Monitoring Attentional State with Functional Near Infrared Spectroscopy

by

Angela Rose Harrivel

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biomedical Engineering)
in the University of Michigan
2014

Doctoral Committee:

Professor Douglas C. Noll, Co-Chair
Associate Research Scientist Scott J. Peltier, Co-Chair
Research Associate Professor Luis Hernandez-Garcia
Associate Professor Theodore James Huppert, University of Pittsburgh
Associate Professor Daniel Howard Weissman

Dedication

To everyone who told me I could.

Acknowledgements

This work was supported by the University of Michigan fMRI Laboratory, the Vehicle Systems Safety Technologies Project in NASA's Aviation Safety Program, and the Glenn Research Center Graduate Studies Program. Colleagues at the NASA Glenn and Langley Research Centers are appreciated for their encouragement, help and support, especially: Terri McKay and Kara Latorella as colleagues in prior work; Eric Baumann, Jeffrey Mackey, Daniel Gotti and Padetha Tin for fNIRS head probe development; and Alan Hylton, Tristan Hearn and Alan Pope for helpful review. We are grateful for the assistance of: members of Ted Huppert's laboratory group at the University of Pittsburgh, the fMRI Laboratory at the University of Michigan for the collection and pre-processing of the fMRI data, Ryan Smith and Keith Newnham for scanning and assistance with fMRI-compatible materials, Krisanne Litinas for fMRI data processing help, Ruth Halsey for study administration, Charles Nicholas for IT help, Yash Shah for assistance with classifier implementation, the fNIRS group within the Center for Human Growth and Development at the University of Michigan, and especially all fMRI Laboratory group members and staff for regular feedback, help, and encouragement. Finally, the participation of all human subject study volunteers is greatly appreciated.

Angela deeply appreciates the patience, support and encouragement of close friends and family throughout the entire PhD process. This is particularly true for Sam Straight.

Table of Contents

List of Tables

Table

List of Figures

Figure

List of Appendices

Appendix

Abstract

Functional Near Infrared Spectroscopy (fNIRS) is a technique for quantifying hemodynamic activity in the brain. Its portability allows application in real world operational contexts. The ability to distinguish levels of task engagement in safety-critical situations is important for detecting and preventing attentional performance decrement. We therefore investigated whether fNIRS can be used to distinguish between high and low levels of task engagement during the performance of a selective attention task, and validated these results using functional magnetic resonance imaging (fMRI) as a gold standard. Participants performed the multi-source interference task (MSIT) while we recorded brain activity with fNIRS from two brain regions. One was a key region of the "task-positive" network, which is associated with relatively high levels of task engagement. The second was a key region of the "task-negative" network, which is associated with relatively low levels of task engagement (e.g., resting and not performing a task). Using activity in these regions as inputs to a multivariate pattern classifier, we were able to predict above chance levels whether participants were engaged in performing the MSIT or resting. Classifier input features were selected from an array of probe channels at each of the two locations based on the fit to a model of expected task activity, or on training data. Standard linear regression was implemented with both static and adaptive general linear models to remove concurrently measured physiological noise. Two types of models were used to process the fNIRS signals. One employed knowledge of the task being performed to determine the system's best capability. The other did not, for a realistic characterization. We were also able to replicate prior findings from fMRI indicating that activity in "task-positive" and "task-negative" regions is negatively correlated during task performance. Finally, data from both companion and simultaneous fMRI experimental trials verified our assumptions about the sources of brain activity in the fNIRS experiment, established a upper bound on classification accuracy expectations for response to the MSIT, and served to validate our fNIRS classification results. Together, our findings suggest that fNIRS could prove quite useful for monitoring cognitive state in real-world settings.

Chapter 1
Introduction

1.1. Problem statement

Monitoring brain activity may provide an effective means for monitoring cognitive state, and in particular for distinguishing between high and low levels of task engagement. Numerous functional magnetic resonance imaging (fMRI) studies have revealed that activity increases in a so-called "task-positive" network (TPN), which includes the dorsolateral prefrontal cortex (DLPFC), dorsal anterior cingulate cortex (dACC), superior and inferior parietal lobe (SPL and IPL), and anterior insula (AI), when participants are engaged in an attention-demanding cognitive task compared to when they rest (Dosenbach et al., 2006; MacDonald et al., 2000). In contrast, activity increases in a so-called "task-negative" network (TNN) or "default mode" network (DMN), which includes the anterior medial frontal gyrus (aMFG), posterior cingulate cortex (PCC), and certain regions of lateral parietal cortex (LPC), when participants rest as compared to perform a task (Greicius et al., 2003; Raichle et al., 2001). Further, activity in the TPN and the TNN can be negatively correlated (Fox et al., 2005; Kelly et al., 2008). For this reason, monitoring activity in key regions of the TPN alone or monitoring activity in key regions of the TPN and TNN together may distinguish between relatively high and relatively low levels of task engagement (Chee et al., 2008; Drummond et al., 2005; Weissman et al., 2006).

Since monitoring hemodynamic activity with fMRI cannot be accomplished outside of a laboratory, we employed functional Near Infrared Spectroscopy (fNIRS) to determine whether monitoring brain activity is an effective method for monitoring cognitive state. fNIRS is a portable optical neuroimaging technique that can be used to quantify hemodynamic activations. Moreover, it is relatively low-cost, non-confining, non-invasive, and safe for long-term monitoring (Gibson et al., 2005; Gratton et al., 2005; Boas et al., 2004; Schroeter et al., 2006; Steinbrink et al., 2005). Finally, temporal resolution is sub-second, and spatial resolution is on the order of 1 cm^2 at best (Bunce et al., 2006; Obrig and Villringer, 2003; Strangman et al.,

2002). Measurements have been shown to be consistent with fMRI (Schroeter et al., 2006; Steinbrink et al., 2005; Emir et al., 2008; Huppert et al., 2006; Kleinschmidt et al., 1996; Strangman et al., 2002) and electroencephalography (EEG) (Li et al., 2007; Moosman et al., 2003). Critically, the ambulatory nature of fNIRS, while still an emerging technique in such application, allows neuroimaging in the field.

1.2. Motivation

The ability to distinguish between high and low levels of task engagement is important for detecting and preventing performance decrements during safety-critical operational tasks in the real world. Examples of such tasks include commercial aviation, monitoring for air traffic control, executing space walks, performing surgery, and driving. Since accident-causing errors can be made even by skilled professionals (Dismukes et al., 2007), the ability to monitor cognitive state measures for low levels of task engagement in real time could be useful for developing an "early warning system" for detecting and preventing performance errors before they occur. Intelligent flight decks of the future can be responsive to state changes to optimally support human performance. Thus, the identification of cognitive performance decrement, such as lapses in operator attention, may be used to predict and avoid error-prone states. The ultimate goal is to reduce errors due to lapses in attention to increase safety.

Although commercial aviation accident rates are currently quite low, their reduction is of great importance in the face of increasing air traffic. Even the most expert and conscientious pilots are susceptible to making errors in stressful or monotonous situations. Intelligent cockpits of the future will interact with operators in ways designed to reduce error-prone states and mitigate dangerous situations at the edges of human performance. One important aspect of achieving such a system is the development of reliable, sensitive and operationally-relevant metrics for the state of the "human in the loop" by non-invasive, portable, safe and inexpensive means. This work broadly aims to reduce the effects of attentional performance decrement and thus improve safety by informing the operator of her/his own state for self-management of attention. In the future, "next generation" intelligent systems are envisioned which can trigger appropriate interjection of mitigations, for example via task automation software, depending on the state of the operator. In this way, proper engagement of the operators with safety-critical tasks may be encouraged.

Another important aspect of such a system is monitoring those metrics "in-task" in a passive manner. Compensating functional activity patterns could indicate effortful performance in the face of impairment (Hugdahl et al., 2001) or increased "top down" attentional effort (Sarter et al., 2006; Sharp et al., 2006). This points to the need for ongoing performance measures beyond stimuli/response tests (such as reaction time testing) or short duration one-time "fit-to-fly" testing, as compensation and extra effort may be employed in the short term but may not be maintainable throughout tasks of long duration.

To achieve these goals, we turn to monitoring brain activity responsible for task execution. A review of fNIRS studies on executive function describes monitoring complex goal-directed behavior as effected by the dorsolateral prefrontal cortex and other frontal areas (Joanette et al., 2008). As one of three components of attention, along with orienting and alerting, executive function is employed for conflict resolution and to handle novel situations (Fan et al., 2005). Workload management and strategic planning also use such high-level executive processes (Oken et al., 2006; Gilbert and Burgess, 2008). Activities performed by rote (such as routine monitoring of flight automation) require little vigilant attention and foster disengagement. Activations in right prefrontal and attentional network (ATN) areas are important to maintaining endogenous support for vigilance (Posner et al., 2004). Conversely, medial frontal and DMN regions deactivate with goal-oriented task engagement (Fox et al., 2005; Weissman et al., 2006). The properties of these networks motivate our research questions, including their accessibility to fNIRS because key regions of the networks reside in the cortical layers of the brain.

1.3. Research questions

We have explored approaches for improving the validity and reliability of using the emerging technique of fNIRS to monitor brain activity and improve the predictive power of fNIRS-based cognitive state classifiers. These approaches include the measurement of functionally-connected networks and the exploration of various fNIRS processing methods. In the present study, we employed fNIRS and fMRI to determine whether it is possible to distinguish between high and low levels of task engagement (i.e., performing a task versus resting). Specifically, we monitored cortical brain activity from the DLPFC (which is also part of ATN) in the TPN and from the MFG (which is also part of the DMN) in the TNN while

participants alternated between performing and not performing a cognitive task. We then employed multivariate pattern classification techniques in an effort to distinguish between periods of task performance and periods of rest using TPN activity alone, and using TPN and TNN activity together. Given recent data indicating that interactions between key regions of the TPN and TNN vary with task engagement (Prado and Weissman, 2011), we predicted that either approach for monitoring brain activity would allow us to distinguish between relatively high and relatively low levels of task engagement, but that the latter approach would likely prove most effective. Analogous classification results using fMRI signals were used to validate the fNIRS findings, and statistical parametric mapping was used to verify TPN and TNN activity. Additionally, to establish the processing methods which would produce the highest state prediction accuracy, we investigated multiple pre-processing methods for removing noise and for producing multiple versions of the fNIRS traces used as classifier input features.

To further investigate the usefulness of monitoring TNN regions in addition to TPN regions, we calculated time-windowed across-network correlation. We reasoned that because TPN and TNN activity has been shown to be negatively correlated, detecting network anti-correlation could be used as classifier input features or to verify correct fNIRS probe placement in practice. Finally, we quantified very low frequency activity in the signals to determine whether such activity may be due to intrinsic resting state and TNN activity (Fox et al., 2007; Biswal et al., 1995) or to physiological confound (Sassaroli et al., 2012).

1.4. Context

The results of fNIRS studies thus far involving higher executive function have been varied, although promising (Arenth et al., 2007). fNIRS measurements also are relevant to improving the efficiency of skill acquisition and expertise development (Hatakenaka et al., 2007; Leff et al., 2008). Distinguishing between workload levels has been accomplished with fNIRS (Izzetoglu et al., 2004). Features sensitive to which cognitive resources are taxed, extracted from fNIRS measurements made on the forehead, have shown promise for applicability to usability testing of various user-interface designs (Hirshfield et al., 2011). Cognitive performance measurement is highly relevant to space flight as well, since adaption to microgravity can cause performance decrement due to motion sickness, lack of sleep, loss of sensorimotor control, increased stress or mood changes (Cowings et al., 2003). Indeed, an fNIRS system for affective

state assessment is under investigation (Callaway et al., 2008). Looking forward, the detection of particular brain activations could be used to assess the efficacy of countermeasures and predict cognitive performance under stress, such as vigilance decrement due to fatigue (Drummond et al., 2005a), stress (Lieberman et al., 2005), or divided attention (Nebel et al., 2005; Vohn et al., 2007).

Despite the challenges as discussed in sections 2.1.2 and 2.1.3, fNIRS provides rich potential to supplement behavioral data with objective neurological measures in simulated and real-world contexts. For example, operator attention could be monitored *during* the performance of tasks. Further, with better motion tolerance than fMRI, and with no claustrophobic and noisy environment, fNIRS enables such monitoring in ecologically valid environments. For example, gait and centrifuge studies can be performed where fMRI studies would be impossible. The portability of fNIRS enables its use for treatment in a physician's office, rehabilitation facility or eventually even in a take-home device. fNIRS has been applied to a train driving simulation task comparing manual and automated modes (Kojima et al., 2005), and to the development of a robot which adapts to the multitasking state of a human operator (Solovey et al., 2011). When combined with near-real-time signal analysis and classification processing, this allows for potential use in brain-computer interface (BCI) applications (Coyle et al., 2007; Matthews et al., 2008) for long-duration monitoring and biofeedback training.

The use of cognitive state measures to optimize human performance (for example by informing flight automation or the operator themselves of a hazardous state) has been of particular importance to aviation safety (Schnell et al., 2004; Pope et al., 1995) and to the Augmented Cognition program (Snow et al., 2006) of the Defense Advanced Research Projects Agency of the United States. More generally, the ability to monitor cognitive state for low levels of task engagement could aid the detection and prevention of vigilance decrements due to sleep-deprivation (DeHavas et al., 2012; Drummond et al., 2005) or distraction (Strayer, et al. 2011).

The results presented here build on previous work showing that frontal oxygenation as measured with fNIRS is sensitive to workload (Izzetoglu et al., 2004) and that fNIRS can reliably detect both resting state physiology (Mehnert et al., 2009; Mesquita et al., 2010) and functionally-connected networks (White and Culver, 2008).

1.5. Importance

The present findings make important contributions to the field of fNIRS. Specifically, they show, for the first time, that it is possible to detect negative correlations between activity in key regions of the TPN and TNN with fNIRS. Further, they show that the detection of activity in the TNN is useful for distinguishing between high and low levels of task engagement. This capability might prove useful in future applications of fNIRS that are aimed at discriminating between optimal behavioral performance (where a negative correlation is expected) and internally-guided thought (where co-activation and, hence, a positive correlation is expected) (Christoff et al., 2009; Smallwood et al., 2012). Thus, it could function to improve the predictive power of a fNIRS-based attentional state monitoring system.

We have performed an extensive, fMRI-validated exploration of attentional state prediction using various fNIRS processing methods and features. We have also extended existing noise removal methods and identified limitations to the attentional state activation model used.

1.6. Summary of methods, results, conclusions

Human participant trials using both fNIRS and functional Magnetic Resonance Imaging (fMRI) were used to monitor brain activity to determine whether it is possible to distinguish between cognitive states associated with high vs. low task engagement. To induce attentive and inattentive states, participants alternately engaged in a psychological task and rested for multiple individual runs of the same task. The Multi-Source Interference Task (MSIT) was used with nine fMRI and twelve fNIRS participants total across a two-part study. Both fNIRS and fMRI were used to quantify hemodynamic activations in the brain regions of interest. Multivariate pattern classification was implemented with Support Vector Machines (SVM). State classification was separately based on processed fNIRS and fMRI signals. Models for classifier state prediction purposes were built using a subset of that participant's runs as the training data. The rest of the runs were used separately for prediction testing purposes. Each time instance is used to either inform the SVM model or to produce a predicted state output.

In part one of the study, the detection of anti-correlated attentional networks in the brain was demonstrated for the first time with fNIRS. Also, a trend was found implying that using TPN and TNN activity together can improve classification accuracy over using TPN activity alone ($t(4) = 1.48$, $p < 0.25$). We also conducted a companion fMRI study for verification and

comparison purposes. First, given that DLPFC and MFG activity was recorded with fNIRS at the scalp surface, we wished to verify that our paradigm actually elicited hemodynamic activations in these regions. fNIRS probe placement on the head for part two was informed by these results. Second, we wished to compare fNIRS with fMRI with regard to the ability to distinguish between high and low levels of task engagement. Given that fMRI detects motor cortex activation reliably enough for routine use (Möller et al., 2005), and that the motor cortex should be activated by the button presses required by our task, we expected that including such activation as a classifier input feature would produce the highest levels of accuracy. Thus, we reasoned that classification accuracy with fMRI when including motor cortex activation would establish an "upper bound" for classification accuracy expectations. As expected, the fMRI study confirmed both of these predictions. Attentional state prediction accuracy with fNIRS was found to be comparable to that found with fMRI for the same task, when fMRI classifier input features were restricted to those from spatial volumes similar to those of fNIRS.

In the second part of the study, the MSIT again was used, but during fMRI scanning while simultaneously recording brain activity in the same two regions with fMRI-compatible fNIRS head probes. Six classifier input features were selected from the arrays of probe channels at the two locations. The features were selected based on how well the signals discriminated task activity for a subset of the data which was used to train the classifier. The balance of the data was used for prediction testing purposes and not for this selection. Classification was performed on signals which were pre-processed using static and adaptive regression. When knowledge of the task being performed was used, dynamic filtering of noise via adaptive regression was found to produce superior results, representing the current best capability of monitoring attention with the fNIRS techniques presented here. Within-participant classification accuracy of task engagement vs. rest was found to be 84% +/- 6% with 20 min of training (81% +/- 5% with 7 min of training) when knowledge of the task being performed is used in signal processing with dynamic regression. Accuracy was 70% +/- 11% (68% +/-12%) when no knowledge of the task is used. This is a more realistic capability which is applicable in situations where, for example, the relative engagement level of a pilot is not known *a priori*. This result is in line with prior results. fMRI was used to verify regional task activation and validate classifier accuracy. Again, and this time for simultaneously-collected time traces of hemodynamic activation, we found comparable performance between fNIRS and fMRI when classifier input features for fMRI were restricted to

spatial volumes which were estimated to overlap with the sensitivity of the fNIRS probes. We also found classification based on whole-brain fMRI data to be the best produced overall, consistent with its use as the gold standard for this study.

Thus, classification accuracies were found to be above chance levels with and without the use of *a priori* knowledge of the task being performed. Further, these results were obtained with classifier training times that reasonably could be implemented in operational settings. However, the best processing methods for classifier feature generation depend on the situation. For example, without *a priori* knowledge of the task being performed, the least pre-processed of the fNIRS signals was found to provide the best performance. With *a priori* knowledge of the task being performed, and thus a strong model for the expected brain activity in each brain region being interrogated, noise is more effectively removed and classifier performance is superior. Such knowledge would be available during controlled laboratory experiments and operational simulations, and possibly in intelligent, real operational environments of the future. We conclude that fNIRS could prove quite useful for monitoring cognitive state in real-world settings.

1.7. Outline of the dissertation

In Chapter 1 we have introduced the motivation and research questions for this work. In Chapter 2 we review the fNIRS technique and discuss aspects of an fNIRS system which are important to cognitive state sensing, focusing specifically on attentional state in real world contexts. Throughout this work, we define attentional state as the relative level of engagement of the participant in an external, goal-oriented task. Additionally, methods used to obtain results from both parts of the study are introduced in Chapter 2. In Chapter 3 we focus on part one of the study, where we demonstrate the detection of relevant functionally-connected brain networks with fNIRS, and present attentional state classification accuracy results when activity in the TNN is included, with both fMRI and fNIRS. In Chapter 4, we focus on part two, where we present classification accuracy results for multiple pre-processing methods of the measured fNIRS signals. Also, we present classification results based on simultaneously-recorded fMRI Blood Oxygen Level Dependent (BOLD) signals from the whole brain, from regions that are significantly activated by the MSIT, and from regions that correspond to estimates of the volumes to which the fNIRS probes are sensitive. Finally in Chapter 5 we conclude with a summary of results and implications for further work in cognitive state monitoring.

Chapter 2
Applying Functional Near Infrared Spectroscopy to Attentional State Monitoring

2.1 The fNIRS Technique

The application of the emerging neuroimaging technique of fNIRS to operator state characterization is promising but challenging. In this chapter we discuss how fNIRS works and present specific challenges to be overcome for use in cognitive state monitoring. fNIRS methods are presented, including approaches to the challenges. Additionally, TPN and TNN activity is covered in greater detail. Finally, behavioral, classification and fMRI methods common to both parts of the study are described.

More generally, the application of fNIRS to functional neuroimaging is heavily researched, historically and currently, by such groups as those led by Britton Chance (University of Pennsylvania), David Boas (Massachusetts General Hospital), Theodore Huppert (University of Pittsburgh), Scott Bunce (Drexel University), and Enrique Gratton (University of Illinois at Urbana-Champaign). With further research, including the present study, fNIRS may evolve into a synergistic and valuable complement to EEG and other psychophysiological measures for monitoring and predicting cognitive states.

2.1.1 Introduction to the technique

fNIRS is used to quantify hemoglobin concentration changes in cortex. It is portable, relatively low-cost, non-confining, non-invasive, and safe for long-term monitoring (Bunce et al., 2006; Gibson et al., 2005; Gratton et al., 2005; Boas et al., 2004; Schroeter et al., 2006; Steinbrink et al., 2005). Measurements have been shown to be consistent with fMRI (Schroeter et al., 2006; Steinbrink et al., 2005; Emir et al., 2008; Strangman et al., 2002a) and EEG measurements (Li et al., 2007; Moosman et al., 2003). Optical absorption increases as hemoglobin concentrations increase, which in turn depends on local neural activity (Boas et al.,

2004; Bunce et al., 2006). This is analogous to the generation of BOLD signals in fMRI via changes in deoxygenated hemoglobin. Both techniques quantify neural activity indirectly via neurovascular coupling. Sustained hemodynamic responses as measured by fNIRS during a finger tapping task can be seen in Figure 14, top panel. Both fNIRS and fMRI sense the hemodynamic response (Huppert et al., 2006), as opposed to the Event Related Optical Signal (EROS) and Electroencephalogram (EEG) measurements, which sense the activity of neurons themselves (through cellular changes which affect optical scattering, or through electrical activity, respectively). EROS as a technique has limited operational application due to lower signal-to-noise and higher susceptibility to motion artifact. (Francescini and Boas, 2004; Bunce et al., 2006)

fMRI is a hemodynamic neuroimaging method capable of indirectly measuring neural activity (via neuro-vascular coupling) in the whole brain by quantifying changes in paramagnetic deoxygenated hemoglobin (HbR), with spatial resolution on the order of a few mm and temporal resolution on the order of seconds. In contrast, fNIRS interrogates the outer surface of the cortex only, and quantifies changes in oxygenated and deoxygenated hemoglobin concentration ([HbO] and [HbR], respectively). Temporal resolution is sub-second, while spatial resolution is on the order of $1cm^2$ at best (Bunce et al., 2006; Obrig and Villringer, 2003; Strangman et al., 2002b). The intensity of infrared light that has diffused through the capillary bed is detected after being differentially absorbed by both species of hemoglobin. For example, light of wavelength 690nm is primarily absorbed by HbR, while light of 830nm is absorbed more by HbO (Tuchin, 2007, fig. 6.3). Absorption in this wavelength range by water and other chromophores is sufficiently low to permit the light to penetrate all the way to cortical tissue through the scalp, skull, and cerebral spinal fluid.

To make functional activity measurements from the surface of the head with continuous wave instruments, optical delivery and detection is required at 2 to 4 cm apart. Thus, a *probe* is made up of a *source* and a *detector*, which straddles its measurement location on the head surface. The measurement location is typically the point through which a line normal to the head passes on its way to the deep brain region of interest. Measured optical intensity changes are due to changes in absorption of the light as it migrates from source to detector. This change in absorption contains information about the change in hemoglobin concentration ([Hb]) depending on the length of the optical path through the tissue undergoing hemodynamic changes (please see

Appendix A, Boas et al., 2001 and Tuchin, 2007, p. 454 for mathematical details). Oxygenated and deoxygenated hemoglobin are the predominant absorbers in biological tissue in the red and near-infrared wavelength region used for fNIRS. Relative changes in the concentrations of both these species of hemoglobin over time are calculated using the Modified Beer Lambert Law (MBLL) (Delpy et al., 1998) from intensity measurements for each of two wavelengths per location. The modification refers to two changes to the well-known Beer Lambert Law, which describes the dependence of light attenuation on absorption, path length and chromophore concentration (Tuchin, 2007, eq. 6.6). One is the subtraction of a background scattering term, which is a significantly greater source of light attenuation than that due to absorption, but is assumed to be constant. The other is the inclusion of a differential path length factor (DPF) (which is discussed further in section 2.1.2) to account for the increase in path length due to multiple scattering effects as the photons migrate through tissue (Tuchin, 2007, p. 454). The DPF depends on wavelength, is typically 5 or 6 for healthy adults, increases with age and increased scattering, and is independent of sex (Duncan et al., 1996). It was fixed at 5.5 for 690nm and at 6 for 830nm for all participants in this study. The resulting [Hb] time traces make up the digital fNIRS signals which are processed for noise removal then used as inputs to classification algorithms.

Activations are quantified for a region of interest by comparing hemoglobin concentration changes between species. A hemodynamic response is indicated by an increase in [HbO] or a decrease in [HbR], over time periods on the order of 10-12 seconds (or longer if sustained) as regional blood flow increases to overcompensate for local metabolism (Buxton, 2002, p. 419). Regional or time averages are taken of the activation quantity, and comparisons are made between measurements obtained during different conditions such as work and rest. Care must be taken when directly comparing activation magnitudes since sensitivity (discussed below in section 2.1.2) is not uniform across probe locations (including across participant and across wavelengths) and depends on the make-up of the capillary bed probed by the photon migration path.

Further signal processing challenges are discussed in section 2.1.3. First, however, we address the sensitivity of the probes depending on their location.

2.1.2 Sensor Location and Sensitivity Challenges

Overlap of the *sensitivity profile* (the shape of the sensed volume of tissue through which the sourced photons are scattered and absorbed on their way to the detection location) with the active vasculature (the capillaries (Liu et al., 1993)) of interest is the goal of sensor localization. In other words, the sensitivity of the measurement technique relies on the degree to which the tissue sensed is that which is affected by hemodynamic changes as evoked by experimental stimuli (Boas et al., 2001; Schweiger et al., 2007; Strangman et al., 2003). Systemic physiology which affects the hemodynamic changes within vasculature sensed anywhere between the source and the detector (that not due to activation in response to the function of interest) confound the measured signals.

To aid understanding of the many factors affecting the sensitivity of the optical probes, we introduce some important terms here. The distance the sourced photons travel overall through the tissue is termed the *optical path length* (depicted in Figure 1 in pink). This distance is greater than the chord distance between the source and detector by the multiplicative *differential path length factor* (DPF). Brain activations do not cause changes in all of the tissue through which the light scatters – only a portion of it. The distance the light travels through the parts that do change with activation is the *partial path length* (depicted in Figure 1 in gray). For more sophisticated modeling of the photon migration path, the reader is referred to Okada and Delpy, 2003. Sensitivity profiles are further discussed in section 4.5.2.2. The light from each probe location encounters different superficial tissue, and the light of each wavelength takes a different path (and therefore has a different differential path length) through that tissue, even if injected at the exact same location. The partial path length is not known. Therefore, because the Modified Beer Lambert Law calculation of hemoglobin concentration (introduced in section 2.1.1 and given in Appendix A) assumes the absorption has occurred along the entire optical path length, the absorption by hemoglobin and thus the calculated change in [Hb] is underestimated. In reality, greater absorption has occurred over the shorter partial path length (Boas et al., 2004; Huppert et al., 2009), contributing to inherent error in the measurements.

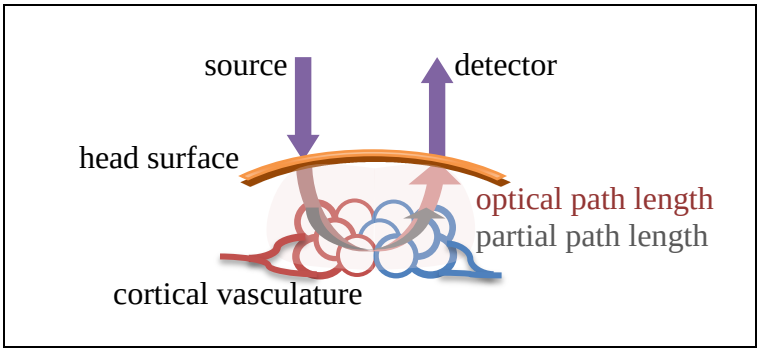


Figure 1. Schematic of the optical and partial path lengths relevant to fNIRS

The spatial extent of the sensitivity profile determines the spatial resolution. The attempt is made at each probe location to maximize the overlap of the sensed volume with the volume of tissue undergoing the hemodynamic activations of interest. This is typically done without detailed information on the underlying vasculature and anatomy. The reduction of the source-detector distance does not necessarily improve spatial resolution, as many factors affect the shape of the sensed volume (which is the combination of all the smaller tissue volumes through which the photons pass on their way to the detector) relative to the location of the volume of interest. Errors are introduced due to vascular and superficial tissue non-uniformity, both across the head and between individuals. With few probes, the extent, position and intensity of measured hemoglobin changes strongly depend on probe arrangement (Kawaguchi et al., 2007). However, if a high spatial density of regularly-spaced sources and detectors is used with image reconstruction methods, the measured activation can be a more accurate reproduction of the actual activation in both location and extent, especially for localized activations (Boas et al., 2001; Kawaguchi et al., 2007). Thus, with a greater number of detectors than currently available for this work, sensitivity and accuracy can be improved by using overlapping and neighboring probes placed symmetrically about the active region of interest. It is thus advantageous to develop probes with reduced spatial footprint so more sources and detectors can fit as unobtrusively as possible on the head.

Varying source-detector separation gives rise to depth resolution (Hillman et al., 2007), because greater source-detector separation probes deeper tissue. However, maximum depth is limited to the outer layer of cortex. Further, signal intensity is lowered (signal to noise is reduced) with longer path lengths, and the sensitivity profile itself changes depending on the non-uniform superficial tissue through which the photons are scattered (Boas et al., 2004; Okada

and Delpy, 2003). For example, as the thickness of the cerebral spinal fluid layer increases, the lateral extent of the sensitivity profile increases. Greater skull thickness contributes to less fNIRS sensitivity and smaller partial path length. Interestingly, Monte Carlo simulations show an inverse effect for thin layers of cerebral spinal fluid. (Okada and Delpy, 2003) Unfortunately, individual measurements of these important superficial tissue thicknesses are not readily made *in vivo*. It is possible to measure the overall optical path length with frequency domain fNIRS instruments (Arridge et al., 1992), but this measurement cannot reveal the individual layer thicknesses (Okada and Delpy, 2003; Hoshi et al., 2005). Thus, some adjustments to source-detector separation can be made to account for the depth of the brain region of interest, but it is not cost-effective to do so with precision for multiple users due to the necessary anatomical MRI or vasculature mapping efforts required for each participant.

Functional and structural MRI also may be used to locate volumetric cortical regions of interest which can be used to inform and verify (as in the present study) or customize fNIRS optode placement on the surface of the head. Functional MRI scans can measure the BOLD signal in response to a localizer task which is known to activate the region of interest. Preliminary fNIRS experiments may be used for the same purpose, to identify the probe from which the processed signal most fits the expected response for the known localizer task. Concurrent structural scans theoretically can improve estimates of the appropriate source-detector separation length (to sense at the proper depth) based on superficial tissue measurements over the particular volume of interest. This would ideally be done on a per-optode basis for each participant or user. Efforts to estimate the spatial correspondence of fMRI voxels and fNIRS sensitivity profiles are described in section 4.5.2. The cost-benefit ratio of such efforts is likely prohibitively high at this time, considering the positive results of this study while using one source-detector separation distance (3cm).

The change in oxygenated versus deoxygenated hemoglobin concentrations is calculated based on the measured optical intensity, depending greatly on the optical path length, which in turn depends on the wavelength, the individual's age (Duncan et al., 1996), and the tissue through which the photons are scattered. However, only the partial path length, along which the light has the opportunity to be absorbed by the hemoglobin-containing vasculature of interest, contributes to the hemodynamic information in the signal. An increase in the partial path length increases sensitivity and therefore the signal to noise ratio. This increase is limited if the active

volume is focal (very small). At the other extreme, probes placed over vasculature of diameter greater than about 1 mm will see too much absorption to be useful (Cannestra et al., 2001; Liu et al., 1993).

Errors due to hemoglobin species cross-talk are also affected by localization. Cross talk is due to absorption by one species being attributed to the other because the hemoglobin concentration calculation for each species depends on the optical intensity measured at both wavelengths. The differential path length is directly used in calculating [Hb] using the MBLL. The partial path length is unknown. Since light of each wavelength takes a different partial path, ideally the appropriate calculation for each species would use partial path lengths specific to each wavelength. Since these are unknown, errors are unavoidable (Boas et al., 2001). These errors can be minimized with (1) proper wavelength selection and (2) the maximization of the partial path length such that as much of the path length as possible is taken up by tissue undergoing the activation of interest. Partial path length can be maximized by placing the probes within ~1cm of the active region of interest (Strangman et al., 2003).

2.1.2.1 Practical considerations for placing optodes

To best approximate this without fMRI or 3D tracking data, optodes have been placed according to the International 10-20 (I 10-20) system used for electroencephalography (as in this study) (Jasper, et al, 1958; Jurcak et al., 2007; Okamoto et al., 2004; Sato et al., 2006), or by trial and error (Snow et al., 2006). Maximizing the detected intensity is one way to attempt to avoid large vascular obstacles, but this decreases the likelihood of accurate measurement of the magnitude of hemoglobin concentration changes (Δ[Hb]) (Boas et al., 2001). Maximum intensity at a reasonable detector gain setting could be due to a lack of tissue absorption. When properly positioned such that the tissue volume sensed by the optode overlaps with hemodynamic absorbers, optical intensity is reduced during activation. During activation, optical intensity increases as [HbR] decreases, offset by intensity decreases as [HbO] increases. In other words, if one moves a probe about the surface of the head searching for a local spot where optical intensity is largest, one may find a spot which likely contains few capillaries of interest. In the other extreme, one may find a vessel that is too large, providing too much absorption and not enough signal intensity. What should be maximized instead is the modulation between baseline (intensity higher, total [Hb] lower) and stimulated activation (intensity lower, total [Hb] higher). Optical

intensity at both these extremes should be within the dynamic range of the optical detectors. This is not readily determined without a controlled experiment or localizer task. Also, a large optical signal could be due to static or dynamic ambient light exposure if direct optode-to-skin contact is not maintained. Thus, the maintenance of baffling, detector sensitivity settings and sensor placement throughout the experimental trial are very important.

Manual probe reapplication can result in about 0.5 cm deviations on the head (Sato et al., 2006), causing slightly different parts of the cortex to be interrogated. Custom headgear fitting could begin to improve repeatability of placement between sessions. Headgear that has been custom-fabricated to fit a particular individual the same and proper way for each application not only would support consistent signal detection, but comfort and long term wear-ability as well.

The overlap of the sensitivity profile between each source-detector pair (or *optode*) with the active vasculature of interest is most difficult to achieve for focal active regions. This is an important challenge for single-optode-per-location layouts where multiple separate measurements are made as opposed to a regular array of sources and detectors appropriate for image reconstruction. For the purposes of state sensing, we employed multiple separate measurements. Imaging or spatial mapping of the activations makes use of multiple-optode-per-location layouts and image reconstruction algorithms to improve spatial resolution (Schweiger et al., 2007; Kawaguchi and Okada, 2007). However, the size of the sensed region (spatial resolution) has been quantified as 1cm across at absolute smallest (best), and so only for the most shallow measurements (Strangman et al., 2003). The use and improvement of image reconstruction techniques to optimize spatial resolution is interesting and important to the field. However, it was not pursued for this study. Fewer optodes, if well-placed surrounding the locations of interest, allow sufficient detection, but not precise mapping, of relevant activations. Adaptive filtering results to be presented in Chapter 4 support the idea that information extraction is more likely to be limited by the non-stationary nature of physiological noise (Kolehmainen et al., 2003; Diamond et al., 2006; Zhang et al., 2007b; Abdelnour and Huppert, 2009) than to be limited by the selected localization scheme. Further, systems with fewer optodes are advantageous for use in operational contexts.

However, optodes spent sensing multiple locations of interest can improve robustness and reliability. Thus there is an important tradeoff to consider in the design of probe arrays for operational use. Due to weight, space, and cost constraints, there may be a limited number of

optodes available, and thus a limited number of locations at which functional activations can be measured. Each location requires a separate probe or array of probes on the head, and dedicated channels for data acquisition. Instrumentation size increases with the number of channels. Also, wear-ability and participant or user compliance may decrease as probes are added to the head. There is some minimum number of probe locations with which one may obtain sufficient information for such comparisons and classification accuracy. In this study, we have employed a total of 6 optodes across two locations, and achieved positive results with both 6 and 2 classifier input features.

Further, comparisons of patterns of relative activations over multiple regions, and between different conditions over time, allow the measurement of network activations and identification of local versus global hemodynamic changes (such as systemic physiology) and global noise (such as motion artefact). The measurement of network activations can improve the specificity of brain state assessment because the attribution of regional activations to particular functional tasks may be inappropriate. Due to a lack of functional specificity, activation in one region may not be specific for the stimuli of interest. One region may be activated (in varying combination with other regions) by many different kinds of stimuli (Gazzaniga, 1989; Kanwisher, 2010).

2.1.2.2 The importance of consistent and stable optode locations

Once selected, obtaining consistent signal at all locations across time and thus across stimulus conditions is important because the quantification of neuronal activation with fNIRS relies on relative measurements. Consistent signal measurement also is important because the loss of signal at one optode could reduce the accuracy of the state classifier by removing an important predictor variable from the classification algorithm inputs. This is heavily influenced by the stability of the head probes. Many errors are introduced at the optical-tissue interface (Schweiger et al., 2007) and signal quality is highly susceptible to mechanical instability. Existing methods for fNIRS probe attachment in adults can be time-consuming and difficult to employ, are susceptible to motion artefact, and can be quite obtrusive and uncomfortable (Harrivel et al., 2009). Indeed, many recent fNIRS studies have been restricted to frontal probe placement (Hirshfield, et al. 2011; Callaway, 2008; Solovey et al., 2011), thus avoiding absorptive and mechanical offset interference from the hair. Currently, many fNIRS systems

employ optical sensors which project for some length from the head, and require a skilled attendant and much time to position and separate the hair beneath each probe in turn. While existing probes are successfully used under the controlled conditions of research laboratories, they are often inappropriate for applications which do not enjoy the benefit of time and attending personnel to place and adjust optodes.

To account for this, signal quality metrics that can, for example, detect high frequency noise artefacts may be used for automated removal of channels on which signal has been lost (Brigadoi et al., 2013; Cooper et al., 2012). If data from a candidate optode location is removed and no significant decrease in classifier accuracy is observed, then the sensor placement design may not need to include this location. Likewise if a location is added and no significant improvement in classification accuracy is obtained. Due to the unpredictable nature of operational use of non-invasive techniques, it may be prudent to retain locations wherever possible and consider redundant measurement locations. Bilateral measurement locations in sensor placement designs provide one opportunity for such redundancy.

In addition to robustly stable attachments for reliable signals, probes for practical application in the field ideally should baffle the detectors such that ambient and cross-channel source light is blocked, have a low profile (height from the surface of the head) to not interfere with helmets or headsets, be comfortable for repeated and extended use, and allow repeated self-application and positioning of light delivery and detection in a quick and easy (yet reliable, stable and repeatable) fashion. Headgear which closely follows the curvature of the head, ideally custom-fit to the individual, will produce more reliable signals due to more consistent optical coupling. Comfortable headgear eliminates significant subject non-compliance and the introduction of distractions or other psychological confounds due to pain and discomfort. Designs have been proposed to improve comfort with transparent elastomer and improve self-applicability with hair-parting comb shapes (Harrivel et al., 2009; Harrivel, et al., 2012b).

The ability to exploit functional connectivity of known networks in the brain, as will further be discussed in section 2.2 below, further motivates the need to collect good signal from a strategic number of locations simultaneously by fitting probes anywhere across the whole head. First, we turn to a discussion of signal processing once probes have been placed and signals have been (or as signals are being) collected.

2.1.3 Signal Processing Challenges

fNIRS is advantageous as a neuroimaging technology in part because the hemodynamic response it measures also underlies the fMRI BOLD response (Huppert et al., 2006). Localized cortical hemodynamic activations that have been shown to be evoked by known functional tasks using well-established fMRI techniques also should be detectable with fNIRS. However, established standards for fNIRS practice are just beginning to emerge (Abdelnour et al., 2009; Huppert et al., 2009; Lloyd-Fox et al., 2010; Matthews et al., 2008; Orihuela-Espina et al., 2010). These standards have yet to be widely adopted by the full fNIRS community.

The quantification of changes over time in hemoglobin concentrations stemming from neural activations in fNIRS signals is confounded by signal contributions from many sources other than hemodynamic activations in the brain. A signal component which covaries across all or a subset of channels (locations) is likely due to systemic physiology (Franceschini et al., 2006), or due to motion (such as that due to impact) if the optical probes are rigidly connected. Variance can masquerade as changes in [Hb] due to an ill-fitting probe or exposure to another source or to ambient light), or can be due to actual changes in [Hb] driven by physiology that is not of interest.

In this study, signals are temporally filtered to include the hemodynamic response of interest and low frequency functional fluctuations (Biswal, 1995; Mantini et al., 2007), and to reduce signal contributions due to motion (high frequency, of the probe with respect to the head), cardiac pulsation (at about 1 Hz) and respiration (at about 0.25 Hz). Since inspiration promotes venous return to the heart from peripheral veins (Vander, 2001, p. 424), total [Hb] and light absorption both decrease with inspiration. Oxygenated and deoxygenated [Hb] both increase with expiration, differentiating such signal components from the hemodynamic response where [HbR] would decrease. These physiological effects can be used to indicate proper probe coupling to the tissue, and may be of interest to other studies. However, it is appropriate and necessary to remove these confounds when activations in the brain are of interest (Tachtsidis et al., 2009). Temporal filtering is the most straightforward method for dealing with these higher frequency artefacts. Literature regarding time series analysis for fNIRS signals shows examples of such artefacts (Franceschini et al., 2006; Huppet et al., 2009, Fig. 3).

