Unsupervised classification of operator workload from brain signals

Matthias Schultze-Kraft\textsuperscript{1,2,6}, Sven Dähne\textsuperscript{3,4}, Manfred Gugler\textsuperscript{5}, Gabriel Curio\textsuperscript{2,5} and Benjamin Blankertz\textsuperscript{1,2}

\textsuperscript{1} Neurotechnology Group, Technische Universität Berlin, Berlin, Germany
\textsuperscript{2} Bernstein Focus: Neurotechnology, Berlin, Germany
\textsuperscript{3} Machine Learning Group, Technische Universität Berlin, Berlin, Germany
\textsuperscript{4} Berlin Big Data Center, Berlin, Germany
\textsuperscript{5} Charité—University Medicine Berlin, Campus Benjamin Franklin, Berlin, Germany

E-mail: schultze-kraft@tu-berlin.de

Received 2 October 2015, revised 1 February 2016
Accepted for publication 10 March 2016
Published 14 April 2016

Abstract

Objective. In this study we aimed for the classification of operator workload as it is expected in many real-life workplace environments. We explored brain-signal based workload predictors that differ with respect to the level of label information required for training, including entirely unsupervised approaches. Approach. Subjects executed a task on a touch screen that required continuous effort of visual and motor processing with alternating difficulty. We first employed classical approaches for workload state classification that operate on the sensor space of EEG and compared those to the performance of three state-of-the-art spatial filtering methods: common spatial patterns (CSPs) analysis, which requires binary label information; source power co-modulation (SPoC) analysis, which uses the subjects’ error rate as a target function; and canonical SPoC (cSPoC) analysis, which solely makes use of cross-frequency power correlations induced by different states of workload and thus represents an unsupervised approach. Finally, we investigated the effects of fusing brain signals and peripheral physiological measures (PPMs) and examined the added value for improving classification performance. Main results. Mean classification accuracies of 94\%, 92\% and 82\% were achieved with CSP, SPoC, cSPoC, respectively. These methods outperformed the approaches that did not use spatial filtering and they extracted physiologically plausible components. The performance of the unsupervised cSPoC is significantly increased by augmenting it with PPM features. Significance. Our analyses ensured that the signal sources used for classification were of cortical origin and not contaminated with artifacts. Our findings show that workload states can be successfully differentiated from brain signals, even when less and less information from the experimental paradigm is used, thus paving the way for real-world applications in which label information may be noisy or entirely unavailable.

Keywords: workload, classification, brain–computer interface, unsupervised, EEG

(Some figures may appear in colour only in the online journal)
of alertness can lead to an overload of the human operator, having critical consequences for health, safety and efficiency aspects. An assessment of operator workload has therefore become crucial for the operation of automation systems. Such an assessment can then be utilized to implement a system that automatically self-regulates the level of human–machine interaction by adapting to changes of the operator’s workload (Parasuraman and Wilson 2008).

The workload state of humans is not directly observable but can be inferred indirectly through various variables, including subjective reports, psycho- and peripheral physiological measures (PPMs) as well as task performance. During the last decades, neurophysiology has proven to be a sensitive and informative modality for the measurement of mental states. In particular, the human electroencephalogram (EEG) has been shown to provide reliable estimators of workload. EEG estimators of workload are based on the fact that changes in workload are associated with characteristic changes in the EEG. These typically amount to modulations in the power of oscillatory activity in particular frequency bands of the EEG (Buzsaki and Draguhn 2004). The most prominent frequency bands with power changes related to workload are theta (4–7 Hz) and alpha (8–13 Hz). Theta power has been shown to be positively correlated with workload, most notably over frontal regions, whereas alpha power is typically found to be negatively correlated with workload, in particular over parietal regions (Govins and Smith 2003, Holm et al 2009).

An EEG-based assessment of workload for adaptive automation can be implemented using a brain–computer interface (BCI) (Wolpaw and Wolpaw 2012). These are systems that use techniques from machine learning in order decode intentions or mental states of individuals in real-time and translate these into computer commands. While most research on BCIs aims at providing a communication channel for users based on brain signals, so called passive BCIs (Zander and Kothe 2011) make use of spontaneously generated brain signals in order to allow for an online detection of the user’s mental state (Blankertz et al 2010). The foundations for the use of BCI for workload detection were laid already two decades ago (Govins et al 1995, Pope et al 1995), showing that an EEG-based workload assessment can be used for adaptive automation purposes in operational environments. Since then, several studies have shown that a moderation of operator task load based on a workload index derived from the ongoing EEG can be successfully used for enhancing task performance (Prinz et al 2000, Scerbo et al 2003, Kohlmorgen et al 2007, Wilson and Russell 2007).

One of the challenges in the development of BCIs for adaptive automation is their application in real-life environments, such as during driving, gaming or industrial workplaces. Typical tasks in such environments involve increased requirements in multiple modalities, from higher cognitive functions such as decision making and response inhibition to the more basic processing of visual input and execution of motor commands. We refer to the workload induced by such tasks as operator workload. In this study, we envisioned an industrial workplace where an automated plant produces parts at a certain speed and a human operator assembles them into a final product as they are transported past him on a conveyor. Such a workplace that requires a high and constant level of manual skills and alertness offers the ideal scenario for employing an adaptive automation system because it involves several competing goals: while a high work speed is desired to maximize productivity, a too high workload of the operator may result both in an increased rate of defective goods and may be detrimental to the operator’s health. Here, a continuous assessment of workload would allow to dynamically adapt the work speed to a level that counterbalances the plant’s productivity and the operator’s performance. We mimicked such a scenario by designing an experimental task on a touch screen that required permanent visual and motor engagement by subjects. By changing task difficulty, the experiment induced two levels of operator workload in subjects. Note that the two workload states in this study are defined by a difference in task performance but are not directly based on a subjective measure (Hart and Staveland 1988). While participants consistently reported the difference between the two workload conditions as clearly distinguishable, error metrics and subjective workload are not identical (Putze et al 2010). In an offline analysis we then investigated to what extent the workload condition could be classified from the acquired EEG data.

