

# Multiscale temporal neural dynamics predict performance in a complex sensorimotor task

Wojciech Samek<sup>a,\*</sup>, Duncan A.J. Blythe<sup>b,d</sup>, Gabriel Curio<sup>c,d</sup>, Klaus-Robert Müller<sup>b,e</sup>, Benjamin Blankertz<sup>f</sup>, Vadim V. Nikulin<sup>c,g</sup>

<sup>a</sup>*Machine Learning Group, Fraunhofer Heinrich Hertz Institute, Berlin, Germany*

<sup>b</sup>*Machine Learning Group, Berlin Institute of Technology, Berlin, Germany*

<sup>c</sup>*Neurophysics Group, Department of Neurology, Charité University Medicine, Berlin, Germany*

<sup>d</sup>*Bernstein Center for Computational Neuroscience, Berlin, Germany*

<sup>e</sup>*Department of Brain and Cognitive Engineering, Korea University, Seoul, Republic of Korea*

<sup>f</sup>*Neurotechnology Group, Berlin Institute of Technology, Berlin, Germany*

<sup>g</sup>*Centre for Cognition and Decision Making, National Research University Higher School of Economics, Russian Federation*

---

## Abstract

Ongoing neuronal oscillations are pivotal in brain functioning and are known to influence subjects' performance. This modulation is usually studied on short time scales whilst multiple time scales are rarely considered. In our study we show that Long-Range Temporal Correlations (LRTCs) estimated from the amplitude of EEG oscillations over a range of time-scales predict performance in a complex sensorimotor task, based on Brain-Computer Interfacing (BCI). Our paradigm involved eighty subjects generating covert motor responses to dynamically changing visual cues and thus controlling a computer program through the modulation of neuronal oscillations. The neuronal dynamics were estimated with multichannel EEG. Our results show that: (a) BCI task accuracy may be predicted on the basis of LRTCs measured during the preceding training session, and (b) this result is spatially specific, since we find that LRTCs measured over the sensorimotor areas were most predictive of the task performance. Our results provide direct empirical evidence in addition to previous theoretical work suggesting that scale-free neuronal dynamics are important for optimal brain functioning.

---

## 1. Introduction

While many neuronal operations occur over specific and often limited time scales, there is growing consensus that in order to develop a complete picture of neuronal functioning, brain activity should be viewed simultaneously over a range of time scales [1, 2, 3, 4, 5], an approach which is justified by recent work suggesting that the brain operates at a "critical" state. The theory of criticality implies that neuronal firing occurs in the form of avalanches [6] consisting of bursts of simultaneous neuronal firing of variable size and duration. The duration and the size of such avalanches at criticality follow a power law distribution [1], implying scale free behavior of neuronal activation which results from the system being poised at a phase transition. In the case of neural systems, its position close to a phase transition implies a balance between excitation and inhibition [7]. A number of theoretical and empirical studies indicate that the functional

---

<sup>\*</sup>WS and DAJB contributed equally to this work

<sup>\*</sup>Corresponding author at: Fraunhofer Heinrich Hertz Institute, Machine Learning Group, Einsteinufer 37, 10587 Berlin, Germany. E-mail address: wojciech.samek@hhi.fraunhofer.de. Telephone/Fax number: +49(0)30 31002 417/190.

implications of such a critical state include maximizing dynamic range [8, 9], information transfer [5] and information capacity [5] in brain functioning.

Criticality manifests itself not only in the presence of aperiodic avalanches but also in slowly decaying autocorrelations in the amplitude dynamics of neuronal oscillations [7, 10, 11]. Correlations such as these are often referred to as Long-Range Temporal Correlations (LRTCs) with the corresponding spectra of the amplitude envelopes having  $1/f^\alpha$  shape. Importantly, avalanche dynamics and LRTCs in neuronal oscillations have been found to occur simultaneously [7]; this connection between LRTCs in the amplitude envelopes of oscillations and critical dynamics, coupled with the fact that such dynamics have been shown to be important for optimal neuronal processing, suggests that *measuring* LRTCs of neuronal oscillations (such as alpha/ mu oscillations) could prove to be an important measure of how close the networks under study are to the "theoretical optimum" at criticality. These considerations motivated our empirical study of the relationship between LRTCs estimated from the amplitude envelopes of oscillations and task performance. The physiological implications of LRTCs were elucidated by [7] where it was shown that LRTCs and critical dynamics arise simultaneously in networks in which excitation and inhibition are balanced; such a balance causes oscillations to arise spontaneously, but not to spiral out of control and likewise not to be completely extinguished by inhibitory neurons—this leads to optimal signal propagation through the network as shown by [8, 9, 5].

As a measure of multi-scale dynamics LRTCs are particularly attractive for the study of the human brain. They can be assessed with non-invasive recordings such as EEG and MEG. Due to the continuous nature of EEG/MEG, a separation of the time-series into avalanches and inactivity is often not feasible [12], whereas LRTCs may be estimated directly from continuous time-series. Interesting is that LRTCs have been observed not only in electrophysiological recordings but also in behavioral measures such as reaction times [13]. Moreover, recent studies have shown that LRTCs in the amplitude dynamics of neuronal oscillations correlate with LRTCs in motor tapping [14] and with LRTCs in perceptual hit rate [15]. However, given the predictions from previous studies regarding the relevance of power-law dynamics for brain functioning, the most important question remains open: whether LRTCs in neuronal oscillations directly predict *cognitive* performance and whether this predictive power extends to activities involving distributed brain areas.

In the present study we used BCI on the basis of motor imagery (MI) which is known to effectively modulate sensorimotor oscillations. MI is defined as mental simulation of an action without the activation of effectors [16, 17, 18]. In the context of BCI subjects perform specific MI depending on the type of visual stimuli, requiring them to move the cursor to a designated part of the screen. Thus performance in BCI depends on complex neuronal activations including processing of the visual stimuli, selection and generation of the appropriate MI, and the following assessment of the intended result according to the actual position of the cursor on the screen. While BCI commands are conventionally generated on the basis of activity lasting a few hundreds of milliseconds [19], it remains open whether cortical states, extending over time-scales up to tens of seconds, may be helpful in the prediction of BCI performance. In order to address this question we use LRTCs, which are consistent with the presence of criticality in neuronal dynamics, in order to predict subjects' performance in sensorimotor BCI.

## 2. Materials and Methods

### 2.1. Data set

We study the Vital BCI data set [20], which comprises the EEG recordings of 80 subjects performing motor imagery with the left hand, right hand or feet. Motor imagery is defined as the mental simulation of a specific action, e.g., the kinesthetic imagination of a movement with the left hand, without any corresponding motor output [21]. All participants (41 female, age  $29.9 \pm 11.5$  years; 4 left-handed) were BCI-novices and had not been diagnosed at the time of the experiment with any neurological diseases. The experiment consisted of a training and a subsequent feedback session which were recorded on the same day with a break of about 10-15 minutes. The experimental procedures were approved by the local ethics committee and all subjects gave informed consent.

Participants were positioned seated in a comfortable chair with arms on armrests. Electroencephalography (EEG) signals were recorded from the scalp using an extended 10-20 electrode system with 119 Ag/AgCl electrodes (EasyCap, Munich, Germany) and multi-channel amplifiers (BrainAmp DC by Brain Products, Munich, Germany). The sampling rate was set to 1000 Hz with a band-pass filter of 0.05 Hz to 200 Hz and the reference was at nasion. The BCI experiment started with one run of real movements and three training runs of imagined movements, each containing 25 trials each of: left hand, right hand, and right foot or feet movements. Each trial lasted 8 seconds with a baseline period of 2 sec, a 4 sec motor imagery period and a 2 sec rest period. The motor task to perform was indicated by visual cues being arrows pointing left, right or down. After every 20 trials there was a break of 15 sec. In total 225 motor imagery trials were recorded during the training session indicating left, right or foot imagery, respectively. Figure 1 summarizes the experimental protocol.

After training the two motor imagery classes yielding highest classification accuracy were selected using cross-validation [19]. Out of 80 subjects analyzed, the optimal pair chosen was left/ right hand for 30 subjects, left hand/ foot for 34 subjects and right hand/ foot for 16 subjects. Subsequently three online BCI feedback runs, each of 100 trials, were performed with these two classes. For some participants only one or two runs were recorded due to fatigue ( $N = 3$ ) or lack of time ( $N = 7$ ). A bias adaptation technique [22] based on the first 20 trials was adopted. During feedback, as in the training stage, each trial began with a baseline period of 2 sec where a black fixation cross was presented in the center of a grey screen. The arrow indicating the target motor imagery class, i.e., the type of movement (left, right hand or feet) the user was asked to imagine in this particular trial, appeared for 1 sec; then the cross turned purple and moved according to the classifier output. After 4 sec of cursor movement the cross froze at the final position and turned black again. Two seconds later the cross was reset to the central position and the next trial began. The classification score was indicated during the 15 second break between every block of 20 trials.

### 2.2. Preprocessing & BCI classification

The preprocessing steps consist of spectral filtering, extraction of informative time segments and artifact removal. For each of the 80 subjects first we downsampled the signal to 100 Hz and selected 62 electrodes densely covering the sensorimotor cortices. Noisy electrodes were then removed using the method proposed in [19]. The EEG was filtered in a restricted frequency range with a 5th order Butterworth filter and cut into epochs. In order to reduce the impact of EEG artifacts on BCI training, we removed outlier trials [19].

Then we selected subject-specific frequency ranges and time segments by cross-validation on the training data [19].

For dimensionality reduction spatial filtering is performed using Common Spatial Patterns (CSP) [23, 19]. This method is a popular preprocessing approach, because it improves the signal to noise ratio by enhancing the amplitude changes relating to event-related synchronization and desynchronization (ERS/ERD) [24]. These effects are induced by motor imagery and reflect sensorimotor brain activity resulting in an amplitude decrease (ERD) or increase (ERS) of oscillatory components, typically in the alpha or beta frequency range. Applied to band passed signals CSP finds spatial filters that maximize the band power differences between motor imagery classes. Technically, CSP solves the generalized eigenvalue problem

$$\Sigma_1 w_j = \lambda_j \Sigma_2 w_j \quad (1)$$

where  $\Sigma_1$  and  $\Sigma_2$  are the average covariance matrices estimated for the two motor imagery classes.

CSP filters are sorted according to their ability to separate motor imagery conditions. Throughout this paper we refer to the most discriminative CSP component as 1st CSP, to the second most discriminative CSP component as 2nd CSP and so on.

Figure 2 displays typical topographies of the two most discriminative CSP [23] components which define the features on which the final classification is performed. The middle panel of Figure 2 also displays power-spectral densities of CSP components with typical peaks in the alpha ( $\sim 10$  Hz) and beta ( $\sim 20$  Hz) frequency ranges. The lower panel of the figure displays time-resolved event-related desynchronization (ERD) of the amplitude of alpha oscillations during left/right hand motor imagery: note stronger attenuation of the oscillations in the left and right hemispheres for the imagery of right and left hand movements, respectively. These transient ERDs were the basis for the BCI classification [25, 19].

After applying the CSP filters  $W = [w_1 \dots w_d] \in \mathbb{R}^{D \times d}$  to the data, each trial  $i$  was brought into the form of a feature vector:

$$f_i = \log \left( \text{diag} \left[ W^\top X^{(i)} (X^{(i)})^\top W \right] \right) \quad (2)$$

The logarithmic transformation was applied in order to render the features approximately normally distributed [19].  $\text{diag}[\cdot]$  denotes the operator extracting the diagonal of a matrix. In the final step of BCI training a classifier  $y = h(x)$  separating the two motor imagery classes was computed. We used *Linear Discriminant Analysis* (LDA) [22] for this task, which classifies trials  $x$  according to the decision criterion:

$$h(x) = \text{sign}(v^\top x + b) \quad (3)$$

with  $v = \Sigma_f^{-1}(\mu_2 - \mu_1)$  and  $b = -\frac{1}{2}v^\top(\mu_1 + \mu_2)$

where  $\mu_1$  and  $\mu_2$  denote the mean feature vectors of class 1 and 2, respectively,  $\Sigma_f$  stands for the covariance matrix of the features and  $\text{sign}(x)$  is a function with +1 for  $x \geq 0$  and -1 otherwise.

