

# Robust EEG Channel Selection across Subjects for Brain-Computer Interfaces

Michael Schröder,<sup>1</sup> Thomas Navin Lal,<sup>2</sup> Thilo Hinterberger,<sup>3</sup> Martin Bogdan,<sup>1</sup>  
N. Jeremy Hill,<sup>2</sup> Niels Birbaumer,<sup>3</sup> Wolfgang Rosenstiel,<sup>1</sup> and Bernhard Schölkopf<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, Eberhard-Karls University Tübingen, Sand 13, 72076 Tübingen, Germany  
Emails: schroedm@informatik.uni-tuebingen.de, bogdan@informatik.uni-tuebingen.de,  
rosenstiel@informatik.uni-tuebingen.de

<sup>2</sup>Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany  
Emails: navin@tuebingen.mpg.de, jez@tuebingen.mpg.de, bs@tuebingen.mpg.de

<sup>3</sup>Institute of Medical Psychology and Behavioral Neurobiology, Eberhard-Karls University Tübingen, Gartenstrasse 29,  
72074 Tübingen, Germany  
Emails: thilo.hinterberger@uni-tuebingen.de, niels.birbaumer@uni-tuebingen.de

Received 11 February 2004; Revised 22 September 2004

Most EEG-based brain-computer interface (BCI) paradigms come along with specific electrode positions, for example, for a visual-based BCI, electrode positions close to the primary visual cortex are used. For new BCI paradigms it is usually not known where task relevant activity can be measured from the scalp. For individual subjects, Lal et al. in 2004 showed that recording positions can be found without the use of prior knowledge about the paradigm used. However it remains unclear to what extent their method of *recursive channel elimination* (RCE) can be generalized across subjects. In this paper we transfer channel rankings from a group of subjects to a new subject. For motor imagery tasks the results are promising, although cross-subject channel selection does not quite achieve the performance of channel selection on data of single subjects. Although the RCE method was not provided with prior knowledge about the mental task, channels that are well known to be important (from a physiological point of view) were consistently selected whereas task-irrelevant channels were reliably disregarded.

**Keywords and phrases:** brain-computer interface, channel selection, feature selection, recursive channel elimination, support vector machine, electroencephalography.

## 1. INTRODUCTION

Brain-computer interface (BCI) systems are designed to distinguish two or more mental states during the performance of mental tasks (e.g., motor imagery tasks). Many BCI systems for humans try to classify those states on the basis of electroencephalographic (EEG) signals using machine learning algorithms.

The input for classification methods is a set of training examples. In the case of BCI one example might consist of EEG data (possibly containing several channels) of one trial and a label marking the class of the trial. Classification methods pursue the objective to find structure in the data and as a result provide a mapping from EEG data to mental states.

For some tasks the relevant EEG recording positions that lead to good classification results are known, especially when the tasks involve motor imagery (e.g., the imagination of limb movements) or the overall activity of large parts of the cortex (so-called slow cortical potentials, SCP) that occurs during intentions or states of preparation and relaxation.

For the development of new paradigms the neural correlates might not be known in detail and finding optimal recording positions for the use in BCIs is challenging. Such new paradigms can become necessary in cases when motor cortex areas show lesions, for the increase of the information rate of BCI systems, or for robust multiclass BCIs.

Algorithms for channel selection (CS) can identify suitable recording sites for individual subjects even in the absence of prior knowledge about the mental task. In this case it is possible to reduce the number of EEG electrodes necessary for the classification of brain signals without losing substantial classification performance.

In addition the CS results<sup>1</sup> can help to understand which part of the brain generates the class-relevant activity and even

---

<sup>1</sup>If an ordered list of channels is given by the CS algorithm that represents the importance of each channel for classification, this result is also called a *ranking*.

