A Brain Computer Interface with Online Feedback based on Magnetoencephalography

Thomas Navin Lal[1]   NAVIN@TUEBINGEN.MPG.DE
Michael Schröder[2]   SCHROEDM@INFORMATIK.UNI-TUEBINGEN.DE
N. Jeremy Hill[1]   JEZ@TUEBINGEN.MPG.DE
Hubert Preissl[3,4]   HUBERT.PREISSL@UNI-TUEBINGEN.DE
Thilo Hinterberger[3]   THILO.HINTERBERGER@UNI-TUEBINGEN.DE
Jürgen Mellinger[3]   JUERGEN.MELLINGER@UNI-TUEBINGEN.DE
Martin Bogdan[2]   BOGDAN@INFORMATIK.UNI-TUEBINGEN.DE
Wolfgang Rosenstiel[2]   ROSENSTIEL@INFORMATIK.UNI-TUEBINGEN.DE
Thomas Hofmann[5]   THOMAS.HOFMANN@IPSI.FRAUNHOFER.DE
Niels Birbaumer[3,6]   NIELS.BIRBAUMER@UNI-TUEBINGEN.DE
Bernhard Schölkopf[1]   BS@TUEBINGEN.MPG.DE

[1] Max-Planck-Institute for Biological Cybernetics, Dept. of Empirical Inference, Tübingen, Germany

[2] Eberhard Karls University, Dept. of Computer Engineering, Tübingen, Germany

[3] Eberhard Karls University, Dept. of Medical Psychology and Behavioral Neurobiology, Tübingen, Germany

[4] University of Arkansas for Medical Sciences, Dept. of Ob/Gyn, Little Rock, AR, USA

[5] Technical University of Darmstadt, Dept. of Intelligent Systems, Darmstadt, Germany

[6] Center for Cognitive Neuroscience, Faculty of Physics, University of Trento, Italy

Abstract

The aim of this paper is to show that machine learning techniques can be used to derive a classifying function for human brain signal data measured by magnetoencephalography (MEG), for the use in a brain computer interface (BCI). This is especially helpful for evaluating quickly whether a BCI approach based on electroencephalography, on which training may be slower due to lower signal-to-noise ratio, is likely to succeed. We apply RCE and regularized SVMs to the experimental data of ten healthy subjects performing a motor imagery task. Four subjects were able to use a trained classifier to write a short name. Further analysis gives evidence that the proposed imagination task is suboptimal for the possible extension to a multiclass interface. To the best of our knowledge this paper is the first working online MEG-based BCI and is therefore a "proof of concept".

1. Introduction

The goal of research into brain-computer interfaces (BCIs) is to build communication and control systems that a person can use to interact with the environment without the need for muscular or peripheral neural activity. The principal application of a BCI is as a form of neural prosthesis for people suffering from severe paralyzing conditions which can be caused by, for example, Amyotrophic Lateral Sclerosis (ALS).

Currently the most successful approaches to patient BCI are still those in which the user rather than the computer has to do most of the learning, and this usually takes many weeks or months in practice. If machine learning (ML) algorithms can be integrated effectively into such systems, they offer the promise of faster identification of the approach that is most suitable for a given patient. Thus ML has the potential to reduce the training time for a patient from several weeks or months to a few days or hours.

Most BCIs using ML techniques require a data collection phase during which the subject repeatedly executes a training task which is reflected by brain signals at clearly separable locations. Algorithms like the Support Vector Machine (SVM) or the Fisher Discrim-

inant can be applied to derive a classifying function from the collected training data. This function can be used in online applications to identify different brain states produced by the subject.

The majority of BCIs are based on extracranial EEG recordings during motor imagery. We restrict ourselves to mentioning just a few publications that use imagined limb movements (Pfurtscheller. et al., 1998; Ramoser et al., 2000; Wolpaw & McFarland, 1994; Schröder et al., 2005).

Besides EEG other recording techniques have been used for human BCIs. These BCIs take advantage of, for example, an increased signal-to-noise ratio as in electrocorticography (ECoG) (Leuthardt et al., 2004; Graimann et al., 2004; Lal et al., 2005) or the high spatial resolution of functional magnetic resonance imaging (fMRI) (Mitchell et al., 2003; Weiskopf et al., 2004). Furthermore extracellular invasive recordings have been used in patients by the team of Philip Kennedy in 1996 and recently by John Donoghue's lab.

