INVITED PAPER

Learning From More Than One Data Source: Data Fusion Techniques for Sensorimotor Rhythm-Based Brain–Computer Interfaces

This paper reviews the use of data fusion techniques for sensorimotor rhythm-based BCIs including machine learning for integration of complementary features of neural activation, multiple previous sessions, and multiple subjects.

By Siamac Fazli, Sven Dähne, Wojciech Samek, *Member IEEE*, Felix Bießmann, and Klaus-Robert Müller, *Member IEEE*

ABSTRACT | Brain–computer interfaces (BCIs) are successfully used in scientific, therapeutic and other applications. Remaining challenges are among others a low signal-to-noise ratio of neural signals, lack of robustness for decoders in the presence of inter-trial and inter-subject variability, time constraints on the calibration phase and the use of BCIs outside a controlled lab environment. Recent advances in BCI research addressed these issues by novel combinations of complementary analysis as well as recording techniques, so called *hybrid BCIs*. In this paper, we review a number of data fusion techniques for BCI along with hybrid methods for BCI that have recently emerged. Our focus will be on sensorimotor rhythm-based BCIs. We will give an overview of the three main lines of research in this area, integration of complementary features of neural activation, integration of multiple previous sessions and of multiple subjects, and show how these techniques can be used to enhance modern BCI systems.

KEYWORDS | Brain–computer interface (BCI); data fusion; electroencephalography (EEG); hybrid BCI; multi-modal; mutual information; near-infrared spectroscopy (NIRS); zero-training

I. INTRODUCTION

Brain–computer interface (BCI) studies have traditionally focussed on predicting brain states in single subjects and single feature types in isolation. Although this approach has been very productive and gave rise to a spectrum of interesting research directions and applications [1]–[7], there were clear limits of this approach: Analyzing just a single aspect or in technical terms a particular feature set in a neural recording may miss important information, for example only looking at event-related potentials (ERPs) or parts of the frequency spectrum would be insufficient to capture, say, a preparatory potential. In addition salient features can differ from subject to subject, from day to day, sometimes even from trial to trial. These challenges are among the prime reasons why statistical learning methods have become very popular in BCI research: Instead of

Manuscript received January 29, 2015; revised March 11, 2015; accepted March 14, 2015. Date of publication May 19, 2015; date of current version May 29, 2015. This work was supported by the Brain Korea 21 Plus Program as well as the SGER Grant 2014055911 through the National Research Foundation of Korea funded by the Ministry of Education. This publication only reflects the authors views. Funding agencies are not liable for any use that may be made of the information contained herein. ***(Corresponding author: Siamac Fazli.)***

S. Fazli is with the Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Korea (e-mail: fazli@korea.ac.kr).
S. Dähne is with the Machine Learning Group, Department of Computer Science, Berlin Institute of Technology, 10623 Berlin, Germany (e-mail: sven.daehne@tu-berlin.de).
W. Samek is with the Machine Learning Group, Department of Video Coding and Analytics, Fraunhofer Heinrich Hertz Institute, 10587 Berlin, Germany (e-mail: wojciech.samek@hhi.fraunhofer.de).
F. Bießmann is with Amazon, Berlin, Germany (e-mail: felix.biessmann@gmail.com).
K.-R. Müller is with the Berlin Institute of Technology (TU Berlin), 10587 Berlin, Germany and also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 136-713, Korea (e-mail: klaus-robert.mueller@tu-berlin.de).

Digital Object Identifier: 10.1109/JPROC.2015.2413993

specifying features *a priori*, machine learning methods are used to extract signals of interest automatically and could thus improve the decoding accuracy significantly. Nonetheless, there are still limits of current state-of-the-art BCI systems that restrict their wide and robust application in clinical and nonclinical contexts. The signal-to-noise ratio of neural recordings is rather low [8]–[11], which is in part due to the high inter-subject and inter-trial variability [12], [13], the limited amount of training data available, measurement noise, artifacts from paradigms and/or subject movements. In recent years, BCI researchers started to explore novel strategies in order to overcome these limitations and to increase information transfer rates as well as robustness of brain state decoders.

One such novel approach to BCI is driven by combinations of complementary recording and analysis strategies. This approach is called *hybrid BCI* and was originally defined by the authors of [14] as a BCI system that must fulfill the following four criteria: 1) The system must rely on brain signals, 2) it must be based on at least one brain signal that the user can intentionally modulate, and 3) processing must occur in real-time and 4) feedback must be provided to the user. The second criterion is often interpreted to mean more than one brain signal, which directly leads to the problem of how to optimally integrate multiple signals in order to derive a single control signal. In *sequential* hybrid BCIs, this task is overcome by switching between different BCI systems such that only one signal source drives the BCI at any given time, while *simultaneous* hybrid BCIs are based on concurrent processing of different inputs. In principle, both types of hybrid BCIs can benefit from a segment of machine learning that is called *data fusion*, because it subsumes techniques that combine information from multiple signal sources as well as associated databases [15], [16].

Typical signal combinations employed in hybrid BCI settings are electroencephalography (EEG) and electromyography (EMG) [17], [18], event related desynchronization (ERD) signals and steady state visually evoked potentials (SSVEP) [19]–[21], ERD and P300 [22], combination of SSVEP with the heart-rate [23], or near-infrared spectroscopy (NIRS) and EEG [24] among others.

In this paper, we examine techniques of combining multiple information sources in general, such as several sessions, several subjects or several underlying physiological processes. The success of complementary recording concepts in BCI is in line with their popularity in neuroimaging in general. In the last decades, combinations of multiple modalities have provided important insights into neural processing [25]–[27]. Combinations of data from multiple subjects enabled researchers to explore novel research directions, in which traditional methods fail as the stimuli used are too complex to be modeled with traditional regression techniques [28], [29].

In this review, we will give an overview over the current state-of-the-art in BCI research with a focus on data fusion techniques for sensorimotor rhythm-based (SMR) BCIs (see Fig. 1). In order to provide a broad overview we do not limit our discussion to studies that fit the definition of hybrid BCIs as given above. Instead we will cover research on data fusion in the context of BCIs that is based on the following:

1) **Different types of features**: Traditional BCIs have used only particular features known to reflect certain cognitive processes, such as EEG band-power in a specific spatial location or frequency band. Combinations of features from multiple frequency bands and spatial locations can increase information transfer rates.

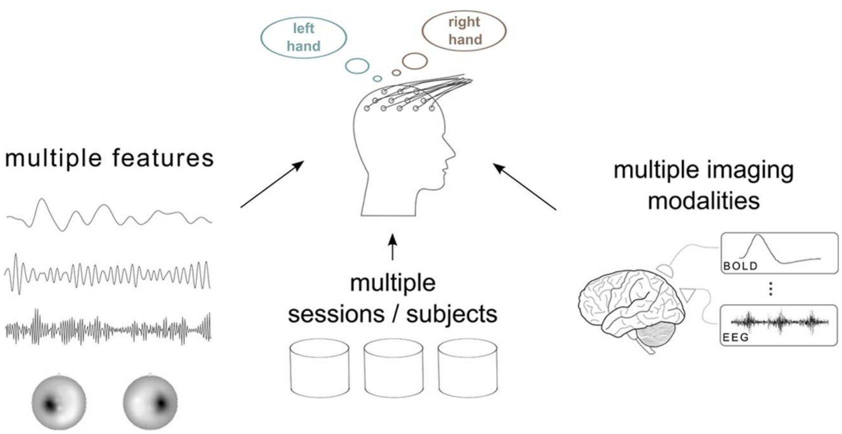

Fig. 1. ***Data fusion techniques can be applied in BCI systems that either integrate complementary features of neural activation, utilize data from previously recorded sessions and additional subjects or combine the advantages of multiple imaging modalities. The main goal is to reduce calibration times, improve classification accuracy, increase robustness, or improve the BCI system performance according to other measures. The combination of multiple EEG features or multiple neuroimaging modalities lies at the heart of hybrid BCI approaches.***

2) **Multiple sessions and subjects**: BCIs need a calibration phase to find features of interest, but these features can change over sessions. We review statistical learning techniques leveraging data from multiple sessions and subjects to increase robustness and reduce training time of BCIs.
3) **Multiple imaging modalities**: Brain activity can be measured by different techniques, e.g., changes in electromagnetic fields induced by neural discharges are picked up by EEG electrodes—but the energy demand of brain cells is not necessarily reflected in EEG recordings. However, some aspects of brain metabolism can be monitored using near infrared spectroscopy (NIRS). It is here where we find hybrid BCI studies.

The neuroimaging approaches enumerated above have a number of benefits. In the case of multiple modalities this is most obvious: each individual neuroimaging method suffers from particular limitations; EEG has high temporal resolution but due to volume condition of the head it has rather low spatial resolution; hemodynamic measurements as obtained from functional magnetic resonance imaging (fMRI) or NIRS can have high spatial resolution and measure changes in metabolism that EEG is blind to, but these measurements suffer from the delay of the vascular response limiting its temporal resolution. Combining these complementary sources of information, it becomes possible to partly overcome these limitations of single modalities [24], [30], [31].

II. BRAIN–COMPUTER INTERFACING REVISITED

BCIs are based on volitional control of neural activity. Neural activation is reflected directly in electrophysiological signals and indirectly in terms of the metabolic response of neural activation, most importantly changes in blood-oxygenation, so called hemodynamic signals. Mentally controlled localized changes in neural activity can be measured invasively and non-invasively. Although invasive recordings can offer higher signal-to-noise ratio, non-invasive recordings bear fewer risks and are thus often preferred in BCIs, in particular for healthy human users. In the following we review a small number of signatures of neural activity that are typically used in non-invasive BCIs.

A. Electrophysiological Signatures Used for BCIs

1) Sensorimotor Rhythms: One popular paradigm for voluntarily inducing different brain states is based on the modulation of sensorimotor rhythms. The μ-rhythm (10–12 Hz) and synchronized components in the β-band (16–22 Hz) are macroscopic idle rhythms that prevail over the postcentral somatosensory cortex and precentral motor cortex, when a given subject is at rest. Imaginations of movements as well as actual movements, e.g., with the right or left hand or foot, are known to suppress these idle rhythms contralaterally. This change in neural oscillation is also known as the ERD/ERS[1] effect [32] and can be detected on single-trial basis when applying state-of-the-art machine learning and signal processing methods.

2) Slow Cortical Potentials: Non-oscillatory physiological signals that are related to movement preparation/execution are the so-called slow cortical potentials (SCP) [33], [34]. SCPs are negative deflections of the EEG prior to the onset of limb movement and are also referred to as readiness potential (RP) or contingent negative variations (CNV) [35] in the literature. Interestingly, the cortical generators—and thus the spatial topography as well as the spatio-temporal dynamics—of SCPs are quite different from sensory motor rhythms [36], [37]. The resulting potential for additional information about the brain state of the BCI user makes SCPs a prime candidate for combination with SMR features.