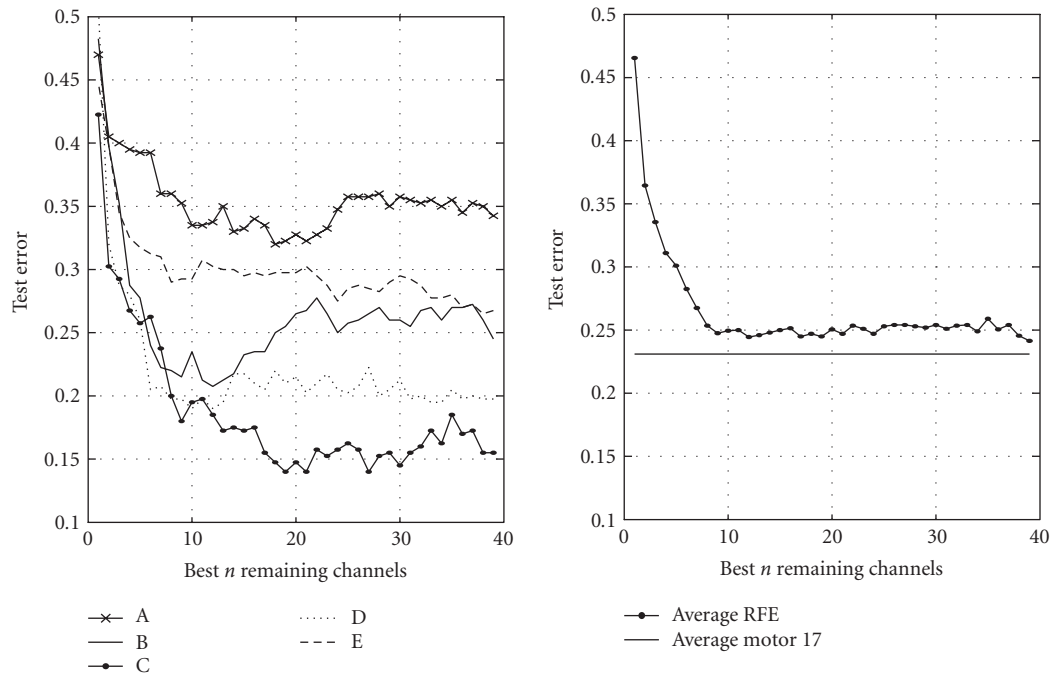


FIGURE 1: Test error of the channel selection method *RCE* for five subjects (A to E) on 39 EEG channels. The left graph shows the development of the test error against the best  $n$  remaining channels determined by *RCE*. For some subjects, the test error can be decreased by selecting fewer than 39 channels. The right graph shows test error of *RCE* averaged over the five subjects. On average, good performance can be obtained by less than 10 channels. The average test error for a set of 17 EEG channels over or close to motor cortex is added as a baseline for comparison.

simplifies the detection of artifact channels.<sup>2</sup> In [2], different channel selection algorithms have been compared for a motor imagery task. Figure 1 shows an example of the change in classification error that is observed applying the winning method recursive channel elimination (*RCE*) to the data of five individuals.

If data from several subjects are available, the questions arise, whether a set of channels selected for one subject is useful also for other subjects and whether generalized conclusions can be drawn about channels relevant for the classification of a certain mental task across subjects.

The paper is organized as follows. Section 2 contains the experimental setup, a description of the mental task, and the basic data preprocessing. In Section 3 the channel selection method and the classification algorithm are described. Results of cross-subject channel selection compared to average individual channel selection are given in Section 4 while the final section concludes.

## 2. DATA ACQUISITION

### 2.1. Experimental setup and mental task

We recorded EEG signals from eight untrained right-handed male subjects using 39 silver chloride electrodes

(see Figure 2). The reference electrodes were positioned at TP9 and TP10. The two electrodes Fp2 and 1 cm lateral of the right eye (EOG) were used to record possible EOG artifacts and eye blinks while two frontotemporal and two occipital electrodes were positioned to detect possible muscle activity during the experiment. Before sampling the data at 256 Hz an analog bandpass filter with cutoff frequencies 0.1 Hz and 40 Hz was applied.

The subjects were seated in an armchair at 1 m distance in front of a computer screen. Following the experimental setup of [3] the subjects were asked to imagine left versus right hand movements during each trial. With every subject, we recorded 400 trials during one single session. The total length of each trial was 9 seconds. Additional intertrial intervals for relaxation varied randomly between 2 and 4 seconds. No outlier detection was performed and no trials were removed during the data processing at any stage.

Each trial started with a blank screen. A small fixation cross was displayed in the center of the screen from second 2 to 9. A cue in the form of a small arrow pointing to the right or left side was visible for half a second starting with second 3. In order to avoid event-related signals in later processing stages only data from seconds 4 to 9 of each trial were considered for further analysis. Feedback was not provided at any time.

### 2.2. Preanalysis

As Pfurtscheller and da Silva have reported [4], movement-related desynchronization of the  $\mu$ -rhythm (8–12 Hz) is not

<sup>2</sup>Some subjects unintentionally use muscle activity that influences the recorded signals when trained in a BCI system, especially if feedback is provided.

equally strong in subjects and might even fail for some subjects due to various reasons (e.g., because of too short inter-trial intervals that prevent a proper resynchronization). Therefore we performed a pre-analysis in order to identify and exclude subjects that did not show significant  $\mu$ -activity at all.

For seven of the eight subjects, the  $\mu$ -band was only slightly different from the 8–12 Hz usually given in the EEG literature. Only one subject showed scarcely any activity in this frequency range but instead a recognizable movement-related desynchronization in the 16–20 Hz band.

Restricted to only the 17 EEG channels that were located over or close to the motor cortex, we calculated the maximum energy of the  $\mu$ -band using the Welch method [5] for each subject. This feature extraction resulted in one parameter per trial and channel and explicitly incorporated prior knowledge about the task.

The eight datasets consisting of the Welch-features were classified with linear SVMs (see below) including individual model selection for each subject. Generalization errors were estimated by 10-fold cross-validation. For three subjects the pre-analysis showed very poor error rates close to chance level, and their datasets were excluded from further analysis.

### 2.3. Data preprocessing

For the remaining five subjects the 5 s windows recorded from each trial resulted in a time series of 1280 sample points per channel. We fitted an autoregressive (AR) model of order 3 to the time series<sup>3</sup> of all 39 channels using forward-backward linear prediction [6]. The three resulting AR coefficients per channel and trial formed the new representation of the data.

The extraction of the features did not explicitly incorporate prior knowledge although autoregressive models have successfully been used for motor-related tasks (e.g., [3]). However, they are not directly linked to the  $\mu$ -rhythm.

Before AR, datasets from several subjects were combined for cross-subject channel selection, an additional centering and linear scaling of the data was performed. This was done individually for each subject and trial in order to maintain the proportion of corresponding AR coefficients in a trial.