There exist only very few EEG based approaches that work for completely paralyzed patients. One major factor is the limited data quality of the EEG which affects the learning ability of both the computer software as well as the patient. Magnetoencephalography (MEG) shows a much better signal quality than EEG and therefore promises increased learning effects. Since the brain signals exploited by MEG and EEG are fundamentally the same, however, the user should find it relatively easy to transfer from an MEG-BCI to a more portable EEG based system at a later training stage. For this to work, a ML technique has to be established that is applicable for both types of data. A promising candidate is the EEG-based BCI described in (Lal et al., 2004) and (Lal et al., 2005) since it shows convincing results. Furthermore, it is able, due to its built-in feature selection procedure, to work with a low number of recording channels which is important for realtime scenarios (see Section 8).

In the present paper we transfer their approach to the MEG domain, prove the feasibility of single-trial MEG signal classification and evaluate its accuracy in a BCI application. To the best of our knowledge this is the first time that the MEG recording technique has been used for feedback BCI systems.

The paper is structured as follows: after a short introduction to the MEG-technology we present the experimental setup and data preprocessing. Section 5 introduces the feature selection and classification methods applied to the recorded data. Offline-results are reported in Section 6. Section 7 discusses a possible multi-class extension. How subjects used an online BCI to write a short name is described in Section 8.
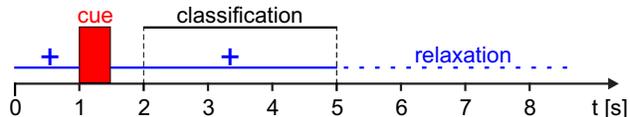

Figure 1. Overview of the trial structure during the data collection phase. The time interval used for classifier training started 0.5 seconds after the cue had ended. Relaxation intervals of randomized duration separated the trials. During the intervals marked with a "+", the fixation cross was visible.

2. Magnetoencephalography (MEG)

Magnetoencephalography (MEG) is a non-invasive measurement of the magnetic fields caused by electrical current dipoles that are generated by neural activity. Due to the orientation of magnetic sensors (coils) and the folding of the cortex surface, MEG primarily is sensitive for currents of tangential orientation generated in sulci, while EEG signals are based on both, tangential currents in the sulci and radial currents in the gyri.

Magnetic fields suffer far less than electric fields from the spatial blurring effect of the skull and intracerebral fluid. As a result MEG signals show a higher signal-to-noise ratio and are much more localized.

An important characteristic of brain signals is the so called mu-rhythm (μ-rhythm), which is found at frequencies of approximately 8-12Hz and 18-22Hz above the sensorimotor cortex. It was shown that measurements of this rhythm are similar for recordings of the human EEG and MEG (Tiihoonen et al., 1989). The intensity of the mu-rhythm is closely related to imagined as well as executed movements. Spectral properties of MEG recordings during executed thumb movements were studied by Salmelin and Hari (1994) and by Georgopoulos et al. (2004). The authors report that it is possible to reconstruct trajectories drawn by subjects using a joystick from the MEG measurements recorded in parallel to the movement execution. First work on the properties of mu-rhythm during *attempted* finger movements in tetraplegic persons was presented by (Kauhanen et al., 2004).

3. Experimental Setup

Ten healthy subjects participated in the experiment. Their MEG signals were recorded at 625Hz sampling rate from 150 channels located over the scalp. The subjects were seated relaxed in front of a projection screen. The subjects' heads were fixed to avoid movements. During the data collection phase of the exper-
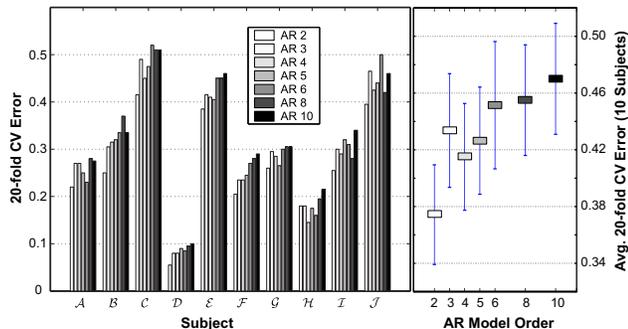
Figure 2. The left plot shows the 20-fold cross validation error of seven different AR model orders for the ten subjects. The estimates were based on data from all 150 channels. The right plot contains the averaged errors of the AR model orders along with standard errors.