Most physiological contributions and artifacts in the signals can be identified by their frequency, but simple temporal filtering can degrade the signal of interest when the frequency

band of the confounding noise overlaps that of the hemodynamic response. Significant physiological noise sources do occur at or near the hemodynamic response function frequency (especially Mayer waves at 0.1 Hz and slower blood pressure regulation (Julien et al., 2006; Matthews et al., 2008; Sassaroli et al., 2012)). These cannot be filtered out without potentially reducing the signal of interest (Matthews et al., 2008; Abdelnour and Huppert, 2009). Blood pressure was not monitored in this study. Principal component analysis (PCA) is commonly used to filter out physiological signals, but the physiology cannot always be separated consistently and cleanly into removable components (Franceschini et al., 2006; Huppet et al., 2009).

Static regression techniques do not always leave the activations of interest throughout the time trace intact (Murphy et al., 2009), and require the use of known stimulus onset times to model the expected hemodynamic response. *A priori* knowledge of the task being performed and thus the stimulus onset times is not likely to be available when psychophysiological monitoring is applied in operational contexts. However, to determine the best capability of fNIRS to monitor the TPN and TNN, knowledge of stimulus onset times is used in part one and in some cases in part two of the study. As discussed in Chapter 4, in part two we explore both: 1) dynamic regression techniques and 2) the determination of more realistic capability by not using *a priori* knowledge of the stimulus onset times.

Thus, signals simultaneously recorded from the superficial tissue are used to clean lower-frequency physiological signals (and other artifact such as gross probe motion and ambient light exposure) from the deeper measurements to isolate the functional activations of interest. The superficial tissue is sampled using an optode with a source-detector separation distance of 1 cm (producing a *shallow* or physiological signal), while the deeper measurements use a source-detector separation distance of 3 cm (producing a *deep* signal which samples the brain as well as superficial tissue). Both shallow and deep measurements are very sensitive to physiological effects from superficial tissue, but to differing degrees at differing times (Zhang et al., 2007b). Presumably only the deep measure includes brain signal, and the shallow signal is assumed to be dominated by noise. Therefore we assume better estimates of brain activity in the deep signal can be obtained by removing shallow contributions.

This method of cleaning the signals may avoid misinterpreting physiology (e.g. Mayer waves (Mesquita et al., 2010; Julien et al., 2006)) as functionally-connected activity. Accordingly, if a close source-detector pair (with shallow sensitivity) is used to measure

physiological signals in the superficial tissue particular to each location (Gagnon et al., 2011; Zhang et al., 2007a), the unwanted physiological contributions to the signals of interest can be removed by subtracting a fraction of that signal based on General Linear Model (GLM) fit parameters. In the static case, these are determined once for the full time series. This was done in both parts of the study, and will be described in more detail in sections 2.1.4.2, 3.2 and 4.3.4.

This removal method has been shown to be most appropriate when there is sufficient correlation between various time lags of the shallow signal and the deep signal (Zhang et al., 2009). However, the shallow signal can contain task-evoked effects due to the influence of the sympathetic nervous system on blood vessel diameter during cognitive responses (Kirilina et al., 2012). Therefore, to avoid removing the effect of interest, we used the shallow source traces (without time lags and without correlation assessment) as regressors in two standard linear regression steps. Task effects were removed from the shallow physiological traces before they were smoothed over 1 s and used in the second regression step as independent nuisance regressors. Any physiological effects in the brain (in the band-pass frequency range) which are not apparent in superficial tissue are not corrected.

Each source-detector pair produces time traces of changes in both hemoglobin species (Δ[HbO] and a Δ[HbR] traces). In part one of this study, this correction was made within-species only. That is, the shallow Δ[HbO] trace was used as a nuisance regressor for the deep Δ[HbO] trace, and the shallow Δ[HbR] trace was used as a nuisance regressor for the deep Δ[HbR] trace. This is further discussed in section 3.2. In part two of the study, we used both species to correct each trace, as discussed in section 4.2. Additionally, an adaptive GLM was applied time-point-by-time-point to dynamically filter out non-static physiological noise (Abdelnour and Huppert, 2009). This will be described in section 4.3.4.

2.1.4 fNIRS Acquisition and Processing Methods

2.1.4.1 fNIRS data acquisition

Hemoglobin concentration changes were measured using an Imagent NIRS instrument and fiber optic cables (ISS, Inc., Champaign, IL). Source light was transmitted from laser diodes in the instrument using paired 400 micron core multi-mode optical fibers with standard SMA-905 connectors. Each source fiber pair delivered 690 nm and 830 nm wavelength light. Detected light was collected at the skin surface and transmitted to the instrument detectors using 3 mm

diameter glass fiber optic bundles (one per region). Rigidly-connected source-detector pairs (comprising one *optode* or one channel) were used. The instrument employs time multiplexing to separate channels by turning on the source-pairs one at a time in a fixed sequence. This sets the data collection rate, which was 6.25 Hz throughout both parts of the study. During each task run, 2500 time points, or feature *instances*, were collected.

Multiple sources (eight in part one and five in part two) were located around one detector placed over the DLPFC region, while three sources were located around a second detector placed over the MFG. Optodes were placed with respect to the I 10-20 locations. The array of probes used to interrogate the MFG in both parts of the study, as shown in Figure 5, contained two sources placed 3 cm from one detector placed between FPz and FP2. The array of probes used to interrogate the DLPFC contained multiple sources placed 3 cm from one detector placed near F4. Motor regions were not simultaneously interrogated due to instrumentation and fiber optic probe limitations.

The probes were located with the aid of an electroencephalography net (64-channel HydroCel Geodesic Sensor Net by EGI, Inc.) applied according to EGI Inc.'s instruction. The nets were used to identify the I 10-20 locations for each participant at each visit, not to record EEG data. A mark was made under the same net pedestals for all participants. The use of the nets made localization reliable, consistent and expeditious. The MFG mark was placed under a pedestal half way between FPZ and FP2. This placed the sources for the MFG array (which were 2 cm from each other) at about the midline, with the detector then about 3 cm from the midline, to avoid the superior sagittal sinus as much as possible. The DLPFC mark was placed at the pedestal for EEG channel 59, which is immediately inferior to F4. The DLPFC array was placed by hand such that the detector and the source for channel 3 straddled its mark. The MFG array was placed similarly for channel 2. The probe arrays were secured with a separate Velcro strap for each source, allowing independent channel adjustment to achieve an acceptable balance between pressure on the head and signal. The probes were not moved between the four runs unless the participant requested adjustment for the purposes of comfort. In those cases, care was taken to relieve pressure on the head without translating the probes.

Each probe array in both parts of the study included a source probe located 1 cm from the detector for superficial physiological and nuisance signal regression as discussed in section 2.1.3. The sources located 3 cm from the detectors provided deep traces of interest which sampled

brain tissue, while the sources at 1 cm provided shallow traces which sampled only superficial tissue due to the proximity to the detector (Zhang et al., 2007a; Gagnon et al., 2011). The shallow channels were primarily sensitive to physiological changes in the superficial tissue, while also being sensitive to nuisance signal contributions such as motion of the rigidly-connected probe and residual ambient light exposure.

The Imagent instrument employs photomultiplier tubes for optical intensity detection. Gain settings for these were set on a per-participant basis after applying the probes and ensuring most of the hair was parted under the probe tips. Gain was increased and probes were re-adjusted to make better contact with the skin, iteratively, until as many channels as possible detected continuous wave intensity signal above a threshold of 500 analog-to-digital counts. None of the channels with low signal produced the best fit to the task model, and thus were not passed to the classification step. The shallow channel sources at 1 cm may damage the detectors at gain settings appropriate for sources at 3 cm due to the lack of signal attenuation over their relatively shorter optical path length. To avoid this without changing laser source intensity during the trial while the participant was wearing the probes, delivered optical intensity was reduced for the shallow channel sources. Adjustment on a per-participant basis was effected quickly by layering optically absorbent pigment on partially-transmissive cosmetic tape between the ends of the fiber probe tips and the skin.

2.1.4.2 fNIRS data processing

Both oxygenated and deoxygenated hemoglobin concentration ([Hb]) changes were calculated from the filtered raw continuous wave intensity measured traces using the Modified Beer Lambert Law (Delpy et al., 1998), then normalized (Huppert et al., 2009). As described in 2.1.3, a combination of temporal filtering and regression techniques were used to process the [Hb] time traces.

Temporal band-pass filtering of the [ΔHb] traces was used to focus on sustained task activations while removing very slow drift and higher frequency physiological and motion contributions. We call these minimally-processed *as-measured* traces, and no task knowledge is used in the processing. Standard linear regression was then used to remove physiological contributions from the as-measured signal to produce the *functional* trace (and to inform the selection of probes from each array to use for classification in part one of the study discussed in

section 3.2.1). The expected task response, when used for analysis, was modeled by convolving the boxcar task indicator function with the canonical hemodynamic response function. The task performance timing is described in section 2.3.3. The expected deoxygenated [Hb] time series was set to the negative of the oxygenated [Hb]. The shallow trace was smoothed using a 6-timepoint (1 s) moving average, and task-like response in it was removed (by a separate regression step) prior to use as a nuisance regressor in the design matrix. The functional task signal was effectively the measured Δ[Hb] signal minus the nuisance signal. The magnitude of nuisance signal to remove was determined by the second regression step (described in section 2.1.3). This was then passed to the classification step.

Within- and across-network regional correlations, respectively, were defined as the time-series correlation coefficients relating activity between different regions of the DLPFC in the task-positive network and between the DLPFC and the MFG in the task-negative network. Each of these correlation values was averaged across all four runs in each participant, before it was averaged across participants. All statistical tests on correlation and accuracy values were one-tailed, with comparisons paired by participant.

2.2 Attentional Performance and Task Engagement

Depending on the specific application, signal must be measured at appropriate brain regions that collectively exhibit some degree of functional specificity for the state or cognitive process of interest. The number of optodes may be limited by both hardware considerations and by the time available in practice for ensuring good signal at each upon application. To use fNIRS with few optodes, as opposed to many across the whole head, the location of these regions must be known *a priori*, and must be located in the outer cortical layers of the brain.

2.2.1 Exploiting Functionally-Connected Networks

Optode layouts are highly configurable, and other networks and regions can be interrogated for other operational and clinical applications, if they are known to be relevant, and are accessible to fNIRS. For this study, the attentional and resting state networks of the brain have been selected as sensor locations for this work to exploit known anti-correlation of the ATN and DMN activations. The DMN, which is one of many resting state networks of the brain, activates during resting or task-free states (Greicius et al., 2003; Mantini et al., 2007; Raichle et al., 2001). During focused attention and task engagement, relative to during a task-free or rest state, the ATN is expected to activate and the DMN is expected to deactivate. This simple model is depicted in Figure 2. Whether this anti-correlation originates due to competition for resources (Kelly et al., 2009) or is a manifestation of intrinsic interactions in the brain (Fox et al., 2005), it has been shown and quantified with fMRI studies. Also, less deactivation of the DMN has been shown to be associated with worse behavioral performance (such as stimulus reaction time) and attentional lapses (Weissman et al., 2006; Kelly et al., 2008; Kineses et al., 2008). Activity differences have been linked to autism (Weng et al., 2010), attention deficit hyperactivity disorder, and Alzheimer's disease (Uddin et al, 2009). Importantly, errors due to lapses in attention have been predicted in less than one minute prior to the error by detecting neural activity in networks (Eichele et al., 2008; O'Connell et al., 2009; Weissman et al., 2006).

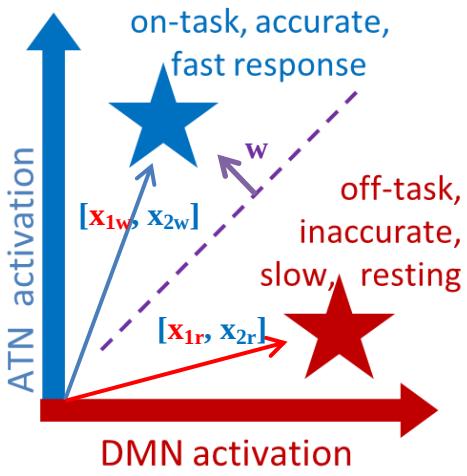


Figure 2. Expected network activations depending on task condition

ATN activations (blue) are expected to be high while DMN activations (red) are low during task engagement. With less task engagement, the opposite response is expected. The vector **w**, introduced in section 2.4, is shown normal to a dashed line separating the two conditions, or classes. A vector instance is illustrated for each class. Each vector instance has two components, or two classifier input feature instances, where component x_{1w} represents the DMN activity in the 'work' or more engaged state, and component x_{1r} represents the DMN activity in the 'rest' or less-engaged state. In this simple example, note that x_{1w} is less than x_{1r}. Similarly, component x_{2w} represents the ATN activity in the 'work' or more engaged state, and component x_{2r} represents the ATN activity in the 'rest' or less-engaged state. Note that x_{2w} is greater than x_{2r}.

The functional connectivity of these networks is an important aspect of the proposed monitoring system. The brain is regionally organized, but functional specificity of particular regions is not absolute. Particular regions of the brain are specialized for certain functional tasks, but can be active as well for a variety of different functional tasks and respond to a variety of different stimuli (Gazzaniga, 1989; Kanwisher, 2010). Thus, as introduced in 2.1.2, activity in one region indicated by a single point measurement cannot be assumed to be completely and only due to the functional task of interest. In this way, such measurements can be sensitive to a state or condition, but not specific and thus not predictive. Multiple point measurements allow for the detection of a functional network and its level of activation. Further, some spatially-separate regions of the brain undergo temporally correlated activation in response to a functional task or stimulus. These regions are said to be functionally connected, and this connectivity can

be quantified by measuring across-region temporal correlation of the hemodynamic activations in response to stimuli. Additionally, functionally-connected resting state networks exhibit activity at low frequencies (below 0.1Hz) during resting conditions (Biswal et al., 1995).

The detection of activity in multiple regions (for example at least one node each in two networks) should provide more confidence that what has been detected actually indicates the functional state that is intended to be measured. Therefore, we proposed that attentional state may be monitored through the measurement of neural hemodynamic activity in cortical layers of functionally-connected attentional and default mode network regions of the brain. If functionally-connected network activations as measured by fNIRS are correlated with behavioral measures indicating performance decrement due to lapses in attention, then sustained attention can be monitored with fNIRS. A first step toward attaining this goal is detecting and predicting task engagement relative to rest, or a lack of task engagement.

Notably no cognitive strategy is required on the part of the participant or user to employ these networks, allowing passive monitoring of the brain without hindering task performance. Further supporting the suitability of these networks for attentional state monitoring, activity is increased in the TPN (and decreased in the TNN) as task demand increases (Fox et al., 2005; McKiernan et al., 2003; Wojciulik and Kanwisher, 1999). This aspect may allow for the reduction of false alarms in non-extreme states, as the extremes of engagement foster hazardous disengagement due to a lack of arousal, or potentially dangerous performance decrement due to task overload (Yerkes and Dodson, 1908).

Additionally, the capability of detecting such network activation might prove useful in future applications of fNIRS that are aimed at discriminating between optimal behavioral performance (where a negative correlation is expected) and internally-guided thought (where co-activation and, hence, a positive correlation is expected) (Christoff et al., 2009; Smallwood et al., 2012).

2.2.2 Specific Probe Locations

The detection of multiple nodes of these networks requires measurement from locations other than just the prefrontal areas accessible through the forehead. This underscores the need for the development of headgear that can efficiently and reliably be used across the whole head. By monitoring networks across the brain instead of single locations of activity, the brain state in a

given time period is truly being monitored, as opposed to only monitoring activity over time at one location, which may not be functionally specific for the state of interest (LaConte et al., 2007). For example, fNIRS measurements made on the forehead only and averaged broadly across each hemisphere may be sensitive to workload while not indicating specific causal states (Izzetoglu et al., 2004). Such measurements should be interpreted with care, given the role of the medial frontal gyrus in the DMN. Separately measuring ATN *and* DMN activations may allow state sensing with improved accuracy by providing additional information to improve state differentiation. This research question is addressed in Chapter 3.

Detailed coordinates of specific surface optode locations are given below and plotted in Figure 3. These coordinates are used to measure the selected volumetric regions of interest. The Brodman Area (BA) is given, along with the depth of the cortex of interest and the Talairach to Ten-Twenty (TT) coordinates. TT coordinates are generated by projecting a line from a volume of cortex in the brain along the normal axis to the scalp surface (Steinstrater et al., 2001). Cz is the origin at the center, with Fz at (0, 1) and C4 at (1, 0). The two dimensional coordinate system and relevant volume to surface projections are also shown in Figure 3. The depth of the tissue is identified by the color bar on the right. The selected regions of interest are bilateral.

Based on fMRI literature (Kelly et al., 2008; Fox et al., 2005; Weissman et al., 2006), we predicted these task negative / resting state network sub-components would show decreasing activations during engagement:

- Medial Frontal Gyrus (BA 10, TT(±0.2,1.5), depth 18mm) *(marked in Figure 3 bilaterally with blue circles between Fp and Fpz; the center of the forehead)*
- Angular Gyrus (BA 39, TT(±1.3,-1.0), depth 15mm) *(marked in Figure 3 with blue circles and by projection to the area between P4 and T6)*
- Superior Frontal Gyrus (BA 8, TT(1.0,0.6), depth 16mm) *(not shown in Figure 3)*

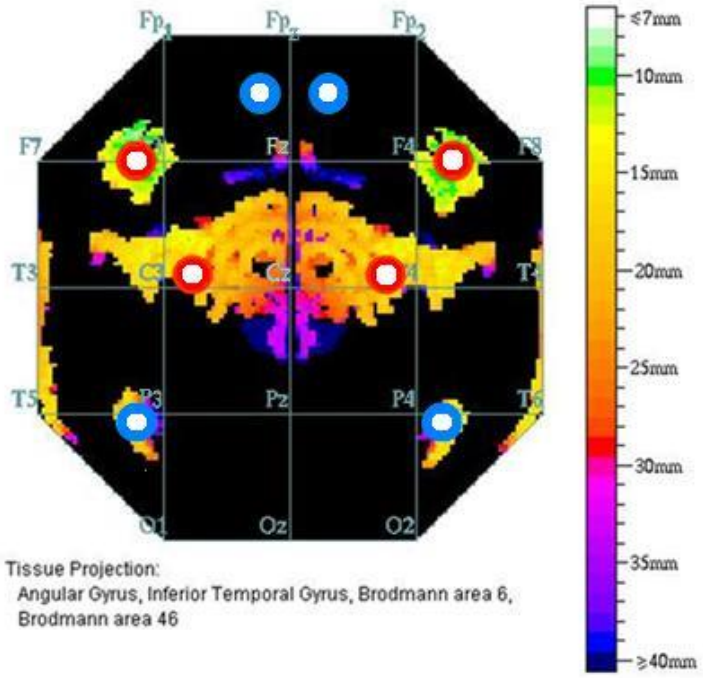


Figure 3. Tissue projections with respect to the 10-20 measurement locations

Eight measurement locations (source-detector pair midpoints) are shown schematically with cortex to be interrogated beneath the marked circles. The color scale indicates depth. The tissue projection tool is available at http://wwwneuro03.uni-muenster.de/ger/t2tconv/conv3d.html .

Similarly, we predicted these task-positive / attentional network sub-components would show increasing activations during task engagement, and together be anti-correlated with the task-negative network activations:

- DLPFC (BA 46, TT(1.3,1.0), depth 10 to 13mm) *(marked in Figure 3 with red circles and by projection to the area between F4 and F8; just into the hairline from the temple)*
- Supplementary Motor (BA 6, TT(0.9,0.3), depth 15mm) *(marked in Figure 3 with red circles and by projection to the area anterior to the line between C3 and C4)*
- Inferior Parietal Lobule (BA 40, TT(1.3,-0.5), depth 19mm) *(not shown in Figure)*

We have chosen to focus initially on the following two locations. The dorsolateral PFC (DLPFC) is of particular interest to us for monitoring task engagement and sustained attention (vigilance) (Posner, 2004), and is also part of the TPN. The Medial Frontal Gyrus (MFG), as a

node of the DMN (Greicius, et al. 2003; Raichle, et al. 2001), is of particular interest for indicating rest or a low level of task engagement, and is also part of the TNN. (Kelly et al., 2008; Fox et al., 2005; Weissman et al., 2006)

Details of the implementation of classification algorithms to predict operator state (engaged or not engaged in a task) based on measures of activity in these regions will be discussed in section 2.4. First, we present methods for inducing the task engagement we wish to quantify.

2.3 Behavioral Methods

To facilitate our ability to make this distinction, we asked participants to perform the multi-source interference task (MSIT) (Bush and Shin, 2006; Stins et al., 2005). The MSIT is a selective attention task in which optimal performance requires participants to suppress multiple sources of interference. Stroop interference may be due to any semantic meaning of the prompt, Eriksen interference is present in non-congruent trials due to flanking distractors, and Simon interference is present in non-congruent trials due to conflict between the placement of the target numbers and the finger used for the correct response. Thus, it reliably and robustly activates the TPN, even in individual task blocks from the same participant (Bush and Shin, 2006). Given these characteristics, we reasoned that the MSIT would provide a strong signal with which to monitor task engagement. Consistent with this reasoning, in the present study we were successful at distinguishing between relatively high and relatively low periods of task engagement.

The variability of the cognitive functions themselves, the task execution strategy, the presence of internally-guided or task-unrelated thoughts, and the physiological state of the operator, are inherent confounding factors (Arenth et al., 2007) that change over time. These were not corrected for in this study, beyond controlling for these variances by collecting the fNIRS and fMRI data simultaneously in part two of the study. Results are presented in sections 3.4 and 4.2. All statistical tests on behavioral results were one-tailed, with comparisons paired by participant.

2.3.1 Participants

Human subject data were collected according to a protocol approved by both the University of Michigan IRB MED and the NASA IRB. Informed consent was obtained from all participants, who were healthy adults between the ages of 21 and 50 years. All participants practiced the behavioral task with performance feedback for 1 minute at the beginning of the study. Next, they performed a set of four 7-minute-long runs. All participants were right-handed but one, who was ambidextrous, determined by self-report. The fact that all participants readily used their right hand for response was convenient for our use of the motor response for one of

the fMRI classification cases in part one of the study. Signals were taken from the left motor cortex for use in a "co-activating" case to be described in section 3.5.1.

2.3.2 Behavioral Task

In each trial of the MSIT (duration, 2 s), participants viewed a horizontally-oriented array of four digits at the center of the screen (duration, 1 s). A target digit was printed in a larger font than each of three distracter digits (72 point vs. 60 point). The stimulus sizes were chosen to make the digits visible within the MRI scanner. Moreover, the large target digit was chosen to be about 20% larger than the small distracter digits to make the task sufficiently difficult. Indeed, Stins et al. (2005) observed a much smaller interference effect in the MSIT when the large target digit was 33% larger than the small distracter digits. Since the goal of the present MRI experiment was to contrast activity during task performance to activity during rest, we wanted to ensure that the task was difficult enough to elicit a good deal of task-related activity.

Participants were instructed to identify the target digit (1, 2, 3, or 4) by pressing a key with one of four fingers (1 = right thumb, 2 = right index finger, 3 = right middle finger, 4 = right ring finger) as quickly as possible without making mistakes. For congruent trials (50%), both the spatial position of the target and the identity of the three distracter digits were mapped to the correct response (**1**111, 2**2**22, 33**3**3, 444**4**). For incongruent trials (50%), both the spatial position of the target and the identity of the three distracter digits were mapped to a conflicting response (**2**111, 2**1**22, 33**4**3, 444**3**). Participants alternated between trials composed of 1s and 2s and trials composed of 3s and 4s across trials to prevent immediate stimulus and response repetitions (Mayr et al., 2003; Jimenez et al., 2013).

Responses were recorded via the keyboard (the 'n', 'u', '9' and '0' keys) of a laptop used in fNIRS experiments of part one of the study, and via a MR-compatible response device (Psychology Software Tools, Inc., Sharpsburg, PA) in all fMRI experiments. The task was implemented using a combination of MATLAB (Natick, MA) and the Psychophysics Toolbox (Kleiner et al., 2007).

2.3.3 Functional Neuroimaging Experimental Design

We employed a block design. In each of four runs (lasting 400s), an initial 16 s rest block was followed by 12 alternations between the MSIT (16 s, 8 trials per block) and rest (16 s).

2.4 Support Vector Machine Classification Methods

Classification was performed using time traces as input features to Support Vector Machines (SVMs). Scripts were implemented in MATLAB using LibSVM (Chang and Lin, 2011). We choose SVMs for good performance with ease of implementation, flexibility in number of input features, processing speed once trained in the interest of future real-time application, and the ability to use tuning parameters to optimize feature separation for each participant (this is discussed further in section 3.2.1). SVMs have shown potential for high accuracy during sustained activity with minimal training, and have been implemented in real time fMRI methods (Laconte et al., 2007).

SVMs determine a class to be assigned to test data based on a model produced using *training* data. The data to be tested (or to be used for *prediction*) and the data to be used for training must be in the same form. That is, they are both vectors of the same number of components. Each component is said to be an input *feature*. Here we focus on SVMs with two classes as used throughout this study with various numbers of features to perform classification in different cases. Every prediction and every training *instance* is thus a vector of values for each feature at that instance of time. A truth label (y_i) is used to indicate the known class for the training instance, indicated by i. Similar labels can be used, if known, to test classifier accuracy by comparing the predicted class to the known class.

In Figure 2, we schematically depict a two-class SVM using two features. One feature quantifies activity in the ATN or TPN, while the second quantifies activity in the DMN or TNN. A vector **w** is shown normal to a dashed line separating the two conditions in the training data into two classes. A vector instance is illustrated for each class. Each vector instance has two components, or two classifier input feature instances. The vector **w** is normal to a line that separates the two classes, which is determined during classifier training and model production. To perform prediction, the classifier takes a dot product between **w** and the instance vector to be predicted. The scalar result determines the class to which that prediction instance best belongs.

In the case of more than two features, the line would be a plane or hyperplane. During training, the placement of this plane is optimally determined to separate the two classes with the widest *margin* between it and the nearest instance on either side. It is optimal because the width of the margin is maximized while the cost of misclassification of labeled training data is minimized. Mathematically, this is expressed as

$$\min_{\mathbf{w},\xi,\mathbf{b}} \{ 1/2 \|\mathbf{w}\|^2 + c \Sigma_i \xi_i \} \qquad \text{subject to } y_i(\mathbf{w'x}_i - b) \geq 1 - \xi_i \qquad (1)$$

where $\mathbf{w'x} = b$ defines the separating plane and ξ_i is a loss > 0 based on the misclassification of the training data instance, where $\mathbf{w'}$ is the transpose of $\mathbf{w}$. The two planes marking the edges of the margin are defined by $\mathbf{w'x} = b + 1$ and $\mathbf{w'x} = b - 1$. Minimizing $\mathbf{w}$ maximizes the margin because the distance between the planes is inversely proportional to $\|\mathbf{w}\|$. (Cortes and Vapnik, 1995)

A radial basis function was employed for the SVM. This uses a Gaussian kernel (K) to map our few input features (x_i) into a new, multidimensional feature space (x_j). (Chang and Lin, 2011)

$$K(x_i,x_j) = \exp(- g \|x_i - x_j\|^2) \qquad \text{for } g > 0 \qquad (2)$$

The classes may be more separable in the new space. Instances to be predicted and instances for training are mapped in the same way, and predictions are determined in the new feature space.

Thus there are two tuning parameters that must be selected prior to training, and thus can be optimized for each participant individually. The first tuning parameter is c, which multiplies the cost of misclassifications as in (1) during model training. A lower c can produce a more flexible and generalizable SVM model by reducing the impact of misclassifications in exchange for finding a wider separation margin between the two classes. If c is too low, constraints are ignored and a useful solution is not found. If c is too high, the training data are over-fit and resulting accuracy upon generalization to the prediction of multiple sets of test data suffers.

The second is kernel parameter gamma (g), which determines the non-linearity of the mapping discussed above as in (2). Non-linear mapping may improve performance depending on g, especially considering the relatively small number of features used in the case of fNIRS (compared to whole-brain fMRI as discussed in sections 4.5.3 and 4.6.3). This is the advantage of a non-linear classifier relative to a linear classifier. Arguably linear classifiers are more

straightforward for interpretation. However, a benefit of machine learning is the identification of predictable patterns that may not be apparent to humans.

A SVM classifier was trained (producing a model) to discriminate high from low levels of task engagement (two classes of attentional *state*) using one or three of each participant's four runs (7 min or 21 min of training data). Its prediction accuracy was tested on the participant's fourth run. Thus, all training and prediction were conducted within participants. All permutations were computed for each participant for each c, g pair, such that the best SVM model prediction accuracy could be determined for each of the 4 runs. The truth labels used for training and accuracy determination purposes were determined from the boxcar task indicator signal, but shifted 4 seconds later to account for the delay of the hemodynamic response (the green trace in Figure 9).

To determine the best possible prediction accuracy achievable for each participant in this study, c and g were tested at various selected values *using the training data only* to optimize accuracy. The highest classification accuracies produced using the same set of parameters across all four runs (the same c and g for all permutations) were reported for each participant and reported in sections 3.2.1 and 4.4. The best c, g parameters found upon optimization were not extreme in value and not identical across participants. This is discussed further in section 3.2.1. Accuracy was determined by comparing the work or rest state predicted by the classifier for each instance to the known behavioral states (i.e., performing the task or resting) at that time.

Prior work has demonstrated the ability to discriminate periods of time during which a participant is engaged or not engaged in a complex task with accuracy of 71% +/- 3%, and potentially up to 88%, using 7 feature traces across the same two locations used in the present study, being the MFG and the DLPFC. Minimal signal processing included temporal filtering only, and artifact-free, within-participant, within-session data was used to train a non-linear support vector machine classifier (Harrivel et al., 2011). In this prior study, the classifier was trained with real data that fit the expected activation model. In other words, an "attending" truth label was applied to example instances where DLPFC activity ([HbO] increase relative to [HbR] as described in section 2.1.1) was relatively high and MFG activity was relatively low, while a "resting" truth label was applied to example instances where the inverse activity occurred.

2.5 fMRI Methods

fNIRS offers the advantage of direct comparison to hemodynamic activations as measured by fMRI. To facilitate such comparison, fMRI data acquisition, processing and classification were executed as follows.

2.5.1 fMRI Data Acquisition

fMRI data were acquired using a 3T GE MR750 MRI scanner with a T2*-weighted spiral-in BOLD sequence with pulse sequence parameters TR/TE/FA = 2s/30ms/90°. The FOV was 22 cm in a 64x64 matrix for 40 slices with 3mm thickness. The total time for each scan (400 s) was matched to the functional task at 200 volumes, after discarding 5 volumes at the beginning of each scan. Physiological signals were collected concurrently using a pulse oximeter and chest plethysmograph. For anatomical reference in the functional data analysis, a high-resolution T1-weighted anatomical image was collected using spoiled-gradient-recalled acquisition (SPGR) in steady-state imaging with pulse sequence parameters TR/TE/FA = 12.2 ms/5.2 ms/15°. The FOV was 26 cm in a 256x256 matrix for 136 slices at thickness 1.2 mm.

2.5.2 fMRI Data Processing

After slice-timing and motion correction, and physiological respiratory and cardiac noise removal with RETROICOR (Glover et al., 2000), fMRI data were co-registered to the participant's anatomical scan, normalized to the MNI template (Collins et al., 1994), smoothed, modeled and estimated with SPM8 (Wellcome Department of Cognitive Neurology, London, UK). Affine registration to the MNI space template was performed with nonlinear regularization set to 1 for smooth deformation fields. Spatial smoothing was performed with a Gaussian kernel of 8x8x8 mm at full width half maximum. Normalized volumes were rewritten with 2x2x2 mm voxels. High pass filtering was set with a cutoff frequency of 0.0078Hz. The resulting image files contained BOLD activation time traces for each voxel of the brain and were passed on to the classification step. Importantly, TPN and TNN anti-correlation is expected to be seen both with and without physiological correction (Chang and Glover, 2009).

Both “rest minus task” and “task minus rest” contrasts were generated and used in second level analyses to verify fNIRS probe placement. No global mean was used. These results are presented in section 2.6.

2.5.3 fMRI Classification

Processed fMRI BOLD traces were used as SVM inputs features in three ways: 1) after averaging across two clusters centered on participant-specific local maxima within the regions of interest in part one, 2) after averaging across six clusters centered at participant-specific locations in the cortex in part two, and 3) by using traces from each of approximately 30,000 voxels coinciding with voxels in a skull-stripped anatomical image for that participant in part two.

For the fMRI cases, each run is predicted three times, once per training run used. For example, run A is predicted using a model trained with run B, then with a model trained with run C, then with a model trained with run D. Three accuracies are produced, and an average is reported as the result for the prediction of run A. The process is repeated for the next permutation, predicting run B by training separately with runs A, C and D, and so forth until all four runs have a reported prediction accuracy. Otherwise, all methods for classification were the same as those for the fNIRS traces as described in sections 2.4 and 3.2.1.

2.6 fMRI Verification of Task Activation

The locations of statistically significant activations for the MSIT, after second-level random effects analysis across nine independent participants from both study parts one and two (using level two results across the four runs for each), are shown for the DLPFC (with a local maximum at [48, 18, 34] in MNI space; Figure 4, left) and for the MFG (with a local maximum at [-20, 68, 0] in MNI space; Figure 4, right). The t-statistic is mapped, with the threshold set at an uncorrected significance level of $p < 0.01$ for the work minus rest contrast, and at $p < 0.1$ for the rest minus work contrast. Of importance, the experimental task activated the networks as expected, with "task-positive" and "task-negative" hemodynamic activations occurring in the same regions that were interrogated by the fNIRS probes. Notably, for a separate fixed effects analysis across the 36 runs, the DLPFC and PCC regions survived a family wise error correction at $p < 0.01$ and $p < 0.05$, respectively; the MFG region, however, did not survive this correction.

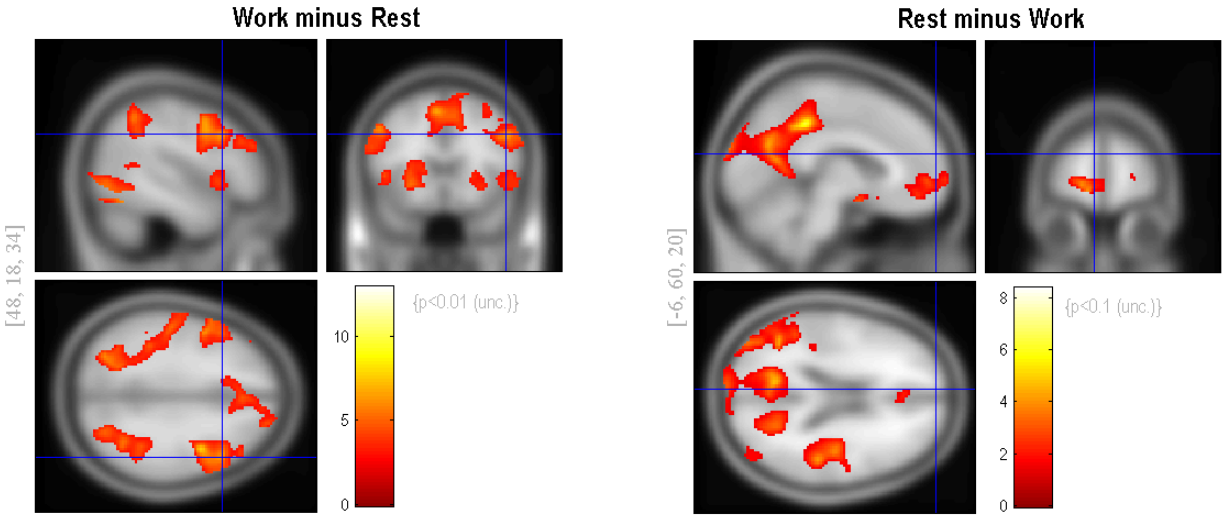


Figure 4. Locations of statistically significant activations for the MSIT

Locations of statistically significant activations for the MSIT are shown after second-level random effects analysis across nine independent participants. The expected "task-positive" and "task-negative" hemodynamic activations occurred in the same regions that were interrogated by the fNIRS probes. Left: The t-statistic is mapped, with the height threshold set at an uncorrected

significance level of $p < 0.01$ and the extent threshold set at 10 voxels. The DLPFC local maximum is shown by the crosshair at [48, 18, 34] in MNI space for the work minus rest contrast. Right: The t-statistic is mapped, with the height threshold set at an uncorrected significance level of $p < 0.1$ and the extent threshold set at 10 voxels. The MFG local maximum is shown inferior to the crosshair at [-6, 60, 20] in MNI space for the rest minus work contrast.

Finally, we note that although other regions of the "task-negative" network were identified in the fMRI analysis, they may be less useful for monitoring task engagement with fNIRS. First, lateral parietal regions were not consistently activated bilaterally across participants. Thus, monitoring both sides with fNIRS would present greater difficulty due to the increased number of optical probes. Second, although precuneus and PCC regions of the task-negative network (Greicius et al. 2003; Raichle et al. 2001) were reliably activated (Figure 4, right), they are too deep to be accessible via fNIRS probes, which can interrogate only the outer layers of the cortex (Boas et al., 2004) as explained in sections 2.1.1 and 2.1.2.

Chapter 3
Quantifying Task Engagement with Functional Near Infrared Spectroscopy

This chapter focuses on methods and results particular to part one of the study (Harrivel et al., 2013). Human participant trials using both fNIRS and functional Magnetic Resonance Imaging (fMRI) were used to monitor brain activity to determine whether it is possible to distinguish between cognitive states associated with high vs. low task engagement. Methods were those of sections 2.1.4, 2.3, 2.4 and 2.5 unless note in this chapter. fNIRS optodes were placed as described in section 3.1. Optode channels were selected and used as classifier features for part one of the study according to the methods presented in 3.2.1. Results are shown and discusses for: task engagement classification accuracy with fNIRS signals in 3.2.2, across-network anti-correlation in 3.3, behavioral effectiveness of the MSIT task in 3.4, and classification with fMRI BOLD signals in 3.5.2 using methods given in 3.5.1. The results of fMRI trials verifying neural activation using the MSIT in the regions interrogated by the fNIRS probes were presented in 2.6.

For part one of the study, seven participants (three female, four male) completed the fMRI study. Five participants (two female, three male) completed the fNIRS study (including two from the fMRI study). Four runs each were completed for the fNIRS and fMRI experiments, which were performed on separate days. Otherwise, behavioral methods were as described in section 2.3.

3.1 fNIRS Probes and Locations

The array of head probes and fiber optic cables used to interrogate the MFG in part one of the study is shown in Figure 5, right, and placed as described in section 2.1.4.1. The source fibers were 400 micron core glass fibers and each detector was a 3mm diameter glass fiber bundle. The sources were held in place using both clear and blacked-out plastic at the locations shown with respect to the I 10-20 locations in Figure 5, left. The array of probes used to interrogate the DLPFC contained seven sources placed 3 cm from the detector placed near F4.

The line toward Cz was used to orient optode channel 1. The clear material improved visual inspection for placement and hair control at the probe-skin interface. Blackout material was applied over the probes to block ambient light. A Velcro strap was used at each source probe to steer the probe along the normal to the head while attaching it.

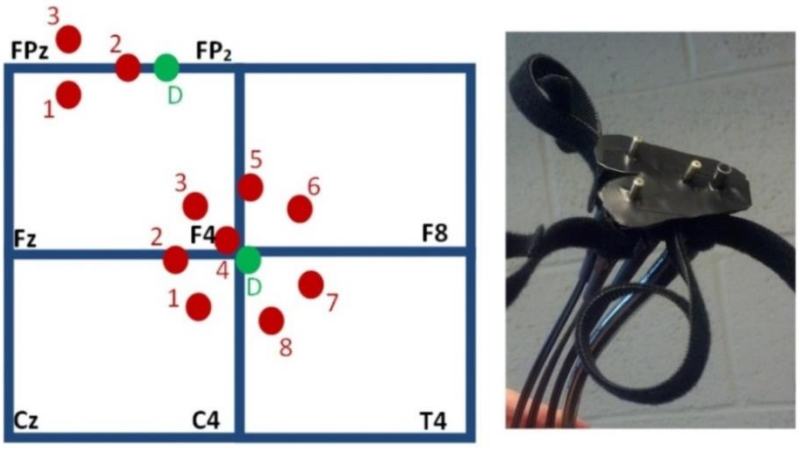


Figure 5. Head probes used to interrogate the DLPFC and MFG in part one of the study

Left: Source channels (red) are arranged at a separation distance of 3 cm from two detectors (green), shown schematically with respect to the I 10-20 locations. The shallow sources, located at 1 cm from the detectors, provide channel 4 for the DLPFC array near F4, and channel 2 for the MFG arrays between FPz and FP_2. The probes consistently producing the traces with the highest task model fit parameters were DLPFC channels 1, 2 and 3 and MFG channel 1 (numbered, left). Right: The skin-side of the head probes and fibers used to interrogate the MFG in study part one.

Figure 6. DLPFC head probe array holder for part two of the study

Velcro strap attachments used in both parts are also shown.

3.2 fNIRS State Prediction using both TPN and TNN Features

3.2.1 fNIRS Processing and State Prediction Methods

Except as noted specifically for part one of the study, signals were processed as described in 2.1.4.2. Specific to part one of the study, temporal band-pass filtering was set to include 0.008Hz to 0.08Hz to capture sustained task-evoked activations and very low frequency components attributable in part to resting state functionally connected networks (Fox et al., 2007; Biswal et al., 1995). In part one of the study, shallow physiological trace removal was performed within-species. That is, the oxygenated [Hb] trace from the shallow channel was regressed from the measured oxygenated trace of each deep channel on that detector's array, and similarly for the deoxygenated traces.

