*Supplementary text 1*: **Error estimation procedure**

Classifier accuracy was estimated with 25-fold cross-validation using all the available data for the given subject. In each iteration of the cross-validation, 70% of the data was used for parameter estimation; each such set was divided into 70% training and 30% validation, the remaining 30% of the data was used for testing.

For a classification problem that uses regularization one typically expects that the (estimated) classifier error as function of regularization parameter exhibits a clear global minimum. In our case, when plotted against the regularization parameter, the classification error clearly revealed such minimum in all subjects (see figure s1).
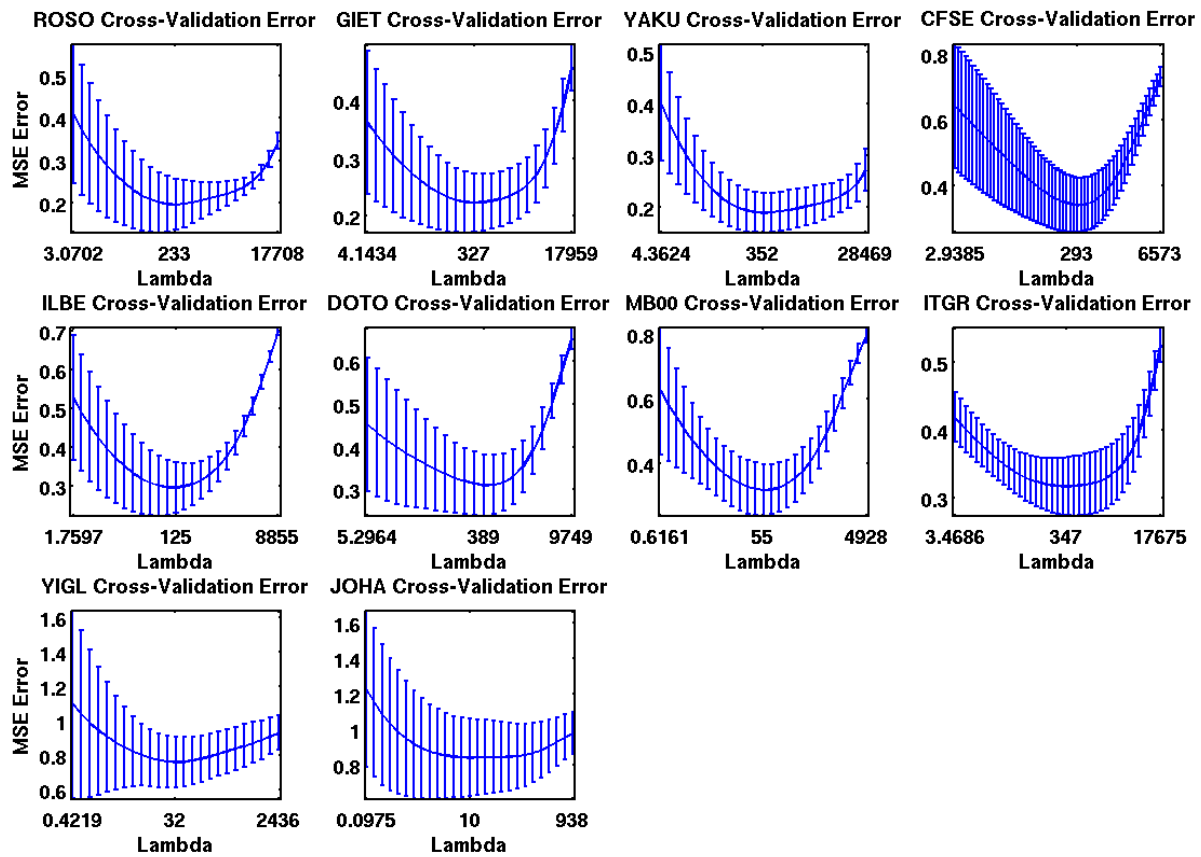
**Figure s1**: Cross-Validation Error graphs for all subjects.

Cross validation error estimate obtained by 25-fold cross validation with Ridge regression. Error bars denote 1-std – wide margin around the error estimate. Only a segment between sub-optimal and above-optimal lambda is shown. The middle value of the lambda axis is the optimal lambda that minimizes cross-validation error. First and last values of lambda are approximately 100 smaller and larger than the optimal lambda, respectively.