Algorithm 1 Error Estimation using a double CV-scheme

Require: preprocessed MEG data of one subject
1: **for** (cntMainFolds = 1 to 50) **do**
2:     split data randomly: 80% training set, 20% test set
3:     with training set do:
4:         20-fold CV: find *good* ridge r
5:         rank channels using r
6:         20-fold CV: estimate the number b of good channels using r (Fig. 3 and 4)
7:         reduce data to best b channels
8:         20-fold CV: find *good* ridge r_red on reduced data
9:         train SVM S on reduced data using ridge r_red
10:     reduce test set to best b channels
11:     test S on the reduced test set
12:     save error and number of good channels b
13: **end for**
Output: mean error + variance, average number of good channels

iment the subjects were instructed to imagine movements of their tongue or left little finger. The choice of these two imaginations was motivated by the relatively great distance of the respective cortical areas on the motor cortex.

A trial began with a small fixation cross displayed at the center of the screen. One second later the randomly chosen task cue (either an image of a tongue or of a left little finger) was displayed for half a second (see Figure 1). The fixation cross appeared again at second 1.5 marking the classification interval and disappeared at second 5.0 marking the relaxation interval of two to four seconds duration (randomized). The subjects were asked to imagine movements during the classification interval (depicted in Figure 1). Each subject performed four blocks, each containing fifty trials with randomly selected cues. After each block the subjects could take a pause of approximately 5min duration to relax.

4. Data Preprocessing

From every trial we extracted second 2 to second 5. We thus obtained 1875 samples for each of the 150 MEG channels. To remove linear trends the least-square linear approximation was determined and subtracted from the original time series[1] before a forward-backward autoregressive model of order 2 was fitted to every detrended signal. Autoregressive (AR) models (Haykin, 1996) are able to describe detrended time

___
[1]If movement imagery paradigms are used, linear trends are unlikely to be useful for classification. However, for other BCI paradigms (e.g. those using *slow cortical potentials*) they may contain class-specific information.

series in a condensed form and capture their spectral characteristics.

The choice of this model order was based on an analysis of model orders 2 to 10. For each model order, the classification error on the data from all 150 channels was estimated with 20-fold cross validation. Figure 2 shows the errors estimated for the ten subjects and for different model orders. For nine subjects model order 2 yielded the lowest error. The data of one subject resulted in minimal errors using model order 4. Model order 2 was used for all subjects to represent uniformly the data during further processing steps.

To represent one trial, a vector of length $150*2$ was composed that contained the concatenated AR coefficients of all channels. The label corresponding to such a vector was defined to be -1 if the imagination task was left little finger movement and $+1$ if it was an imagined tongue movement. For every subject $\mathcal{A},\mathcal{B},...,\mathcal{J}$ we used 200 training points $(x,y) \in \mathbb{R}^{150\cdot2} \times \{-1,1\}$ for further analysis.

5. Feature Selection and Classification

The motivation for feature selection in BCI research is twofold. The calculation of online feedback is quite time consuming since it involves preprocessing and classification of the data. Being able to work with a subset of the data is therefore favorable. Furthermore, identifying relevant recording positions may help in understanding the underlying cognitive processes.

Recursive Feature Elimination (RFE) (Guyon et al., 2003) is an iterative greedy backward embedded feature selection method. It is based on the training of several SVMs and exploits their margin characteristics to determine good features. In this paper we use
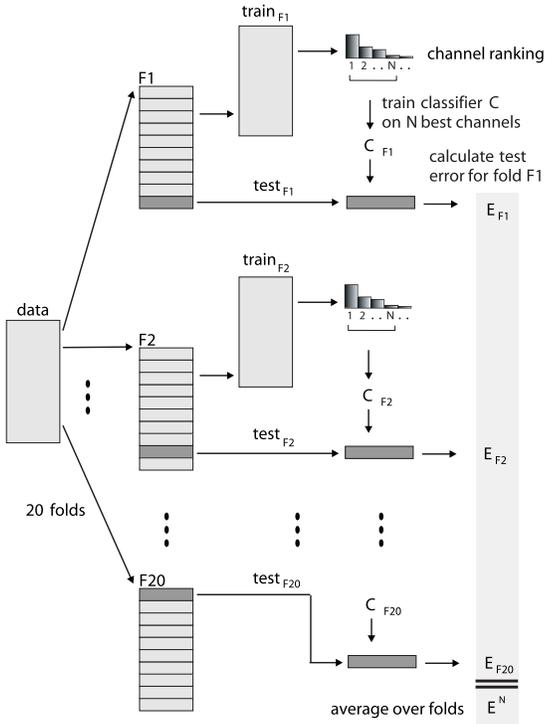
Figure 3. This plot describes the process of calculating an estimate of the expected risk when only using the best N ranked channels for classification. First the data is split into 20 train/test folds. The channels are ranked on each training set using Recursive Channel Elimination, a classifier is trained using the best N features only and tested on the corresponding test set. The 20 test errors are averaged. Please note that the set of channels used by the different classifiers might vary (figure from (Lal et al., 2004)).