### 2.4. Notation

Let  $n$  denote the number of training vectors (trials) of the datasets ( $n = 400$  for each of the five datasets) and let  $d$  denote the data dimension ( $d = 3 \cdot 39 = 117$  for all five datasets). The training data for a classifier is denoted as  $X = (x^{(1)}, \dots, x^{(n)}) \in \mathbb{R}^{n \times d}$  with labels  $Y = (y_1, \dots, y_n) \in \{-1, 1\}^n$ . For the task used in this paper  $y = -1$  denotes imagined left hand movement and  $y = 1$  denotes imagined

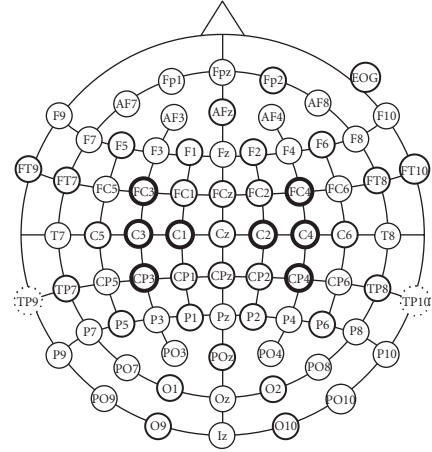


FIGURE 2: The positions of 39 EEG electrodes used for data acquisition are marked by black circles. The two referencing electrodes are marked by dotted circles. Eight electrodes over or close to the motor cortex are shown in bold circles (positions C1, C2, C3, C4, FC3, FC4, CP3, and CP4).

right hand movement. The terms *dimension* and *feature* are used synonymously.

## 3. CHANNEL SELECTION AND CLASSIFICATION METHODS

Channel selection algorithms as well as feature selection algorithms can be characterized as either filter or wrapper methods [7]. They select or omit dimensions of the data that correspond to one EEG channel depending on a performance measure.

The problem of how to rate the relevance of a channel if nonlinear interactions between channels are present is not trivial, especially since the overall accuracy might not be monotonic in the number of features used. Some methods try to overcome this problem by optimizing the selection for feature subsets of fixed sizes (plus-1 take-away-r search) or by implementing floating strategies (e.g., floating forward search) [7]. Only few algorithms like genetic algorithms can choose subgroups of arbitrary size during the selection process. They have successfully been used for the selection of spatial features [8] in BCI applications but are computationally demanding.

For the application of EEG channel selection, it is necessary to treat certain groups of features homogeneously: numerical values belonging to one and the same EEG channel have to be dealt with in a congeneric way so that a spatial interpretation of the solution becomes possible.

In [2] three state-of-the-art algorithms were compared for the problem of channel selection in BCI. As the method of *recursive channel elimination* (RCE), which is closely related to support vector machines (SVM), performed superior compared to other methods, we will use RCE for the cross-subject channel selection experiments described in this paper.

<sup>3</sup>For comparison reasons this choice of the model order is the same as in [2]. For this work different model orders had been compared in the following way. For a given order we fitted an AR-model to each EEG sequence. After proper model selection a support vector machine with 10-fold cross-validation (CV) was trained on the AR coefficients. Model order 3 resulted in the best mean CV error.

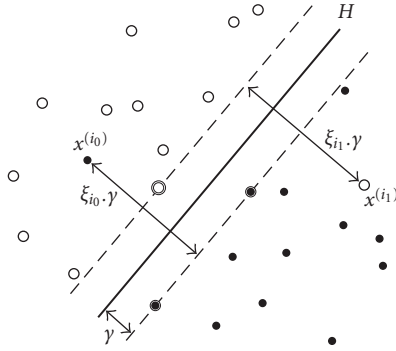


FIGURE 3: Linear SVM. For nonseparable datasets, slack variables  $\xi_i$  are introduced. The bold points on the dashed lines are called support vectors (SVs). The solution for the hyperplane  $H$  can be written in terms of the SVs. For more detail see Section 3.1.

### 3.1. Support vector machines

The support vector machine is a relatively new classification technique developed by Vapnik [9] which has shown to perform strongly in a number of real-world problems, including BCI [10]. The central idea is to separate data  $X \subset \mathbb{R}^d$  from two classes by finding a weight vector  $w \in \mathbb{R}^d$  and an offset  $b \in \mathbb{R}$  of a hyperplane

$$\begin{aligned} H : \mathbb{R}^d &\longrightarrow \{-1, 1\}, \\ x &\longmapsto \text{sign}(w \cdot x + b) \end{aligned} \quad (1)$$

with the largest possible margin,<sup>4</sup> which apart from being an intuitive idea has been shown to provide theoretical guarantees in terms of generalization ability [9]. One variant of the algorithm consists of solving the following optimization problem:

$$\begin{aligned} \min_{w \in \mathbb{R}^d} & \|w\|_2^2 + C \sum_{i=1}^n \xi_i^2 \\ \text{s.t. } & y_i(w \cdot x^{(i)} + b) \geq 1 - \xi_i \quad (i = 1, \dots, n). \end{aligned} \quad (2)$$

The parameters  $\xi_i$  are called slack variables and ensure that the problem has a solution in case the data are not linear separable<sup>5</sup> (see Figure 3). The margin is defined as  $\gamma(X, Y, C) = 1/\|w\|_2$ . In practice one has to trade off between a low training error, for example,  $\sum \xi_i^2$ , and a large margin  $\gamma$ . This trade off is controlled by the regularization parameter  $C$ . Finding a good value for  $C$  is part of the model selection procedure. If no prior knowledge is available  $C$  has to be estimated from the training data, for example, by using cross-validation. The value  $2/C$  is also referred to as the *ridge*. For a detailed discussion please refer to [11].

<sup>4</sup>If  $X$  is linear separable the margin of a hyperplane is proportional to the distance of the hyperplane to the closest point  $x \in X$ .

<sup>5</sup>If the data are linear separable the slack variables can improve the generalization ability of the solutions.

