

Omurtag, A., Aghajani, H., and Keles, H.O. (2017). Decoding human mental states by whole-head EEG+ fNIRS during category fluency task performance, *J. Neural Eng.* 14 066003

Decoding Human Mental States by Whole-Head EEG+fNIRS During Category Fluency Task Performance

Ahmet Omurtag

Department of Engineering, Nottingham Trent University, Nottingham, United Kingdom

Haleh Aghajani

Department of Biomedical Engineering, University of Houston, Houston, TX 77204, United States

Hasan Onur Keles

Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, United States

ABSTRACT

Objective. Concurrent scalp electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS), which we refer to as EEG+fNIRS, promises greater accuracy than the individual modalities while remaining nearly as convenient as EEG. We sought to quantify the hybrid system's ability to decode mental states and compare it with unimodal systems.

Approach. We recorded from healthy volunteers taking the category fluency test and applied machine learning techniques to the data.

Main results. EEG+fNIRS's decoding accuracy was greater than that of its subsystems, partly due to the new type of neurovascular features made available by hybrid data.

Significance. Availability of an accurate and practical decoding method has potential implications for medical diagnosis, brain-computer interface design, and neuroergonomics.

Keywords

functional near-infrared spectroscopy (fNIRS)

Electroencephalography (EEG)

category fluency

mental state decoding

multi-modal brain recording

support vector machine

hybrid functional neuroimaging

INTRODUCTION

There are many reasons for wanting to have accurate and practical methods for decoding human mental states noninvasively. Developing passive and active brain-computer interfaces (BCI) for everyday life and for the clinic (Nicolas-Alonso and Gomez-Gil, 2012), improving human performance and well-being through better design (Mehta and Parasuraman, 2013; Zander and Kothe, 2011), investigating the mechanisms of perception and cognition (Haynes and Rees, 2006), and diagnosing and predicting disease (Orrù et al., 2012; Zhang and Raichle, 2010) to name a few. High fidelity and convenience are obvious technical desiderata for many of these applications. In this paper we aimed to study the characteristics of an emerging multimodal (or hybrid) technique, concurrent electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRS), which we refer to as EEG+fNIRS. This technique promises greater accuracy than the individual modalities while remaining nearly as convenient as traditional scalp EEG. Our primary aims were to quantify the hybrid system's performance in relation to unimodal approaches and better understand the interplay between the different types of features it provides.

EEG arises mainly from the synchronization of post-synaptic potentials in dendrites, and represents the input to a large neuronal population rather than its action potentials (Buzsáki et al., 2012). It is measured as the voltage difference (typically $<100 \mu\text{V}$) between an electrode attached to the scalp and a reference electrode placed elsewhere on the head or body. EEG may reflect the momentary activity of a population located near the recording electrodes, but also distal populations if they are sufficiently active. fNIRS, on the other hand, provides information that is complementary to EEG, by measuring the changes in cerebral blood flow (CBF) and related hemoglobin concentrations through near-infrared light source/detectors on the scalp. It is comparable to EEG in portability. We have used the continuous wave method with two distinct wavelengths to extract the concentration changes of oxy- and deoxy-hemoglobin (HbO and HbR) in the outer layers of the cortex (Durduran, 2010; Scholkmann et al., 2014; Ugilaloro, 2014). The light emitted by an infrared source placed on the scalp produces a signal at a detector placed a few centimeters away from the source. This allows the calculation of local HbO and HbR by means of the modified Beer-Lambert law (Delpy et al., 1988). Following neural activation, local blood flow and volume typically increase on a time scale of seconds, causing a rise in HbO and a decrease in HbR of smaller magnitude. These closely agree with the blood oxygen level dependent (BOLD) response from fMRI (Huppert et al., 2006; Kleinschmidt, 1996; Steinbrink et al., 2006; Strangman et al., 2002). fNIRS has been

used in numerous studies of the hemodynamics of adult and developing human brains (e.g. Gentili et al., 2013; Lloyd-Fox et al., 2010; Mesquita et al., 2010; White et al., 2009).

EEG and fNIRS signals differ greatly in the underlying physiology and the physical principles they utilize. This implies that particular shortcomings such as artifacts of various origin, possible weakly coupled sensors, and subject variability that lead to challenges in analysis and classification, would only affect one modality rather than both at the same time. The process of pooling of the signals from two distinct systems is therefore expected to enhance the hybrid system's robustness. EEG+fNIRS also offers additional information based on neurovascular coupling (NVC), the cascade of processes by which neural activity modulates local blood flow and oxygenation (Keles et al., 2016; Ma et al., 2016). NVC related features are not resolvable by a unimodal signal sensitive to only neural activity (e.g. EEG) or only hemodynamics (e.g. BOLD).

Although there is an increasing number of EEG+fNIRS based decoding studies currently available, none to our knowledge has systematically investigated the separate effects of various types of features (Buccino et al., 2016; Coffey et al., 2012; Fazli et al., 2012; Liu et al., 2017; Putze et al., 2015; Tomita et al., 2014). Another primary aim of this study was, therefore, to rigorously compare the performance of multiple subsets of the hybrid system. In this study we used co-located EEG and fNIRS channels at all standard 10-20 sites bilaterally covering the frontopolar, frontal, central, temporal, parietal, and occipital areas. Many fNIRS studies utilize small frontally located optodes (e.g. Gentili et al., 2013; Kahlaoui et al., 2012; Takahashi et al., 2011; Yeung et al., 2016) and this provides a convenient arrangement since the forehead has little hair that can impede optical coupling. In fact smaller sets of optodes are better than whole-head montages in terms of rapid deployment. In applications where accuracy is paramount, such as BCI, however, this raises a question about the extent of the effectiveness of small fNIRS montages. Our whole-head setup (not to be confused with high-density) allowed us to address this question as well.

A suitable testbed for our study was provided by the semantic verbal fluency task, also known as the category fluency task (CFT). CFT engages multiple mental abilities, as individuals generate words in a particular semantic category (e.g. flowers) while continuously searching their memory for verbal information and self-monitoring in order to avoid repetitions and external distractions. There is extensive information regarding the patterns of neural activity in healthy

subjects accompanying CFT coming from multiple neuroimaging studies (Costafreda et al., 2006; Frith et al., 1991; Schecklmann et al., 2008; Schlösser et al., 1998). Verbal fluency performance and alterations in related brain activity are being linked to clinical ramifications such as traumatic brain injury (Henry and Crawford, 2004), depression (Takizawa et al., 2014), and cognitive decline (Heinzel et al., 2013). Versions of the verbal fluency task have also been widely used in BCI development and other research (Schecklmann et al., 2010; Schudlo and Chau, 2015a). However we are not aware of studies of multimodal brain function measurement during verbal fluency. We therefore chose to investigate EEG+fNIRS in the context of verbal fluency where an accurate and practical method can potentially make a significant contribution.

METHODS

Study Design

The experiments, approved by the Institutional Review Board of the University of Houston, were performed according to standard block design. After written informed consent, a subject was seated in a comfortable chair in front of a desk. A block of the experiment contained a 30 second rest period when the subject relaxed without moving, thinking of anything particular, or falling asleep. This was followed by a 30 s Category Fluency Task. In the beginning of a Task, the experimenter verbally conveyed to the subject the name of a category such as "flower" or "car" selected randomly from a predetermined list. The subject proceeded to find words that belonged to that category. The subject wrote the word on a small white board on the desk. The subject was asked to move only one hand, move as little as possible, and adhere to a rate of about 6 words per 30 s. Subjects who went faster were instructed by the experimenter by saying "slow down." If no word was found the subject was to write "no idea" at about the same rate. Task was followed by another 30 s of Rest. The final portion of the block consisted of a control condition when the subjects wrote the days of the week at the same rate as during the task. We included the control condition since it was less likely to be contaminated by thought than rest. Each recording contained three repetitions of the block. Henceforth we use the term *episode* to refer to a Task, Rest, or Control time segment. Ten healthy adult volunteers participated in the study (8 male, age mean 26 years (range 24-28)). All participants could communicate in English and were asked not to consume alcoholic or caffeinated beverages during the 3 hours prior to the experiment. Inclusion criteria included no history of psychological illness or substance dependence (determined by self-reporting).

Data Collection

EEG and fNIRS data were acquired simultaneously by using microEEG (Bio-Signal Group Inc., Brooklyn, New York) and NIRScout extended continuous wave system (NIRx Medical Technologies, New York). The microEEG is a wireless and battery operated miniature (80 g) device that is approved by the Food and Drug Administration under 510(k). It was recording 19 channels at 250 Hz with 16 bit resolution. The fNIRS dual wavelengths were set at 760 and 850 nm and NIRStar software by NIRx was used to verify the signal quality before each recording. EEG and fNIRS signals were synchronized by means of event markers sent by the Presentation software (Neurobehavioral Systems Inc.). The EEG and fNIRS (6.25 Hz) sample rates were sufficiently rapid for the purposes of our study. The EEG and fNIRS channels were co-located at the standard International 10-20 sites Fp1, Fp2, F7, F3, Fz, F4, F8, C3, Cz, C4, T3, T4, T5, T6, P3, P4, Pz, O1, and O2, covering all areas of the scalp. The EEG reference and ground electrodes were, respectively, FCz and Fpz. In order to stably maintain an EEG sensor together with a fNIRS source-detector pair at every site, we utilized an array of triplet holders (Keles et al., 2016).

Preprocessing

The signals were band-pass filtered (0.5-80 Hz for the EEG and 0.01-0.5 Hz for fNIRS) and the EEG was also notch filtered at 60 Hz to minimize the power line noise. Power spectrograms were created from the EEG signals using non-overlapping $\Delta W = 1$ s Hamming windows. This resulted in a frequency resolution of $\Delta f = 1 / \Delta W = 1$ Hz which was sufficient for subsequently determining the frequency band power during feature extraction. The fNIRS detector readings at two wavelengths were converted into concentration changes of oxy- and deoxy-hemoglobin by using the modified Beer-Lambert Law (MBL) (Delpy et al., 1988). One of the authors (AO) visually inspected the preprocessed signals for artifacts such as eye blinks, muscle activity, motion effects, and sleep patterns. Two recordings that contained excessive amounts of such artifacts were omitted from the study and their data were discarded. The remaining 8 recordings contained 10% or less of such artifact contamination and were included in the subsequent analysis. The parameter values employed in the MBL and other details of hardware and preprocessing were the same as in a previous study (Keles et al., 2016). Figure 1 shows an example of preprocessed data. Most of the EEG spectrograms in the figure contain a strong alpha rhythm (8-12 Hz), especially in the parietal and occipital areas. The EEG patterns, as well as the oxy- (red traces) and deoxy-hemoglobin changes (blue), contain fluctuations which may

be related to varying levels of cognitive effort. Calculations described in this paper used Matlab v.8.2.0.701 (The MathWorks, Inc., Natick, Massachusetts, United States).

