# FEATURE SELECTION FOR BRAIN-COMPUTER INTERFACES

**Marcel van Gerven**[1], **Ali Bahramisharif**[1], **Christian Hesse**[2], **Ole Jensen**[2], **Tom Heskes**[1]

[1]Intelligent Systems, Institute for Computing and Information Sciences, Radboud University Nijmegen
[2]F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen

## 1. Brain-Computer Interfaces

A brain-computer interface (BCI) is a device that allows a user to interact with the external world based on measurements of brain activity alone [1]. This has diverse applications such as communication and control by patients that suffer from severe disabilities, new forms of of human-computer interaction, as well as novel methods of data-analysis in cognitive neuroscience. We mainly focus on the use of electroencephalography (EEG) and magnetoencephalography (MEG).

## 2. Experimental Paradigms

In order to achieve BCI control, we need to identify to which experimental condition single-trial data belongs. In our research, we use imagined movement and covert attention as our experimental paradigms.

- **Imagined movement:** Imagined movement of the left or right hand leads to modulations of alpha (8-12 Hz) and beta (16-24 Hz) band power over the motor cortex contralateral to the side of imagined movement.
- **Covert attention:** Covert attention to the left or right visual field leads to modulations of alpha band power in posterior channels.

## 3. Feature Selection

- In order to build a robust BCI, we need to identify which components of brain activity are modulated by the experimental conditions.
- There are thousands of components (sensors $\times$ frequency bands $\times$ time windows).
- We need to perform feature selection in order to
  - increase classification performance
  - improve model interpretability
- We use a regularization approach that minimizes $E(\Theta) = L(\Theta) + R(\Theta)$ where
  - $L(\Theta)$ is the negative log-likelihood of the parameters given the data
  - $R(\Theta)$ is the negative logarithm of the prior on the parameters

## 4. Groupwise regularization

- Our goal is to select groups of features instead of individual features.
- We use a Laplace prior on feature *groups* $\boldsymbol{\theta}_i \subseteq \Theta$:

$$p(\Theta) = \lambda \prod_i \frac{1}{2} \exp\left(-\lambda \parallel \boldsymbol{\theta}_i \parallel_p\right) .$$

- Known as $\ell_1/\ell_p$ regularization since $R(\Theta)$ is equivalent to the $\ell_1$ norm of $\ell_p$ norms on each group.
- The regularization parameter $\lambda$ controls the number of selected groups.
- We constructed an algorithm that efficiently samples the regularization path by evaluating only those values of $\lambda$ for which feature groups become unstable [2].

## 5. Examples

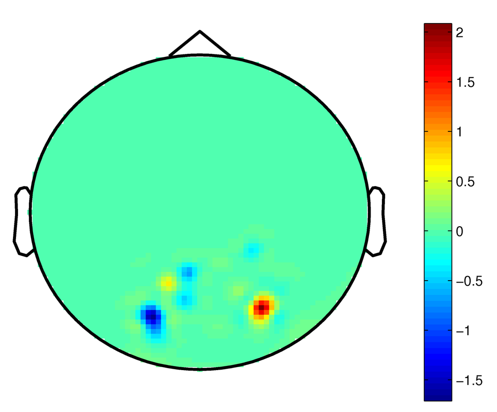

**Figure 1:** *Standard $\ell_1$ regularization allowed us to identify regions of interest for the covert attention paradigm (80% correctly classified trials for this subject). Measurements were made using a 275 channel MEG system. Note the correct identification of activity in posterior sites. The figure shows a scalp-projection of estimated classification parameters.*

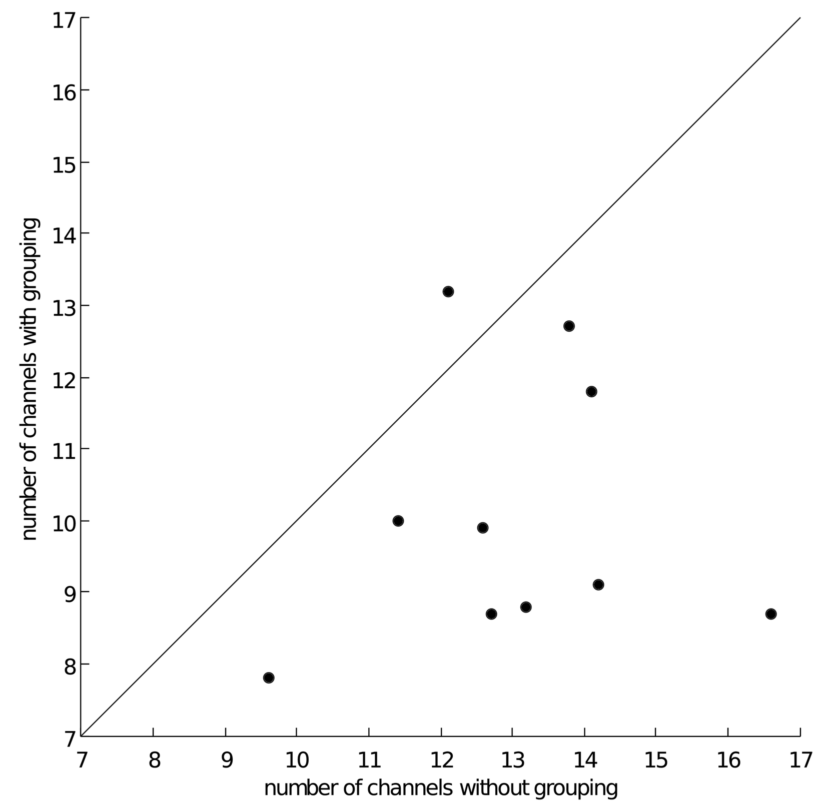


**Figure 2:** *Groupwise regularization allowed us to select a smaller number of EEG sensors that perform well for the imagined movement paradigm. This was realized by placing components that belong to the same sensors into the same group. Measurements were made using a 24 channel EEG setup. We observed a 17% decrease in the number of used channels on average.*