*Supplementary text 2*: **Error bias estimation**

Since minimizing the error over any free parameters biases the error estimate downwards, we compared the estimated error to the estimate obtained by applying exactly the same algorithm to the data with randomly scrambled class labels. Classification error of the data with scrambled labels was, on average, at chance level for all subjects (see figure s2). The difference between the mean error estimates for true and shuffled labels, was significant for all subjects (p<0.0001, FDR corrected) using a student's t-test.

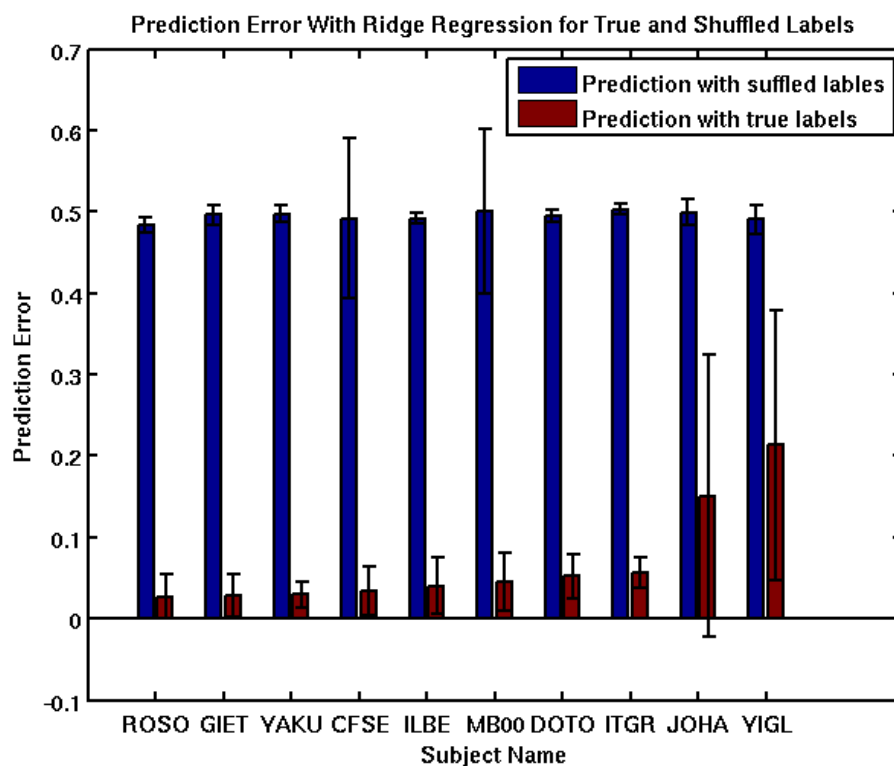*Supplementary figure 2*: **Prediction Error with shuffled labels**



**Figure s2:** Prediction Error with shuffled labels

Classifier error rates of Ridge regression for all subjects. Regularization parameter and electrode were selected to minimize the classification error using 25-fold cross-validation. In blue: Control results obtained using the same algorithm on data with randomly scrambled target labels. Error bars denote 1-std – wide margin around the error estimate. Average estimated prediction error (red) for most subjects is below 10% which is well below the values that might be expected by chance (blue), which are around 50%.

