# EEG signal classification

## Introduction

In this assignment, we use simple binary classifiers (e.g., NCC and LDA) to accomplish a task in the field of Brain Computer Interface (BCI) called P300 speller *paradigm*. Our goal is to predict the cognitive state corresponding to the character (target vs. non-target) on the speller where the subject was paying attention through P300 (a kind of deflection in EEG caused by certain stimuli).

### Paradigm

The subject (or user) is presented with a speller which could be a matrix of characters. The user's task is to focus attention on characters in a word that was prescribed by the investigator (i.e., one character at a time). When the speller starts, each single character is intensified in a random order. A single repeat (round) includes a number of flashes corresponding to a number of individual characters (e.g., 36). The subject needs to focus on one target character in the matrix at a time. In each run, the EEG brain signals are recorded for further analysis. Specifically, by extracting the features from the EEG signals, the single target character could be detected after several rounds of intensifications using machine learning classifiers. Here, the character can be considered as the cognitive state which could be either target or non-target stimulus corresponding to character. The typical workflow is shown in Fig. 1.

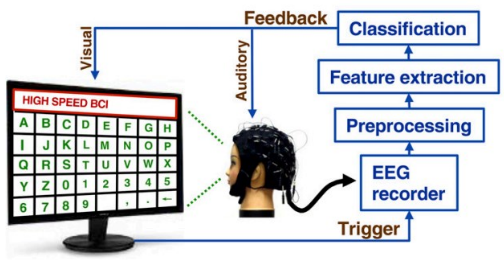


**Figure 1.** Workflow for predicting the cognitive states from EEG data.

### Data collection and feature extraction

EEG signals were recorded using 64 electrodes as shown in Fig. 2. In this assignment, we consider 62 electrodes and exclude AF3,4 because they are unlikely to contribute substantially to the classification task.

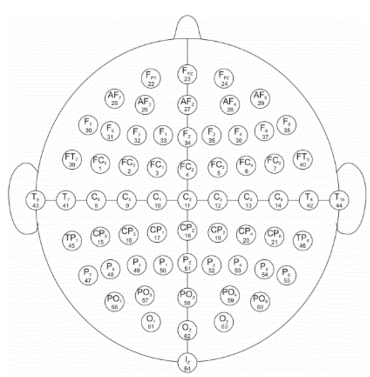

**Figure 2**: This diagram illustrates electrode designations.

Subsequently, the recorded 62 channel EEG signal is organized in one big matrix (Signal) as illustrated in Figure 3. In this matrix, the row corresponds to the electrode channel while the column corresponds to the time recording the EEG data. For feature selection, we select 5 temporal windows of EEG activity, and voltages are averaged within these windows. Hence, a feature can be represented as 62×5 matrix. The 2$^{nd}$ row of Fig. 3 shows the scalp maps of 5 temporal windows.

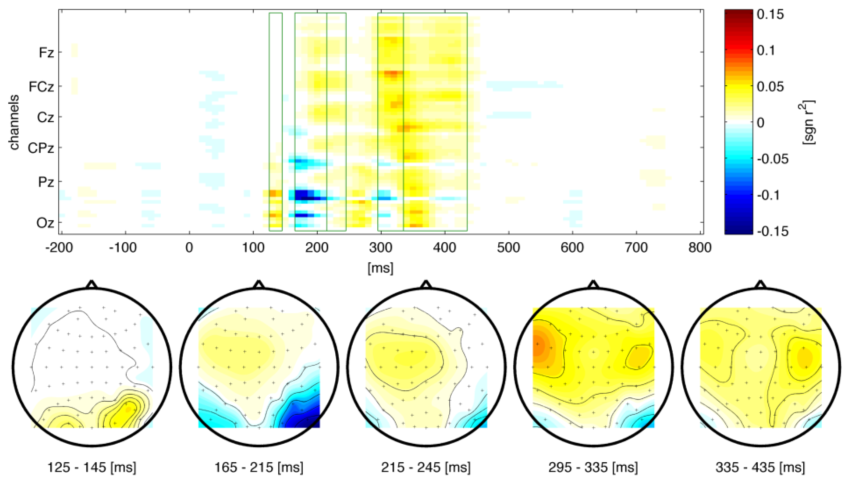


**Figure 3**: Illustration of EEG signal

Figure 4 shows an example of event-related potentials (ERPs), which is brain response to an internal or external event such as a sensory stimulus, for targets and non-targets for two selected electrodes Cz and PO8. Note that the neural processing of a sensory stimulus is associated with positive and negative ERP components that can be extracted from the electroencephalogram (EEG).