B. Hemodynamic Signatures for BCI

Brain activations are changes in membrane potentials of neurons. In order to restore their resting state potentials, neurons consume energy, which is provided by nearby blood vessels. The induced localized changes in oxygen are an often used indirect measure of neural activity. Oxygenated blood has different light absorption and magnetic properties, which can be measured with optical and fMRI, respectively. While fMRI offers an exquisite spatial resolution and is successfully used for BCIs [38], the application of fMRI technology in many contexts is somewhat difficult, due to the size of the recording equipment. For BCIs used in everyday settings, near-infrared spectroscopy (NIRS) is a popular alternative to fMRI, as it is a relatively simple and cheap way of measuring blood oxygenation non-invasively. Hybrid BCI studies have shown that these measurements can yield information about neural signals that is complementary to electrophysiological signals obtained with EEG measurements.

III. INTEGRATION OF MULTIPLE FEATURES

A. Multiple Frequency Bands

A simplistic approach would be to compute features in one specific frequency band at a single spatial location, selected by looking at previous studies which showed that certain cognitive processes of interest are associated with this particular feature. This approach can be suboptimal. Motor imagery causes the macroscopic idle rhythms to desynchronize. However, the exact frequency range of this ERD effect is highly subject-dependent and needs to be estimated. Already very early BCI research showed that the

[1]Event-related desynchronization/Event-related synchronization.

estimation of a subject-depedent frequency range leads to highly beneficial classification accuracies [39]–[41]. A number of strategies have been proposed for the successful estimation. Some more recent methods include a heuristic [42], [43], filter-banks [44], [45] and a probabilistic Bayesian framework [46], [47]. The filter bank CSP (FBCSP) algorithm [44] computes CSP features in various frequency bands and applies a feature selection method in order to identify the most informative bands. Several strategies have been proposed for this task, among the most common is the mutual information-based best individual feature algorithm (MIBIF) [44], [48]. An extensive evaluation of this algorithm on two BCI data competition IV data sets has recently demonstrated that a data fusion BCI approach based on integration of multiple frequency bands significantly outperforms single-feature systems in terms of classification accuracy [49]. In contrast to the FBCSP algorithm, the Bayesian framework [46] does not rely on predefined filter banks (which may be chosen suboptimally), but directly constructs discriminative features in a data-driven manner. Mathematically, this approach models the frequency band as a random vector and applies particle-based approximation methods to maximize its posterior probability.

B. Multiple Spatial Locations

EEG measurements are known to have a very high temporal resolution, but poor spatial resolution due to volume conduction. Combining EEG features from multiple spatial locations can reverse the volume conduction effects to some degree. Such spatial filters for band-power features computed in narrow frequency bands are the basis for classification in SMR-based BCI systems. Using a z-transform notation for digital signals, for any trial, the band-power is computed as

$$\mathbf{y}(z) = \mathrm{var}\left[W^\top \ H(z) \ \mathbf{x}(z)\right] \tag{1}$$

where $\mathrm{var}[\cdot]$ denotes the variance operator, $\mathbf{x}$ is the raw EEG signal, $H(z)$ is a diagonal matrix of identical bandpass filter transforms and the columns of matrix W represent spatial filters. A popular technique for computing spatial filters is termed common spatial pattern (CSP) [42], [50]–[52] and allows one to focus on spatial locations with the highest ERD/ERS effect. Mathematically, a CSP filter is a projection that maximizes the variance of one class, while minimizing the variance of the other class simultaneously. In order to compute W, the CSP algorithm jointly diagonalizes the covariance matrix Σ_i of the trial-concatenated matrix of class $i \in 1, 2$

$$W^\top \Sigma_1 W = D \quad \text{and} \quad W^\top \Sigma_2 W = I - D \tag{2}$$

where I is an identity matrix and D is a diagonal matrix with entries d_i $(0 \leq d_i \leq 1)$. As eigenvalue d_i is equal to the power ratio of signals of class 1 by class 2 in the corresponding CSP filter (ith column of matrix W), best discrimination is provided by filters with very high (i.e., near 1) or very low (i.e., near 0) eigenvalues. Typically, one would retain projections corresponding to two or three of the highest eigenvalues d_i, i.e., CSP filters for class 1, and projections corresponding to the two or three lowest eigenvalues, i.e., CSP filters for class 2. For a more detailed description of CSP and its application to BCI we would like to refer the reader to [42].

The temporally and spatially filtered data may then be classified by means of simple linear classifiers, such as linear discriminant analysis (LDA). LDA assumes the classes to be normally distributed with different means μ_1 and μ_2 but with an identical covariance matrix Σ of full rank. Assuming these quantities to be known, the hyperplane, given by the normal vector $\mathbf{w}$, can be calculated by

$$\mathbf{w} = \Sigma^{-1}(\mu_1 - \mu_2). \tag{3}$$

Given that these assumptions hold, the separating hyperplane is *Bayes optimal*.

The above approach computes band-power in multiple spatial locations (specified by different CSP filters); however, it is based on one global (in the sense of all electrodes) CSP computation. The authors of [53] extended this basic approach and proposed to locally compute spatial filters, termed *CSP patches*, for multiple locations. The integration of features computed from local neighborhoods has several important advantages such as robustness and has been shown to outperform the global CSP baseline.

C. Multiple Physiological Processes

Although extracted both from EEG recordings, SMR and SCP features are very different and often analyzed independently. Perhaps the first successful attempt to combine SMR and SCP feature was reported in [54], in the context of the BCI competition 2003. The authors used a simple classification approach (LDA), which was trained on the concatenated feature vectors of gamma band activity and slow cortical activity from only two EEG channels each. The proposed approach outperformed classifiers based solely on features of either domain.

Dornhege *et al.* [55] then systematically investigated several fusion strategies to combine SMR features (extracted by standard CSP) with features obtained from SCP. Based on an independence assumption between the two involved physiological processes and modeling the extracted features using Gaussian distributions, the Bayes optimal classifier was derived. Using this classifier, performance gains [measured via information transfer rate (ITR)] of up to 50% were achieved in offline analysis, outperforming classifiers based only on SMR or SCP alone, as well as outperforming the simple feature concatenation approach presented in [54]. The results presented in [55], were then replicated by the authors of [56], which underlines the potential increase in BCI performance that can be achieved by using machine learning to optimize the combination of SMR and SCP features.

Most recently, a BCI framework that employs integration of SMR (multiple frequency bands), beta rebound, as well as SCP features was presented in [57]. This study was conducted with severely motor-impaired patients. Thus, a strong emphasis was placed on high flexibility of the BCI processing system in order to quickly adapt it to the requirements of each individual participating patient. Multiple oscillatory features of several frequency bands were automatically combined into a single classifier output. In a concurrent processing stream, the spatial-temporal features of the SCPs were channeled into a separate classification output. The classifier outputs of the two physiological processes were then combined by a meta classifier.

When a user recognizes an erroneous system response, a so-called error-related potential (ErrP) can be detected by the EEG [58]–[61]. Detecting this ErrP in real-time feedback sessions allows to improve the robustness and speed of EEG-based communication. The ErrP has initially been successfully detected in choice reaction tasks [61]–[64], but also for BCI related paradigms, such as motor imagery [65], [66], ERP-based spelling [67] and more recently for neuroprostetic applications, such as controlling an artificial arm [68].

In summary, multiple physiological processes, such as ERD/ERS, SCPs as well as ErrPs can occur during real-time motor imagery feedback sessions. Ideally, all these processes are monitored and their information combined for optimal feedback performance. Another recent line of research has focussed on combining multiple feedback paradigms, each of which depending on different physiological processes. Here the subject needs to focus on two (or more) mental tasks simultaneously. The combination of an SMR-based paradigm with SSVEP has previously been employed for the application of a brain switch [19], [69]. While the SSVEP signal quality remained unchanged, it lead to impaired ERD effects [20]. Nonetheless, the authors of [20] and [21] find that this type of hybrid approach is beneficial for the majority of subjects, especially for those with poor prior classification accuracy (so-called *BCI illiterates* or weak BCI performers [70], [71]). A number of researchers have also looked into the possibility of combining motor imagery with a P300-based paradigm [22], [72], [73] for applications such as 2-D cursor control and wheelchair operation, among others.

IV. INTEGRATION OF MULTIPLE SESSIONS AND SUBJECTS

In recent years, a number of approaches have been established, that enable users to start a high-speed BCI feedback session without the need of recording any calibration data [44], [48], [74]–[76]. These so-called *zero-training BCI systems* have large advantages for patient studies where any additional recording session is associated with significant costs and patient effort. Apparently, many other applications, e.g., computer games, also benefit from off-the-shelf BCI technology that works with any user at any time. Besides the development of zero-training BCI systems there are also other reasons why one would profit from integration of data from multiple subjects and sessions, e.g., robustness to artifacts and nonstationarity. The following paragraph will review a number of recent approaches that improve robustness and reduce calibration time leveraging data from multiple subjects and sessions.

A. Session-to-Session Transfer

The first of such approaches aimed at reusing spatial filters from previous sessions of the same user, a so-called *session-to-session* transfer [74]. Since CSP filters are subject-specific, similar filters should be found across different sessions for a given subject. It can be assumed that regions with a high density of CSP filters, so-called *clusters*, contain examples which are particularly stable and informative across sessions [77], [78]. Since CSP filters are solutions to a generalized eigenvalue problem and any multiple of the eigenvector is also a solution, it is therefore sufficient to consider only normalized CSP filters on the $(C-1)$ dimensional hypersphere [74]. The CSP space is inherently non-euclidean and an appropriate metric between two columns $\mathbf{w_1}$ and $\mathbf{w_2}$ of a CSP filter matrix W is the angle between the two lines these vectors form:

$$m(\mathbf{w_1}, \mathbf{w_2}) = \arccos\left(\frac{\mathbf{w_1}\mathbf{w_2}}{|\mathbf{w_1}||\mathbf{w_2}|}\right). \tag{4}$$

These distances can then be used to find clusters within the CSP filter space. Points that are located at cluster centers can then be selected as typical CSP filters by using a so-called γ-index [74].

An alternative approach to construct session-independent spatial filters is based on regularization. The authors of [79] recently proposed a divergence-based framework for spatial filter computation and showed that different regularization schemes can be easily implemented within this framework. Mathematically, spatial filter computation reduces to the following maximization problem:

$$\begin{aligned} &\arg\max_W\ (1-\lambda)\tilde{D}(x_1 \parallel x_2) - \lambda\Delta \\ &\text{with } x_1 \sim \mathcal{N}(0; W^\top \Sigma_1 W) \\ &\text{and } x_2 \sim \mathcal{N}(0; W^\top \Sigma_2 W) \end{aligned} \tag{5}$$

where Δ is a regularization term, λ is a tradeoff parameter, $\tilde{D}(p\|q)$ is a symmetric divergence between probability distributions p and q, and $\mathcal{N}(0; W^\top \Sigma W)$ denotes a zero mean Gaussian distribution with covariance matrix $W^\top \Sigma W$. The authors of [79] prove that this formulation reduces to CSP when using symmetric Kullback–Leibler divergence and $\lambda = 0$. Spatial filters can be computed solely on historical sessions by setting λ to 1 and using the following regularization term:

$$\Delta = -\sum_i \tilde{D}\left(\mathcal{N}\left(0; W^\top \Sigma_1^i W\right) \parallel \mathcal{N}\left(0; W^\top \Sigma_2^i W\right)\right) \quad (6)$$

where Σ_1^i and Σ_2^i are class-covariance matrices from the ith historical session. The spatial filters maximizing the above objective try to maximize the average ERD/ERS effect on the historical recordings.