An example of the data processing method is shown in Figure 7 for one participant for one run. The task indicator signal is shown in green in each of the four panels, with +1 indicating a work period of the MSIT. Expected activity is shown in thin black in the top three panels. The deep [HbO] signal as measured at the MFG is shown in blue (after temporal band-pass filtering only) in the top, third and fourth panels. The shallow [HbO] physiological is shown in red in the second panel, where the solid line shows the signal used as the nuisance regressor and the dotted line is the signal prior to removal of task effects. The functional signal is shown by the thick black line in the third panel. The functional signal is the measured signal minus a fraction of the shallow signal of the same Hb species, with that fraction determined by the regression fit. The residual trace is shown in thin black in the fourth panel. Since the shallow-signal-producing optode is participant to the same gross probe motion and ambient light environment as the deep-signal-producing optodes, multiple types of artifact remaining in the signal after temporal filtering are reduced via this data processing method.

The oxygenated or deoxygenated [Hb] species functional trace from one optode was selected for use in classification based on how well the measured deep traces fit the expected functional activation task response, quantified by their beta fit parameter. The resulting parameters for each Hb species on each channel are given by participant and run in Appendix B.

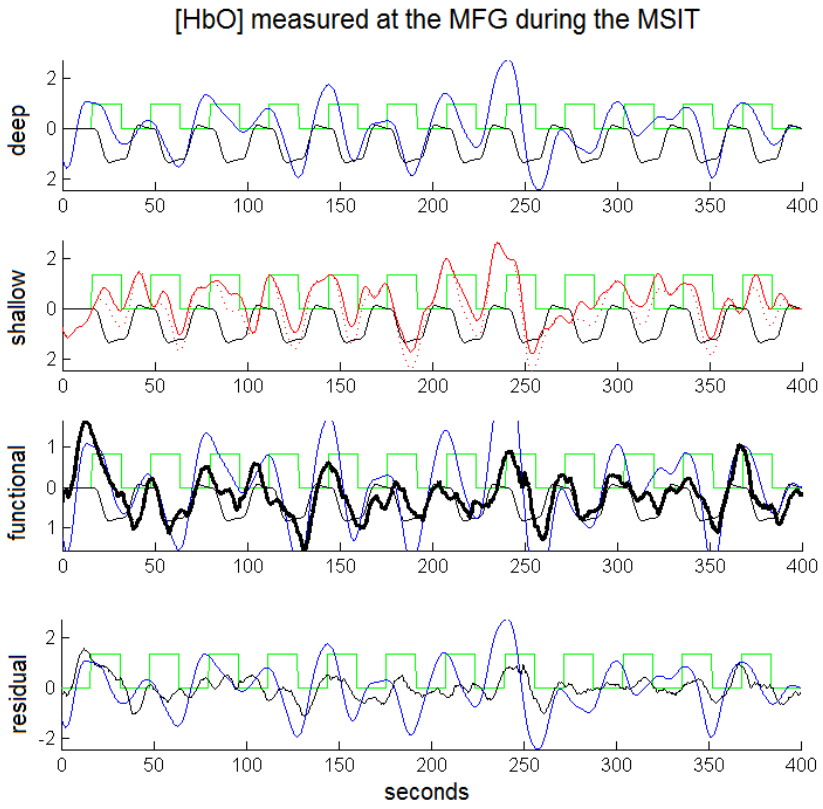


Figure 7. Signal processing for study part one fNIRS data

The task indicator signal is shown in green in each of the four panels, with +1 indicating a work period of the MSIT. Expected activity is shown in thin black in the top three panels. The deep [HbO] signal as measured at the MFG is shown in blue (after temporal band-pass filtering only) in the top, third and fourth panels. The shallow [HbO] physiological is shown in red in the second panel, where the solid line shows the signal used as the nuisance regressor and the dotted line is the signal prior to removal of task effects. The functional signal is shown by the thick black line in the third panel. The functional signal is the measured signal minus a fraction of the shallow signal, with that fraction determined by the regression fit. The residual trace is shown in thin black in the fourth panel.

As described in section 2.4, classification was done separately for within- and across-network input pairs. First, the best two traces were selected from the array of seven deep DLPFC channels (a within-network pair, noted as DLPFC and DLPFC2 in Figures 8 and 10), based on which traces (oxygenated or deoxygenated) best fit the functional activation model (which had the highest beta fit parameter considering all four runs). The traces which best fit the task model

were assumed to make the best input features for producing the highest classification accuracy. The same best DLPFC trace was then paired with the best trace from the array of two deep MFG probes (an across-network pair, noted as DLPFC and MFG in Figures 8 and 10) and a separate, second classification step was performed. Trace selections were not changed across runs.

To determine the best possible prediction accuracy achievable with the methods of this part of the study, c was tested at 0.1, 1 and 10, and g was tested at 0.001, 0.01 and 0.1 (for a total of 9 cases) in each classification step to maximize accuracy for both within- and across-network cases. These values were selected after running a test case varying c and g each across 8 orders of magnitude (please see Appendix C). Accuracy did not change appreciably across four orders of magnitude, and the parameter ranges were selected in those regions (e.g. $0.1 < c < 10$ and $0.001 < g < 0.1$ in Appendix C). The c, g pair producing the highest accuracy for the training data for that participant was selected for prediction purposes for that participant.

Despite relative insensitivity to the specific c and g values in the selected range, tuning classifier parameters between participants allows individual differences to be taken into account for optimal accuracy. The fNIRS signals and therefore the classifier input features and associated optimal classifier parameters will be different per participant. Since the best parameters depend on the separability of the data and the shape of the boundary between data points in the two conditions, or classes, the optimal c, g pair might change per participant and per optode application. Classifier features may be more or less readily separated into the two classes, depending on the strength of the signal of interest relative to other components in the features which are less informative to the classifier. This signal contrast can be affected by different psychological response strategies, different physiological processes (contributing to varying degrees in both space and time), and different partial tissue volumes sampled by the optodes. There is no reason to force the threshold separating the two classes to be the same across participants, unless one seeks to generalize and apply SVM models across participants, which was not done in this study.

Training and generalization to prediction was accomplished within-participant. Further, tuning may be implemented for realistic operations, allowing use of the best capability of the classifier by customizing the system to each particular subject's training data set. Across-task and across-visit accuracy (discussed further in section 3.6.3), and the parameters which best

achieve them, are left to additional investigations. In across-participant application of a SVM model, variability in c and g does become important.

Correlations were calculated (as described in section 2.1.4.2) between the two functional time traces used as classifier input features. These results were averaged across that participant's four runs, and are presented in Table 1.

Table 1. Classification parameters and correlation values for study part one fNIRS data
The best c, g parameters found after optimization for each participant for fNIRS classification, and the time trace correlations (r) averaged across that participant's 4 runs.

participant	c	g	r (within)	r (across)
1	1	0.1	0.84	-0.21
2	10	0.001	0.23	-0.01
3	10	0.01	0.82	-0.16
4	10	0.1	0.95	-0.07
5	0.1	0.001	0.91	-0.15

3.2.2 fNIRS State Prediction Results

The four-run average classification accuracy for each of five participants is presented in Figure 8. As expected, we were able to distinguish between task engagement and rest. In particular, both averages differed significantly from chance at 50% (across: $t(4) = 9.65$, $p < 0.0005$; within: $t(4) = 4.95$, $p < 0.005$). Also in line with predictions, there was a non-significant trend toward greater classification accuracy for across-network pairs (M = 69.1%, SD = 4.4%) than for within-network pairs (M = 66.0%, SD = 7.2%), ($t(4) = 1.48$, $p < 0.25$). The probe channels resulting in oxygenated or deoxygenated [Hb] traces with the highest task model fit parameters were 1, 2 and 3 (as numbered in Figure 5, left; data in Appendix B). The [Hb] species of the best-fitting traces were nearly evenly split: eight were oxygenated and seven were deoxygenated [Hb] traces. Thus, we found no universally best Hb species for fitting the task model. This result is consistent with prior suggestions that a probe's sensitivity to one species or the other depends on whether it mostly samples the arterial or venous compartment (Strangman et al., 2003), which is likely to change for every probe application.

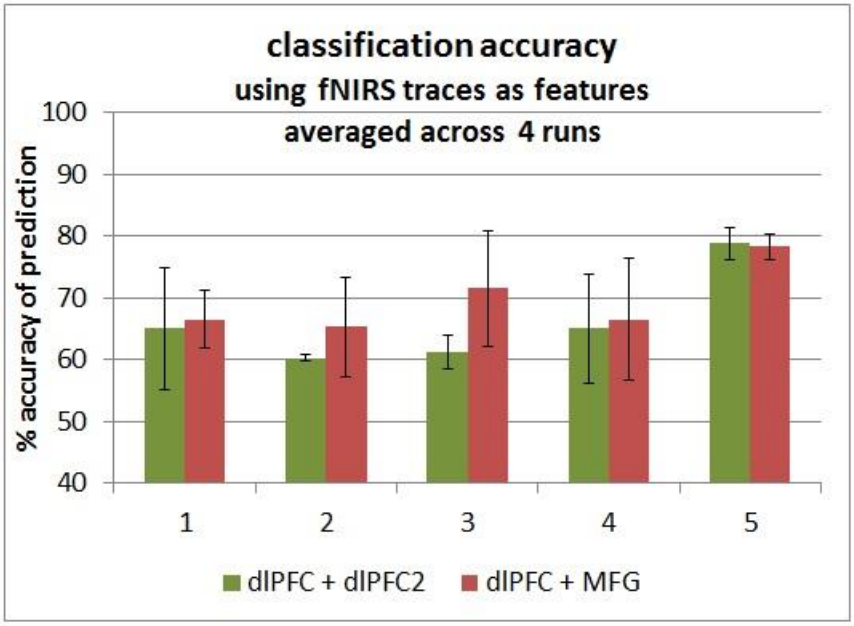


Figure 8. Classification accuracy for study part one fNIRS data

Classification accuracy averaged across four runs for each of five participants. Two fNIRS time traces were used as support vector machine input features in each case. Left-hand bars: within-network pairs (green). Right-hand bars: across-region pairs (red). Accuracy is the number of time points for which the prediction matched the truth label out of all 2500 time points. An accuracy of 50% represents prediction by chance. Error bars represent ± one standard deviation.

Prior work indicates that activity in key regions of the "task-positive" network is positively correlated while activity in key regions of the "task-positive" and "task-negative" networks is negatively correlated (Fox et al., 2005; Kelly et al., 2008). Consistent with such findings, the group-averaged correlations for within-network pairs were significantly greater than zero ($r = 0.75$, $SD = 0.29$; $t(4) = 5.72$, $p < 0.005$), while those for across-network pairs were significantly less than zero ($r = -0.12$, $SD = 0.08$; $t(4) = 3.41$, $p < 0.025$). Further, the correlation averages for within-network pairs were significantly higher than those of the across-network pairs ($t(4) = 5.56$, $p < 0.005$; see Table 1 for a participant-specific list of correlation values). These findings suggest that our fNIRS probes accurately measured activity in the "task-positive" and "task-negative" networks.

3.3 Across-Network Anti-Correlation

Finally, since across-network correlation was determined on a per-participant basis, we also wished to verify that negatively correlated across-network activity was observed at the group level. To this end, we averaged the across-network functional task signals across all 20 runs. For the purposes of this group average, all deoxygenated traces were first inverted, consistent with a reduction of deoxygenated [Hb] during activation. The group-averaged time traces are presented in Figure 9, which shows filtered, normalized and corrected [Hb] changes (see fNIRS methods; truth labels (green trace) show task at +1 and rest at -1). As expected, the across-network correlation determined in this fixed effects analysis was significantly less than zero ($r(18) = -0.58$, $p < 0.01$). This finding illustrates that, even at the group level, DLPFC activity (blue trace) increased during task performance while MFG activity (red trace) decreased.

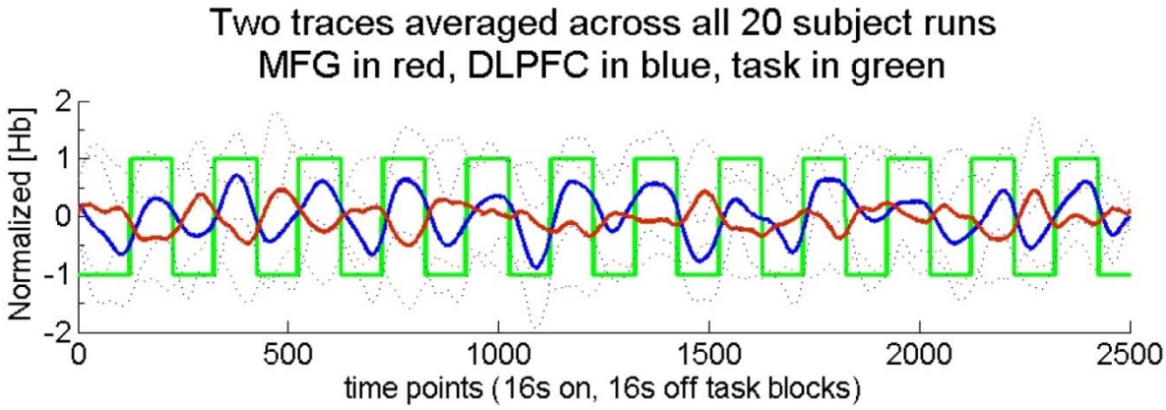


Figure 9. Two across-region fNIRS time traces

Two across-region fNIRS time traces averaged across all 20 runs. These traces illustrate filtered, normalized, and corrected (see fNIRS methods) hemoglobin concentration changes. For the purposes of this group average, all deoxygenated traces were first inverted. Truth labels are indicated by the green trace with task at +1 and rest at -1. The across-network correlation for the group was -0.58. Task blocks were 16 seconds each, with equal rest time. Dotted lines show ± one standard deviation for each time point.

3.4 Behavioral Results

One participant performed the task incorrectly for incongruent trials. Thus, this participant's data were excluded from the behavioral analyses. However, since this participant was engaged in the task and responding to stimuli, these trials were not excluded from the fNIRS and fMRI analyses. Including these data was appropriate because the fNIRS and fMRI analyses were aimed at distinguishing between performing a task and resting, rather than distinguishing between incongruent and congruent trials. Mean accuracy was 98% (SD = 2.2%, N = 4) for the four fNIRS participants who performed the task correctly, and 97% (SD = 4.7%, N = 6) for the six fMRI participants who performed the task correctly. As expected, mean accuracy was relatively high.

Also as expected, performance was worse in incongruent than in congruent trials (the analysis of the data included only the participants who performed correctly in most incongruent trials). In a random effects analysis, mean reaction time was significantly higher in the incongruent condition (fNIRS: M = 0.639s, SD = 0.035, N = 4; fMRI: M = 0.789s, SD = 0.114s, N = 6) than in the congruent condition (fNIRS: M = 0.552s, SD = 0.033s, N = 4; fMRI: M = 0.714s, SD = 0.096s, N = 6), (fNIRS: $t(3) = 26$, $p < 0.0005$; fMRI: $t(5) = 9.1$, $p < 0.0005$). Likewise, mean error rate was significantly higher in the incongruent condition (fNIRS: M = 2.0%, SD = 1.8%, N = 4; fMRI: M = 4.9%, SD = 6.4%, N = 6) than in the congruent condition (fNIRS: M = 0.38%, SD = 0.81%, N = 4; fMRI: M = 2.1%, SD = 3.5%, N = 6), (fNIRS: $t(3) = 2.9$, $p < 0.05$; fMRI: $t(5) = 2.2$, $p < 0.05$).

Errors of omission were rare and thus not analyzed. Most runs were free of such errors. One subject committed a maximum of three, two, two and none during the four runs while all other subjects committed none. Each run contained 96 trials.

3.5 fMRI Validation of State Prediction

Both “rest minus task” and “task minus rest” contrasts from section 2.5 were used in additional second level analyses performed separately for each participant across that individual participant’s four runs to guide voxel selection for fMRI classification. No global mean was used.

3.5.1 fMRI Processing Methods

For part one of the study, all fMRI traces used as SVM inputs were processed BOLD responses, averaged across clusters centered on local maxima within the regions of interest (DLPFC, primary motor cortex, and MFG). Eighteen voxels were symmetrically selected in a cubic shape around participant-specific centers to be included in the average. Participant-specific locations were selected using the second-level statistical maps generated using that individual participant’s four runs. The location for the DLPFC region was selected using the second level analysis of the work-rest contrast. The location for the MFG region was selected using the second level analysis of the rest-work contrast. The voxels used were not contiguous, and the region of interest spanned 1 cm per side in MNI space. In this way, the SVM input features for fMRI were restricted to traces averaged across local tissue. This is analogous, for fairness of comparison between the modalities, to the volume of tissue interrogated by one fNIRS probe, which is on the order of centimeters (Boas et al., 2004).

Classification was performed with a fMRI trace selected from the contralateral motor area, which was paired with one from the DLPFC region (a co-activating pair, noted as DLPFC and Motor in Figure 10). The same DLPFC trace was paired with one from the MFG region (an across-network pair, noted as DLPFC and MFG in Figure 10), and classification was performed again. The same DLPFC trace was then paired with a second DLPFC trace from a region 2 cm superior and 2 cm medial to the first DLPFC trace in MNI space (a within-network pair, noted as DLPFC + DLPFC2 in Figure 10), and classification was performed a third time. To optimize accuracy, c was tested at 0.0001, 0.001, 0.01, 0.1, 1, 5, 10 ,100 ,1000 , and g was tested at 0.00001, 0.0001, 0.001, 0.01 ,0.1 ,1, 5, 10, 100 (a total of 81 cases) in each classification step.

The motor region was selected for its robust and reliable response during task periods. We treat classification based on time traces from the motor region as a gold standard. That is, we do not expect fNIRS classification accuracies to exceed those attainable using motor cortex activations measured with fMRI.

Across-network, within-network and co-activating correlations were defined, respectively, as the correlation coefficients between across-network (i.e., the DLPFC and the MFG), within-network (i.e. DLPFC and DLPFC2), or co-activating (i.e., the DLPFC and motor cortex) pairs of functional task signals. The fMRI traces were additionally smoothed across 10 seconds to reduce the impact of noise in the signal, before the correlation coefficient was calculated. These were averaged across runs by participant, then averaged across participants.

The SVM tuning parameters producing the best classification accuracies after optimization, and the tie trace correlations, are presented in Table 2.

Table 2. Classification parameters and correlation values for study part one fMRI data
The best c, g parameters found after optimization for each participant for fMRI classification, and the time trace correlations (r) averaged across that participant's four runs.

participant	c	g	r (co-acting)	r (within)	r (across)
6	5	0.1	0.76	0.78	-0.23
7	1	1	0.06	0.06	0.10
8	0.1	0.01	0.57	0.48	0.34
9	1	1	0.72	0.85	-0.03
10	1	0.01	0.50	0.52	0.16
11	1	1	0.51	0.47	0.47
12	1000	0.00001	0.62	0.62	0.41

3.5.2 fMRI Classification Results

The four-run average classification accuracy for each of the seven participants is presented in Figure 10. Replicating the fNIRS results, group averaged classification accuracy was significantly greater than chance at 50% for all three types of region pairs (across: $t(6) = 3.56$, $p < 0.01$; within: $t(6) = 2.56$, $p < 0.025$; co-activating: $t(6) = 7.96$, $p < 0.0005$), indicating we were able to distinguish task engagement from rest. Also as expected, group averaged

classification accuracy was significantly higher for the co-activating (DLPFC and motor) pairs (M = 74.1%, SD = 8.0%) than for the across-network pairs (M = 65.6%, SD = 11.6%), ($t(6) = 4.83$, $p < 0.005$), consistent with the high reliability of motor cortex activation detection in fMRI studies (Möller et al., 2005) and with the motor cortex activation associated with button press responses in the present task. Within-network pair accuracy (M = 63.0%, SD = 13.5%) tended to be lower than across-network pair accuracy, but not significantly ($t(6) = 1.57$, $p < 0.1$).

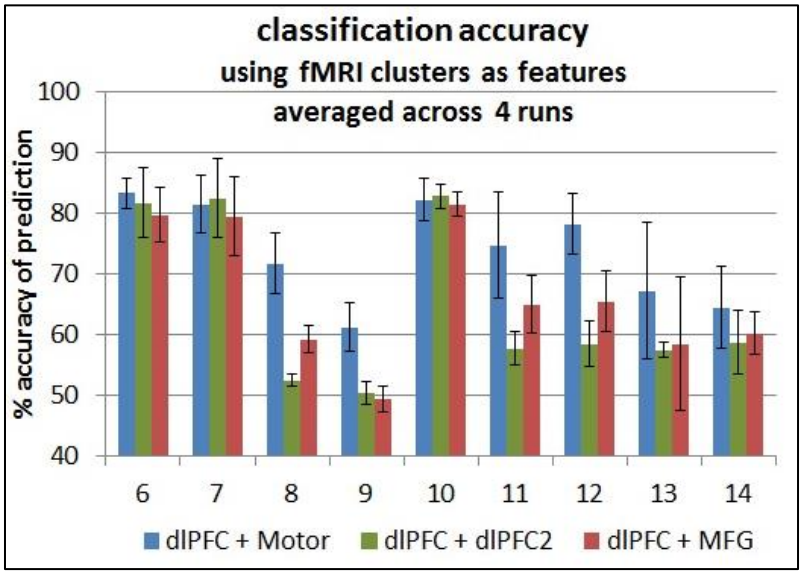


Figure 10. Classification accuracy for study part one fMRI data

Classification accuracy averaged across four runs for each of seven participants. Two fMRI time traces were used as support vector machine input features in each case. Accuracy is the number of time points for which the prediction matched the truth label out of all 200 volumes. An accuracy of 50% represents prediction by chance. Error bars represent ±one standard deviation. Left-hand bars: co-activating pairs (blue). Center bars: within-network pairs (green). Right-hand bars: across-region pairs (red).

Also consistent with the fNIRS results, the group-averaged correlations for co-activating ($r = 0.53$, SD = 0.23) and within-network pairs ($r = 0.54$, SD = 0.26) were significantly greater than zero (within: $t(6) = 6.07$, $p < 0.0005$; co-activating: $t(6) = 5.49$, $p < 0.005$). In contrast, those for the across-network pairs ($r = 0.17$, SD = 0.25) did not differ from zero ($t(6) = 1.81$, $p < 0.1$). As predicted, however, they were significantly lower than those for the within-network pairs ($t(6) = 2.29$, $p < 0.05$; see Table 2 for a participant-specific list of correlation values).

3.6 Discussion

In part one of the present study, we investigated whether functional neuroimaging methods (i.e., fNIRS and fMRI) can be employed to distinguish periods of task engagement from periods of rest. As described below, our findings support this view. They also provide valuable information about which brain activations may prove most useful for monitoring task engagement in the field.

3.6.1 Findings of part one of the present study

Our first set of findings came from the fNIRS experiment. Here, we found that multivariate pattern classification techniques could distinguish between periods of task performance and periods of rest based on brain activity recorded from (a) different regions of the DLPFC in the task-positive network (a within-network pair) or (b) the DLPFC in the task-positive network and the MFG in the task-negative network (an across-network pair). Further, there was a trend toward higher classification accuracy for across-network pairs than for within-network pairs. Indeed, accuracy with across-network pairs approached 70%, even with the basic processing methods described here (with adaptive physiological filtering and additional probes, accuracy may further improve). This result fits with prior data suggesting that variability in task engagement is associated with variability in activity and/or functional connectivity involving both the task-positive and the task-negative networks (e.g., Prado and Weissman, 2011; Weissman et al., 2006). Most important, our fNIRS findings indicate that online recordings of brain activity via fNIRS may provide a valuable tool for detecting varying levels of task engagement in the real world.

Our second set of findings came from an fMRI study. Of importance, these findings both verified and extended the results of the fNIRS study discussed earlier. First, we observed activations and deactivations, respectively, in the DLPFC and the MFG, which verified that our functional neuroimaging paradigm engaged the task-positive and task-negative networks. Second, further analyses revealed that these activations occurred in the same DLPFC and MFG regions that were activated in the fNIRS experiment, wherein activity was measured with probes

on the scalp. Third, the trend toward higher classification accuracy for across-network pairs than for within-network pairs observed with fNIRS was also observed with fMRI. Fourth, the fMRI findings indicated a possible upper bound on classification accuracy for distinguishing between task performance and rest: as expected, the highest classification accuracy was observed when the DLPFC trace and co-activating (for this task) motor cortex trace served as inputs to the SVM classifier. Together, these fMRI findings verified that our fNIRS recordings reflected activity in the key regions under investigation (DLPFC and MFG). They also replicated the fNIRS results and extended them by suggesting an upper bound for classification accuracy based on relatively focal hemodynamic activity.

3.6.2 Novel contributions of part one of the present study

The present findings make an important contribution to the field. Specifically, they show, for the first time, that it is possible to detect negative correlations between activity in key regions of the "task-positive" and "task-negative" networks with fNIRS. Further, they show that the detection of activity in the "task-negative" network is useful for distinguishing between high and low levels of task engagement. This capability might prove useful in future applications of fNIRS that are aimed at discriminating between optimal behavioral performance (where a negative correlation is expected) and internally-guided thought (where co-activation and, hence, a positive correlation is expected) (Christoff et al., 2009; Smallwood et al., 2012). Thus, it could function to improve the predictive power of a fNIRS-based cognitive state monitoring system. Our novel findings build on previous work showing that frontal oxygenation is sensitive to workload (Izzetoglu et al., 2004) and that fNIRS can reliably detect both resting state physiology and functionally-connected networks (Mehnert et al., 2009; Mesquita et al., 2010; White and Culver, 2009).

3.6.3 Limitations of part one of the present study

While stronger negative correlations were associated with higher classification accuracy for across-network pairs, relative to within-network measures, in both modalities (data shown in Appendix D), negative correlations between DLPFC and MFG activity were not observed in every participant. This lack of consistency may stem from a variety of sources, including non-optimal fNIRS probe localization, variation in participant compliance or strategy, interference

from physiological or motion artifact, variable fMRI voxel selection, and co-activation of key regions in the task-positive and task-negative networks during mind wandering or internally-guided thought (Christoff et al., 2009; Smallwood et al., 2012). Future studies should be conducted to distinguish among these possibilities and to determine which methodologies provide more consistent measures of negatively correlated activity in the task-positive and task-negative networks.

Also regarding the consistency of our measures, classification accuracy varied considerably across runs (see the error bars in Figures 8 and 10). Future studies might therefore be conducted to investigate the source(s) of this variability as well as the impact of other sources of variability (e.g., across-visit, across-participant and across-task) on classification accuracy. Such studies might also investigate the impact of using the known task model to clean the measured traces when producing functional task signals for use in classification (see fNIRS data processing), which may have biased the classifier toward higher accuracy in the present study. We explore this topic in Chapter 4 by considering classification accuracy using feature traces after various pre-processing treatments that do not use such knowledge.

Another study limitation stems from the fact that some task-evoked systemic signals are measureable on the scalp surface, and that at least one such signal - skin blood volume - depends on cognitive state (Kirilina at al., 2012). Since fNIRS is sensitive to hemodynamics in superficial tissue at all source-detector separation distances, it is possible that signals in the superficial tissue due to drainage or systemic confounds may have driven the negative across-network correlations that we observed. To investigate this possibility, we quantified correlations for the traces taken from the across-network shallow source-detector pairs (see fNIRS data acquisition). Of importance, no association was observed between the correlation values for the superficial traces and the correlation values for the deep traces, whether corrected or not (data shown in Appendix E). Thus, the negative correlations that we measured between the across-network deep traces likely reflected [Hb] related changes in brain tissue rather than superficial physiological signals. However, if skin blood changes provide additional information about the task engagement, then it could be useful in future studies to include such changes directly as classifier inputs. We explore this topic further in Chapter 4 by considering classification accuracy using the shallow physiological signals themselves.

Finally, we note that motor activation is not always a reliable component of task engagement as some tasks require sustained attention over long periods in the absence of overt responses (e.g., instrument cross-checking and visual display searches). Thus, in part two we focus on our fNIRS finding that classification accuracy was slightly higher for across-network pairs (i.e., pairs in which one region came from the task-positive network while the other came from the task-negative network) than for within-network pairs (i.e., pairs in which both regions came from the task-positive network). As we mentioned earlier, this finding fits with the view that task engagement is determined by interactions between these networks (Fox et al., 2005; Kelly et al., 2008; Weissman et al., 2006; Prado and Weissman, 2011).

3.6.4 Future work

Future work could address whether adaptive filtering of fNIRS traces over smaller time windows improves the activation-based classification measures reported here. This may include investigating adaptive physiological noise removal driven by the correlation between the deep and shallow traces (Harrivel et al., 2012a) after proper consideration of task-evoked effects that may be present in the superficial tissue, or motion artifact reduction based on the frequency domain phase signal (Harrivel and Hearn, 2012). An increase in probe density could also be used to improve localization within the regions of interest after proper consideration of the tradeoffs discussed in section 2.1.2. Finally, measures of network correlation could be quantified over shorter time scales (Thompson et al., 2013) to determine whether transient internally-guided thought can be distinguished from periods of "zoning out." The simultaneous fNIRS and fMRI trials of part two of the study further examine these and other possible methods for improving our ability to discriminate between varying levels of task engagement.

Chapter 4

Explorations in the Prediction of Task Engagement with fNIRS and fMRI

In part two of the study, we wished to perform simultaneous fNIRS and fMRI data collection to facilitate a controlled comparison of attentional state prediction based on the same spatial patterns of hemodynamic activity. fNIRS and fMRI measure the same hemodynamic response, but filtered differently in time and space (Huppert et al., 2006). Patterns of these activations may be influenced by the physiological state of the participant (such as fatigue), hemodynamic changes due to physiological variations in the cortex, participant motivation, effort and response strategy (Sarter et al., 2006).

We used the same MSIT as in part one to induce relatively high and relatively low levels of external, goal-oriented task engagement. Additionally, both fNIRS and fMRI time traces were again used (separately) as input features to SVM classifiers. For a more in-depth exploration of the prediction of task engagement, we built on the results of part one and explored multiple fNIRS data processing methods. These employed both static (sections 4.3.2 and 4.4.2) and adaptive (sections 4.3.4 and 4.4.4) methods to remove noise and artifact. We assessed the impact of each on average state prediction accuracy and present the results in Table 3. This exploration also included more realistic implementations which do not rely upon known task activations, as discussed in section 3.6.3, to clean the traces prior to their use as input features (section 4.4.1 and 4.4.5). To further consider the practical implications of our methods, we investigated the effect of reducing the amount of training data (section 4.4.6). Additionally, because task-evoked effects can be observed in physiological changes in superficial tissue, we determined the prediction accuracy based on shallow fNIRS traces (as introduced in 2.1.3) (section 4.4.3). Finally, we predicted task engagement using fMRI features taken from regions selected similarly to that done in part one (section 3.5.1), from regions local to the estimated sensitivity of the fNIRS probes, and from the whole brain. We present these methods in section 4.5 and the results in section 4.6.

4.1 fNIRS Probes and Locations

In part two of the study, the probes were placed as in part one (as in sections 2.1.4 and 3.1), except that the line toward Fz was also marked to orient the DLPFC probe array. Also, because optodes 1, 2 and 3 were the most informative in part one, only 4 optodes were used in the DLPFC probe array for part two. The placement is shown in Figure 11. The four sources were placed with 45° between them as can be seen in Figure 12, top. This array of 4 optodes was rotated upon application to each participant such that the mark for Fz fell between optodes 2 and 3. The same two optodes used in part one at the MFG were duplicated for part two, for a total of six optodes and thus twelve [Hb] time traces ([HbO] and [HbR] per optode). The MFG probe array is shown in Figure 12.

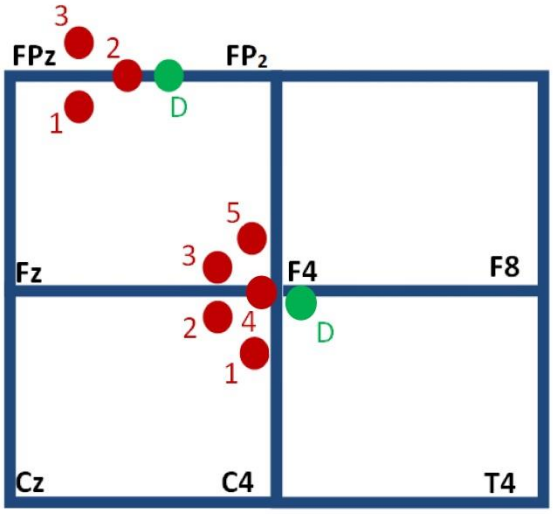


Figure 11. Head probes used to interrogate the DLPFC and MFG in part two of the study
Source channels (red) are arranged at a separation distance of 3 cm from two detectors (green), shown schematically with respect to the I 10-20 locations. The line toward Fz fell between optode channels 2 and 3. The shallow sources, located at 1 cm from the detectors, provide channel 4 for the DLPFC array, which is near F4, and channel 2 for the MFG array, which is between FPz and FP2.

The headgear application process was as follows. The probe arrays were placed on the participants as they lay supine on the scanner bed with their head in position for the receiver coil. Participants were instructed at this time to remain still through the end of data collection. The Velcro straps were adjusted for each source probe prior to the final application of padding and

attachment of the top half of the head receiver coil. Intensities were checked in an iterative fashion as described in section 2.1.4. If adjustments were required, the head coil was removed to facilitate the adjustment of the hair, the addition or removal of attenuating material over the shallow source probes, or the reattachment of probes such that they contacted the skin. Once signals were satisfactory, the participant was translated into the bore of the scanner. This was a lengthy dressing process inappropriate for operational situations. Participants tolerated indentations in the skin due to the probes. However, multiple indentations per probe were not seen upon removal, indicating the probes remained stationary.

The headgear performed well and was compatible with all aspects of the fMRI laboratory environment. We built fMRI-compatible probes using a combination of detector fiber bundles from ISS, Inc. and customized source fibers (Harrivel et al., 2009; Mackey et al., 2012). The source fibers were 400 micron core glass fibers and each detector was a 3mm diameter glass fiber bundle. These employed glass right-angle prisms in periscope assemblies to facilitate placement within the space limitations of the fMRI head receiver coil. Source fibers were held in place with a nylon set screw within the periscopes, which were machined out of black Delrin. Additionally, 3 mm diameter light pipes (15 mm in length and bonded to the other face of the prisms) provided offset for hair penetration and visual inspection of mark locations. Light pipe baffling all the way to the skin surface protected the measurements from ambient light exposure. The light pipes fit snugly into holes punched in hard rubber array holders such as that shown empty in Figure 6. The array holders were designed such that each source-light pipe assembly could be individually adjusted with its own Velcro strap to allow for an individualized fit. The detector's light pipe was inserted in the central hole. Prototype probe sets were scanned and any fMRI-artifact-generating material was replaced. A vitamin E marker was attached next to each detector for use in part two as discussed in section 4.5.2.

Figure 12. DLPFC head probe array for part two of the study

Prisms and light pipes can be seen (lower). A thin gold coating can be seen on the central detector housing which was later removed. Each optode has a separate Velcro attachment.

Figure 13. MFG head probe array for part two of the study

4.2 Behavioral Methods and Results

For part two, there were seven participants (3 female, 4 male) who completed four runs each of the MSIT during simultaneous fNIRS and fMRI data acquisition. Except for the addition of the calculation of the coefficient of variation of reaction time (CVRT), behavioral methods were identical to those for part one (section 2.3). All participants performed the MSIT correctly for all trials. As expected, and in line with part one results (section 3.4), mean behavioral accuracy was relatively high at 97.6% (SD = 2.3%, N = 7).

As discussed in section 2.2, associations between behavioral performance and activity in the TPN and TNN have been reported (Fox et al., 2005; Kelly et al., 2008; Weissman et al., 2006; Prado and Weissman, 2011). We wished to assess the association, if any, between behavioral performance throughout the runs and across-network time trace correlations (sections 4.4.9 and 4.5.4), and between behavior and the low-frequency spectral content of the fNIRS and fMRI signals (section 4.6.4). Thus, CVRT was calculated as the standard deviation of reaction time divided by the mean reaction time, each over 8 trials. Twelve CVRT values were determined per run. Mean CVRT was 0.240 (SD = 0.225%, N = 7).

Also as expected, performance was worse in incongruent than in congruent trials. In a random effects analysis, mean reaction time was significantly higher in the incongruent condition (M = 0.761 s, SD = 0.126 s, N = 7) than in the congruent condition (M = 0.680 s, SD = 0.106 s, N = 7), ($t(6) = 7.6$, $p < 0.0005$). Likewise, mean error rate was significantly higher in the incongruent condition (M = 3.91%, SD = 3.03%, N = 7) than in the congruent condition (M = 1.43%, SD = 2.25%, N = 7), ($t(6) = 3.3$, $p < 0.01$). Errors of omission were rare and thus not analyzed. One subject committed a maximum of three per run while all other subjects committed two or fewer per run. Each run contained 96 trials.

4.3 fNIRS Processing Methods

4.3.1 Classification Methods

In part one of the study, two time trace features were taken, one from each probe array, from the optode which produced the [ΔHb] trace which best fit the modeled task response. In part two, to take advantage of additional channels of presumably informative signals, six time trace features were used. These were selected from the total of twelve (six optodes by two [Hb] species each) available across both probe arrays based on how well the static functional trace from that optode-species combination discriminated the task in the training data. In all cases, data used for prediction was completely separate from training data used to create the SVM model. Also in all cases, training data were processed in the same manner as that used for data to be predicted. When the shallow traces were used as input features, all four available [Hb] time traces were used (sections 4.3.3 and 4.4.3).

Time trace features were selected according to a discrimination score, F(j) (Chang and Lin, 2008). The score considers each training feature independently. For i time instances and j features:

$$F(j) \equiv \frac{\left(\bar{x}_j^{(+)} - \bar{x}_j\right)^2 + \left(\bar{x}_j^{(-)} - \bar{x}_j\right)^2}{\frac{1}{n_+ - 1}\sum_{i=1}^{n_+}\left(x_{i,j}^{(+)} - \bar{x}_j^{(+)}\right)^2 + \frac{1}{n_- - 1}\sum_{i=1}^{n_-}\left(x_{i,j}^{(-)} - \bar{x}_j^{(-)}\right)^2} \tag{3}$$

where $^{(+)}$ indicates one class and $^{(-)}$ indicates the other class (in a binary state classifier as used throughout this study). The numerator quantifies inter-class variance. The greater the difference between: the mean of all instances with truth labels for the (+) class (e.g. indicating an engaged state) minus the mean of all instances, and the mean of all instances with truth labels for the (—) class (e.g. indicating a rest state) minus the mean of all instances, the better the state discrimination. The denominator is the sum of the variance in the (+) class and the variance in the (—) class. Less variance within each class indicates better discrimination. This method of

quantifying state discrimination for the purposes of down-selecting uninformative candidate features is simple and effective (Chang and Lin, 2008).

Ranking according to this discrimination score was performed across all training runs used. In other words, we did not restrict the selection, for example, to four from the DLPFC array and two from the MFG array. The six optode-species combinations with the greatest discrimination scores were selected. Then, the same optode-species combinations were used for both training and predicting (see section 2.4 for additional classification methods). For that participant and that set of training data, the other time traces were disregarded. This ranking process was performed for each case of different classifier input *feature types* described in sections 4.4.1, 4.4.2, 4.3.4, 4.3.5 and 4.4.6.

4.3.2 Static Regression

Standard linear regression was then used to remove physiological contributions from the minimally-processed (that is, temporally filtered [ΔHb]) *as-measured* signal to produce the *functional* task signal as described in section 2.1.4.2. To differentiate this from the dynamic filtering to be introduced in section 4.3.4, we call this *static* regression. For part two of the study, filtering was set to include 0.008 Hz to 0.1 Hz to focus on sustained task activations while removing very slow drift and higher frequency physiological and motion contributions and while retaining more of the frequency components below 0.1 Hz which contribute to our sustained task and resting state hemodynamic signals of interest. Functional traces to be used as inputs to the classifier are generated in the same manner as part one of the study. That is, the functional trace is the measured trace minus the nuisance components determined by the GLM fit, as described in section 2.1.4.2.

Additionally, the shallow trace removal was not performed within-species. That is, the oxygenated *and* deoxygenated [Hb] traces from the shallow channel were regressed from the measured traces of each Hb species of each deep channel on that detector's array. This is appropriate because shallow and deep traces can be produced by optodes sampling different mixes of arterial and venous compartments. A probe's sensitivity to one species or the other depends on whether it mostly samples the arterial or venous compartment (Strangman et al., 2003), which is likely to change between probe applications due to changes in probe location. Additionally, physiological interference has been shown to be inhomogeneous in space (Gagnon

et al., 2012b). Given that the shallow and deep probes sample nearby but different locations, they likely do not sample the same mix of arterial and venous compartments. The measured traces thus may contain nuisance contributions stemming from variations in either species of [Hb] to varying degrees across locations. Therefore, we did not restrict this method to within-species correction only, as was done in part one. We hypothesized that both the beta fit parameters for these two regressors would be relatively high independent of the species of the measured trace being cleaned.

4.3.3 Physiological Traces as Classifier Input Features

Typically, systemic physiological confounds must be removed to avoid false positive results (Tachtsidis et al., 2009), as we have done in both parts of the study as described in sections 2.1.3, 3.2.1 and 4.4.2. However, if physiology detected in the shallow signal may be due to task engagement, then these signals may be useful on their own as classifier input features. Indeed, task-evoked changes in systemic blood pressure and autonomic nervous system responses do manifest in the skin layer, to which fNIRS is highly sensitive (Boas et al., 2004; Kirilina et al., 2012). For example, systolic blood pressure has been shown to increase and peak with task demand, and then decreases once the task is too challenging (Fairclough et al., 2013). To test the possibility of predicting task engagement using physiological fNIRS traces alone, the physiological traces themselves were used as classifier input features. These were the as-measured shallow traces, without the first regression step as described in section 2.1.3. In this case, all four available [ΔHb] traces were used without ranking or pruning.