Recursive Channel Elimination (RCE), an adaptation of RFE for the special case of EEG data (Lal et al., 2004). The pseudo-code of Algorithm 1 contains an overview of the data analysis.

We analyzed the data of the subjects separately. In step 2 of Algorithm 1 the data is randomly split into a training set which contains 80% of the data and a test set which contains the remaining 20%. Throughout the paper we use linear SVMs which are regularized using a ridge on the (linear) kernel matrix. Note that this is equivalent to a C-SVM formulation with quadratic slack variables (Cortes & Vapnik, 1995). On the basis of the training data the ridge which leads to the smallest CV-error is selected (see step 3 of Algorithm 1). This ridge is then used by the RCE procedure which produces a ranking of the MEG channels (see step 5).

The question arises, how many of the best ranked channels should be used as an input to a classifier.
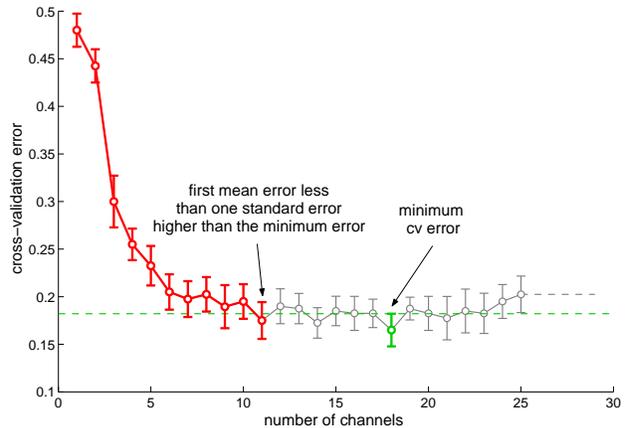

Figure 4. The graph shows the cross validation error estimates plotted against the number N of the best ranked channels. The error estimate of ,e.g., $N = 8$ was calculated based on the data from channels ranked 1th to 8th (and not only from the single channel at rank 8). Bars denote the standard errors. In this example, a combination of the 18 best ranked channels yielded the lowest average error. We are interested in the minimal number of channels such that if one further channel is dropped the resulting error rate will be significantly higher than the minimal error. We estimate this subset of channels by finding the smallest number of ranked channels such that the resulting error rate still lies within the two standard errors (α) or one standard error interval (β) of the minimal error. In case of (β) the 11 best ranked channels would be selected.

To answer this question we restrict the training data to the $N \in \{1, ..., 150\}$ best ranked channels and estimate a generalization error of an SVM trained with the reduced data using the cross validation technique shown in Figure 3. Figure 4 contains a schematic plot of these estimates (for $N \in \{1, ..., 25\}$). We select the number b of best channels as the minimum number of channels yielding an error estimate which deviates less than

(α) two standard errors, or

(β) one standard error

from the minimal error estimate.

In step 7 the data of the b best ranked channels is extracted. The ridge is optimized again and an SVM using the best ridge is trained. Finally the SVM is tested on the reduced test set. We repeat this procedure fifty times to obtain stable results.

6. Results of Offline Analysis

The error rates obtained with Algorithm 1 on data from ten subjects are summarized in Table 1. When using all channels the cross-validation error ranges

Table 1. Classification errors and standard deviations for the ten subjects when using Algorithm 1. The rightmost column contains CV-errors when using the data of all channels. The error rates obtained when using the subset of channels suggested by the (α) or (β)-estimates (see Sec. 5) are contained in the first two columns.