### 3.2. Recursive channel elimination

This channel selection method is derived from the recursive feature elimination method proposed by Guyon et al. [12]. It is based on the concept of margin maximization. The importance of a channel is determined by the influence it has on the margin of a trained SVM. Let  $W$  be the inverse of the margin

$$W(X, Y, C) := \frac{1}{\gamma(X, Y, C)} = \|w\|_2. \quad (3)$$

Let  $X^{-j}$  be the data with features  $j$  removed and  $Y^{-j}$  the corresponding labels. In the original version one SVM is trained during each iteration and the features  $\hat{j}$  which minimize  $|W(X, Y, C) - W(X^{-j}, Y^{-j}, C)|$  are removed (typically, i.e., one feature only); this is equivalent to removing the dimensions  $\hat{j}$  that correspond to the smallest  $|w_j|$ . For channel selection this method was adapted in the following way.

Let  $F_k \subset \{1, \dots, d\}$  denote the features from channel  $k$ . For each channel  $k$  we define the score  $s_k := (1/|F_k|) \sum_{l \in F_k} |w_l|$ . At each iteration we remove the channels with the lowest score. If no prior knowledge is available the parameter  $C$  has to be estimated from the training data.

### 3.3. Generalization error estimation

For model selection purposes we estimated the generalization error of classifiers via 10-fold cross-validation.

If the generalization error of a channel selection method had to be estimated, a somewhat more elaborated procedure was used. An illustration of this procedure is given in Figure 4.

The whole dataset is split up into 10 folds ( $F1$  to  $F10$ ) as for usual cross-validation. In each fold  $F$ , the channel selection (CS in Figure 4) is performed based on the training set of  $F$  only, leading to a specific ranking of the 39 EEG channels. For each fold  $F$ , 39 classifiers  $C_F^h$ ,  $h = 1, \dots, 39$ , are trained as follows:  $C_F^h$  is trained on the  $h$  best<sup>6</sup> channels, respectively, of the train set of  $F$  and tested on the corresponding channels of the test set of  $F$ . For each fold, this results in 39 test errors ( $E_F^1$  to  $E_F^{39}$ ).

During the last step, the corresponding test errors are averaged over all folds. This leads to an estimate of the generalization error for every number of selected channels.

## 4. EXPERIMENTS AND RESULTS

The successful transfer of EEG channel rankings of one subject to another can be difficult for several reasons.

- (i) The head shapes might vary between subjects. This limits the comparability of electrode positions and channel selection outcomes.
- (ii) Subjects might use different mental representations for a task, even if they are instructed carefully.

<sup>6</sup>In this context, *best* means according to the ranking calculated for that fold.

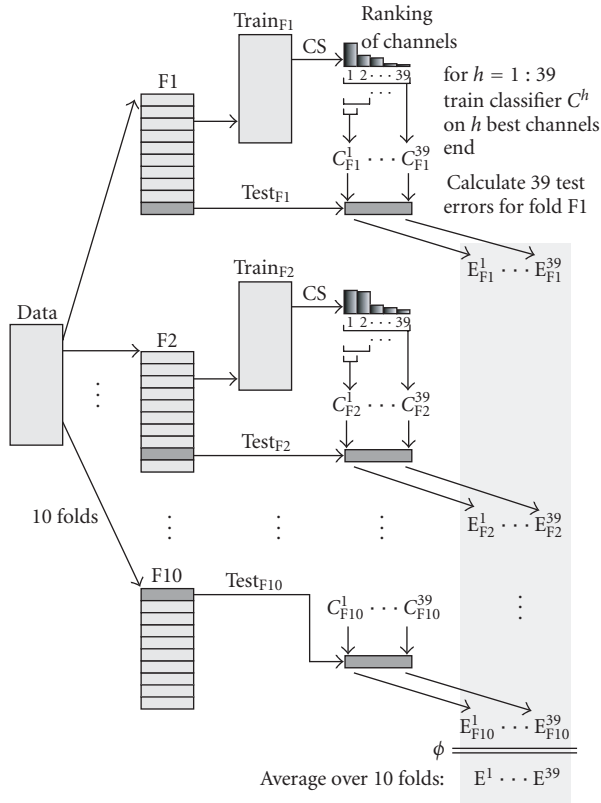


FIGURE 4: Illustration of the procedure for channel selection and error estimation using cross-validation.

- (iii) Cortex areas important for the mental task are probably organized slightly differently between subjects. This limits the comparability of localized activity patterns.

Luckily motor imagery tasks involve a comparably big part of the cortex. As a result small dislocations of EEG electrodes (e.g., around typical motor positions C3 and C4, see Section 2) usually do not lead to profound error increase for the classification of brain activity.

Nevertheless it is very important to investigate the reliability of cross-subject channel selection: on the one hand, even a slightly increased classification error leads to a large drop in the information rate for a BCI system [13]; on the other hand, mental tasks that do not show the advantages of motor imagery will more and more be focused on by BCI research in order to expand existing systems to multiclass BCIs or for increasing the information rate of patients whose motor areas are not intact.

The following subsections show results for the recursive channel elimination method on cross-subject data. In Section 4.1 RCE is applied to combined data of all five subjects. Results are compared with the individual channel rankings obtained from the five subjects. In Section 4.2 the transfer of rankings is investigated: RCE calculates rankings of data combined from 4 subjects before these rankings are tested on the corresponding remaining unseen dataset of the last subject.

#### 4.1. Channel selection on combined data

We applied the channel selection method of recursive channel elimination (RCE) introduced in Section 3 on a training dataset that was combined from the five AR datasets.