4.3.4 Adaptive Regression

4.3.4.1 Dynamic Functional Connectivity and Non-stationary Physiology

To improve the capability of an attentional state prediction system based on our methods, we wished to improve the removal of noise from the measured signals. To do so, we recognize that nuisance physiological contributions to the deep signals of interest are non-stationary in time (Kolehmainen et al., 2003; Diamond et al., 2006; Zhang et al., 2007b; Zhang et al., 2009; Abdelnour and Huppert, 2009). Thus, static regression and beta fit parameters which are not allowed to change in time is a limitation. Given a good model of the expected task-related

activations, it is possible to better identify dynamic noise in the signals for removal at each time instance.

Further, we recognize that functional connectivity of the TPN and TNN is also dynamic. Indeed, short time windows have been used to assess correlation of fMRI BOLD series in these networks (Thompson et al., 2013), as anti-correlated network activations are variable in time and frequency, whether due to noise or variations in cognitive state, on the scale of seconds to minutes (Chang and Glover 2010; Deco et al., 2011; Majeed et al., 2011). Also, the temporal synchronicity of functionally connected networks, which would have to be variable to function as such, has been proposed as an integrating mechanism for introspective and extrospective functions (Fransson et al., 2006).

4.3.4.2 Dynamic Regression Using the Kalman Filter

We implemented a discrete Kalman filter (Kalman, 1960) to improve the removal of noise from our measured [Hb] signals. Kalman filtering is used to estimate, with minimum error, the state of a process that can be well represented by a linear model. Its recursive nature makes it attractive for real time autonomous applications (Welch and Bishop, 2006).

Kalman filters previously have been implemented for processing fNIRS data. A state space model was used with a Kalman filter to estimate simulated responses added to real auxiliary physiological (noise) data collected during rest, and showed improvement of analysis with physiological modeling (Diamond et al., 2006). An adaptive GLM was implemented with a Kalman filter to classify right vs. left single-trial finger tapping after removing physiological noise from real response signals by modeling the cardiac signal, respiration and Mayer waves (Abdelnour and Huppert, 2009). A GLM was also implemented with a Kalman filter using simulated responses and real data from short source-detector separation optodes, showing time point adaption is best for signal cleaning (Gagnon et al., 2011). Finally, a tapped delay filter (with coefficients optimized on a sample-by-sample basis and [HbR] filtered separately) (Zhang et al., 2007b) has been used to recover response to a visual task with real responses and real data from short-separation optodes. Zhang, et al. showed signals with poor signal to noise ratio do not benefit from cleaning, that cleaning can degrade the signal if it is not dominated by "global interference" or physiological contributions, and saw contrast improvement for [HbO] only (Zhang et al., 2009). A recent example of a similar variable regression technique applied to fMRI

data is given in Kang et al. (2011), and a channel-by-channel adaptive GLM was implemented with fNIRS data by Aqil, et al. (2012) using a recursive least squares estimation method to recover response to a finger tapping task. The current study applies an adaptive Kalman filter using real fNIRS data for both the noise and the task response to be recovered, and extends the application of adaptive filtering methods beyond robust tapping and visual tasks to a cognitive attention task.

We implement an extension to the work of Abdelnour and Huppert (2009) by using real short-separation optodes with an adaptive GLM. Further, as explained in section 4.3.2, we do not filter Hb species separately because of inhomogeneous sampling of the vasculature. In the filter, the current state is set to the prior state plus process noise (with covariance Q), with transition matrix set to identity and no driving input. We consider the process noise to be strongly related to the shallow physiological signals. The measurement is set to the observation matrix times the state plus measurement noise (with covariance R). We consider the measurement noise to be strongly related to the difference of the measurement from the expected task activity model. Stated in GLM terms, the observation matrix is the design matrix, and the state is a vector of beta fit parameters. The measured signal, as in the static case, is the as-measured trace (filtered to include 0.008 Hz to 0.1 Hz). The observation matrix is four columns wide and the state vector has four components. The columns are populated with four regressors, the last three of which are nuisance regressors. These four are the expected task activity, the shallow [HbO] trace, the shallow [HbR] trace, and constants to model DC (or constant) offset. Functional traces to be used as inputs to the classifier are generated in the same manner as part one of the study. That is, the functional trace is the measured trace minus the nuisance components determined by the GLM fit, as described in section 2.1.4.2.

Each time instance, the new state estimate (new values for the beta parameters) is determined based on the prior state and the Kalman gain times the difference between the measurement value and the predicted value based on the observation matrix. The Kalman gain is updated prior to being used for the next state estimation. This adjustment depends on the measurement noise (which drives the gain down if it is large) and on the process noise (which drives the gain up if it is large). The gain determines how much influence the measurement has on the state estimate, depending on the prediction determined via the observation matrix. If the process noise is high, the gain is high and the measurement has more influence.

For optimal filter performance, the process noise and measurement noise are assumed to be independent, normal and white (Welch and Bishop, 2006). These assumptions are not fully met because there is some dependence of the measurement noise on the process noise, so the error minimization may not be optimal. However, the filter is still useful for our purposes in producing cleaned functional traces for use as classifier input features.

Q and R (introduced above) are tuning parameters for the Kalman filter which must be set. A variety of values for Q and R were tested with an idealized response and real physiological measures. The standard error of regression increased with larger R and smaller Q. However, if R is too small, the measurements may be over-trusted and dynamic filtering will not be employed. Likewise, if Q is too large, very large changes will be allowed from time point to time point, and the measurements will be over- tracked. At the other extremes, a very large R does not allow the measurements to be used, and a very small Q approaches the static regression case (Abdelnour and Huppert, 2009). Q and R should give performance that mixes dynamic filtering with balanced use of the measurements.

For our study, Q was set to 0.1 times the variance of the shallow physiological traces, and R was set to 10 times the variance of the difference between the measurement trace and the expected task activity. This was done based on the entire time traces prior to adaptive regression, and allows for some adjustment based on the real data for each participant and each run. We note that, for online processing, these quantities may be based on recent windows of data to base Q and R on real variance in the time-local response rather than on expectations. These values of Q and R allowed for a combination of periods when the measurement was tracked and when it was not tracked. Also, these values avoided state estimate traces which spiked steeply in a manner unexpected for hemodynamic responses.

In Figure 14, sample adaptive regression outputs are shown to demonstrate the effect of changing Q and R. The two output traces are [ΔHbO] shown in red, and [ΔHbR] shown in blue, with their difference ([ΔHbO] minus [ΔHbR]) also shown in magenta. Importantly, the robust tapping response acquired here from an optode located over the left motor cortex is appropriately processed with static regression and the parameters selected for the study (Figure 14, right). This adaptive method further will be applied to the more challenging detection of attentional responses for the current study.

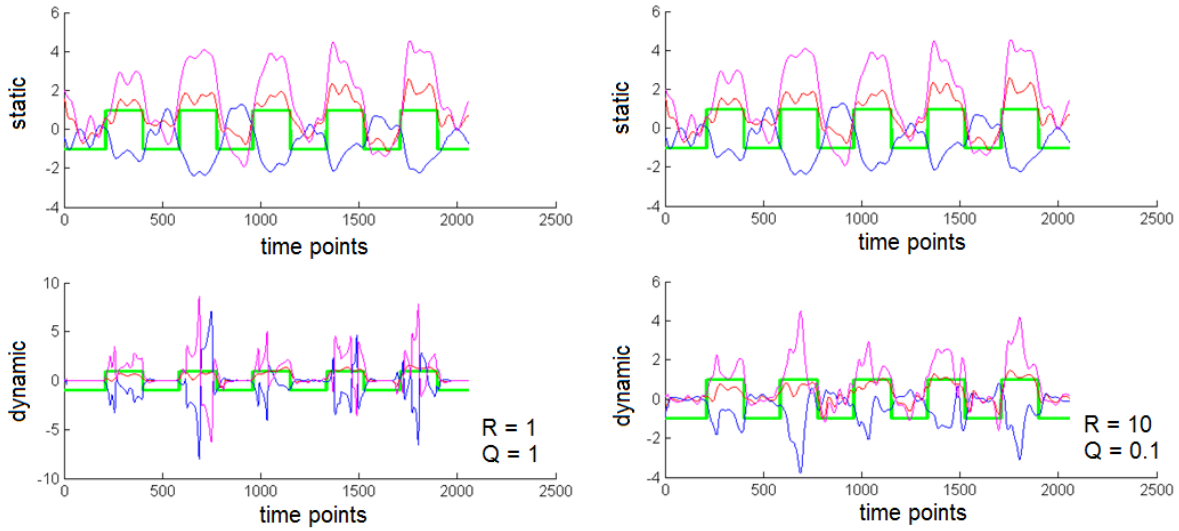


Figure 14. Adaptive regression Kalman filter outputs for two different values of Q and R

The input traces were collected from the left motor cortex during five 30-second epochs (green) of right-hand tapping, and band-pass filtered from 0.008 Hz to 0.1 Hz. The two output traces are [ΔHbO] shown in red, and [ΔHbR] shown in blue, with their difference ([ΔHbO] minus [ΔHbR]) shown in magenta. In the top panel, the static regression result is shown. This is repeated in the top panel on both sides for ease of visual comparison. The lower panel shows the output of the adaptive regression for different values of Q and R. Q on the right is 0.1 times that on the left. Similarly, R on the right is 10 times that on the left.

4.3.4.3 Residual Error and Spectral Analysis

To better understand the noise in the measured signals, we analyzed the standard error of regression (SER) and the spectral content of the residual error traces after both static and adaptive regression. SER was calculated as

$$SER = (\mathbf{e}'\mathbf{e}/d)^{\wedge}(1/2) \quad (4)$$

where **e** is the time trace difference between the measurement and the modeled response, **e**' is the transpose, and d is the degrees of freedom, which was set to the number of time points minus the number of regressors.

We did not correct d for the bandwidth of the as-measured traces. Thus, a corrected value for the degrees of freedom was not used in this assessment. This is an appropriate comparison because the sampling frequency was the same for both the static and adaptive cases. Further, we were mainly interested in quantifying the difference between the as-measured trace and all

contributions explained by our general linear models. SER was determined as described above for the two regions, by two Hb species, six optode channels and four runs.

Also, the spectral content of the residual error traces was quantified for both the static and adaptive regression methods to check for components not effectively removed by our model. In addition to the two residual traces, spectral amplitude peaks were also quantified for the shallow traces (which represent the fNIRS physiological case), the as-measured traces (as described in section 2.1.4.2 as minimally processed, with band-pass filtering from 0.008 Hz to 0.1 Hz in part two of the study). All traces were averaged across the four runs for each participant prior to this analysis. The amplitude spectrum was computed via Fast Fourier Transform (FFT) with 0.00152 Hz resolution. The largest peak in the frequency range of 0.0076 Hz to 0.0168 Hz was identified to investigate the amplitude of very low frequency components observable in the residual traces. These are potentially attributable to intrinsic activity in resting state functionally connected networks (Fox et al., 2007; Biswal et al., 1995) or local vasomotion of the cerebrovascular (Sassaroli et al., 2012), either of which may contribute to signal content in this frequency range. The peak value was recorded for each of the two regions (and for fNIRS for each of the two Hb species), for each participant. The average across species and regions for each participant is presented in section 4.4.4. Additionally, we investigated the association of the average CVRT with the strength of this activity for each participant to identify any association of this behavioral metric (Fox et al., 2005; Kelly et al., 2008; Weissman et al., 2006; Prado and Weissman, 2011) with very low frequency activity that may be attributable to the TPN or TNN.

4.3.5 Physiology-cleaned Traces as Classifier Input Features, Static and Dynamic

In the static and adaptive regression methods described above, we rely upon our knowledge of the task being performed by the participant to produce functional task signals cleaned of physiological noise and other artifact for use in classification. In real world settings, the knowledge of that task model generally will not be available. Thus, to further explore the real world capability of an attentional state prediction system based on our methods, we explore static and adaptive regression using nuisance regressors alone. We call these "*physiology-cleaned only*" cases, and no task knowledge is used in the processing. The functional traces used as classifier input features have been generated in the same manner as described in sections 4.3.2 and 4.3.4, except the known task indicator function is not used to produce regressors of expected

activity. Thus, three beta fit parameters are used in these cases as opposed to four. This case should not be confused with that of section 4.3.3, where the shallow physiological traces themselves are used as the classifier input features.

4.3.6 The Effect of Varying the Amount of Training Data

To further assess the impact of moving from a controlled laboratory to more realistic implementations of an attentional state prediction system based on our methods, we explore the effect of a reduction in training data used on the classification accuracy. This is important because shorter and less frequent training means more efficient and simpler operational use.

For the static, dynamic and as-measured cases, SVM classifier models were developed based on one run's worth of training instances (2500 time points, or approximately 7 min). Each participant performed four runs. Thus, instead of training with three and predicting the fourth as described in section 2.4, each run is predicted three times, once per training run used. For example, run A is predicted using a model trained with run B, then with a model trained with run C, then with a model trained with run D. Three accuracies are produced, and an average is reported as the result for the prediction of run A. The process is repeated for the next permutation, predicting run B by training separately with runs A, C and D, and so forth until all four runs have a reported prediction accuracy.

4.3.7 Features Selected per Network by PCA

To investigate a more data-driven method of selecting classifier input features, we used principal component analysis (PCA) on the functional traces of each probe array instead of pruning as described in section 4.3.1. We reasoned that after processing as described in section 4.3.1, the greatest variance in the functional traces may be due to task effects. Indeed, model-free independent component analysis has been used to find task- and network-related activations in fMRI BOLD signals (Kineses et al., 2008). A plot of principal components used as classifier features is shown in Figure 27. We selected the top four principal components after performing PCA on the eight [Hb] functional traces from the DLPFC region array. Likewise, we selected the top two components after performing PCA on the four [Hb] functional traces from the MFG region array. In this way we matched the number of features used in all the previous functional trace cases (six) for fair comparison across SVMs. These six features were generated for each of

the five processing methods (as-measured, static, adaptive, physiology-cleaned only static, and physiology-cleaned only adaptive).

4.3.8 Features Produced per Network by Multivariate Multi Scale Entropy

Complexity analysis has been applied to fNIRS signals, although studies are currently limited. Neural complexity of a system of fNIRS time series was monitored during mental tasks with inconclusive but promising increases found during active tasks (Çiftçi et al., 2005). Other studies have applied Lyapunov analysis to quantify the chaotic nature of fNIRS signals. Signals measured by fNIRS during imaginary motor tasks were shown to be chaotic (Soe et al., 2008) and useful for the identification of active brain states across experiments (Khoa et al., 2009). Finally, tactile-stimulus evoked fast optical responses in the somatosensory cortex were found to be more chaotic in the active period (Hu et al., 2011). Therefore, we wished to determine the accuracy of engaged state prediction based on a measure of complexity of the fNIRS signals in each network.

Multiscale entropy (MSE) analysis extends traditional complexity measurements to consider the randomness or orderliness of a time series on multiple time scales, is applicable to statistically independent time series of finite length, and appropriately finds minimal complexity for both deterministic and completely unpredictable signals (Costa et al., 2005; Ahmed and Mandic, 2011). MSE has been applied in biomedical areas such as heart rate variation, EEG, and complexity loss with the stress of aging or disease (Costa et al., 2005). However, multi-channel time traces taken from a local region which sample the same process, such as those used in this study, are not necessarily independent and uncorrelated. The multivariate MSE (MMSE) approach of Ahmed and Mandic (2011) extends MSE for use with such multi-channel data. In the technique, delay vector pairs are created using time lags of the signal and are compared for similarity, at coarse and fine time scales, both across and within channels.

In this study, MMSE was computed using the author's freely-available MATLAB functions. A value for MMSE was determined at each time instance using a rolling time window of 5 s, which is roughly half the expected hemodynamic response time, to capture MMSE information pertaining to activation state changes. Six time lags of one time point each were used, and the similarity threshold parameter was set to 0.2 times the standard deviation of the time series (Ahmed and Mandic, 2011).

MMSE traces were generated in this manner for use as classifier input features. The twelve available [Hb] traces were ranked according to the discrimination score described in section 4.3.1. One MMSE trace was calculated using the four top -ranked traces from among the eight traces of the DLPFC probe array. A second MMSE trace was calculated using the two top -ranked traces from among the four traces of the MFG probe array. Thus, two classifier input features are used (one from each network) for each of the five processing methods (as-measured, static, adaptive, physiology-cleaned only static, and physiology-cleaned only adaptive).

4.3.9 Features Produced by Cross Network Rolling Window Correlation

Measurements at one or more locations within each network allow temporal correlations of activations to be used for prediction. This may be useful for classification purposes because the strength of functional connectivity has been shown to be related to behavioral performance in attention-demanding cognitive tasks (Fox et al., 2005; Kelly et al., 2008; Thompson et al., 2013), and less deactivation of the DMN has been shown to interfere with optimal performance on a trial-by-trial basis (Weissman et al., 2006) and to be involved in the recovery from errors (Prado and Weissman, 2011). The TPN and TNN could be more tightly anti-correlated during task engagement with respect to that during rest. Therefore, we wished to determine whether traces of time-windowed across-network anticorrelation, as previously proposed (Harrivel et al., 2011), would be useful as classifier input features.

Pearson's linear correlation coefficient was calculated over a ten second rolling window between the functional traces after both static and adaptive regression. This was done for the top-ranked (according to the discrimination score as described in section 4.3.1) DLPFC trace and the top-ranked MFG trace, and then repeated for the second-ranked traces from each probe array. Indeed, such analysis with a 12.3 s rolling window recently has been applied to fMRI data to study the relationship between anti-correlation and behavior (Thompson et al., 2013). Thus, two rolling window correlation coefficient (RWCC) traces were generated for each case (static and adaptive) for use as classifier input features.

4.4 fNIRS Classification Results

We wished to determine which fNIRS pre-processing methods and classifier feature types would provide the highest accuracy of state prediction. Thus, six pre-processing methods were implemented with four feature types each as described in section 4.3. Processing methods which did not use knowledge of the task were the as-measured (section 4.4.1), physiological (section 4.4.3), and physiology-cleaned only cases (section 4.4.5). Two physiology-cleaned only cases were investigated: static and adaptive.

The resulting mean accuracies and standard deviations (SD) are presented in Table 3 and in the remainder of section 4.4. Statistical significance with respect to chance prediction at 50% for each case is shown in Table 4. All of the feature types produced using adaptive processing and knowledge of the task generated classification accuracies which were significantly greater than chance prediction. Further, all of the functional feature traces generated classification accuracies which were significantly greater than chance prediction regardless of processing method and regardless of the use of task knowledge. One other case in Table 3 reached significance, being the PCA features produced using the static regression processing method. The best c and g parameters for each participant (as described in section 2.4) for the functional and physiological features are given in Table 5. These SVM parameters for the other feature types (PCA, MMSE RWCC) are given in sections 4.4.7, 4.4.8 and 4.4.9 (respectively).

4.4.1 As-measured Traces

Both static and adaptive cases used knowledge of the task for production of the functional traces used as input classifier features. The as-measured traces, as introduced in section 2.1.4, produced the best classification accuracy of all cases for which no information about the task being performed is used. As presented in Table 3, this result was 70.2% +/- 11.3% when averaged across all seven participants. This processing method achieved significance for the functional feature type only ($t(6) = 6.978$, $p<0.0005$). The accuracy for each participant is presented in Figure 15.

Table 3. State prediction accuracy for various fNIRS processing methods and features

Accuracy was averaged across seven participants. The number of classifier features is shown in square brackets. Those significantly greater than chance prediction are marked with daggers.

feature type:	**functional [6]**		**PCA [6]**		**MMSE [6]**		**RWCC [6]**	
processing method:	**% accuracy**	**SD**	**% accuracy**	**SD**	**% accuracy**	**SD**	**% accuracy**	**SD**
as-measured *	70.2 †[4]	11.3	62.6	11.1	56.1	8.3	-	-
static	71.8 †[4]	14.8	65.1 †[1]	12.6	54.1	3.4	52.3	2.9
physiological *	65.1 †[3]	12.3	-	-	-	-	-	-
adaptive	83.8 †[4]	5.7	79.9 †[4]	5.8	77.0 †[4]	7.2	56.0 †[4]	2.4
pco static *	64.7 †[3]	7.4	56.1	7.1	52.5	2.4	-	-
pco adaptive *	62.0 †[1]	9.5	59.2	8.8	59.9	11.0	-	-

* no task knowledge used pco: physiology-cleaned only

†[1] $p<0.01$ †[2] $p<0.001$ †[3] $p<0.005$ †[4] $p<0.0005$

Table 4. Statistical significance of accuracy for fNIRS classification results

Accuracy was averaged across seven participants. One-tailed paired t-tests were used. The number of classifier input features is shown in square brackets.

feature type:	**functional [6]**		**PCA [6]**		**MMSE [6]**		**RWCC [6]**	
processing method:	**t(6)**	**p <**	**t(6)**	**p <**	**t(6)**	**p <**	**t(6)**	**p <**
as-measured *	6.978	0.0005	3.006	0.025	1.928	0.1	-	-
static	6.109	0.0005	3.190	0.01	3.124	0.025	2.102	0.05
physiological *	3.963	0.005	-	-	-	-	-	-
adaptive	20.584	0.0005	13.536	0.0005	9.936	0.0005	6.526	0.0005
pco static *	5.231	0.005	2.275	0.05	2.751	0.025	-	-
pco adaptive *	3.317	0.01	2.760	0.025	2.382	0.025	-	-

* no task knowledge used pco: physiology-cleaned only

Table 5. SVM parameters for fNIRS classification for functional and physiological features

Parameters are shown for various fNIRS processing methods for each of the seven participants.

participant	processing method:	**as-measured**	**static**	**physiological**	**adaptive**	**pco static**	**pco adaptive**
15	**c**	1	1	0.01	1	0.1	1
	g	0.1	0.5	0.1	0.5	0.1	0.5
16	**c**	0.01	0.001	0.001	0.01	0.01	0.001
	g	0.1	0.1	0.1	0.1	0.1	0.1
17	**c**	0.01	0.1	0.01	0.01	0.1	1
	g	0.5	0.1	0.1	0.5	0.5	0.1
18	**c**	0.01	0.01	0.1	0.1	0.001	1
	g	0.1	0.1	0.1	0.5	0.5	0.5
19	**c**	0.01	0.01	0.1	0.1	0.001	0.001
	g	0.1	0.1	0.5	0.5	0.1	0.1
20	**c**	0.01	0.1	0.001	0.1	0.1	0.01
	g	0.1	0.1	0.1	0.5	0.1	0.1
21	**c**	0.001	0.1	1	0.001	0.001	0.01
	g	0.5	0.1	0.5	0.5	0.5	0.1

pco: physiology-cleaned only

4.4.2 Static Regression

The functional traces produced via static regression, as described in section 4.3.2 and as presented in Table 3, generated classification accuracy of 71.8% +/- 14.8% when averaged across all seven participants. Significance was achieved ($t(6) = 6.109$, $p<0.0005$), but with an increase of less than 2% over the as-measured case. The accuracy for each participant is presented in Figure 15. Exemplary plots of prediction data and the corresponding classifier outputs at every time instance for one run are presented in Figure 16 and Figure 17.

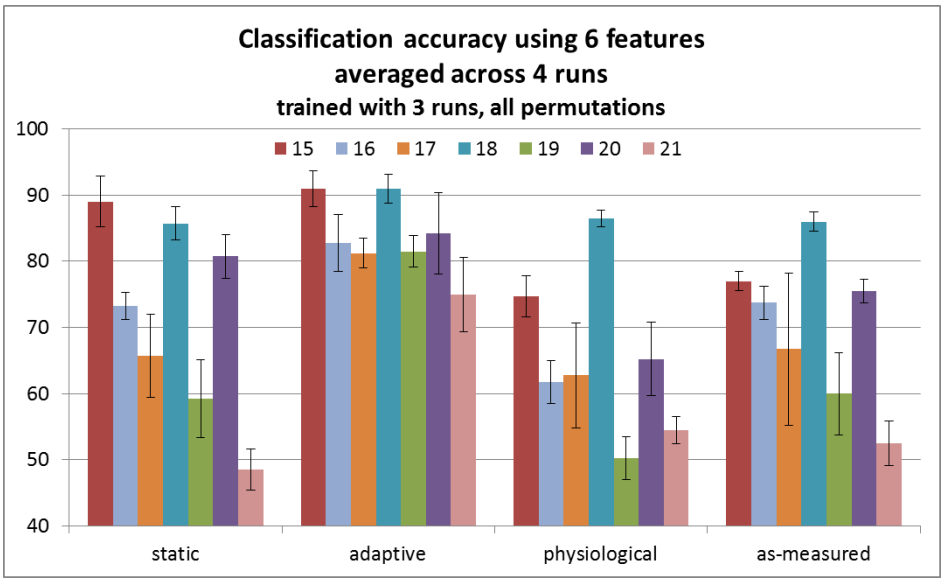


Figure 15. State prediction accuracy for various fNIRS processing methods

Accuracy is shown for cases using functional traces as features, for seven participants. Processing methods which did not use knowledge of the task were the as-measured (of section 4.4.1), and physiological cases (of section 4.4.3).

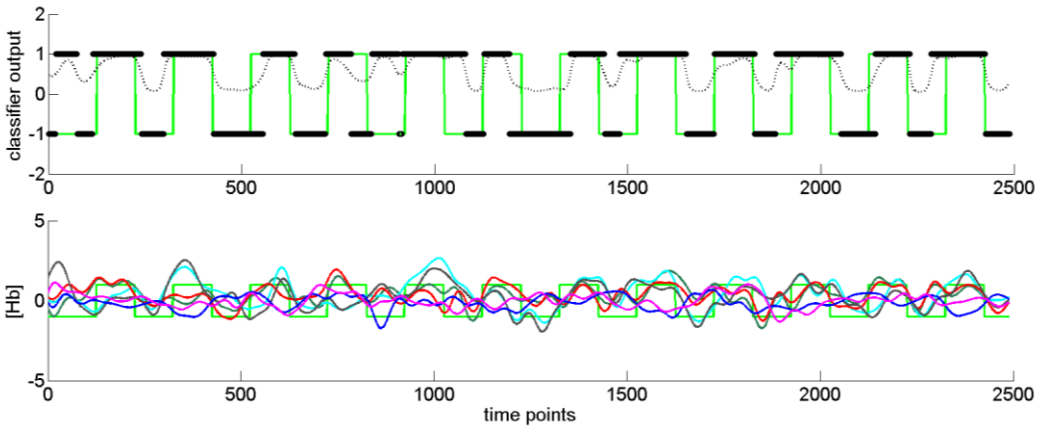


Figure 16. Prediction data and classifier outputs for one run, static case (a)

Prediction accuracy was 74.8% for this run, and three training runs were used. In the top panel, black markers at +1 indicate a classifier output of a more engaged or 'working' state, while those at -1 indicate a less engaged or 'resting' state. A prediction is made at every instance. The dotted black line indicates the probability estimates of the classifier predictions. In the bottom panel, the six classifier input features are plotted. Truth labels are indicated by the green task indicator function in both the top and bottom panels.

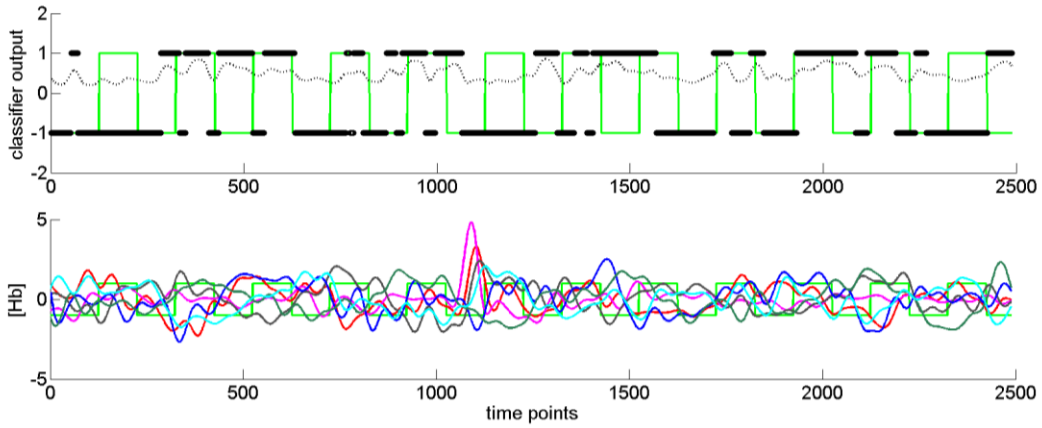


Figure 17. Prediction data and classifier outputs for one run, static case (b)

Prediction accuracy was 50.9% for this run, and three training runs were used. In the top panel, black markers at +1 indicate a classifier output of a more engaged or 'working' state, while those at -1 indicate a less engaged or 'resting' state. A prediction is made at every instance. The dotted black line indicates the probability estimates of the classifier predictions. In the bottom panel, the six classifier input features are plotted. Truth labels are indicated by the green task indicator function in both the top and bottom panels.

We hypothesized that both the beta fit parameters for the two shallow physiological regressors would be relatively high independent of the species of the measured trace being cleaned. Regression was not performed on the shallow trace channels themselves. The beta fit parameters were greater than zero for both species of the shallow nuisance regressors. These values are shown in Figure 18.

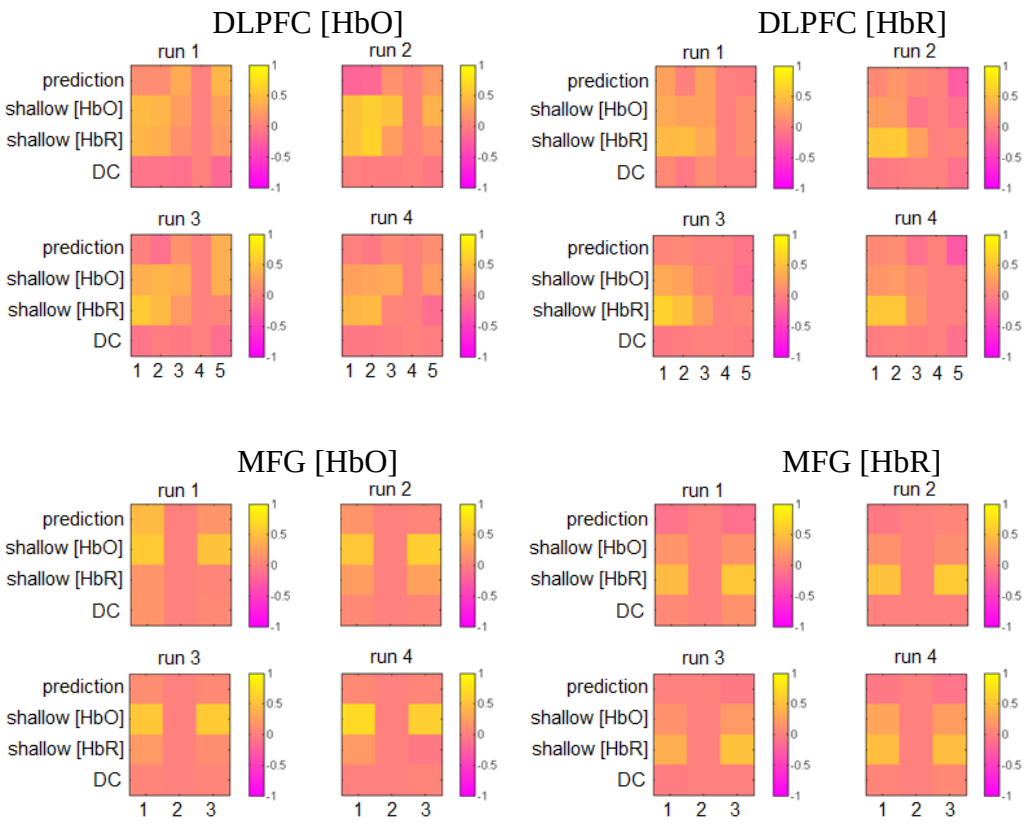


Figure 18. Beta fit parameters for the static regression fNIRS processing

Parameters are given for the two regions, two Hb species, six optode channels and four runs, after averaging across seven participants. DLPFC channel 4 and MFG channel 2 were used for the shallow traces. Regression was not performed on the shallow trace channels.

4.4.3 Physiological Traces as Classifier Input Features

Task-evoked effects in the superficial tissue are clearly real and have been reported elsewhere (Kirilina 2012), but were not found to be reliable across participants. This case, as described in section 4.3.3, generated classification accuracy of 65.1% +/- 12.3% when averaged across all seven participants. This result was significantly greater than chance accuracy ($t(6)=3.963$, $p<0.005$). The accuracy for each participant is presented in Figure 15. The best c and g SVM parameters are given in Table 5.

These physiological traces were produced using no information about the task being performed, and this case represents the second-best accuracy for cases where no task knowledge was used. For one outlying participant, classification accuracy averaged 86.5% +/- 1.3%. The four [Hb] traces and output state predictions at every time point for run 1 are shown in Figure 19.

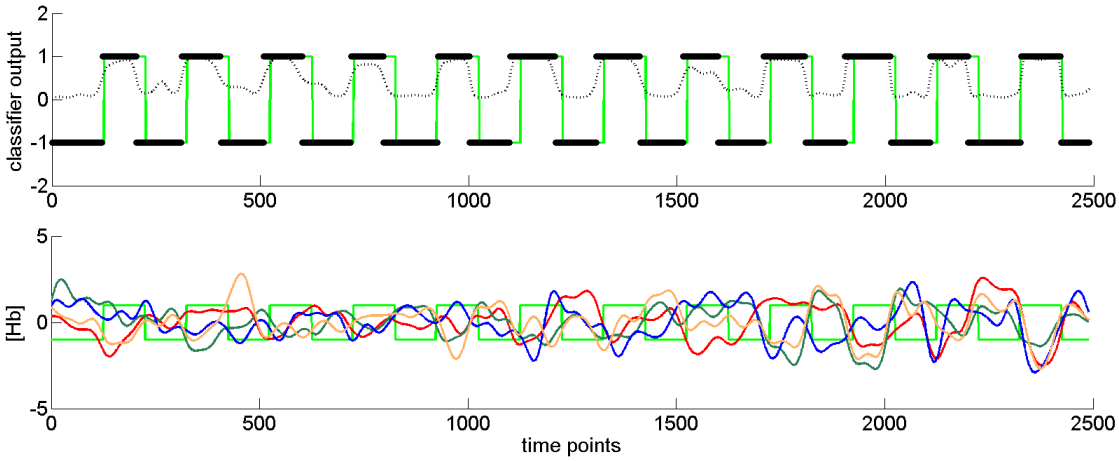


Figure 19. Four shallow physiological traces used as classifier input features

Prediction accuracy was 84.5% for this run, and three training runs were used. Traces for two Hb species per region are shown. In the top panel, black markers at +1 indicate a classifier output of a more engaged or 'working' state, while those at -1 indicate a less engaged or 'resting' state. A prediction is made at every instance. The dotted black line indicates the probability estimates of the classifier predictions. In the bottom panel, the four classifier input features are plotted (red: DLPFC [HbO], orange: MFG [HbO], green: DLPFC [HbR], blue: MFG [HbR]). Truth labels are indicated by the green task indicator function in both the top and bottom panels.

4.4.4 Adaptive Regression and Signal Spectral Content

The functional traces produced via adaptive regression, as described in section 4.3.4 and as presented in Table 3, generated classification accuracy of 83.8% +/- 5.7% when averaged across all seven participants. This case represents the best classification accuracy found in this study. Further, all feature types produced via adaptive regression produced mean accuracies statistically significantly different from chance prediction. The accuracy for each participant is presented in Figure 15. Exemplary plots of prediction data and the corresponding classifier outputs at every time instance for one run are presented in Figure 20 and Figure 21.

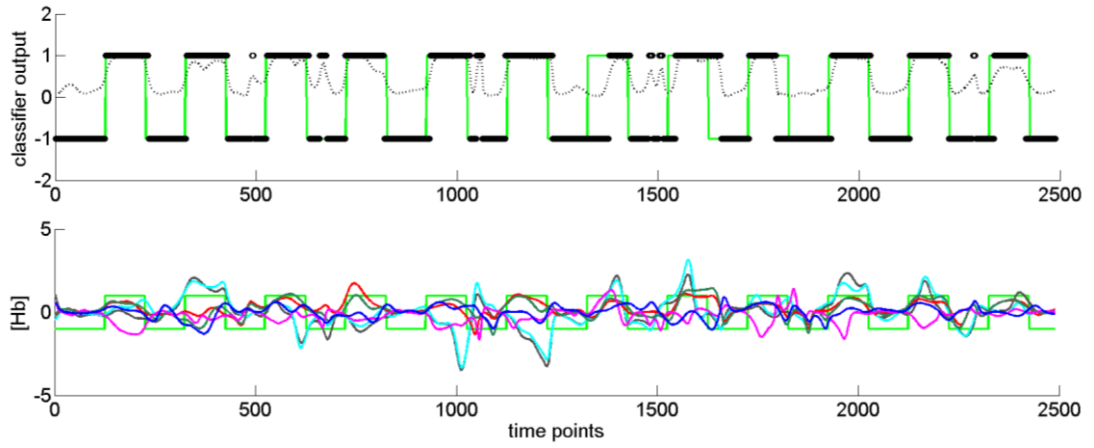


Figure 20. Prediction data and classifier outputs for one run, adaptive case (a)

Prediction accuracy was 87.5% for this run, and three training runs were used. In the top panel, black markers at +1 indicate a classifier output of a more engaged or 'working' state, while those at -1 indicate a less engaged or 'resting' state. A prediction is made at every instance. The dotted black line indicates the probability estimates of the classifier predictions. In the bottom panel, six classifier input features are plotted. Truth labels are indicated by the green task indicator function in both top and bottom panels. This is the same participant and run as in Figure 16.

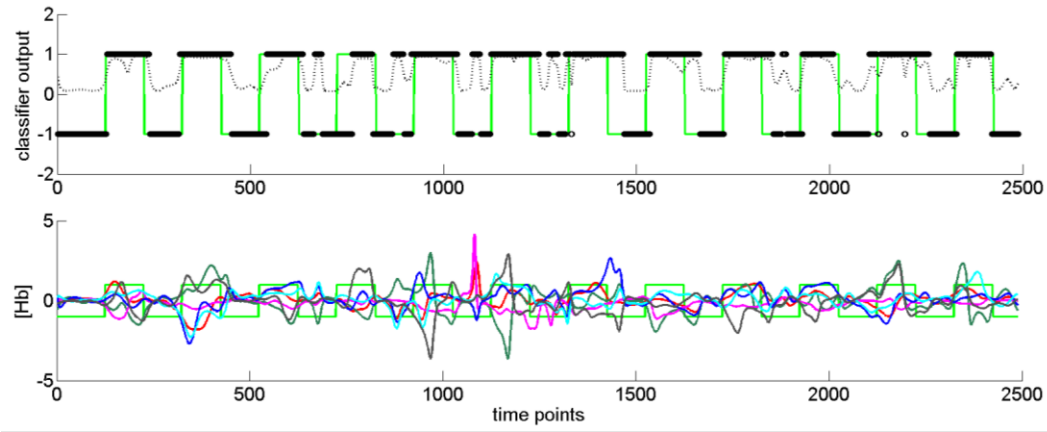


Figure 21. Prediction data and classifier outputs for one run, adaptive case (b)

Prediction accuracy was 79.5% for this run, and three training runs were used. In the top panel, black markers at +1 indicate a classifier output of a more engaged or 'working' state, while those at -1 indicate a less engaged or 'resting' state. A prediction is made at every instance. The dotted black line indicates the probability estimates. In the bottom panel, six classifier input features are plotted. Truth labels are indicated in green. This is the same participant and run as in Figure 17.

In this section we also present the results of the residual error and spectral analysis discussed in section 4.3.4.3. SER is shown in Figure 22 for both the static and adaptive regression fNIRS processing after averaging across seven participants. SER is shown for the two regions, by two Hb species, six optode channels and four runs. DLPFC channel 4 and MFG channel 2 were used for the shallow traces. Regression was not performed on the shallow trace channels themselves. In every case, SER for adaptive regression is less than that for static regression, indicating improved fit to the task and physiological noise model employed.

Peaks in spectral amplitude in the very low frequency range (around 0.01Hz), quantified as described in section 4.3.4.3, were seen in the static and adaptive regression residual traces. Therefore, we wished to determine whether such activity was due to (1) slow oscillations in neuronal activation related to functional network connectivity or intrinsic resting state activity (0.01 – 0.08 Hz) (which may be present before and after the removal of task effects (Fransson et al., 2006; Fair et al., 2007)), (2) very low frequency physiological confounds such as spontaneous slow changes in cerebrovascular tone (0.01 Hz to 0.1 Hz), or (3) some combination. Aside from such signal content in the traces from the short-source-detector optodes, covariates in this frequency range were not specifically accounted for in our models.

The spectral amplitude peak values determined as described in section 4.3.4.3 are shown in Figure 23, normalized to the as-measured case. Very low frequency activity is present in the physiological and as-measured traces, is present after static regression, but is reduced after adaptive regression. The amount of this activity in the static residual traces is significantly greater than that in the adaptive residual traces ($t(6) = 16.0$, $p<0.0005$). There is not a significant difference between the activity in the physiological traces and the as-measured traces, while the activity in the as-measured traces is greater than that in the static residual traces ($t(6) = 6.15$, $p < 0.0005$). Correlations between the CVRT and the activity in the fNIRS adaptive residual ($r(5) = -0.3$) and in the fNIRS static residual ($r(5) = -0.4$) were not significant.