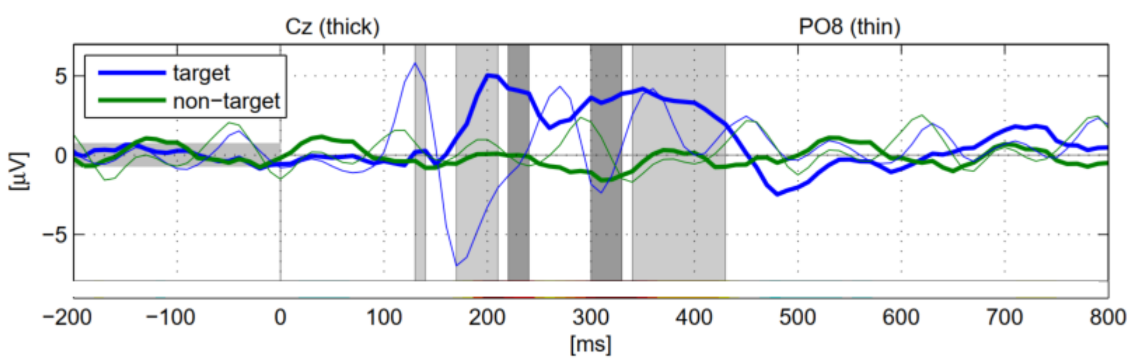


**Figure 4**. ERPs for targets and non-targets for two selected electrodes Cz and PO8. The five shaded areas mark the selected windows.

## TASK: NCC and LDA implementations.

Data Electroencephalography (EEG) data was recorded during a copy-spelling BCI Experiment at the Berlin BCI group (http://www.bbci.de). You can find a video demo attached with this assignment, or the original paper and demo at this link: http://iopscience.iop.org/1741-2552/8/6/066003/media

The data set consists preprocessed EEG data $\mathbf{X} \in \mathbb{R}^{5\times62\times5322}$ and stimulus labels $\mathbf{Y} \in \mathbb{R}^{5\times5322}$ during a copy-spelling paradigm with a P300 speller. The data matrix $\mathbf{X}$ contains 5 selected time windows of EEG activity at 62 electrodes after a visual stimulus was presented on the screen in front of the subject. If the first row of Y is 1, the stimulus was a target stimulus, if the second row of Y is 1, the stimulus was a non-target stimulus. Hence, in total, we have 5322 EEG data samples where each is associated with a label y (target or non-target).

1. How well can we predict the cognitive state from the EEG data?
   a. Implement a nearest centroid classifier.
   b. Implement a linear discriminant analysis (LDA) classifier

   Train both classifiers on the $5 \cdot 62 = 310$ concatenated (across time and electrodes) features to predict target stimuli from the EEG data. Train both classifiers on 70% of the data and test it on 30% of the data. Select the data points for training and test randomly, but make sure they do not overlap!

2. Compare the prediction accuracy for each classifier. For both classifiers: plot histogram of classifier outputs for targets/non-targets using pylab.hist (see Fig. 5)

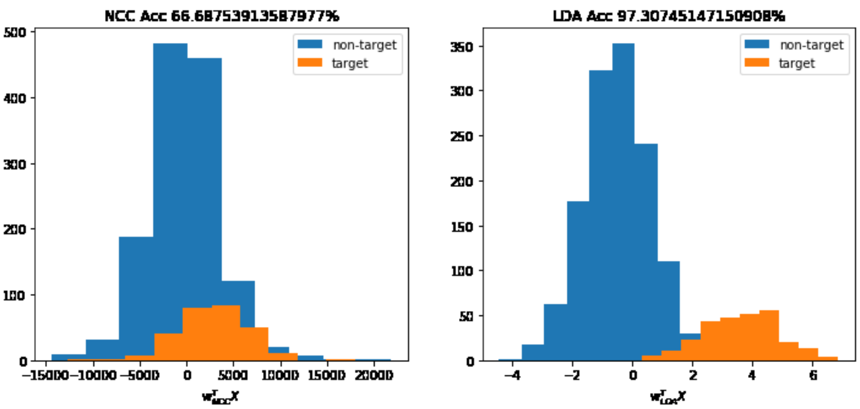


**Figure 5.** Left: Classifier outputs for nearest-neighbor classifier. Right: Classifier outputs for linear discriminant classifier

3. Estimate the generalization performance of each classifier. Test each classifier on the training set data and on the test data data. Compare the prediction accuracies (see fig.6).

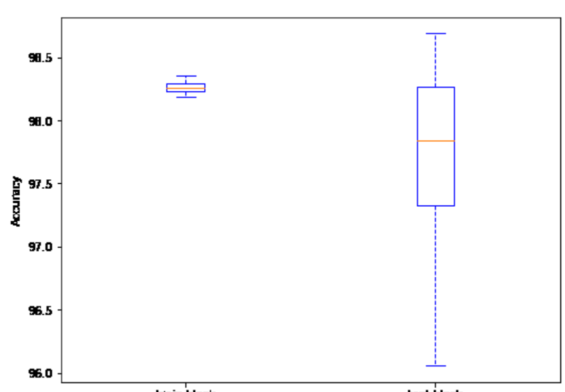


**Figure 6:** Example of LDA classification accuracy for training and test blocks in 10-fold cross-validation.