We computed the out-of-sample classification accuracy

$$acc = \frac{\text{number of correctly classified trials}}{\text{total number of trials}} \quad (4)$$

on the online feedback data for each of the 80 subjects recorded in the experiment. More methodological

details are included in the appendix.

### 2.3. Performance predictors

We use signal-to-noise ratio and Hurst exponents, both computed on the training data, to predict the performance of the subjects in the separate online feedback session. Both predictors are computed on the six most discriminative CSP components.

#### 2.3.1. Signal-to-noise ratio estimation

In order to compute the signal-to-noise ratio ( $SNR$ ) in a frequency range  $[f - \Delta, f + \Delta]$  we spatially filter the data by applying a CSP filter and compute the power spectrum  $\mathcal{P}[f]$  with Welch’s method [26] using 2 sec windows. We considered narrow frequency bins of width 5 Hz, with lower limit  $f - \Delta = 1, 2, \dots, 40$  Hz and  $\Delta = 2.5$  Hz. After discarding the 15 sec pause periods and outliers windows (see description in appendix) we compute  $SNR$  as the degree to which a spectral peak at  $f$  stands above the  $1/f^\alpha$  background neural noise, i.e.,

$$SNR(f) = \max \{ \log(\mathcal{P}[f']) - q(f') \mid f' \in [f - \Delta, f + \Delta] \} \quad (5)$$

where  $q(\cdot)$  is the linear polynomial which intercepts  $\log(\mathcal{P}[\cdot])$  at  $f - \Delta$  and  $f + \Delta$ . Large values correspond to pronounced peaks in the spectrum at a given frequency  $f$  (Figure 2 middle panel), while for small  $SNR$  there are usually no peaks in the spectrum<sup>1</sup>

#### 2.3.2. Hurst exponent and their estimation with Detrended Fluctuation Analysis

The Hurst exponent  $H$ , quantifies the fluctuations in a stochastic process. For a stationary process,  $H$  may be defined formally via the auto-covariance function of the process  $X(t)$ :

$$\mathbb{E}(X(t)X(t+s)) - \mathbb{E}(X(t))^2 \sim \frac{1}{s^{2-2H}}, \quad (6)$$

which holds for  $1 > H > \frac{1}{2}$ . In this case the process is said to be *long-range temporally correlated* (LRTC)<sup>2</sup>. Note that there are several Hurst exponents associated with a process  $X(t)$ . These include the Hurst exponent of  $X(t)$  but also the Hurst exponent of time-series obtained from  $X(t)$  by narrowband filtering in some frequency range. It is exactly these *narrowband* exponents which we study in the following: we obtain one exponent per frequency range, using the same frequency ranges as for  $SNR$  estimation.

Figure 3 displays amplitude dynamics of sensorimotor mu oscillations from a typical CSP component over a range of time scales. The upper panel displays several minutes of the training run, on the basis of which  $H$  was calculated. One observes “bursts” of alpha oscillations lasting for a few minutes, although closer inspection at a higher temporal resolution reveals that in fact within these larger bursts, transient bursts occur interleaved with periods of relative quiescence—behaviour which continues towards increasingly higher temporal resolutions (lower panel).

<sup>1</sup>Note that negative values correspond to troughs in the power-spectrum. In practice, however, we find that significant troughs do not occur.

<sup>2</sup>An instance of a process with  $H \in (0, 1)$  is fractional Gaussian noise, which is defined as the increments of fractional Brownian motion [27, 28, 29].

In order to quantify these fluctuations we estimated the Hurst exponent of a time-series by Detrended Fluctuation Analysis (DFA) [30], using a window spanning over a range of time-scales. Using our outlier detection procedure (see appendix) DFA epochs containing outliers and scales with insufficient data were rejected. The advantage over covariance analysis or analysis of the power-spectrum are its robustness to trends contaminating the empirical time-series and its desirable convergence properties [31, 32]. Robustness to trends implies that if the data is contaminated by shifts in voltage then these be ignored by the DFA estimation. The steps involved in DFA [30] are as follows. First one forms the aggregate sum of the empirical time-series  $Y(t)$ : in our case  $Y(t)$  is the amplitude envelope of the narrowband filtered EEG obtained with the Hilbert transform.

$$X(t) = \sum_{i=1}^t Y(i) \quad (7)$$

Analysis of the fluctuations in  $X(t)$  is then performed by measuring the variance of  $X(t)$  in windows of varying size  $n$  after linear detrending, i.e.,  $X(t)$  is split into windows of length  $n$ ,  $X_n^{(1)}, \dots, X_n^{(j)}, \dots, X_n^{(\lfloor N/n \rfloor)}$  and the average variance after subtracting the linear fit to the data  $X_n^{(j)}$  in these windows is formed; i.e. let  $P(X_n^{(j)})$  be the least-squares estimated linear fit of  $X_n^{(j)}$ , then the DFA coefficients or detrended variances are:

$$\begin{aligned} F_{DFA}^2(n) &= \frac{1}{\lfloor N/n \rfloor \cdot n} \sum_j \left( X_n^{(j)} - P(X_n^{(j)}) \right)^\top \left( X_n^{(j)} - P(X_n^{(j)}) \right) \\ &=: \frac{1}{\lfloor N/n \rfloor \cdot n} \sum_j f_{DFA}^2(n, j) \end{aligned} \quad (8)$$

Crucially, in the limit of data the slope of  $\log(F_{DFA}^2(n))$  against  $\log(n)$ , the DFA estimate of the Hurst exponent  $\hat{H}$ , converges to  $H$ . Thus notice that there is a distinction between the estimate  $\hat{H}$  using DFA and  $H$  itself; in the limit of data these converge, however, for finite data, there will always be a difference which depends on the amount of data. For simplicity and due to the absence of any ambiguity in the interpretation, we use simply  $H$  instead of  $\hat{H}$  throughout the paper. In Appendix D.1 we show how  $H$  can be affected by decimation and filtering of the EEG. In Appendix D.2 linearity of the least-squares fit in DFA is assessed. In Appendix D.3 we provide information on the effect of the window length on the estimation of  $H$ .

### 3. Results

#### 3.1. BCI task performance and classification

The mean accuracy computed was  $acc = 72 \pm 2\%$  (standard deviation estimated from 500 bootstrap iterations) (median  $acc = 75 \pm 3\%$ ). In particular, we find that 59 of the 80 subjects performed significantly above chance according to the threshold ( $acc > 56\%$ ) determined by a binomial test. Notice, however, that we include all 80 subjects in the subsequent analysis because we aim at predicting a wide range of individual performances.

The spatial distribution of the topographies correspond to neuronal generators located in the sensorimotor cortices; a grand-average of the individual inverse solutions (weighted minimum norm [33]) of CSP components confirmed the presence of maxima in the left and right sensorimotor cortices for the right and left hand motor imagery, respectively, while in a more symmetric medial location for foot motor imagery

(see Figure 4). The median ranges of the lower and higher frequencies were 9.5 Hz and 13.5 Hz, respectively, indicating that alpha oscillations were most relevant for classification, which was further confirmed by the calculation of biserial- $R^2$  values (Figure 5).

### 3.2. Signal-to-noise ratio

$SNR$  was most pronounced in the alpha frequency range, especially for the first, most discriminative CSP component. For all frequency ranges, we report in Figure 6 the grand-average values for the first and last CSP component out of the six CSP components as well as the average value over all six CSP components (denoted ‘Average CSP’).

### 3.3. Hurst exponent values and statistics

Based on the criterion that a time-scale should display at least 10 epochs of outlier free data, for 99% of the Hurst exponents estimated on the first CSP filter, the largest window size lay between 29 and 55 seconds long. Figure 7 displays typical DFA plots for different CSP components for a representative subject in the alpha/ mu frequency range. Note a highly accurate linear relationship between the length of the window  $n$  (logarithmically scaled on the  $x$ -axis) and the fluctuations measured by the logarithm of DFA coefficients  $\log(F_{DFA}(n))$  (see Equation 8 for the definition of  $F_{DFA}$ ), thus indicating the presence of LRTC extending for at least 50 seconds in the amplitude dynamics of sensorimotor alpha oscillations. We found that the goodness of fit (based on  $R^2$ ) across all subjects was on average  $> 91\%$  and in the alpha/mu frequency range  $>99\%$ . To avoid confusion with other goodness-of-fit measures used in this paper, we refer to the goodness of the linear fit of the log-DFA coefficients with the log-scale as  $DFA-R^2$ . In a manner similar to the  $SNR$  analysis, we calculated the dependency of Hurst exponents on the frequency of neuronal activity. Figure 8 shows that the largest Hurst exponents are detected for alpha and beta oscillations, especially for the most discriminative CSP component. All six CSP components displayed significant LRTC behavior ( $H > 0.5$ ) for all frequencies ( $p \ll 0.001$ ; Wilcoxon sign rank test on exponent values), see Figure 8.

### 3.4. Dependency between signal-to-noise ratio and Hurst exponent values

We computed correlations between signal-to-noise ratio and Hurst exponents ( $SNR$  and  $H$ ). As already noted by [34],  $SNR$  confounds the measurement of Hurst exponents—a higher  $SNR$  results in the measurement of *larger* Hurst exponents than in the presence of higher noise levels. We find indeed that  $H$  and  $SNR$  are positively correlated (see Figure 9), with the most pronounced correlations in the alpha range ( $r = 0.56$ ,  $p \ll 0.001$ ; Student  $t$  approximation to distribution of Spearman’s correlation coefficient). After a Bonferroni correction [35] controlling for multiple testing the correlation is significant ( $p < 0.001$ , number of tests=40 (number of frequency bands)).

### 3.5. Sensorimotor Cortex Hurst exponent values predict task performance

We calculated Spearman correlations between Hurst exponents,  $H$ , and BCI task performance,  $acc$ . This correlation analysis was performed for all frequencies between 1–40 Hz in analogy to the analysis performed for  $H$  and  $SNR$  statistics. For the comparison we measured correlations between  $SNR$  and  $acc$ .