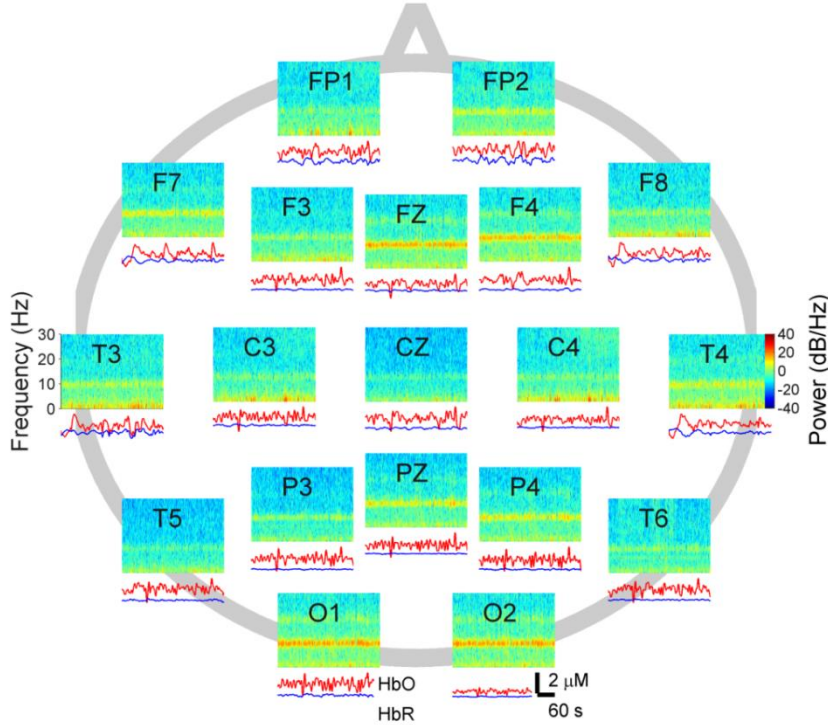


Figure 1 Example of preprocessed data from a recording. At each of the 10-20 locations the EEG spectrogram and the oxy- (red) and deoxy-hemoglobin concentration changes (blue) are shown.

Feature Extraction

This paper investigated the performance of three systems (EEG, fNIRS, and Hybrid) each with two sets of features (Basic and Expanded). The contents of these systems and feature sets are summarized in Table 1. The abbreviations such as PSD, PLV, etc. in the table denote particular feature types that will be explained in this subsection. The motivation behind the distinction Basic v Expanded sets was as follows. Some features such as frequency band power of EEG or the mean amplitude of changes in hemoglobin concentration from fNIRS are widely used and their utility is well established in functional neuroimaging. These were gathered under the Basic feature set. Other features, e.g. derived from phase synchronization, are less commonly found in the literature while yet others, such as neurovascular features, are being used for the first time in this paper to our knowledge. These helped form the Expanded set. In computing most features we have also utilized principal components analysis, as will be described below.

		System Type		
		EEG	fNIRS	Hybrid
Feature Set	Basic	<i>PSD</i>	<i>HBO, HBR</i>	<i>PSD, HBO, HBR</i>
	Expanded	<i>PSD, PLV</i>	<i>HBO, HBR, COR</i>	<i>PSD, PLV, HBO, HBR, COR, NVO, NVR</i>

Table 1 System types and associated feature sets.

The list below shows the definition for each feature type in Table 1, along with an indication of its dependence on variables such as observation window, frequency, and sensor pair. Following the list will be detailed descriptions for how each feature is calculated.

- $PSD(f, t)$: EEG frequency band power at frequency band f and observation window t .
- $PLV(f, p, t)$: EEG based phase locking value at frequency f for the sensor pair p .
- $HBO(t)$ and $HBR(t)$: fNIRS based oxy- and deoxy-hemoglobin concentration changes.
- $COR(t)$: fNIRS based correlation between oxy- and deoxy-hemoglobin.
- $NVO(f, t)$ and $NVR(f, t)$: EEG+fNIRS based correlation between EEG frequency band power and oxy- and deoxy-hemoglobin.

PSD: We chose to denote this feature as PSD because it is based on the power spectral density of EEG. A widely used and informative variable derived from an EEG signal is the frequency band power, calculated by integrating the power spectral density within a particular frequency range, and averaging the integral over the feature window, ΔT . The frequency band power of EEG is thought to represent the extent of recruitment into locally synchronized neuronal populations (Gevins and Smith, 2000). From the spectrogram of each channel we calculated the frequency band power, $p_j(f, t)$, where $j = 1, 2, \dots, N$ denoted a channel, $f = 1, 2, \dots, N_{freq}$ denoted one of the frequency ranges, and t represented the non-overlapping time windows used in feature generation. In our setup $N = 19$, corresponding to the total number of 10-20 sites. $N_{freq} = 8$ corresponded to the total number of frequency ranges chosen in 4 Hz increments from 0 to 32 Hz, namely delta (0–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta1 (12–16 Hz), beta2 (16–20 Hz), beta3 (20–24 Hz), beta4 (24–28 Hz),

beta5 (24 – 28 Hz) , and low gamma (28 – 32 Hz) . The upper bound was chosen to be low gamma since cerebral activity contributes negligibly to scalp EEG beyond this range (Petroff et al., 2016). Preliminary studies indicated that lumping or partitioning the beta range in different ways did not affect our results.

PLV: From the EEG signals across pairs of sites we calculated the Phase Locking Value (PLV) at the frequencies 4, 10, 20, and 40 Hz. PLV is a practical method for quantifying long-range, frequency-specific synchronization and a potential measure of inter-area neural integration during cognitive tasks (Varela et al., 2001). The signal pairs were selected to be intra-hemispheric (F7-T5, F3-P3, F4-P4, F8-T6, FP1-O1, FP2-O2), inter-hemispheric symmetric (FP1-FP2, F7-F8, F3-F4, T3-T4, T5-T6, O1-O2), and inter-hemispheric asymmetric (FP1-O2, FP2-O1, F7-T6, F8-T5). To calculate the frequency-specific PLV for each pair, the corresponding EEG signals were band-pass filtered at that frequency within a 2 Hz wide band, their phases extracted via Hilbert transform, the phase difference $\Delta\theta_m$ computed at every data point m , and the PLV for a particular window was found as the absolute value of the mean of $\exp\{i\Delta\theta_m\}$ over the window (Le Van Quyen et al., 2001). (Note $i = \sqrt{-1}$.) PLV values close to unity thus represent the tendency of the phase difference to remain fixed, while the PLV approaches zero if the phase difference tends to uniformly sample the interval 0 to 2π . The 2 Hz width of the bands was chosen to be sufficiently narrow for the phase to be well-defined (Dvorak and Fenton, 2014).

HBO, HBR, and COR: We averaged (over the feature window) the amplitudes of the oxy- and deoxy-hemoglobin concentration changes in order to generate features from the fNIRS data. These provided the features denoted HBO and HBR. Preliminary results suggested that the correlation between the oxy- and deoxy-hemoglobin was exhibiting regional and time dependent patterns. We therefore also computed the correlation between the two quantities for each window, yielding the feature denoted COR.

NVO and NVR: Multimodal data allowed us to investigate the utility of features derived from relationships between neural (based on EEG) and hemodynamic activations (fNIRS). Previous research suggested the existence of frequency- and area-specific lagged correlation patterns in EEG+fNIRS. In particular the oxy-hemoglobin response to spontaneous alpha power fluctuations in the resting state typically follow an anti-correlation with a negative peak around 8

s (Keles et al., 2016). Since this study uses relatively shorter feature time windows we only computed the zero-lag correlation between EEG frequency band power and oxy- and deoxy-hemoglobin changes (to obtain NVO and NVR, respectively). These helped generate what we refer to as *neurovascular* features. They utilized the same set of 4 Hz wide frequency ranges that were described above. The calculation of the correlation was performed without removing the signal means. The neurovascular quantities therefore provide a measure of the degree of neuronal synchronization and the local instantaneous hemodynamic state.

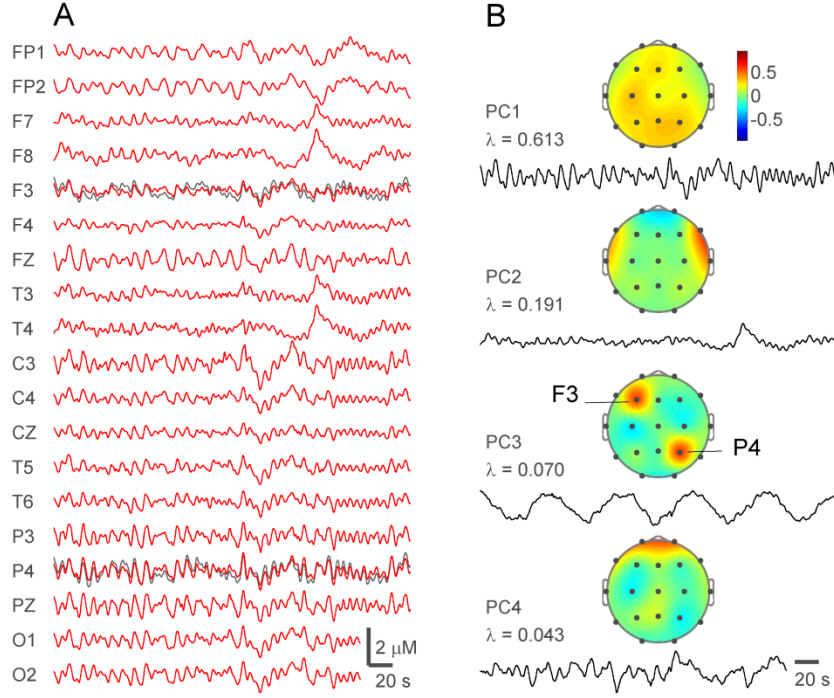


Figure 2 Principal components analysis of 19 channels of resting state fNIRS data. (A) Oxy-hemoglobin concentration changes (red curves) with simulated sinusoidal (60 s period) small amplitude perturbations superimposed at channels F3 and P4 (gray curves). (B) The first four principal components (PC) and their time courses. The perturbation was identified by the third PC with energy $\lambda = 0.07$.