Unfortunately, this approach is not very robust to "outlier sessions," i.e., instead of finding spatial filters that work well for the majority of the sessions one may obtain a solution which works extremely well for few outlier sessions (e.g., due to artifacts) but fails for the majority of other sessions. By using a different divergence, namely the symmetric Beta divergence [79], we can significantly robustify the solution. Note that conceptually this approach is related to the clustering method (which is also robust to outliers) that we presented before.

Robust filters obtained with the above methods allow to construct *Zero-Training* classifiers, which use calibration data recorded days before the actual BCI experiment. However, for optimal feedback performance the bias of the classifier needs to be adjusted. Since the mental state of the user can be very different during the feedback phase compared to the calibration phase, also the non task related brain activity differs [80]–[82]. The authors of [74] showed that by using the first 20 trials for bias adaptation one can construct zero-training systems that show no significant difference in the feedback accuracy compared to calibration-based systems.

Another approach to more robust spatial filters, the *invariantCSP* method [83], aims to increase robustness to artifacts such as eye movements or blinks by utilizing recordings from a so-called artifact session. The spatial filters computed on the calibration data are regularized away from activity which occurs in the artifact session, i.e., when explicitly inducing different types of artifacts in the EEG.

B. Subject-to-Subject Transfer

The previous section discussed *session-to-session* transfers. With these type of algorithms is it possible for expert BCI users to start feedback sessions without the need of recording any prior calibration session. In the following section we will review more recent algorithms that enable expert as well as novice BCI users to start feedback sessions without any prior subject-dependent calibration data.

For high-speed real-time feedback it is of paramount importance to identify a subject-dependent optimal band-pass filter. While a number of methods have been proposed, if a calibration dataset is available [42], [43], [46], [47], these cannot be applied when there is no calibration data present. One possibility would be to choose a broad-band filter or one that is centered around the μ-rhythm, but previous work has shown that any one specific filter will lead to suboptimal performance, if applied to a number of subjects [75]. An alternative that has proven successful is an ensemble based framework, which construct finite collections of individually weak classifiers from potentially very large ensembles [84]–[86]. Here, a large number of basis functions are generated from a database of many subjects to generate a *subject-independent Zero-Training* classifier [75], [87], [88].

Each dataset is first preprocessed by a number of predefined temporal filters (i.e., band-pass filters) in parallel (see upper part of Fig. 2). These temporal filters are chosen with prior knowledge from neurophysiology: most of them lie within the μ-band, some in the β-band, some in between μ and β-band and one broadband 7–30 Hz filter is included. For each basis function a spatial filter (CSP) as well as a linear discriminant analysis (LDA) classifier is estimated. The appropriate weighting of the basis functions is highly important for obtaining a low generalization error on unseen subjects and needs to be estimated only on the training data, termed *leave-one-subject-out (LOSO)* cross-validation (see lower part of Fig. 2). In order to ensure maximal interpretability ℓ_2 regression with a ℓ_1 norm as a regularizer is chosen, which is known to lead to a high level of sparsification [89], [90]:

$$\begin{aligned}\underset{w_{ij}^{(k)}}{\arg\min} & \sum_{x \in X \setminus X_k} \left(h_k(x) - y(x)\right)^2 \\ & + \alpha \sqrt{\sum_{i=1}^{B} \sum_{j \in S \setminus S_k} \sum_{x \in X \setminus X_k} c_{ij}(x)^2} \\ & \times \left(\sum_{i=1}^{B} \sum_{j \in S \setminus S_k} \left|w_{ij}^{(k)}\right| + |b|\right) \end{aligned} \quad (7)$$

$$h_k(x) = \sum_{i=1}^{B} \sum_{j \in S \setminus S_k} w_{ij}^{(k)} c_{ij}(x) - b \quad (8)$$

where $c_{ij}(x) \in [-\infty; \infty]$ is the continuous classifier output, before thresholding, obtained from the session j by applying the bandpass filter i, B is the number of frequency bands, S the complete set of sessions, X the complete data set, S_k the set of sessions of subject k, X_k the dataset for subject k, $y(x)$ is the class label of trial x, and w_{ij}^k in (8) are

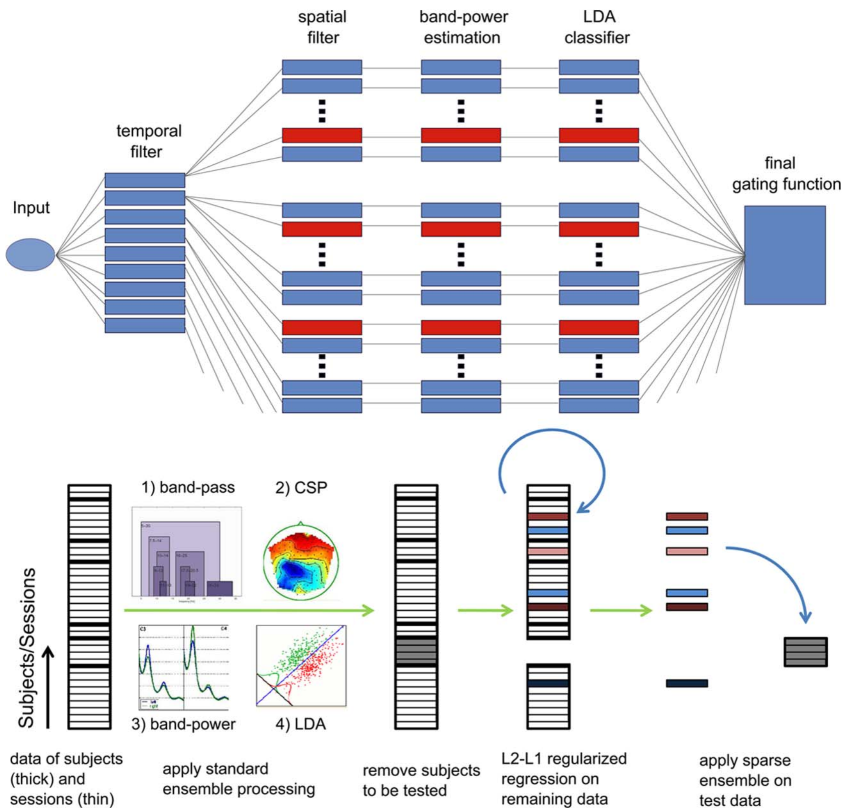

Fig. 2. *A subject-independent zero-training BCI system. Upper part: Flowchart of the ensemble processing. Data is fed through a bank of temporal and spatial filters as well as LDA classifiers in parallel, followed by a final gating function. Red patches illustrate inactive nodes after sparsification. Lower part: Flowchart for the leave-one-subject-out (LOSO) cross-validation. Figure adapted from [88].*

the weights given to the LDA outputs. The hyperparameter α in (7) is varied on a logarithmic scale and multiplied by a dataset scaling factor which accounted for fluctuations in voting population distribution and size for each subject. The dataset scaling factor is computed using $c_{ij}(x)$, for all $x \in X \setminus X_k$. For computational efficiency reasons the hyperparameter can be tuned on a small random subset of subjects whose labels are to be predicted from data obtained from other subjects such that the resulting test/train error ratio is minimal, which in turn affects the choice (leave in/out) of classifiers among the basis functions. Following this methodology allows to obtain a very sparse set of voting classifiers which perform as well as standard, state-of-the-art subject calibrated methods. While spatial and temporal filters can be estimated for unseen subjects, the bias still needs to be set, prior to the start of an online experiment. This could be done similarly as previously discussed in Section IV-A.

As explained in Sections IV-A, utilizing data from other subjects may also be valuable for reasons such as robustness. The authors of [91]–[94] use other subjects' data as regularization target for covariance matrix estimation and demonstrate significant improvements when data is scarce. The authors of [95] jointly train the spatial filters of several subjects by applying a multi-task learning algorithm. A Bayesian method for subject-to-subject information transfer has been proposed in [96] and [97]. An unsupervised BCI based on inter-subject information has been proposed in [98]. Finally, as mentioned before, the authors of [79] and [99] transfer information about nonstationarities in the data between subjects.

C. Integration of Multiple Subjects in the Absence of Stimulus Information

An avenue that is less pursued and yet rather promising is the unsupervised training of subject-specific models by integrating information from multiple subjects or from multiple concurrent processes present in individual subjects. Traditionally, neuroscientific studies, including BCI research, need a target variable, in order to analyze brain activity. In many settings, we cannot model this target variable. A prominent example are complex stimuli, such as naturalistic movies. Despite decades of research on the visual system, how exactly a movie is processed, i.e., *which stimulus features* are relevant for neural processes evoked by subjects watching a movie, remain unclear. This is why analysis of brain data in response to realistic stimuli is still an active area of research. A popular approach to these kind of experiments is to assume that whatever neural activity is evoked by complex stimuli, it is fair to assume that the temporal signature of these neural processes is similar across subjects. Using this assumption, models can be trained to decode the mental states of subjects exposed to complex stimuli without modeling the stimulus explicitly.

Inspired by the works of [28], the authors of [100], [101], and [29] apply *canonical correlation analysis* (CCA) [102], or versions thereof, to find components in the neuroimaging data that exhibit maximally correlated activation across subjects. In these examples, part of the brain activity is correlated across subjects because the subjects were exposed to the same stimulus. Note that these approaches make no assumption about the relation between stimulus and brain response. Relevant components are extracted solely based on the assumption that brain activity be consistent across subjects. This assumption is formalized in the following objective function:

$$\operatorname*{arg\,max}_{\mathbf{w}_\mathbf{x},\mathbf{w}_\mathbf{y}} = \operatorname{corr}\left(\mathbf{w}_\mathbf{x}^\top \mathbf{x}, \mathbf{w}_\mathbf{y}^\top \mathbf{y}\right) \tag{9}$$

where $\mathbf{x}$ and $\mathbf{y}$ are data points which can be from two subjects who were exposed to the same stimulus, or data from a single subject but from two separate expositions to the same stimulus (e.g., two viewings of the same movie). The coefficients of $\mathbf{w}_\mathbf{x}$ and $\mathbf{w}_\mathbf{y}$ are optimized such that the projections $\mathbf{w}_\mathbf{x}^\top \mathbf{x}$ and $\mathbf{w}_\mathbf{x}^\top \mathbf{x}$ yield maximal (positive or negative) correlation. The objective presented in (9) can be transformed into a generalized eigenvalue problem and thus solved efficiently using standard software packages.