The estimation of the average generalization error for all 39 stages of the channel selection process with RCE was carried out using linear SVMs as classifiers with parameter  $C$  previously determined by 10-fold cross-validation.<sup>7</sup> Details about the 10-fold cross-validation process for channel selection are described in Section 3.3 and Figure 4. Figure 5 shows the development of the estimated classification error for all 39 steps of the RCE.

For this combined dataset the test error was minimal (26.9%) when using data from 32 or more EEG channels but further reduction down to 24 channels increased the test error only marginally. Reducing the number of channels to fewer than the best 17 channels leads to a strong increase of the test error.

Throughout the ranking in the table of Figure 5, artifact or task-irrelevant channels appear only in the last ranks (e.g., EOG, occipital channels, FT9, FT10, etc.). Direct comparison between Figures 1 and 5 reveal that the curve in Figure 1 shows smaller error rates. The performance of a classifier trained on the RCE channels of combined data is worse than the average performance of classifiers trained on the individual RCE channels of single subject data.

#### 4.2. Transfer of channel selection outcomes to new subjects

In this section we analyze whether there exists a general good subgroup of EEG channels (i.e., a subgroup of channels that perform well for all subjects) for a fixed mental task and whether this subgroup can be determined by the RCE method. We describe different methods to obtain channel rankings, some of which include the data of more than one subject. However these rankings are always tested on the data of one subject only. Table 1 provides an overview over all ranking modes.

##### Cross-subject modes

We iterate the following process. One subject is removed from the combined data base. We perform the RCE on the remaining data which leads to a channel ranking.

We use this ranking in two different ways to obtain test errors via 10-fold cross-validation on the data of the removed subject.

(i) *Best 8 (cross)*. The channel subset used for testing consists of the eight best-ranked channels. The resulting 8 best channels are plotted in Figure 6.

(ii) *Best  $n$  (cross)*. The channel subset used for testing consists of the  $n$  best-ranked channels. The number  $n$  is chosen such that the expected cross-validation error on the four

<sup>7</sup>Estimating the parameter for each number of channels in the process of channel selection might improve the accuracy but was not performed.

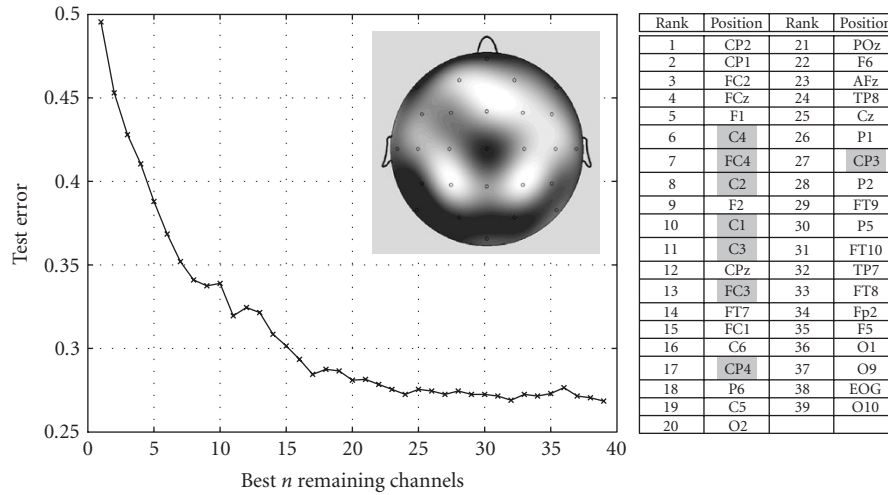


FIGURE 5: RCE results for a combined dataset of all 5 subjects. The graph shows a test error estimation for the  $n$  best channels. The error values were estimated by 10-fold crossvalidation. The table on the right shows the channel ranking performed on the combined data. Eight channels which are located over or close to the motor cortex (see Figure 2) are printed with grey background. The surface map visualizes this ranking. The 24 best-ranked electrodes were mapped to grey scale values. Bright areas of the surface map correspond to relevant channels (according to RCE) whereas dark areas show less-relevant electrodes.

TABLE 1: Ranking modes overview: explanation of the ranking modes used for the comparison shown in Figure 7. The rankings were calculated on different kinds of datasets: on data from single subjects or (for cross-subject tests) on combined datasets (4-fold cross-validation). Testing of the ranking modes was always performed on the data of one single subject.

Mode	Ranking method	Ranking based on	Description
<i>Motor 8</i>	A priori knowledge	Single subject	8 channels over or close to motor cortex
<i>Random 8</i>	(Random)	Single subject	8 channels
<i>Best n (single)</i>	RCE	Single subject	$n$ channels with highest rank that minimize CV error
<i>Best 8 (single)</i>	RCE	Single subject	8 channels with highest rank
<i>Best n (cross)</i>	RCE	Four subjects	$n$ channels with highest rank that minimize CV error
<i>Best 8 (cross)</i>	RCE	Four subjects	8 channels with highest rank

subjects is minimized. Note that this choice does not depend on the data of the fifth test subject.

As this process is repeated for every subject that was left out, we can average the error values of the modes *Best 8 (cross)* and *Best n (cross)* over five repetitions.

#### For comparison: single-subject modes

For the fixed mental task of motor activity and imagery, the EEG literature suggests the channels CP3, CP4, and adjacent electrodes (e.g., [3]). Our guess at generally good subgroup of EEG channels is thus the electrode set: FC3, FC4, C1, C2, C3, C4, CP3, CP4 (see electrodes marked in boldface in Figure 2). The corresponding test mode is referred to as *Motor 8*.

