp. 123–135.

A. Jiang, J. Shang, X. Liu, Y. Tang, H. K. Kwan, and Y. Zhu, "Efficient CSP Algorithm With Spatio-Temporal Filtering for Motor Imagery Classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 1006–1016, 2020.

H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *IEEE Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, 2009.

H. Ramoser, J. Muller-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, 2000.

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.