In this study, we sought to investigate the performance of workload classification from EEG spectral features, using progressively less label information from the experiment. In most experiments the labels of the experimental condition at given points in time are well-known. Thus, the classical approach is to train a linear classifier on the extracted features using those exact labels. In our industrial workplace scenario, however, a situation is very well conceivable in which the different workload conditions are neither externally induced nor known (but result e.g. from a self-regulation mechanism) but instead the performance of the operator (e.g. error rate) is known. This variable is expected to reflect the workload state and can be considered a noisy version of the true labels. The obvious approach in this case is to employ a linear regression on the EEG features, using the error rate as target variable. In a third scenario we assume that also the error rate is unknown and no other information about the workload condition is available. This scenario requires an approach that combines EEG features using only the prior knowledge about the spectral changes in the EEG associated with workload and their spatial localization (Govins and Smith 2003, Holm et al 2009). Because no label information is available, such an approach must be inherently unsupervised.

In order to classify (or predict) workload levels we implemented six different predictive models, which fall into one out of two categories: (i) channel-based and (ii) spatial-filter-based. In the channel-based models, feature extraction is done for each recording channel separately. In the spatial-filter-based models the data are first projected onto a set of optimized spatial filters and features are then extracted from the output of the spatial filters. Furthermore, each of the channel-based and spatial-filter-based models fall into one out
of three sub-categories, depending on the amount of information required by the approach. As outlined above, these sub-categories include (a) the use of binary class labels, (b) the use of a continuous error measure, and (c) no use of a supervision signal at all. We applied each of the six predictive models to the recorded EEG data and validated their performance in discriminating between induced workload states and in predicting the subjects’ error rate. Since we sought to implement a ‘true’ BCI that uses exclusively brain signals for workload classification, before validation we removed confounding signals that might stem from artifacts. Finally, in addition to EEG, we recorded the subjects’ heart rate, respiration rate and skin conductance and examined the potential of such non-cerebral signals in improving the BCI performance.

2. Methods

2.1. Experimental task

Ten healthy male subjects, aged 26–40, participated in the experiments. All participants gave their informed oral and written consent. Subjects were instructed to carry out a task on a 21 inch touch screen lying on a table in front of them (figure 1(a)). The task was designed as a computer game: objects consisting of three vertically aligned screws (screw triplets) were falling vertically with equal velocity from random positions at the top of the screen, approaching the bottom of the screen. Each screw in a triplet was randomly tagged and colored with one out of four predefined colors, multiple occurrences of one color were not allowed. At the bottom border of the screen was a bucket consisting of three vertical segments. Using their index fingers, subjects could tag (and untag) the bucket segments with colors by pressing colored buttons positioned both at the left and right borders of the screen. Furthermore they could move the bucket horizontally along the entire bottom screen border by sliding the bucket with one of the index fingers. The task was to catch each falling screw triplet with the bucket before it reached the bottom, ensuring that each time the bucket was tagged with the same colors and in the same order as the caught screw triplet. Catching with wrong colors was considered an error as well as letting a triplet hit the bottom of the screen. The falling speed of the triplets was constant throughout the experiment, however the interval between the occurrence of the triplets varied in two different conditions. In the low workload condition (L) the interval between each set was constant, whereas in the high workload condition (H) the intervals were shorter and varied randomly.

2.2. Experimental setup

Before the experiment started the game’s parameters were calibrated for each subject individually in three 5 min runs. The first run was a free trial run during which subjects could familiarize themselves with the game. All participants quickly reached a fairly constant error rate. In the second run the falling velocity of the screw triplets and the intervals between them were adjusted such that subjects where able to accomplish the task with an error rate of approximately 10%. Participants consistently reported the task being moderately demanding but not stressful. This setting was then used during the experiment for the low workload condition. In the third run the game parameters for the high workload condition were determined by adjusting the variance of the randomly occurring intervals between the screw triplets such that an increased sense of stress was reported and yielding error rates between 20% and 25%. Eventually, each subject performed four sessions of 24 min each during which EEG was recorded (figure 1(b)). Each session consisted of 16 blocks of 90 s each of alternating L and H conditions, starting with condition L. In order to mimic the conditions of an industrial workplace, during the whole experiment a closed loop recording of a real acoustic scenery of an industrial work environment was played through speakers at realistic volume. The touch screen task was implemented in the open source framework Pyff (Ventur et al 2010).
2.3. Data acquisition

EEG data was recorded at 1000 Hz using BrainAmp amplifiers and 64 electrode actiCAP (Brain Products GmbH, Gilching, Germany). In addition to the EEG three peripheral physiological signals were recorded: electrocardiogram (ECG) was recorded bipolarly with two surface Ag/AgCl electrodes positioned at breast height at the front and back of the body, respiration activity was recorded using a Sleepmate respiratory effort belt (Ambu, Ballerup, Denmark), skin conductance was recorded with a GSR module (Becker Meditec GmbH, Karlsruhe, Germany).

2.4. Data analysis

The data analysis employed in this study is schematically described in figure 2 and detailed in the following subsections. First, several preprocessing steps including artifact removal and bandpass filtering were performed on all data sets. Next, one of six predictive models was trained on part of the data and subsequently tested on data that was not used for training. The output of a predictive model was evaluated in two separate ways, namely by computing the model’s correct classification rate of workload conditions and the correlation of the model’s output with the error rate. Some of the predictive models employ so-called spatial filtering methods which are in turn based on a linear generative model of EEG. In order not to interrupt the flow of reading, we have summarized the background theory on the generative model, as well as a detailed description of spatial filtering methods CSP, SPoC, cSPoC and spatio-spectral decomposition (SSD) in appendix. The appendix also details how to derive spatial activation patterns that can be interpreted with respect to the physiological origin of the signal that was extracted by a spatial filtering method.

We now continue with a detailed account of the preprocessing steps and thereafter describe the predictive models.