	α-ESTIMATE	β-ESTIMATE	ALL CHANNELS
$\mathcal{A}$	0.297 ± 0.076	0.285 ± 0.076	0.258 ± 0.076
$\mathcal{B}$	0.327 ± 0.072	0.328 ± 0.072	0.287 ± 0.072
$\mathcal{C}$	0.484 ± 0.076	0.473 ± 0.076	0.441 ± 0.076
$\mathcal{D}$	0.098 ± 0.052	0.085 ± 0.052	0.080 ± 0.052
$\mathcal{E}$	0.378 ± 0.071	0.395 ± 0.071	0.403 ± 0.071
$\mathcal{F}$	0.230 ± 0.071	0.227 ± 0.071	0.239 ± 0.071
$\mathcal{G}$	0.339 ± 0.076	0.310 ± 0.076	0.313 ± 0.076
$\mathcal{H}$	0.237 ± 0.065	0.218 ± 0.065	0.193 ± 0.065
$\mathcal{I}$	0.335 ± 0.070	0.333 ± 0.070	0.323 ± 0.070
$\mathcal{J}$	0.460 ± 0.078	0.470 ± 0.078	0.456 ± 0.078
$\varnothing$	0.318 ± 0.113	0.312 ± 0.119	0.299 ± 0.116

Table 2. This table summarizes the average sizes and standard deviations of subsets suggested by the (α) or (β)-estimates (see Sec. 5) for the ten subjects.
On average the β method suggested subsets of size 16.6. Using the α-estimate resulted on average in subsets of size 7.1.

	α-ESTIMATE	β-ESTIMATE	ALL CHANNELS
$\mathcal{A}$	6.1 ± 3.7	13.5 ± 11.2	150
$\mathcal{B}$	15.0 ± 21.7	33.3 ± 30.6	150
$\mathcal{C}$	13.1 ± 17.2	30.9 ± 30.1	150
$\mathcal{D}$	6.8 ± 3.6	15.7 ± 17.7	150
$\mathcal{E}$	1.2 ± 0.8	5.4 ± 16.8	150
$\mathcal{F}$	1.7 ± 1.0	4.1 ± 10.9	150
$\mathcal{G}$	3.3 ± 2.5	8.5 ± 6.0	150
$\mathcal{H}$	8.2 ± 5.0	21.0 ± 22.9	150
$\mathcal{I}$	5.6 ± 9.6	12.5 ± 17.2	150
$\mathcal{J}$	10.0 ± 15.4	21.3 ± 23.9	150
$\varnothing$	7.1 ± 4.6	16.6 ± 10.0	150

from chance level (subjects $\mathcal{C}$ and $\mathcal{J}$) to 8% (subject $\mathcal{D}$). When using the (α) or (β)-estimate to determine a feature subset, the average (taken over the subjects) error increases slightly from 29.9% to 31.8% (α) or 31.2% (β).

The number of channels was reduced significantly as can be seen in Table 2. The β-estimate suggested on average channel sets of size 16.6 and the α-estimate suggested channel sets of size 7.1. Note that this is less than 5% of the original 150 channels.

As an example of which channels are suggested by the recursive channel elimination method, we plotted the ten best ranked channels of subject $\mathcal{H}$ in Figure 5. This subset of channels lies over or close to the motor cortex and thus agrees well with the underlying cognitive process.

It is very difficult to compare results from this study with other studies since subject pools are usually very small and experimental setups vary. However, (Lal et al., 2004) conducted a comparable study using EEG-instead of MEG-recordings. Their subjects were asked to perform a left hand versus right hand movement imagery task. The authors applied the same machine learning algorithms. The average error rate using all EEG-channel was $33,75\%$. The average error rate reported in the present MEG-paper is $29,9\%$, which is smaller although only half as many training points were used. This finding supports the initial hypothesis that compared to an EEG-based BCI learning might be easier in an MEG-environment.
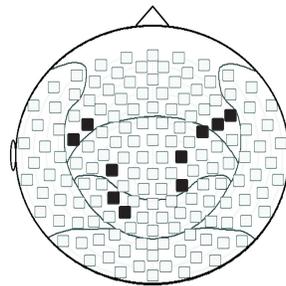

Figure 5. This figure contains the ten best ranked MEG channels from subject $\mathcal{H}$. They are located over or close to the motor cortex.