Principal Components Analysis (PCA): We adopted PCA as a method of generating more illuminating and potentially better discriminating features. Each type of quantity described above is defined at all the 10-20 locations, with the exception of PLV which is for electrode pairs only. We performed PCA separately on each of the groups of variables (except PLV). For example the window averaged oxy-hemoglobin concentration change constituted a set of N time series. This yielded a set of principal components (PC), $\phi_j^{(n)}$ $n = 1, 2, \dots, N$, with corresponding time series $a_n(t)$. PCA guarantees that distinct pairs of such time series are uncorrelated. For this

reason PCA (and the related method of independent components analysis (ICA)) have been used by researchers to identify simultaneous independent contributors to the EEG signals. Such sources may arise from spontaneous or task driven cerebral activity, muscle or motion artifacts, eye blinks, or non-physiological ambient noise. Identification of distinct sources is thus useful for artifact rejection (Jung et al., 2000). We have selected PCA because of its computational stability and speed, and its ability to calculate the relative amplitudes of the sources (unlike ICA). The variance of the time series associated with a PC is denoted λ (also called its energy). We normalize them so that the set of energies for a particular type of feature sum to unity.

In order to investigate the utility of PCA for our study, we tested its ability to segregate an extraneous small perturbation that was artificially introduced into a resting state (RS) fNIRS recording (Figure 2). The 4 min segment of oxy-hemoglobin concentration changes (A, red curves) were obtained from a previous study (Keles et al., 2016). The perturbation was a small amplitude sinusoid with a period of 60 s added onto the oxy-hemoglobin concentration changes in channels F3 and P4 (gray curves). These channels were selected randomly. The first four PCs are shown on the right as activation maps. Directly below each PC is shown the corresponding time series. The first PC (with the highest energy $\lambda = 0.613$) displayed a distributed activation pattern with amplitudes higher in the posterior areas. The corresponding $a_1(t)$ was consistent with spontaneous fluctuations of oxy-hemoglobin during RS. The second PC identified the temporalis muscle oxygenation that appears particularly prominently in the electrode sites F7, F8, T3, and T4. The time course of this activation shown as a sharp spike decaying in about 10 s was consistent with a jaw clenching event (Keles et al., 2016). The third PC consisted exclusively of the artificially added sinusoidal perturbation as seen both in the activation map and the time series. The variation distributed across areas (in this case F3 and P4) was therefore encapsulated, through PCA, into a single variable $a_3(t)$. Furthermore its topographic distribution was visible in $\phi_j^{(3)}$. The implication was that distributed processes such as cerebral activity due to task performance (potentially similar to the perturbation simulated here) could be better represented by the PC time series than by the signals from individual sites. The topographic plots in Figure 2B and elsewhere in this paper used EEGLAB (Delorme and Makeig, 2004). The locations of the channels in Figure 2B are shown as small dark filled circles and the values are spline interpolated over the head.

Classification

We used support vector machine (SVM) in this setting to make inferences about the mental states of individual subjects on the basis of their functional neuroimaging data. After feature extraction the data are in the form of a matrix whose columns are different features and rows are time windows (or an observation). Every observation in our experiments has an associated label such as Rest, Task, or Control, thus the problem is suitable for analysis by supervised machine learning (ML). Supervised ML approaches in general (and SVM in particular) take advantage of distributed and subtle patterns of activation (otherwise possibly undetectable) to allow inference at the level of the individual subject and specific time window, rather than at the group level. This makes them suitable for applications in brain computer interfaces (BCI). SVM classify data points by maximizing the margin between classes in a high-dimensional feature space (for review and references, Orrù et al., 2012). SVM has long been used for decoding mental states of healthy subjects (Haynes and Rees, 2006). In some applications it may be advantageous to nonlinearly transform the feature space before training, however in our preliminary exploration of nonlinear kernels the accuracy did not increase. In order to avoid possible overfitting we utilized linear kernels only. We used SVM for binary classification on 1) Task and Control; and 2) Task and Rest observations.

When there was a need to reduce the system sizes (e.g. Figure 9), we utilized, R^2 , square of Pearson's correlation between a feature and class labels, in order to rank features with regard to their discrimination ability. This is a straightforward and widely used univariate filter for feature reduction (Gudiño-Mendoza et al., 2016; Mwangi et al., 2014). We examined the accuracy of classification under two distinct conditions: with subject-specific SVM and with group trained SVM. In the latter observations from all subjects were pooled before training. For testing we used leave-one-out cross-validation (LOOCV) in the former and k-fold cross-validation (k=10) for the latter case. The LOOCV was implemented by leaving out one observation and training and testing an SVM, then repeating this process for each observation. In the group trained SVM, the k-fold CV was implemented where about 90% of the observations were used for training and the remaining used for testing, and the procedure repeated with mutually exclusive and exhaustive sets of test partitions. The Matlab functions `svmtrain`, `svmclassify` and `cvpartition` were utilized for implementing the above calculations. To evaluate performance, sensitivity (fraction of true positives correctly identified), specificity (fraction of true negatives correctly identified), and accuracy (fraction of observations correctly identified) were determined.

For example in the Task v Control problem, sensitivity was defined as the fraction of Task windows correctly identified, while specificity was the fraction of Control windows correctly identified.

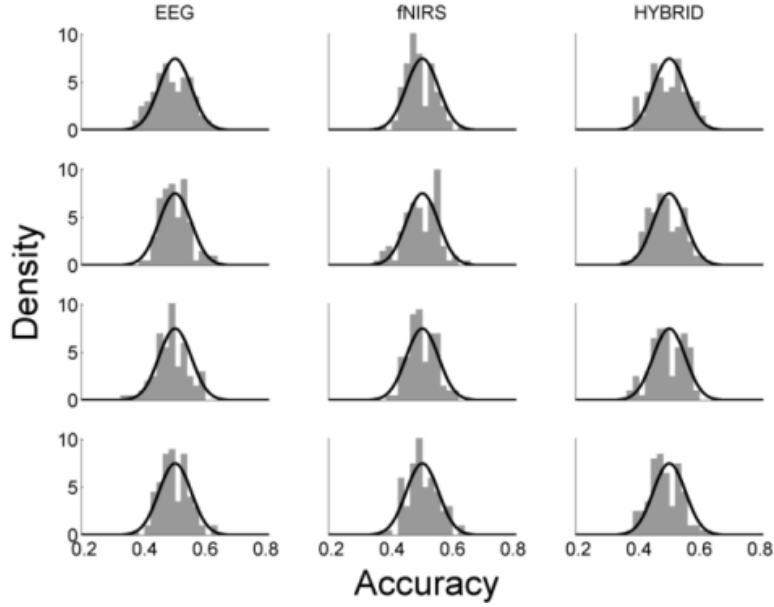


Figure 3 Null density of accuracies of 4 subjects predicted by the permutation test (shaded) and binomial model (solid curve). The Expanded feature set was used. The x-axis bin width: 0.02. Each row is a different subject.

The statistical significance of classification from a specific system (e.g. EEG) and a specific subject was determined by the binomial test (Combrisson and Jerbi, 2015). The number n of correct results was given by $P(n) = \binom{m}{n} q^n (1-q)^{m-n}$, where m was the total number of observations in the training set, and $q = 0.5$. To compute the binomial density we let $a = n/m$, and approximated its density as $\rho(a) \approx P(na) / \Delta a$ for compartments of size Δa . This satisfied the requirement that the integral of $\rho(a)$ over $0 \leq a \leq 1$ equals unity as $m \rightarrow \infty$. However before proceeding to use the binomial test in our study, we chose to validate our implementation of it by performing a set of permutation tests as follows. The labels of the training observations were randomly reshuffled, followed by classification and CV. Then this procedure was repeated a large number of times. The resulting set of values was used to construct the null density of accuracy. The normalized histograms from the permutation tests were calculated using $\Delta a = 0.02$. Figure 3 shows the results of the binomial model (solid curve) and permutation tests with 100 repetitions (shaded area) for 4 representative subjects. These data from all subjects

suggested that the null density followed the binomial model. We defined the p-value of a particular accuracy a as the area under the binomial density for x-axis $\geq a$.

This procedure allowed us to calculate individual subjects' p-values. However the significance of the group result still remained to be determined. To investigate this we formulated the global null hypothesis that the classification was no better than chance for at least one of the tests.

Rejecting this hypothesis amounted to the strong statement that none of the accuracy values was drawn from the null distribution. In testing this hypothesis we utilized the classical Bonferroni procedure which insures that if each of the k tests has $p < \alpha / k$, then the global null hypothesis will be falsely rejected with a probability no greater than α . We set $\alpha = 0.05$ and k was taken as either the number of subjects (for subject specific trained SVM) or the number of cross validations (for group trained SVM).

In order to test the statistical significance of the difference in accuracy between pairs of systems (e.g. EEG and fNIRS) we formulated a separate null hypothesis. This stated that the subject dependent accuracies of both systems were drawn from the same distribution. We chose the relatively conservative, non-parametric Kolmogorov-Smirnov (KS) test to evaluate this null hypothesis.

RESULTS

We now present the results of mental state decoding which implemented the methods described in the previous section. Since decoding was based on features generated from short time windows we will first present examples of time series of representative feature types (Figure 4). The accuracy results will be presented for subject specific trained (Figure 5) and group trained (Figure 8) classifiers. We will present the variation of accuracy in time (Figure 6), accuracy will be examined further in the form of receiver operating characteristics (Figure 7), and the dependence of fNIRS accuracy on small subsets of frontal sensors will be shown (Figure 9). Our approach allowed us to interpret PCs as topographic activation maps (Figure 10) and also examine the most discriminating feature types for individual subjects (Table 2).