2.4.1. Preprocessing of EEG.

Suppression of motion- and eye movement artifacts. Owing to the design of the experimental task, which involved a persistent moving of eyes and head, it was expected that the recorded EEG data was contaminated by muscular and ocular artifacts. Electromyogram (EMG) activity is most strongly manifested in frequency bands higher than 20 Hz (Whitham et al 2007). In our study we made use of theta (4–7 Hz) and alpha (8–13 Hz) frequency bands only, and therefore contamination with EMG was of little concern in our study. Artifacts caused by eye movements (also known as Electro-oculogram, short EOG, activity) on the other hand may pose a problem, because their frequency content overlaps strongly with the frequency bands of interest in this study. Due to the lack of electrodes that directly recorded this activity, we used the difference between electrodes F9 and F10 and the average of electrodes Fp1 and Fp2 to estimate horizontal and vertical EOG activity, respectively. Those four electrodes were excluded from all subsequent analyses. The estimated horizontal and vertical EOG activity was then removed from the EEG data by means of a regression approach, which is described in detail in Parra et al (2005). In this approach, the contribution of the EOG signals to each of the remaining recording channels is estimated and subsequently removed. Nevertheless, we investigated the possible influence of residual EOG activity on our predictive models, see section 3.3.

Segmentation into epochs. The bandpass filtered EEG data were then segmented into non-overlapping time windows of 90 s length. We refer to the data within such a time window as an epoch.
Bandpass filtering. The scalp EEG was bandpass filtered in two separate frequency bands, namely the theta range (4–7 Hz) and the alpha range (8–13 Hz). For each of the two frequency bands, a Butterworth filter of order 5 and non-causal IIR filtering was employed. This operation yielded two bandpass filtered datasets per subject, which served as input to the channel-based prediction models.

Bandpass filtering and dimensionality reduction (spatial-filter-based models only). For the spatial filtering methods (see next section) an additional preprocessing step consisting of dimensionality reduction was applied. We used SSD (Nikulin et al. 2011, Haufe et al. 2014a) (see also section 2.1 in the appendix) in order to extract a maximally oscillatory subspace for the theta range and the alpha range, respectively. Thus, SSD was applied twice to the EEG data in order to optimize for theta and alpha separately. From both applications of SSD, the first 15 SSD components were retained while the remaining components were discarded, yielding two 15 dimensional datasets per subject, which served as input to the spatial-filter-based predictive models.

2.4.2. Preprocessing of PPMs.

Extraction of cardiac- and respiration frequency signal. A subject-specific heart beat template was extracted from the raw ECG signal and subsequently correlated with the signal in a sliding window. The times of high correlation peaks detected by a threshold represented the times of heart beats, from which the time resolved frequency was obtained. Similarly, the respiratory activity recorded by the respiratory effort belt was lowpass filtered and the peaks and troughs of each inhalation and exhalation threshold detected, yielding a time resolved respiration frequency.

2.4.3. Predictive models. We implemented three predictive models based on single-channel features (channel-based models) and three corresponding models based on features derived from the output of spatial filters (spatial-filter-based models). The models are tailored to one of three scenarios in which different amount of information about the current workload state is available for training the models.

In all models, we used log-variance of bandpass filtered data, computed within 90 s epochs, as features that represent spectral modulations. For the channel-based models, log-var features are computed according to

\[ P_{i,f}(e) = \log(\text{Var}(x_{i,f}(e))) \]

where \( x_{i,f}(e) \) denotes the signal of a single EEG channel indexed by \( i \in \{1, \ldots, N_{\text{channels}}\} \), bandpass filtered for frequency band \( f \in \{\theta, \alpha \} \), within the epoch indexed by \( e \in \{1, \ldots, N_{\text{epochs}}\} \). For the spatial-filter-based models, log-var features are computed according to

\[ P_{i,f}(e) = \log(\text{Var}(w_{i,f}^T X_{i}(e))) \]

where \( X_{i}(e) \) denotes the signal of all bandpass-filtered EEG channels within the epoch \( e \), while \( w_{i,f} \) denotes the frequency-band-specific spatial filter indexed by \( i \in \{1, \ldots, K\} \) with \( K \) being the number of filters to be used per frequency band.

In order to suppress drifts and fluctuations that are outside of the relevant time-scale, the log-var features were high-pass filtered at 1/600 Hz (thereby removing changes slower than 10 min). Once the log-var features have been computed and filtered, they are combined according to

\[ \hat{y}(e) = \sum_{f=\{\theta, \alpha\}} \sum_{i=1}^{K} b_{i,f} \cdot P_{i,f}(e), \]

where \( \hat{y}(e) \) denotes the scalar output of the model for the \( e \)'th epoch.

The parameters that all of the predictive models have to estimate using training data are the weighting coefficients \( b_{i,f} \). Additionally, spatial-filter-based models have to estimate the spatial filters \( w_{i,f} \). The predictive models differ in the algorithms that are used to optimize the spatial filters and weighting coefficients. In the following we describe the properties of each predictive model in detail.

Classification models:

If the true exact labels from the experiment are available, it is possible to employ a classification based approach.

Model ChCsp (classification on channels) uses regularized linear discriminant analysis (LDA) to train the weighting coefficients based on labels. The regularization is based on shrinkage of the feature covariance matrix (Blankertz et al. 2011).

The corresponding spatial-filter-based model SFCsp (spatial filtering and classification) uses the common spatial patterns (CSPs) algorithm (Blankertz et al. 2008) to train the spatial filters and LDA to combine the resulting log-var features. CSP finds filters that increase the power contrast between two experimental conditions and is a popular method in BCI research. The spatial filters were optimized using the CSP algorithm separately for each frequency band, such that the resulting CSP signals have maximal variance difference between the two workload classes. From each of the two resulting set of filters, the \( K = 3 \) filters with the largest absolute Eigenvalue were selected, yielding six filters in total.

Regression models:

Whenever a continuous measure, such as the the error rate, is available a regression based approach can be used.

In the model ChReg (regression on channels) we use the error rate as a supervision signal to optimize the weighting coefficients using ridge regression, which is a regularized version of ordinary regression. The regularization is based on shrinkage of the feature covariance matrix.

The corresponding spatial-filter-based model is denoted by SFReg (spatial filtering and regression). In order to train the spatial filters, we employ a novel method called the source power co-modulation (SPoC) analysis, which optimizes the correlation of band power to a given target function, in this case the error rate (Dihne et al. 2014a). The spatial filters were optimized using the SPoC algorithm separately for each frequency band. Note that the labels are defined by the experimental paradigm, i.e. they correspond directly to the \( L \) and \( H \) condition of the experiment.
frequency band. From each of the two resulting sets of filters, the \( K = 3 \) filters with the largest absolute Eigenvalue were selected, yielding six filters in total. The resulting log-var features were then combined using ridge regression.