7. Possible Multi-class Extensions

Motor action or motor imagery is associated with decreasing spectral energy of the motor specific frequency bands (mu-rhythm, Sec. 2). This decrease is moderate on ipsilateral regions of the motor cortex and stronger on contralateral regions. After the motor action or imagination has ended, the motor rhythms re-establish within a few seconds and the associated frequency bands regain their previous intensity. The question arises, to what extent the MEG measurements captured the class-specific mu-rhythm changes of both the imagined finger and the imagined tongue movements. To answer this question we proceeded as follows: we generated spectrograms of the raw MEG signals in the time interval from second two through second three and from 0 Hz to 100 Hz. We then estimated the predictive ability of every point in the
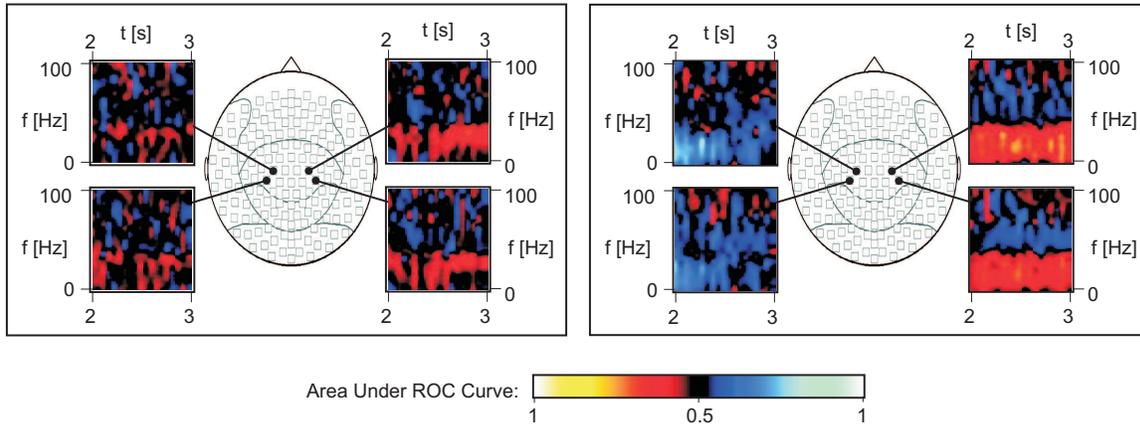
Figure 6. For four MEG channels of subject $\mathcal{H}$ that are situated symmetrically over motor cortex areas spectrograms of the time interval two to three seconds and 0 Hz to 100 Hz were generated. For every point of the time/frequency space, a ROC-score is plotted in the four boxes of the top graphic. Red regions code for decreased mu-activity of cortex regions that are involved in the planning of left little finger movements. Blue regions would be associated with brain regions involved in the planning of tongue movements (please see text for an explanation). Since clearly clustered blue regions cannot be found, we conclude that the MEG measurements did not capture possible decreases of mu-activity associated with imagined tongue movements. In the lower part of the plot blue and red regions are clearly visible. These ROC-scores were generated from data of subject $\mathcal{H}$ during a separate experiment using left hand versus right hand imagination.

time/frequency space of the spectrogram by calculating the area under its receiver-operating characteristic curve (further called the ROC-score). The upper plot of Figure 6 contains these ROC-scores of four selected MEG channels of subject $\mathcal{H}$. A bright color[2] indicates ROC-score values close to zero or close to one. For a point in this plot this means that the specific frequency at a particular time contains class specific information. If the score is close to 0.5 it denotes that this feature (taken by itself) does not carry useful information for the classification task at hand[3].

Increased ROC-scores (> 0.5 plotted in hot colors) can technically be explained in two ways. The energy at the particular time and frequency either

- increases in class $+1$ (tongue), or
- decreases in class -1 (left little finger).

In the same manner ROC-scores (< 0.5 plotted in cool colors) can result from energy at a particular time and frequency

- decreasing in class $+1$ (tongue), or
- increasing in class -1 (left little finger).

Since motor imagery is associated with decreased energy we expect that red colors appear at cortical regions which are involved in the planning of finger movements and blue regions at locations involved in the planning of tongue movements.

The two right boxes of the left plot of Figure 6 show low ROC-scores for the mu-band over the right motor cortex throughout the full time interval. Following the previous argument we conclude that the scores are generated by the part of cortex involved in the planning of left finger movements. However neither of the four ROC plots of the left plot shows clustered blue regions. It seems that there was either no decrease in mu-activity during the imagination phases of tongue movement, or the MEG measurements did not capture possible decreases. This finding holds true for all ten subjects. Most probably the classifiers (only) have learned to detect whether a finger movement was imagined or not.

For classification scenarios in two-class problems it is sufficient to know under which conditions a training point belongs to one particular class. When dealing with more classes however, the class membership can not be inferred from knowing that a point does not belong to a particular class. The tongue versus finger paradigm might therefore be suboptimal with regard to a possible extension to a multi-class system.