Regarding individual features, Figure 4 presents examples of the subject and block averaged quantities (thick solid curves) obtained from EEG (left column), fNIRS (middle), and multimodal data (right). Every block contains an initial Rest episode, followed by Task, Rest, and Control

episodes, each lasting 30 s, as described earlier in the subsection Study Design. During the Task the subjects were asked to generate words belonging to a category and during Control they went over the days of the week, as described earlier in the subsection Study Design. The standard deviation of variability across subjects and blocks is shown by the shaded areas. To better illustrate the variability of response, the traces for individual subjects during one block of the experiment are individually displayed in different colors (thin solid curves) in the top row. These individual traces are omitted from the remaining rows for clarity. The sharp and sustained decline in alpha power during the Task period (C), slight increase in the theta (B), and the hemodynamic response patterns (D and E) were consistent with previous literature (Gevins and Smith, 2006; Schecklmann et al., 2008). The right column presents the new type of hybrid feature made possible by EEG+fNIRS. Although they look similar to the hemodynamic data, they are not identical to them and one of the goals of our study was to quantify the additional information they potentially contained. The data in Figure 4 were computed for each non-overlapping 2 s window and smoothed by using 4 neighboring windows for clarity.

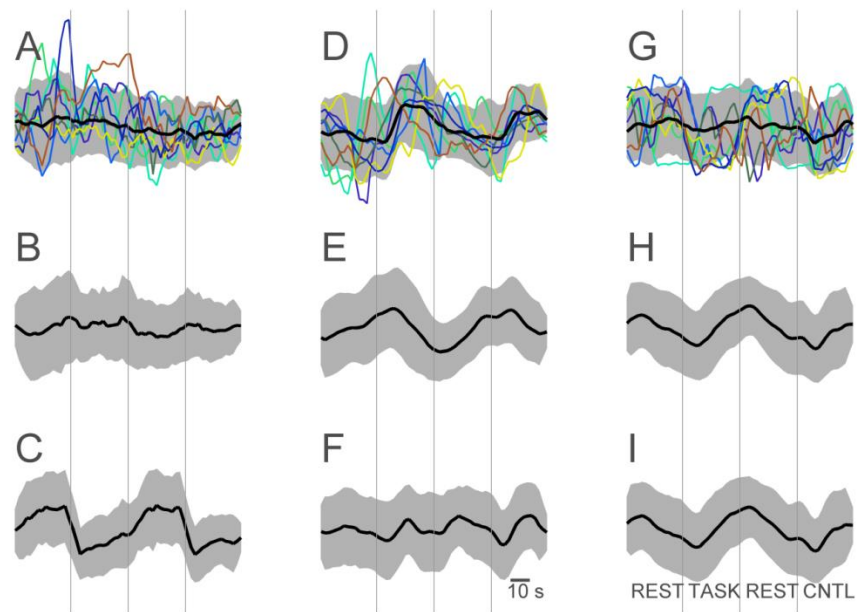


Figure 4 Block and subject averaged measures (thick solid curves) derived from EEG (A-C), fNIRS (D-F), and EEG+fNIRS (G-I). Shaded areas represent the standard deviation of variability across subjects and blocks. (A) Frequency band power (FBP) 0-4 Hz at F8; (B) FBP 4-8 Hz F8; (C) FBP 8-12 Hz PZ; (D) Oxy-hemoglobin concentration changes at T3; (E) Deoxy-hemoglobin concentration changes at T3; (F) Correlation between oxy- and deoxy-hemoglobin; (G) Correlation between FBP 0-4 Hz and deoxy-hemoglobin at T3; (H) Correlation between FBP 4-8 Hz and deoxy-hemoglobin at T3; (I) Correlation between FBP 8-12 Hz and deoxy-hemoglobin at T3. All quantities

have been centered and normalized, hence the vertical axis is in arbitrary units. Vertical gray lines represent boundaries between the Rest, Task, and Control periods. The top row displays responses of individual subjects during one block (thin solid curves in different colors). Individual responses are omitted from other panels.

We performed binary classification of Task v Control (Figure 5A) and Task v Rest (B) with three systems and two feature sets (see Table 1), with features calculated from 2 s windows. The classifier training was subject specific and validated by LOOCV. The bars indicate the subject mean of accuracy while the error bars indicate the standard deviation of subject variability. Any pairs of bars which differ beyond chance (via KS test) are indicated by asterisk. fNIRS has greater accuracy than EEG and Hybrid has the highest accuracy, in Task v Control (A). In the Expanded feature set the Hybrid accuracy is significantly higher than that of either system ($p < 0.05$). Comparing Basic with Expanded, the EEG and fNIRS results show that PLV and COR features contributed small increases to the accuracy. Similarly comparing the Hybrid results show that including the neurovascular features helped increase the accuracy from about 70% to 77%. In classifying Task v Rest (B), fNIRS and Hybrid performed about equally but both did significantly better than EEG. On the whole the accuracy of classifying Task v Rest (B) was higher than that of Task v Control (A).

In accordance with the binomial model (see Methods) we calculated a total of 96 ($= 8 \times 3 \times 2 \times 2$) distinct p-values for the accuracies of each subject, system type, feature set (Basic and Expanded), and classification type (Task v Control and Task v Rest). The strong claim that all subjects' classification results were better than random was satisfied only by the Hybrid system in Task v Control, and by all three systems in Task v Rest.

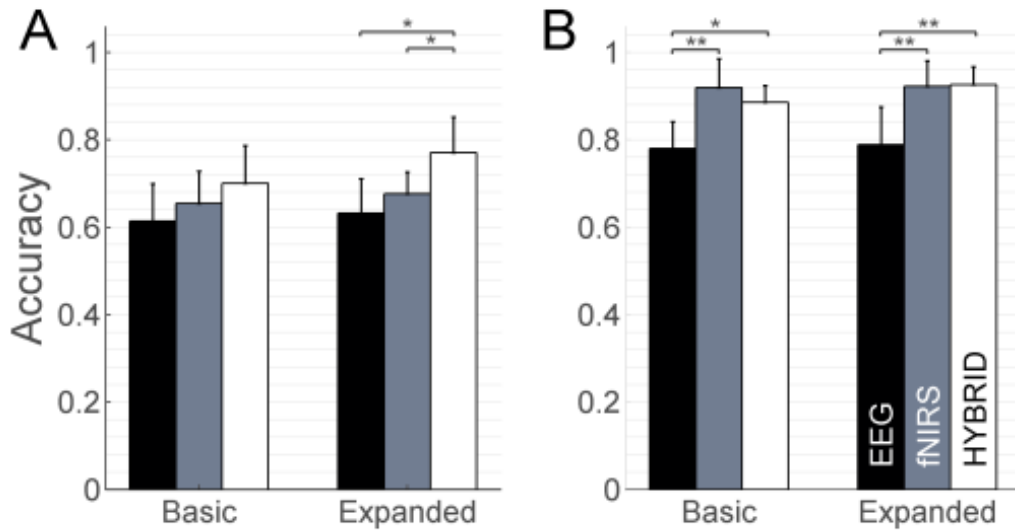


Figure 5 Subject averaged accuracy of classification with subject specific trained SVMs. (A) Task v Control. (B) Task v Rest. Feature window size $\Delta T = 2$ s. System types are EEG (black), fNIRS (gray), and Hybrid (white). Error bars indicate the standard deviation of inter-subject variability. Pairs that were significantly different (according to Kolmogorov-Smirnov test) are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$).

It is well known that fNIRS measures hemodynamic responses that are substantially slower than the neuronal responses measured by EEG (e.g. Buccino et al., 2016). This suggests that the accuracy of EEG should be higher in the initial segment of an episode, before fNIRS signals have had time to respond. In order to investigate this effect, we examined the accuracy of individual feature windows. Figure 6 shows that in decoding Task v Control, fNIRS and Hybrid accuracies tended to increase with time after the start of an episode, while that of the EEG system remained approximately constant. In decoding Task v Rest, however, all three systems' accuracies increased with time. In these calculations the standard deviation of variability across subjects and blocks were relatively large (>0.2) and not shown in this figure in order to avoid visual crowding. The large standard deviations resulted from the fact that separately examining individual feature windows reduced the size of available samples.

So far we presented only subject averaged results and overall accuracies. Next we investigated the ability to identify Task windows against a baseline. The ability to do so can be characterized by sensitivity and specificity (also called the receiver operating characteristics (ROC)). **Figure 7** shows the ROC of individual subjects for baseline as Control (A) and Rest (B). Each subject is shown by a unique marker type. The systems shown separately are EEG (black), fNIRS (red),

and Hybrid (green). Figure 7A indicates that EEG and fNIRS both had low specificity while fNIRS had a substantially higher sensitivity than EEG. That is they performed roughly equally in identifying Control windows, while fNIRS did better in identifying Task. It is also clear that Hybrid generally maintained the higher sensitivity of fNIRS while improving on its specificity. Hybrid results were closest to the upper left corner of the plot which was the optimal location for a system to be. Tracking individual subjects' ROC mostly confirm this conclusion (e.g. square, diamond, inverted triangle, and star markers) with one exception (triangle). In the Task v Rest problem, the sensitivity and specificity values appeared to be similar and improved in going from EEG to fNIRS or Hybrid.

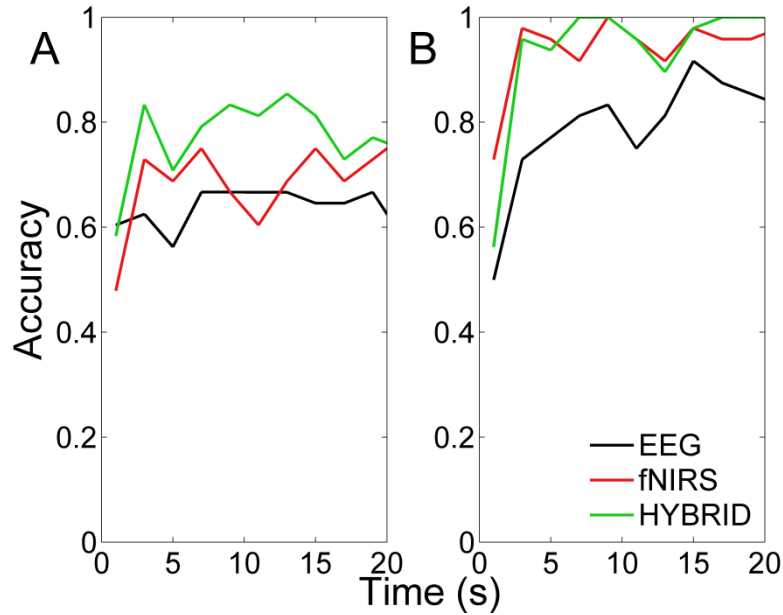


Figure 6 Subject and block averaged accuracy as a function of time. (A) Task v Control. (B) Task v Rest. Feature window size $\Delta T = 2$ s. System types are EEG (black), fNIRS (red), and Hybrid (green).