Traces of Hurst exponent values over frequencies for the first CSP component are plotted in Figure 5 as an image, when sorted in an ascending order with respect to BCI performance,  $acc$ . The image clearly shows that higher exponents correspond to improved performance, which is especially visible for oscillations in the alpha

frequency range. Less pronounced but still visible is the peak in the beta frequency range. Interestingly, these peaks resemble the location of the peaks in the biserial- $R^2$  distribution (Figure 5) showing which frequencies were most discriminative for classification (see Appendix C.2 for the definition of biserial- $R^2$  values—these should be distinguished from the DFA- $R^2$  values). This may be explained as follows: since we require variability in  $H$  to predict performance and since a larger mean in a frequency band also coincides with greater variability, then the mean exponent per frequency also coincides with predictive power (i.e. biserial- $R^2$ ). The left panel of Figure 10 displays a scatter-plot between Hurst exponents and BCI performance,  $acc$ , for the first CSP component ( $r = 0.56$ ,  $p < 0.001$ ; frequency band 10-15 Hz). The right panel of Figure 10 displays a corresponding scatter-plot between  $SNR$  and  $acc$  ( $r = 0.52$ ,  $p < 0.001$ ). When training a linear regression model on  $N - 1$  subjects using their  $H$  (or  $SNR$ ) and  $acc$  values and using it to predict the performance of a new subject (leave-one-subject-out approach), we found that  $H$  and  $SNR$  explains 28% and 27% of the variance of  $acc$ , respectively. Even when  $H$  was averaged across frequencies (for the first CSP component), we found it remains predictive of  $acc$  ( $r = 0.45$ ,  $p < 0.001$ ). Finally we found these results to be robust even to removing subjects who did not perform significantly above the chance level, 56% ( $r = 0.35$ ,  $p < 0.01$ ; frequency band 10-15 Hz). We also tested whether the correlation between  $H$  and  $acc$  can be explained in a binary way, i.e. depending on whether a subject performed above chance or not. For that we constructed a binary variable by dividing subjects into two groups on the basis of their performance:  $acc \leq 56\%$  and  $acc > 56\%$  (56% is the significantly-above-chance performance level). The partial correlation between  $H$  and  $acc$  remains significant when regressing out this binary variable ( $r = 0.30$ ,  $p < 0.01$ ; frequency band 10-15 Hz). Also bootstrapping and application of Wilcoxon signed-rank test confirmed that the correlation between continuous  $acc$  values and  $H$  is significantly larger ( $p \ll 0.001$ ) than the correlation between  $H$  and the binary variable (the latter calculations were based on point-biserial correlation).

A comprehensive overview of the correlation between  $H$  and  $acc$  by frequency bin is displayed in the first column of Figure 11. We found that correlations peak in the mu/alpha range are particularly pronounced for the first and average CSP components, where the correlations are significant even after a very conservative Bonferroni correction. We found that the peak is negligible for the last (i.e. sixth most discriminative) CSP component.

The correlation between  $SNR$  and  $acc$  by frequency is displayed in the second column of Figure 11. The figure demonstrates that  $SNR$  values are also predictive of BCI performance in the alpha frequency range. Likewise, we found that there is a significant correlation between average  $SNR$  over frequencies and  $acc$  ( $r = 0.32$ ,  $p < 0.01$ ); however the correlation is much weaker than the correlation between  $H$  and  $acc$ .

### 3.6. Hurst exponent predictions cannot be explained in terms of signal-to-noise ratio

Since we observed that  $SNR$  and Hurst exponents are cross-correlated, and that  $SNR$  also predicts performance, we used partial correlations to determine whether the predictive power of  $H$  is independent of the predictive power of  $SNR$  values; see appendix for a review of partial correlation analysis.

We found that even after controlling for  $SNR$ , the predictive power of Hurst exponents of BCI performance is significant. For the average  $H$  over frequencies on the first CSP component we found that after controlling for the effect of the average  $SNR$ , a significant correlation remains ( $r = 0.42$ ,  $p < 0.001$ ). The same analysis performed for  $SNR$ , controlling for  $H$  yields a much weaker correlation ( $r = 0.24$ ,  $p < 0.05$ ).

Partial correlations for individual frequency bands are displayed in the third and fourth columns of Figure 11. In order to rigorously check whether the correlations between frequency-wise  $H$  values and  $acc$  do not depend on  $SNR$ , we controlled not only for  $SNR$  and  $H$  values in the same frequency range (for each bin in the range 1–40 Hz) but also for separate frequencies. To do this for each pair of frequency bins  $(i, j)$  for  $i, j \in \{1 \text{ Hz}, 2 \text{ Hz}, \dots, 40 \text{ Hz}\}$ , we calculated a partial correlation

$$r_{i,j} = \text{corr}(H_i, acc | SNR_j) \quad (9)$$

The most stringent control for correlation between  $H_i$  and  $acc$  is given by  $r_i^* = \min_j \{r_{i,j}\}$ .  $r_i^*$  may be thought of as the lowest partial correlation which results from controlling for the potentially confounding effect of  $SNR$  in all frequencies.

Having performed this analysis for the first CSP filter we found that the largest worst case partial correlation for individual frequency bands is nevertheless highly significant ( $r^* = 0.37$ ,  $p < 0.001$ ). After a Bonferroni correction controlling for multiple testing this result is still significant ( $p = 0.005$ ,  $n = 40$ , number of frequency bands). An analogous result also holds for the average  $H$  across all CSP components. On the other hand, after controlling for  $H$ , we found that after multiple testing correction  $SNR$  over frequencies is not significantly predictive of performance for any of the CSP components considered ( $p > 0.05$ ).

To make sure that the correlations are not due to a binary effect we performed similar analysis on the subjects who performed significantly above the chance level ( $acc > 56\%$ ). Figure 12 shows the partial correlation between BCI classification accuracy and  $H$  or signal-to-noise ratio with the effects of  $SNR$  or  $H$  respectively being removed. As before after controlling for  $SNR$ , the Hurst exponent is significantly (after multiple testing correction,  $p < 0.05$ ) predictive of the performance; this is not the case for  $SNR$ . This result not only holds for the average CSP, but also for the first CSP component.

## 4. Discussion

Using multichannel EEG, we showed that the performance in a complex sensorimotor task can be predicted on the basis of the neuronal dynamics of neuronal oscillations. This prediction is frequency specific and involves oscillations generated in the Rolandic brain area. Below we discuss the robustness of the results obtained and provide a tentative neurophysiological explanation for this predictive power, emphasizing the role of power-law dynamics in neuronal information processing.

### 4.1. Task performance and classifiers

We trained a classifier for each of 80 subjects, obtaining a median accuracy of 75%, which corresponds closely to the performance of state-of-the-art classification procedures for naïve BCI subjects [20]. The classifiers were based on well-tested standards for sensorimotor BCI [36] and included optimization of spatial filters, frequency ranges and feature selection. CSP patterns could be modeled with sources located in the sensorimotor areas of the cortex, which is in agreement with previous BCI studies [37, 38] and demonstrated a somatotopy relating to hand and foot activations. While LRTCs were systematically calculated for all frequency ranges, we found that Hurst exponents in the alpha-frequency range have the highest predictive power, with the second best frequency window in the beta range; this finding relates to the central role of  $\mu$  oscillations in sensorimotor processing [39, 40, 41, 14]. Note that, although for the evaluation of LRTCs

we used data from the training condition, requiring subjects to perform motor imagery, the temporal scales of DFA analysis far exceeded (Figure 7) the time scale over which the transient modulation of oscillations (4 seconds) due to the performance of motor-imagery occurred, with DFA windows extending up to 55 seconds. Such estimation of LRTCs during task conditions has also been performed in other studies [10, 42] leading to the insight that transient modulation of neuronal dynamics does not abolish LRTCs. In fact, task-related performance may even allow subjects to remain more vigilant, preventing them from drowsiness, which is a risk factor during purely resting-state conditions [43].

#### 4.2. Task prediction on the basis of SNR and LRTCs

We showed that the predictive power of Hurst exponents cannot be explained by the confounding effect of *SNR*. The authors of [34] showed that the estimation of Hurst exponents may be biased by *SNR*: smaller *SNR* leads to an attenuation of exponents, moving estimates closer to the 0.5 value characteristic of white noise. Thus it might be argued that Hurst exponents are predictive of the BCI performance only due to their correlation with *SNR*, which has been shown to predict BCI performance [20, 44]. As an estimator for *SNR* we used the amplitude of the spectral peaks relative to the  $1/f^\alpha$  noise floor. Importantly, however, such a prediction does not carry neurophysiological information since it may be reduced to how well neuronal oscillations may be extracted from the background neuronal noise. Figure 11 in fact shows that when SNR from all CSP components is taken into account (third column, third row), the task accuracy can still be predicted from the Hurst exponents, while when the Hurst exponents from all CSP components are taken into account, no prediction could be obtained between SNR and task accuracy (third column, fourth row). This provides a strong evidence for the task-performance prediction not being explained in terms of SNR. This furthermore implies that the predictive power gained by the Hurst exponents cannot be simply due to the signal quality on the scalp and that the exponents rather reflect neurophysiological aspects of dynamic neuronal activity which predict performance.

#### 4.3. Neuronal oscillations predict task performance

A number of previous studies showed that the performance in different tasks might relate to the amplitude [45, 46] or phase [47] of ongoing oscillations immediately preceding the stimulus requiring behavioral response. The usual interpretation of such dependencies is based on changes of cortical excitability or in general on a neuronal propensity to respond given a neuronal state directly preceding the stimulus. In the present study, however, we were interested instead in how the task performance can be predicted from the neuronal dynamics recorded in a *separate* experimental session, thus relating performance with subject-specific persistent features of neuronal activation. BCI systems, which use sensorimotor oscillations, require the monitoring of visual stimuli and continuous generation of covert motor activity depending on the position of the cursor on the screen. An advantage of the BCI paradigm is that it allows for the study of motor behavior without the confounding factors associated with overt movements, such as proprioceptive feedback, which can strongly modulate ongoing neuronal oscillations [40]. Successful BCI performance is thus primarily based on the ability of subjects to quickly generate the appropriate modulation of ongoing oscillations in response to the visual cues. Thus, in the case of BCI, performance is directly defined by very specific neuronal events, which, in our case, correspond to event-related desynchronization (ERD) of the sensorimotor rhythms [36, 48]. This in turn allows us to be specific in interpreting the factors defining performance. ERD of sensorimotor oscillations has long been associated with the active participation of a given cortical area in the task [24],

while an increase in the amplitude of oscillations (event-related synchronization) is thought to correspond to active inhibition [49, 46]. Preventing a cursor moving in one direction and immediately moving it in another direction requires swift switching between ERD in one somatotopic cortical location to the subsequent generation of an ERD response in another location. In neurophysiological terms this translates to rapidly alternating states of cortical excitation and inhibition leading to a selective modulation of neuronal activity. Such coordination between excitation and inhibition is likely to benefit from the underlying neuronal networks moving fluently from one state to the other. Interestingly, it has recently been shown in a computational study [7], that stronger LRTCs in neuronal oscillations may relate to a balance between excitation and inhibition. Thus stronger LRTCs may represent a suitable background condition for the rapid modulation of neuronal oscillations. This in turn provides a possible explanation for the positive correlation between LRTCs and BCI performance which requires rapid switching between the cortical states.

#### *4.4. Implications for BCI Practitioners*

For BCI researchers the question arises how to make use of the results presented in this paper in practice. Since Hurst exponent estimates were shown to correlate with performance, one possible practical application would be to use these predictors for subject pre-screening. This would save time as poorly performing subjects would be identified and excluded from the experiment and a different experimental paradigm could be used in their case. Another possible field of application of the presented results is the improvement of the BCI signal processing methods. For instance, one might apply the Hurst exponent for selecting the number of spatial filters by only selecting filters with  $H$  values larger than a predefined threshold. Alternatively the CSP algorithm could be regularized towards sources exhibiting large  $H$  values, similar to the approach in [50] where CSP was regularized towards stationarity. Hurst exponents could also be used for selecting the most informative frequency band for each subject. In addition to these technical applications, Hurst exponents combined with neurofeedback techniques could be used to improve the subjects capability to perform BCI. More precisely, on the basis of our findings we conjecture that if subjects were to learn how to optimize their neuronal states leading to increased Hurst exponents, then they could potentially improve in control of a BCI.

#### *4.5. Effect Size*

The strength of the correlation reflects the effect size and in our study even after the most stringent control for the possible confounding effect of SNR, the strength of the correlation between  $H$  and the task performance was at least 0.37. The interpretation of the effect size may vary from discipline to discipline, but it is not uncommon to refer to the correlation of 0.3 or higher as moderate [51]. Given the non-invasive character of the recordings, the correlation values obtained in the present study are not insubstantial. Clearly temporal dynamics are not the only factor determining the task performance; however, the obtained correlation establishes a link between LRTCs in neuronal oscillations and task performance.