*Supplementary figure 3:* **Comparison between Ridge and Logistic Regression**
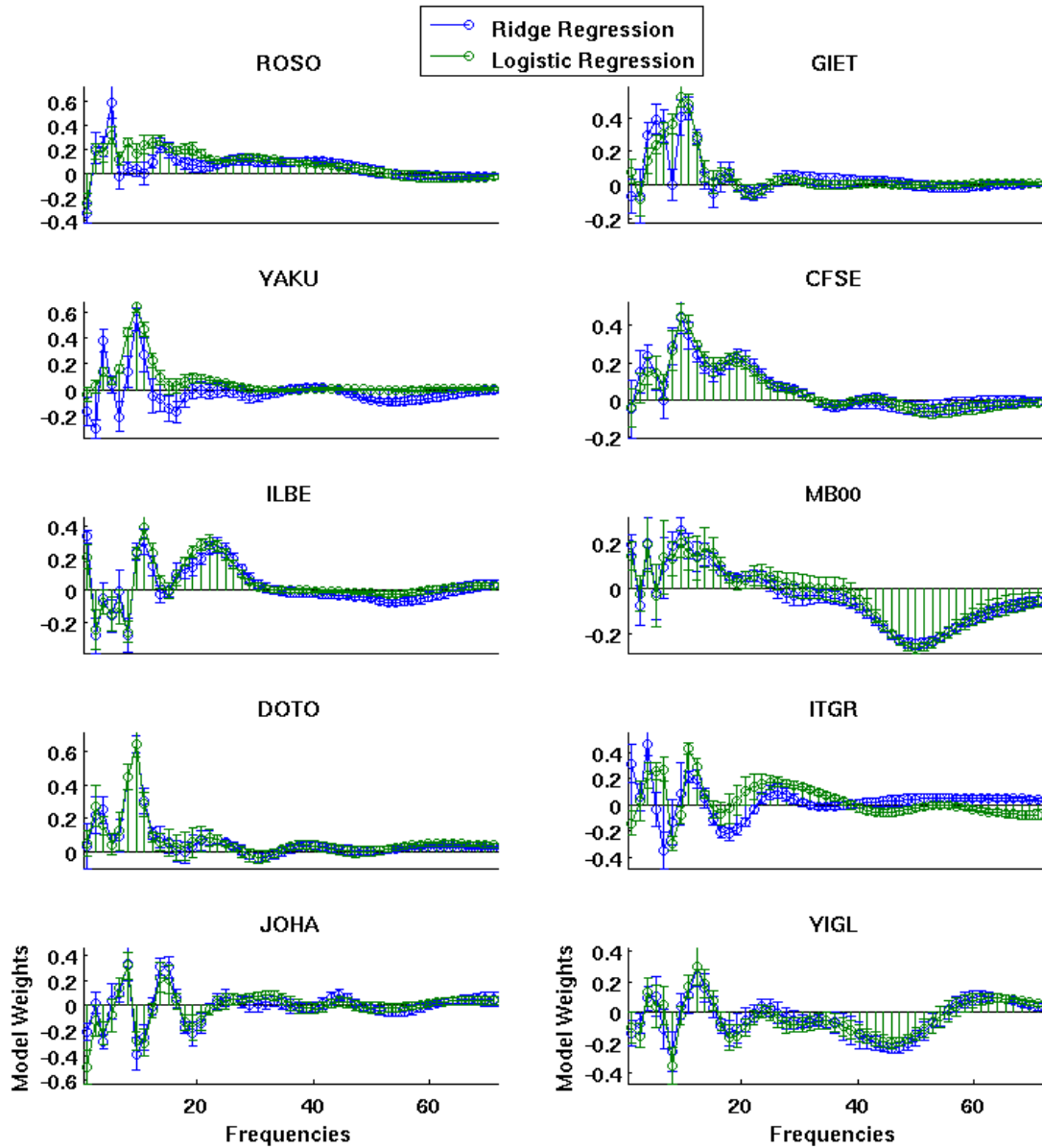


**Figure s3:** Comparison of regression weights with Logistic and Ridge Regression.

Regression frequencies weights for Linear Ridge (blue) and Regularized Logistic (green) Regression for each subject. Subjects are sorted by performance of Ridge Regression Classifier (best first). Each weight is normalized by magnitude. Distribution of weights across frequencies is similar between the two methods. Highest weights are distributed in the Alpha frequency range as expected from the paradigm of experiment, but higher frequencies also receive significant weights. Each weight is normalized by magnitude for comparison.

*Supplementary text 3*: **Relation between classifier error and data normalization**

As part of the classifier construction, the importance of data normalization was explored.

The data matrix given to the classifier consists of columns of frequencies across time. Each frequency column is mean-centered to zero but the difference of amplitudes between various frequencies is large, mainly frequencies around 10Hz have very high amplitudes compared to higher frequencies such as the 40Hz range. This led us to hypothesize there may be important information in the higher frequency range that gets unaddressed by the classifier because of the information it gets from the high amplitude frequencies. This problem could be solved by normalizing each frequency column by its standard deviation, this way all frequencies would have the same amplitude and only their modulation according to the experimental paradigm would contribute to the classifier. However this strategy may artificially introduce noise into the classifier since frequencies with low standard deviation might contain more noise than frequencies with high standard deviation. This question of weather normalization of amplitude is needed was addressed experimentally. Ridge Regression classifier was trained to predict eyes state in two cases:

1. with data where all frequencies were normalized by standard deviation

2. with data where only frequencies with standard deviation higher than $70^{th}$ percentile were normalized, assuming that large amount of noise will exist in frequencies with standard deviation lower than $70^{th}$ percentile.