The classifiers used in the foregoing results were trained individually for each subject. We next performed classification for the same feature groups and systems while the data from all subjects were pooled and the SVM trained only once. The bar height and error bars in Figure 8 represent the mean and standard deviation of the accuracies found from 10-fold CV. Compared to Figure 5 accuracies were generally lower. In the case of Task v Rest, all CV results had better-than-chance accuracy (i.e. the global null hypothesis was rejected) for all three systems and both feature sets. For Task v Control, no system or feature set was able to reject the global null hypothesis since there was at least one CV result with a p-value greater than the Bonferroni

corrected threshold. The pairs of accuracies that differed significantly (via KS test) are indicated by asterisk in Figure 8. In Task v Rest, the highest accuracy was obtained with the Hybrid system using the Expanded feature set, but it was not significantly more accurate than fNIRS.

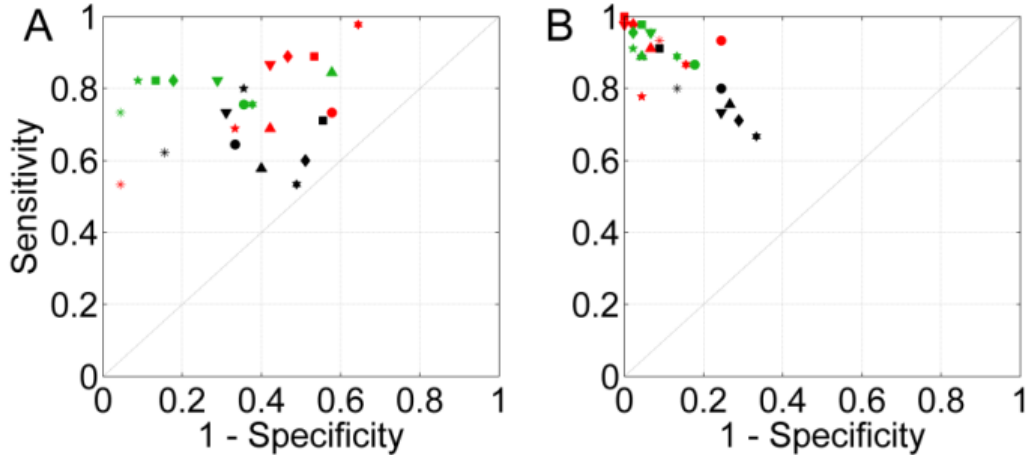


Figure 7 Receiver operating characteristics of individual subjects. (A) Task v Control. (B) Task v Rest. Feature window size $\Delta T = 2$ s. System types are EEG (black), fNIRS (red), and Hybrid (green). Each subject is represented by a different marker type.

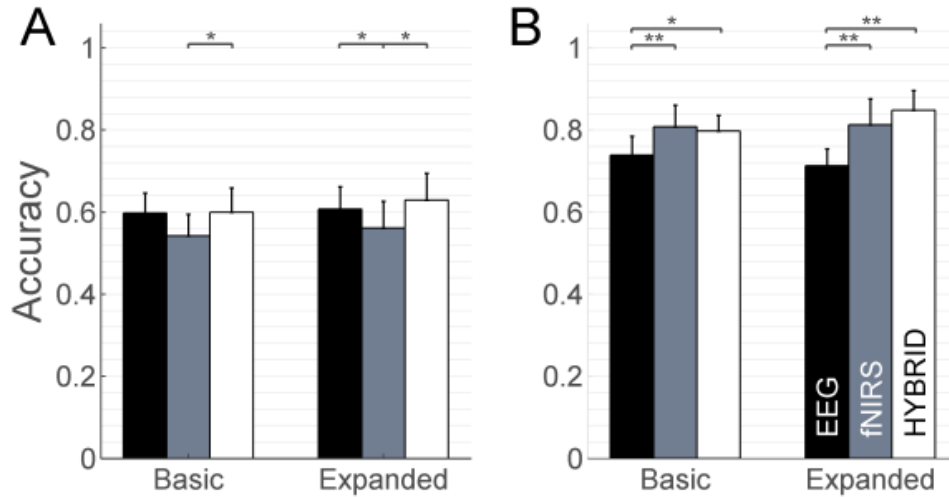


Figure 8 Subject averaged accuracy of classification with non-subject specific SVMs. (A) Task v Control. (B) Task v Rest. Feature window size $\Delta T = 2$ s. System types are EEG (black), fNIRS (gray), and Hybrid (white). Error bars indicate the standard deviation of the variability from 10-fold CV. Pairs that were significantly different (according to Kolmogorov-Smirnov test) are indicated by asterisks (* $p < 0.05$; ** $p < 0.01$).

Results presented so far utilized whole-head EEG+fNIRS channels. We next focused only on the fNIRS system and investigated the performance of small subsets of fNIRS channels. We chose the small subset from the frontal area since prefrontal activations accompany cognitive tasks and many researchers collect fNIRS only frontally. We chose a 7 channel montage consisting of F7, F3, FZ, F4, F8, FP1, and FP2, and an even smaller one only containing only 5 channels: F3, FZ, F4, FP1, and FP2. Figure 9 shows that there were small decreases in accuracy as compared to the full set (red curve, with shaded area which indicates the standard deviation of subject variability). In Task v Control, both of the small frontal montages performed about as well as the full one. In Task v Rest, the 4 channel montage (blue) showed a substantially smaller accuracy.

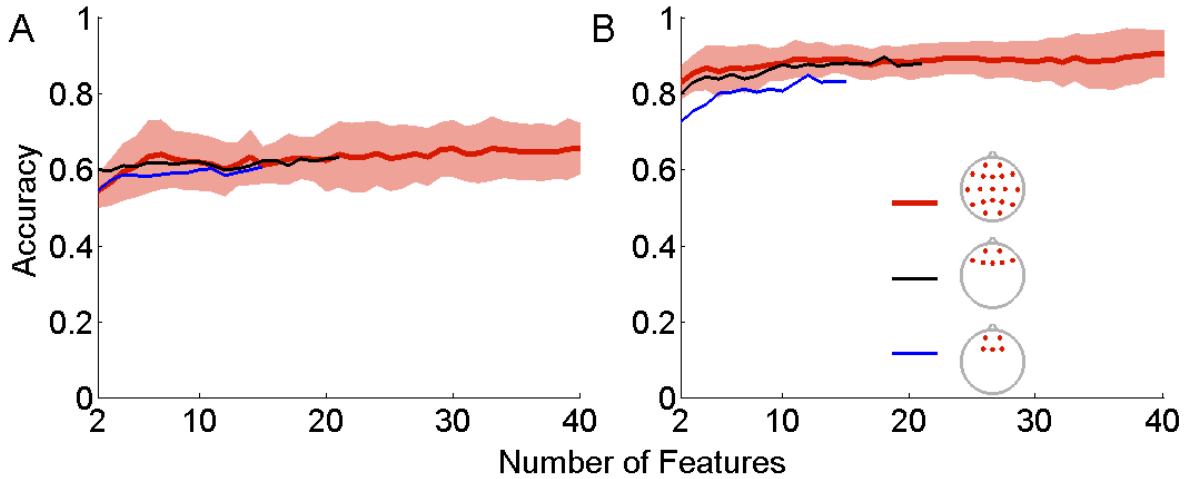


Figure 9 Subject averaged accuracy of the fNIRS from whole-head montage compared to those of frontal montages. Classifier training was subject specific. (A) Task v Control. (B) Task v Rest. Feature window size $\Delta T = 2$ s. Shaded area shows inter-subject variability of the full montage. The x-axis represents the number n of features included for each montage. The set of n features is selected by including only the top features sorted according to R^2 .

Figure 10 shows the most discriminating PCs for multiple types of features from non-subject specific classification. Under each PC the histogram of the associated time series are shown for Task (black curve) and Rest (gray). The degree to which these histograms are separated is an indication of the discrimination ability of the PC. The PCs in Figure 10 are chosen from items listed in the rightmost section of Table 2. Figure 10 contains the top three PCs from among EEG features (left column), fNIRS features (middle), and Hybrid (right). The EEG features are the frequency band powers in various ranges. Third from top in the left column is the alpha power

(Figure 10C) with a mainly posterior presence. Its histograms show that the alpha power declines in the Task relative to Rest (black curve is shifted to the left relative to the gray curve). The first PC in the middle column (Figure 10D) is *HBO* and the histogram indicates that it becomes more positive in Task relative to Rest. The PC has positive values in the frontal and occipital areas, and mainly negative ones elsewhere. Figure 10E and F show activations under the inferior electrode sites that overlap with the temporalis muscle. The right column shows that the most discriminating Hybrid features are the correlation of deoxy-hemoglobin with the EEG frequency bands in various beta ranges of frequency. The PCs indicate that these correlations are positive in the frontal and parietal-occipital areas and negative elsewhere, and the correlations tend to switch their sign during the transitions between Task and Rest. These neurovascular features also show a pronounced left-right asymmetry.