#### *4.6. Are dynamics genuinely scale-free?*

The time-scales studied in this paper ranged between 4 seconds and 1 minute. On this basis alone we may not be able to claim that the dynamics studied are scale-free. Due to the nature of the task, we were unable to use longer time-scales in all subjects, because of the occasional outliers in EEG activity induced by task performance. However, in Appendix D.3, we in addition estimated Hurst exponents in the range

0.5–81 seconds (in subjects where this data was available) and found that they were strongly correlated ( $r = 0.90$ ) with the exponents estimated in the 4–54 seconds range. The range 0.5–81 seconds covers two orders of magnitude reflecting a similar range of magnitudes reported for alpha oscillations elsewhere in the literature (for example [52]). However, as we show in Appendix D.1, using very short windows may introduce a small positive bias in the estimation of Hurst exponents and that is why we rather considered time windows starting from 4 seconds. Moreover, it is known from studies at rest that alpha/ mu oscillations are scale-free over at least two orders of magnitude [14, 15] and remain so even during sensory stimulation [53]. Discussion of the limitations of finite-size data for the estimation of Hurst exponents can also be found in other studies [54, 55]. While acknowledging these limitations, the authors in the abovementioned studies argue that shorter time segments may allow the extrapolation of neuronal dynamics over longer scales. While unequivocal demonstration of the scale-free dynamics requires very long time series, DFA is known to reliably estimate scaling exponents when using even short data segments [55].

Moreover in Appendix D.2 we analyse whether the scaling analysed with DFA is scale-free or non-linear over the range of magnitudes studied. We find that there is a small systematic departure from genuine monofractal scaling. However the increase in DFA- $R^2$  is less than 0.005, thus not yielding a large increase in explanatory power.

#### 4.7. Task performance and power-law dynamics

Although the presence of LRTCs does not constitute conclusive proof of the existence of criticality in a neural system, it is nonetheless consistent with this hypothesis [10, 7]. The main finding of our study is that LRTCs are advantageous for brain function, as has been conjectured, for critical dynamics. Previous research has revealed that Hurst exponents in the amplitude dynamics of ongoing oscillations correlate with Hurst exponents in motor tapping [14] and with Hurst exponents of subliminal-stimulus detection [15]. These studies, however, did not provide *direct* evidence that the presence of LRTCs predicts task performance as shown in our study. Indeed, it is exactly performance itself not its variability which is most important for understanding the functional significance of LRTCs. In general our work contributes to growing evidence that power-law neuronal dynamics might be beneficial for brain functioning as shown previously for dynamic stimulus range [8, 9], information capacity [5] and information transmission [5]. The fact, that we could predict task performance from LRTCs determined during the training session, indicates that power-law dynamics are subject-specific. This is in line with findings that LRTCs are genetically defined [54] and have significant test-retest reliability [56] showing stability of individual LRTCs across many days. In addition, the existence of subject-specific LRTCs is further confirmed by demonstrating that LRTCs are sensitive biomarkers of pathological neuronal activity in a variety of diseases such as Schizophrenia [57], Parkinson’s [58] and Alzheimer Disease [59].

In conclusion, we have shown that power-law dynamics in the amplitude of neuronal oscillations, estimated collectively over a range of time scales, predict subjects’ performance in a complex sensorimotor task. This finding provides a direct demonstration for the functional relevance of scale-free critical organization of neuronal activation in the human brain.

#### Acknowledgement

This work was supported in part by the Federal Ministry of Education and Research (BMBF) under the projects Adaptive BCI (FKZ 01GQ1115) and Berlin Big Data Center BBDC (01IS14013A), in part by the

German Research Foundation (GRK 1589/1), the Berlin Bernstein Center for Computational Neuroscience (BCCN 01GQ 1001C), by the Brain Korea 21 Plus Program through the National Research Foundation of Korea funded by the Ministry of Education, and by the Russian Academic Excellence Project ‘5-100’. This publication only reflects the authors views. Funding agencies are not liable for any use that may be made of the information contained herein.

## References

- [1] J. M. Beggs, D. Plenz, Neuronal avalanches in neocortical circuits, *The Journal of neuroscience* 23 (35) (2003) 11167–11177.
- [2] D. R. Chialvo, Psychophysics: Are our senses critical?, *Nature physics* 2 (5) (2006) 301–302.
- [3] R. Barbieri, M. Shimono, Criticality in large-scale brain fmri dynamics unveiled by a novel point process analysis, *Networking of Psychophysics, Psychology and Neurophysiology* (2012) 61.
- [4] T. Petermann, T. C. Thiagarajan, M. A. Lebedev, M. A. Nicolelis, D. R. Chialvo, D. Plenz, Spontaneous cortical activity in awake monkeys composed of neuronal avalanches, *Proceedings of the National Academy of Sciences* 106 (37) (2009) 15921–15926.
- [5] W. L. Shew, H. Yang, S. Yu, R. Roy, D. Plenz, Information capacity and transmission are maximized in balanced cortical networks with neuronal avalanches, *The Journal of neuroscience* 31 (1) (2011) 55–63.
- [6] P. Bak, C. Tang, K. Wiesenfeld, Self-organized criticality: An explanation of the 1/f noise, *Physical review letters* 59 (4) (1987) 381.
- [7] S.-S. Poil, R. Hardstone, H. D. Mansvelder, K. Linkenkaer-Hansen, Critical-state dynamics of avalanches and oscillations jointly emerge from balanced excitation/inhibition in neuronal networks, *The Journal of Neuroscience* 32 (29) (2012) 9817–9823.
- [8] O. Kinouchi, M. Copelli, Optimal dynamical range of excitable networks at criticality, *Nature physics* 2 (5) (2006) 348–351.
- [9] W. L. Shew, H. Yang, T. Petermann, R. Roy, D. Plenz, Neuronal avalanches imply maximum dynamic range in cortical networks at criticality, *The Journal of Neuroscience* 29 (49) (2009) 15595–15600.
- [10] K. Linkenkaer-Hansen, V. V. Nikouline, J. M. Palva, R. J. Ilmoniemi, Long-range temporal correlations and scaling behavior in human brain oscillations, *The Journal of neuroscience* 21 (4) (2001) 1370–1377.
- [11] A. Zhigalov, G. Arnulfo, L. Nobili, S. Palva, J. M. Palva, Relationship of fast-and slow-timescale neuronal dynamics in human meg and seeg, *The Journal of Neuroscience* 35 (13) (2015) 5385–5396.
- [12] D. A. Blythe, V. V. Nikulin, Meta-universality classes at criticality, *arXiv preprint arXiv:1502.07885*.
- [13] C. T. Kello, G. D. Brown, R. Ferrer-i Cancho, J. G. Holden, K. Linkenkaer-Hansen, T. Rhodes, G. C. Van Orden, Scaling laws in cognitive sciences, *Trends in cognitive sciences* 14 (5) (2010) 223–232.
- [14] D. J. Smit, K. Linkenkaer-Hansen, E. J. de Geus, Long-range temporal correlations in resting-state alpha oscillations predict human timing-error dynamics, *The Journal of Neuroscience* 33 (27) (2013) 11212–11220.

- [15] J. M. Palva, A. Zhigalov, J. Hirvonen, O. Korhonen, K. Linkenkaer-Hansen, S. Palva, Neuronal long-range temporal correlations and avalanche dynamics are correlated with behavioral scaling laws, *Proceedings of the National Academy of Sciences* 110 (9) (2013) 3585–3590.
- [16] J. Decety, Do imagined and executed actions share the same neural substrate?, *Cognitive brain research* 3 (2) (1996) 87–93.
- [17] M. Jeannerod, The representing brain: Neural correlates of motor intention and imagery, *Behavioral and Brain sciences* 17 (02) (1994) 187–202.
- [18] M. Jeannerod, Neural simulation of action: a unifying mechanism for motor cognition, *Neuroimage* 14 (1) (2001) S103–S109.
- [19] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K.-R. Müller, Optimizing spatial filters for robust EEG single-trial analysis, *IEEE Signal Processing Magazine* 25 (1) (2008) 41–56.
- [20] B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K.-R. Müller, G. Curio, T. Dickhaus, Neurophysiological predictor of SMR-based BCI performance, *NeuroImage* 51 (4) (2010) 1303–1309.
- [21] A. Guillot, C. Collet, V. A. Nguyen, F. Malouin, C. Richards, J. Doyon, Brain activity during visual versus kinesthetic imagery: an fmri study, *Human brain mapping* 30 (7) (2009) 2157–2172.
- [22] C. Vidaurre, M. Kawanabe, P. von Büna, B. Blankertz, K.-R. Müller, Toward unsupervised adaptation of LDA for brain-computer interfaces, *IEEE Transactions on Biomedical Engineering* 58 (3) (2011) 587–597.
- [23] H. Ramoser, J. Müller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, *IEEE Transactions on Rehabilitation Engineering* 8 (4) (1998) 441–446.
- [24] G. Pfurtscheller, F. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, *Clinical Neurophysiology* 110 (11) (1999) 1842–1857.
- [25] S. Lemm, B. Blankertz, T. Dickhaus, K.-R. Müller, Introduction to machine learning for brain imaging, *NeuroImage* 56 (2) (2011) 387–399.
- [26] P. Welch, The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms, *IEEE Transactions on audio and electroacoustics* (1967) 70–73.
- [27] B. B. Mandelbrot, J. W. Van Ness, Fractional brownian motions, fractional noises and applications, *SIAM review* 10 (4) (1968) 422–437.
- [28] A. Eke, P. Herman, J. Bassingthwaite, G. Raymond, D. Percival, M. Cannon, I. Balla, C. Ikrényi, Physiological time series: distinguishing fractal noises from motions, *Pflügers Archiv* 439 (4) (2000) 403–415.
- [29] A. Eke, P. Herman, L. Kocsis, L. Kozak, Fractal characterization of complexity in temporal physiological signals, *Physiological measurement* 23 (1) (2002) R1.
- [30] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, A. L. Goldberger, Mosaic organization of dna nucleotides, *Physical Review E* 49 (2) (1994) 1685.
- [31] J.-M. Bardet, I. Kammoun, Asymptotic properties of the detrended fluctuation analysis of long-range-dependent processes, *Information Theory, IEEE Transactions on* 54 (5) (2008) 2041–2052.
- [32] M. S. Taqqu, V. Teverovsky, W. Willinger, Estimators for long-range dependence: an empirical study, *Fractals* 3 (04) (1995) 785–798.