**Unsupervised models:**

In this setting no supervision signal is present.

A well-known, workload-modulated interaction between frontal theta and parietal alpha power (Holm et al. 2009) forms the basis of the model \( \text{ChUns} \) (unsupervised on channels). The model output is constructed by subtraction of theta log-var features at channel \( Fz \) from alpha log-var features at channel \( Pz \).

The filters for the corresponding spatial-filter-based model were optimized using the cSPoC algorithm (Dühne et al. 2014b). Given two datasets, cSPoC maximizes pairs of filters such that the bandpower dynamics between pairs are highly (anti-)correlated. In this setting, cSPoC was employed to maximize anti-correlation between theta and alpha bandpower dynamics, i.e. it builds on the same assumption as (Holm et al. 2009). Three filter pairs (i.e. \( K = 3 \) filters per frequency band) were optimized and the resulting log-var features are combined by averaging them within bands and then subtracting the resulting averaged theta cSPoC features from the averaged alpha cSPoC features.

The six models are summarized in table 1. In this table, the models are categorized according to the level at which features are extracted and to how much information is required during training.

In addition to the EEG-based features and predictive models outlined above, we tested whether features derived from the PPMs constitute an added value.

**PPM model:** This model is comprised of the PPMs recorded during the experiment. An inspection of their change over time confirmed that they were modulated by the induced workload state with equal direction (figure 3 and section 3.1). The mean heart rate, respiration rate and skin conductance in each data epoch were computed, high-pass filtered at 1/600 Hz (thereby removing changes slower than 10 min), z-scored, and finally summed up in order to generate the model output.

What all models have in common is that they yield a one-dimensional output with one value for each data epoch. This model output was then used two-fold for validation: One validation approach consisted in treating the sign of the output as the classification of workload into either the low (positive sign) or high (negative sign) state and computing the average classification accuracy. The other approach consisted in computing the correlation of the model output with the error rate of the subject. In order to test for generalization, the validation was done in a chronological four-fold cross-validation. In this cross-validation scheme, each of the four sessions became the test set once, while the remaining three were used for training. The classification accuracy and correlation value were then computed from the average across test folds.

### 3. Results

**3.1. Behavioral and physiological data**

The task induced workload had a clear impact on subjects’ error rate. As shown in figure 3(a), the error rate was modulated by the condition in a step function manner. During the low workload condition the rate was approximately 10%. When the condition was changed from low to high workload (at \( t = 90 \) s) the error rate rapidly increased to about 20%–25% after which it remained approximately constant throughout the high workload condition. Accordingly, the opposite occurred when switching from high to low workload. We found that also the three PPMs respiration, heart beat and electrodermal response were modulated by the task (figures 3(b)–(d)). Similar to the error rate, the respiratory and cardiac frequency were modulated by the condition in a step function manner, with lower rates during the low workload condition. Electrodermal response, on the other hand, constantly decreased during low workload and increased during high workload condition.

**3.2. Validation of models**

We assessed the ability of each of the six predictive models to classify between both workload conditions and to predict the error rate. Therefore, according to the analysis pipeline described above, each model was trained and validated individually. Figure 4(a) shows the mean classification accuracies for the six models, averaged across subjects. Among the channel-based models, \( \text{ChCla} \) performs best (90.4%), followed by \( \text{ChReg} \) (87.5%) and \( \text{ChUns} \) (75.7%). Among the spatial-filter-based models, the \( SF_{Cla} \) performs best (94.1%), followed by \( SF_{Reg} \) (91.8%) and \( SF_{Uns} \) (82.3%). The correlations of model outputs with the error rate show a very similar picture (figure 4(b)), with mean correlations of 0.65, 0.63 and 0.47 for the channel-based models and 0.68, 0.67 and 0.60 for the spatial-filter-based models, respectively. These results reveal that both across model type (i.e. without or with spatial filtering) and across validation type (i.e. classification accuracy or correlation with error rate) there is a successive decrease of performance with respect to the type of labels used by the model: \( \text{Cl} > \text{Reg} > \text{Uns} \). An inspection of classification

### Table 1. The six models used to predict workload and which algorithms they use. The models are categorized with respect to the level at which log-var features are extracted (channel-based vs spatial-filter-based) and the amount of information required during training (binary labels for classification, continuous measure for regression, no supervision signal for unsupervised).

<table>
<thead>
<tr>
<th>Classification</th>
<th>Channel-based</th>
<th>Spatial-filter-based</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChCla (LDA)</td>
<td>SFCla (CSP + LDA)</td>
<td></td>
</tr>
<tr>
<td>ChReg (Regression)</td>
<td>SFReg (SpO + Regression)</td>
<td></td>
</tr>
<tr>
<td>ChUns (Power difference)</td>
<td>SFUns (cSpO + Power difference)</td>
<td></td>
</tr>
</tbody>
</table>

**M Schultze-Kraft et al**
accuracies of single subjects for the spatial-filter-based models (figure 4(c)) shows that this performance gradient is most evident for subjects with lowest classification accuracies (e.g., subjects icb, icc and ik) but is otherwise less pronounced or absent. Considering that a classification accuracy of 70% has been suggested as a minimum criterion for BCIs (Kübler et al 2001), figure 4(c) shows that this value is achieved in 10 out of 10 subjects with the SF_Cla model, followed by SF_Reg (9/10) and SF_Uns (8/10). Figures 4(a) and (b) furthermore reveal that, compared to the channel-based models, employing the spatial filtering methods CSP, SPoC and cSPoC in the spatial-filter-based models results in a performance increase of 3.8%, 4.2% and 8.2%, respectively. A statistical assessment on classification accuracies shows that this increase is not significant for classification models Ch_Cla and SF_Cla (one-sided, paired $t(9) = -1.60$, $p = 0.07$) and the unsupervised models Ch_Uns and SF_Uns (one-sided, paired $t(9) = -1.32$, $p = 0.11$) but is significant for the regression models Ch_Reg and SF_Reg (one-sided, paired $t(9) = -2.27$, $p = 0.025$).