4.4.5 Physiology-Cleaned Only Traces as Classifier Input Features

The functional traces produced via static and adaptive regression, as described in section 4.3.5 and as presented in Table 3, generated classification accuracy of 64.7% +/- 7.4% and

62.0% +/- 9.5%, respectively, when averaged across all seven participants. The accuracy for each participant is presented in Figure 24.

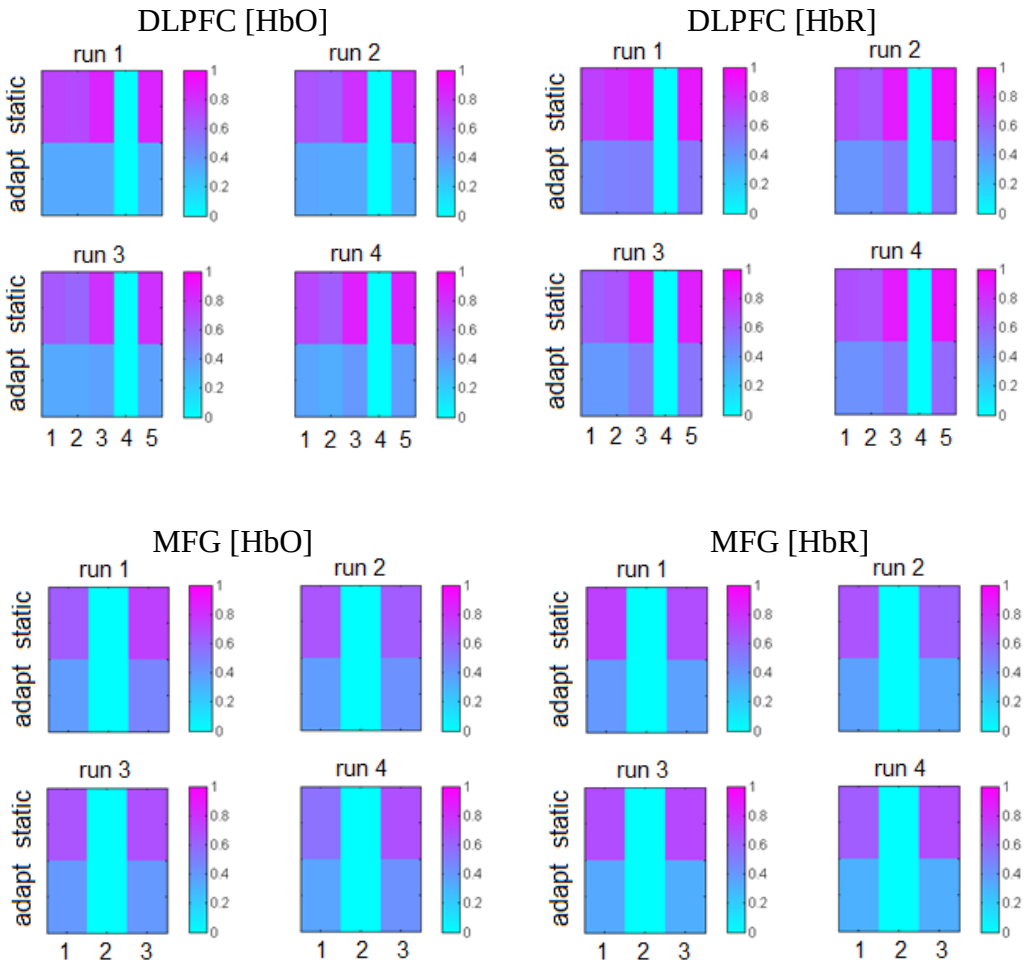


Figure 22. Standard error of regression for static and adaptive regression

Standard error of regression is shown for the static and adaptive regression fNIRS processing for the two regions, two Hb species, six optode channels and four runs, after averaging across 7 participants. DLPFC channel 4 and MFG channel 2 were the shallow traces.

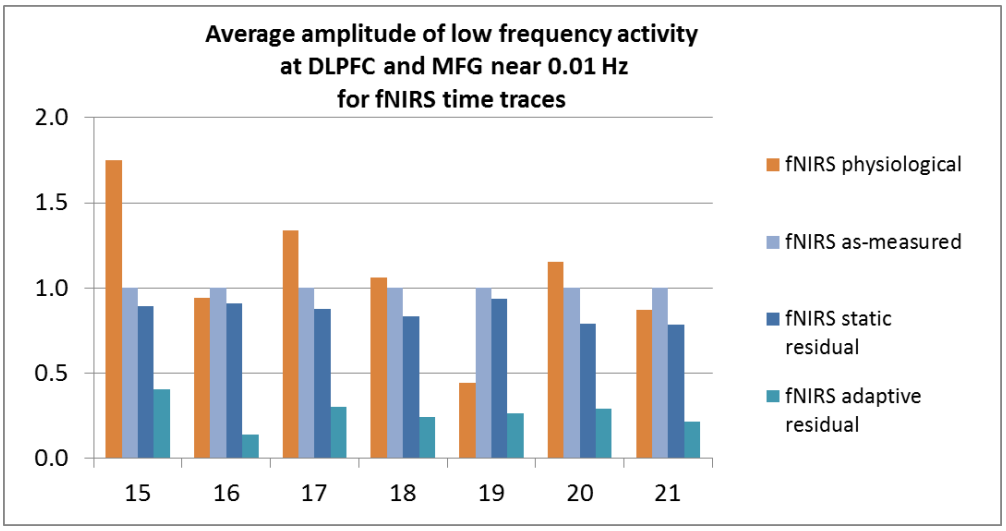


Figure 23. Spectral amplitude in the very low frequency range

Amplitude normalized to the "as measured" unprocessed case is shown in the shallow physiological traces, and in residual traces after static and adaptive regression for fNIRS.

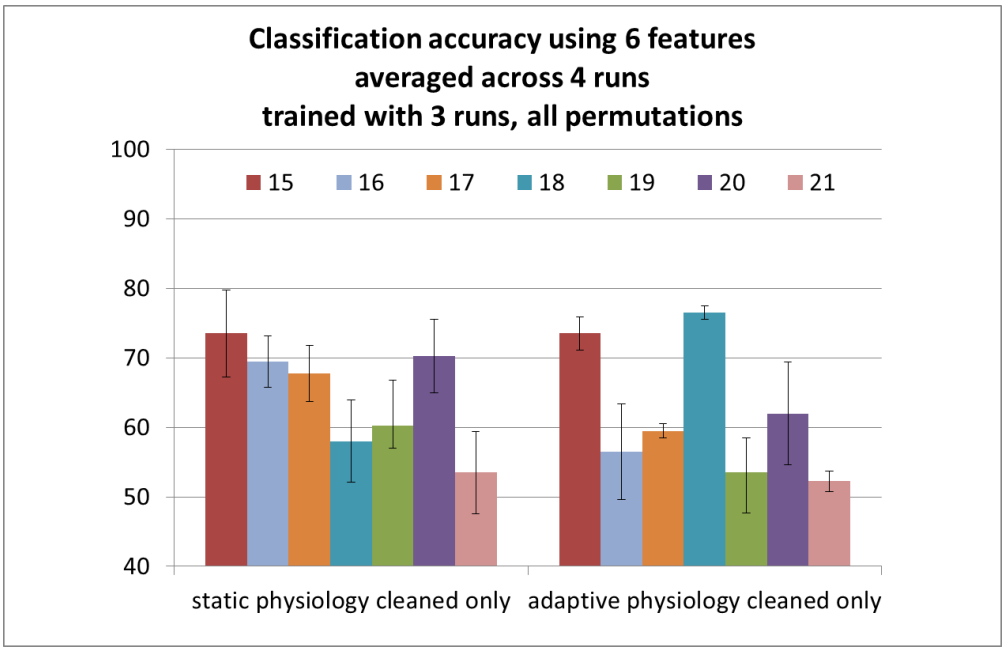


Figure 24. State prediction accuracy for fNIRS functional traces cleaned of physiology only

Accuracy is shown for both the static and dynamic regression cases, for seven participants. These processing physiology-cleaned only methods did not use knowledge of the task. Two cases were investigated: static and adaptive.

4.4.6 The Effect of Varying the Amount of Training Data

The functional traces produced via static and adaptive regression, and the as-measured traces, were classified using SVM models built using only one run of training data as described in section 4.3.6. The classification accuracy was 69.5% +/- 13.4% in the static case, 81.2% +/- 4.6% in the adaptive case, and 67.6% +/- 11.6% in the as-measured case when averaged across all seven participants. The accuracy for each participant is presented in the right-most three sets of bars in Figure 25. Only a few percentage points are lost going from 21 min to 7 min of within-session training data. All cases continued to achieve statistical significance with respect to chance prediction. State prediction accuracy and statistics are presented in Table 6, while SVM parameters for each participant are presented in Table 7.

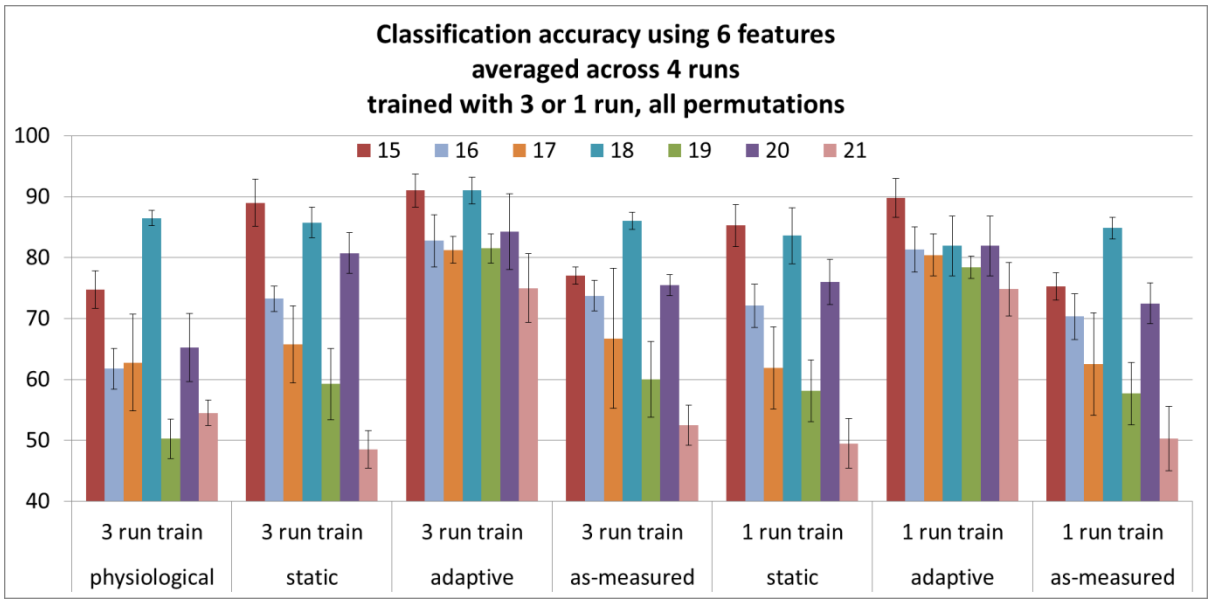


Figure 25. State prediction accuracy with reduced training time

Accuracy is shown for various fNIRS processing methods, for seven participants. The data of Figure 15 are reproduced here to facilitate visual comparison. Processing methods which did not use knowledge of the task were the as-measured and physiological cases. The physiological case was not investigated for 1 run of training data.

Table 6. State prediction accuracy and statistics for fNIRS classification with less training

	functional [6] reduced training time			
processing method:	**% accuracy**	**SD**	**t(6)**	**p <**
as-measured	67.6 †[3]	11.6	4.015	0.005
static	69.5 †[3]	13.4	3.843	0.005
adaptive	81.2 †[4]	4.6	18.146	0.0005

†[1] p<0.01 †[2] p<0.001 †[3] p<0.005 †[4] p<0.0005

Table 7. SVM parameters for fNIRS classification with less training, for each participant

participant	processing method:	**as-measured**	**static**	**adaptive**
16	**c**	0.01	0.001	0.01
	g	0.1	0.1	0.1
17	**c**	0.1	0.001	0.001
	g	0.1	0.1	0.5
18	**c**	0.01	0.01	0.001
	g	0.1	0.1	0.1
19	**c**	0.01	0.01	0.01
	g	0.1	0.1	0.5
20	**c**	0.1	0.001	0.001
	g	0.1	0.1	0.1
21	**c**	0.01	0.001	0.001
	g	0.1	0.1	0.5

4.4.7 Features Selected per Network by PCA

The functional traces produced via static and adaptive regression, via static and adaptive regression when cleaning with the shallow physiological traces only, and the as-measured traces were classified using PCA to select six of the available twelve [Hb] traces as described in section 4.3.7. As presented in Table 3, the classification accuracy was 65.1% +/- 12.6% in the static case, 79.9% +/- 5.8% in the adaptive case, 56.1% +/- 7.1% in the static physiology-cleaned only case, 59.2% +/- 8.8% in the adaptive physiology-cleaned only case, and 62.6% +/- 11.1% in the as-measured case, when averaged across all seven participants. The accuracy for each participant

is presented in Figure 26. Prediction data and classifier outputs for one run are shown in Figure 27. Accuracies achieved using this technique were nearly as high as those obtained via pruning with the discrimination score. SVM parameters for each participant are presented in Table 8. Only the static and adaptive cases achieved statistical significance with respect to chance prediction ($t(6) = 3.190$, $p < 0.01$; $t(6) = 13.536$, $p < 0.0005$, respectively). Statistical significance for each case is presented in Table 4.

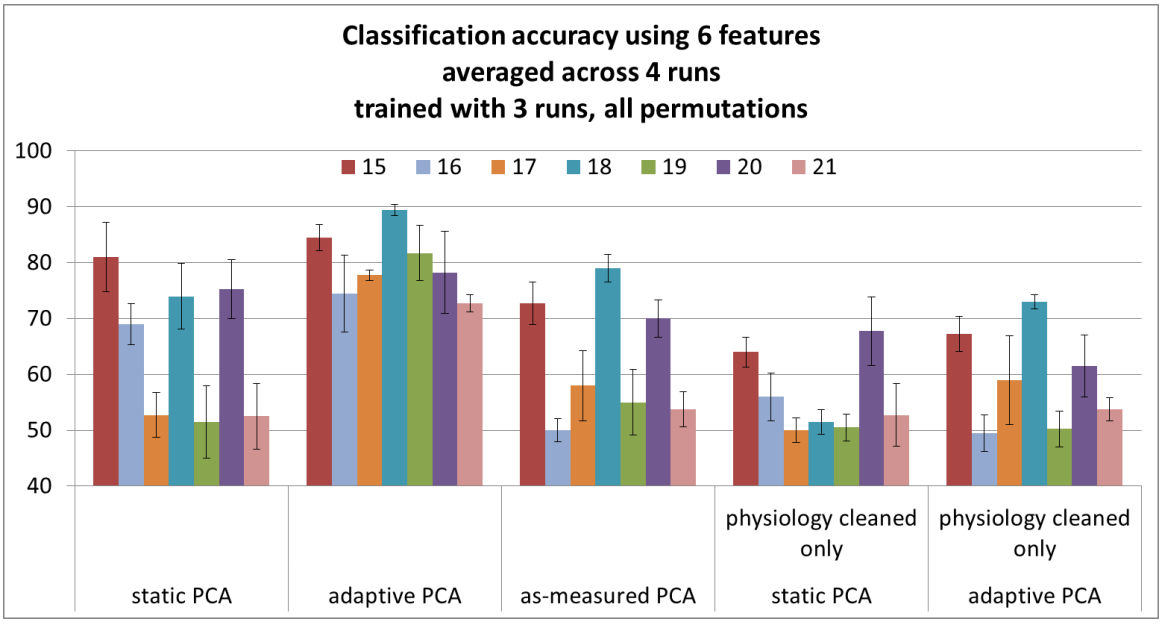


Figure 26. State prediction accuracy using PCA for feature selection

Accuracy is shown for various fNIRS processing methods, for seven participants. PCA processing methods which did not use knowledge of the task were the as-measured and physiology-cleaned only cases. Two physiology-cleaned only cases were investigated: static and adaptive.

4.4.8 Features Produced per Network by Multivariate Multi Scale Entropy

The functional traces produced via static and adaptive regression, and the as-measured traces, were used for creating two classifier feature traces as described in section 4.3.8. The classification accuracy was 54.1% +/- 3.4% in the static case, 77.0% +/- 7.2% in the adaptive case, 56.1% +/- 8.3% in the as-measured case, 52.5% +/- 2.4% in the static physiology-cleaned only case, and 59.9% +/- 11% in the adaptive physiology-cleaned only case, when averaged across all seven participants. Accuracies achieved using this technique achieved significance

with respect to chance prediction only in the adaptive case ($t(6) = 9.936$, $p<0.0005$). Statistical significance for each case is presented in Table 4. SVM parameters for each participant are presented in Table 9.

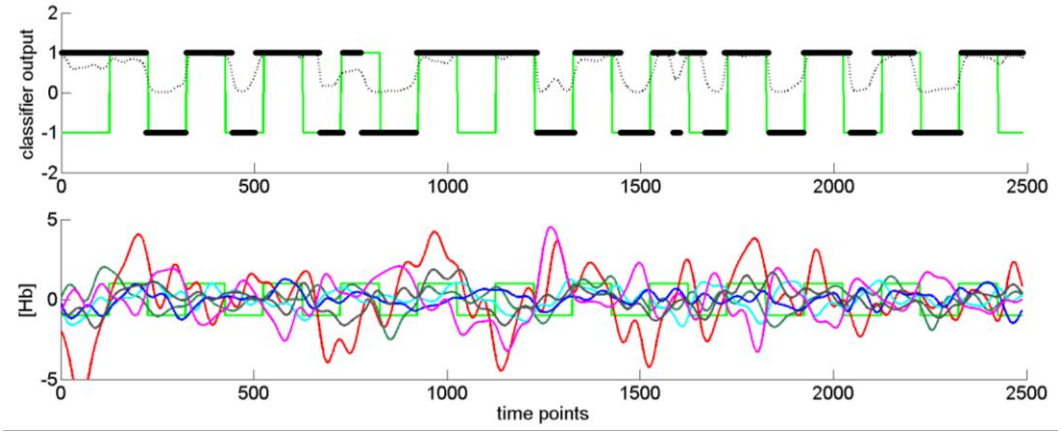


Figure 27. Prediction data and classifier outputs for one run, PCA case

Prediction accuracy was 75.5% for this run, and three training runs were used. In the top panel, black markers at +1 indicate a classifier output of a more engaged or 'working' state, while those at -1 indicate a less engaged or 'resting' state. A prediction is made at every instance. The dotted black line indicates the probability estimates of the classifier predictions. In the bottom panel, the six classifier input features are plotted. The four top principal components, in order, are red, green, cyan and blue. The two top principal components, in order, are magenta and grey. Truth labels are indicated by the green task indicator function in both the top and bottom panels.

MMSE during the MSIT task based on functional traces from the DLPFC and the MFG produced using adaptive regression, averaged across seven participants, is shown in Figure 28. These are shown for each participant in Appendix F. In line with the literature cited in 4.3.8, we found complexity to reliably increase, for both the TPN and the TNN region, during active periods at the group level in the adaptive regression case. However, our classification results are determined within-participant, and thus the responses used for the classifier predictions do not benefit from group averaging. Accuracy only reaches 77%, which is less than the 84% obtained using the functional traces produced by adaptive regression directly, as presented in section 4.4.4.

Table 8. SVM parameters for fNIRS classification with PCA features

participant	processing method:	as-measured	static	adaptive	pco static	pco adaptive
15	**c**	0.01	0.01	1	0.01	0.01
	g	0.1	0.5	0.1	0.5	0.5
16	**c**	1	0.01	0.01	0.01	0.001
	g	0.5	0.1	0.1	0.1	0.5
17	**c**	1	1	0.01	0.01	1
	g	0.5	0.5	0.5	0.1	0.1
18	**c**	0.01	0.01	0.1	1	0.1
	g	0.1	0.1	0.5	0.5	0.5
19	**c**	1	1	0.1	0.1	0.1
	g	0.1	0.5	0.1	0.1	0.5
20	**c**	0.1	0.001	0.01	0.001	0.1
	g	0.1	0.1	0.1	0.1	0.1
21	**c**	0.01	1	0.01	0.01	0.1
	g	0.1	0.5	0.5	0.1	0.5

pco: physiology-cleaned only

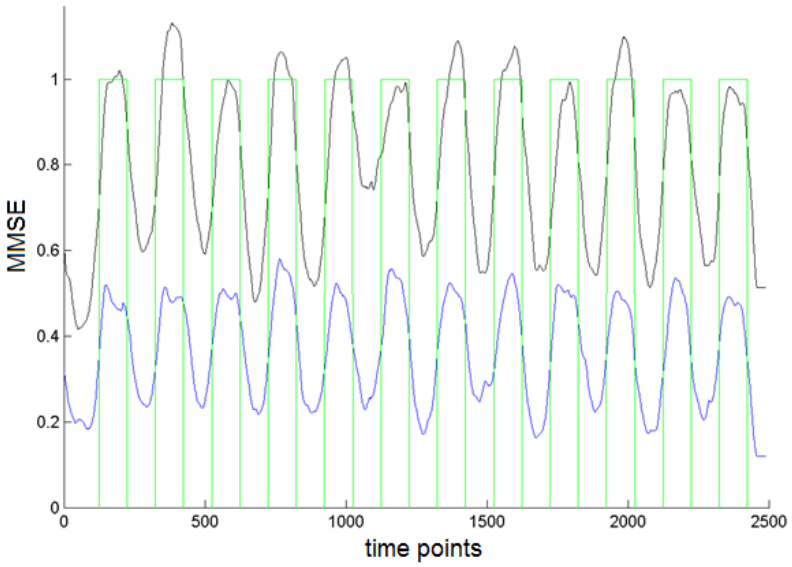


Figure 28. MMSE during the MSIT based on adaptive regression functional traces

Functional traces were taken from the DLPFC (black) and the MFG (blue), during the MSIT, averaged across seven participants. The task indicator signal is shown in green, with +1 indicating a work period of the MSIT.

Table 9. SVM parameters for fNIRS classification with MMSE features

Parameters are given for various fNIRS processing methods.

participant	processing method:	**as-measured**	**static**	**adaptive**	**pco static**	**pco adaptive**
15	**c**	1	1	50	10	5
	g	1	0.1	0.1	0.1	5
16	**c**	1	10	5	1	10
	g	0.1	0.1	0.1	0.1	0.1
17	**c**	50	10	1	10	1
	g	5	0.1	0.1	0.1	0.1
18	**c**	50	10	1	1	5
	g	0.1	0.1	0.1	0.1	1
19	**c**	1	1	50	10	1
	g	0.1	0.1	5	0.1	0.1
20	**c**	50	1	1	1	1
	g	0.1	0.1	0.1	0.1	0.1
21	**c**	10	10	5	10	1
	g	5	0.1	0.1	0.1	0.1

pco: physiology-cleaned only

4.4.9 Features Produced across Networks by Rolling Window Correlation

RWCC traces were calculated as described in section 4.3.9 based on functional traces produced using static and adaptive regression. Classification accuracy was 52.3% +/- 2.9% in the static case, and 56% +/- 2.4 % in the adaptive case. Only the adaptive case achieved statistical significance with respect to chance prediction ($t(6) = 6.526$, $p < 0.0005$). Statistics for each case are presented in Table 4. The as-measured case was not predicted. SVM parameters for each participant are presented in Table 10.

To investigate possible reasons for a lack of significant classification accuracy in the RWCC case, the traces were averaged across seven participants, and are shown in Figure 29. Their spectral content is shown in Figure 30, determined as described in section 4.3.4.3. Task response is seen at 0.03125 Hz but the magnitude is not greater than that of other low frequency components. Interestingly, but of no positive consequence for classification accuracy, more prominent spikes are seen at 0.0625 Hz, which is equivalent to the task switching frequency.

CVRT for each task on-off epoch was plotted with the across-network RWCC for each participant and is shown in Appendix G. Consistent with a lack of activity at the task frequency, no relationship was apparent.

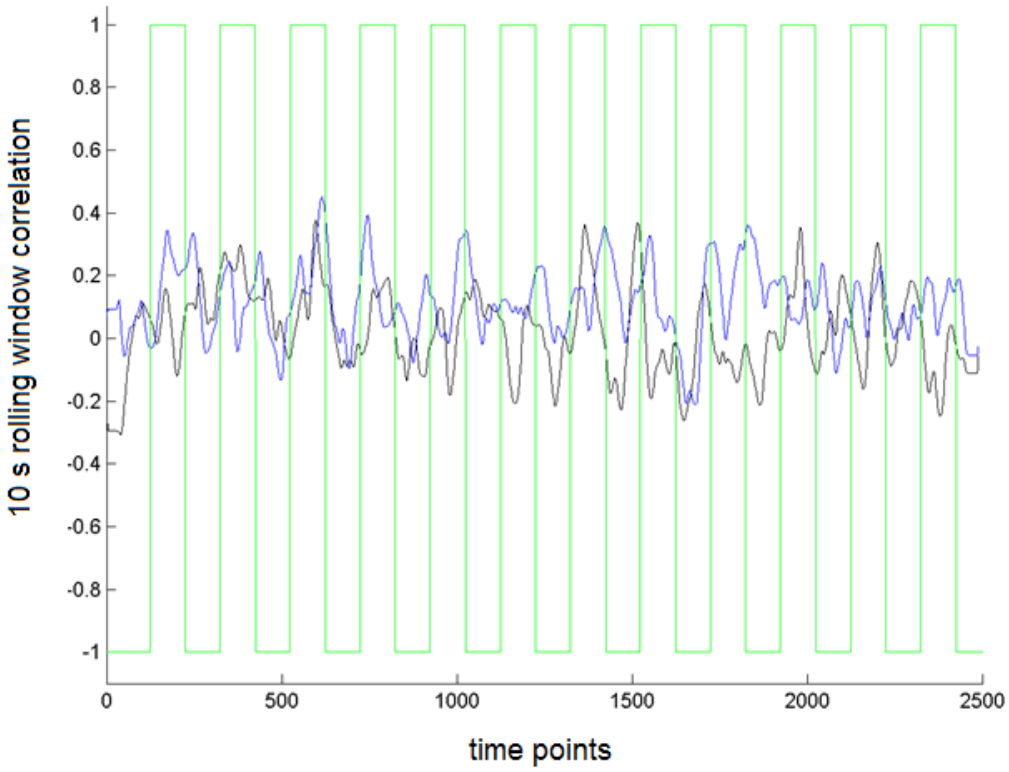


Figure 29. RWCC during the MSIT based on adaptive regression functional traces

Across-network RWCC during the MSIT based on adaptive regression functional traces from the DLPFC and the MFG. The top-ranked pair (black) and the second-ranked pair (blue) are shown, averaged across seven participants. The task indicator signal is shown in green, with +1 indicating a work period of the MSIT.

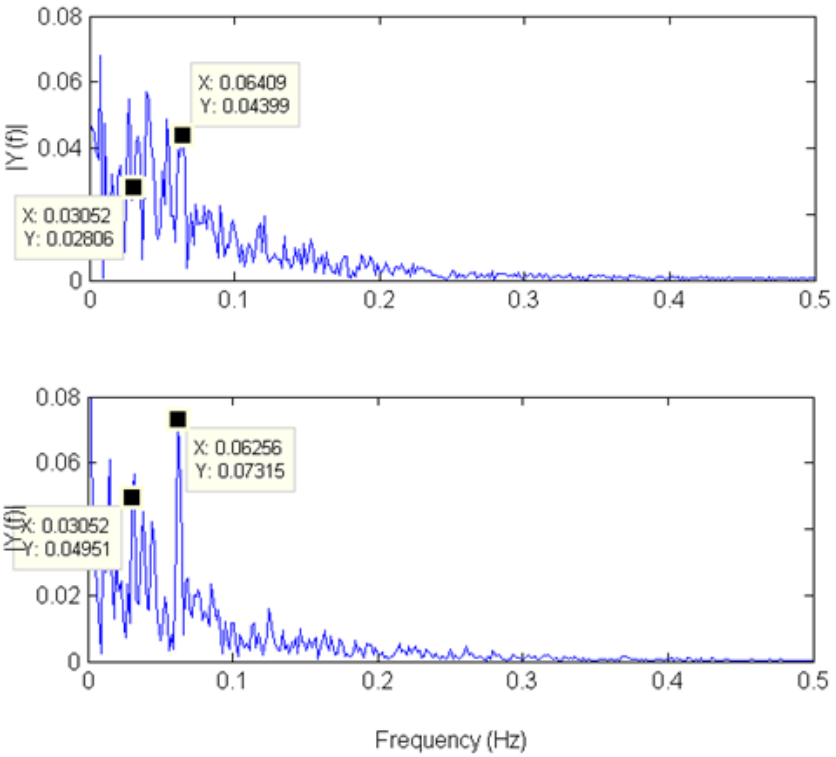


Figure 30. Spectral content of the across-network RWCC

The RWCC traces in figure 29 are shown (top: top-ranked pair, bottom: second-ranked pair).

Table 10. SVM parameters for fNIRS classification with RWCC features

RWCC was determined using the static and adaptive functional traces.

participant	processing method:	**static**	**adaptive**
15	**c**	0.001	1
	g	0.1	0.5
16	**c**	0.1	1
	g	0.1	0.5
17	**c**	0.01	1
	g	0.5	0.5
18	**c**	0.001	1
	g	0.5	0.5
19	**c**	0.001	0.1
	g	0.1	0.5
20	**c**	0.001	1
	g	0.1	0.5
21	**c**	1	0.1
	g	0.1	0.5

4.5 fMRI Processing and Classification Methods

In part two of the study which we discuss here, simultaneous fMRI data were acquired because the prediction of task engagement based on this data provides the standard for comparison in our study. This guarantees within-participant comparison of results between the two techniques for the same spatial patterns of hemodynamic activations during the same task performance, controlling for such variables as effort, motivation, practice, recruitment and task performance strategy, anatomy, time of day and physiological state. fMRI data were acquired, processed and classified as descried in section 2.5.

We predicted task engagement using fMRI in three ways. First, two features were created from regions according to participant-specific statistical maps (similarly to that done in part one, section 3.5.1, but with a spherical, contiguous volume) to determine classification accuracy based on statistically *significant* fMRI BOLD traces. Second, six features were created from regions local to the estimated sensitivity of the six fNIRS probes to determine classification accuracy based on BOLD traces from restricted spatial volumes analogous to the *sensitivity* of fNIRS probes. Third, nearly 30,000 features were taken from individual voxels across the *whole brain* to determine classification accuracy based on information of spatial extent and resolution not achievable with fNIRS. We expect this level of information to produce superior state prediction accuracy. We present these methods here and the results in section 4.6. The fMRI data were not used for one participant due to artifact which did not affect the fNIRS data. Thus, results are presented for six participants.

4.5.1 Voxels Selected by Significantly Activated Clusters

A 1 cm-radius spherical volume of contiguous voxels was used at each location, centered at participant-specific locations. Participant-specific locations were selected using the second-level statistical maps generated using that individual participant's four runs. The location for the DLPFC region was selected using the second level analysis of the work minus rest contrast. The location for the MFG region was selected using the second level analysis of the rest minus work contrast. Figure 31 shows the locations selected for one participant. Two classifier input features

were generated by averaging across each of the two contiguous clusters of voxels. Additionally, a grey matter mask for that participant was applied to ensure no out-of-brain voxels were used. Figure 32 shows the resulting BOLD traces after averaging across the six participants.

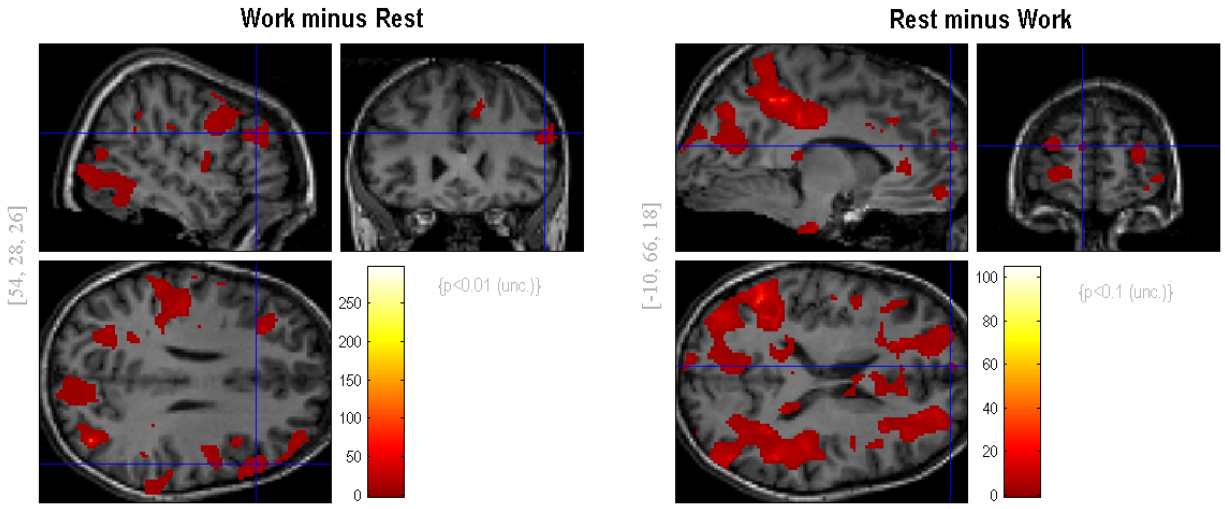


Figure 31. Participant-specific locations of significant BOLD responses for one participant
Locations of statistically significant activations for the MSIT are shown after second-level analysis across four runs for one participant near the same regions that were interrogated by the fNIRS probes. Left: The t-statistic is mapped, with the height threshold set at an uncorrected significance level of $p<0.01$ and the extent threshold set at 10 voxels. The DLPFC location is shown by the crosshair at [54, 28, 26] in MNI space for the work minus rest contrast. Right: The t-statistic is mapped, with the height threshold set at an uncorrected significance level of $p<0.1$ and the extent threshold set at 10 voxels. The MFG location is shown at the crosshair at [-10, 66, 18] in MNI space for the rest minus work contrast.

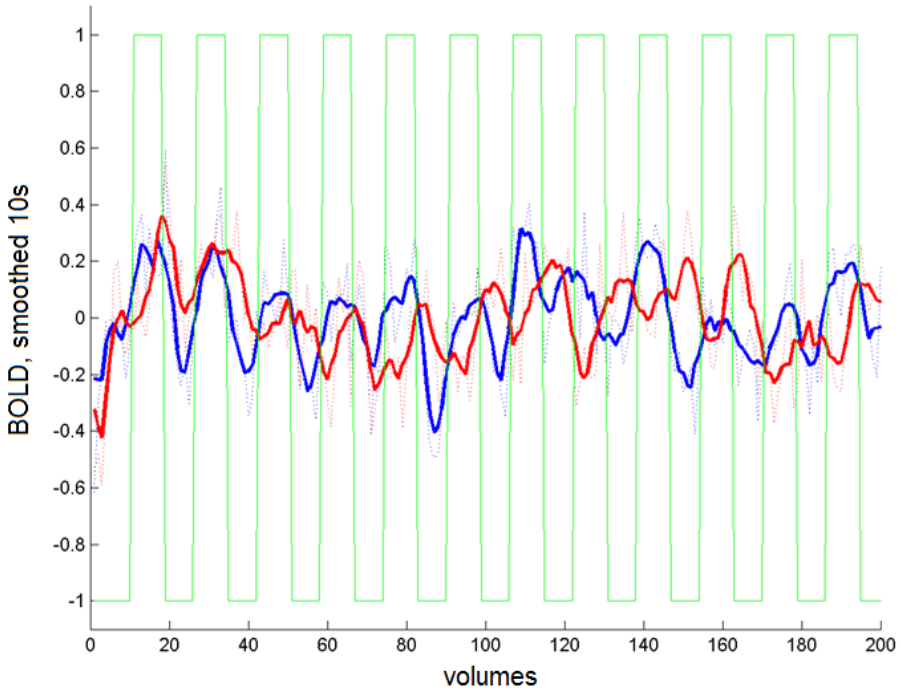


Figure 32. BOLD traces after averaging across significantly activated clusters

A spherical ROI of 1cm radius was used to determine the voxels in the cluster at each location. After averaging across voxels for each participant at each location, the traces were averaged across six participants for each location. The two resulting traces are shown (red: MFG, blue: DLPFC). Dotted lines indicate +/- one standard deviation. The task indicator signal is shown in green, with +1 indicating a work period of the MSIT.

4.5.2 Voxels Local to the Sensitivity of fNIRS Probes

We wished to determine classification accuracy based on BOLD traces from restricted spatial volumes analogous to the sensitivity of fNIRS probes. Our purpose is to determine classification accuracy expectations when information is restricted to local tissue volumes for a fair comparison of fNIRS and fMRI results. Similar to the case of section 4.5.1, individual classifier input features were created by averaging time traces over the voxels from contiguous 1 cm-radius spheres. However in this case, six such spheres were used. This resulted in six classifier input features (one per fNIRS optode), and no pruning was implemented. Additionally in part two of the study, a grey matter mask for that participant was applied to ensure no out-of-brain voxels were used, with the goal of approximating a volume which overlaps the portion of the fNIRS optode sensitivity profile that is in the brain. We aim for brain overlap only because

our fNIRS traces have been cleaned of contributions from the superficial tissue as discussed in section 2.1.3.

4.5.2.1 Sphere location method

The centers of these spheres were located at a point 1 cm into the brain from each optode's estimated scalp intersection point, along a line between that point and the anterior commissure. The scalp intersection point was determined by moving one step in each dimension within that participant's anatomical overlay image from the channel middle point toward the anterior commissure until intensity above a participant-specific threshold was encountered. The thresholds were adjusted until the scalp intersection points fell on the outside surface of the head as opposed to on the surface of the brain. An example of such a point is shown in Figure 33 (right). The channel middle point was defined as halfway between that optode's source and the detector. The anatomical images, from which the skull had not been stripped, were normalized along with the functional volumes as described in section 2.5.2.

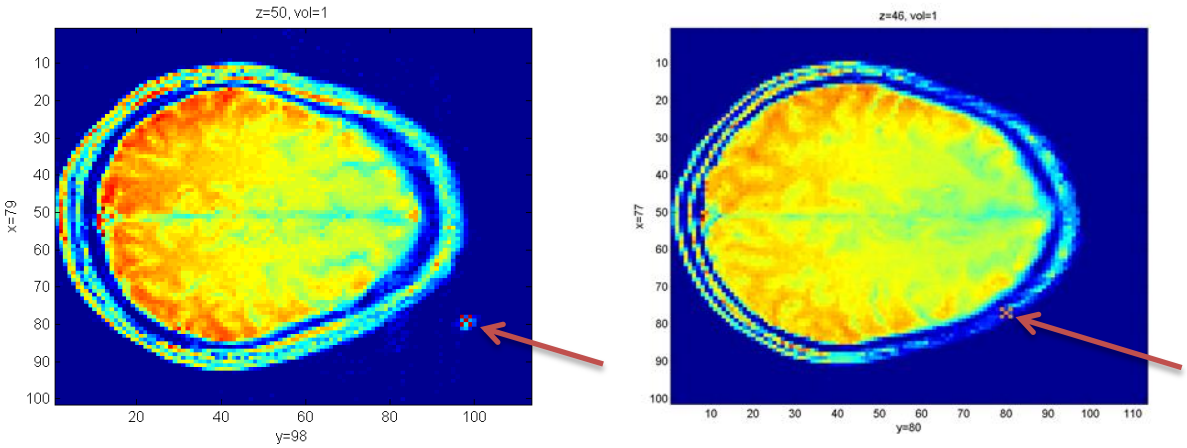


Figure 33. NIfTI images of a marker and a scalp intersection point

Points are marked for one participant (left) for one optode (right) are shown. An arrow points to a small x identifying the point in each.

Six channel middle points were determined based on their locations with respect to the vitamin E marker, which were measured by hand once for all participants. Four of these were on the DLPFC probe array, and two were on the MFG array. The DLPFC probe array was assumed to be first in a plane parallel to the sagittal fMRI slices, then rotated counter clockwise through

45 degrees about the axial or z axis. The vectors between the marker and the channel middles were then assumed to be in MNI space, where, for this study, each voxel covered 2 mm in each of the three dimensions. The maximum translation in one dimension after rotation was 15 mm. Due to non-linear normalization field deformations outside the head volume, differences between real space and normalized space were quantified and found to be on average 6% of the separation distances, as presented in Appendix H. Thus we were able to move in MNI space from the markers to the channel middles via the rotated vectors. Since it was placed on the head nearly parallel to the coronal fMRI slices, no such rotation was used for the MFG probe array. A marker on the DLPFC probe array is shown in the anatomical overlay in Figure 33 (left).

Thus, one sphere was constructed per optode, as shown in Figure 34, for a total of six clusters and therefore six classifier input features.

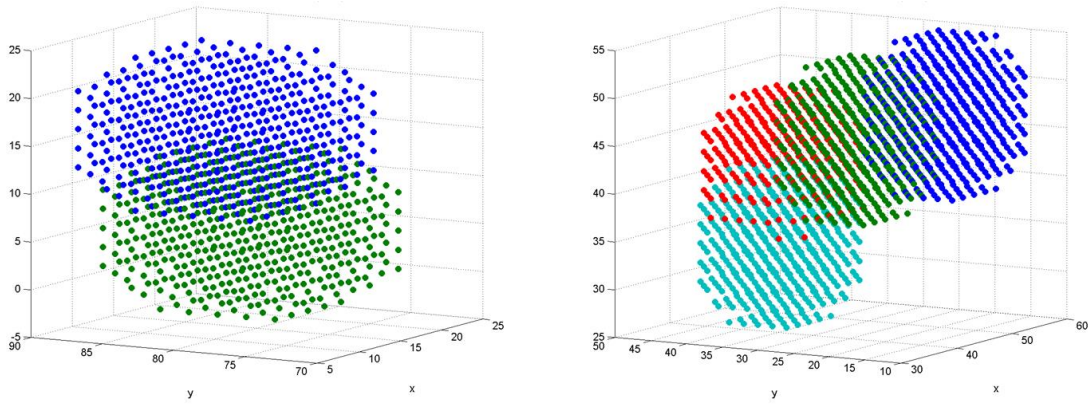


Figure 34. Spherical volumes used at the MFG and the DLPFC

Volumes for one participant are shown. When combined with a grey matter mask, these volumes are used to approximate the sensitivity profiles of two fNIRS MFG optodes (left), and four DLPFC optodes (right).