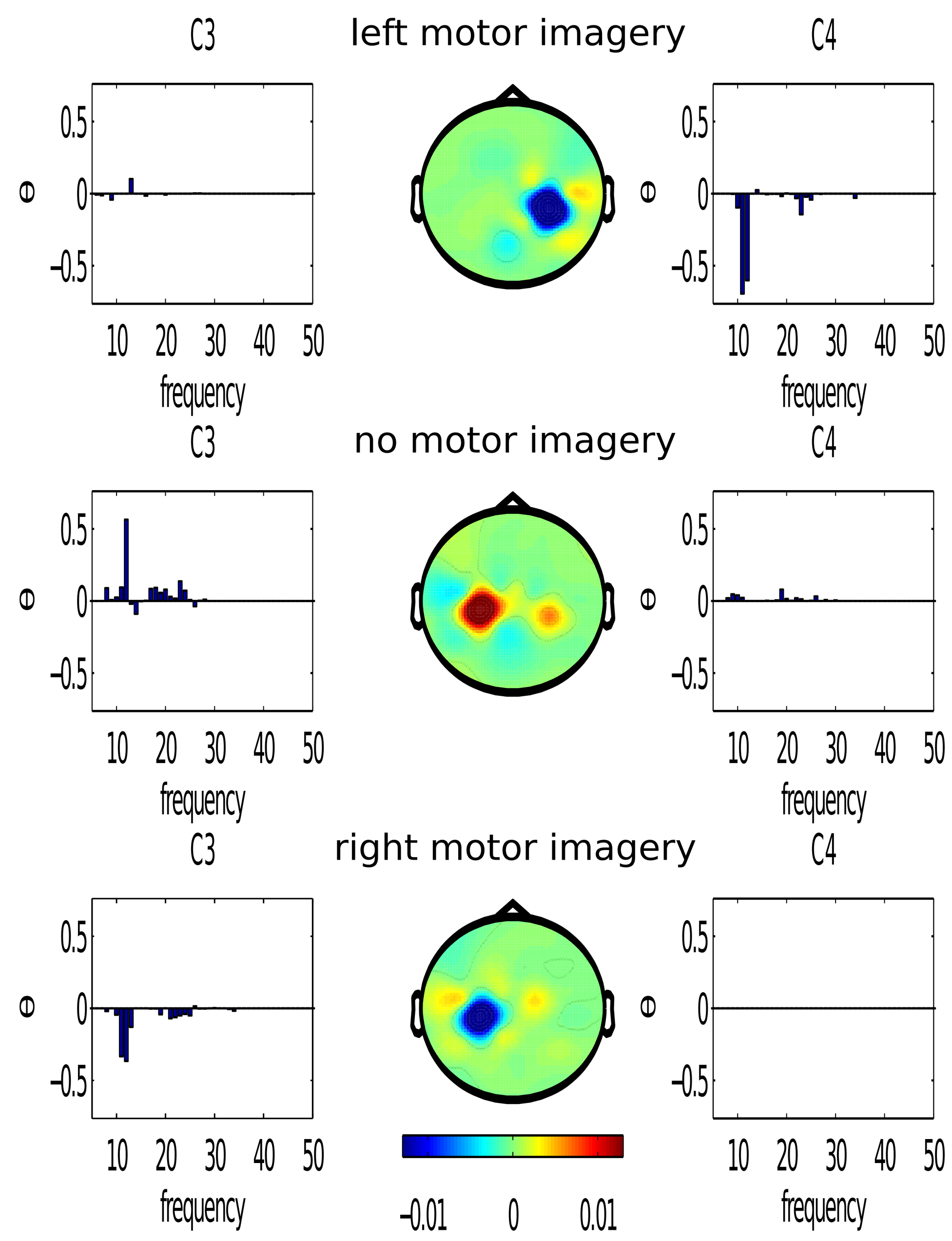


**Figure 3:** *Groupwise regularization can be used for transfer learning. By defining groups over subjects, we learned a model that uses the same features for each subject while adapting parameters per subject. The model is shown here for one subject using the imagined movement paradigm. Note the correct identification of alpha and beta modulations. The figure shows estimated classification parameters projected back onto the scalp per experimental condition.*

## 6. Future goals

- Examine other regularization strategies to learn the commonalities and differences between subjects.
- Use transfer learning to generalize better over multiple BCI sessions.
- Extend the covert attention paradigm to allow for 2-D BCI control.

## References

[1] J. R. Wolpaw and N. Birbaumer and D. J. McFarland and G. Pfurtscheller and T. M. Vaughan. Brain-computer interfaces for communication and control. Clinical Neurophysiology, **113**, 767–791, 2002.

[2] T. Heskes and M.A.J. van Gerven. Stability conditions for $\ell_1/\ell_p$ regularization. Technical Report ICIS-R07034. Radboud University Nijmegen, The Netherlands. 2008

project number 07050 Netherlands Organisation for Scientific Research VICI grant 639.023.604 BrainGain