If no prior knowledge of a task and no channel selection were available, a random choice of channels would be the single solution. For comparison reasons we include the mode *Random 8*. Its test error is the average of ten repetitions of choosing eight random channels, optimizing the regularization parameter  $C$  and testing this random subset via 10-fold cross-validation on the data of one subject.

For the two modes *Best 8 (single)* and *Best n (single)* the RCE method was applied to the individual data of single subjects only. These modes used subgroups of the eight best channels and  $n$  best channels (see above) for calculating the test error via 10-fold cross-validation. It can be expected that the ranking for data from single subjects leads to more accurate classification results and can reveal task-related artifact

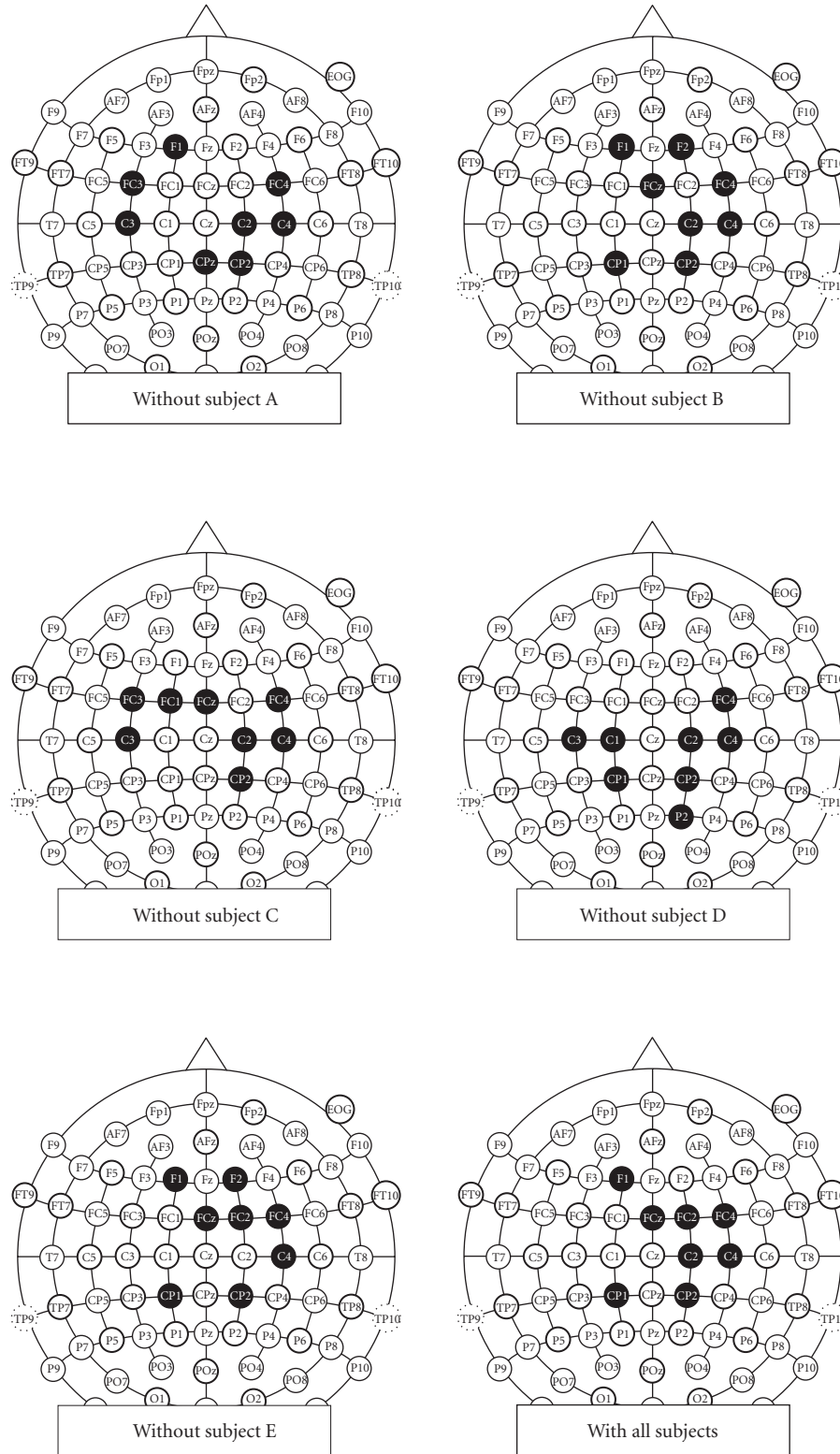


FIGURE 6: The database consists of data from 5 subjects. The channels were ranked 5 times using the channel selection method recursive channel elimination (RCE), each time using the data of four subjects only. The electrode positions marked in bold are the 8 best-ranked ones and are consistently located over or close to the motor cortex although the method was not provided with prior knowledge about the motor imagery task. This type of ranking is referred to as *Best 8 (cross)*.