A left hand versus right hand task seems to be better suited for a multi-class extension. The right plot of Figure 6 shows the ROC-scores of the same subject during a imagined left hand versus right hand task (all other experimental parameters were the same as in the finger versus tongue experiment). Here the decrease of the mu-activity was clearly captured by the MEG recordings.
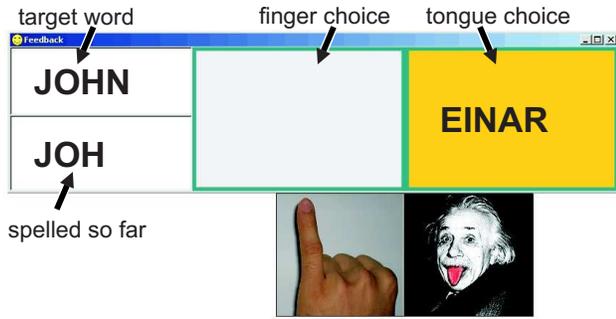
---

[2]A color version of this paper is available online on the ICML website.

[3]Note that this way of analyzing the data does not take into account combination of features.

Figure 7. Screenshot of the projection screen while a subject writes the word "JOHN" (shown at the upper left box). The lower left box contains the letters "JOH" spelled so far. About every seven seconds, the subject imagines a movement to maneuver within the binary spelling tree. In the present situation the next letter "N" to write is among the letters of the right box. The subject can chose this box by imagining tongue movements (indicated by the thumbnail pictures added underneath the speller as reminders). In this case a subset of letters "EINAR" would appear next. An imagined finger movement would communicate that the intended letter is not among the displayed ones and the empty left box would be chosen. In this case the speller changes to the sibling branch and a different set of letters will appear next. After a few iterations only one letter remains. If it is chosen by the subject, the letter will be appended to the word spelled so far.

The computer system decoding the imagined movements is a combination of the feature selection technique recursive channel elimination and linear SVMs.

8. Online Brain Computer Interface

During the part of the experiment described in the previous sections, online feedback was not presented to the subjects. In this section we describe how four of the subjects used a trained SVM to write a short name. This part of the experiment was carried out directly after the first part. Since the subjects had to wait while we analyzed their data we could not estimate all parameters as described in the previous section. Instead we used slightly suboptimal parameter settings.

For every subject we used the 200 training points (Section 3). The detrended data were preprocessed using an AR model of order 6. Similar to Algorithm 1, we selected a ridge, ranked the channels using RCE, restricted the data set to the best 20 ranked channels, selected a ridge again and trained an SVM on the restricted data.

We asked the subjects to perform the same task as before. After every trial that was classified correctly by the SVM a "Smiley" was displayed. Depending on their performance the subjects completed two to four blocks of fifty trials.

The subjects $\mathcal{B}$, $\mathcal{D}$, $\mathcal{F}$, $\mathcal{H}$ and $\mathcal{I}$ obtained an accuracy higher than 70% in their last block. For these subjects we combined the collected data of both experimental parts and trained an SVM as described above. The SVM was then used in the third part of the experiment during which the subjects spelled a short name. At the beginning of the spelling experiment the spelling method displayed one half of the letters of the alphabet (including some special characters). If the letter to be spelled was among the displayed ones, the subject had to imagine a tongue movement. To communicate that the letter was not displayed, the subject imagined a finger movement (see Figure 7). To help the subjects concentrate on the imagination task, the box of correct choice was highlighted (this spelling variant is sometimes referred to as "copy spelling" and useful for training subjects before proceeding to "free spelling"). In the next step the selected subset of the alphabet was split into two parts again, one of which was displayed. On the last stage of this process the letter had to be confirmed and was displayed on the left part of the screen. The procedure started over again to allow the selection of further letters.

The spelling algorithm allows deletion of already selected letters. Furthermore the splitting algorithm was optimized such that it reflects letter frequencies of the language. For more details please refer to (Birbaumer et al., 1999).

Four out of five subjects succeeded in spelling a short name (4.25 letters on average). The fifth subject aborted the experiment after successfully spelling the first letter of a name.