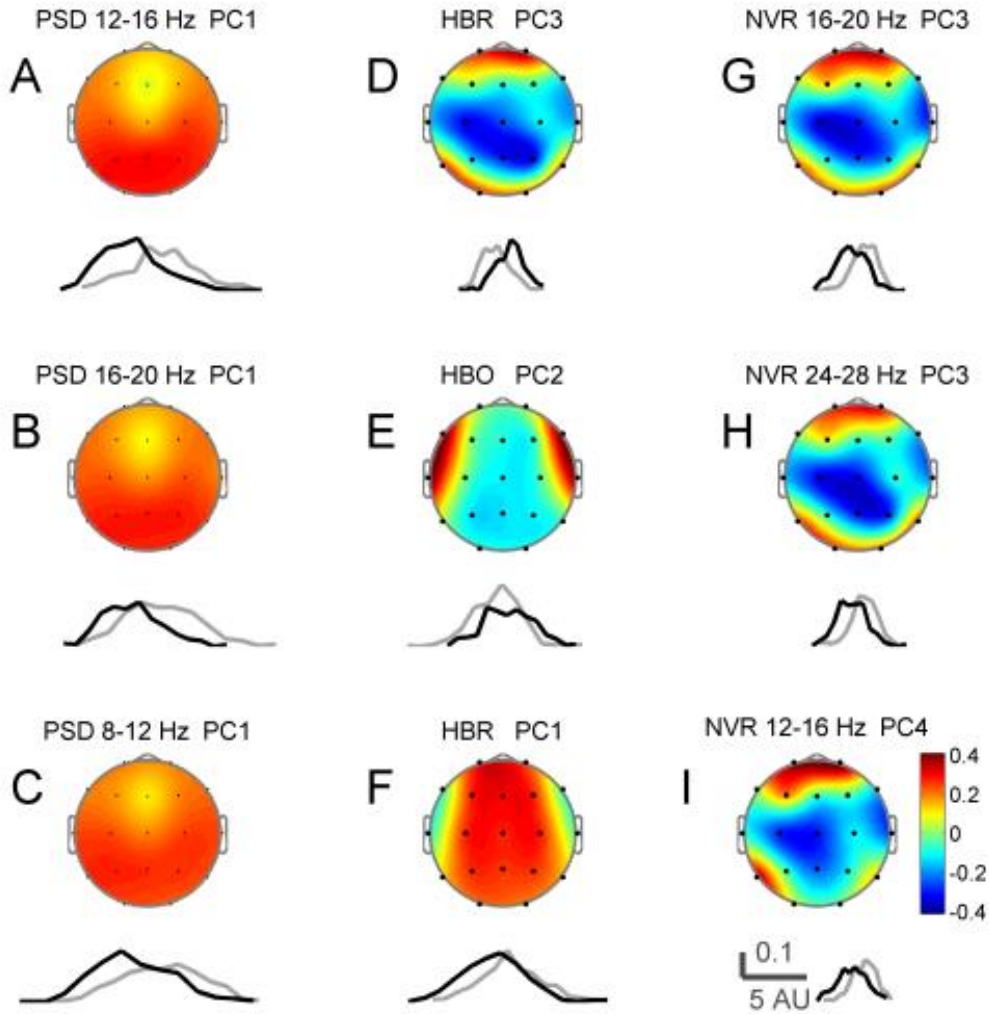


Figure 10 Principal components for the non-subject specific classification of Task v Rest. The top 3 PCs (according to R^2) from each type of system are shown. These PCs can be found in the rightmost section of Table 2. Below each PC are shown the histograms of its associated time series for Task (black curve) and Rest (gray).

We show in Table 2 the top ten most discriminating (according to R^2) features for three individual subjects and for the non-subject specific Task v Rest classification problem. The table indicates that different feature types tended to be the most effective ones for different subjects. For example top few features in Subject 1 were based on EEG frequency power, in Subject 2 and 3 they were based on deoxy-hemoglobin concentration change and neurovascular correlations. The table also indicates that in every subject the neurovascular features have a strong presence among the most discriminating features.

Subject 1				Subject 2				Subject 3				Non-Subject Specific			
Type	Freq Range (Hz)	PC #	R^2	Type	Freq Range (Hz)	PC #	R^2	Type	Freq Range (Hz)	PC #	R^2	Type	Freq Range (Hz)	PC #	R^2
PSD	12-16	1	0.58867	HBR	n.a.	3	0.39850	NVO	12-16	1	0.39369	NVR	16-20	3	0.18100
			0.48829				0.35305				0.38268				0.17945
PSD	20-24	1	0.47814	NVR	28-32	3	0.34137	NVO	20-24	1	0.36410	HBR	n.a.	3	0.17454
PSD	8-12	1	0.45967	NVR	24-28	3	0.32333	HBO	n.a.	2	0.35592	PSD	12-16	1	0.17057
PSD	16-20	1	0.33577	NVR	20-24	3	0.29164	NVO	16-20	1	0.34908	PSD	16-20	1	0.14416
HBR	n.a.	1	0.33299	NVR	16-20	3	0.28582	NVO	24-28	1	0.34638	NVR	24-28	3	0.13990
PSD	24-28	1	0.32835	NVR	8-12	2	0.28537	PSD	16-20	1	0.34381	NVR	12-16	4	0.13849
PSD	28-32	1	0.28116	NVR	12-16	3	0.27594	NVO	28-32	1	0.32710	NVR	20-24	3	0.13461
NVR	16-20	1	0.27428	HBO	n.a.	2	0.23226	NVO	0-4	1	0.32707	NVR	28-32	3	0.11562
NVR	20-24	1	0.26495	NVR	0-4	1	0.19173	NVO	8-12	1	0.32241	PSD	8-12	1	0.11438
NVR	4-8	1		NVO	24-28	2		NVO	4-8	1		PSD	20-24	1	

Table 2 Top ranked (according to R^2) features for the classification of Task v Rest. The columns are type of feature, associated frequency range (if applicable), the rank of the PC according to its energy (λ), and the squared Pearson correlation with the labels (R^2).

DISCUSSION

In this paper we have shown that the Hybrid system tends to perform significantly better than the EEG or fNIRS systems. On the whole our results indicated that concurrent EEG and fNIRS work synergistically rather than dilute each other's capabilities. (We would have inferred dilution had the Hybrid accuracy been intermediate between EEG and fNIRS.) By examining subsets of

features we could resolve the Hybrid advantage into mainly two factors. Firstly, pooling of neural and hemodynamic features provides robustness to the system since different types of features perform best in different subjects (Table 2). And secondly, neurovascular features that are only available to the Hybrid system likely enhance the accuracy (Figure 5 Figure 8). The fact that a similar mutual enhancement of EEG and fNIRS was observed during n-back task performance (Aghajani et al., 2017) suggests that the effect is generalizable. The neuropsychological demands of verbal fluency performance are shared by several other standard tests as well as everyday tasks (Unsworth et al., 2010). Nevertheless, the reliability and reproducibility of Hybrid methods' accuracy in tasks other than verbal fluency remain to be explored.

Our results have also suggested that (1) due to considerable inter-subject variability, subject specific training of automated classifiers are required for high accuracy; (2) the well-known slowness of the hemodynamic response lowers the accuracy of the fNIRS system during the initial several seconds of an episode (Figure 6); (3) the higher accuracy of fNIRS relative to that of EEG is in some cases due only to increased sensitivity of detection of the Task state (Figure 7); and (4) that frontal fNIRS montages may forfeit a substantial fraction of their accuracy relative to that of the whole-head montage (Figure 9). This is not surprising since cognitive processes typically activate multiple regions in coordination.

The results we have obtained for EEG and fNIRS are separately consistent with those found in previous unimodal studies. For example the survey by Mehta and Parasuraman (2013) place the accuracy of automated classification by EEG broadly in the range 70-85%. Our values for Task v Rest classification (Figure 5) fit within this scope, although our Task v Control accuracies were somewhat lower. The fNIRS based classification of VFT v Rest in Schudlo and Chau (2015) is an average of 80% which is surpassed by our values in Figure 5B. These are intended only as rough comparisons, as the experiments and types of analysis are not closely matched. The purpose of our study was not to improve or use the best available methods of classification, but rather to compare on an equal footing the performance of EEG+fNIRS with its own subcomponents.

Did the additional features in the Expanded set help enhance performance? We can explore this question in reference to Figure 5. In Task v Control the EEG system's accuracy went from 0.61 to 0.63 when PLV was included, so PLV may be having a small beneficial effect. Similarly, it went from 0.78 to 0.81 in Task v Rest. In neither case did these changes reach statistical significance. There were similar small increases brought about by COR in the fNIRS system,

however they also did not reach statistical significance. Figure 5A shows that neurovascular features may be having a beneficial effect in Task v Control. In the Basic feature set, the accuracy went from 0.65 for fNIRS to 0.70 for Hybrid. In the Expanded feature set, it went from 0.68 for fNIRS to 0.76 for Hybrid. Some of this increase may be due to pooling of the EEG and fNIRS features mentioned above. However the frequent appearance of NVO and NVR among the top ranked features (e.g. Table 2 but also other subjects (not shown)) suggests that neurovascular features may be specifically contributing the high accuracy of the Hybrid system. To our knowledge ours is the first study to use neurovascular features in multimodal classification of mental states.

We have further found that Pearson correlation based feature ranking is an effective method for feature selection, as the first few top ranking features account for most of the accuracy of a system (Figure 9). This also indicates that system size can be drastically reduced if needed without a correspondingly drastic loss in accuracy. This need may arise, for example, if classification needs to be carried out on-board in a wearable system. Figure 9 displays no indication of overfitting, which would have resulted in a decrease in accuracy with increasing number of features when the number of features was sufficiently high. Other results from our data (not shown) for the EEG and Hybrid systems also displayed no evidence of overfitting. This is consistent with the finding of Chu et al. (2012) that feature selection in fMRI does not improve accuracy unless based on domain specific knowledge about brain regions of interest.

We determined that the classification of Task v Rest was easier than that of Task v Control. This is consistent with expectations since Rest may be considered as being further away than Control from the Task state. However this expectation is contingent on an assumption that during Rest the subject's mind remains restful and does not wander excessively. Somewhat contrary to the present result, in a working memory task (n-back) the easiest level (0-back) was better discriminated from more difficult levels than the resting state (Aghajani et al., 2017). The similarity of responses during Task and Control (Figure 4) is expected to the extent that word finding and weekday reciting could have overlapping functional patterns associated with autonomic response, simple mnemonic recall, and language generation (Schecklmann et al., 2010).