In the studies presented in [100] and [101], the authors constrained $\mathbf{w}_\mathbf{x}$ and $\mathbf{w}_\mathbf{y}$ to be identical. Using this variant of CCA, the authors were able to extract physiologically meaningful components from ongoing EEG activity of several subjects who viewed the same movie stimuli. Interestingly, peaks in time-resolved intra- and inter-subject-correlation between the extracted EEG activity co-occurred with arousing moments in the movie stimuli and thus point to markers of emotionally laden attention. In a follow-up study on expression of interest in television programs, the authors of [101] were able to show a strong correspondence between inter-subject-correlations on the one side and social media activity as well as audience ratings on the other. Using hemodynamic signals obtained by functional magnetic resonance imaging and a methodology based on CCA, the authors of [29] show that inter-subject-correlations are predictive of the degree of immersion and thus a marker of how strongly a stimulus is experienced by human observers. Taken together, these findings lay the foundation for brain–computer interfaces and mental state monitoring based on correspondence between signals of groups of subjects, rather than individual subjects.

A nonlinear version of canonical correlation analysis is the recently developed *canonical source power correlation* analysis (cSPoC, [103]). Similar to CCA, cSPoC finds a set of spatial filters that maximize a correlation coefficient. However, the crucial difference between CCA and cSPoC is that in case of cSPoC, the correlation coefficient is not defined between the projected signals but between nonlinear functions thereof. Specifically, cSPoC optimizes the correlation between the envelopes of the projected signals, which is useful if the recorded data is of oscillatory nature and the changes in spectral properties are the informative aspects. The cSPoC objective reads

$$\operatorname*{arg\,max}_{\mathbf{w}_\mathbf{x},\mathbf{w}_\mathbf{y}} = c_{\text{sg}} \cdot \operatorname{corr}\left(\Phi\left(\mathbf{w}_\mathbf{x}^\top \mathbf{x}\right), \Phi\left(\mathbf{w}_\mathbf{y}^\top \mathbf{y}\right)\right) \tag{10}$$

where $\mathbf{x}$ and $\mathbf{y}$ are bandpass filtered multivariate datasets, the constant $c_{\text{sg}} \in \{+1, -1\}$ decodes whether positive or negative correlations are desired and the function $\Phi(\cdot)$ represents the envelope of its argument.

The structure of classical SMR-based BCI paradigms induces negative correlations between the spectral power of brain rhythms. This is due to the fact that ideally during a given trial only the neural source coding for the desired class of this trial should exhibit an ERD/ERS, while the neural source coding for the other class should be in an idle state, and vice-versa for trials of the other class. The authors of [103] make use of this fact to show that relevant physiological sources can be extracted from the EEG even without knowledge of class labels. Instead, they optimized spatial filters using the cSPoC objective in (10) to maximize negative correlations between the envelopes of the extracted components. Fig. 3 shows the three most class-discriminative components obtained with CSP and cSPoC from a representative subject. The resulting components bear strong similarity with those obtained from standard CSP (which requires label information) and yield comparable—though not improved—classification accuracy in a population of 80 subjects. Nonetheless, these results point to an interesting alternative approach in order to extract oscillatory neural sources that can drive a BCI application.

V. INTEGRATION OF MULTIPLE IMAGING MODALITIES

There are many ways in which recording brain activity with one neuroimaging modality alone can lead astray researchers' conclusions. EEG measurements require neural discharges to happen in a highly synchronized manner across large populations of neurons, and EEG will only detect these discharges if the population has a certain orientation with respect to the recording electrode. Certain cell types, which do not have an appropriate

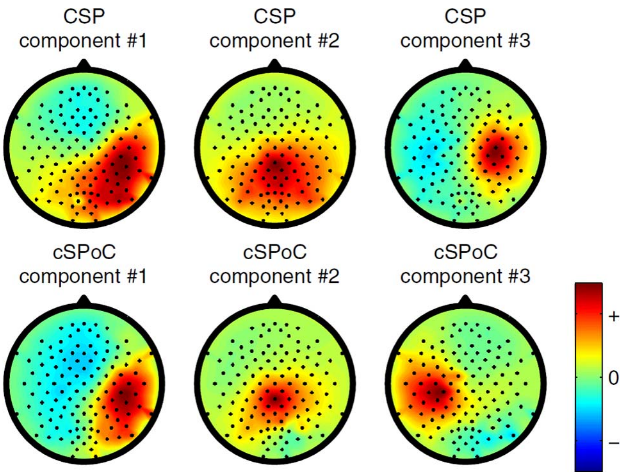

Fig. 3. ***Spatial activation patterns of components extracted from the same representative subject using CSP (top row) and cSPoC (bottom row). The data was acquired during a classical SMR-based BCI paradigm in which during each trial the subject was cued to imagine movement of either the right or left hand. CSP makes use of class label information while cSPoC does not. Yet for both methods, the resulting components are consistent with neural sources in sensorimotor-related as well as parietal areas of the brain. Figure adapted from [103].***

morphology cannot be measured at all, but no matter how neurons are oriented, neural activation will induce metabolic activity which can be measured in the hemodynamic signal. This does not mean that it would be better to use hemodynamic activity only—many changes in hemodynamic activity are not related to neural activity at all; hemodynamic signals are largely affected by changes in blood pressure, blood volume and blood oxygenation that is simply due to breathing or also to the amount of coffee drunken before the experiment [104]. Because these neuroimaging modalities offer complementary views on neural activation, combinations of neuroimaging techniques have become increasingly important in basic research and clinical applications. In basic research, integration of multiple modalities were extremely important to establish a better understanding of what the single modalities actually measure [25]–[27]. In clinical research, multi-modal neuroimaging is a very promising approach for better localization of epileptic seizures: EEG has the spatial resolution to detect the seizure, and fMRI offers the spatial resolution to localize it. These needs for multi-modal recordings in clinical and basic research were the basis for the advances in recording technology [105]–[107] and analysis strategies [108], [109] that gave rise to their increasing popularity in BCI research [110], [111]. A recently proposed multi-modal setup for BCI combines EEG with NIRS [24], [112]. Both recording setups are relatively cheap and do not require expensive infrastructure, as is the case for fMRI. In the following we review some work that illustrates the advantages of multi-modal recordings for BCIs.

A. NIRS-EEG

The successful combination of data from multiple domains has motivated researchers to also examine the potential benefits of combining near-infrared spectroscopy (NIRS) with EEG for the purpose of BCI [24]. NIRS measures concentration changes of oxygenated and deoxygenated hemoglobins ([HbO] and [HbR]) in the superficial layers of human cortex. The concentration of [HbO] is expected to increase after focal activation of the cortex due to higher blood flow, while [HbR] is washed out and decreases [113], [114]. NIRS measures a comparable effect to the blood oxygenation level dependent (BOLD) contrast in fMRI, since also for fMRI the washout of [HbO] is the major constituent [105]. At the same time it compares favorably to fMRI in terms of low costs, portability and easiness to handle [115].

In a recent study, simultaneous measurements of NIRS and EEG (Fig. 4(a) and (b)) were recorded during a real-time SMR-based BCI feedback experiment (Fig. 4(c)) [24]. The feedback during recording was EEG-based, but in offline simulations the classification results for each signal domain were evaluated separately, and in addition also for their combination. The combination was achieved by means of meta-classifiers, which weight the linear classifiers of the individual measurements according to training data samples [24]. While the EEG classification accuracy across subjects is superior, when compared to [HbO] as well as [HbR] (see Fig. 4(d)), the results show that usage of meta-classifiers, combining features from EEG and NIRS, can improve the classification accuracy of SMR-based BCI systems (Fig. 4(e)). This increase is obtained in over 90% of the considered subjects and led to a significant performance increase of 5% on average.

To examine the degree of independence between the NIRS and EEG-based classifier outputs, their outputs can be restricted to values 0 and 1 and their mutual information $\mathcal{I}$ can be estimated. Mutual information is an information theoretic measure, which estimates the information that two random variables share. It can be expressed in terms of conditional entropies of random variables X and Y:

$$\mathcal{I}(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \qquad (11)$$

The conditional entropy $H(X|Y)$ quantifies the remaining entropy of X, after the value of Y is known. If $H(X|Y) = H(X)$, then $\mathcal{I}(X;Y) = 0$: the variables are independent. On the other hand, if X and Y are identical, then $H(X|Y) = 0$ and hence $\mathcal{I}(X;Y) = H(X)$. $\mathcal{I}(X;Y)$ is symmetric and its values are in the range of 0 and 1: $\mathcal{I}(X;Y) = \mathcal{I}(Y;X) \in [0;1]$ [116].

The left part of Fig. 4(f) shows $\mathcal{I}(\mathrm{EEG};\mathrm{NIRS})$ as a function of classification accuracy of [HbO] and [HbR]. The mutual information rises with increasing accuracy of

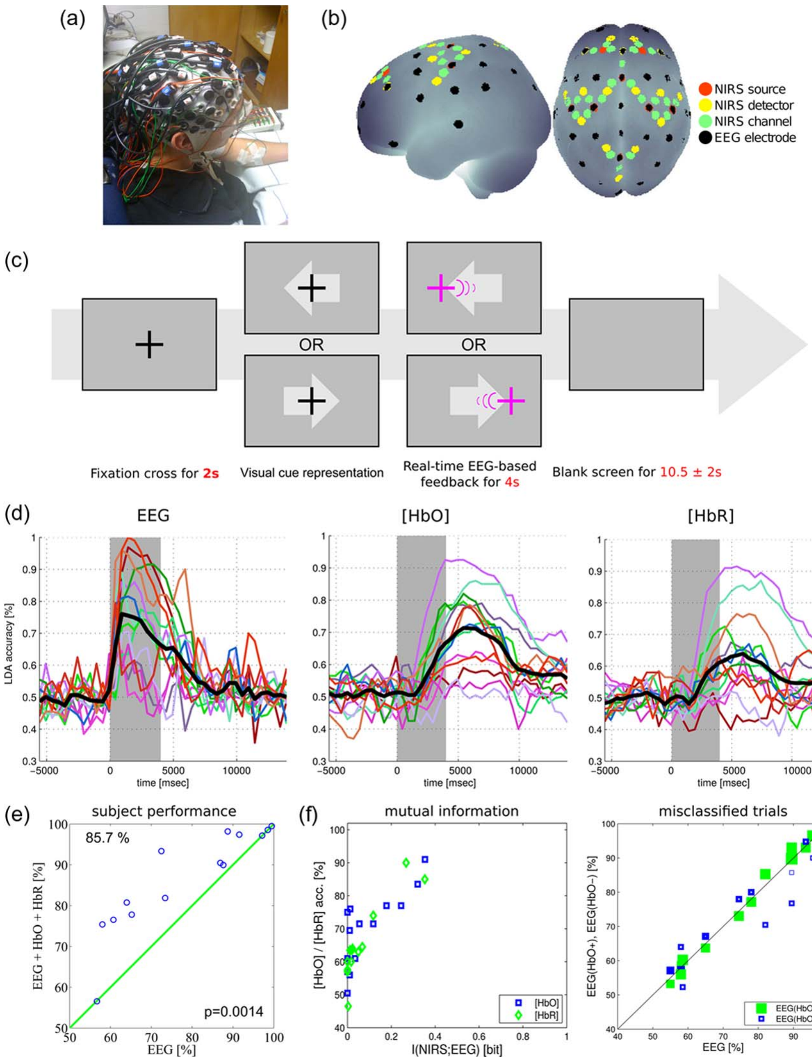

Fig. 4. ***BCI system based on EEG and NIRS. (a) and (b) EEG Electrode and NIRS optode setup. (c) Paradigm flowchart. (d) EEG and NIRS classification accuracy [%] (LDA) computed within a sliding windows of 1 s length. The x-axis denotes the center of the moving window. Colored lines show the accuracy for the single subjects while the black line is the average over subjects. The gray bar indicates the time interval of cue presentation. (e) Scatter plot comparing classification accuracies and significance values of EEG (x-axis) versus EEG + [HbO] + [HbR] (y-axis). (f) Left: Mutual information of EEG and NIRS classifier outputs (x-axis) is compared to [HbO]/[HbR] classification performances (y-axis). Each triangle/square pair represents the results of a single subject. Right: Compares EEG classification accuracy of all trials to EEG classification accuracy of trials, where [HbO] was correct/incorrect. The sizes of squares encodes the number of trials.***

the NIRS chromophores. Similarly, the mutual information also rises with increasing EEG accuracy (not shown here, please refer to [24]). This makes intuitive sense, since with increasing accuracy both classifier outputs will correctly predict and thus share the majority of class labels. However, please note that the mutual information does not reach values above 0.4 bit for any subject. To investigate, whether EEG and NIRS classifiers misclassify the same trials, the EEG classification accuracy of all trials is plotted in relation to the EEG classification accuracy of trials, where [HbO] was correct/incorrect (see Fig. 4(f), right part). As can be seen the EEG classification is largely invariant to the classification performance of [HbO] (and [HbR], not shown). In other words, EEG and NIRS mostly misclassify different trials. It can therefore be deduced that the individual methods in fact complement each other in terms of *information content*.