4.5.2.2 Rationale for the use of 1 cm-radius spheres centered at points 1 cm deep

We use a sphere at depth and radius informed by theoretical calculations, empirical combined fNIRS and fMRI studies, and the lateral spread of simulated sensitivity profiles as described below. Tuchin, 2007, p. 454, eq. 6.5 gives the depth from the surface or scalp, z, of the modal line of the maximum depth of the most probable photon migration path for highly scattering media at half the source-detector separation distance as:

$$z\ (y= r_{sd}/2) = r_{sd} / 2\sqrt{2} \quad (5)$$

where r_{sd} is the source-detector separation distance. This is 11mm for r_{sd} = 30mm. Since this is the modal line, there will be some photon paths that reach shallower and larger depths.

Three relevant empirical studies were considered to further inform our selection of depth and extent of volumes for comparison of fNIRS and fMRI signals. In the first, Huppert, et al. (2006) found the mean depth of functional contrast for a finger tapping task to be 14 mm +/- 4 mm from the surface of the head. In the second, Cui, et al. (2011) examined the correlation between fNIRS and fMRI signals across cognitive tasks. The best correlation was found using an elliptical shape drawn between points where projections from the source and detector intersect the brain. For spherical shapes, the best correlation was for spheres of radii 17.5 mm centered at the brain surface. No variation of correlation was found across tasks. In that study, no physiological correction was performed for either modality beyond temporal filtering. However, MRI is not subject to the same sensitivity to the superficial tissue as fNIRS. This is important in the frequency range of interest (that of the HRF and slower sustained activations) due to slowly varying local and systemic physiological effects (as discussed in section 4.4.4). Therefore, surface contribution removal for fNIRS, as discussed in section 2.1, is warranted for proper comparison of fNIRS to cortical fMRI signals. In the third, Sasai, et al. (2012) considered axial slices only, and looked at the correlation of fMRI and fNIRS signals considering voxels within a 3cm radius of the projection point on the cortical surface. The best correlation was found within 2 voxels (2 mm) of this point. No fNIRS superficial tissue correction was made.

Finally, we considered simulation results to refine our volume characteristics. Okada and Delpy (2003) presented Monte-Carlo simulations of NIRS sensitivity profiles in modeled brain slabs with varying Cerebral Spinal Fluid (CSF) and skull layer thicknesses. They showed the thickness of the CSF layer influences overall scalp-to-brain distances and therefore the shape of the sensitivity profile in the direction along the depth. Also, they show that the thickness of the skull layer influences the partial optical path length through the brain. For greater skull thicknesses, less brain is sampled. Laterally, both skull and CSF do not affect the spread of the sensitivity profile, being relatively constant at 15 mm at FWHM and extending approximately +/- 10 mm at 20% of the normalized intensity. Considering all their modeled cases, fNIRS sensitivity reached maximally up to 20 mm of depth from the head surface.

We selected a spherical volume of 10 mm radius centered 10 mm from the scalp along a line toward the anterior commissure deep inside the brain. Thus the maximum sampled depth for our purposes is estimated to be 20 mm. We did not adjust for skull and CSF thickness in our study. Our selected spherical volume encompassed more than the maximally-correlated voxels found by Sasai et al., reached a lower depth than that found by Cui et al., included slightly deeper than that found by Huppert et al. and the theoretical value, and reached the same maximal depth but slightly less than the lateral spread as that found by Okada and Delpy. Additionally, this is in line with the results of Schroeter et al. (2006). Given the proximity of the neighboring optodes in our probe arrays, there is some lateral overlap of the volumes (as seen in Figure. 34) which counters this. At the risk of including non-informative voxels in the cluster average, we cast a relatively wide net to capture voxels of fMRI measurements likely to fall within the maximal reach of the fNIRS measurements along the depth dimension.

4.5.3 Voxels from the Whole-Brain

Finally, we wished to determine classification accuracy expectations when BOLD traces from voxels of spatial extent and resolution not achievable with fNIRS are used. In this whole-brain case, no averaging of time traces was done. fMRI data were processed as described in section 2.5.2 and classified as described in section 2.5.3. Participant-specific skull-stripped anatomical overlay images were used to select voxels from the functional images which were located in the brain superior to z = -20 mm in normalized SPM coordinates. The number of voxels in this mask, and thus the number of classifier input features, averaged 28926 +/- 876 across the six participants.

Separately, a linear SVM was employed (testing c = 1, 2, 5, 10, 100 with g = 0) to produce a map of the support vector weights associated with the time trace at every other voxel. The weights indicate which voxels most influenced the placement of the separation hyper-plane during classifier training and SVM model generation. The weights were taken from the generated model and converted to z-scores. The scores were multiplied by 1000 for visualization purposes before being plotted voxel by voxel in greyscale on the anatomical overlay for that participant.

4.5.4 Across-Network Rolling Window Correlation

To complement the rolling window correlation fNIRS analysis of sections 4.3.9 and 4.4.9, a similar rolling window correlation analysis was performed using the two fMRI classifier input feature traces of the significant case described in sections 4.5.1 and 4.6.1. These features provide time traces based on information in key regions of the TPN and TNN as verified in section 3.5. An even number of TRs, specifically 6 for a window of 12 s, was used for the window length to quantify across-network correlation. The traces were averaged across the four runs for each participant, and then averaged across all six participants.

4.6 fMRI Classification Results

We predicted task engagement using fMRI with two features created from regions according to participant-specific statistical maps to determine classification accuracy based on statistically significant fMRI BOLD traces, with six features created from regions local to the estimated sensitivity of the six fNIRS probes to determine classification accuracy with spatially-restricted volumes, and with nearly 30,000 features taken from individual voxels across the whole brain to determine classification accuracy based on information of spatial extent and resolution not achievable with fNIRS. We expected this level of information to produce superior state prediction accuracy. The state prediction accuracies and the standard deviations are given in Table 11, and shown in Figure 35 per participant, for the significant, sensitivity, whole brain and fNIRS static physiology cleaned only cases. All accuracies were determined to be significantly greater than chance prediction at 50%, as also shown in Table 11. In Table 12, we present the SVM parameters for each participant for the same cases.

Table 11. State prediction accuracy and statistics for fMRI classification

Accuracy and statistics are given in the significant, sensitivity, whole brain and fNIRS static physiology cleaned only cases, averaged across six participants. The number of classifier input features is shown in square brackets.

processing method:	**% accuracy**	**SD**	**t(5)**	**p <**
significant [2]	59.4 †[1]	6.5	3.563	0.01
sensitivity [6]	66.5 †[1]	12.0	3.367	0.01
whole brain [~30k]	94.2 †[4]	3.2	34.061	0.0005
fNIRS static physiology cleaned only* [6]	63.2 †[3]	6.9	4.671	0.005

†[1] $p<0.01$ †[2] $p<0.001$ †[3] $p<0.005$ †[4] $p<0.0005$

* This accuracy differs from the value presented in Table 3 because here we consider only the 6 participants with useful data in both modalities.

Table 12. SVM parameters for fMRI classification

Parameters are given in the significant, sensitivity whole brain and fNIRS static physiology cleaned only cases, for six participants. The number of classifier input features is shown in square brackets.

participant	processing method:	**significant [2]**	**sensitivity [6]**	**whole-brain [~30k]**	**fNIRS static physiology cleaned only [6]**
16	**c**	200	100	5	200
	g	0.01	0.001	0.00001	0.01
17	**c**	1	2	100	1
	g	5	0.01	0.00001	5
18	**c**	100	10	100	100
	g	0.01	0.1	0.00001	0.01
19	**c**	100	100	100	100
	g	0.01	0.005	0.00001	0.01
20	**c**	10	200	100	10
	g	0.01	0.005	0.00001	0.01
21	**c**	200	10	5	200
	g	10	0.1	0.00001	10

4.6.1 Voxels selected by Significantly Activated Clusters

Task engagement was predicted using two features based on statistically significant fMRI BOLD traces for each participant. The accuracy of 59.4% +/- 6.5% is within the range of our result from part one of 65.6% /- 11.6%.

4.6.2 Voxels Local to the Sensitivity of fNIRS Probes

Task engagement was next predicted using six features based on spatially-restricted volumes for each optode as described in section 4.5.2. This facilitates a relatively more direct comparison between classification done with fNIRS and with fMRI. The accuracy was 66.5% +/- 12.0. In this case, fMRI classification has been done using traces cleaned of physiology but not produced using a GLM analysis. Thus, the static physiology-cleaned-only fNIRS case from section 4.3.5 is compared. That fNIRS accuracy was 63.2% +/- 6.9%. The difference between the accuracies for these two cases was not statistically significant ($t(5) = 0.62$, $p<0.4$).

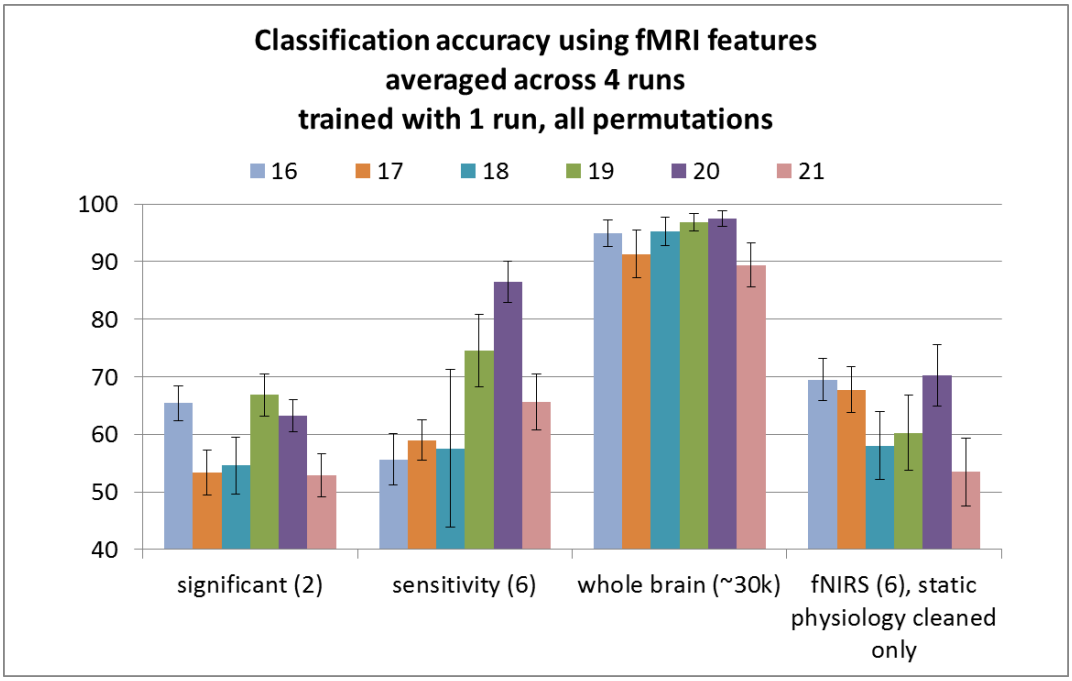


Figure 35. State prediction accuracy for significant, sensitivity and whole brain fMRI cases
Accuracy for each of the six fMRI participants is shown, averaged across four runs each. The number of classifier input features is shown parenthetically. The fNIRS static physiology cleaned only case is shown for comparison to the fMRI sensitivity case.

4.6.3 Voxels from the Whole-brain

Finally, task engagement was predicted using nearly 30,000 input features to determine classification accuracy based on hemodynamic information taken from the whole-brain. As expected, classification accuracies in this case were quite high, averaging 94.2% +/- 3.2%. This result validates our task paradigm and classification methods as a positive control, and represents classification accuracies likely not achievable with the ambulatory technique of fNIRS.

Accuracy using the linear SVM did not differ from that determined using the radial basis function kernel with very low g values. Support vector weight maps described in section 4.5.3 are presented in Appendix I. Weights are shown in greyscale on the images. The greater the weight's absolute value (white: positive, black: negative), the more influential that voxel's time trace was on the SVM model. Greater weight values indicate relatively more influence in the determination of the vector **w** as described in section 2.4. Greater positive weights are in motor, visual and prefrontal cortex regions in each subject. This is as expected considering the motor response and visual stimulus presentation used in the work period of our cognitive task.

4.6.4 Across-Network Rolling Window Correlation

The RWCC between the two classifier input feature traces used in the significant fMRI classification case is presented in Figure 36. The spectral content of the RWCC trace is shown in Figure 37. Similar to the results presented in section 4.4.9, minimal task response is seen at 0.03125 Hz among other low frequency components. Relatively more prominent spikes seen at 0.06 Hz in section 4.4.9 are not seen here, suggesting there they may be a result of the smaller window size.

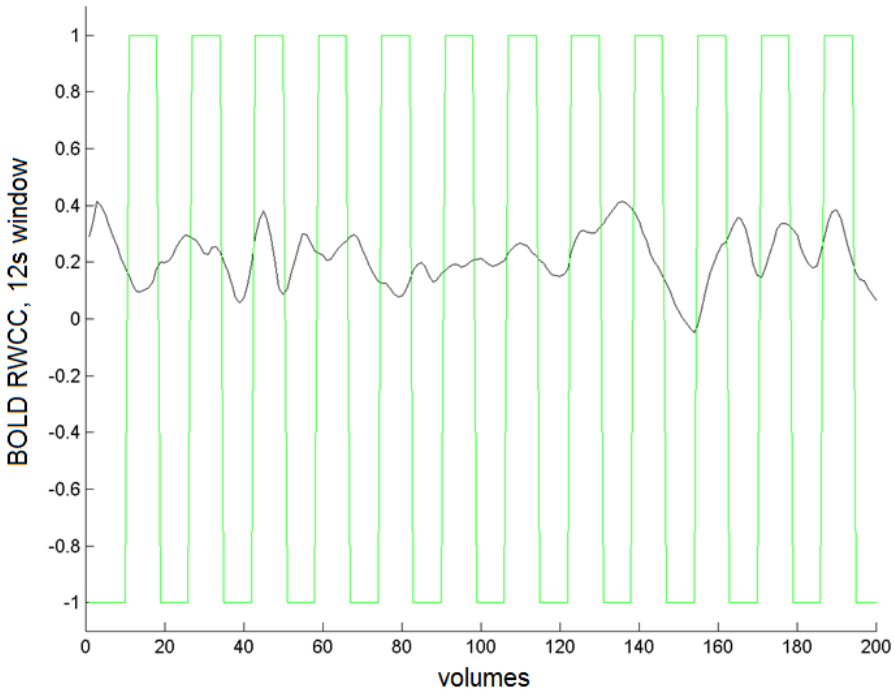


Figure 36. RWCC between the two fMRI classification input features in the significant case The task indicator signal is shown in green, with +1 indicating a work period of the MSIT.

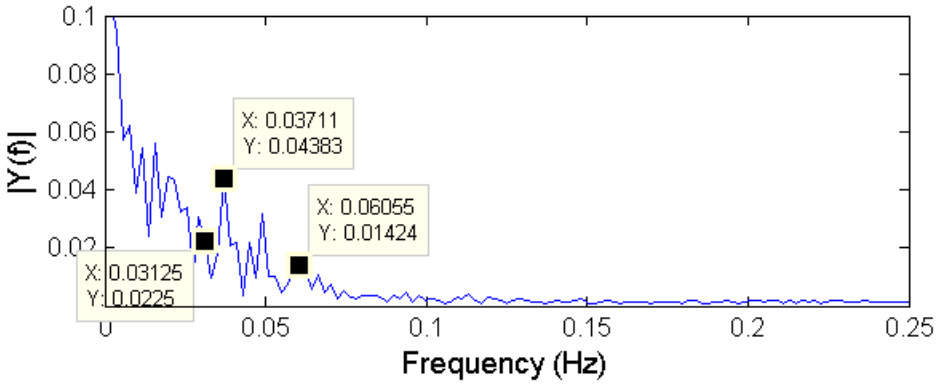


Figure 37. Spectral content of the RWCC between the two significant case fMRI features.

4.7 Discussion

In part two of the present study, we performed simultaneous fNIRS and fMRI data collection and compared attentional state prediction based on spatial patterns of hemodynamic activation as measured by both modalities. Multiple fNIRS data processing methods were employed to determine which of these methods and which classifier feature types would produce the best state prediction accuracy. The as-measured and physiology-cleaned only cases explored real world application where knowledge of the task being performed is not known. Other feature types based on both static and adaptive processing methods explore the best capability of an attentional state monitoring system when employed in controlled laboratory settings. Our findings are discussed here.

4.7.1 Findings of part two of the present study

Both static and adaptive methods were used to remove noise from the measured signals. State prediction accuracy performed using signals after noise removal via adaptive regression produced the highest mean classification accuracy overall. In every case, SER for adaptive regression was less than that for static regression, indicating that noise removal was superior with adaptive regression and a strong model of expected task activity. We conclude that state prediction accuracy of 84% represents the best capability of an attentional state monitoring system using the optode localization methods presented in this study, and when employed in controlled laboratory settings when stimulus onset times are known. The superior performance of the adaptive filtering method indicates that information extraction is more likely to be limited by the non-stationary nature of physiological noise (Kolehmainen et al., 2003; Diamond et al., 2006; Zhang et al., 2007b; Abdelnour and Huppert, 2009) than to be limited by regional optode localization precision.

Additionally, adaptive regression reduces variable low frequency activity. The significant reduction in amplitude of very low frequency activity in the adaptive regression residuals with respect to those of the static regression, and the lack of significant correlation with a behavioral metric, may indicate that observed peaks in the static regression residual traces in the very low

frequency range are mostly due to un-modeled time-varying physiology, such as vasomotion, that is present in the shallow tissue and effectively removed by adaptive processing. The fNIRS technique is very sensitive to changes in the superficial tissue, as its non-invasive nature inherently involves sensing of the vasculature in the skin and through the skull. Thus, it is highly sensitive to both systemic Mayer waves and local vasomotion (Sassaroli, 2012). However, intrinsic resting state activity also may be contributing to signal content in this frequency range and cannot be ruled out with our study. Whether variability in the very low frequency range is due to physiology or due to varying intrinsic resting state functional activity, adaptive filtering effectively removed it to focus on task effects.

The as-measured traces produced the best classification accuracy of all cases for which no information about the stimulus onset times was used. This was 70%, which is in line with previous results obtained using a more operational and arousing gaming task (Harrivel et al., 2011). Other processing methods which also did not use stimulus onset information cannot separate noise from signal effectively. Indeed, shallow signal removal can actually reduce signal contrast if applied inappropriately (e.g. when the signal is not dominated by physiological confound) (Zhang et al., 2009) or when task effects (arising from skin or brain layers) cannot be removed from the nuisance regressors (as in the physiology-cleaned only cases). Thus, is it perhaps not surprising that other methods are outperformed by the minimally-processed as-measured traces. These trends remained even with a reduction in the amount of training data.

The shallow fNIRS traces themselves produced the second-best accuracy for cases where no knowledge of the task was used. Brain response may be seen even at 1 cm separation. Additionally, task-evoked effects in the superficial tissue are clearly real and have been reported elsewhere (Kirilina et al., 2012), but generally were not reliable for all participants in our study.

In the static regression case, evidence of some mixing of shallow contributions from each species, due to unequal sampling of arterial and venous compartments by each optode, was seen for both species and at both the DLPFC and MFG regions as we predicted. This mixing was most pronounced for the [HbO] traces at the DLPFC. Despite this improvement in the model used for nuisance regression, static regression provided only a minimal increase in state prediction accuracy across participants, and was far out-performed by the adaptive regression method.

It is encouraging that the data-driven method of PCA produced state classification accuracy of 80% with adaptive regression and also achieved significant prediction accuracy with

static regression. Also, the application of MMSE did very well, but only for the adaptive case. This indicates that these methods do best when physiological confounds unrelated to brain activity are effectively removed. Currently, due to a lack of knowledge of task performance timing to improve such removal, this finding does not bode well for their application in real world situations.

Real world applications present further complication due to the dynamic nature of functional connectivity (Chang and Glover 2010; Deco et al., 2011; Hutchinson, et al., 2013; Majeed et al., 2011; Thompson et al., 2014), as well as uncontrolled task-unrelated thoughts and mind wandering (Gruberger et al., 2011). This is because the hemodynamic activations of interest in the TPN and TNN are not neatly driven to be more or less anti-correlated in-task vs. out-of-task at the single-participant, moment-to-moment level. Indeed, Kelly et al. (2008) describe tightly-coupled anti-correlated network activations that explain the variance of overall participant performance levels, not participant performance on a moment-to-moment or time-windowed basis. Further, co-activation and therefore positive correlation is expected during internally-guided thought (Christoff et al., 2009; Smallwood et al., 2012). Therefore, a lack of prediction accuracy based on RWCC traces does not mean we are not measuring intrinsicly-driven activity in functionally connected networks. However, anti-correlated network activity may be too highly variable at the single-participant level to be useful for the purposes of fNIRS probe placement verification in practice.

Finally, we predicted task engagement using fMRI to validate classifier accuracy. fMRI BOLD traces were taken from statistically significant regions, from regions local to the estimated sensitivity of the fNIRS probes, and from the whole brain. All produced state prediction accuracies which achieved statistical significance with respect to chance prediction. For the whole-brain fMRI classification case, state prediction was 94% accurate, as expected for the very high number of classifier input features. The best g parameters were extremely low. This corresponds as expected to linear classification, for which performance is expected to be high without non-linear mapping of the input features into a new, multi-dimensional space. The non-linear mapping otherwise used throughout this study was accomplished via the radial basis function kernel, which depends on the value of the parameter g (Chang and Lin, 2011). However, when fMRI is restricted spatially in the same manner as fNIRS, classification accuracy is not significantly different. While neither the fMRI nor the fNIRS case compared here

employed GLM processing, recall that the fNIRS functional traces were ranked with the top six selected of the twelve [Hb] traces available. This would tend to improve prediction accuracy. One might argue that pruning gave an advantage to the fNIRS classification in this case. However, it is impossible to selectively "prune out" portions of the optical path for one optode. Therefore, removing less-discriminating voxel traces from contributing to the fMRI classifier input feature trace generation would be inappropriate.

When fMRI data are available, support vector weight maps, such as those produced as described in section 4.5.3 and presented in section 4.6.3 and Appendix I, can be used to inform fNIRS probe positioning for specific tasks. Classification accuracy with whole-brain fMRI of 94% is boosted by separable activity across the two conditions in motor, visual, prefrontal and possibly other regions not accessible to fNIRS as implemented in this study.

4.7.2 Novel contributions of part two of the present study

This work builds on the novel contributions of part one to improve methods for the prediction of task engagement using fNIRS to detect key regions of functionally-connected networks. We accomplished and reported an extensive exploration of fMRI-validated attentional state prediction based on various fNIRS processing methods and features. Additionally, we extended existing adaptive filtering methods to use both real physiological nuisance traces and real task-evoked hemodynamic responses. Further, we applied the technique to the detection of cognitive responses as opposed to more robust and well-characterized motor and visual responses. We also have improved upon existing physiological correction methods to consider unequal sampling of arterial and venous vascular compartments across optodes. We have shown that RWCC traces are not informative for attentional state classification using the simple, two-state model of Figure 2. Finally, we demonstrate the usefulness of PCA to accomplish feature selection, and the application of multivariate entropy measurements to multi-channel fNIRS data.

4.7.3 Limitations of part two of the present study

4.7.3.1 Behavioral task model simplicity

Some limitations to this study are a result of the simplicity of our model for expected task-evoked network activation. The adaptive regression method does perform well using the model to remove noise, and contributions due to physiology in the shallow tissue are dynamic, so

dynamic filtering is appropriate. However, the model used in this study does not allow flexibility for states other than the two expected. Accuracy found to be better in the dynamic case may be because in this study we are forcing the classifier features (the functional traces and features created using them), via GLM analysis, to adhere more closely to a model that is derived from the expected two-state system of Figure 2. A model expecting anti-correlation when either on-task or off-task, with other participant performance strategies and task-unrelated thoughts discarded as noise, is too simple considering the adaptive nature of functional connectivity and potential network cooperation.

A potential alternate explanation for variation in intrinsic activity is variation in the cognitive resources each participant has available, while performing well on the task, for task-unrelated thoughts (MacKiernen 2006; Fransson 2006). Participants performed well on the task based on behavioral accuracy, and errors of omission were rare. However, task-unrelated and internally-guided thoughts vary across time throughout the task run, and across participants. These cannot be completely controlled, and alternate experimental paradigms would be required to study them appropriately (Gruberger et al., 2011). Finally, intrusive task-unrelated thoughts may interfere with performance due to the participant's inability to deactivate default mode network activity, which may contribute to errors and momentarily-worse behavioral performance (DeHavas et al., 2012; Drummond et al., 2005b; Prado et al., 2011; Weissman et al., 2006) not detected with fNIRS in this study. However, this study intended to investigate the classification of task engagement vs. resting periods, and individual moment-to-moment behavioral performance was not investigated.

4.7.3.2 fNIRS optode location estimations

There are also limitations to our estimation of volumes in the brain which approximate fNIRS optode sensitivity profiles, as used in section 4.6.2. First, one fMRI marker was used per probe array. To avoid approximations of the channel middle locations, markers should be used at each optode. Also, the precision of overlap will be limited in any case by the localization challenges discussed in section 2.1.2. Second, we did not adjust for skull and CSF thickness in our study, which varies by both location and participant and affects the size and depth of the optode sensitivity profile in the brain. Third, scalp intersection points were not on the very surface of the head in every case due to variability in the intensities of the anatomical overlay images. Fourth, the line between the channel middle locations and the anterior commissure only

approximates a line normal to the head. However, since our goal was to determine classification accuracy based on BOLD traces from restricted spatial volumes analogous to the sensitivity of fNIRS probes, we feel exact overlap is not necessary.

4.7.3.3 fNIRS classification specificity, sensitivity and relevance

Receiver Operator Characteristic (ROC) plots as shown in Appendix J for the fMRI sensitive case data consistently show moderately good (low) false positive rates but poor (low) positive predictive values (PPV, or sensitivity). This makes sense if the classifiers are not identifying task engagement (the positive or attending state) at all time points when it is the "true" state. Psychological confounds such as self-specific or introspective recall during a task, or the performance of a task that is relatively less-demanding, could be labeled as an inattentive state in error if the signal to noise ratio is too low. As explained above, such thoughts cannot be fully controlled and thus represent a limitation to the study. However, ROC plots as shown in Appendix K for fNIRS static, adaptive and as-measured cases indicate improvement in both sensitivity and specificity for the adaptive case, likely stemming from improved noise removal via the model discussed in section 4.7.3.1. False positives (states falsely classified as attentive) due to poor specificity may allow, for example, hazardous operator states to go unmitigated. False negatives (states falsely classified as inattentive) due to poor sensitivity may produce nuisance alarms that are ignored. Finally, unless behavioral measures or other contextual clues are included, whether a task being attended to is actually relevant to the operational task at hand cannot be distinguished.

Chapter 5
Conclusions and Future work

5.1 Conclusions

We have explored approaches for improving the validity and reliability of fNIRS-based cognitive state classifiers. In the present study, we employed fNIRS and fMRI and determined that it is possible to distinguish attentional state by measuring functionally-connected networks, and further that it is possible to detect negative correlations between activity in key regions of the TPN and TNN at the group level with fNIRS. fMRI was used to validate TPN and TNN task activity in response to the task paradigm.

Various fNIRS processing methods were explored and multivariate pattern classification techniques were successfully used to distinguish between periods of task performance and periods of rest. We found a trend that the detection of activity in both the TPN and TNN is useful for distinguishing between high and low levels of task engagement. Also, we extended adaptive regression techniques for fNIRS by using them to recover real cognitive task-evoked responses from measures contaminated with dynamic physiological artifacts. These artifacts were prominent on either hemoglobin species, and originated in the superficial tissue, to which fNIRS is highly sensitive. These novel methods, now demonstrated, can be used to improve the predictive power of fNIRS-based attentional state monitoring systems. Additionally, data driven and multivariate feature types were created. Their use is promising, but currently they do not provide a significant increase in performance.

Classification results using fMRI signals set accuracy expectations when information is restricted to local tissue volumes for a fair comparison of fNIRS and fMRI results. Alternately, the whole brain case produced state prediction accuracies far superior to those achievable using fNIRS, and indicates that more information from more regions in the brain is what produces superior results. Yet even with the spatial restrictions inherent in the fNIRS technique (poor spatial resolution and large partial volume errors), classification accuracies of 84% can be achieved using a known model of expected task activity and thus better noise removal

techniques. This represents the current best capability of an attentional state monitoring system based on our methods. We conclude that classification accuracy is limited by the availability of appropriately-complicated task models more than by regional optode localization precision. When we consider more realistic capability, the best classification accuracy is 70%. This level of accuracy is likely insufficient for monitoring and predicting cognitive state using fNIRS alone in real-world applications, unless task activity is measured *in situ*. However, future work (described in section 5.2) should improve classification accuracy. Further, these techniques are ready for investigation as a synergistic hemodynamic measure which may inform or complement EEG and other psychophysiological measures in multi-modal classification schemes.

Finally, we investigated the use of time-windowed across-network correlation as a classifier input feature. Positive classification results were not obtained using these features and the two-state model of Figure 2, and we found them to be too variable to be useful for the verification of correct fNIRS probe placement in practice. This variability is likely due to the dynamic nature of functional connectivity and moment-to-moment changes in internally-guided thought as discussed in section 4.7.1. Limitations to the simple attentional state activation model we employed were apparent due to variations between across-network correlation values at the single participant level. However, variation in correlation over time implies the information being measured in each network does not provide redundant information to the classifier. This further supports our conclusion from part one of the study that measuring both networks may be beneficial with more complicated models of activation.

5.2 Future work

The fNIRS and fMRI analyses in this study were aimed at distinguishing between performing a task and resting, rather than distinguishing between correct responses and errors. It was most important to the present analysis whether the participant was engaged in performing the task. All participants performed well on the task and errors were minimal. In the future, it also may be interesting to compare results of novice vs. practiced participants in more complex and demanding tasks. Training the classifier based on behavioral performance in the moment should be explored. Also, testing trained classifier models across tasks may be investigated by attempting to classify attentional state during one task (such as a more complex, task) using training data obtained during the performance of another task (such as the MSIT used here). This

would provide useful information regarding the complicated transfer of classification capabilities to real-world situations (Kingstone et al., 2003; Smilek et al., 2007). Ongoing commercial development of small-sized, robust and portable multi-channel fNIRS instrumentation will support the translation of these techniques from the laboratory through simulations and eventually to real-world applications.

The detection of multiple nodes of the TPN and TNN may provide an opportunity for improvements in accuracy and generalizability of trained classifier models by providing additional information to the classifier. Additional optodes may be employed where practicable for dense array detection (Eggebrecht et al., 2014) of network activations. However, considering our positive results and relatively sparse optode layouts used in this work, we note that such detailed mapping is not a requirement for cognitive state monitoring, and should only be implemented when practical limitations and resources allow.

To improve classification methods, using more than two classes can be explored to accommodate complicated, sometimes cooperative TPN and TNN activations. For example, network cooperation vs. anti-correlation can be explored against relatively high vs. relatively low network activity in a four-state classification scheme. As promising features for the indication of brain state, MMSE features may represent another opportunity for classifier performance improvement. However, they have only been minimally explored in this study. In particular, across-network MMSE should be investigated to quantify entropy in both networks across task condition, rather than that in each network. Finally, across-modality coherence and seed-based correlation analysis may be used to further examine the spatial and temporal details of the correspondence of fNIRS and fMRI data across task and rest conditions.

In general, we have demonstrated operationally-viable fNIRS methods for monitoring cognitive state, and have taken strides toward combining hemodynamic activation measurements with those of complimentary psychophysiological modalities. Employing multiple simultaneous methods for detecting the same state may lead to more reliable state identification, especially when the probability of being in a state, as predicted via each modality, is considered. This has positive implications for supporting optimal human performance in real-world contexts. We hope that future studies combining fNIRS recordings with other psychophysiological measurements and multivariate classification methodologies will build upon the present work to produce robust systems for detecting cognitive state changes in real-world settings.

Appendix A

Hemoglobin concentration change calculations

For each of two wavelengths, the time trace of raw intensity measurements are normalized to the mean and the optical density changes are calculated via:

$$\Delta O.D. = -\log (I(t)/Imean)$$

For each time trace, the optical density changes are used to calculate the hemoglobin concentration ([Hb]) changes in μmol based on the following relations:

$$\Delta O.D. = \mu_a * DPF * L$$

$$\mu_a = \varepsilon(\lambda) * \Delta[Hb]$$

where:

μ_a is the scattering absorption coefficient in cm^{-1},

L is the chord length between the source and detector in cm,

$\varepsilon(\lambda)$ is the known wavelength-dependent extinction coefficient in $\mu mol^{-1}\ cm^{-1}$,

and the differential path length factor (DPF) is unitless and accounts for the scattered optical path length being greater than L.

Absorption over the optical path length changes in time, generating changes in optical density which are detected as changes in the optical signal.

Two measurements are made (O.D.), and two unknowns (oxygenated hemoglobin concentration ([HbO]) and deoxygenated hemoglobin concentration ([HbR])) are calculated:

$$\Delta O.D.(\lambda 1) = (\varepsilon_{HbR\lambda 1} * \Delta[HbR] + \varepsilon_{HbO\lambda 1} * \Delta[HbO]) * DPF_{\lambda 1} * L$$

$$\Delta O.D.(\lambda 2) = (\varepsilon_{HbR\lambda 2} * \Delta[HbR] + \varepsilon_{HbO\lambda 2} * \Delta[HbO]) * DPF_{\lambda 2} * L$$

Please refer to the text of section 2.1 for context and references.

Appendix B

Beta fit parameters for DLPFC channels, study part one

Conditional formatting maps greater negative values to more intense red and greater positive values to more intense green. Participants are indicated in the "SN" column. Note that channels 1, 2 and 3 remain as the channels producing the most positive beta fit parameters even after excluding participant 281. Please refer to section 3.2.2 for context.

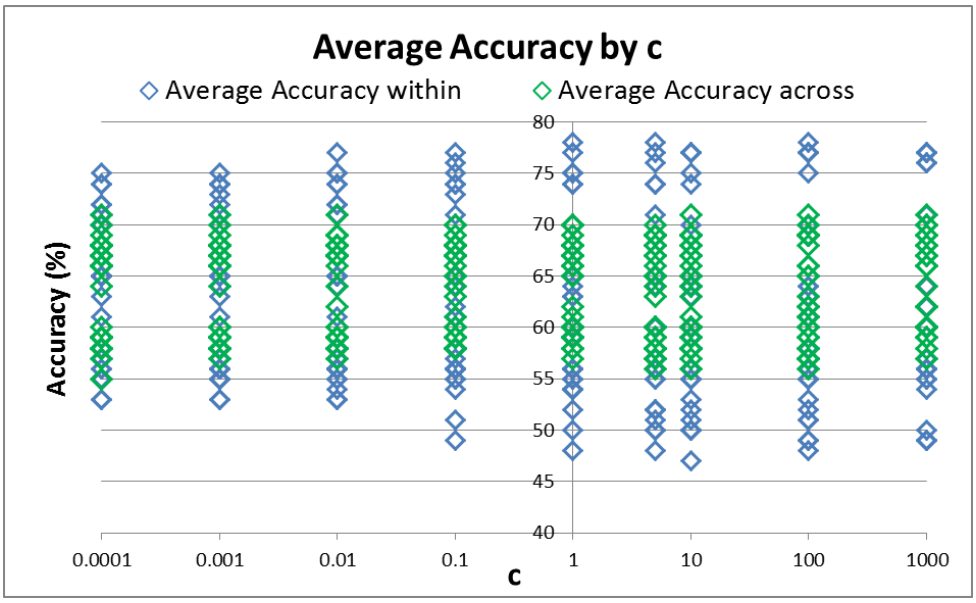
		sum:	15.32	17.63	14.04		2.87	3.76	-1.22	8.05
		optode:	1	2	3		5	6	7	8
SN	run	species								
081	1	oxy	-0.19	0.80	**1.02**		0.16	0.00	0.37	0.66
	2	oxy	0.51	1.03	**1.01**		0.13	1.13	0.55	0.87
	3	oxy	0.01	0.19	**0.16**		0.04	0.28	0.11	0.11
	4	oxy	0.58	0.90	**0.84**		-0.19	0.10	0.23	0.71
	1	deoxy	-0.53	-0.44	-0.76		-0.43	-0.42	-1.05	-1.04
	2	deoxy	-0.23	-0.07	0.23		0.60	0.03	-0.12	-0.26
	3	deoxy	-0.02	0.01	-0.02		0.22	0.11	0.06	0.02
	4	deoxy	0.02	0.06	-0.19		-0.22	-0.25	-0.54	-0.43
152	1	oxy	**0.73**	-0.18	-0.31		-0.24	0.52	-0.44	-0.45
	2	oxy	**0.33**	-0.49	-0.15		-0.50	-0.34	-0.38	-1.11
	3	oxy	**0.22**	-0.18	-0.56		-0.33	-0.21	-0.42	-0.26
	4	oxy	**0.28**	-0.60	-0.49		-0.32	0.37	-0.54	-0.72
	1	deoxy	-0.23	-0.76	-0.83		-0.26	0.22	0.03	0.07
	2	deoxy	0.57	0.19	0.25		-0.68	-0.53	0.05	-0.07
	3	deoxy	0.39	0.44	-0.34		-0.25	0.09	-0.20	-0.27
	4	deoxy	0.27	0.30	0.15		-0.45	0.63	0.35	-0.02
190	1	oxy	**0.37**	0.24	-0.48		0.18	-0.27	-0.11	0.19
	2	oxy	**0.60**	0.49	0.12		0.48	0.05	0.11	0.46
	3	oxy	**0.70**	0.76	0.64		0.77	0.50	0.12	0.38
	4	oxy	**0.77**	0.74	-0.82		0.76	-0.22	0.11	0.40
	1	deoxy	-0.51	0.19	-0.17		-0.16	0.24	-0.06	-0.24
	2	deoxy	-0.31	0.12	-0.04		-0.19	0.31	-0.18	-0.13
	3	deoxy	-0.83	-0.11	0.93		-0.78	0.37	-0.41	-0.55

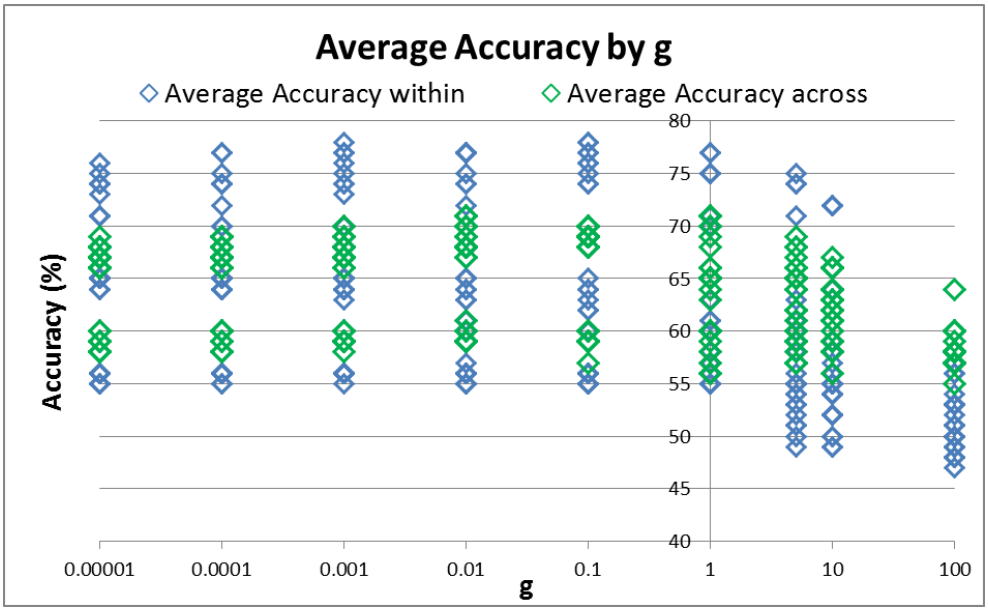
	4	deoxy	-1.29	-0.05	-0.85	-1.44	0.28	-0.45	-0.60
221	1	oxy	0.17	0.18	0.25	0.82	0.14	-0.58	0.24
	2	oxy	0.03	0.06	-0.07	0.23	-0.37	-0.71	-0.52
	3	oxy	0.32	0.33	0.14	0.08	-0.52	-1.00	0.01
	4	oxy	-0.08	-0.10	-0.13	-0.07	-0.58	-0.74	-0.31
	1	deoxy	**0.62**	0.36	0.13	-0.11	0.54	1.25	0.40
	2	deoxy	**1.48**	1.30	1.16	0.28	0.40	0.63	1.18
	3	deoxy	**0.64**	0.60	0.54	0.60	0.00	-0.15	0.60
	4	deoxy	**0.90**	0.74	0.74	-0.10	0.20	0.20	0.67
281	1	oxy	1.06	1.34	1.49	0.76	0.19	0.33	0.88
	2	oxy	0.97	1.44	1.49	0.53	0.23	0.28	0.79
	3	oxy	1.25	1.21	1.48	0.42	0.03	0.01	0.97
	4	oxy	1.27	1.22	1.31	0.58	0.10	-0.01	1.13
	1	deoxy	1.05	1.25	**1.59**	0.83	0.28	0.73	1.15
	2	deoxy	0.94	1.40	**1.56**	0.34	0.16	0.63	1.13
	3	deoxy	1.51	1.39	**1.50**	0.56	0.12	0.51	1.04
	4	deoxy	0.99	1.34	**1.51**	0.22	-0.15	0.19	0.99
		sum:	15.32	17.63	14.04	2.87	3.76	-1.22	8.05

Appendix C

Results of a test case varying c and g each across 8 orders of magnitude

Please see section 3.2.1 for context.