- [33] A. Ioannides, J. Bolton, C. Clarke, Continuous probabilistic solutions to the biomagnetic inverse problem, *Inverse Problems* 6 (4) (1990) 523.
- [34] D. A. Blythe, S. Haufe, K.-R. Müller, V. V. Nikulin, The effect of linear mixing in the eeg on hurst exponent estimation, *NeuroImage* 99 (2014) 377–387.
- [35] J. M. Bland, D. G. Altman, Multiple significance tests: the bonferroni method, *BMJ* 310 (6973) (1995) 170.
- [36] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, G. Curio, The non-invasive berlin brain–computer interface: fast acquisition of effective performance in untrained subjects, *NeuroImage* 37 (2) (2007) 539–550.
- [37] S. Haufe, R. Tomioka, T. Dickhaus, C. Sannelli, B. Blankertz, G. Nolte, K.-R. Müller, Localization of class-related mu-rhythm desynchronization in motor imagery based brain-computer interface sessions, in: 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2010, pp. 5137–5140.
- [38] S. Haufe, R. Tomioka, T. Dickhaus, C. Sannelli, B. Blankertz, G. Nolte, K.-R. Müller, Large-scale eeg/meg source localization with spatial flexibility, *NeuroImage* 54 (2) (2011) 851–859.
- [39] G. Pfurtscheller, A. Berghold, Patterns of cortical activation during planning of voluntary movement, *Electroencephalography and clinical neurophysiology* 72 (3) (1989) 250–258.
- [40] V. V. Nikulin, F. U. Hohlefeld, A. M. Jacobs, G. Curio, Quasi-movements: A novel motor–cognitive phenomenon, *Neuropsychologia* 46 (2) (2008) 727–742.
- [41] S. R. Jones, C. E. Kerr, Q. Wan, D. L. Pritchett, M. Hämäläinen, C. I. Moore, Cued spatial attention drives functionally relevant modulation of the mu rhythm in primary somatosensory cortex, *The Journal of Neuroscience* 30 (41) (2010) 13760–13765.
- [42] F. U. Hohlefeld, F. Ehlen, L. K. Krugel, A. A. Kühn, G. Curio, F. Klostermann, V. V. Nikulin, Modulation of cortical neural dynamics during thalamic deep brain stimulation in patients with essential tremor, *Neuroreport* 24 (13) (2013) 751–756.
- [43] B. A. Diaz, S. Van Der Sluis, S. Moens, J. S. Benjamins, F. Migliorati, D. Stoffers, A. Den Braber, S.-S. Poil, R. Hardstone, D. Van’t Ent, et al., The amsterdam resting-state questionnaire reveals multiple phenotypes of resting-state cognition, *Frontiers in human neuroscience* 7 (2013) 446.
- [44] H.-I. Suk, S. Fazli, J. Mehnert, K.-R. Müller, S.-W. Lee, Predicting bci subject performance using probabilistic spatio-temporal filters, *PloS one* 9 (2) (2014) e87056.
- [45] R. Schubert, S. Haufe, F. Blankenburg, A. Villringer, G. Curio, Now you’ll feel it, now you won’t: Eeg rhythms predict the effectiveness of perceptual masking, *Journal of cognitive neuroscience* 21 (12) (2009) 2407–2419.
- [46] S. Haegens, L. Luther, O. Jensen, Somatosensory anticipatory alpha activity increases to suppress distracting input, *Journal of Cognitive Neuroscience* 24 (3) (2012) 677–685.
- [47] N. A. Busch, R. VanRullen, Spontaneous eeg oscillations reveal periodic sampling of visual attention, *Proceedings of the National Academy of Sciences* 107 (37) (2010) 16048–16053.
- [48] G. Pfurtscheller, C. Neuper, C. Guger, W. Harkam, H. Ramoser, A. Schlogl, B. Obermaier, M. Pregenzer, et al., Current trends in graz brain-computer interface (bci) research, *IEEE Transactions on Rehabilitation Engineering* 8 (2) (2000) 216–219.

- [49] S. Palva, J. M. Palva, New vistas for  $\alpha$ -frequency band oscillations, *Trends in Neurosciences* 30 (4) (2007) 150–158.
- [50] W. Samek, C. Vidaurre, K.-R. Müller, M. Kawanabe, Stationary common spatial patterns for brain-computer interfacing, *Journal of Neural Engineering* 9 (2) (2012) 026013.
- [51] J. Cohen, *Statistical power analysis for the behavioral sciences.*, Lawrence Erlbaum Associates, 1988.
- [52] R. Ton, A. Daffertshofer, Model selection for identifying power-law scaling, *NeuroImageIn press*.
- [53] K. Linkenkaer-Hansen, V. V. Nikulin, J. M. Palva, K. Kaila, R. J. Ilmoniemi, Stimulus-induced change in long-range temporal correlations and scaling behaviour of sensorimotor oscillations, *European Journal of Neuroscience* 19 (1) (2004) 203–218.
- [54] K. Linkenkaer-Hansen, D. J. Smit, A. Barkil, T. E. van Beijsterveldt, A. B. Brussaard, D. I. Boomsma, A. van Ooyen, E. J. De Geus, Genetic contributions to long-range temporal correlations in ongoing oscillations, *The Journal of Neuroscience* 27 (50) (2007) 13882–13889.
- [55] E. Tagliazucchi, F. von Wegner, A. Morzelewski, V. Brodbeck, K. Jahnke, H. Laufs, Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep, *Proceedings of the National Academy of Sciences* 110 (38) (2013) 15419–15424.
- [56] V. V. Nikulin, T. Brismar, Long-range temporal correlations in alpha and beta oscillations: effect of arousal level and test–retest reliability, *Clinical neurophysiology* 115 (8) (2004) 1896–1908.
- [57] V. V. Nikulin, E. G. Jönsson, T. Brismar, Attenuation of long-range temporal correlations in the amplitude dynamics of alpha and beta neuronal oscillations in patients with schizophrenia, *Neuroimage* 61 (1) (2012) 162–169.
- [58] F. Hohlefeld, J. Huebl, C. Huchzermeyer, G.-H. Schneider, T. Schönecker, A. Kühn, G. Curio, V. Nikulin, Long-range temporal correlations in the subthalamic nucleus of patients with parkinson’s disease, *European Journal of Neuroscience* 36 (6) (2012) 2812–2821.
- [59] T. Montez, S.-S. Poil, B. F. Jones, I. Manshanden, J. P. Verbunt, B. W. van Dijk, A. B. Brussaard, A. van Ooyen, C. J. Stam, P. Scheltens, et al., Altered temporal correlations in parietal alpha and prefrontal theta oscillations in early-stage alzheimer disease, *Proceedings of the National Academy of Sciences* 106 (5) (2009) 1614–1619.

## Appendix A. Selection of Spatial Filters

We selected three filters with the lowest and highest median variance scores  $s_j$ , which we then sorted according to discriminativity  $\max(s_j, 1 - s_j)$  in decreasing order. Note that the median variance score  $s_j$  of the  $j$ th CSP filter was computed as

$$s_j = \frac{\text{median}_i(\text{var}(w_j X^{(i)}))}{\sum_{k=1}^D \text{median}_i(\text{var}(w_k X^{(i)}))} \quad (\text{A.1})$$

where  $X^{(i)}$  denotes the band-pass filtered signal of  $i$ th trial and  $\text{var}(\cdot)$  stands for the variance operator. We preferred the median score over using the eigenvalues  $\lambda_j$  (mean score) in virtue of its greater robustness to outliers. We used the heuristic of [21] to choose the optimal number of filters; note that despite this

variable number of filters for computation of the optimal classifier, we always compute six CSP filters using fix parameters (i.e., frequency range, time segment) for computation of  $SNR$  and  $H$  values.

## Appendix B. Outlier Removal

In our subsequent analysis we estimate Hurst exponents on the amplitude envelope time-series of oscillations during execution of a training session. Although transient modulations of oscillations occur during motor imagery, we consider ongoing EEG activity over long time periods of about 30-50 seconds which allows the estimation of amplitude variations over a range of time scales. Estimation of LRTC during tasks have also been previously performed in a number of studies. It is also important to note that Hurst exponents were estimated in the training session, while we estimated BCI performance in the subsequent feedback phase of the experiment.

We applied the following outlier detection procedure prior to the DFA:

1. Find all outlier time-points  $X(t_i)$  s.t.  $|X(t_i) - \mu| > 4\sigma$  where  $\sigma$  is the standard deviation and  $\mu$  the mean of  $X(t)$ .
2. Remove these outliers and repeat 1. recursively until no outliers remain.
3. Record all outliers obtained from every step of the recursion.

The motivation for the recursive step was that, since outliers may be present on a range of scales, the less pronounced outliers may be missed during the first iteration. After this outlier detection step, we rejected all DFA epochs  $f_{DFA}^2(n, j)$  which spanned time indices including outliers. In order to avoid frequent exclusion of large segments, we restrict DFA estimation to a range  $n$  such that there are at least 10 segments per scale available after outlier rejection.

This approach to outlier rejection for DFA was necessary to ensure that the time-series structure of the data was preserved; simply cutting out portions of data may have introduced spurious autocorrelations which might bias the DFA procedure. We dealt with pauses in the experimental session in a similar manner by requiring that all DFA epochs  $f_{DFA}^2(n, j)$  should range over time-indices from a continuous stretch of the experiment’s duration, i.e. not spanning stretches separated by pauses.

## Appendix C. Mathematical Details

### Appendix C.1. Partial correlation analysis

Several steps of our analysis require the estimation of a partial correlation between two variables  $(X, Y)$  but controlling for the effect of additional variables  $\mathbf{Z}$ . This amounts to measuring the correlation between the residuals  $R_X$  and  $R_Y$  of regression fits between resp.  $X$  and  $\mathbf{Z}$  and  $Y$  and  $\mathbf{Z}$ . Under the null hypothesis  $\mathcal{H}_0$  that  $X$  and  $Y$  are correlated only through  $\mathbf{Z}$ ,  $R_X$  and  $R_Y$  are uncorrelated. In other words, partial correlation is a test for conditional independence (dependence measured by correlation) of  $X$  and  $Y$  given  $\mathbf{Z}$ . Formally, under  $\mathcal{H}_0$  we assume that, for uncorrelated random variables  $\epsilon_1, \epsilon_2$ :

$$X = \beta_1 \mathbf{Z} + \epsilon_1 \quad \text{and} \quad Y = \beta_2 \mathbf{Z} + \epsilon_2 \tag{C.1}$$

Since  $R_X \approx X - \beta_1 \mathbf{Z}$  and  $R_Y \approx Y - \beta_2 \mathbf{Z}$ , this implies that:

$$R_X \approx \epsilon_1 \quad \text{and} \quad R_Y \approx \epsilon_2 \quad (\text{C.2})$$

So these residuals are independent in the limit of data points. In our calculations we use `partialcorr.m` in MATLAB, which allows for the control of correlation over an arbitrary number of variables.

Because we do not wish to assume the presence of Gaussian or linear relationships between the variables we measure, we consider non-parametric *Spearman* correlations. These are related to Pearson partial correlations in exactly the same manner as the standard Spearman correlation is related to a Pearson correlation. Thus, one considers the Pearson correlation between the residuals resulting from regressing the ranks of the values of  $X$  and  $Y$  respectively against the ranks of  $\mathbf{Z}$ . See [48] for further details.

### Appendix C.2. Biserial $R^2$ values

The biserial- $R^2$  value is a general measure indicating how well data fit a statistical model. Here we use this quantity to measure of how much variance of the joint distribution can be explained by class membership. In this case the biserial- $R^2$  value is equivalent to the squared bi-serial correlation coefficient

$$R^2 = \left( \frac{(\mu_1 - \mu_2)\sqrt{n_1 n_2}}{\sigma(n_1 + n_2)} \right)^2 \quad (\text{C.3})$$

where  $\mu_1$  and  $\mu_2$  denote the mean value of class 1 and 2, respectively,  $\sigma$  represents the overall standard deviation and  $n_1$  and  $n_2$  stand for the number of samples in the respective class. Note that the biserial- $R^2$  value should be distinguished from the DFA- $R^2$  value.

## Appendix D. Supplementary Analyses

### Appendix D.1. The effect of filtering on Hurst exponent estimation

Here we present a simulation which aims at checking the size of the bias induced by the temporal smoothing yielded by decimation, filtering the signal using Butterworth filter and calculating amplitude envelope. This corresponds to the same pipeline as applied to the EEG data for Hurst exponent estimation.

To this end, we simulate 1000 uncorrelated Gaussian noise time-series in MATLAB of length  $2 \times 10^6$ . We then decimate the signals to 1/10th of their sampling rate, filter using Butterworth filters of order 5 in a frequency range corresponding to 8–10 Hz assuming the sampling rate of the signal is now 100 Hz (corresponding to the signal analyzed) and calculate the amplitude envelope using the Hilbert transform. We then estimate Hurst exponents using DFA with log-spaced window lengths between 4 and 30 seconds and between 4 and 220 seconds.