Comparison of prediction error for the classifier without normalization and with the two types of normalization by standard deviation is shown in figure s4. It is clear that performance without amplitude normalization is better. This indicates that normalization introduced more noise than information into the classifier.

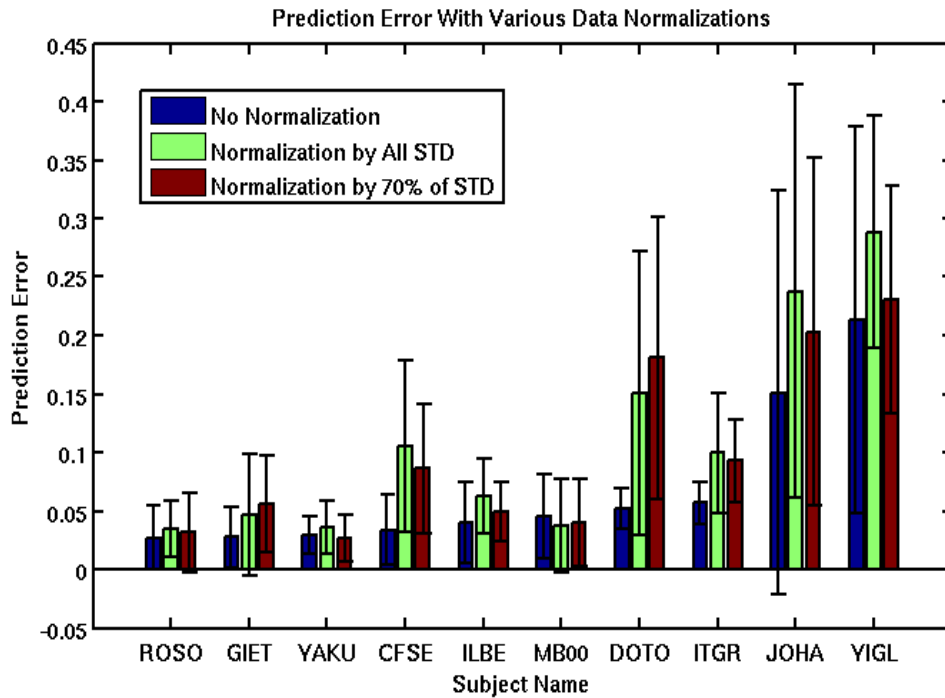*Supplementary figure 4:* **Prediction Error with various normalization techniques**



**Figure s4:** Prediction Error with various normalization techniques

Performance of Ridge Regression classifier with various types of data normalization. Without normalization prediction error is significantly lower for all but one subject (t-test, $p<0.05$). Error bars denote 1-std – wide margin around the error estimate determined by cross-validation.

*Supplementary text 4*: **Ridge regression weights with various regularization parameters**

The classification model describes the relative contribution of each frequency to the prediction and the sign of each weight describes which of the two classification labels is supported by a positive value of the corresponding frequency. In figure s5, the average value and the standard deviation of weights across cross validation folds are presented (denoted by a circle and error bar at each frequency).

Resulting frequency weights with optimal regularization parameter (chosen with cross validation) and with above and below optimal regularization parameters had significantly distinct distributions (paired t-test with optimal weights set, $p < 0.05$, FDR corrected) as can be seen in figure s5.

The over regularized classifier produced very similar models for all subjects, its weights distribution is similar to the classical FFT power distribution of the EEG signal in this experiment, showing mainly high power of the Alpha band frequency (8-13 Hz). The optimal regularization model revealed high contribution of the Alpha band to the prediction as well as lower but significant contribution of other frequencies such as Beta and Gamma. In addition the optimal model revealed a detailed division of the frequencies into bands that contribute positively and negatively to the prediction. This division is not seen by the classical FFT method or by the over constrained model.

