# III. Albany dataset

**D.M.J.Tax and Benjamin Blankertz**
**Fraunhofer FIRST.IDA**
**Kekuléstr. 7, D-12489 Berlin, Germany**
davidt@first.fraunhofer.de

**Problem**

A user is presented a $6 \times 6$ matrix of characters. Randomly each row and column is highlighted 15 times. When the user focuses on one letters, a highlighting of the correct row or column evokes a P300 response. The task is to find the letter the user was focusing on from the user EEG responses of the 15 repetitions of the 12 highlights.

**Preprocessing**

From the original data, the channels $F3, F1, Fz, F2, F4, FC3, FC1, FCz, FC2, FC4, C5, C3, C1, Cz, C2, C4, C6, CP3, CP1, CPz, CP2, CP4, P7, P8, AFz$ were taken (25 channels in total), and epochs around the row/column highlights were extracted (from -50 to 500 ms). A baseline-correction was made by subtracting the average signal taken between 0 and 150 ms. Then the signal between 250 and 451 ms. was 10 times subsampled, and 5 values per channel were obtained. This $5 \times 25$-dimensional feature vector was used in the classifier. In total 7560 objects were present in the training set (that means 42 letters).

**Classifier and combining**

The best classifier for the individual P300 responses is the regularized linear discriminant, with a regularization constant of 0.0375. This was obtained using 10 times 10-fold cross-validation.

The $12 \times 15$ epochs for each letter were classified. Each repetition of 12 epochs was re-sorted to get the 6 row and 6 column response in standard order. Thus a matrix x of $12 \times 15$ outcomes were obtained. The lowest and highest 3 classifier outcomes were ignored, and from the rest the average was taken:

```
sortx = sort(x,2);              % sort the 15 classifier outputs
x_ave = mean(sortx(:,4:12),2); % remove the extrema and average
```

Finally, the maximum for the 6 rows and the maximum for the 6 columns was taken as the final output. These indices were used to find the letters in the letter-matrix.

**The results**

The results (using 5-fold cross-validation) on the training set was 3 errors out of 42:

```
C G* T D O G F I S H W A T E R B O W L H A T G* A T G L O V E S H O E Y*
F I S H R A T
```

(the letters indicated by * are errors, the classifier outputs do not show much hope for improvement here).

The outcome of the test set is

```
F O O D M O O T H A M P I E C A K E T U N A Z Y G O T 4 5 6 7
```