We find that the mean Hurst exponent estimated in the range 4 to 30 seconds is 0.52 and the standard deviation over estimates is  $\sigma = 0.020$ . Over the range 4 to 220 seconds we find the mean estimate is 0.51 and the standard deviation over estimates is  $\sigma = 0.022$ .

Since the simulated noise is uncorrelated, the true Hurst exponent of its amplitude envelopes is 0.5. This simulation shows that there is a small positive bias in estimation induced by our preprocessing pipeline but which is not large enough to explain the values of Hurst exponents we observe in our data analysis ( $H$  often larger than 0.6).

### *Appendix D.2. Systematic non-linearity in the DFA fluctuation plot*

Here we present additional data analysis to investigate whether there is a significant systematic departure from monofractal (linear log-log) scaling over the range of time-scales analysed using DFA. If the log-log plot is significantly non-linear, then a quadratic polynomial fit should provide a substantially better fit than a linear fit. Moreover, if the non-linearity consists of a saturation of scaling (lower-gradient at larger scales) then the  $x^2$  coefficient in the quadratic fit should be negative. To this end we fit quadratic polynomials to the DFA log-log plot, i.e. we fit  $\log(F^2(n))$  as a function of  $\log(n)$  in the alpha frequency range (8–13 Hz).

We find that there is indeed a systematic saturation of scaling towards larger time-scales, as we find that the quadratic coefficient of the polynomials fitted has mean less than zero ( $p < 0.001$ ; Wilcoxon sign-rank). However, we find that there is not a large change in the explained variance between the quadratic fit (mean DFA- $R^2 = 0.99 \pm 7.6 \times 10^{-4}$ ) and the linear fit (mean DFA- $R^2 = 0.99 \pm 0.0011$ ).

Thus we conclude that while there is indeed a small systematic saturation for large time-scales, this saturation does not lead to a substantially better model-fit over the linear model.

### *Appendix D.3. The effect of largest window length on Hurst exponent estimation*

Due to the presence of outliers in the data, the longest window length available for Hurst exponent estimation is approximately one minute. We show here, however that Hurst exponents measured over a wider range of time-scales correlate with Hurst exponents measured over a longer range of time-scales.

To this end, we estimate Hurst exponents in the alpha range (8–13Hz) between 0.5 and a maximum window length of 81 seconds and between 4 and 54 seconds respectively for all subjects. We find a Spearman correlation of 0.90 ( $p < 0.001$ ) between the two estimates. However, due to the positive bias introduced by the filtering (see Appendix D.1) we performed our main analysis using DFA analysis with time windows starting at 4 seconds.

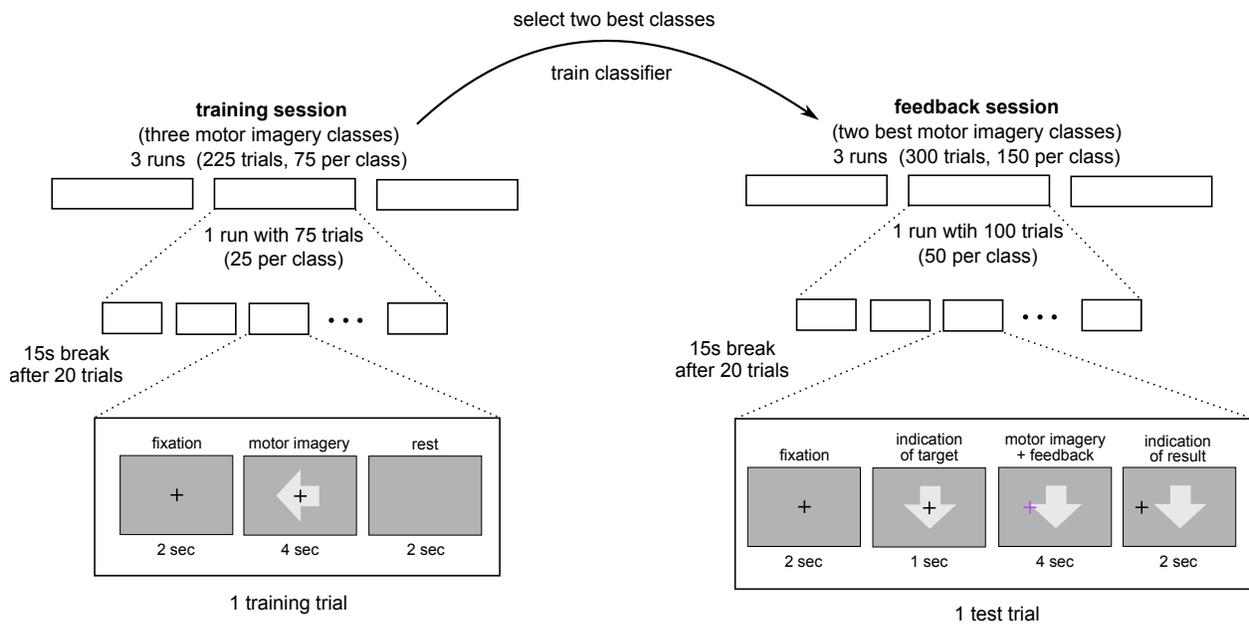


Figure 1: Overview of the experimental protocol.

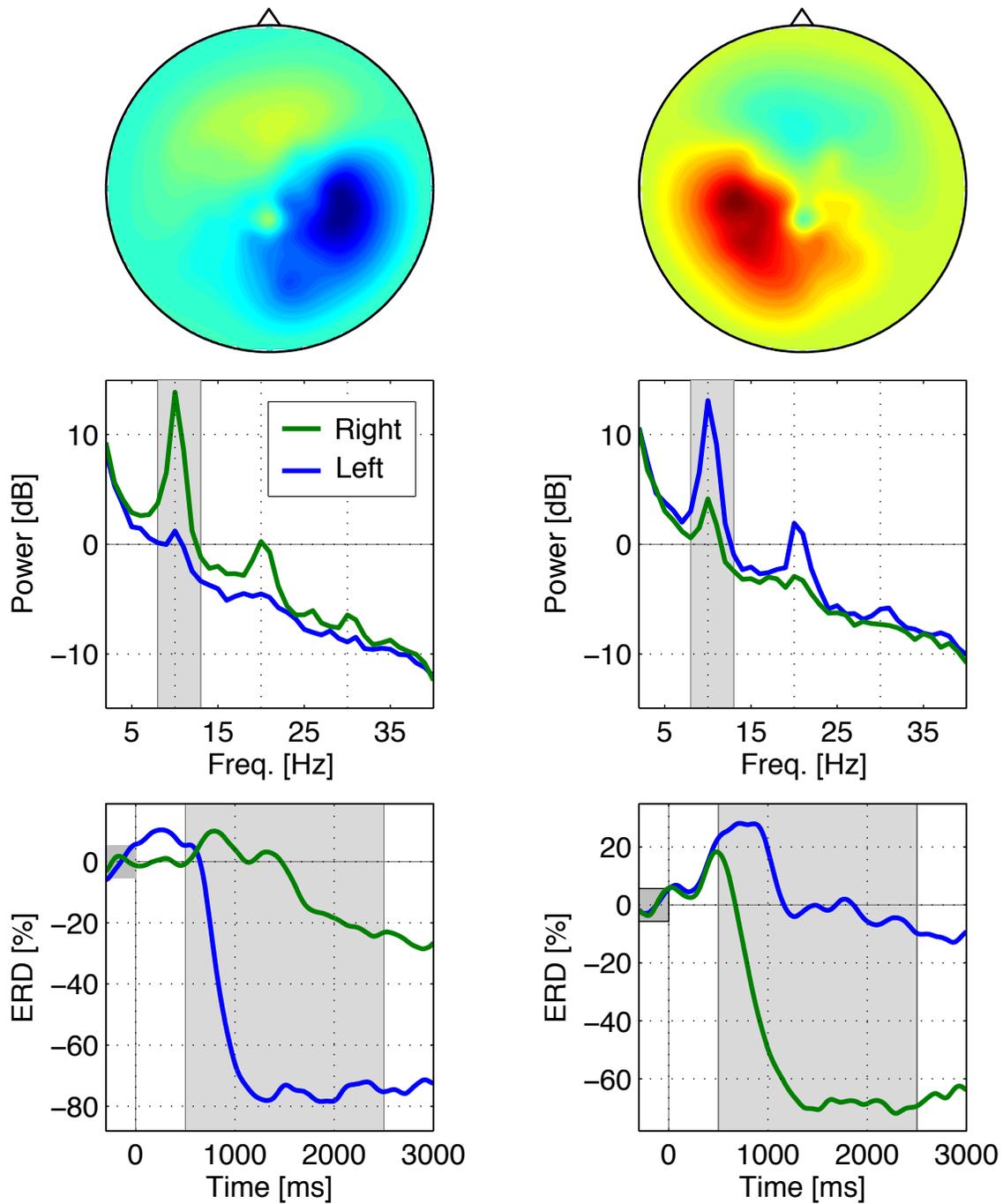


Figure 2: An example of EEG patterns relating to the task performance: the top panels display two sensorimotor CSP topographies, the middle panels their corresponding power-spectra during training epochs, with blue and green lines indicating left and right hand imagery, respectively, and the bottom row displays event related desynchronization (ERD) responses. For left hand motor imagery (left column) the CSP pattern shows an activation over the right sensorimotor cortex and the power spectrum (blue line) displays a strong power decrease in the alpha and beta band. The event related desynchronization response of the alpha band filtered signal depicts the time course of the power decrease. For right hand motor imagery (right column, green lines) the responses are analogous.