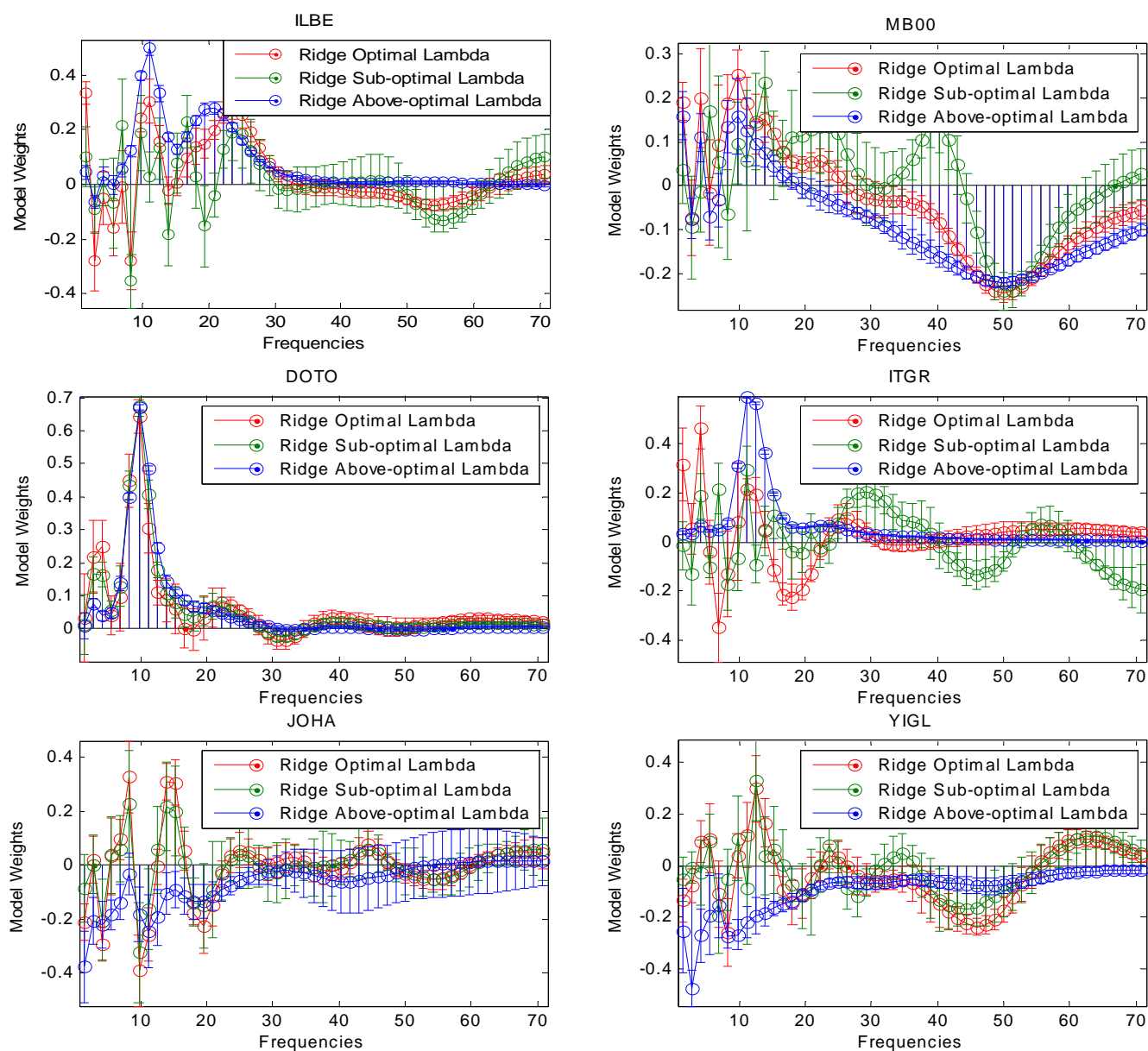*Supplementary figure 5:* **Ridge regression weights with various regularization parameters**

**Figure s5:** Ridge regression weights with various regularization parameters.

In red: Result with optimal regularization parameter determined by cross validation. In green: result with regularization parameter 100 times smaller than the optimal one. In blue: result with regularization parameter 100 times larger than the optimal one. The two un-optimal weight sets differ significantly from the optimal one. Each weight is normalized by magnitude for comparison.

*Supplementary text 5*: **Spatial Distribution of performance**

Linear Ridge regression classifier was applied to the data of each subject from each electrode separately. The classifier was trained to predict subjects' eyes state (opened or closed). Regularization parameter was chosen for each electrode separately and classifier performance was assessed by means of cross validation.

Good performance was achieved with data from only one electrode, however performance differed between electrodes. See figure s6 for distribution of prediction performance across electrodes. For all subjects electrodes with best prediction strength are located in the occipital area. This finding is consistent with known findings about the occipital origin of the Alpha band that modulates most evidently with eyes state.

For each subject, one electrode with the best classifier performance was chosen for further analysis (for further details see Table s1).