A more recent study [112] investigated long-term training effects of SMR-based BCIs and found significant enhancement of activation patterns in form of an [HbO] increase (NIRS) as well as a stronger ERD in the upper β-band (EEG) over the course of ten sessions.

B. EEG-EMG

While BCI technology has seen swift advances in recent years, information transfer rates are still not on par when compared to non-BCI control. For patients who still retain some control of their body, these functions should be employed in parallel with the BCI system in order to increase robustness and interaction speed of their communication with the outside world. In their paper, Leeb *et al.* [17], [18] examine the possibility of parallel usage of EEG and electromyographic (EMG) activity, whereby the control abilities of both channels are fused. In their synchronous SMR-based BCI paradigm they test two fusion techniques: one where the classifier outputs of both modalities are balanced equally and another, where they are combined with naïve Bayesian approach [117]. According to Bayes' rule

$$P(C|O_1, O_2) \propto P(C)P(O_1, O_2|C) \tag{12}$$

where C denotes the class (left or right) and O_1 and O_2 are the classifier outputs of EEG and EMG, respectively. If independent sources are assumed (12) becomes

$$P(O_1, O_2|C) = P(O_1|C)P(O_2|C). \tag{13}$$

Then

$$C_{out} = \arg\max_{c} P(C = c)P(O_1|C = c)P(O_2|C = c) \tag{14}$$

where $c \in$ left, right. The prior $P(C = c)$ can be assumed to be equal, if classes have similar size and $P(O_1|C = c)$ and $P(O_2|C = c)$ can be estimated from the training data.

Their results show that the multimodal fusion approach of muscular and brain activity yielded better and more stable performance compared to the single conditions [18]. In addition they show that increasing muscular fatigue leads only to moderate degradation of performance, a common phenomenon for early stage amyotrophic lateral sclerosis patients. The system thus offers users reliable BCI control, even though she/he is getting more and more exhausted or fatigued during the day.

VI. WHICH HYBRID BCI AND/OR WHICH DATA FUSION TECHNIQUES SHOULD BE SELECTED WHEN?

Let us now briefly analyze and discuss hybrid BCIs from an abstract systems perspective. When do we expect a BCI, which relies on combinations of complementary analysis as well as recording techniques, to perform better or worse than a classical unimodal system? What are the essential factors for this and how can a practitioner extract this information to design the setup which is required of his hybrid BCI from the given BCI data?

Let us consider two (complementary) variables X_1 and X_2, that are driven by some unknown neural process Z. Then X_1 and X_2 can be combined to predict a target variable $Y \in \{\pm 1\}$, which drives the latent neural process, if their mutual information is low but they both carry meaningful information $\mathcal{I}$ with respect to the target Y. Mathematically, these two conditions can be summarized as

$$(i) \quad \mathcal{I}(X_1, X_2) \text{ is low} \tag{15}$$

$$(ii) \quad \mathcal{I}(X_1, Y) \text{ and } \mathcal{I}(X_2, Y) \text{ are high} \tag{16}$$

where $\mathcal{I}$ denotes the mutual information between variables, see (11).

Note that if both variables X would be more and more correlated then the gain achieved to predict Y will become lower and lower due to their dependencies. Similarly, X_1 and X_2 could contain different signal-to-noise ratios, an optimal predictor combining both will gauge them according to this informativity. In the extreme, only the high SNR variable will be considered while the low SNR variable will be ignored. If SNR and information content are known, then Bayesian inference allows for optimal combination. Under the assumption that the variables X_1 and X_2 are Gaussian distributed with equal covariances, one can show that the expected misclassification risk r_i of the variable X_i is

$$r_i = g\left(\frac{\mu_i}{\sigma_i}\right) \tag{17}$$

with

$$g(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} \exp\left(-\frac{1}{2}(x - z)^2\right) dx \tag{18}$$

where μ_i and σ_i are the parameters specifying the Gaussian distribution of classification scores under the optimal Bayes classifier. The constructed classifier for the combined features is also Gaussian distributed with expected misclassification risk

$$r = g\left(\frac{\mu_i}{\sqrt{2}}\right) = g\left(\frac{g^{-1}(r_1) + g^{-1}(r_2)}{\sqrt{2}}\right). \tag{19}$$

Thus, if both variables have the same misclassification risk $r_1 = r_2$, then the SNR increases as $\sqrt{2}$. If, on the other hand, one variable is very uninformative due to low a SNR, then it does not contribute to the decrease of r and should thus not be included. Finally, X_1 and X_2 may be informative with respect to Y but with different temporal signature, i.e., NIRS and EEG for BCI. Then it is important to find a predictor that makes optimal use of combining informative high SNR in time. A suboptimal timing in combining the variables may in the extreme render one variable useless.

To examine the potential benefits of the proposed combinatory approaches in terms of mutual information we would like to point the readers' attention to Fig. 5, where correlation coefficients (top row) as well as mutual information (bottom row) is depicted for a number of feature combinations. The left column shows the combination of channel-wise band-power of common spatial pattern components (CSP), channelwise band-power (BP), and slow cortical potential (SCP) features (as previously described in Section III-C). As one may expect, CSP and BP features show relatively high correlation coefficients/mutual information, while CSP and BP do not share information with SCP. The middle column shows features from a subject-to-subject transfer. The considered trials stem from subject 3 (S3). While features derived from subject 3 themselves have high correlation coefficients/mutual information with each other, features derived from other subjects' spatial filters and classifiers (please refer to Section IV-B for more details on this procedure) have mostly low correlation coefficients/mutual information. In contrast, features from subjects S5, S8, and S4 show high correlation coefficients/mutual information with S3. In other words, the features derived from the data of these subjects are viable candidates for predicting trials of S3. The right column shows feature combinations from the hybrid neuroimaging study. Band-power features have low correlation coefficients/mutual information with [HbO] as well as [HbR]. As can be seen in the upper part [HbO] and [HbR] are anticorrelated, which is in line with current knowledge from neurophysiology. Note that this effect is not visible in the lower part.

Let us now briefly discuss the BCI systems presented in this paper in the light of the systems reasoning put forward in this section. Multimodal neuroimaging as in NIRS-EEG

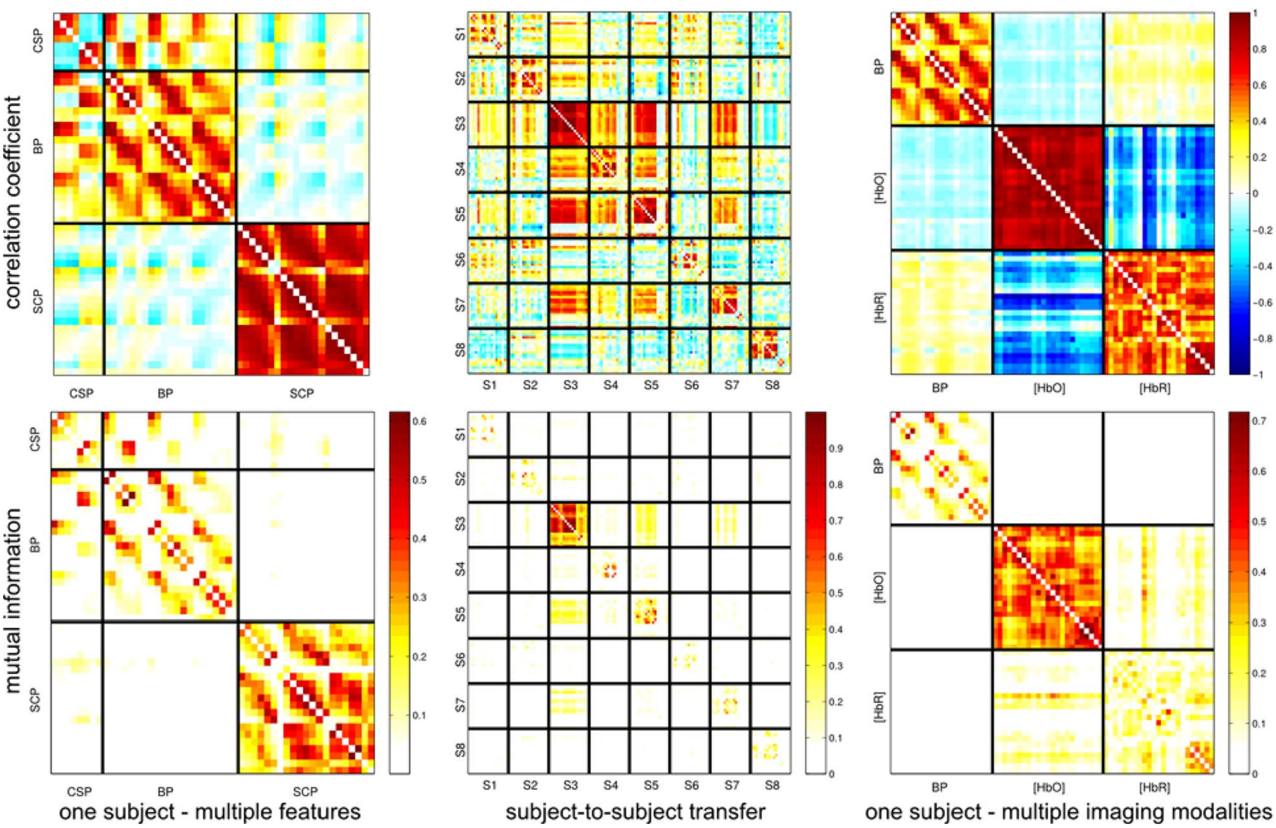

Fig. 5. ***Correlation coefficients (top row) and mutual information (bottom row) for features of various BCI approaches, where data from multiple sources is available. Left column: Combination of channel-wise band-power of common spatial pattern components (CSP), channelwise band-power (BP) and slow cortical potential (SCP) features. BP and CSP features share common information (high correlation coefficients), but both do not share information with SCP (low correlation coefficients). Middle column: Features from subject-to-subject transfer. Trials are from subject 3 (S3). Features derived from subject 3 themselves have high correlation coefficients/mutual information with each other, while features derived from other subjects' classifiers have mostly low correlation coefficients/mutual information, exceptions are features from subjects S5, S8, S4. Right column: BP has low correlation coefficients/mutual information with [HbO] as well as [HbR]. [HbO] and [HbR] are anticorrelated.***

or fMRI-EEG result in significantly increased costs, setup times and possibly calibration times. Their benefits in terms of classification accuracy should therefore be carefully evaluated. In extreme cases, where BCIs serve as a communication tool for severely disabled or patients in a vegetative state these costs should of course be a minor concern. While the classification accuracy of NIRS-EEG systems have been shown to be superior to sole EEG, the intrinsic hemodynamic delay of NIRS (and fMRI) pose significant constraints on the information transfer rates of such systems. However, some very recent results indicate that this intrinsic delay may be diminished substantially [111], [118]–[120]. The NIRS-EEG hybrid BCI had to be carefully gauged to reflect the two signals different spatiotemporal resolution and reliability. Generically EEG is the main resource for decoding, although NIRS can be very usefully added to catch difficult trials and to correct errors.