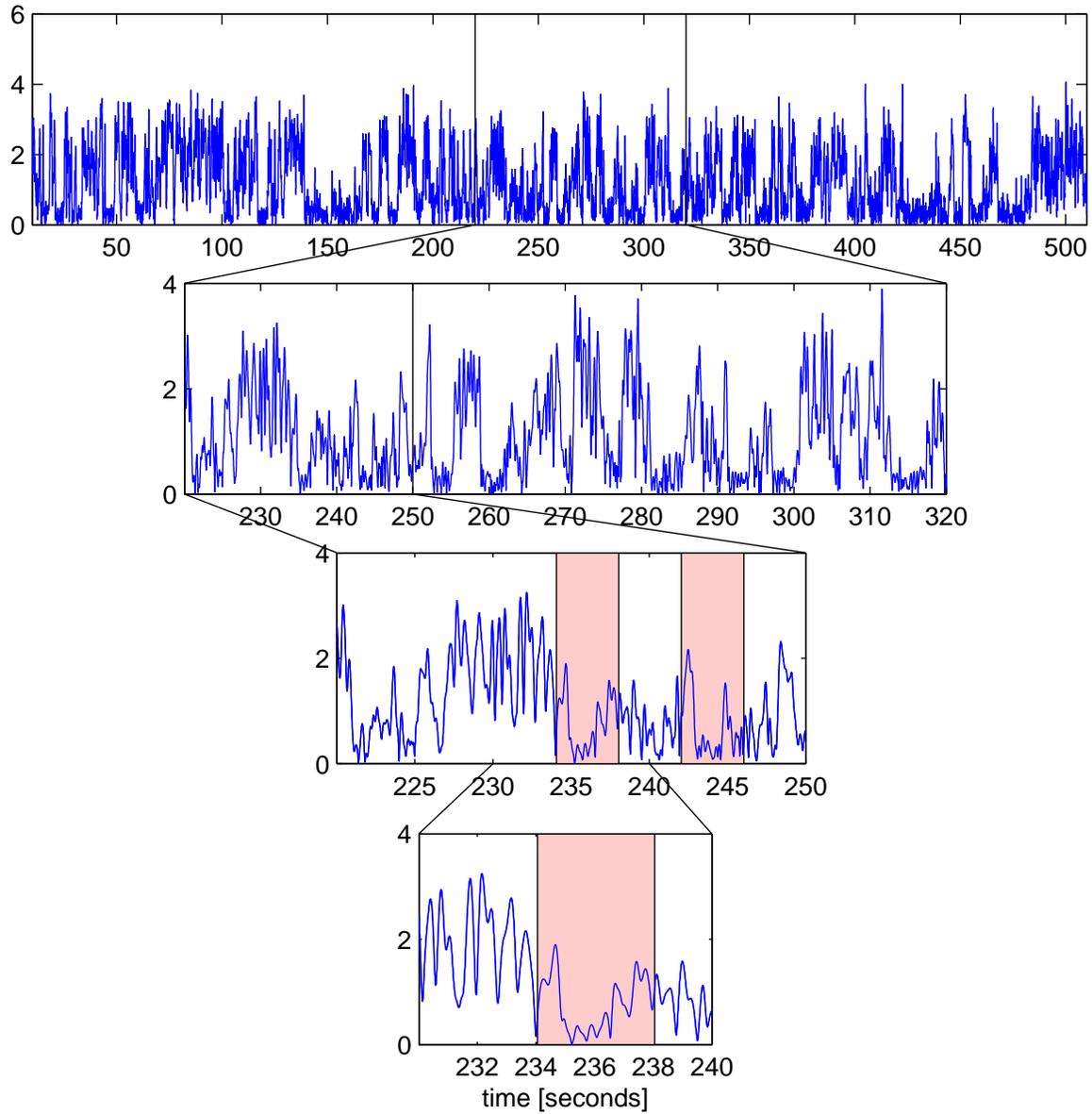


Figure 3: The scale free nature of oscillation amplitudes. The time-series are amplitude envelopes from a typical CSP component, where single trials are marked in red at the smallest time scale. On the time scales of minutes bursts of neuronal activity are clearly distinguishable (top row). Within these bursts finer bursts lasting less than a minute can be identified (second row). Furthermore within these finer bursts one can observe bursts of activity on the time scale of single trials (bottom row). These scale-free dynamics of neural activity indicate that the signal contains information on multiple time-scales.

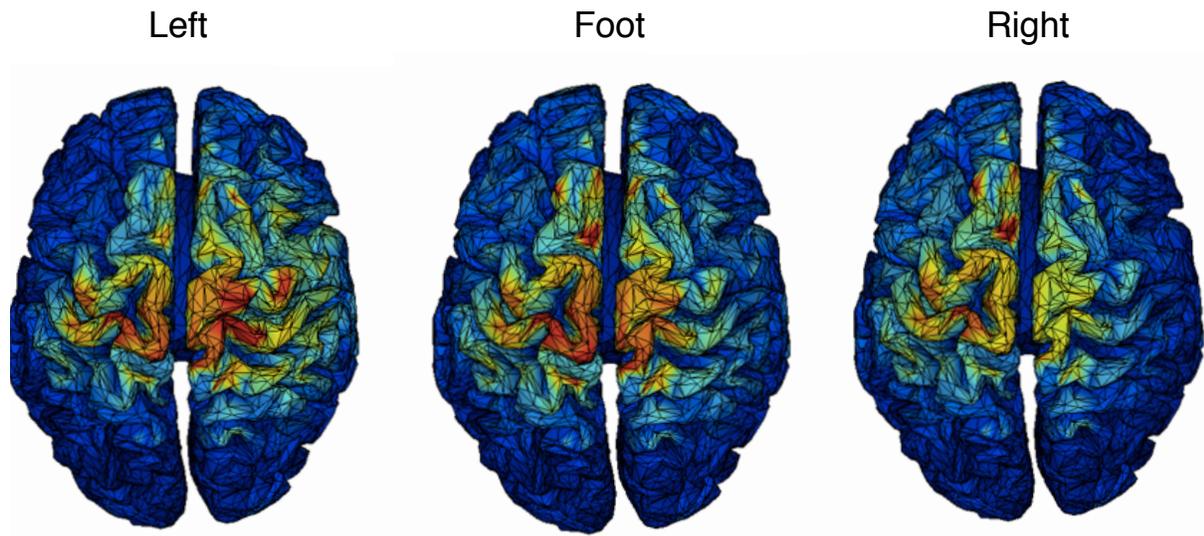


Figure 4: Grand-average weighted minimum norm inverse solutions for the most discriminative CSP patterns from each class: left, foot, and right hand motor imagery. The results confirm the sensorimotor origin of the most discriminative CSP components considered.

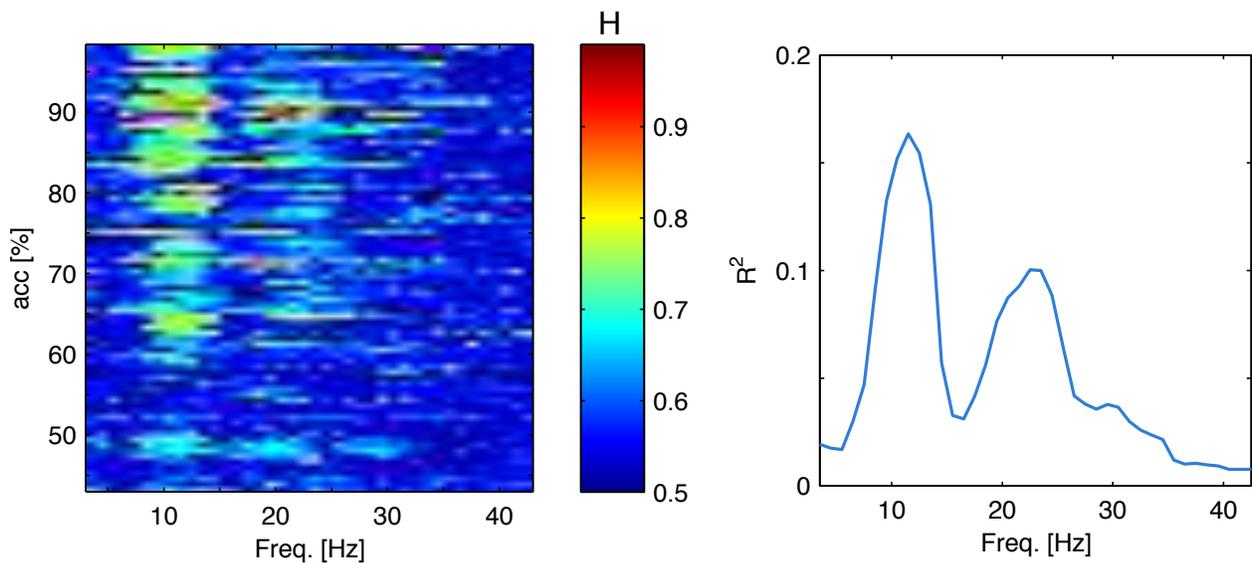


Figure 5: Sorting frequency-specific Hurst exponents according to individual BCI performance. The left panel displays Hurst exponents  $H$  (in color) for each frequency bin of the first CSP component. Subject-specific components are sorted according to the BCI performance ( $y$ -axis). The right panel depicts the average biserial- $R^2$  in the explanation of  $acc$  by  $H$  values for a range of frequency bands of the first CSP component.

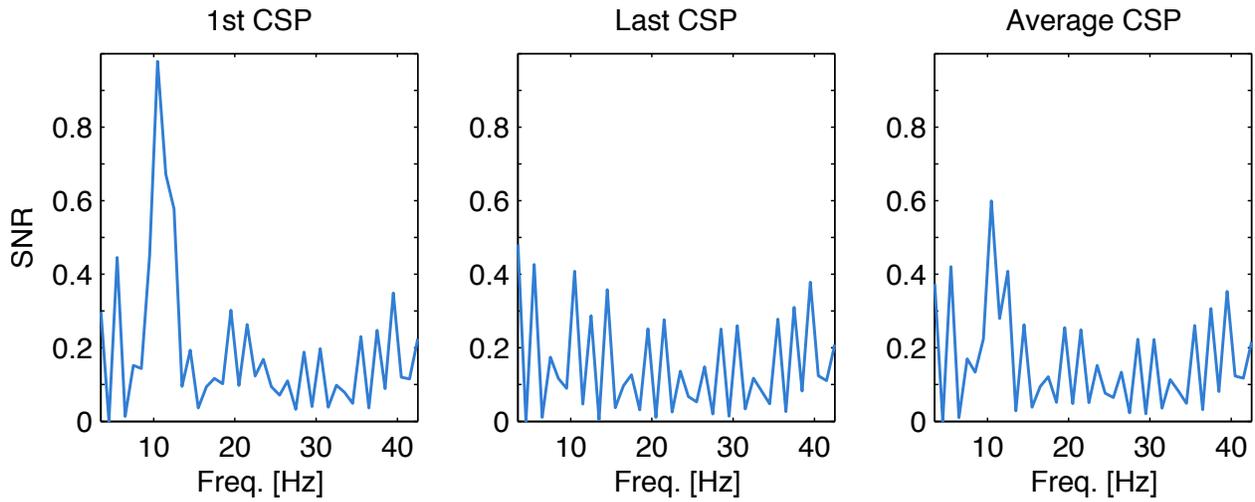


Figure 6: Grand-average signal-to-noise ratios for CSP components for each frequency. Left: The most discriminative CSP component; middle: the sixth CSP component; right: the average across the six most discriminative CSP components. The results show that signal-to-noise ratio is highest in the mu/alpha range  $\approx 10$  Hz). For the most discriminative CSP component the  $SNR$  value is close to 1 whereas for the last CSP component it is significantly smaller and not frequency specific.

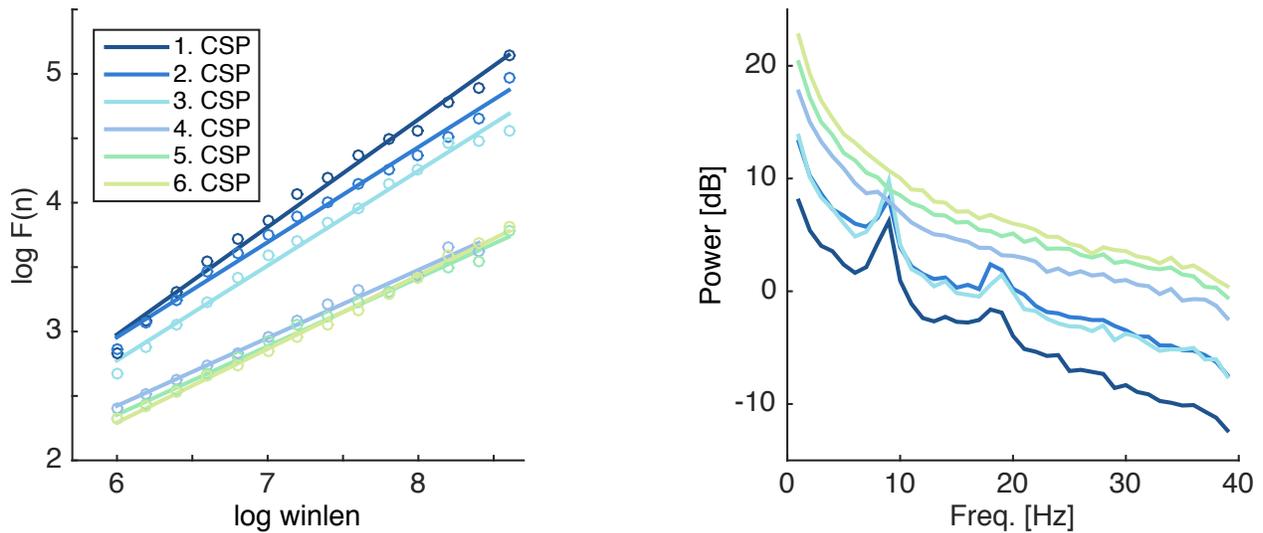


Figure 7: Sample DFA log-log plots for estimation of the Hurst exponent on six CSP components of a representative subject. The angle of each line is a Hurst exponent. The right panel displays the corresponding power-spectra.