*Supplementary figure 6:* **Spatial distribution of classifier performance in light condition**
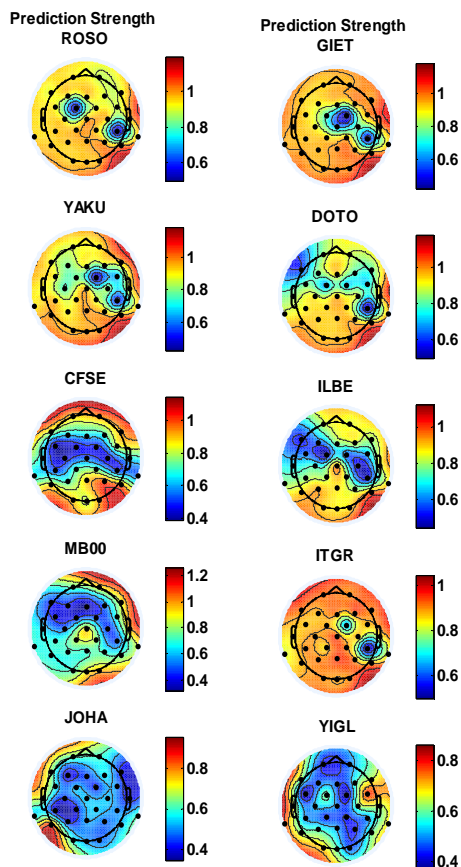


**Figure s6**: Spatial distribution of classifier performance in light condition.

Linear Ridge regression Classifier was applied to each channel separately to identify channels with best prediction strength. This figure shows distribution of classifier performance across channels for each subject

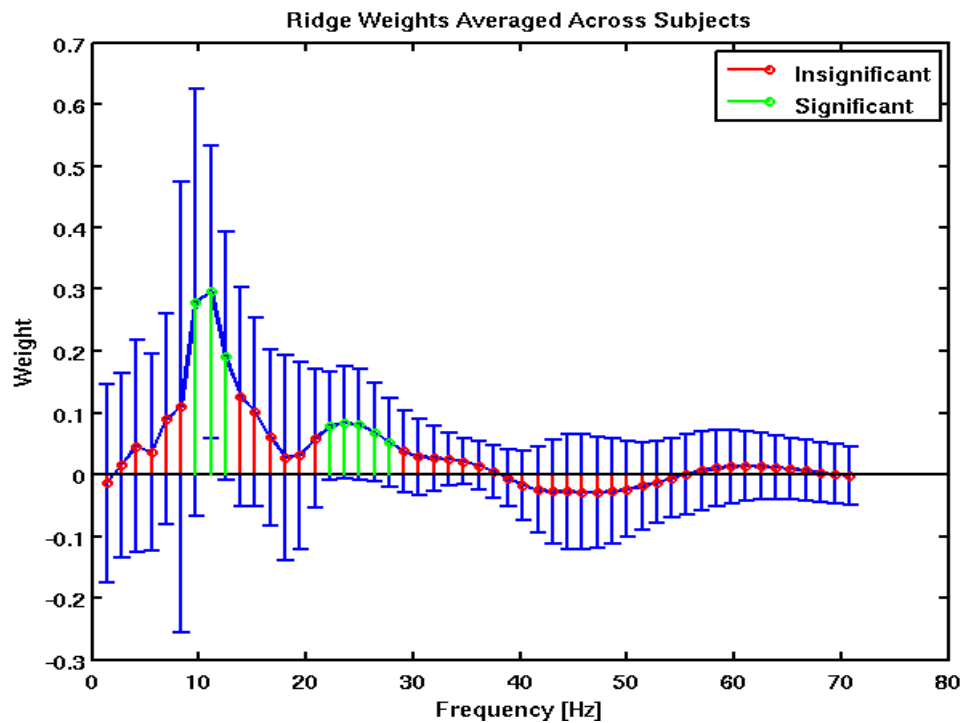*Supplementary figure 7:* **Ridge Regression weights averaged across subjects**



**Figure s7:** Ridge Regression weights averaged across subjects

Weights denoted in green significantly differ from zero (t-test, p<0.05, FDR corrected). Error bars denote 1-std – wide margin around the error estimate computed with cross-validation. The dominant frequency band contributing to the classification is high Alpha (10-15Hz) and Beta (20-30Hz).