Left lateralization of brain activation during verbal fluency is commonly found in healthy individuals (Herrmann et al., 2005; Hutchinson et al., 1999; Schlösser et al., 1998). This is consistent with the small asymmetries observed in our subject-averaged PCs (Figure 10).

Tupak et al. (2012) found that CF task results in greater left lateralization in healthy controls than phonemic verbal fluency. Cortical reorganization of brain function during verbal fluency in aging, and in disorders such as schizophrenia and depression has been found in other imaging studies (Heinzel et al., 2013, 2015; Kahlaoui et al., 2012; Marumo et al., 2014; Shimodera et al., 2012; Takizawa et al., 2008, 2014; Yeung et al., 2016). Using PET, Reuter-Lorenz et al. (2000) found significant left lateralized frontal activation in younger participants but bilateral activations in older participants, suggesting that older subjects may be recruiting other brain areas to compensate for neural declines. Redistribution of brain activity in aging initially occurred without changes in performance. In our study the PCs for individual subjects (not shown) contained inter-subject variability, finer spatial variation, and significant asymmetry particularly in HBO and HBR for some subjects. In preliminary studies we explored the use of asymmetry based fNIRS features but observed no obvious effects on accuracy and therefore excluded them from this paper.

Our study was limited in a number of ways. For example we did not attempt to separate contributions of cerebral from noncerebral sources. PCA can segregate noncerebral activity (typically of higher amplitude than cortical) into distinct PCs as suggested by **Figure 10E** and **F**, which show that the temporalis muscle may have been activated during task performance when subjects clenched their jaw (see Keles et al. (2016)). This may have enhanced the classification performance if changes in muscle activity correlated with different episodes. Classification based on cortical information alone may be more robust and therefore preferable (Zander et al., 2016), however for some passive BCI the origin of the discriminating ability may be less important than its accuracy.

In the present study we used only 2 s windows for feature generation. The choice of window size is customarily driven by a compromise between speed and accuracy. Our preliminary results with the Expanded system (not shown) suggested that the EEG accuracy remained nearly constant and the fNIRS and Hybrid accuracies monotonically decreased as window size was increased from 2 to 10 s. The decreases may be attributable to the availability of fewer observations with increasing window size, and the resulting inability to train the classifier well. However we did not explore this issue further.

Our experimental protocol eliminated variation in subject performance by constraining word generation rate to approximately one word every 5 s. However individuals are known to differ widely in cognitive capacities and recent studies have associated such differences with

differences in brain function and gene expression (Parasuraman and Jiang, 2012). For related reasons our study excluded the analysis of long-term and within-day variability of performance and its neurophysiological correlates (Gevins et al., 2012). These variations are important for personnel selection, training, and design (e.g. chapter 11 of Wickens et al. (2015)).

We used a sparse 10–20 coverage which did not lend itself to an adequate estimate of the current density through the Laplace montage. We have not used any additional EEG electrodes which could have provided a local Laplacian estimate and improve space resolution (Nunez and Pilgreen, 1991). In fNIRS, the effects of systemic artifacts may be investigated, for example by subtracting the global signal from all channels from each channel (common reference correction) or other methods (Bauernfeind et al., 2013; Heinzel et al., 2013, 2015). In addition, the effects of superficial tissue on the fNIRS signals during our experimental procedure may be investigated (Dehaes et al., 2011). We did not pursue these lines of investigation although we used bandpass filtering to remove some physiological artifacts. None of our subjects reported discomfort with the headset used in our experiments but our research did not focus on comfort and convenience. We expect that as prolonged hybrid recordings become increasingly more feasible, the design and usability of hybrid equipment will likely gain in importance in the near future.

The neurovascular features used in this paper consisted of simple zero-lag correlations between hemoglobin concentration changes and EEG frequency band power. Previous work on subjects in the resting state indicates that additional topographically distributed patterns exist in the time lagged correlations (Keles et al., 2016). We have not computed time lagged correlations as extracting such features may require longer feature windows than used in this study. However there is a growing literature that addresses techniques for analyzing multimodal data (Biessmann et al., 2011; Dahne et al., 2013; de Munck et al., 2009) and these may be leveraged for feature extraction in EEG+fNIRS.

CONCLUSION

We investigated EEG+fNIRS as a hybrid technique for decoding mental states in category fluency, a popular psychology test that has implications for medical diagnosis, BCI design, and neuroergonomics. Results suggest that the hybrid system's decoding accuracy is greater than that of its subsystems, and specifically neurovascular features play a prominent role in its enhanced performance. We hope that our results will provide further impetus for developing

mobile/wearable hybrid hardware, decoding methods tailored for hybrid data, and a wide range of practical applications.

REFERENCES

- Aghajani, H., Garbey, M., and Omurtag, A. (2017). Measuring Mental Workload with EEG+fNIRS, under review.
- Bauernfeind, G., Boeck, C., Wriessnegger, S.C., and Müller-Putz, G.R. (2013). Physiological noise removal from fNIRS signals. *Biomed. Eng. Tech.*
- Biessmann, F., Plis, S., Meinecke, F.C., Eichele, T., and Muller, K. (2011). Analysis of Multimodal Neuroimaging Data. *Biomed. Eng. IEEE Rev. In 4*, 26–58.
- Buccino, A.P., Keles, H.O., and Omurtag, A. (2016). Hybrid EEG-fNIRS Asynchronous Brain-Computer Interface for Multiple Motor Tasks. *PloS One 11*, e0146610.
- Buzsáki, G., Anastassiou, C.A., and Koch, C. (2012). The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci. 13*, 407–420.
- Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C., Initiative, A.D.N., and others (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage 60*, 59–70.
- Coffey, E.B., Brouwer, A.-M., and van Erp, J.B. (2012). Measuring workload using a combination of electroencephalography and near infrared spectroscopy. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, (SAGE Publications), pp. 1822–1826.
- Combrisson, E., and Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *J. Neurosci. Methods 250*, 126–136.
- Costafreda, S.G., Fu, C.H., Lee, L., Everitt, B., Brammer, M.J., and David, A.S. (2006). A systematic review and quantitative appraisal of fMRI studies of verbal fluency: role of the left inferior frontal gyrus. *Hum. Brain Mapp. 27*, 799–810.
- Dahne, S., Biessmann, F., Meinecke, F.C., Mehnert, J., Fazli, S., and Muller, K.-R. (2013). Integration of multivariate data streams with bandpower signals. *Multimed. IEEE Trans. On 15*, 1001–1013.
- Dehaes, M., Gagnon, L., Lesage, F., Pélégriani-Issac, M., Vignaud, A., Valabregue, R., Grebe, R., Wallois, F., and Benali, H. (2011). Quantitative investigation of the effect of the extra-cerebral vasculature in diffuse optical imaging: a simulation study. *Biomed. Opt. Express 2*, 680–695.
- Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods 134*, 9–21.
- Delpy, D.T., Cope, M., van der Zee, P., Arridge, S.R., Wray, S., and Wyatt, J.S. (1988). Estimation of optical pathlength through tissue from direct time of flight measurement. *Phys. Med. Biol. 33*, 1433.

- Durduran, T. (2010). Optical measurement of cerebral hemodynamics and oxygen metabolism in neonates with congenital heart defects. *J. Biomed. Opt.* 15.
- Dvorak, D., and Fenton, A.A. (2014). Toward a proper estimation of phase–amplitude coupling in neural oscillations. *J. Neurosci. Methods* 225, 42–56.
- Fazli, S., Mehnert, J., Steinbrink, J., Curio, G., Villringer, A., Müller, K.-R., and Blankertz, B. (2012). Enhanced performance by a hybrid NIRS–EEG brain computer interface. *Neuroimage* 59, 519–529.
- Frith, C.D., Friston, K.J., Liddle, P.F., and Frackowiak, R.S.J. (1991). A PET study of word finding. *Neuropsychologia* 29, 1137–1148.
- Gentili, R.J., Shewokis, P.A., Ayaz, H., and Contreras-Vidal, J.L. (2013). Functional near-infrared spectroscopy-based correlates of prefrontal cortical dynamics during a cognitive-motor executive adaptation task. *Front. Hum. Neurosci.* 7.
- Gevins, A., and Smith, M.E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cereb. Cortex* 10, 829–839.
- Gevins, A., and Smith, M.E. (2006). Electroencephalography (EEG) in neuroergonomics. *Neuroergonomics Brain Work* 15–31.
- Gevins, A., McEvoy, L.K., Smith, M.E., Chan, C.S., Sam-Vargas, L., Baum, C., and Ilan, A.B. (2012). Long-term and within-day variability of working memory performance and EEG in individuals. *Clin. Neurophysiol.* 123, 1291–1299.
- Gudiño-Mendoza, B., Sossa, H., Sanchez-Ante, G., and Antelis, J.M. (2016). Classification of Motor States from Brain Rhythms Using Lattice Neural Networks. In *Mexican Conference on Pattern Recognition*, (Springer), pp. 303–312.
- Haynes, J.-D., and Rees, G. (2006). Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7, 523–534.
- Heinzel, S., Metzger, F.G., Ehlis, A.-C., Korell, R., Alboji, A., Haeussinger, F.B., Hagen, K., Maetzler, W., Eschweiler, G.W., Berg, D., et al. (2013). Aging-related cortical reorganization of verbal fluency processing: a functional near-infrared spectroscopy study. *Neurobiol. Aging* 34, 439–450.
- Heinzel, S., Metzger, F.G., Ehlis, A.-C., Korell, R., Alboji, A., Haeussinger, F.B., Wurster, I., Brockmann, K., Suenkel, U., Eschweiler, G.W., et al. (2015). Age and vascular burden determinants of cortical hemodynamics underlying verbal fluency. *PLoS One* 10, e0138863.
- Henry, J.D., and Crawford, J.R. (2004). A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology* 18, 621.
- Herrmann, M.J., Ehlis, A.-C., Scheuerpflug, P., and Fallgatter, A.J. (2005). Optical topography with near-infrared spectroscopy during a verbal-fluency task. *J. Psychophysiol.* 19, 100–105.