VII. DISCUSSION AND CONCLUSIONS

Integrating complementary sources of information about neural activity in a meaningful way can significantly increase the overall amount of information extracted. This is why data fusion techniques have been highly successful in neuroimaging in general and in brain–computer interfacing in particular. For BCI this amounts to performance gains in subject communication and cognitive state decoding. Neuroimaging in general can leverage multimodal recording and analysis setups for better medical diagnosis or an improved understanding of neurophysiological processes.

In this paper, we examined how gathering data from different sources and applying data fusion techniques can improve BCI performance as well as our understanding of the underlying physiology. Classically the term hybrid BCIs [14] implies the usage of multiple signal types/sources (e.g., different aspects of EEG only, or combinations of EEG and EMG, or NIRS, and EEG etc.). While not all systems that employ fusion techniques strictly fall into the hybrid BCI category, we have seen that exploiting complementary information that was previously analyzed in isolation can boost the performance (e.g., [17], [20], [24], [55]), robustness (e.g., [57], [79], [94], [99]), and ease of use (e.g., [44], [74]–[76]).

While we have mainly focused on sensorimotor rhythm-based BCI paradigms, the concepts introduced and reviewed do also transfer to other BCI systems (such as, e.g., ERP based BCIs). Note furthermore that many of the concepts discussed in this review are ubiquitous in signal processing, robotics and machine learning, under keywords such as information integration [121], [122], sensor fusion [16], [123]–[125], or simply as data fusion [16], [126]–[128], among others.

Finally, we have outlined systematic information theoretic measures that can be used to judge and quantitatively reason upon the usefulness and limits of systems based on data fusion in general and hybrid BCIs in particular. Future work will focus on developing a generic BCI framework in the spirit of [129] and on also including concepts such as nonstationarity [130], [131] and transfer learning [132]–[134] into such a framework. ■

REFERENCES

[1] G. Dornhege, J. d. R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, Eds., *Toward Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Press, 2007.

[2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.: Official J. Int. Federat. Clinical Neurophysiol.*, vol. 113, pp. 767–791, 2002.

[3] J. R. Wolpaw and E. W. Wolpaw, Eds., *Brain-Computer Interfaces: Principles and Practice*. Oxford, U.K.: Oxford Univ. Press, 2012, ISBN: 13:978-0195388855.

[4] J. d. R. Millán *et al.*, "Combining brain-computer interfaces and assistive technologies: State-of-the-art and challenges," *Frontiers Neurosci.*, vol. 4, p. 161, 2010.

[5] E. Donchin, "The P300 as a metric for mental workload," *Electroencephalography Clin. Neurophysiol., Supplement*, vol. 39, pp. 338–343, 1987.

[6] B. Blankertz *et al.*, "The Berlin brain-computer interface: Non-medical uses of BCI technology," *Frontiers Neurosci.*, vol. 4, p. 198, 2010.

[7] A. M. Owen *et al.*, "Detecting awareness in the vegetative state," *Science*, vol. 313, no. 5792, pp. 1402–1402, 2006.

[8] T. Ball, M. Kern, I. Mutschler, A. Aertsen, and A. Schulze-Bonhage, "Signal quality of simultaneously recorded invasive and non-invasive EEG," *NeuroImage*, vol. 46, no. 3, pp. 708–716, 2009.

[9] D. M. Goldenholz *et al.*, "Mapping the signal-to-noise-ratios of cortical sources in magnetoencephalography and electroencephalography," *Human Brain Mapping*, vol. 30, no. 4, pp. 1077–1086, 2009.

[10] F. Lazeyras, I. Zimine, O. Blanke, S. H. Perrig, and M. Seeck, "Functional MRI with simultaneous EEG recording: Feasibility and application to motor and visual activation," *J. Magn. Reson. Imag.*, vol. 13, no. 6, pp. 943–948, 2001.

[11] X. Cui, S. Bray, D. M. Bryant, G. H. Glover, and A. L. Reiss, "A quantitative comparison of NIRS and fMRI across multiple cognitive tasks," *NeuroImage*, vol. 54, no. 4, pp. 2808–2821, 2011.

[12] B. Blankertz, G. Dornhege, M. Krauledat, K.-R. Müller, and G. Curio, "The non-invasive Berlin Brain-Computer Interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539–550, 2007.

[13] B. Blankertz *et al.*, "The Berlin brain-computer interface: EEG-based communication without subject training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 14, no. 2, pp. 147–152, Jun. 2006.

[14] G. Pfurtscheller *et al.*, "The hybrid BCI," *Frontiers Neurosci.*, vol. 4, p. 42, 2010.

[15] E. L. Waltz and J. Llinas, *Multisensor Data Fusion*. Norwood, MA, USA: Artech House, 1990.

[16] D. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, Jan. 1997.

[17] R. Leeb, H. Sagha, R. Chavarriaga, and J. d. R. Millán, "A hybrid brain-computer interface based on the fusion of electroencephalographic and electromyographic activities," *J. Neural Eng.*, vol. 8, no. 2, p. 025011, 2011.

[18] R. Leeb, H. Sagha, R. Chavarriaga, and J. d. R. Millán, "Multimodal fusion of muscle and brain signals for a hybrid-BCI," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2010, pp. 4343–4346.

[19] G. Pfurtscheller, T. Solis-Escalante, R. Ortner, P. Linortner, and G. Müller-Putz, "Self-paced operation of an SSVEP-based orthosis with and without an imagery-based brain switch: A feasibility study towards a hybrid BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 4, pp. 409–414, Aug. 2010.

[20] C. Brunner *et al.*, "Improved signal processing approaches in an offline simulation of a hybrid brain-computer interface," *J. Neurosci. Meth.*, vol. 188, no. 1, pp. 165–173, 2010.

[21] B. Z. Allison *et al.*, "Toward a hybrid brain-computer interface based on imagined movement and visual attention," *J. Neural Eng.*, vol. 7, no. 2, p. 026007, 2010.

[22] Y. Li *et al.*, "An EEG-based BCI system for 2-D cursor control by combining Mu/Beta rhythm and P300 potential," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 10, pp. 2495–2505, Oct. 2010.

[23] R. Scherer, G. Müller-Putz, and G. Pfurtscheller, "Self-initiation of EEG-based brain-computer communication using the heart rate response," *J. Neural Eng.*, vol. 4, no. 4, p. L23, 2007.

[24] S. Fazli *et al.*, "Enhanced performance by a hybrid NIRS-EEG brain computer interface," *NeuroImage*, vol. 59, no. 1, pp. 519–529, 2012, open access.

[25] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, "Neurophysiological investigation of the basis of the fMRI signal," *Nature*, vol. 412, pp. 150–157, 2001.

[26] R. I. Goldman, J. M. Stern, J. Engel, and M. S. Cohen, "Simultaneous EEG and fMRI of the alpha rhythm," *Neuroreport*, vol. 13, no. 18, pp. 2487–92, 2002.

[27] F. Bießmann, Y. Murayama, N. K. Logothetis, K.-R. Müller, and F. C. Meinecke, "Improved decoding of neural activity from fMRI signals using non-separable spatiotemporal deconvolutions," *NeuroImage*, vol. 61, no. 4, pp. 1031–1042, 2012.

[28] U. Hasson, Y. Nir, I. Levy, G. Fuhrmann, and R. Malach, "Intersubject synchronization of cortical activity during natural vision," *Science*, vol. 303, no. 5664, pp. 1634–1640, 2004.

[29] M. Gaebler *et al.*, "Stereoscopic depth increases intersubject correlations of brain networks," *NeuroImage*, vol. 100, pp. 427–434, 2014.

[30] K. J. Friston, "Modalities, modes, models in functional neuroimaging," *Science*, vol. 326, no. 5951, pp. 399–403, 2009.

[31] F. Bießmann, S. M. Plis, F. C. Meinecke, T. Eichele, and K.-R. Müller, "Analysis of multimodal neuroimaging data," *IEEE Rev. Biomed. Eng.*, vol. 4, pp. 26–58, 2011.

[32] G. Pfurtscheller and F. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," *Clinical Neurophysiol.*, vol. 110, no. 11, pp. 1842–1857, 1999.

[33] R. Q. Cui, D. Huter, W. Lang, and L. Deecke, "Neuroimage of voluntary movement: Topography of the Bereitschaftspotential, a 64-channel DC current source density study," *NeuroImage*, vol. 9, no. 1, pp. 124–134, 1999.

[34] G. Dirnberger, U. Fickel, G. Lindinger, W. Lang, and M. Jahanshahi, "The mode of movement selection. Movement-related cortical potentials prior to freely selected and repetitive movements," *Experiment. Brain Res. Experimentelle Hirnforschung. Experiment. Cerebrale*, vol. 120, pp. 263–272, May 1998.

[35] N. Birbaumer, "Slow cortical potentials: Plasticity, operant control, behavioral effects," *Neuroscientist: Rev. J. Bringing Neurobiol., Neurol. Psychiatry*, vol. 5, pp. 74–78, 1999.

[36] C. Babiloni *et al.*, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: A high-resolution EEG study," *NeuroImage*, vol. 10, pp. 658–665, 1999.

[37] C. Toro *et al.*, "Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG," *Electroencephalogr. Clin. Neurophysiol.*, vol. 93, pp. 380–389, 1994.

[38] N. Weiskopf *et al.*, "Real-time functional magnetic resonance imaging: Methods and applications," *Magn. Resonan. Imag.*, vol. 25, no. 6, pp. 989–1003, 2007.