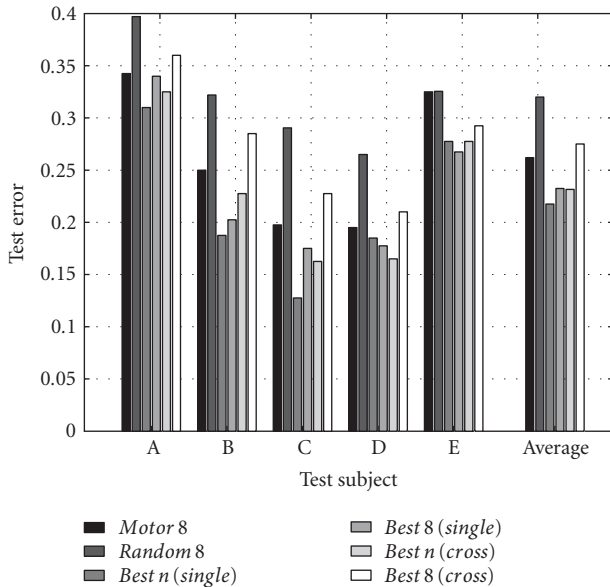


FIGURE 7: Comparison of the test errors of six different ranking modes for single subjects (A to E) and the test errors of these modes averaged over the five subjects (Average). For each mode and subject, the regularization parameter  $C$  was estimated separately. All test errors were obtained using 10-fold CV. The first mode *Motor 8* tests the classification error for 8 channels over or close to the motor cortex, whereas *Random 8* is based on 8 randomly chosen channels. Modes *Best n (single)* and *Best 8 (single)* test channel sets whose rankings were calculated based on the specific subject only. Modes *Best n (cross)* and *Best 8 (cross)* test channel sets whose rankings were calculated based on all other subject's data but did not incorporate data from the test subject.

channels [2] that might not be present in data from other subjects.

Figure 7 shows the results for the 6 modes. The right-most block contains an average taken over subjects for each of the modes. From the average results we observe the following.

(i) The 8 motor channels are not optimal: *Best 8 (single)* performs much better.<sup>8</sup>

(ii) Mode *Best 8 (cross)* performs almost as well as the motor channel mode. Although we conclude that the RCE method fails to find an optimal channel subset, the results suggest that when transferring channel positions across subjects the expected performance is not much worse than the one using prior knowledge.

(iii) The subset of 8 random channels performs surprisingly well. This finding suggests that the structure of the data can successfully be captured by the SVM even if only few channels close to the motor cortex are contained in the channel subset. However all other modes show better error estimations.

<sup>8</sup>In Figure 1 the choice of motor channels results in a lower classification error than the error from the RCE method. This is due to the fact that the regularization parameter  $C$  or ridge was not optimized for a specific ranking as was done in this study.

(iv) The performance of *Best n (cross)* mode is comparable to the results of the *Best 8 (single)* mode (23%); nevertheless this comparison is unfair since on average 27 channels were used. The cross-validation averaged over the five subjects is 26% for the choice of 27 random channels (not plotted in Figure 7).

(v) The best performing mode is *Best n (single)*. On average it only uses  $n = 14$  channels and yields an error as low as 21.8%.

## 5. CONCLUSION

The recursive channel elimination (RCE) method was applied to EEG channel selection in the context of signal classification for a Brain-Computer interface (BCI) system.

All experiments were based on data from five subjects recorded during a motor imagery task comprising imagined left and right hand movement.

For individual subject we analyzed the performance of three different types of rankings: (i) ranking including channels over the motor cortex only, (ii) ranking obtained by RCE from the data of that subject, (iii) ranking obtained by RCE from the data of the other four subjects.

We obtained best results with RCE rankings from single subjects. A comparison reveals that they outperform motor rankings (including prior knowledge about the task) by about 5% absolute error.

The transfer of RCE rankings from the data of multiple subjects to a new subject leads to a small decrease in performance. The difference to the performance of motor rankings turns out to be less than 2% on average.

We conclude that individual channel ranking is preferable over cross-subject ranking for the experimental paradigm investigated here.

However for the first time, it could be shown that RCE cannot only successfully be used to select channels for *individual* subjects, but that RCE rankings on the combined data of *multiple* subjects are consistently in agreement with the EEG literature on motor imagery tasks, and can still yield error rates as low as 17% on unseen subjects.

## ACKNOWLEDGMENTS

The authors would like to thank Bernd Battes and Professor Dr. Kuno Kirschfeld for their help with the EEG recordings. Special thanks to Dr. Jason Weston for his help on feature selection topics. This work have been supported in part by DFG (AUMEX RO 1030/12), NIH, and the IST Programme of the European Community, under the PASCAL Network of Excellence, IST-2002-506778. Thomas Navin Lal was supported by a grant from the Studienstiftung des deutschen Volkes.

## REFERENCES

- [1] E. E. Sutter, "The brain response interface: communication through visually-induced electrical brain responses," *Journal of Microcomputer Applications*, vol. 15, no. 1, pp. 31–45, 1992.



- [2] T. N. Lal, M. Schröder, T. Hinterberger, et al., "Support vector channel selection in BCI," *IEEE Trans. Biomed. Engineering*, vol. 51, no. 6, pp. 1003–1010, 2004.
- [3] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehab. Eng.*, vol. 6, no. 3, pp. 316–325, 1998.
- [4] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.
- [5] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, 1967.
- [6] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, USA, 1996.
- [7] P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler, "Floating search methods for feature selection with nonmonotonic criterion functions," in *Proc. 12th International Conference on Pattern Recognition (ICPR '94)*, vol. 2, pp. 279–283, Jerusalem, Israel, October 1994.
- [8] M. Schröder, M. Bogdan, W. Rosenstiel, T. Hinterberger, and N. Birbaumer, "Automated EEG feature selection for brain computer interfaces," in *Proc. 1st International IEEE EMBS Conference on Neural Engineering*, pp. 626–629, Capri, Italy, March 2003.
- [9] V. N. Vapnik, *Statistical Learning Theory*, John Wiley & Sons, New York, NY, USA, 1998.
- [10] B. Blankertz, G. Curio, and K. Müller, "Classifying single trial EEG: towards brain computer interfacing," in *Advances in Neural Information Processing Systems*, T. K. Leen, T. G. Dietterich, and V. Tresp, Eds., vol. 14, MIT Press, Cambridge, Mass, USA, 2001.
- [11] B. Schölkopf and A. Smola, *Learning with Kernels*, MIT Press, Cambridge, Mass, USA, 2002.
- [12] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine Learning*, vol. 46, no. 1-3, pp. 389–422, 2002.
- [13] A. Schlögl, C. Keinrath, R. Scherer, and G. Pfurtscheller, "Information transfer of an EEG-based brain-computer interface," in *Proc. 1st International IEEE EMBS Conference on Neural Engineering*, pp. 641–644, Capri, Italy, March 2003.