3.3. Interpretation of components

We further aimed to investigate whether the EEG features used to train the CSP, SPoC and cSPoC filters were of cortical origin or whether they might possibly stem from ocular or other artifact sources. For this purpose we examined the spatial activation patterns that correspond to the components found by the three methods as well as the corresponding power envelopes of the components’ time series. In contrast to spatial filters, the spatial activation patterns of components can be interpreted physiologically, as they allow to draw conclusions about the spatial location of the cerebral source that generated the component’s activity (Haufe et al 2014b). Figure 5 shows the patterns and power envelopes of all three EEG methods. The components were computed on the whole data set and then projected on the band-pass filtered data that had been segmented into epochs of 10 s length. The resulting power envelope was then smoothed with a 50 s sliding boxcar window and averaged over all 32 pairs of consecutive L–H blocks. For the sake of clarity, we show results only
exemplarily for four different subjects. Subjects were selected such that they roughly represent the spectrum of classification performance in our study, from best (subject gaa), over moderate (subject lh) to lowest (subjects icc and ik). Furthermore, for the same reason we show only the component pairs that correspond to the highest value of the corresponding optimization criterion, that is, for example, the component with largest absolute-value eigenvalue in the case of CSP.

3.3.1. Spatial activation patterns. The form of the CSP, SPoC and cSPoC components’ activation patterns shown in figure 5—but also the patterns of the other components not shown in the figure, as well as all those of the other subjects (also not shown)—shows none of the characteristics of patterns related to EOG or muscle activity, hence suggesting that the components found by the three methods are of cortical origin. An examination of the patterns in the theta frequency band shows that all three methods consistently found a characteristic theta mid-frontal component, either among the component with the highest value of the corresponding optimization criterion (figure 5: cSPoC for subject gaa, all methods for subject lh, CSP and SPoC for subjects icc and ik), or among the components with lower values of the optimization criterion. Regarding the patterns of the alpha band components, although no particular consistency is observable across subjects, the spatial distribution of these patterns suggest that the sources that generated them are of cortical origin as well.

3.3.2. Power envelopes. The right column in the four panels of figure 5 shows the power envelopes of the corresponding components’ time series, averaged over all L–H block pairs. The envelopes of theta components (blue) shown in figure 5 all show a positive correlation between band power and the task difficulty, i.e. low power during the L, high power during the H condition. Also a majority amongst all other theta components used for classification shows this positive correlation. This finding is consistent with workload literature (Gevins and Smith 2003, Holm et al. 2009) and substantiates the assumption that the theta components found by the three models are sound. The power envelopes of alpha components (red), on the other hand, do not show a consistent tendency towards a negative correlation of alpha power against induced workload state, as is often reported in...
workload literature. For instance, while for subjects gaa and lh the expected negative correlation is indeed observed in all models, for subjects icc and ik the correlation has the same sign as that of the its theta counterpart.

3.4. Added value of physiological measures

Given the modulation of PPMs by the workload condition (figures 3(b)–(d)) and given that PPM features can be extracted from the data as an unsupervised signal, we assessed whether PPM features consitute an added value to the features extracted in the unsupervised models ChUns and SFUns. We first of all found that the mean classification accuracy using only PPM features was 81.8% (figure 6(a), white bar). We then repeated the cross-validation with models ChUns and SFUns, this time however augmenting the EEG features with PPM features, resulting in classification accuracies of 79.3% for model ChUns (3.6% increase) and 88.2% for model SFUns (4.3% increase). While the increase for model ChUns is not significant (paired, one-sided t(9) = -0.90, p = 0.19), the increase for model SFUns is significant (paired, one-sided t(9) = -3.03, p = 0.007). These results support the assumption that peripheral physiology can indeed provide an added value to the unsupervised model SFUns for the classification of workload.

Figure 5. Spatial activation patterns and power envelopes of components extracted by the three EEG spatial filtering methods. Shown are four exemplary results from subjects gaa (a), lh (b), icc (c) and ik (d). The shown activation patterns (scalp maps) and power envelopes correspond to the components with the highest value of the optimization criterion of the respective method. The left and middle column show the activation patterns of components obtained from the theta (blue) and alpha (red) bandpassed data, respectively. The color coding and sign of the activation patterns were adjusted to be consistent across methods and subjects but are arbitrary otherwise. The power envelopes (right column) are color coded accordingly (theta: blue, alpha: red), the x-axis shows time in seconds. Due to standardizing to z-scores, the amplitudes of the curves do not relate to discriminative power.
We found that the three predictive models SFCla, SFReg and fi models, three of which use state-of-the-art spatial supervised approach. We therefore employed six predictive progressively condensed. We investigated the limits of classifying workload states by

**4. Discussion**

**Goal of study**

We investigated the limits of classifying workload states by progressively confining the information about the experiment available to the BCI, ultimately striving for a fully unsupervised approach. We therefore employed six predictive models, three of which use state-of-the-art spatial filtering methods, on EEG data that were recorded while subjects performed a task that alternately induced low and high operator workload. In order to classify both workload conditions, all models exploit the known relationship between workload and power modulations in the theta and alpha frequency bands in the EEG. However, they differ in the type and amount of information they require about the experiment. While the classification and the regression models required either direct or indirect label information from the experiment, the unsupervised models do not require any information at all, thus constituting virtually unsupervised approaches.

**Performance of predictive models**

We found that the three predictive models SFCla, SFReg and SFUns were highly successful in their task to classify both workload states, with mean classification accuracies of 94%, 92% and 82%, respectively (figure 4). The progressive restriction of information is reflected in a decrease of classification performance. The predictive models SFCla and SFReg (and the spatial filtering methods used therein, CSP and SPoC) both represent typical examples of supervised approaches. However, while CSP and LDA need to be trained using binary labels of induced workload states, SPoC in combination with regression may use any continuous variable or measurement that reflects the induced state. In our study the error rate of subjects is an obvious candidate for a target variable for SPoC, since it was found to be clearly modulated by the workload state (figure 3(a)). However, it is likely that changes in the error rate are additionally influenced by other, perhaps more stochastic, aspects of brain activity. The error rate can thus be regarded as a noisy version of the workload condition labels. Therefore, a decline in classification accuracy as well as error rate prediction performance is to be expected when going from the SFCla model (trained with noise free workload condition labels) to the SFReg model (trained with the error rate).