[39] M. Pregenzer, G. Pfurtscheller, and D. Flotzinger, "Automated feature selection with a distinction sensitive learning vector quantizer," *Neurocomputing*, vol. 11, pp. 19–29, 1996.

[40] G. Pfurtscheller *et al.*, "Current trends in Graz brain-computer interface (BCI)," *IEEE Trans. Rehabil. Eng.*, vol. 8, pp. 216–219, Jun. 2000.

[41] F. Babiloni *et al.*, "Linear classification of low-resolution EEG patterns produced by imagined hand movements," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 186–188, Jun. 2000.

[42] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Process. Mag.*, vol. 25, no. 1, pp. 41–56, Feb. 2008.

[43] S. Lemm, B. Blankertz, G. Curio, and K.-R. Müller, "Spatio-spectral filters for improving classification of single trial EEG," *IEEE Trans. Biomed. Eng.*, vol. 52, no. 9, pp. 1541–1548, Sep. 2005.

[44] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, 2008, pp. 2390–2397.

[45] S. Fazli *et al.*, "Ensembles of temporal filters enhance classification performance for ERD-based BCI systems," in *Proc. 4th Int. Brain-Comput. Interface Workshop and Training Course 2006*, 2008, Verlag der Technischen Univ. Graz.

[46] H.-I. Suk and S.-W. Lee, "A novel Bayesian framework for discriminative feature extraction in Brain-Computer Interfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 286–299, Feb. 2013.

[47] H.-I. Suk, S. Fazli, J. Mehnert, K.-R. Müller, and S.-W. Lee, "Predicting BCI subject performance using probabilistic spatio-temporal filters," *PloS ONE*, vol. 9, no. 2, p. e87056, 2014.

[48] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Mutual information-based selection of optimal spatialtemporal patterns for single-trial EEG-based BCIs," *Pattern Recogn.*, vol. 45, no. 6, pp. 2137–2144, 2012.

[49] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, "Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b," *Frontiers Neurosci.*, vol. 6, no. 00039, 2012.

[50] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2nd ed. Boston, MA, USA: Academic, 1990.

[51] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalogr. Clini. Neurophysiol.*, vol. 79, no. 6, pp. 440–447, 1991.

[52] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, pp. 441–446, Apr. 1998.

[53] C. Sannelli, C. Vidaurre, K.-R. Müller, and B. Blankertz, "Common spatial pattern patches—An optimized filter ensemble for adaptive brain-computer interfaces," *J. Neural Eng.*, vol. 8, no. 2, p. 025012 (7pp), 2011.

[54] B. D. Mensh, J. Werfel, and H. S. Seung, "BCI Competition 2003-Data set Ia: Combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 1052–1056, Jun. 2004.

[55] G. Dornhege, B. Blankertz, G. Curio, and K.-R. Müller, "Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms," *IEEE Trans. Biomed. Eng.*, vol. 51, no. 6, pp. 993–1002, Jun. 2004.

[56] M. Krauledat *et al.*, "Improving speed and accuracy of brain-computer interfaces using readiness potential features," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2004, vol. 4, pp. 4511–4515.

[57] J. Höhne *et al.*, "Motor imagery for severely motor-impaired patients: Evidence for brain-computer interfacing as superior control solution," *PLoS ONE*, vol. 9, no. 8, p. e104854, 2014.

[58] M. Falkenstein, J. Hohnsbein, J. Hoormann, and L. Blanke, "Effects of crossmodal divided attention on late ERP components. II. error processing in choice reaction taks," *Electroencephalogr. Clin. Neurophysiol.*, vol. 78, no. 6, pp. 447–455, 1991.

[59] W. J. Gehring, B. Goss, M. G. Coles, D. E. Meyer, and E. Donchin, "A neural system for error detection and compensation," *Psychol. Sci.*, vol. 4, no. 6, pp. 385–390, 1993.

[60] P. S. Bernstein, M. K. Scheffers, and M. G. Coles, "'Where did I go wrong?' A psychophysiological analysis of error detection," *J. Exper. Psychol.: Human Percept. Perform.*, vol. 21, no. 6, p. 1312, 1995.

[61] C. S. Carter *et al.*, "Anterior cingulate cortex, error detection, the online monitoring of performance," *Science*, vol. 280, no. 5364, pp. 747–749, 1998.

[62] L. Parra, C. Spence, A. Gerson, and P. Sajda, "Response error correction—A demonstration of improved human-machine performance using real-time EEG monitoring," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 173–177, Jun. 2003.

[63] B. Blankertz *et al.*, "Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial EEG analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 127–131, Jun. 2003.

[64] P. Ferrez and J. d. R. Millán, "You are wrong!—Automatic detection of interaction errors from brain waves," in *Proc. 19th Int. Joint Conf. Artif. Intell.*, 2005, pp. 1413–1418.

[65] G. Schalk, J. R. Wolpaw, D. J. McFarland, and G. Pfurtscheller, "EEG-based communication: Presence of an error potential," in *Clin. Neurophysiol.: Official J. Int. Fed. Clin. Neurophysiol.*, 2000, vol. 111, pp. 2138–2144.

[66] P. Ferrez and J. d. R. Millán, "Error-related EEG potentials generated during simulated brain-computer interaction," *IEEE Trans. Biomed. Eng.*, vol. 55, pp. 923–929, Mar. 2008.

[67] B. Dal Seno, M. Matteucci, and L. Mainardi, "Online detection of P300 and error potentials in a BCI speller," *Comput. Intell. Neurosci.*, vol. 2010, pp. 11–16, 2010.

[68] A. Kreilinger, C. Neuper, and G. R. Müller-Putz, "Error potential detection during continuous movement of an artificial arm controlled by brain-computer interface," *Med. Biol. Eng. Comput.*, vol. 50, no. 3, pp. 223–230, 2012.

[69] T. Solis-Escalante, G. Müller-Putz, C. Brunner, V. Kaiser, and G. Pfurtscheller, "Analysis of sensorimotor rhythms for the implementation of a brain switch for healthy subjects," *Biomed. Signal Process. Control*, vol. 5, no. 1, pp. 15–20, 2010.

[70] C. Vidaurre and B. Blankertz, "Towards a cure for BCI illiteracy," *Brain Topogr.*, vol. 23, pp. 194–198, 2010, open access.

[71] B. Z. Allison and C. Neuper, "Could anyone use a BCI?" in *Brain-Comput. Interfaces*, 2010, pp. 35–54, Springer.

[72] B. Rebsamen *et al.*, "Hybrid P300 and mu-beta Brain Computer Interface to operate a brain controlled wheelchair," in *Proc. 2nd Int. Conv. Rehabil. Eng. Assist. Technol.*, 2008, pp. 51–55. Singapore Therapeutic, Assistive Rehabil. Technol. (START) Centre.

[73] H. Riechmann, N. Hachmeister, H. Ritter, and A. Finke, "Asynchronous, parallel on-line classification of P300 and ERD for an efficient hybrid BCI," in *Proc. Neural Eng. (NER), 5th Int. IEEE/EMBS Conf.*, 2011, pp. 412–415.

[74] M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, "Towards zero training for brain-computer interfacing," *PLoS ONE*, vol. 3, p. e2967, Aug. 2008.

[75] S. Fazli *et al.*, "Subject-independent mental state classification in single trials," *Neural Netw.: Official J. Int. Neural Netw. Soc.*, vol. 22, pp. 1305–1312, Jun. 2009.

[76] P.-J. Kindermans, M. Tangermann, K.-R. Müller, and B. Schrauwen, "Integrating dynamic stopping transfer learning and language models in an adaptive zero-training ERP speller," *J. Neural Eng.*, vol. 11, no. 3, p. 035005, 2014.

[77] S. Harmeling, G. Dornhege, D. Tax, F. C. Meinecke, and K.-R. Müller, "From outliers to prototypes: Ordering data," *Neurocomputing*, vol. 69, no. 13–15, pp. 1608–1618, 2006.

[78] F. C. Meinecke, S. Harmeling, and K.-R. Müller, "Inlier-based ICA with an application to super-imposed images," *Int. J. Imag. Syst. Technol.*, vol. 15, no. 1, pp. 48–55, 2005.

[79] W. Samek, M. Kawanabe, and K.-R. Müller, "Divergence-based framework for common spatial patterns algorithms," *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 50–72, 2014.

[80] M. Krauledat, P. Shenoy, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Adaptation in CSP-based BCI systems," in *Toward Brain-Computer Interfacing*, G. Dornhege, J. d. R. Millán, T. Hinterberger, D. McFarland, and K.-R. Müller, Eds. Cambridge, MA, USA: MIT Press, 2007, pp. 305–309.

[81] P. Shenoy, M. Krauledat, B. Blankertz, R. P. N. Rao, and K.-R. Müller, "Towards adaptive classification for BCI," *J. Neural Eng.*, vol. 3, no. 1, pp. R13–R23, 2006.

[82] M. Sugiyama, M. Krauledat, and K.-R. Müller, "Covariate shift adaptation by importance weighted cross validation," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, 2007.

[83] B. Blankertz *et al.*, "Invariant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing," in *Advances in Neural Information Processing Systems 20*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. Cambridge, MA, USA: MIT Press, 2008, pp. 113–120.

[84] G. Rätsch, A. Demiriz, and K. Bennett, "Sparse regression ensembles in infinite and finite hypothesis spaces," *Mach. Learn. Spec. Iss. "New Meth. Model Sel. Model Combinat.*, vol. 48, pp. 193–221, 2002.

[85] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits Syst. Mag.*, vol. 6, no. 3, pp. 21–45, 3rd Quart., 2006.

[86] R. Meir and G. Rätsch, "An introduction to boosting and leveraging," in *Advanced Lectures on Machine Learning*. New York, NY, USA: Springer, 2003, pp. 118–183.

[87] S. Fazli *et al.*, "Subject independent EEG-based BCI decoding," in *Advances in Neural Information Processing Systems 22*, Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Cambridge, MA, USA: MIT Press, 2009, pp. 513–521.

[88] S. Fazli, M. Danóczy, J. Schelldorfer, and K.-R. Müller, "L1-penalized Linear Mixed-Effects Models for high dimensional data with application to BCI," *NeuroImage*, vol. 56, no. 4, pp. 2100–2108, 2011.

[89] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc.: Ser. B*, vol. 58, no. 1, pp. 267–288, 1996.

[90] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY, USA: Springer, 2008.

[91] H. Lu, K. Plataniotis, and A. Venetsanopoulos, "Regularized common spatial patterns with generic learning for EEG signal classification," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, 2009, pp. 6599–6602.

[92] H. Kang, Y. Nam, and S. Choi, "Composite common spatial pattern for subject-to-subject transfer," *Signal Process. Lett.*, vol. 16, no. 8, pp. 683–686, 2009.

[93] H. Lu, H.-L. Eng, C. Guan, K. Plataniotis, and A. Venetsanopoulos, "Regularized common spatial pattern with aggregation for EEG classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.

[94] F. Lotte and C. Guan, "Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 2, pp. 355–362, Feb. 2011.