9. Summary

We demonstrated how machine learning techniques can be used to set up an online brain computer interface on the basis of magnetoencephalographic (MEG) recordings. We reported results of a tongue versus left little finger imagined movement task from ten healthy subjects. The classification performance ranged across subjects from chance level up to 92%. We showed that it is possible to reduce the number of used MEG channels to less than 5% of the original 150 channels without significant loss of classification performance. The proposed method worked well enough to allow four subjects within only one session to write a short name using imagined movements only.

Furthermore, we gave evidence that a left hand versus right hand task might be better suited for future MEG based brain computer interfaces.

Our results encourage the use of MEG-technology for

screening as well as initial training with a later transfer to a portable and EEG based brain computer interface. Although the current design certainly has room for improvement, its main message is a prove of concept: to our best knowledge this is the first demonstration of a functioning MEG based brain computer interface.

Acknowledgements

References

Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., Kübler, A., Perelmouter, J., Taub, E., & Flor, H. (1999). A spelling device for the paralysed. *Nature*, *398*, 297–298.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.

Georgopoulos, A. P., Leuthold, A. C., & Langheim, F. J. P. (2004). Motor trajectory from magnetoencephalographic (MEG) data. *abstract, Society for Neruoscience*, *884.1*.

Graimann, B., Huggins, J. E., Levine, S. P., & Pfurtscheller, G. (2004). Towards a direct brain interface based on human subdural recordings and wavelet packet analysis. *IEEE Transactions on Biomedical Engineering*, *51*, 954–962.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2003). Gene selection for cancer classification using support vector machines. *Journal of Machine Learning Research*, *3*, 1439–1461.

Haykin, S. (1996). *Adaptive filter theory*. Upper Saddle River, NJ, USA: Prentice-Hall International, Inc.

Kauhanen, L., Rantanen, P., Lehtonen, J. A., Tarnanen, I., Alaranta, H., & Sams, M. (2004). Sensorimotor cortical activity of tetraplegics during attempted finger movements. *Biomedizinische Technik*, *49 Supp 1*, 59–60.

Lal, T. N., Hinterberger, T., Widman, G., Schröder, M., Hill, N. J., Rosenstiel, W., Elger, C. E., Schölkopf, B., & Birbaumer, N. (2005). Methods towards invasive human brain computer interfaces.

In L. K. Saul, Y. Weiss and L. Bottou (Eds.), *Advances in neural information processing systems 17*. Cambridge, MA: MIT Press.

Lal, T., Schröder, M., Hinterberger, T., Weston, J., Bogdan, M., Birbaumer, N., & Schölkopf, B. (2004). Support Vector Channel Selection in BCI. *IEEE Transactions on Biomedical Engineering. Special Issue on Brain-Computer Interfaces*, *51*, 1003–1010.

Leuthardt, E., Schalk, G., Wolpaw, J., Ojemann, J., & Moran, D. (2004). A brain computer interface using electrocorticographic signals in humans. *Journal of Neural Engineering*, *1*, 63–71.

Mitchell, T., Hutchinson, R., Just, M., Niculescu, R., Pereira, F., & Wang, X. (2003). Classifying instantaneous cognitive states from fMRI data. *Proceedings of the American Medical Informatics Association Annual Syposium* (pp. 465–469).

Pfurtscheller., G., Neuper, C., Schlögl, A., & Lugger, K. (1998). Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. *IEEE Transactions on Rehabilitation Engineering*, *6*, 316–325.

Ramoser, H., Müller-Gerking, J., & Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering*, *8*, 441–446.

Salmelin, R., & Hari, R. (1994). Spatiotemporal characteristics of sensorimotor neuromagnetic rhythms related to thumb movement. *Neuroscience*, *60*, 537–550.

Schröder, M., Lal, T. N., Hinterberger, T., Bogdan, M., Hill, N. J., Birbaumer, N., Rosenstiel, W., & Schölkopf, B. (2005). Robust EEG channel selection across subjects for brain computer interfaces. *EURASIP Journal on Applied Signal Processing*. in press.

Tiihoonen, J., Kajola, M., & Hari, R. (1989). Magnetic mu rhythm in man. *Neuroscience*, *32*, 793–800.

Weiskopf, N., Mathiak, K., Bock, S. W., Scharnowski, F., Veit, R., Grodd, W., Goebel, R., & Birbaumer, N. (2004). Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). *IEEE Transactions On Biomedical Engineering*, *51*.

Wolpaw, R., & McFarland, D. (1994). Multichannel EEG-based brain-computer communication. *Electroencephalography and Clinical Neurophysiology*, *90*, 444–449.