- Huppert, T.J., Hoge, R.D., Diamond, S.G., Franceschini, M.A., and Boas, D.A. (2006). A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans. *Neuroimage* 29, 368–382.
- Hutchinson, M., Schiffer, W., Joseffer, S., Liu, A., Schlosser, R., Dikshit, S., Goldberg, E., and Brodie, J.D. (1999). Task-specific deactivation patterns in functional magnetic resonance imaging. *Magn. Reson. Imaging* 17, 1427–1436.
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M.J., Iragui, V., and Sejnowski, T.J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37, 163–178.
- Kahlaoui, K., Di Sante, G., Barbeau, J., Maheux, M., Lesage, F., Ska, B., and Joanne, Y. (2012). Contribution of NIRS to the study of prefrontal cortex for verbal fluency in aging. *Brain Lang.* 121, 164–173.
- Keles, H.O., Barbour, R.L., and Omurtag, A. (2016). Hemodynamic correlates of spontaneous neural activity measured by human whole-head resting state EEG + fNIRS. *NeuroImage* 138, 76–87.
- Kleinschmidt, A. (1996). Simultaneous recording of cerebral blood oxygenation changes during human brain activation by magnetic resonance imaging and near-infrared spectroscopy. *Journal of cerebral blood flow and metabolism : official. J. Int. Soc. Cereb. Blood Flow Metab.* 16–5.
- Le Van Quyen, M., Foucher, J., Lachaux, J.-P., Rodriguez, E., Lutz, A., Martinerie, J., and Varela, F.J. (2001). Comparison of Hilbert transform and wavelet methods for the analysis of neuronal synchrony. *J. Neurosci. Methods* 111, 83–98.
- Liu, Y., Ayaz, H., and Shewokis, P.A. (2017). Mental workload classification with concurrent electroencephalography and functional near-infrared spectroscopy. *Brain-Comput. Interfaces* 1–11.
- Lloyd-Fox, S., Blasi, A., and Elwell, C.E. (2010). Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neurosci. Biobehav. Rev.* 34, 269–284.
- Ma, Y., Shaik, M.A., Kozberg, M.G., Kim, S.H., Portes, J.P., Timmerman, D., and Hillman, E.M. (2016). Resting-state hemodynamics are spatiotemporally coupled to synchronized and symmetric neural activity in excitatory neurons. *Proc. Natl. Acad. Sci.* 201525369.
- Marumo, K., Takizawa, R., Kinou, M., Kawasaki, S., Kawakubo, Y., Fukuda, M., and Kasai, K. (2014). Functional abnormalities in the left ventrolateral prefrontal cortex during a semantic fluency task, and their association with thought disorder in patients with schizophrenia. *Neuroimage* 85, 518–526.
- Mehta, R.K., and Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.* 7, 889.
- Mesquita, R.C., Franceschini, M.A., and Boas, D.A. (2010). Resting state functional connectivity of the whole head with near-infrared spectroscopy. *Biomed. Opt. Express* 1, 324–336.
- de Munck, J.C., Gonçalves, S.I., Mammoliti, R., Heethaar, R.M., and Da Silva, F.L. (2009). Interactions between different EEG frequency bands and their effect on alpha-fMRI correlations. *Neuroimage* 47, 69–76.

- Mwangi, B., Tian, T.S., and Soares, J.C. (2014). A review of feature reduction techniques in neuroimaging. *Neuroinformatics* 12, 229–244.
- Nicolas-Alonso, L.F., and Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors* 12, 1211–1279.
- Nunez, P.L., and Pilgreen, K.L. (1991). The spline-Laplacian in clinical neurophysiology: a method to improve EEG spatial resolution. *J. Clin. Neurophysiol.* 8, 397–413.
- Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G., and Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. *Neurosci. Biobehav. Rev.* 36, 1140–1152.
- Parasuraman, R., and Jiang, Y. (2012). Individual differences in cognition, affect, and performance: Behavioral, neuroimaging, and molecular genetic approaches. *Neuroimage* 59, 70–82.
- Petroff, O.A., Spencer, D.D., Goncharova, I.I., and Zaveri, H.P. (2016). A comparison of the power spectral density of scalp EEG and subjacent electrocorticograms. *Clin. Neurophysiol.* 127, 1108–1112.
- Putze, F., Hesslinger, S., Tse, C.-Y., Huang, Y., Herff, C., Guan, C., and Schultz, T. (2015). Hybrid fNIRS-EEG based classification of auditory and visual perception processes. *Using Neurophysiol. Signals Reflect Cogn. Affect. State* 180.
- Reuter-Lorenz, P.A., Jonides, J., Smith, E.E., Hartley, A., Miller, A., Marshuetz, C., and Koeppel, R.A. (2000). Age differences in the frontal lateralization of verbal and spatial working memory revealed by PET. *J. Cogn. Neurosci.* 12, 174–187.
- Schecklmann, M., Ehlis, A.-C., Plichta, M.M., and Fallgatter, A.J. (2008). Functional near-infrared spectroscopy: a long-term reliable tool for measuring brain activity during verbal fluency. *Neuroimage* 43, 147–155.
- Schecklmann, M., Ehlis, A.C., Plichta, M.M., and Fallgatter, A.J. (2010). Influence of muscle activity on brain oxygenation during verbal fluency assessed with functional near-infrared spectroscopy. *Neuroscience* 171, 434–442.
- Schlösser, R., Hutchinson, M., Joseffer, S., Rusinek, H., Saarimaki, A., Stevenson, J., Dewey, S.L., and Brodie, J.D. (1998). Functional magnetic resonance imaging of human brain activity in a verbal fluency task. *J. Neurol. Neurosurg. Psychiatry* 64, 492–498.
- Scholkman, F., Kleiser, S., Metz, A.J., Zimmermann, R., Pavia, J.M., Wolf, U., and Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. *Neuroimage* 85, 6–27.
- Schudlo, L.C., and Chau, T. (2015a). Towards a ternary NIRS-BCI: single-trial classification of verbal fluency task, Stroop task and unconstrained rest. *J. Neural Eng.* 12, 066008.
- Schudlo, L.C., and Chau, T. (2015b). Single-trial classification of near-infrared spectroscopy signals arising from multiple cortical regions. *Behav. Brain Res.* 290, 131–142.

- Shimodera, S., Imai, Y., Kamimura, N., Morokuma, I., Fujita, H., Inoue, S., and Furukawa, T.A. (2012). Mapping hypofrontality during letter fluency task in schizophrenia: a multi-channel near-infrared spectroscopy study. *Schizophr. Res.* 136, 63–69.
- Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S., and Obrig, H. (2006). Illuminating the BOLD signal: combined fMRI–fNIRS studies. *Magn. Reson. Imaging* 24, 495–505.
- Strangman, G., Culver, J.P., Thompson, J.H., and Boas, D.A. (2002). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *Neuroimage* 17, 719–731.
- Takahashi, T., Takikawa, Y., Kawagoe, R., Shibuya, S., Iwano, T., and Kitazawa, S. (2011). Influence of skin blood flow on near-infrared spectroscopy signals measured on the forehead during a verbal fluency task. *Neuroimage* 57, 991–1002.
- Takizawa, R., Kasai, K., Kawakubo, Y., Marumo, K., Kawasaki, S., Yamasue, H., and Fukuda, M. (2008). Reduced frontopolar activation during verbal fluency task in schizophrenia: a multi-channel near-infrared spectroscopy study. *Schizophr. Res.* 99, 250–262.
- Takizawa, R., Fukuda, M., Kawasaki, S., Kasai, K., Mimura, M., Pu, S., Noda, T., Niwa, S., Okazaki, Y., Group, J.P. for P.A. of N.-I.S. (JPSY-N., et al. (2014). Neuroimaging-aided differential diagnosis of the depressive state. *Neuroimage* 85, 498–507.
- Tomita, Y., Vialatte, F.-B., Dreyfus, G., Mitsukura, Y., Bakardjian, H., and Cichocki, A. (2014). Bimodal BCI using simultaneously NIRS and EEG. *IEEE Trans. Biomed. Eng.* 61, 1274–1284.
- Tupak, S.V., Badewien, M., Dresler, T., Hahn, T., Ernst, L.H., Herrmann, M.J., Fallgatter, A.J., and Ehlis, A.-C. (2012). Differential prefrontal and frontotemporal oxygenation patterns during phonemic and semantic verbal fluency. *Neuropsychologia* 50, 1565–1569.
- Ugialoro, A. (2014). Cerebral Monitoring and Surveillance Using High-Resolution Functional Optical Imaging. *Neurovascular Coupling Methods, Neuromethods*, eds Zhao. M Ma H Schwartz TH Springer N. Y. Vol 88, 307–330.
- Unsworth, N., Spillers, G.J., and Brewer, G.A. (2010). Variation in verbal fluency: A latent variable analysis of clustering, switching, and overall performance. *Q. J. Exp. Psychol.* 64, 447–466.
- Varela, F., Lachaux, J.-P., Rodriguez, E., and Martinerie, J. (2001). The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239.
- White, B.R., Snyder, A.Z., Cohen, A.L., Petersen, S.E., Raichle, M.E., Schlaggar, B.L., and Culver, J.P. (2009). Resting-state functional connectivity in the human brain revealed with diffuse optical tomography. *Neuroimage* 47, 148–156.
- Wickens, C.D., Hollands, J.G., Banbury, S., and Parasuraman, R. (2015). *Engineering psychology & human performance* (Psychology Press).

- Yeung, M.K., Sze, S.L., Woo, J., Kwok, T., Shum, D.H., Yu, R., and Chan, A.S. (2016). Altered frontal lateralization underlies the category fluency deficits in older adults with mild cognitive impairment: a near-infrared spectroscopy study. *Front. Aging Neurosci.* 8.
- Zander, T.O., and Kothe, C. (2011). Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general. *J. Neural Eng.* 8, 025005.
- Zander, T.O., Shetty, K., Lorenz, R., Leff, D.R., Krol, L.R., Darzi, A.W., Gramann, K., and Yang, G.-Z. (2016). Automated Task Load Detection with Electroencephalography: Towards Passive Brain–Computer Interfacing in Robotic Surgery. *J. Med. Robot. Res.* 1–10.
- Zhang, D., and Raichle, M.E. (2010). Disease and the brain’s dark energy. *Nat. Rev. Neurol.* 6, 15–28.
- Zhang, S., Chen, H.-S., and Pfeiffer, R.M. (2013). A combined p-value test for multiple hypothesis testing. *J. Stat. Plan. Inference* 143, 764–770.