**Michael Schröder** received his Diploma in computer science in 2000. Currently he is a Ph.D. student at the Department for Computer Engineering (Professor Rosenstiel) at the Eberhard-Karls-Universität Tübingen in Germany. His research interests include machine learning, brain-computer interface systems, and signal processing.



**Thomas Navin Lal** received his Diploma in mathematics in 2001 and spent one year with the machine learning group of Professor Dr. Thomas Hofmann at Brown University, Providence, RI. He is currently a Ph.D. student of Professor Dr. Bernhard Schölkopf at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany. He is a researcher at the PASCAL network of excellence and is currently supported by a grant from the Studienstiftung des deutschen Volkes.



**Thilo Hinterberger** received his Diploma in physics from the University of Ulm, Germany, and received his Ph.D. degree in physics from the University of Tübingen, Germany, in 1999, on the development of a brain-computer interface, called "Thought Translation Device." He is currently a Research Associate with the Institute of Medical Psychology and Behavioral Neurobiology at the University of Tübingen, Germany. His primary research interests focus on the further development of brain-computer interfaces and their applications and also on the development of EEG classification methods and the investigation of neuropsychological mechanisms during the operation of a BCI using functional MRI. He is a Member of the Society of Psychophysiological Research and the Deutsche Physikalische Gesellschaft (DPG).



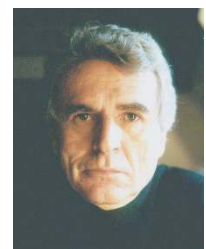
**Martin Bogdan** received the Engineer Diploma in signal engineering from the Fachhochschule Offenburg, Germany, in 1993, and the Engineer Diploma in industrial informatics and instrumentation from the Université Joseph Fourier Grenoble, France, in 1993. In 1998, he received the Ph.D. degree in computer science (computer engineering) from the University of Tübingen, Germany. In 1994, he joined the Department of Computer Engineering at the University of Tübingen, where, since 2000, he has headed the research group NeuroTeam. This research group deals mainly with signal processing based on artificial neural nets and machine learning focused on, but not limited to, biomedical applications.



**N. Jeremy Hill** graduated in experimental psychology at the University of Oxford, UK, in 1995. Until 2001 he was a Research Assistant, Programmer, and finally a doctoral student in the psychophysics laboratory of Dr. Bruce Henning in Oxford. He received the Ph.D. degree in 2002, for a doctoral thesis on psychophysical statistics entitled "Testing hypotheses about psychometric functions." Since then he has been part of Professor Bernhard Schölkopf's Department for Empirical Inference for Machine Learning and Perception at the Max Planck Institute for Biological Cybernetics in Tübingen, Germany, and now he focuses on brain-computer interface research.



**Niels Birbaumer** was born in 1945. He received his Ph.D. degree in 1969, in biological psychology, art history, and statistics, from the University of Vienna, Austria. From 1975 to 1993, he was a Full Professor of clinical and physiological psychology, University of Tübingen, Germany. From 1986 to 1988, he was a Full Professor of psychology, Pennsylvania State University, USA. Since 1993, he has been a Professor of medical psychology and behavioral neurobiology at the Faculty of Medicine, the University of Tübingen, and Professor of clinical psychophysiology, University of Padova, Italy. Since 2002, he has been the Director of the Center of Cognitive Neuroscience, University of Trento, Italy.



**Wolfgang Rosenstiel** is Professor at the University of Tübingen and is the Chair of Computer Engineering. He is also the Managing Director of the Wilhelm Schickard Institute at Tübingen University, and the Director of the Department for System Design in Microelectronics at the Computer Science Research Centre (FZI). He is on the Executive Board of the German Edacentrum. His research areas include artificial neural networks, signal processing, embedded systems, and computer architecture.



**Bernhard Schölkopf** received an M.S. degree in mathematics (University of London, 1992) and a Diploma in physics (Eberhard-Karls-Universität Tübingen, 1994), and a Ph.D. degree in computer science (Technical University Berlin, 1997). He won the Lionel Cooper Memorial Prize of the University of London, the Annual Dissertation Prize of the German Association for Computer Science (GI), and the Prize for the Best Scientific Project at the German National Research Center for Computer Science (GMD). He has researched at AT&T Bell Labs, at GMD FIRST, Berlin, at the Australian National University, Canberra, and at Microsoft Research Cambridge, UK. He has taught at the Humboldt University and the Technical University Berlin. In July 2001, he was elected Scientific Member of the Max Planck Society and Director at the MPI for Biological Cybernetics. In October 2002, he was appointed Honorary Professor for Machine Learning at the Technical University Berlin.