The pattern of weight distributions is clearly seen when looking at the average weights across subjects (figure s6). It is evident that the dominant frequency band contributing to the classification is high Alpha (10-15Hz) and Beta (20-30Hz).
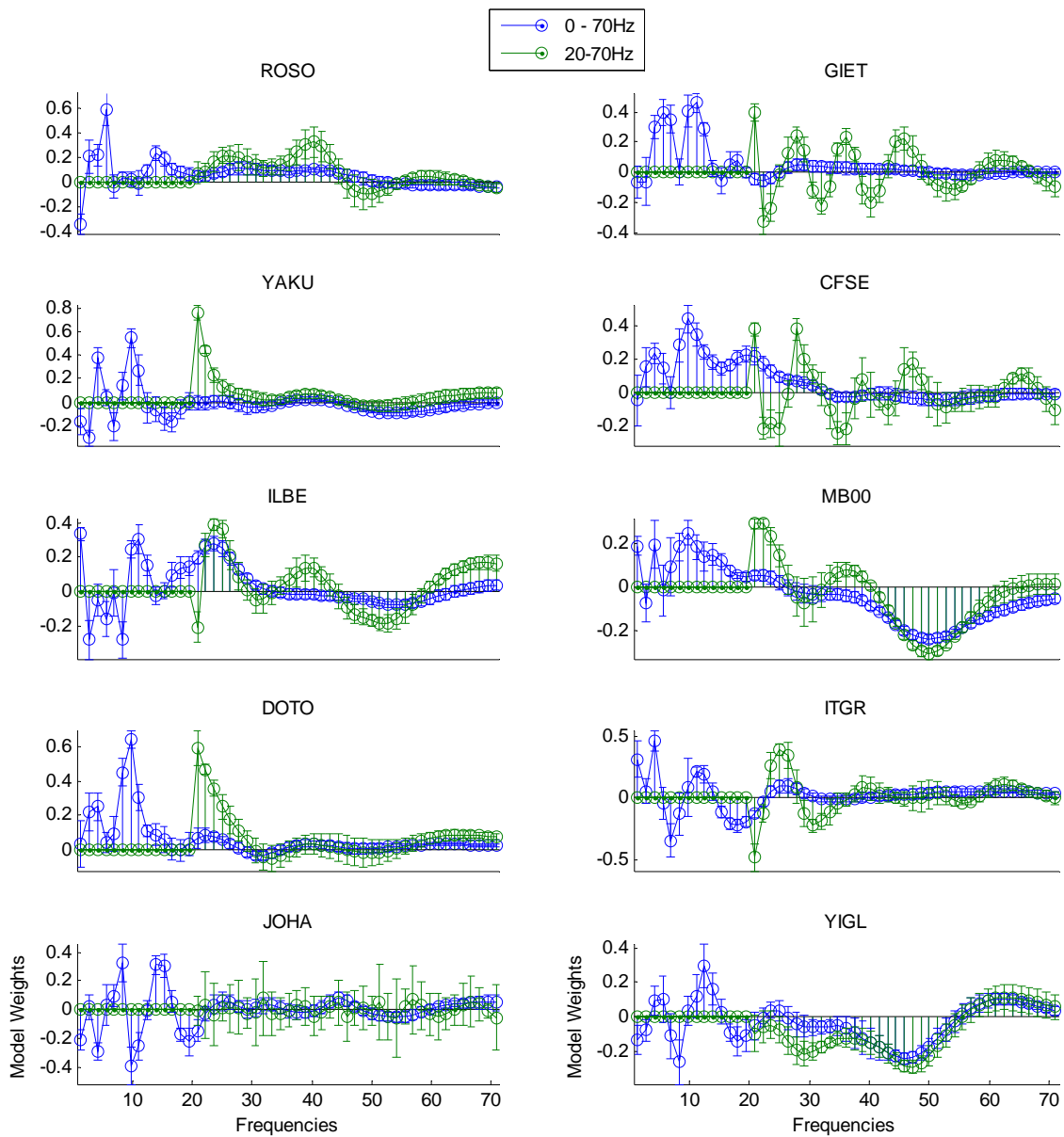
**Figure s8:** Comparison of weights resulting from prediction using all frequencies and prediction using only frequencies above 20Hz.

Subjects are sorted by strength of prediction with frequencies above 20Hz. Note that for subjects with good prediction strength there are large weights for frequencies above 20 Hz for both methods, indicating there is significant information beyond the Alpha band.