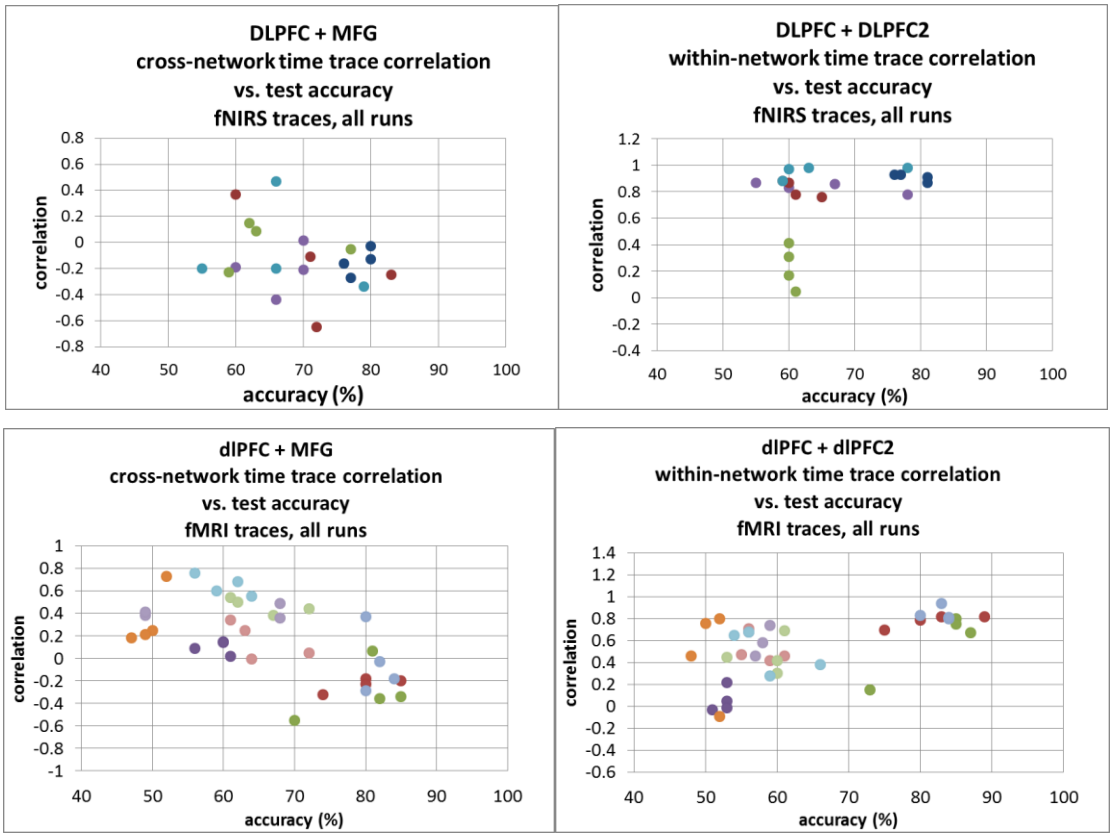


Appendix D

Correlation vs. classification accuracy, study part one

Colors indicate individual participants.

Please refer to section 3.6.3 for context.

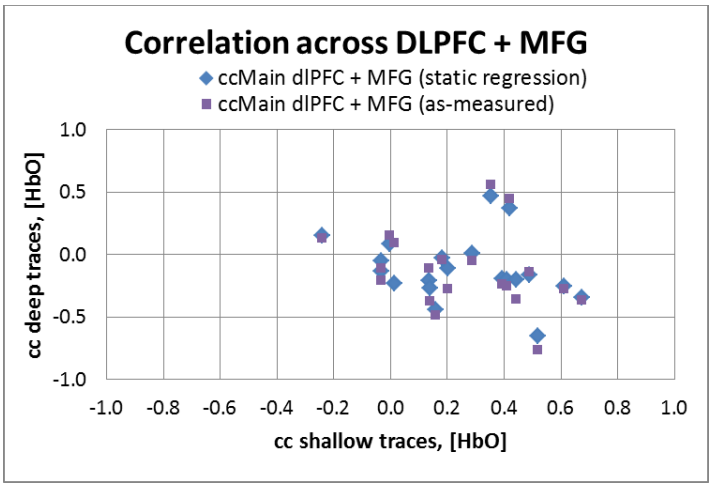


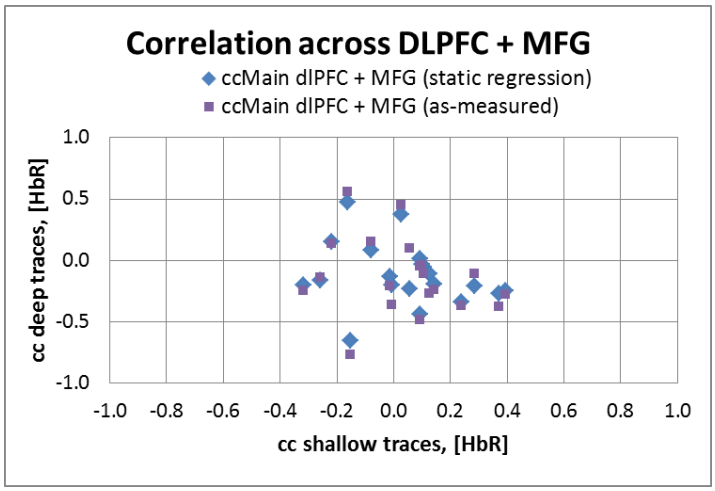
Appendix E

Correlation of deep vs. correlation of shallow traces, study part one

Negative correlations of shallow traces are not associated with negative correlations of deep traces, corrected (via static regression) or not (as-measured). Indeed, positive correlations of shallow traces, perhaps due to systemic physiology similarly influencing the shallow sensors at distant brain locations, may be weakly associated with positive correlations of the deep traces. Given this, we conclude we are not seeing an influence of anti-correlation from the shallow optodes.

Please see section 3.6.3 for context.



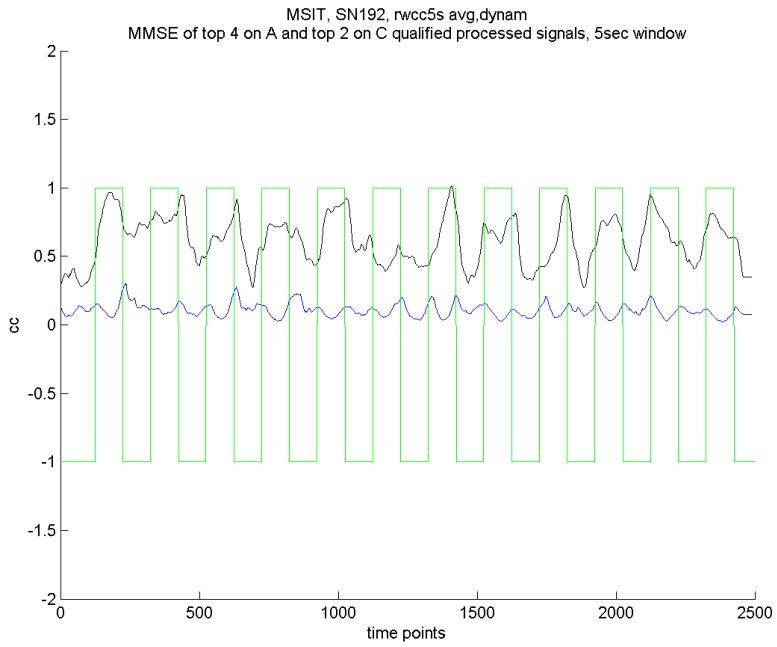


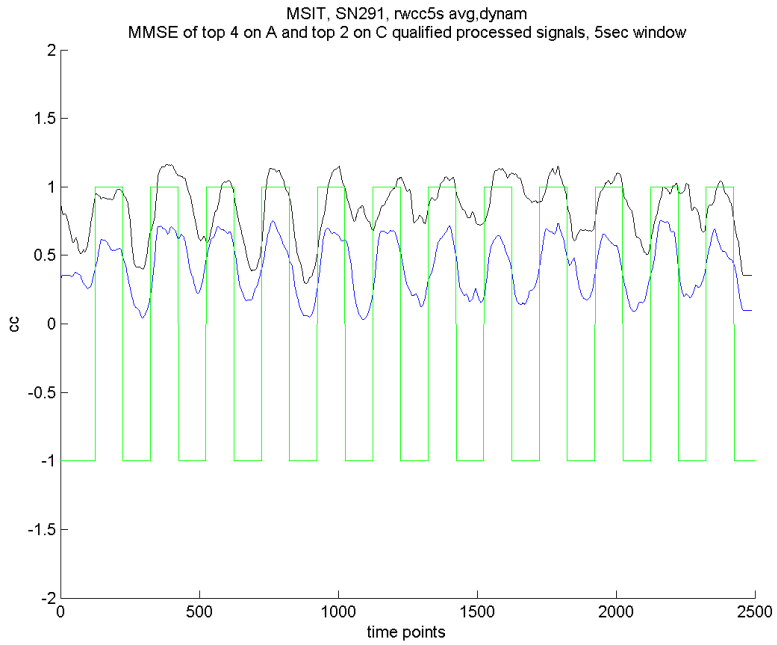
Appendix F

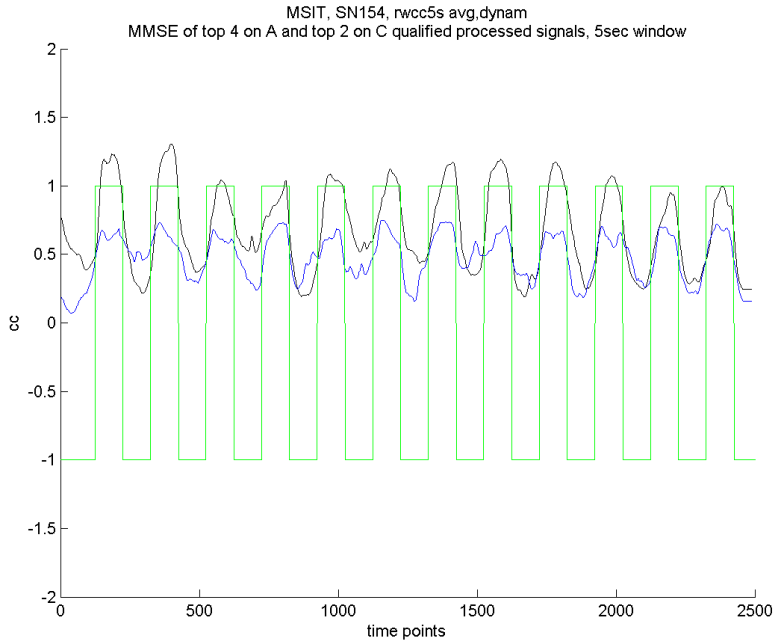
MMSE for each participant, adaptive regression case

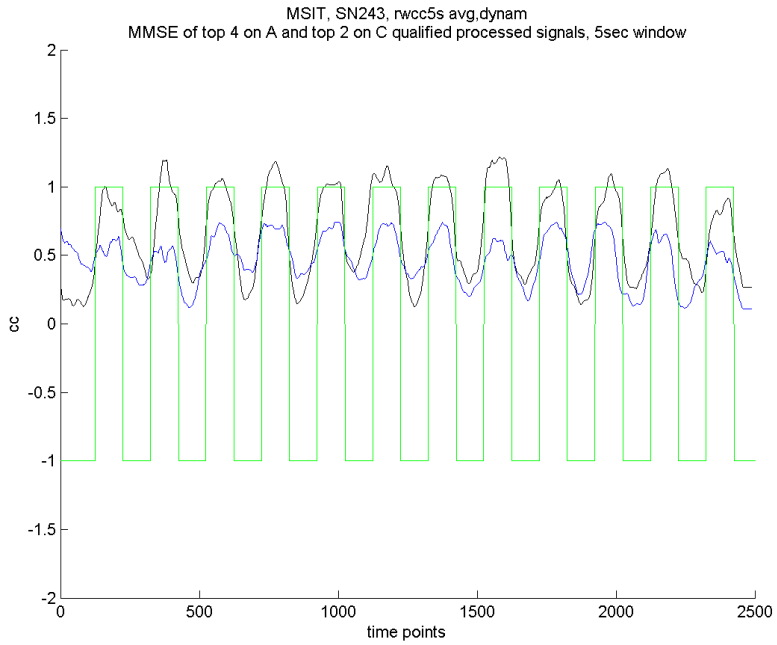
Functional traces were taken from the DLPFC (black) and the MFG (blue), during the MSIT. The task indicator signal is shown in green, with +1 indicating a work period of the MSIT.

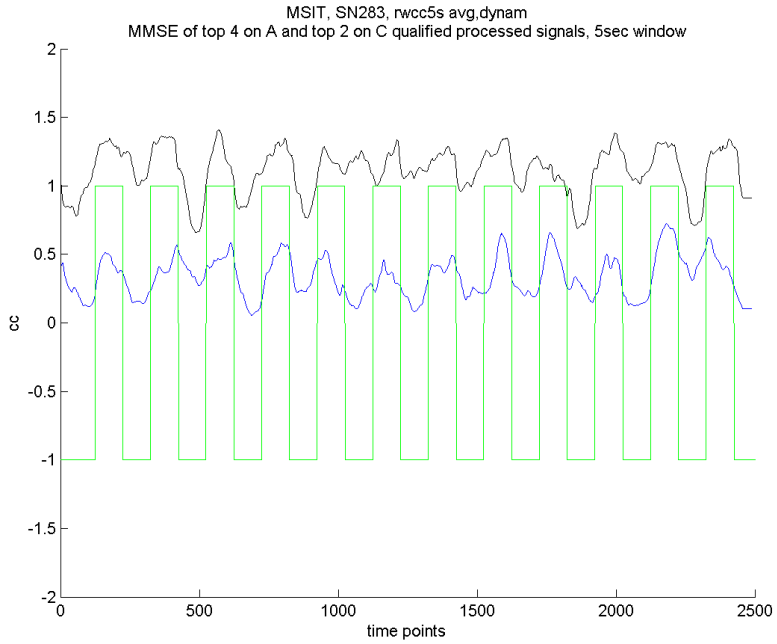
Please see section 4.4.8 for context.

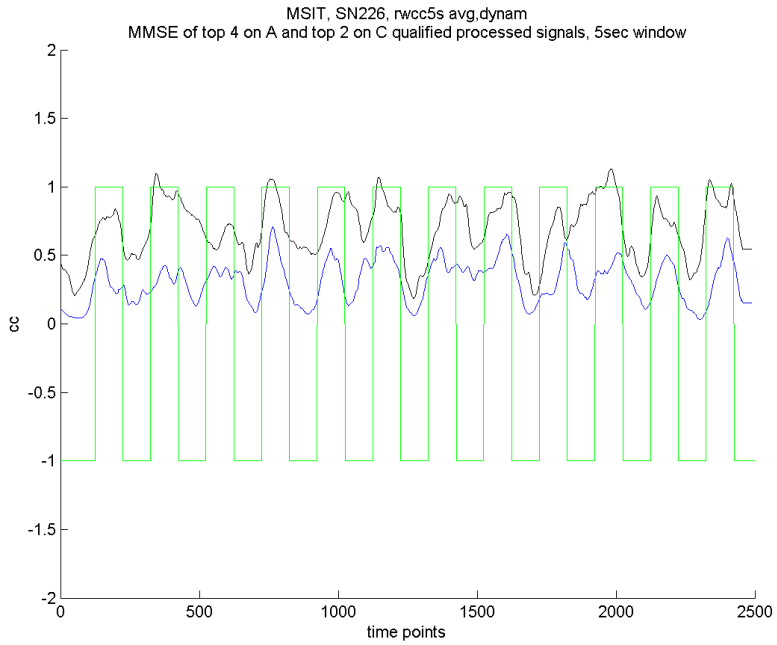


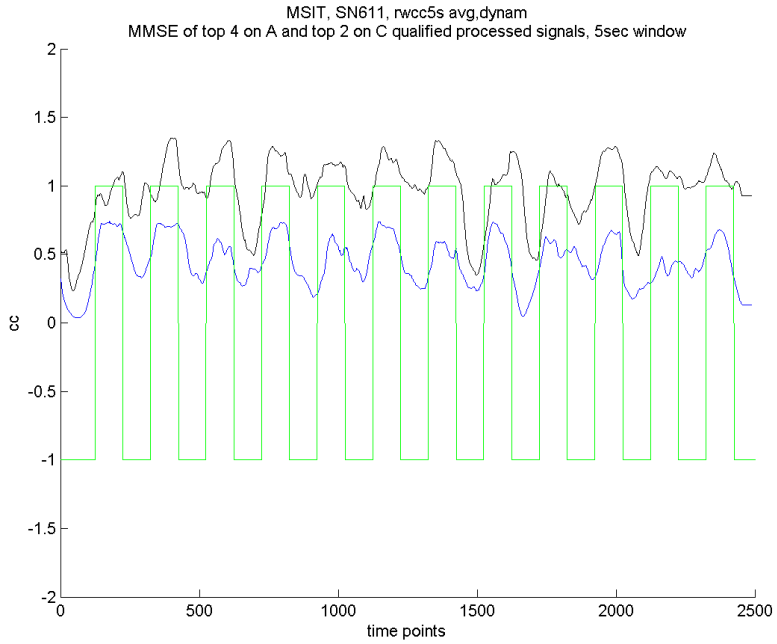

MSIT, SN291, rwcc5s avg,dynam
MMSE of top 4 on A and top 2 on C qualified processed signals, 5sec window
cc
time points
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
500
1000
1500
2000
2500


MSIT, SN154, rwcc5s avg,dynam
MMSE of top 4 on A and top 2 on C qualified processed signals, 5sec window
cc
time points
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
500
1000
1500
2000
2500


MSIT, SN243, rwcc5s avg,dynam
MMSE of top 4 on A and top 2 on C qualified processed signals, 5sec window
cc
time points
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
500
1000
1500
2000
2500


MSIT, SN283, rwcc5s avg,dynam
MMSE of top 4 on A and top 2 on C qualified processed signals, 5sec window
cc
time points
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
500
1000
1500
2000
2500

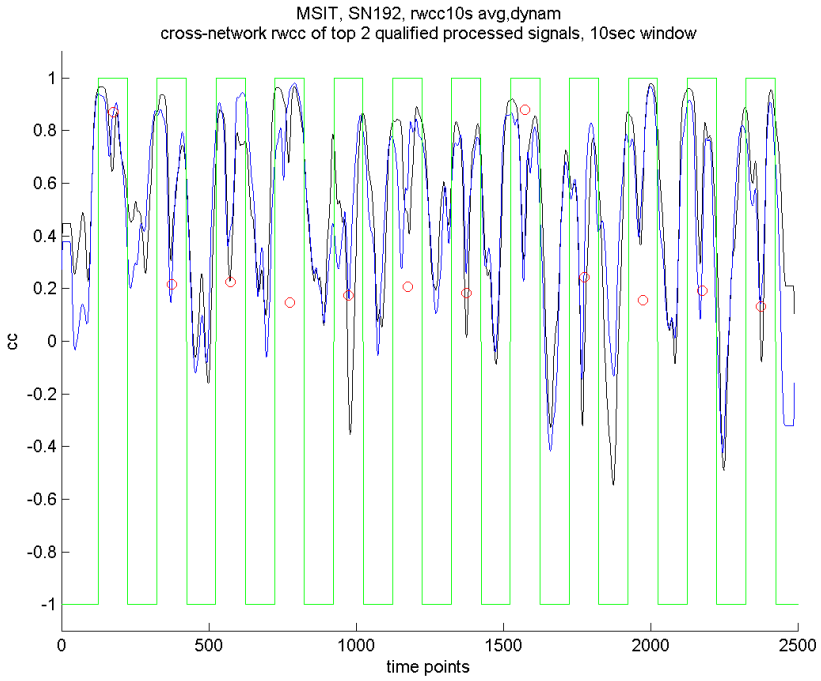

MSIT, SN226, rwcc5s avg,dynam
MMSE of top 4 on A and top 2 on C qualified processed signals, 5sec window
cc
time points
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
500
1000
1500
2000
2500

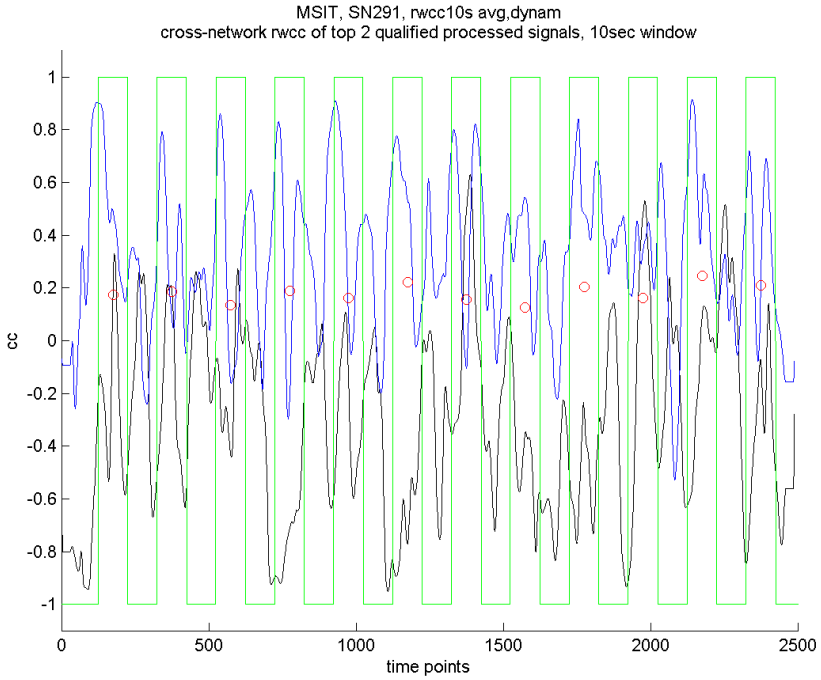

MSIT, SN611, rwcc5s avg,dynam
MMSE of top 4 on A and top 2 on C qualified processed signals, 5sec window
cc
time points
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
500
1000
1500
2000
2500

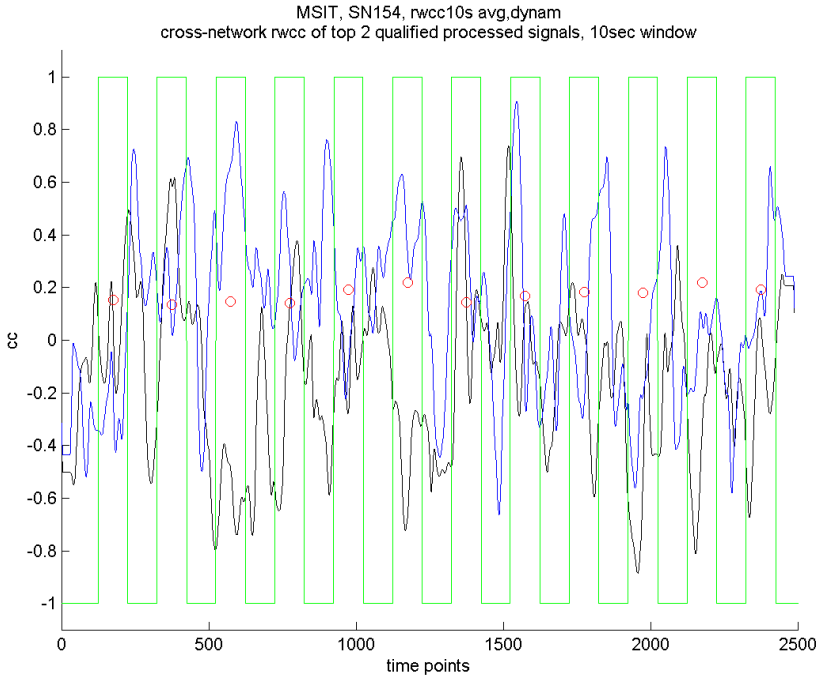
Appendix G

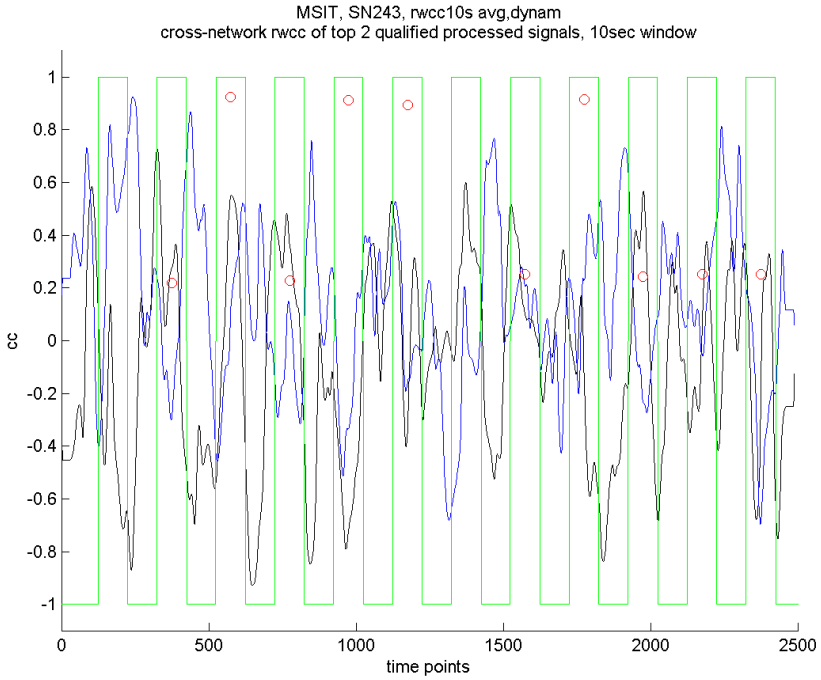
CVRT and RWCC for each participant, adaptive regression case

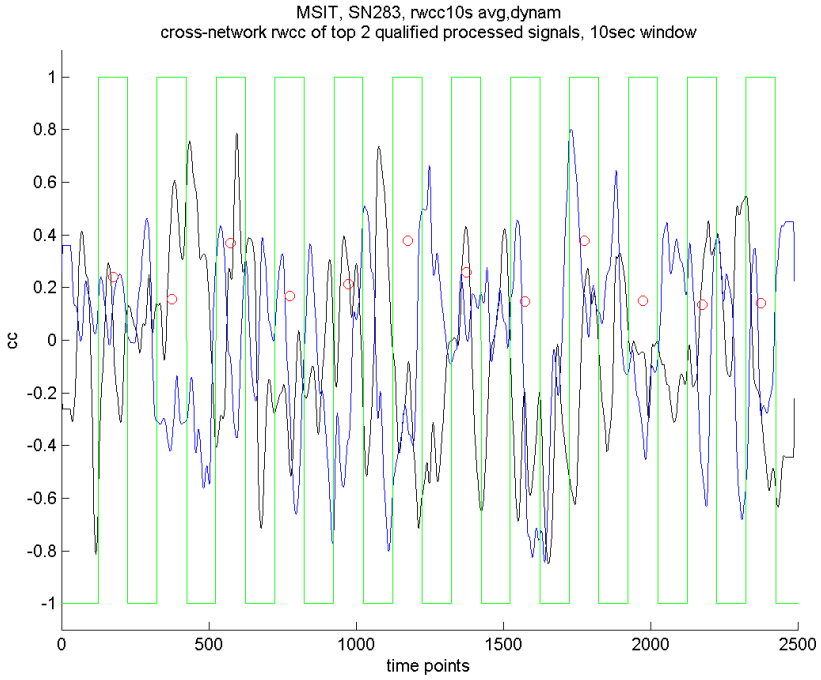
The top-ranked pair (black) and the second-ranked pair (blue) are shown. The task indicator signal is shown in green, with +1 indicating a work period of the MSIT. Please see section 4.4.9 for context.

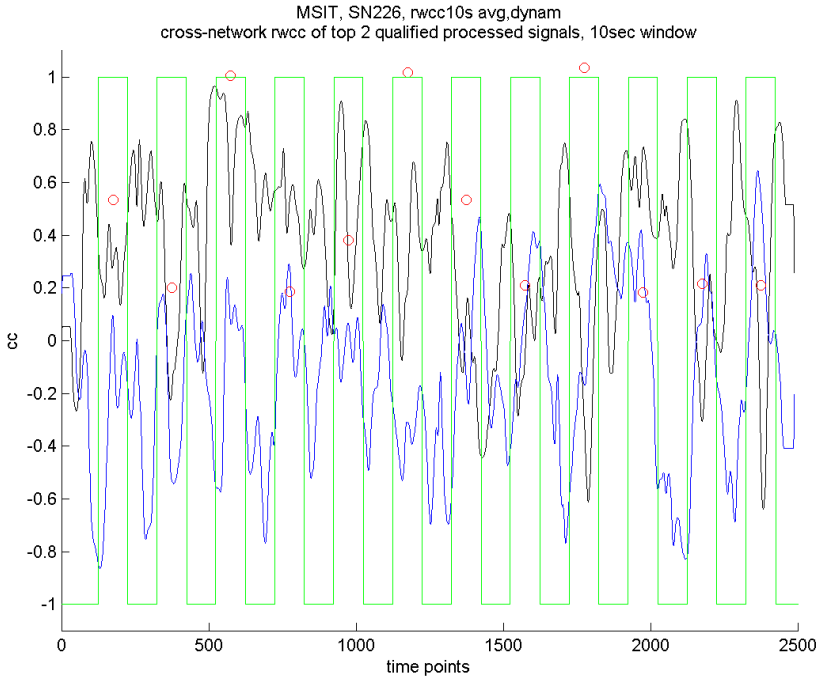


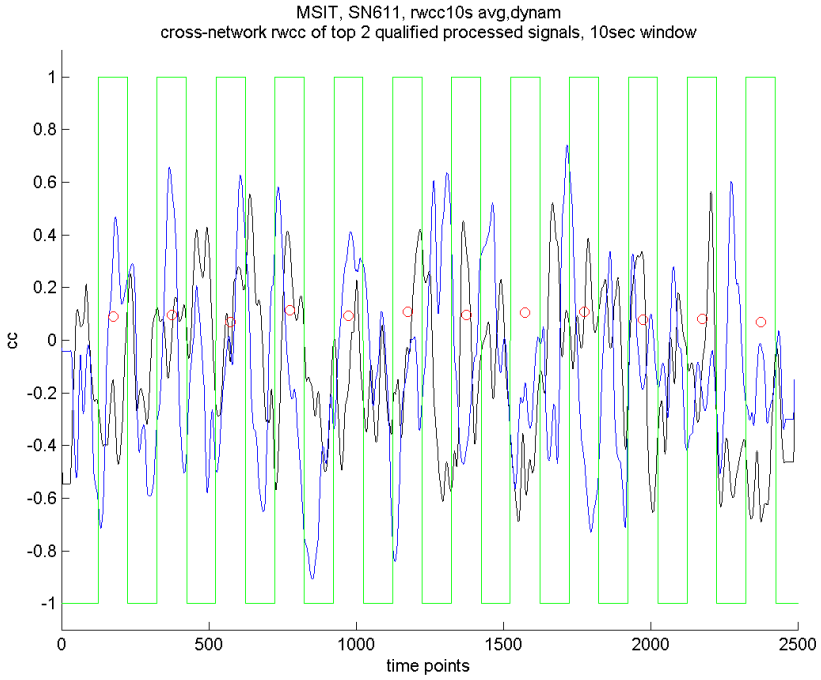

MSIT, SN291, rwcc10s avg,dynam
cross-network rwcc of top 2 qualified processed signals, 10sec window
cc
time points
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
500
1000
1500
2000
2500


MSIT, SN154, rwcc10s avg,dynam
cross-network rwcc of top 2 qualified processed signals, 10sec window
cc
time points
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
500
1000
1500
2000
2500


MSIT, SN243, rwcc10s avg,dynam
cross-network rwcc of top 2 qualified processed signals, 10sec window
cc
time points
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
500
1000
1500
2000
2500


MSIT, SN283, rwcc10s avg,dynam
cross-network rwcc of top 2 qualified processed signals, 10sec window
cc
time points
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
500
1000
1500
2000
2500


MSIT, SN226, rwcc10s avg,dynam
cross-network rwcc of top 2 qualified processed signals, 10sec window
cc
time points
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
500
1000
1500
2000
2500


MSIT, SN611, rwcc10s avg,dynam
cross-network rwcc of top 2 qualified processed signals, 10sec window
cc
time points
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
500
1000
1500
2000
2500

Appendix H

Differences between real space and normalized space outside the head volume

Differences between real space and normalized space MNI space were measured to quantify spatial differences near but outside the head region, due to the non-linearity of normalization field deformations in this region. For one participant, probe array marker-to-marker separation, and separations between the marker and the scalp in three slices (local separations), were measured for the normalized* and non-normalized anatomical overlay images. For local separations, average differences were 2.6 mm +/- 1.4mm in the PFC region and 2.9 mm +/- 2.1 in the MFG region. This is < 15% of the diameter of the sampling sphere. The non-local marker-to-marker distance differed by < 10 mm. Thus, local markers are necessary such that long marker-to-head distances do not need to be estimated. In this study, only separations local to one probe array (up to 15mm) were estimated.

*Affine registration to ICBM space template with non-linear regularization = 1, written in [2 2 2] mm voxels

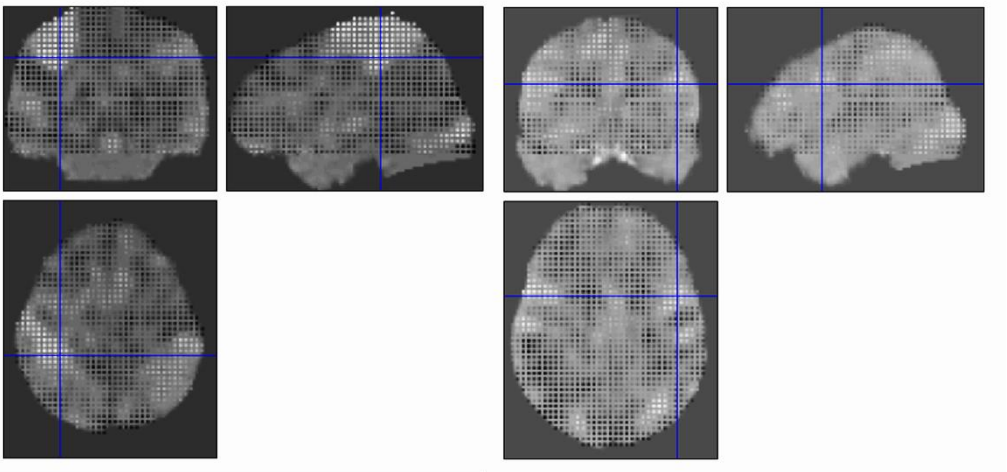
Euclidean distance in slice:	**normalized (mm)**	**not normalized (mm)**	**separation difference (mm)**	**% of separation**
PFC axial	22.4	23.2	1.7	4%
PFC sagittal	32.2	31.7	3.5	5%
PFC coronal	21.6	20.5	2.5	6%
PFC average			**2.6**	**5%**
MFG axial	24.8	27.0	2.6	5%
MFG sagittal	24.8	22.3	3.6	8%
MFG coronal	26.4	28.5	2.5	4%
MFG average			**2.9**	**6%**
marker-to-marker	**79.0**	**70.6**	**8.5**	**6%**

Appendix I

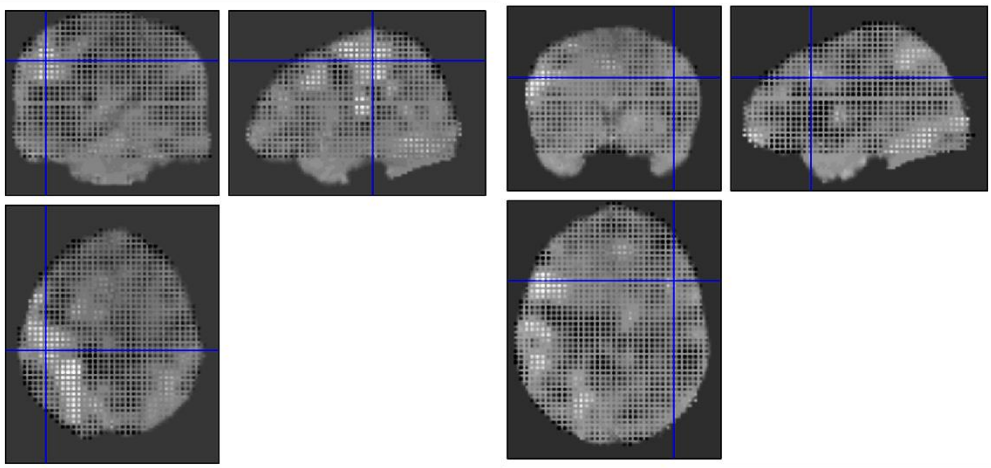
Weight maps for whole brain fMRI Linear Classification

Whole brain weight maps are shown for each participant on their anatomical overlay. The greater the absolute weight (white: positive, black: negative), the more influential that voxel's time trace was on the SVM model. Contrast has been increased for visualization. Please see section 4.6.3 for context.