[95] D. Devlaminck, B. Wyns, M. Grosse-Wentrup, G. Otte, and P. Santens, "Multi-subject learning for common spatial patterns in motor-imagery BCI," *Comput. Intell. Neurosci.*, vol. 2011, no. 217987, pp. 1–9, 2011.

[96] H. Kang and S. Choi, "Bayesian multi-task learning for common spatial patterns," in *Proc. Int. Workshop Pattern Recogn. NeuroImag. (PRNI)*, 2011, pp. 61–64.

[97] H. Kang and S. Choi, "Bayesian common spatial patterns for multi-subject EEG classification," *Neural Netw.*, vol. 57, pp. 39–50, 2014.

[98] S. Lu, C. Guan, and H. Zhang, "Unsupervised brain computer interface based on intersubject information and online adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 2, pp. 135–145, Apr. 2009.

[99] W. Samek, F. C. Meinecke, and K.-R. Müller, "Transferring subspaces between subjects in brain-computer interfacing," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2289–2298, Aug. 2013.

[100] J. P. Dmochowski, P. Sajda, J. Dias, and L. C. Parra, "Correlated components of ongoing EEG point to emotionally laden attention-a possible marker of engagement?" *Frontiers Human Neurosci.*, vol. 6, 2012.

[101] J. P. Dmochowski *et al.*, "Audience preferences are predicted by temporal reliability of neural processing," *Nature Commun.*, vol. 5, 2014.

[102] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3, pp. 321–377, 1936.

[103] S. Dähne *et al.*, "Finding brain oscillations with power dependencies in neuroimaging data," *NeuroImage*, vol. 96, pp. 334–348, 2014.

[104] T. T. Liu *et al.*, "Caffeine alters the temporal dynamics of the visual bold response," *NeuroImage*, vol. 23, no. 4, pp. 1402–1413, 2004.

[105] A. Kleinschmidt *et al.*, "Simultaneous recording of cerebral blood oxygenation changes during human brain activation by magnetic resonance imaging and near-infrared spectroscopy," *J. Cerebral Blood Flow Metab.*, vol. 16, pp. 817–826, 1996.

[106] M. Schmid, A. Oeltermann, C. Juchem, N. K. Logothetis, and S. Smirnakis, "Simultaneous EEG and fMRI in the macaque monkey at 4.7 tesla," *Magn. Resonance Imag.*, vol. 24, no. 4, pp. 335–342, 2006.

[107] S. Vulliemoz *et al.*, "Simultaneous intracranial EEG and fMRI of interictal epileptic discharges in humans," *NeuroImage*, vol. 54, no. 1, pp. 182–190, 2011.

[108] F. Bießmann *et al.*, "Temporal kernel CCA and its application in multimodal neuronal data analysis," *Mach. Learn. J.*, vol. 79, no. 1–2, pp. 5–27, 2010.

[109] S. Dähne *et al.*, "Integration of multivariate data streams with bandpower signals," *IEEE Trans. Multimedia*, vol. 15, no. 5, pp. 1001–1013, Aug. 2013.

[110] F. D. Martino, A. W. de Borst, G. Valente, R. Goebel, and E. Formisano, "Predicting EEG single trial responses with simultaneous fMRI and Relevance Vector Machine regression," *NeuroImage*, vol. 56, no. 2, pp. 826–836, 2011.

[111] S. Fazli and S.-W. Lee, "Brain computer interfacing: A multi-modal perspective," *J. Comput. Sci. Eng.*, vol. 7, no. 2, pp. 132–138, 2013.

[112] V. Kaiser *et al.*, "Cortical effects of user training in a motor imagery based brain–computer interface measured by fNIRS and EEG," *NeuroImage*, vol. 85, no. 0, pt. 1, pp. 432–444, 2014.

[113] K. K. Kwong *et al.*, "Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation," *Proc. Nat. Acad. Sci. United States Amer.*, vol. 89, pp. 5675–5679, 1992.

[114] A. Villringer, J. Planck, C. Hock, L. Schleinkofer, and U. Dirnagl, "Near infrared spectroscopy (NIRS): A new tool to study hemodynamic changes during activation of brain function in human adults," *Neurosci. Lett.*, vol. 154, pp. 101–104, 1993.

[115] S. K. Piper *et al.*, "A wearable multi-channel fNIRS system for brain imaging in freely moving subjects," *NeuroImage*, vol. 85, pp. 64–71, 2014.

[116] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*. New York, NY, USA: Cambridge Univ. Press, 2002.

[117] G. L. Rogova and V. Nimier, "Reliability in information fusion: Literature survey," in *Proc. 7th Int. Conf. Inf. Fusion*, 2004, pp. 1158–1165.

[118] X. Cui, S. Bray, and A. L. Reiss, "Speeded near infrared spectroscopy (NIRS) response detection," *PLoS ONE*, vol. 5, no. 11, p. e15474, 2010.

[119] S. Fazli, J. Mehnert, J. Steinbrink, and B. Blankertz, "Using NIRS as a predictor for EEG-based BCI performance," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2012, 2012, pp. 4911–4914.

[120] S.-K. Yeom *et al.*, "Multimodal imaging technique for rapid response brain-computer interface feedback," in

Proc. Int. Winter Workshop Brain-Comput. Interface (BCI), 2013, pp. 92–94, IEEE.

[121] N. H. Anderson, *Foundations of Information Integration Theory*. New York, NY, USA: Academic, 1981.

[122] V. Torra, *Information Fusion in Data Mining*. New York, NY, USA: Springer-Verlag, 2003.

[123] N. Xiong and P. Svensson, "Multi-sensor management for information fusion: Issues and approaches," *Inf. Fusion*, vol. 3, no. 2, pp. 163–186, 2002.

[124] J. L. Crowley and Y. Demazeau, "Principles and techniques for sensor data fusion," *Signal Process.*, vol. 32, no. 1, pp. 5–27, 1993.

[125] D. L. Hall and S. A. McMullen, *Mathematical Techniques in Multisensor Data Fusion*. Norwood, MA, USA: Artech House, 2004.

[126] M. A. Abidi and R. C. Gonzalez, *Data Fusion in Robotics and Machine Intelligence*. New York, NY, USA: Academic, 1992.

[127] R. Alur *et al.*, "The algorithmic analysis of hybrid systems," *Theoret. Comput. Sci.*, vol. 138, no. 1, pp. 3–34, 1995.

[128] S. Ben-Yacoub, "Multi-modal data fusion for person authentication using SVM," *Audio Video Based Person Authent.*, pp. 25–30, 1999.

[129] R. Tomioka and K.-R. Müller, "A regularized discriminative framework for EEG analysis with application to brain-computer interface," *NeuroImage*, vol. 49, no. 1, pp. 415–432, 2010.

[130] P. von Bünau, F. C. Meinecke, F. J. Király, and K.-R. Müller, "Finding stationary subspaces in multivariate time series," Phys. Rev. Lett., vol. 103, no. 21, p. 214101, 2009.

[131] W. Samek, C. Vidaurre, K.-R. Müller, and M. Kawanabe, "Stationary common spatial patterns for brain-computer interfacing," *J. Neural Eng.*, vol. 9, no. 2, p. 026013, 2012.

[132] S. Thrun, "Is learning the nth thing any easier than learning the first?" *Adv. Neural Inf. Process. Syst.*, , pp. 640–646, 1996.

[133] R. Caruana, *Multitask Learning*. New York, NY, USA: Springer, 1998.

[134] M. E. Taylor and P. Stone, "Transfer learning for reinforcement learning domains: A survey," *J. Mach. Learn. Res.*, vol. 10, pp. 1633–1685, 2009.

ABOUT THE AUTHORS

Siamac Fazli received the B.Sc. degree in physics from the University of Exeter, Exeter, U.K., in 2002, the M.Sc. degree in medical neurosciences from the Humboldt University Berlin, Berlin, Germany, in 2004, and the Ph.D. degree from the Berlin Institute of Technology, Berlin, Germany, in 2011.

From 2011 to 2013, he was a Postdoc Researcher at the Berlin Institute of Technology for the Bernstein Focus Neurotechnology. Since 2013, he has been an Assistant Professor at Korea University. His current research interests include neuroscience, machine learning, multimodal neuroimaging and brain-computer interfacing.

Sven Dähne received the B.Sc. degree in cognitive science from the University of Osnabrück, Osnabrück, Germany, in 2007 and the M.Sc. degree in computational neuroscience from TU Berlin and the Bernstein Center for Computational Neuroscience (BCCN), Berlin, Germany, in 2010. He is currently working towards the Ph.D. degree at TU Berlin and the BCCN.

He has been working on machine learning methods for online-adaptation of EEG-based BCIs and investigations in the causes for nonstationarity in BCI performance. His current research focuses on the analysis of spectral nonstationarities in the context of uni- and multi-modal neuroimaging data.

Wojciech Samek (M'13) received a Diploma degree in computer science from Humboldt University of Berlin, Berlin, Germany, in 2010 and the Ph.D. degree in machine learning from the Technische Universität Berlin, Berlin, in 2014.

In the same year he founded the Machine Learning Group at Fraunhofer Heinrich Hertz Institute which he is currently directing. He was Scholar of the German National Academic Foundation and a Ph.D. Fellow at the Bernstein Center for Computational Neuroscience Berlin. He was visiting Heriot-Watt University in Edinburgh and the University of Edinburgh from 2007 to 2008 and in 2009 he was with the Intelligent Robotics Group at NASA Ames Research Center in Mountain View. His research interests include machine learning, biomedical engineering, neuroscience, and computer vision.

Felix Bießmann received the B.Sc. degree in cognitive science from the University of Osnabrück, Osnabrück, Germany, in 2005, the M.Sc. degree in neuroscience from the International Max-Planck Research School, Tübingen, Germany, and the Ph.D. degree in machine learning from the Berlin Institute of Technology, Berlin, Germany.

From 2013 to 2014, he was an Assistant Professor at Korea University. He currently is with the Machine Learning Group at Amazon Development Center Berlin, Berlin, Germany. His research interests include statistical learning methods for multimodal data with a focus on neuroscientific and biomedical data.

Klaus-Robert Müller (M'12) studied physics at Karlsruhe Institute of Technology, Karlsruhe, Germany, from 1984 to 1989 and received the Ph.D. degree in computer science from Karlsruhe Institute of Technology in 1992.

He has been a Professor of computer science at Technische Universität Berlin, Berlin, Germany, since 2006. At the same time he has been the Director of the Bernstein Focus on Neurotechnology Berlin. After completing a postdoctoral position at GMD FIRST in Berlin, he was a Research Fellow at the University of Tokyo from 1994 to 1995. In 1995, he founded the Intelligent Data Analysis group at GMD-FIRST (later Fraunhofer FIRST) and directed it until 2008. From 1999 to 2006, he was a Professor at the University of Potsdam.

Dr. Müller was awarded the 1999 Olympus Prize by the German Pattern Recognition Society, DAGM, and, in 2006, he received the SEL Alcatel Communication Award. In 2014 he received the Berliner Wissenschaftspreis des regierenden Bürgermeisters. In 2012, he was elected to be a member of the German National Academy of Sciences-Leopoldina. His research interests are intelligent data analysis, machine learning, signal processing, and brain-computer interfaces.