The mean classification accuracy of the SFUns model showed a further decline in performance relative to that of the SFReg model. This decrease was expected as well, since—as opposed to models SFCla and SFReg—the optimization procedure in CSPoC (the spatial filtering method used in model SFUns) is agnostic of the precise experiment structure and relies solely on the known workload induced modulations of cross-frequency couplings in the EEG. To some degree this method may suffer from the fact that the sign of workload induced power modulations in the alpha band is not so clear (see discussion below). This assumption is supported when inspecting the power envelopes of the cSPoC components of subjects icc and ik (figures 5(c) and (d)). Although it is intrinsic to the optimization procedure of CSPoC to find component pairs whose power envelopes are maximally negatively correlated, the power envelopes of the cSPoC alpha components of these subjects show a rather positive correlation with the envelopes of the theta components and therefore with the workload condition. Correspondingly, those two subjects have the lowest classification accuracies with the predictive model SFUns. With a mean classification accuracy of 82% across subjects and achieving single subject accuracies above 70% in 8 out of 10 subjects, SFUns still performs remarkably well considering its unsupervised nature. Our results thus provide evidence for the general feasibility of a reliable classification of workload states using a practically unsupervised approach.

For similar reasons as stated above, the three channel-based models ChCla, ChReg and ChUns show the same decline of performance when going from using class labels to the unsupervised case. While model ChCla uses precise binary labels to combine the extracted features, model ChReg performs a regression on the error rate, a noisy version of the class labels. Finally, model ChUns comprises a fundamentally simple unsupervised approach that makes use solely of the known workload-induced, localized anticorrelation of theta and alpha power. The more interesting finding however is that in comparison to the spatial-filter-based models, the use of spatial filtering leads to classification accuracy increases of roughly 4%, 4% and 8%, respectively. Due to the small number of subjects and—particularly for the unsupervised models—due to the large spread of accuracies across subjects, this increase is not statistically significant for all models. Nevertheless, this finding substantiates the argument that the use of spatial filtering is essential for extracting meaningful features from oscillatory EEG signals and can result in considerable performance increases in the classification of workload states from EEG.

**Comparison to existing work**

Numerous studies have shown the suitability of using EEG for the decoding and prediction a variety of mental states, such as workload (Gevins and Smith 2003, Holm et al 2009, 2009,
Borghini et al (2014), alertness (Jung et al 1997) and fatigue (Stikic et al 2011). An index for workload can be derived from the absolute power in the theta and alpha frequency bands in the EEG (Gevins and Smith 2003, Berka et al 2007) or from the ratio between them (Pope et al 1995, Holm et al 2009). Other studies have additionally made use of power changes in other frequency bands such as delta, beta and gamma (Brouwer et al 2012, Christensen et al 2012). EEG-based workload indices can be further obtained using not only spectral but also features from event-related potentials (ERPs) (Prinzel et al 2003, Brouwer et al 2012, Martel et al 2014). Other studies have strives to make workload state detection robust against affective contexts (Mühl et al 2014) and day-to-day variability (Christensen et al 2012). And finally, studies have aimed at using optical measures as functional near-red spectroscopy for the assessment of mental workload (Ayaz et al 2012). It has been suggested that a continuous assessment of workload can be used to automatically moderate operator task load with the goal to enhance task performance (Parasuraman 2003, Parasuraman and Wilson 2008). Several studies have successfully demonstrated this idea, assessing mental workload as induced by cognitive effort, such as in the n-back task (Prinzel et al 2000, Scerbo et al 2003, Wilson and Russell 2007). However, to the best of our knowledge, only a few studies have undertaken the challenge of assessing operator workload as induced in many real-life environments. Kohlmorgen et al (2007) assessed the workload of subjects driving a vehicel in real traffic conditions. While they demonstrated the technical feasibility in a noisy environment, they assessed mental workload as modulated by a secondary task during driving and not by the driving task itself. Another recent study aimed at directly classifying visuomotor workload states in a simulated driving environment (Dijkstra et al 2013). Although they report very high classification accuracies their results strongly suggest that a majority of signals used for classification were movement artifacts associated with the driving task.

Using brain signals only

When validating the predictive models we placed emphasis on removing any non-cortical signals from the EEG that might be correlated with the task, hence aiming for a ‘true’ BCI in its classical meaning. It was therefore important to ensure that the EEG data were not contaminated with artifacts that could potentially confound the extracted EEG features. The task required participants to constantly execute both head and eye movements, which were both expected to be correlated with task difficulty. Thus, the two main artifact sources in the EEG are electrooculogram (EOG) and EMG activity. Because EMG activity is largely absent from frequencies below 20 Hz (Whitham et al 2007), EMG contamination was negligible in our study since the highest frequencies used for classification lay in the alpha band (8–13 Hz). EOG activity on the other hand may well pose a problem, moreover because workload has been shown to have an impact on eye blink frequency and duration (Brookings et al 1996, Veltman and Gaillard 1996). During preprocessing we therefore estimated horizontal and vertical eye movements and removed it from the EEG data via regression. An inspection of the patterns corresponding to the filters found by the methods CSP, SpOc and CSPOC shows that they all lack the characteristics of components associated with EOG activity (figure 5). This finding suggests that the signals used by the four EEG models were indeed generated by cortical sources and were not confounded by other (non-EEG) variables related to workload or to the task.