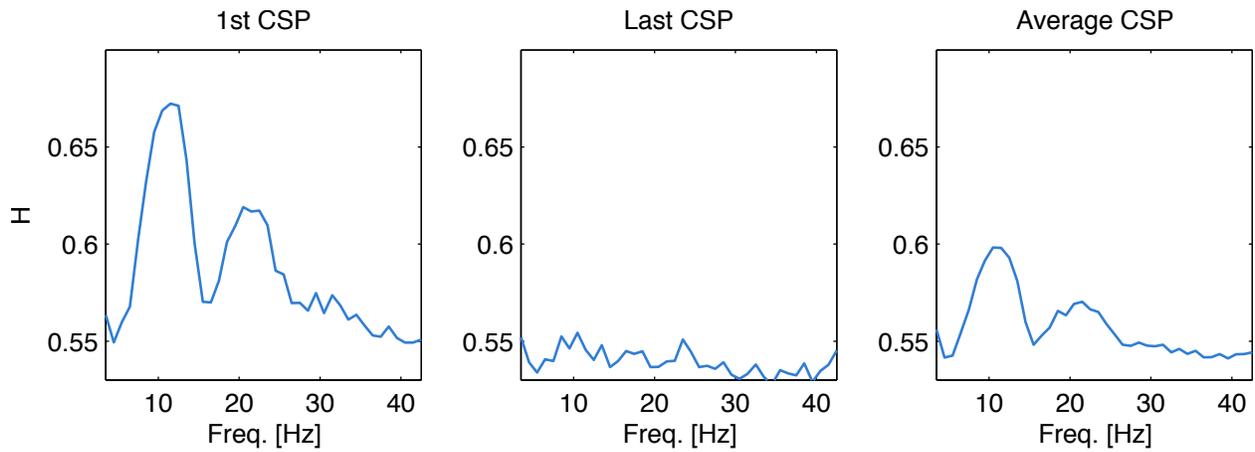


Figure 8: Grand-average Hurst exponents for the first CSP component at each frequency (left), the last CSP component (middle) and the average across CSP components (right). The results show that Hurst exponents are highest in the mu/alpha range, but are also elevated in the beta range. For the last CSP component the  $H$  values do not show discernible peaks in these frequency ranges.

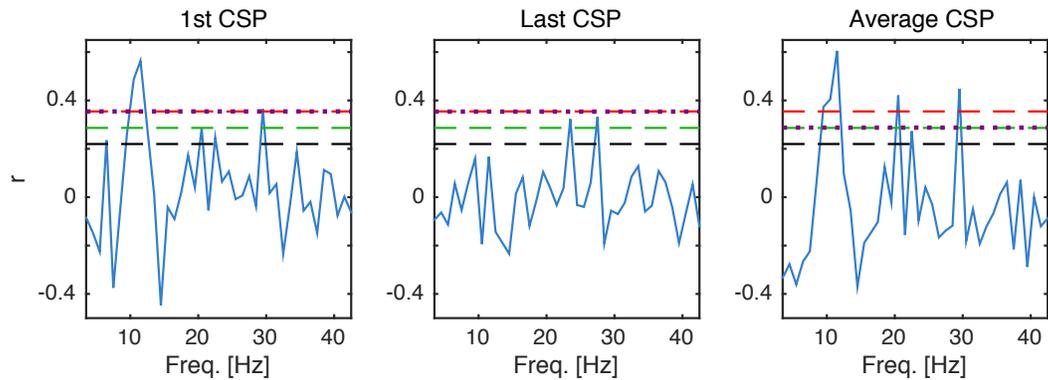


Figure 9: Correlations between signal-to-noise ratio and Hurst exponents for the first CSP component at each frequency (left), the last CSP component (middle) and the average across CSP components (right). In each case the black line denotes the  $p = 0.05$  threshold; the green line,  $p = 0.01$ ; the dotted purple line, the FDR corrected  $p = 0.05$  threshold; and the red line, the Bonferroni corrected  $p = 0.05$  threshold for  $n = 40$ . The results show that Hurst exponents and signal-to-noise ratio are significantly correlated in the alpha range but not outside this range. For the last CSP component significant correlation is lacking for all frequencies.

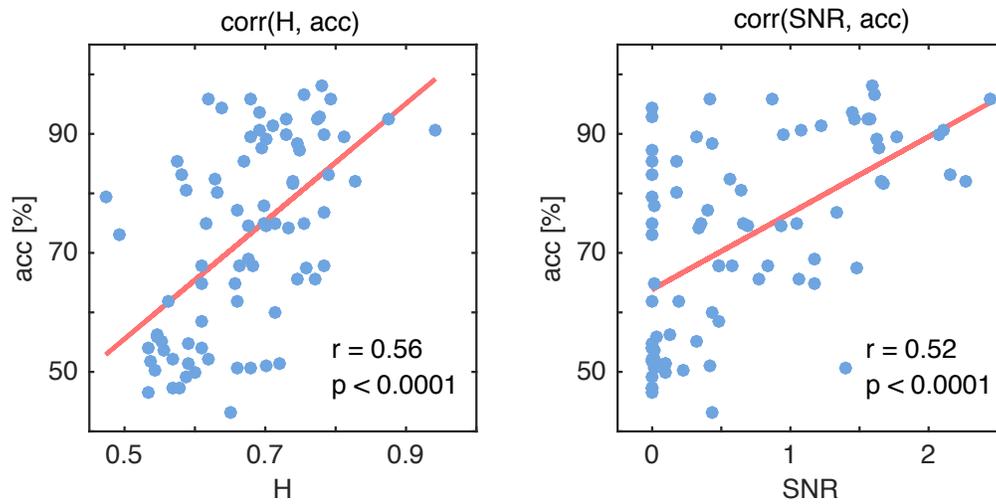


Figure 10: Prediction of BCI classification accuracy. The figure displays scatter plots of Hurst exponents,  $H$ , BCI classification accuracy  $acc$  and signal-to-noise ratio,  $SNR$ . The  $H$  and  $SNR$  values were estimated in the 10–15 Hz range on the first (most discriminative) CSP component, whereas  $acc$  was computed on multiple CSP components and a subject-specific frequency band. The results show that both  $H$  and  $SNR$  are highly correlated with classification accuracy.

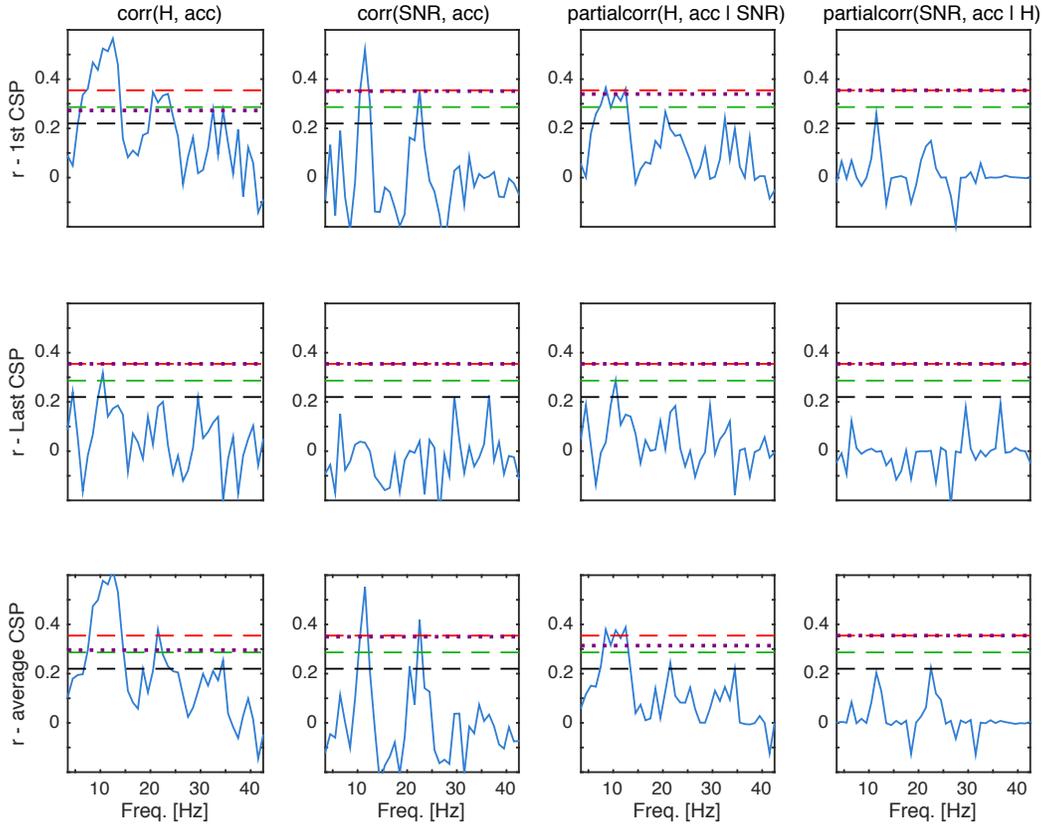


Figure 11: Prediction of BCI classification accuracy by Hurst exponents and signal-to-noise ratio. Based on the first CSP component, first row displays the correlation of  $acc$  with  $H$  and  $SNR$  as well as a partial correlation with the effects of  $H$  or  $SNR$  having been removed. Middle row—as per the first row but for the last CSP component. Bottom row—as per the first row but for the average of CSP components. In each case the black line denotes the  $p = 0.05$  threshold; the green line,  $p = 0.01$ ; the dotted purple line, the FDR corrected  $p = 0.05$  threshold; and the red line, the Bonferroni corrected  $p = 0.05$  threshold for  $n = 40$ .

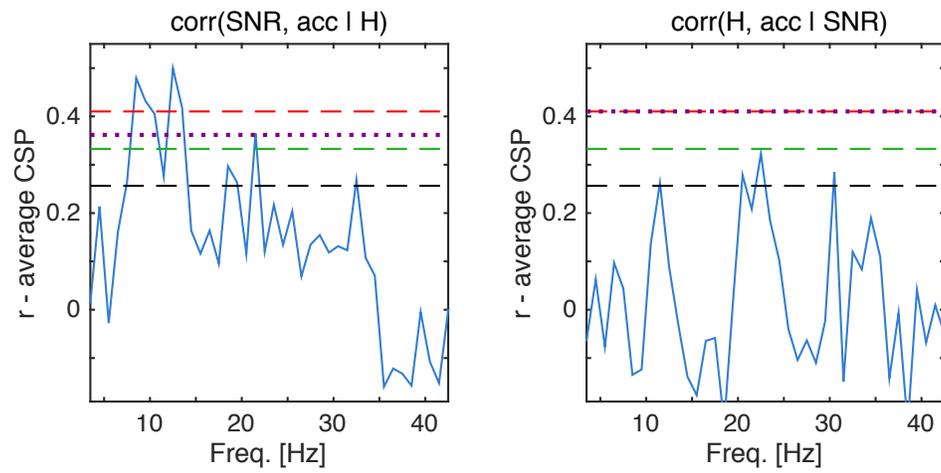


Figure 12: Partial correlation between BCI classification accuracy and Hurst exponents or signal-to-noise ratio with the effects of  $SNR$  or  $H$  respectively being removed for subjects who performed significantly above the chance level. In each case the black line denotes the  $p = 0.05$  threshold; the green line,  $p = 0.01$ ; the dotted purple line, the FDR corrected  $p = 0.05$  threshold; and the red line, the Bonferroni corrected  $p = 0.05$  threshold for  $n = 40$ .