Interpretation of model components

Findings from neurophysiology indicate a relationship between theta and alpha oscillations in the EEG on the one hand and cognitive effort, task engagement and workload on the other hand (Klimesch 1999). An increase of theta power has been shown to be associated with working memory, task requirements and cognitive control, predominantly over frontal regions (Jensen and Tesche 2002, Onton et al 2005, Cavanagh and Frank 2014). Our results confirm the positive correlation with workload (figure 5). Not only did all three spatial filtering methods found at least one component that shows the characteristic frontal midline theta activation pattern, but also the power envelope corresponding to those patterns was positively correlated with the workload condition. The exact role of the alpha rhythm with respect to workload, on the other hand, is still not very clear. Findings from numerous studies have lead to the prevailing idea of alpha band synchronization as a cortical ‘idling’ mechanism. Accordingly, a decrease in alpha power has been shown to be associated with an increase in resource allocation or workload (Klimesch 1999, Keil et al 2006), thus representing a marker for workload that is opposite to that of the theta band. However, several studies have questioned a clear negative relationship between alpha power and workload, showing that alpha power can indeed increase with memory load (Jensen et al 2002) and that the exact direction depends on the specific task and the strategy of subjects (Klimesch et al 1999, Cooper et al 2003) and even depends on the precise frequency subrange (Fink et al 2005). This unclear role of the alpha band is a likely explanation for the fact that our data shows neither an apparent consistency across subjects in the shape of the activation patterns, nor in the sign of the correlation between the envelopes and the workload condition. Since the spectral features resulting from the spatial filters in the predictive models SFClm and SFlog are subsequently optimized via LDA and regression, respectively, the inconsistency in the sign of the relationship between alpha power and workload is irrelevant for these models. For the unsupervised model SFlin, however, the unsupervised subtraction of CSPOC features may have a detrimental effect on its performance.

Peripheral physiology

Previous studies have shown that workload is not only associated with changes in the EEG but also with PPMs such as heart rate (Vogt et al 2006), respiration frequency
that re

virtually unsupervised approach. While the predictive model
different states of workload by means of brain signals, using a

Our

Conclusions and outlook

findings demonstrate the general feasibility to detect
different states of workload. Previous studies have reported only small and
non-significant classification increases when fusing EEG
features with features from physiological measures, as com-
pared to using only EEG (Christensen et al. 2012, Hogervorst
et al. 2014). In contrast, we found that even an unsupervised
fusion of PPM features with the cSPoC features in model
SFUns resulted in a significant increase of classification
performance of 4.3%. A possible explanation for this disagree-
ment is that the type of workload induced in this study, which
involved constant motor engagement, has a stronger effect on
the vegetative system than the mental workload as induced by
an n-back task (Hogervorst et al. 2014). Hence, physiological
measures can indeed constitute an added value to EEG-derived
signals for the classification of workload states, even with an
unsupervised approach.

Acknowledgments

Support was provided by grants 01IB001A, 01IB001C,
01GQ0850, 01GQ0851, and 01IS14013A-E from the German
Federal Ministry of Education and Research (BMBF) and by
grant MU 987/19-1 from the German Research Foundation
(DFG). The authors declare no conflicts of interest.

Appendix. The generative model of EEG and spatial
filtering methods

1. Generative model for EEG

All of the spatial filters methods that we applied here are
based on what is called a linear forward model or linear
generative model for EEG data. The generative model states
that the observed data can be decomposed into a limited set of
components. Each component, in turn, is characterized by a
fixed spatial activation pattern and a corresponding time
course of activity. The physics of electrophysiology implies
that the scalp measurements are a linear superposition (i.e., a
weighted sum) of the individual component activities. Thus,
the mapping of component activity to the EEG recording
channels is modeled by the following equation. Let
x(t) ∈ RN denote the N-dimensional EEG recording at time
point t, then

\[ x(t) = \sum_{i=1}^{K} a_i^t \cdot s_i(t) + \epsilon(t) \]

= As(t) + \epsilon(t), \quad (4)

where a_i ∈ RN denotes the spatial activation pattern and
s_i(t) the temporal activation of the i-th component, for
i ∈ {1, ..., K}. The matrix A = [a_1, ..., a_K] ∈ RN×K contains
the spatial activation patterns of the K components in its
columns and the vector s(t) = (s_1(t), ..., s_K(t))^T ∈ RK
contains the temporal activity of all components at time point t.
Activity that is not explained by the K components is captured
by ϵ(t) ∈ RN and considered noise.

The estimation of components from the data, i.e., the
inversion of the generative model, is called backward mod-
ing. In this approach, the estimation of the time courses and
spatial activation patterns is paramterized by a so-called
spatial filter matrix, here denoted by W ∈ RN×K. An estimate
of the component time courses, denoted by ŝ, is extracted by
projecting the data onto the columns of W. Thus we have

\[ \hat{s}(t) = W^s(x(t). \quad (5) \]

An estimate of the corresponding spatial activation patterns
can be derived from the spatial filters using the filter/pattern
transformation presented in (Haufe et al. 2014b)

\[ \hat{\Lambda} = CW(W^T CW)^{-1}, \quad (6) \]

where C denotes the covariance matrix of the EEG data x.

From equations (5) and (6) it follows that the estimation
of components is reduced to finding appropriate spatial filters,
i.e., columns of matrix W. However, in order to find W
additional assumptions about the nature of (some or all of) the
components have to be made. For further details on spatial filters and spatial patterns, as well as the relationship between spatial activation patterns and EEG/MEG source localization we refer the interested reader to (Dähne et al 2015).

2. Methods for optimizing spatial filters

We now discuss a selection of methods that optimize spatial filters based on certain assumptions about the statistics of the components that are to be extracted from the data. In particular, these methods make assumptions about the spectral properties of the component time courses. It is important to note however that all of the methods we discuss below are examples of backward modeling techniques and they are all based on the linear generative model shown in equation (4).

In order to optimize spatial filters that extract oscillatory components it is useful to express the spectral power of a signal in terms of variance. Let \( P_f(\hat{s}) \) denote the spectral power of the estimated component time course \( \hat{s} \) in a frequency band \( f \). Then \( P_f(\hat{s}) \) is well approximated by first bandpass filtering the time-domain signal and then computing its variance, which we denote by \( \text{Var}(\hat{s}) \). Note that the variance of the (bandpass filtered) component time series can be expressed in terms of the spatial filter \( \mathbf{w} \) that is necessary to extract \( \hat{s} \) from the data:

\[
P_f(\hat{s}) \approx \text{Var}(\hat{s}) = \mathbf{w}^\top \mathbf{C}_f \mathbf{w},
\]

where \( \mathbf{C}_f \) denotes the covariance matrix of the bandpass filtered data.

A time-resolved expression for component bandpower can be derived by simply computing the variance in short consecutive time windows. Let these time windows indexed by the index \( e \), then we have

\[
P_f(\hat{s})_e \approx \mathbf{w}^\top \mathbf{C}_f(e) \mathbf{w},
\]

where \( \mathbf{C}_f(e) \) denotes the covariance matrix of the bandpass filtered data within the \( e \)th time window.

Spatio spectral decomposition (SSD). The aim of SSD is to optimize spatial filters \( \mathbf{W} \) for components that concentrate most of their spectral power in a given frequency band. The idea is to separate components with a ’peaky’ power spectrum from components which exhibit more of a ’1/f’ spectrum. While the former is attributed to genuine brain processes, the latter is considered background noise. Thus, it is useful to define the spectral signal-to-noise ratio (SNR) of a component for a given frequency band \( f \) as

\[
\frac{P_f(\hat{s})}{\text{Var}(\hat{s})} \approx \frac{\mathbf{w}^\top \mathbf{C}_f \mathbf{w}}{\mathbf{w}^\top (\mathbf{C}_{f-\delta} + \mathbf{C}_{f+\delta}) \mathbf{w}},
\]

SSD optimizes the above expression for a set of weight vectors under the constrained that the component time courses are decorrelated to each other. Haufe et al (2014a) have demonstrated the usefulness of SSD as a dimensionality reduction tool for the analysis of oscillatory processes in the EEG.

Source power co-modulation (SPoC). The bandpower of neural oscillations changes over time and these power dynamics have been related to switching of mental states. A method that optimizes spatial filters based on a presumed covariation between band power dynamics and an external target signal is SPoC (Dähne et al 2014a). Let the variable \( z \) denote the target signal and let us further assume that \( z \) has zero mean. Then the covariance between the bandpower dynamics of a component \( \hat{s} \) and the target signal \( z \) can be expressed in terms of the spatial filter \( \mathbf{w} \) as

\[
\text{Cov}(P_f(\hat{s}), z) = \frac{1}{N_e} \sum_{e=1}^{N_e} P_f(\hat{s})_e \cdot z(e)
= \mathbf{w}^\top \left( \frac{1}{N_e} \sum_{e=1}^{N_e} \mathbf{C}_f(e) \cdot z(e) \right) \mathbf{w}. \tag{10}
\]

The SPoC algorithm optimizes spatial filters that maximize the above expression under the constraint that extracted source time courses have unit variance and are mutually decorrelated.

Common spatial patterns (CSPs). If the bandpower dynamics of a component are to be used in a classification setting, one can base the optimization of the spatial filters on the assumed difference of bandpower between classes. This then leads to an algorithm called CSPs (Koles 1991, Blankertz et al 2008), which has become one of the corner stones of sensorimotor rhythm based BCIs (Fazli et al 2015). CSP seeks components that maximize the class difference of bandpower. Let \( P_f(\hat{s}_i) \) and \( \mathbf{C}_{f,i} \) respectively denote the power and component covariance of component \( \hat{s} \) for class \( i \in \{1, 2\} \). The CSP objective function can be formalized as

\[
P_{f,1}(\hat{s}) - P_{f,2}(\hat{s}) \approx \mathbf{w}^\top (\mathbf{C}_{f,1} - \mathbf{C}_{f,2}) \mathbf{w}, \tag{11}
\]

subject to unit variance and mutual decorrelation constraints. Note that CSP can be obtained as a special case of SPoC by encoding the class labels in the SPoC target function \( z \).

canonical source power co-modulation (cSPoC). While CSP and SPoC used label information for the extraction of relevant components, cSPoC (Dähne et al 2014b) is an unsupervised method. The method optimizes spatial filters solely based on assumptions about (positive or negative) co-variation between bandpower dynamics of individual components. In addition to \( P_f(\hat{s})_e \) let \( P_f(\hat{s})_e \) denote the time-resolved bandpower dynamics of a separate component \( \hat{s} \) at a frequency band \( \hat{f} \), with \( \hat{f} \approx f \). As before, let \( \mathbf{C}_f(e) \) denote the corresponding timeseries of covariance matrices. With these definitions the notion of bandpower covariance is formalized as

\[
\text{Cov}(P_f(\hat{s}), P_f(\hat{s})) = \frac{1}{N_e} \sum_{e=1}^{N_e} P_f(\hat{s})_e \cdot P_f(\hat{s})_e
= \frac{1}{N_e} \sum_{e=1}^{N_e} \mathbf{w}^\top \mathbf{C}_f(e) \mathbf{w} \cdot \mathbf{w}^\top \mathbf{C}_f(e) \mathbf{w}. \tag{12}
\]
cSpOc optimizes the above expression for pairs of filters w and w̃ under unit variance and mutual decorrelation constraints. The algorithm can be set to either maximize or minimize the objective function, thereby searching for components with positive or negative bandpower covariance, respectively.

References

Psychophysiological responses to changes in workload during drowsiness Measuring neurophysiological signals in aircraft pilots and car driving using EEG spectral power and ERPs in the vigilance, learning, and memory tasks Aviat. Space Environ. Med. 78 231–44
Blankertz B et al 2010 The Berlin brain–computer interface: non-medical uses of BCI technology Frontiers Neurosci. 4 198
Buzsaki G and Draguhn A 2004 Neuronal oscillations in cortical networks Science 304 1926–38
Cavanagh J F and Frank M J 2014 Frontal theta as a mechanism for cognitive control Trends Cogn. Sci. 18 414–21
Dijksterhuis C, de Waard D, Brookhuis K, Mulder B and de Jong R 2013 Classifying visuomotor workload in a driving simulator using subject specific spatial brain patterns Frontiers Neurosci. 7 149
Haufe S and Nikulin V V 2014a Dimensionality reduction for the analysis of brain oscillations NeuroImage 101 583–97
Hogervorst M A, Brouwer A-M and Van Erp J B F 2014 Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload Frontiers Neurosci. 8 322
Putze F, Jarvis J-P and Schultz T 2010 Multimodal recognition of cognitive workload for multitasking in the car 2010 20th Int. Conf. on Pattern Recognition (ICPR) pp 3748–51
Reimer B and Mehler B 2011 The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation *Ergonomics* **54** 932–42

Whitham E M et al 2007 Scalp electrical recording during paralysis: quantitative evidence that EEG frequencies above 20 Hz are contaminated by EMG *Clin. Neurophysiol.* **118** 1877–88