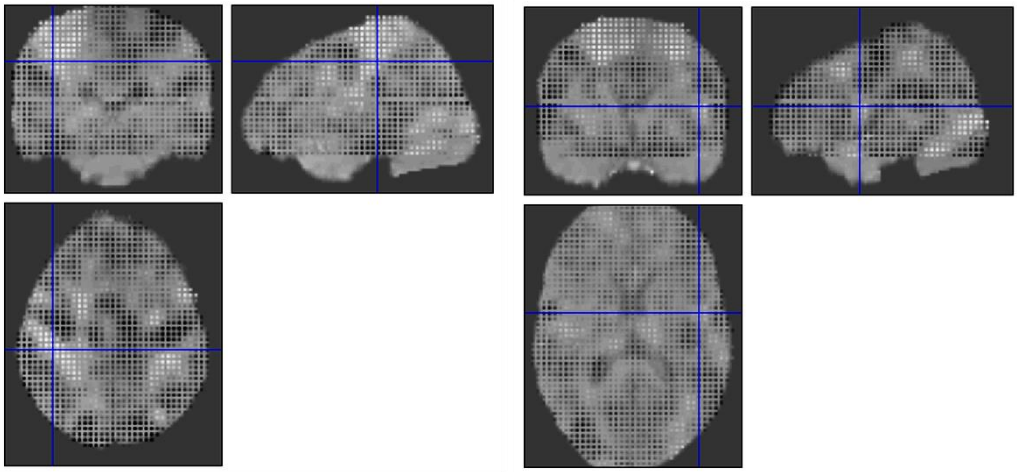
SN291:



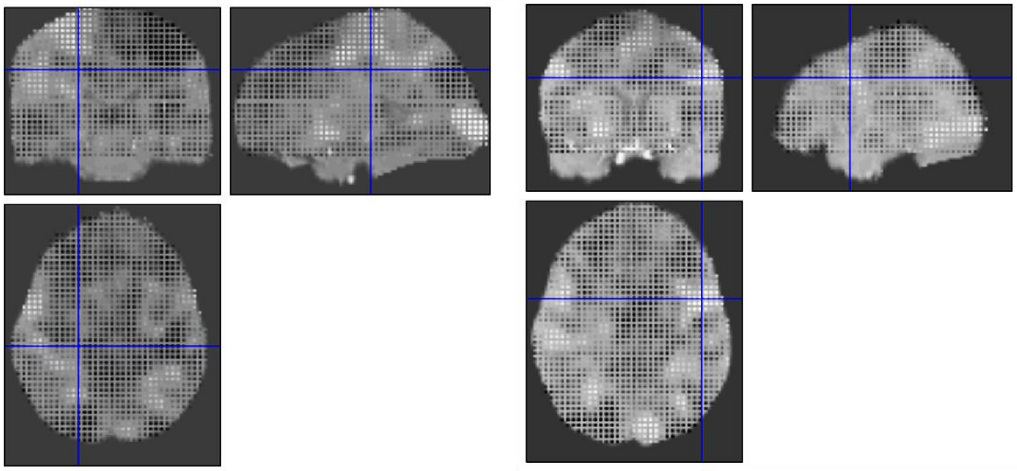
SN154:



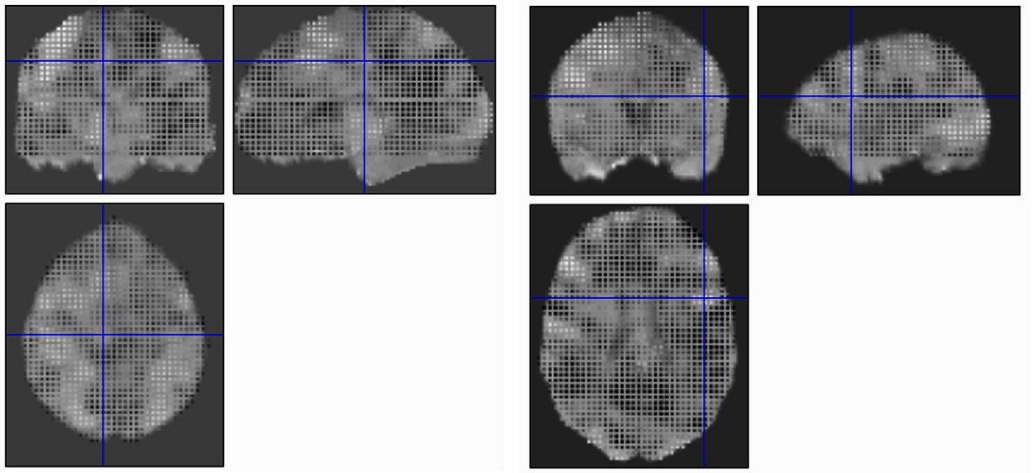
SN243:



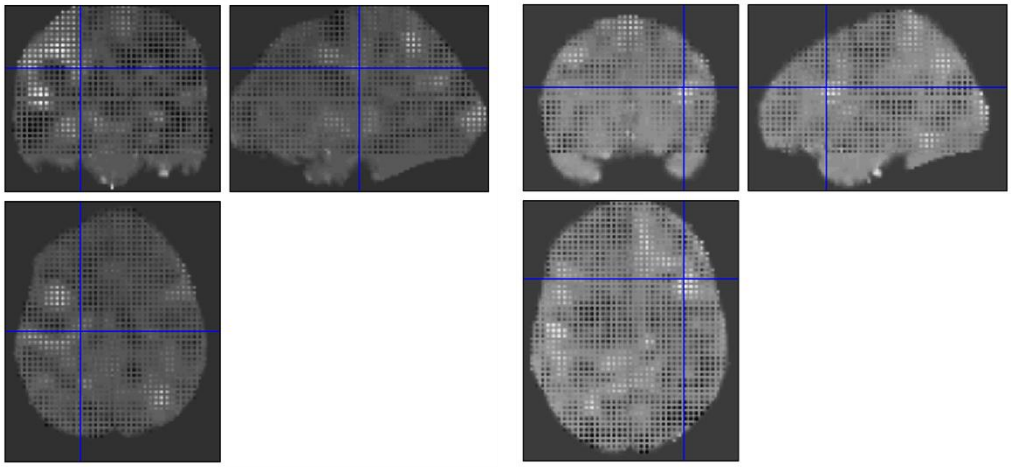
SN283:



SN226:



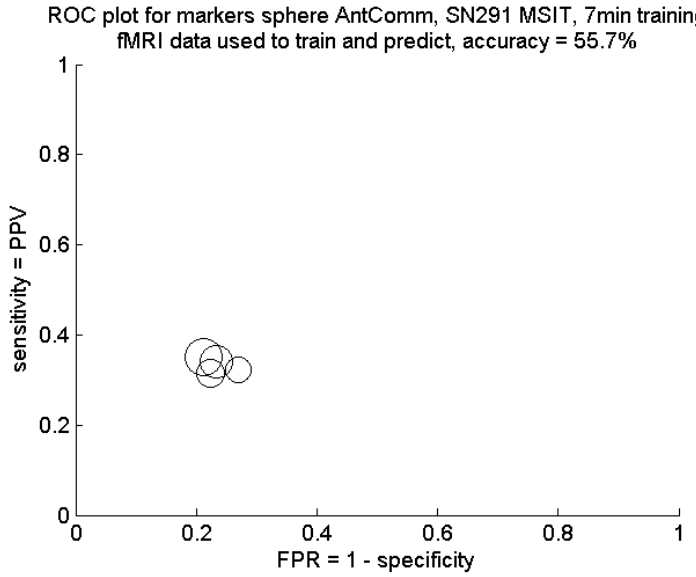
SN611:

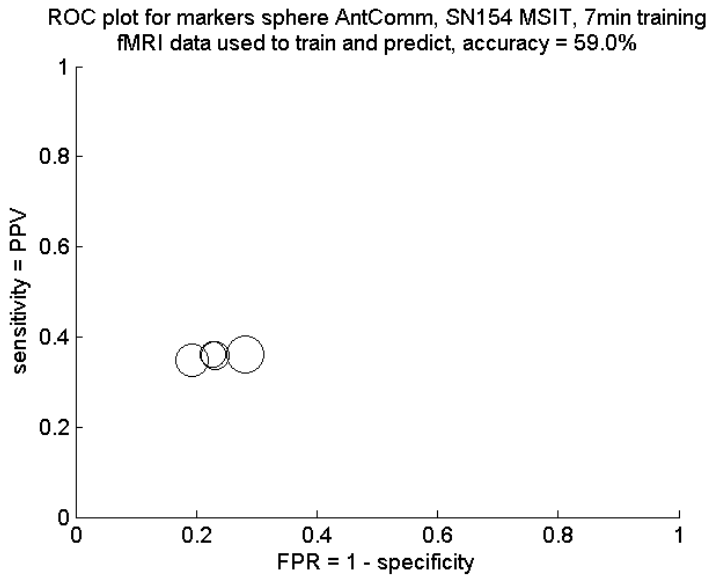


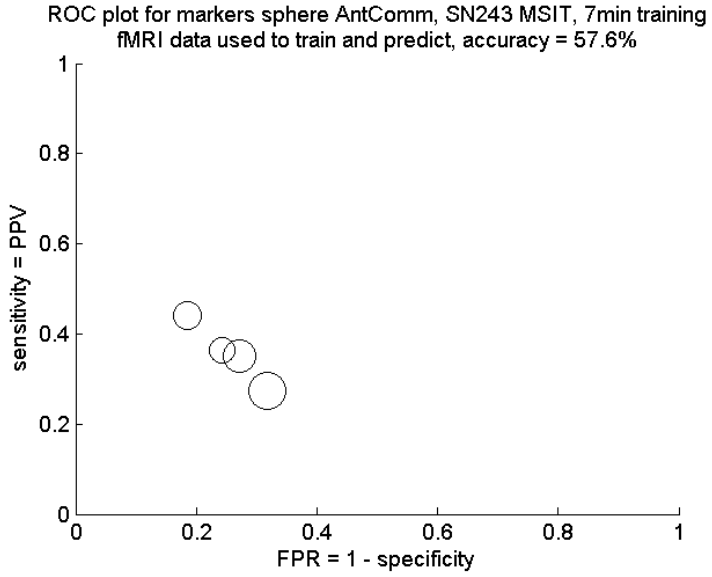
Appendix J

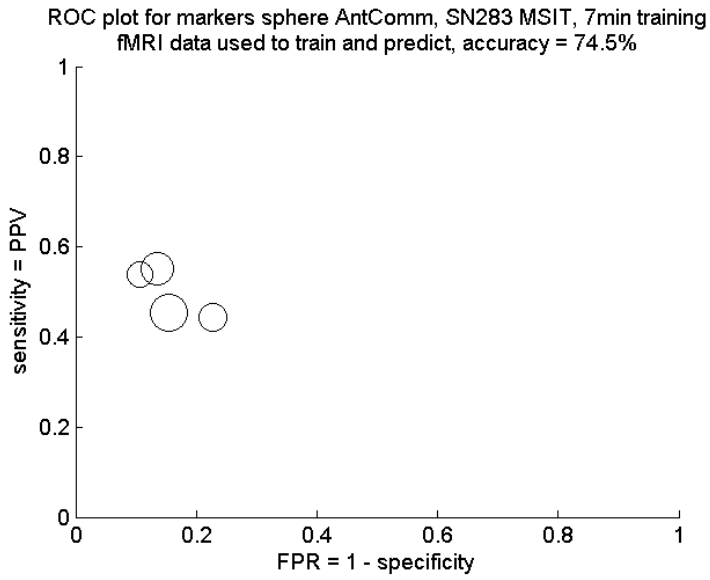
Receiver Operator Characteristic plots for fMRI Results

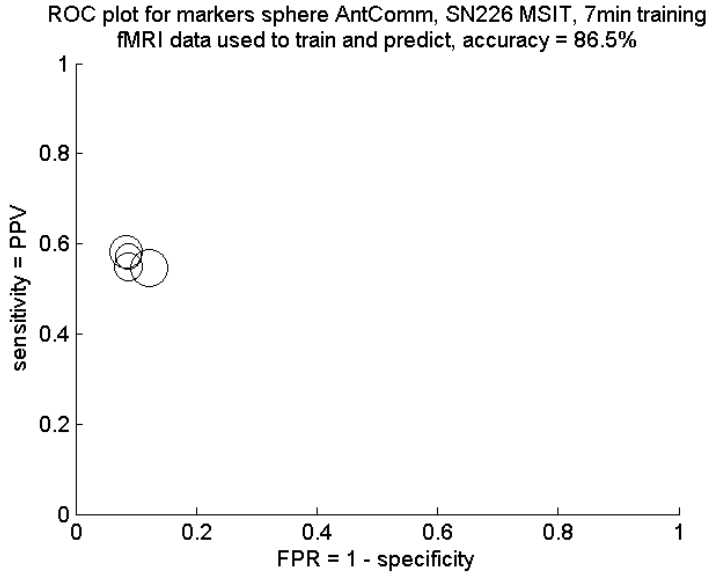
Receiver Operator Characteristic plots are shown for each participant for the sensitivity case and prediction with 7 minutes of training data. Marker size increases with run number. Please see section 4.7.3 for context.

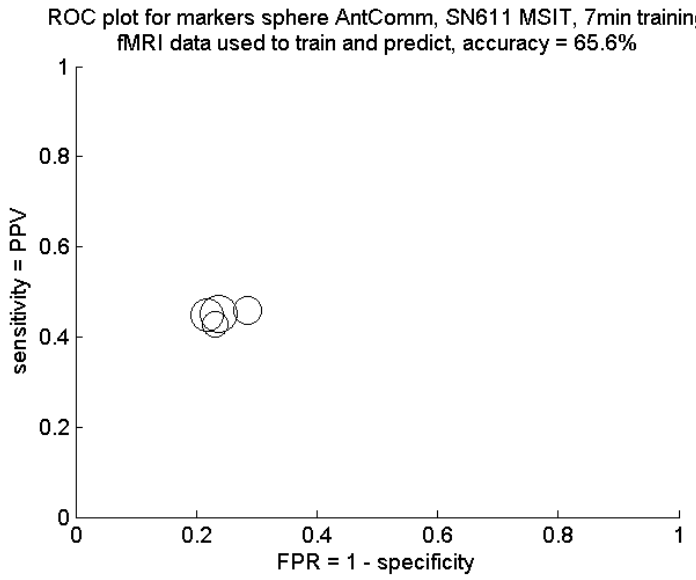



ROC plot for markers sphere AntComm, SN154 MSIT, 7min training
fMRI data used to train and predict, accuracy = 59.0%
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1


ROC plot for markers sphere AntComm, SN243 MSIT, 7min training
fMRI data used to train and predict, accuracy = 57.6%
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1


ROC plot for markers sphere AntComm, SN283 MSIT, 7min training
fMRI data used to train and predict, accuracy = 74.5%
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1

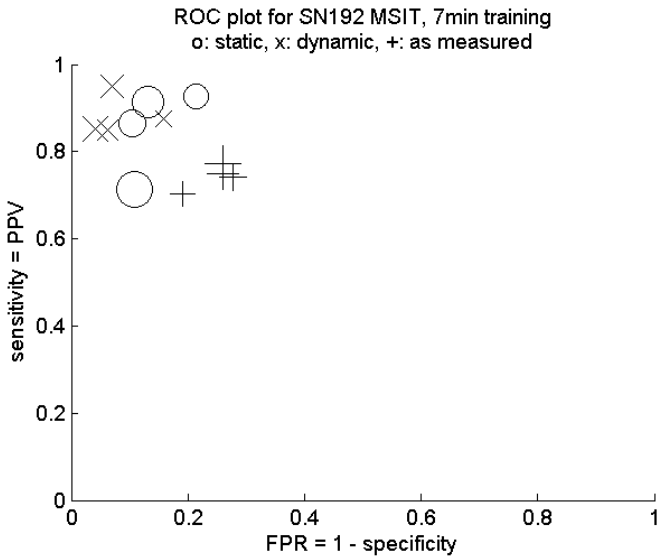

ROC plot for markers sphere AntComm, SN226 MSIT, 7min training
fMRI data used to train and predict, accuracy = 86.5%
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1

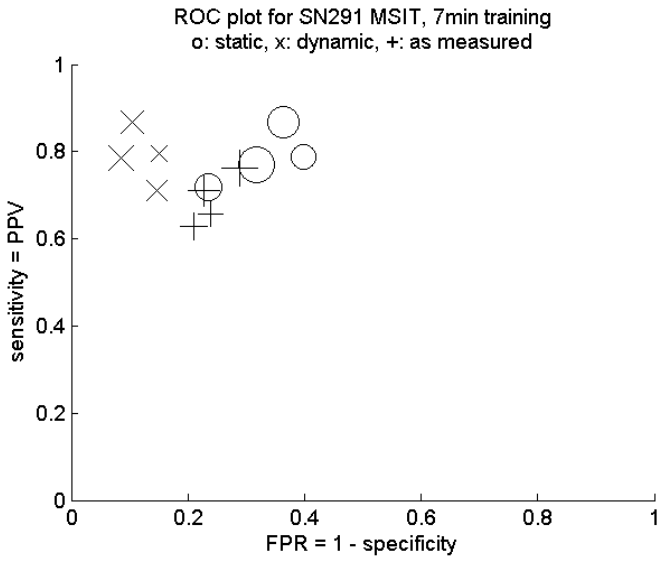

ROC plot for markers sphere AntComm, SN611 MSIT, 7min training
fMRI data used to train and predict, accuracy = 65.6%
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1

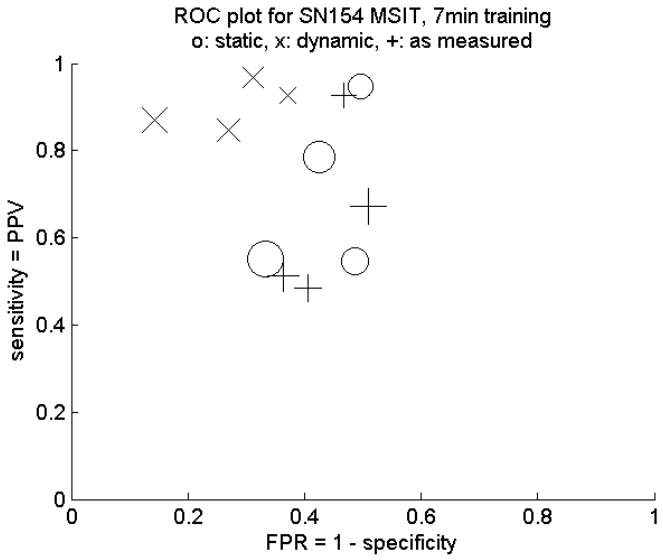
Appendix K

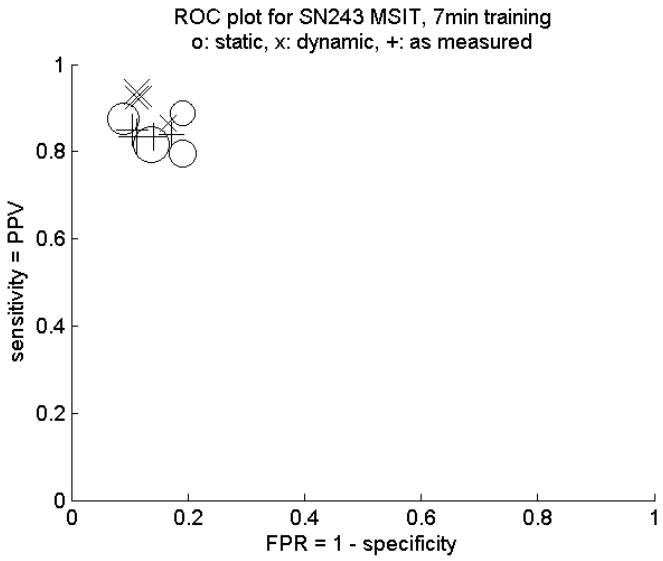
Receiver Operator Characteristic plots for fNIRS Results

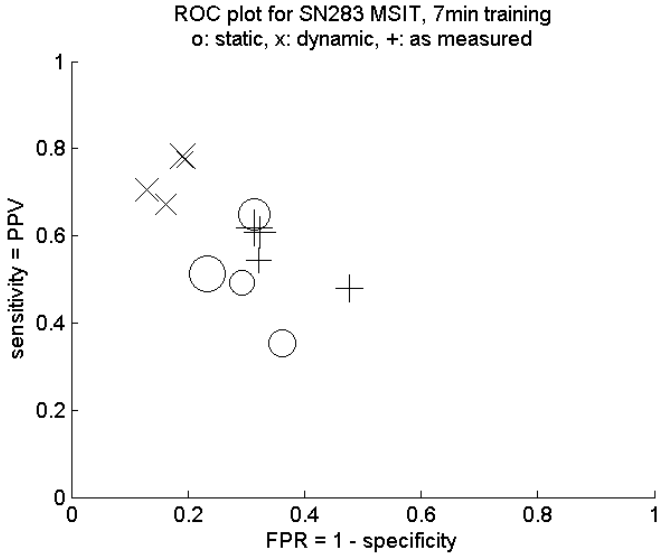
Receiver Operator Characteristic plots are shown for each participant for the static (o), dynamic (x) and as-measured (+) fNIRS processing cases, and prediction with 7 minutes of training data. Marker size increases with run number. Please see section 4.7.3 for context.

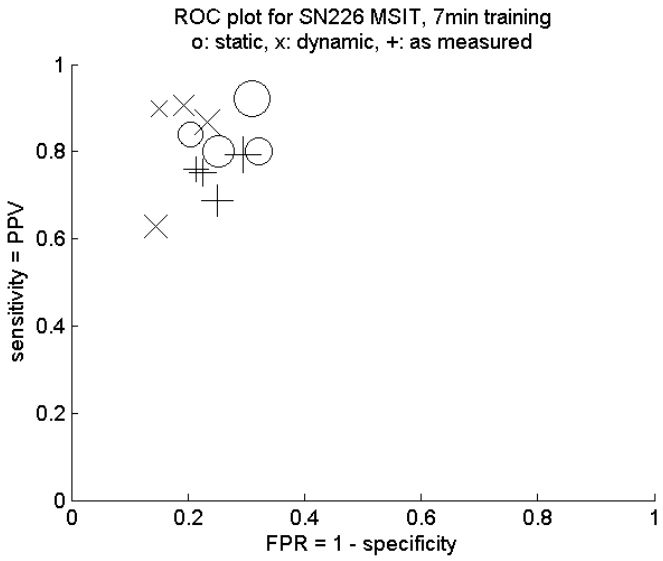


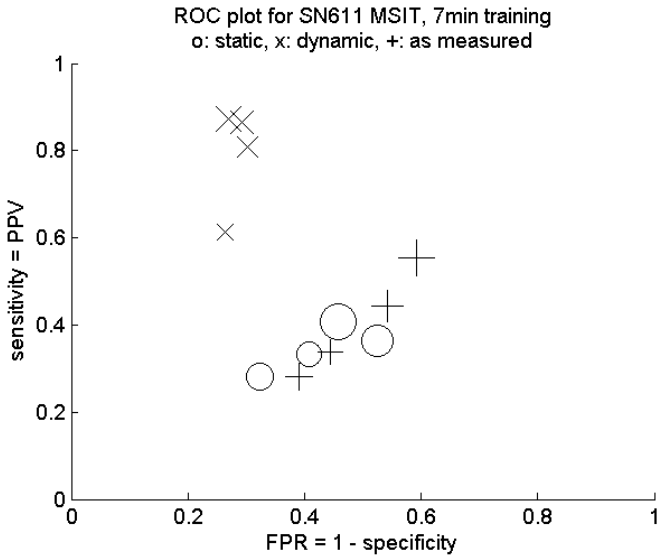

ROC plot for SN291 MSIT, 7min training
o: static, x: dynamic, +: as measured
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1


ROC plot for SN154 MSIT, 7min training
o: static, x: dynamic, +: as measured
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1


ROC plot for SN243 MSIT, 7min training
o: static, x: dynamic, +: as measured
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1


ROC plot for SN283 MSIT, 7min training
o: static, x: dynamic, +: as measured
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1


ROC plot for SN226 MSIT, 7min training
o: static, x: dynamic, +: as measured
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1


ROC plot for SN611 MSIT, 7min training
o: static, x: dynamic, +: as measured
sensitivity = PPV
FPR = 1 - specificity
0
0.2
0.4
0.6
0.8
1

References

Abdelnour, A. F. and Huppert, T. (2009). Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model. *Neuroimage*. 46:1, 133-143.

Ahmed, M. U., and Mandic, D. P. (2011). Multivariate multiscale entropy: A tool for complexity analysis of multichannel data. *Physical Review E*. 84, 061918. DOI: 10.1103/PhysRevE.84.061918

Aqil, M., Hong, K., Jeong, M., Ge, S. S. (2012). Cortical brain imaging by adaptive filtering of NIRS signals. *Neuroscience Letters*. 514:1, 35–41.

Arenth, P. M., Ricker, J. H., and Schultheis, M. T. (2007). Applications of Functional Near-Infrared Spectroscopy (fNIRS) to neurorehabilitation of cognitive disabilities. *Clinical Neuropsychologist*. 21:1, 38-57.

Arridge, S. R., Cope, M., Delpy, D. T., (1992). The theoretical basis for the determination of optical pathlengths in tissue: temporal and frequency analysis. *Phys. Med. Biol.* 37:7, 1531-1560.

Biswal, B., Yetkin, F. Z., Haughton, V. M., and Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn. Reson. Med.* 34:4, 537–541.

Boas, D.A., Dale, A.M., and Franceschini M.A. (2004). Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *Neuroimage*. 23, S275–S288.

Boas, D. A., Gaudette, T., Strangman, G., Cheng, X., Marota, J. J. A., and Mandeville, J. B. (2001). The accuracy of near infrared spectroscopy and imaging during focal changes in cerebral hemodynamics. *Neuroimage*. 13:1, 76-90.

Brigadoi, S., Ceccherini, L., Cutini, S., Scarpa, F., Scatturin, P., Selb, J., Gagnon, L., Boas, D. A., and Cooper, R. J. (2014). Motion artifacts in functional near-infrared spectroscopy: A

comparison of motion correction techniques applied to real cognitive data. *Neuroimage*. 85:1, 181-191. doi: 10.1016/j.neuroimage.2013.04.082

Bunce, S., Izzetoglu, M., Izzetoglu, K., Onaral, B., and Pourrezaei, K. (2006). Functional near-infrared spectroscopy - An emerging neuroimaging modality. *IEEE Engineering in Medicine and Biology Magazine*. 25:4, 54-62.

Bush, G., and Shin, L. (2006). The Multi-Source Interference Task: an fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. *Nature Protocols*. 1:1, 308-313.

Buxton, R. (2002). *Introduction to Functional Magnetic Resonance Imaging: Principles and Techniques*. New York City, NY: Cambridge University Press.

Callaway, E., *Tiny scanner may monitor astronauts' mental health.* 03 July 2008. Available from: http://space.newscientist.com/article/dn14247-brain-breathalysers-may-scan-astronauts-for-stress.html.

Cannestra, A. F., Pouratian, N., Bookheimer, S. Y., Martin, N. A., Becker, D. P., and Toga, A. W. (2001). Temporal spatial differences observed by functional MRI and human intraoperative optical imaging. *Cerebral Cortex*. 11:8, 773-782.

Chang, C., and Glover, G. H. (2009). Effects of model-based physiological noise correction on default mode network anti-correlations and correlations. *Neuroimage.* 47:4, 1448–1459. doi: 10.1016/j.neuroimage.2009.05.012.

Chang, C., and Glover, G. H. (2010). Time–frequency dynamics of resting-state brain connectivity measured with fMRI. *Neuroimage.* 50:1, 81–98.

Chang, C.C., and Lin, C.J. (2008). Feature Ranking Using Linear SVM. JMLR: Workshop and Conference Proceedings. 3: 53-64.

Chang, C.C., and Lin, C.J. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 2:27, 1-27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

Chee, M.W.L., Tan, J.C., Zheng, H., Parimal, S., Weissman, D.H., Zagorodnov, V. and Dinges, D.F. (2008). Lapsing during Sleep Deprivation Is Associated with Distributed Changes in Brain Activation. *Journal of Neuroscience*. 28:21, 5519-5528.

Christoff, K., Gordonb, A.M., Smallwood, J., Smith, R., and Schooler, J.W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. PNAS. 106:21, 8719-8724.

Çiftçi, K., Kahya, Y. P., Sankur, B., Akin, A. (2005). Complexity Analysis of Functional Near Infrared Spectroscopy Signals. 13th European Signal Processing Conference (EUSIPCO 2005), 4-8 September 2005, Antalya, Turkey.

Collins, D.L., Neelin, P., Peters, T.M., and Evans, A.C. (1994). Automatic 3D interparticipant registration of MR volumetric data in standardized Talairach space. *Journal of Computer Assisted Tomography*. 18, 192– 205.

Cooper, R. J., Selb, J., Gagnon, L., Phillip, D., Schytz, H. W., Iversen, H. K., Ashina, M., and Boas D. A. (2012). A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy. Front. Neurosci. 6:147. doi: 10.3389/fnins.2012.00147.

Cortes, C., Vapnik, V. (1995). Support-vector networks. *Machine Learning*. 20:3, 273. doi:10.1007/BF00994018

Costa, M.,. (2005). Multiscale entropy analysis of biological signals. *Physical Review E*. 71, 021906.

Cowings, P.S., Toscano, W.B., Taylor, B., DeRoshia, C.W., Kornilova, L.. Kozlovskya, I. and Miller, N.E. (2003). Psychophysiology of Spaceflight. *14th IAA Humans In Space Symposium, Living in Space: Scientific, Medical and Cultural Implications*. Banff, Alberta, Canada.

Coyle, S.M., Ward, T.E. and Markham C.M., (2007). Brain-computer interface using a simplified functional near-infrared spectroscopy system. *Journal of Neural Engineering*. 4:3, 219-226.

Deco, G., Jirsa, V. K., McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nat Rev Neurosci.* 12:43–56.

De Havas, J.A., Parimal, S., Soon, C.S., Chee, M.W.L. (2012). Sleep deprivation reduces default mode network connectivity and anti-correlation during rest and task performance. *NeuroImage*, 59, 1745–1751.

Delpy, D.T., Cope, M., van der Zee, P., Arridge, S., Wray, S., and Wyatt, J. (1998). Estimation of optical pathlength through tissue from direct time of flight measurement. *Physics in Medicine and Biology*, 33:12, 1433-1442.

Diamond, S. G., Huppert, T. J., Kolehmainen, V., Franceschini, M. A., Kaipio, J. P., Arridge, S. R., Boas, D. A. (2006). Dynamic physiological modeling for functional diffuse optical tomography. *NeuroImage*. 30:1, 88-101.

Dismukes, R.K., Berman, B.A., Loukopoulos, L.D. (2007). *The Limits of Expertise: Rethinking Pilot Error and the Causes of Airline Accidents (Ashgate Studies in Human Factors for Flight Operations)*. Burlington, VT: Ashgate Publishing Company.

Dosenbach, N.U.F., Visscher, K.M., Palmer, E.D., Miezin, F.M., Wenger, K.K., Kang, H.C., Burgund, E.D., Grimes, A.L., Schlaggar, B.L., and Petersen, S.E. (2006). A core system for the implementation of task sets. *Neuron*. 50:5, 799-812.

Drummond, S. P., Meloy, M. J., Yanagi, M. A., Orff, H. J., and Brown, G. G. (2005a) Compensatory recruitment after sleep deprivation and the relationship with performance. *Psychiatry Research-Neuroimaging*. 140:3, 211-223.

Drummond S.P.A., Bischoff-Grethe, A., Dinges, D.F., Ayalon, L., Mednick, S.C., and Meloy, M. J. (2005b). The neural basis of the psychomotor vigilance task. Sleep. 28:9, 1059-1068.

Duncan, A., Meek, J. H., Clemence, M., Elwell, C.E., Fallon, P., Tyszczuk, L., Cope, M., and Delpy, D. T. (1996). Measurement of cranial optical path length as a function of age using phase resolved near infrared spectroscopy. *Pediatric Research.* 39:5, 889-894.

Eggebrecht, A. T., Ferradal, S. L., Robichaux-Viehoever, A., Hassanpour, M. S., Dehghani, H. Snyder, A. Z. Hershey, T. and Culver, J. P. (2014). Mapping distributed brain function and networks with diffuse optical tomography. *Nature Photonics*. *Published online 18 May 2014*. doi:10.1038/nphoton.2014.107

Eichele, T., Debener, S., Calhoun, V. D., Specht, K., Engel, A. K., Hugdahl, K., von Cramon, D. Y., and Ullsperger, M. (2008). Prediction of human errors by maladaptive changes in event-related brain networks. *Proceedings of the National Academy of Sciences of the United States of America.* 105:16, 6173-6178.

Emir, U.E., Ozturk, C. and Akin, A. (2008). Multimodal investigation of fMRI and fNIRS derived breath hold BOLD signals with an expanded balloon model. *Physiological Measurement*. 29:1, 49-63.

Fair, D. A., Schlaggar, B. L., Cohen, A. L., Miezin, F. M., Dosenbach, N. U., Wenger, K. K., Fox, M. D., Snyder, A. Z., Raichle, M. E., and Petersen, S. E. (2007). A method for using blocked and event-related fMRI data to study "resting state" functional connectivity. *NeuroImage*. 35:1, 396–405

Fairclough, S., and Gilleade, K. (2013) Capturing user engagement via psychophysiology: measures and mechanisms for biocybernetic adaptation. *International Journal of Autonomous and Adaptive Communications Systems*. 6:1, 63–79.

Fan, J., McCandliss, B. D., Fossella J., Flombaum J. I., and Posner, M. I. (2005). The activation of attentional networks. *Neuroimage*. 26:2, 471-479.

Franceschini, M.A., and Boas, D.A., 2004. Noninvasive measurement of neuronal activity with near-infrared optical imaging. *NeuroImage*. 21, 372– 386.

Franceschini M. A., Joseph, D. K., Huppert, T. J., Diamond, S. G., Boas, D. A. (2006). Diffuse optical imaging of the whole head. *Journal of Biomedical Optics*. 11:5, 054007.

Fransson, P. (2006). How default is the default mode of brain function? Further evidence from intrinsic BOLD signal fluctuations. *Neuropsychologia*. 44:14, 2836-45.

Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., and Raichle, M.E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *PNAS*. 102:27, 9673-9678.

Fox, M. D. and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci*. 8:9, 700–711.

Gagnon, L., Perdue, K., Greve, D. N., Goldenholz, D., Kaskhedikar, G., and Boas, D. (2011). Improved recovery of the hemodynamic response in diffuse optical imaging using short optode separations and state-space modeling. *NeuroImage.* 56, 1362–1371.

Gagnon, L., Cooper, R. J., Yücel, M. A., Perdue, K. L., Greve, D. N., and Boas, D. A., (2012a). Short separation channel location impacts the performance of short channel regression in NIRS. *NeuroImage.* 59:3, 2518–2528.

Gagnon, L., Yücel, M. A., Dehaes, M., Cooper, R. J., Perdue, K. L., Selb, J., Huppert, T. H., Hoge, R. D., Boas, D. A. (2012b). Quantification of the cortical contribution to the NIRS

signal over the motor cortex using concurrent NIRS–fMRI measurements. doi: 10.1016/j.neuroimage.2011.08.095.

Gazzaniga, M. S. (1989). Organization of the human brain. *Scienc.* 245:4921, 947-952.

Gibson, A. P., Hebden, J. C., and Arridge, S. R. (2005). Recent advances in diffuse optical imaging. Physics in Medicine and Biology. 50:4, R1-R43.

Gilbert, S. J. and Burgess, P. W. (2008). Executive function. *Current Biology*. 18:3, R110-R114.

Glover, G. H., Li, T. Q., and Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine*. 44:1, 162-167.

Gratton, G., Fabiani, M., Corballis, P. M., Hood, D.C., Goodman-Wood, M.R., Hirsch, J., Kim, K., Friedman, D., Gratton, E., (1997). Fast and localized event-related optical signals (EROS) in the human occipital cortex: comparisons with the visual evoked potential and fMRI. *Neuroimage*. 6, 168– 180.

Gratton, E., Toronov, V., Wolf, U., Wolf, M., Webb, A. (2005). Measurement of brain activity by near-infrared light. *Journal of Biomedical Optics,* 10:1, 011008-1-13.

Greicius, M. D., Krasnow, B., Reiss, A. L., and Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *PNAS.* 100:1, 253-258.

Gruberger, M., Ben-Simon, E., Levkovitz, Y., Zangen, A., Hendler, T. (2011). Towards a neuroscience of mind-wandering. *Front Hum Neurosci.* 5:56. doi: 10.3389/fnhum.2011.00056.

Harrivel, A. R., Hearn, T., Carp, J., Weissman, D. H., and Peltier, A. J. (2011). Monitoring attentional state using functional near infrared spectroscopy: A pilot study. *17th Annual Meeting of the Organization for Human Brain Mapping*. Quebec City, Canada.

Harrivel, A. R., and Hearn, T. (2012). Functional Near Infra-red Spectroscopy: Watching the brain in flight. *4th International Conference on Applied Human Factors and Ergonomics*. San Francisco, CA. LEW-18952-1, *patent pending*

Harrivel, A. R., Hernandez-Garcia, L., Peltier, S., and Noll, D. (2012a). Artifact removal for assessment of cross-network anticorrelation and attentional classification with fNIRS. *2nd Biennial fNIRS Meeting*. London, UK.

Harrivel, A. R., Hylton, A., Hearn, T. (2012b). Best Practices for the Operational Application of fNIRS. July, 2012, NASA/TM-2012-217615.

Harrivel, A. R., McKay, T., Hylton, A., King, J., Latorella, K., Peltier, S. J., and Noll, D. C. (2009). Toward Improved Headgear for Monitoring with Functional Near Infrared Spectroscopy, *NeuroImage*, 47: S141. *15th Annual Meeting of the Organization for Human Brain Mapping*, June 2009, San Francisco, CA. NASA LEW-18280-1.

Harrivel A. R., Weissman, D. H., Noll, D. C. and Peltier, S. J. (2013) Monitoring attentional state with fNIRS. *Front. Hum. Neurosci*. 7:861. doi: 10.3389/fnhum.2013.00861

Hatakenaka M., Miyai I., Mihara, M., Sakoda, S., and Kubota, K. (2007). Frontal regions involved in learning of motor skill - A functional NIRS study. *Neuroimage*. 34:1, 109-116.

Heekeren, H. R., Obrig, H., Wenzel, R., Eberle, K., Ruben, J., Villringer, K., Kurth, R., and Villringer, A. (1997). Cerebral haemoglobin oxygenation during sustained visual stimulation—A near-infrared spectroscopy study. *Philos. Trans. R. Soc. Lond., B Biol. Sci.* 352, 743– 750.

Hillman, E. M., Devor, A., Bouchard, M. B., Dunn, A. K., Krauss, G. W., Skoch, J., Bacskai, B. J., Dale, A. M., and Boas, D. A. (2007). Depth-resolved optical imaging and microscopy of vascular compartment dynamics during somatosensory stimulation. *Neuroimage*. 35:1, 89-104.

Hirshfield, L. M. Gulotta, R., Hirschfield, S., Hincks, S., Russel, M., Ward, R., Willians, T., Jacob, R. (2011). This is Your Brain on Interfaces: Enhancing Usability Testing with Functional Near-Infrared Spectroscopy. *ACM CHI 2011, Session: Brain & Bio-sensor Interactions,* 372-383. Vancouver, BC, Canada.

Hoshi, Y., Shimada, M., Sato, C., Iguchi, Y. (2005). Reevaluation of near-infrared light propagation in the adult human head: implications for functional near-infrared spectroscopy. *Journal of Biomedical Optics*. 10**:**6, 064032.

Hu, X.-S., Hong, K.-S., Ge, S.S. (2011). Recognition of stimulus-evoked neuronal optical response by identifying chaos levels of near-infrared spectroscopy time series. *Neuroscience Letters.* 504, 115– 120

Hugdahl K., Rund, B. R., Lund, A., Asbjørnsen, A., Egeland, J., Ersland, L., Landrø, N. I., Roness, A., Stordal, K. I., Sundet, K., and Thomsen, T. (2004). Brain activation measured with fMRI during a mental arithmetic task in schizophrenia and major depression. *Am J Psychiatry.* 161:2, 286-293.

Huppert, T.J., Hoge, R.D., Dale, A.M., Franceschini, M.A., and Boas, D.A. (2006). Quantitative spatial comparison of diffuse optical imaging with blood oxygen level-dependent and arterial spin labeling-based functional magnetic resonance imaging. *Journal of Biomedical Optics*. 11:6, D280-D298.

Huppert, T.J., Diamond, S.G., Franceschini, M.A., and Boas, D.A. (2009). HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Applied Optics*. 48:10, D280-D298.

Hutchison, R. M., Womelsdorf, T., Allen, E. A., Bandettini, P. A., Calhoun, V. D., Corbetta, M., Della Penna, S., Duyn, J. H., Glover, G. H., Gonzalez-Castillo, J., Handwerker, D. A., Keilholz, S., Kiviniemi, V., Leopold, D. A., de Pasquale, F., Sporns, O., Walter, M., and Chang, C. (2013). Dynamic functional connectivity: Promise, issues, and interpretations. *Neuroimage*. 80, 360-378. doi: 10.1016/j.neuroimage.2013.05.079.

ISS, Inc. (2008). *Functional Brain Imaging System - Imagent*. [product information]. Available from: http://www.iss.com/Products/imagent/index.html.

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., and Chance, B. (2004). Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks, *International Journal of Human-Computer Interaction*. 17:2, 211 - 231.

Jiménez, L., and Méndez, A. (2013). It is not what you expect: Dissociating conflict adaptation from expectancies in a Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, 39:1, 271-284.

Joanette, Y., Ansaldo, A. I., de Mattos Pimenta Parente, M. A., Fonseca, R. P., Kristensen C. H., and Scherer, L. C. (2008). Neuroimaging investigation of executive functions: evidence from fNIRS. *PSICO*. 39:3, 267-274.

Kalman, R. E., (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82: 35-45.

Kang, J., Wang, L., Yan, C., Wang, J., Liang, X., He, Y. (2011). Characterizing dynamic functional connectivity in the resting brain using variable parameter regression and Kalman filtering approaches. *Neuroimage*. 56:3, 1222-1234. doi: 10.1016/j.neuroimage.2011.03.033.

Kanwisher, N. (2010). Functional specificity in the human brain: A window into the functional architecture of the mind. *PNAS*. 107:25, 11163–11170.

Kawaguchi, H. and E. Okada. (2007). Evaluation of image reconstruction algorithm for near infrared topography by virtual head phantom - art. no. 662906. *Conference on Diffuse Optical Imaging of Tissue*. Munich, Germany.

Kelly, A.M.C., Uddin, L.Q., Biswal, B.B., Castellanos, F.X. and Milham, M.P. (2008). Competition between functional brain networks mediates behavioral variability. *NeuroImage*. 39, 527-537.

Khoa, T. Q. D., Yuichi, N., Masahiro, N. (2009). Recognizing brain motor imagery activities by identifying chaos properties of oxy-hemoglobin dynamics time series. *Chaos, Solitons and Fractals.* 42, 422–429.

Kineses, Z. T. Johansen-Berg H., Tomassini V., Bosnell R., Matthews P. M., Beckmann C. F. (2008). Model-free characterization of brain functional networks for motor sequence learning using fMRI. *Neuroimage*. 39:4, 1950-1958.

Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., and Eastwood, J. D. (2003). Attention, researchers! It is time to take a look at the real world. *Current Directions in Psychological Science.* 12:5, 176-180.

Kirilina, E., Jelzowb, A., Heinea, A., Niessinga,M., Wabnitzb, H., Brühlb, R., Ittermannb, B., Jacobsa, A.M., and Tachtsidis, I. (2012). The physiological origin of task-evoked systemic artefacts in functional near infrared spectroscopy. *NeuroImage*. 61, 70-81.

Kleiner M., Brainard D., Pelli D. (2007). What's new in Psychtoolbox-3? *Perception 36 ECVP*. Abstract Supplement.

Kleinschmidt, A., Obrig, H., Requardt, M., Merboldt K., Dirnagl, U., Villringer, A., and Frahm, J. (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*. 16:5, 817-826.

Kojima, T., Tsunashima, H., Shiozawa, T., Takada, H., and Sakai, T. (2005). Measurement of train driver's brain activity by functional near-infrared spectroscopy (fNIRS). *Optical and Quantum Electronics.* 37:13, 1319-1338.

Kolehmainen, V., Prince, S., Arridge, S. R., Kaipio, J. P., (2003). State-estimation approach to the nonstationary optical tomography problem. *J. Opt. Soc. Am. A.* 20:5, 876-889.

LaConte, S. M., Peltier, S. J., X. P. P. Hu, (2007). Real-time fMRI using brain-state classification. *Human Brain Mapping*. 28:10, 1033-1044.

Leff, D. R., Elwell, C. E., Orihuela-Espina, F., Atallah, L., Delpy, D. T., Darzi, A. W., and Yang G. Z. (2008). Changes in prefrontal cortical behaviour depend upon familiarity on a bimanual co-ordination task: An fNIRS study. *Neuroimage*. 39:2, 805-813.

Li, L., Du, P., Li, T., Luo, Q., and Gong, H. (2007). Design and evaluation of a simultaneous fNIRS/ERP instrument. *Conference on Optical Tomography and Spectroscopy of Tissue VII. Proc. of SPIE*. 6434, 643429, San Jose, CA.

Lieberman, H. R., Bathalon, G. P., Falco, C. M., Morgan, C. A. 3rd, Niro, P. J., and Tharion, W. J. (2005). The fog of war: Decrements in cognitive performance and mood associated with combat-like stress. *Aviation Space and Environmental Medicine.* 76:7, C7-C14.

Liu, H., Boas, D. A., Zhang, Y., Yodh, A. G., Chance, B. (1995). Determination of optical-properties and blood oxygenation in tissue using continuous NIR light. *Physics in Medicine and Biology*. 40:11, 1983-1993.

Lloyd-Fox, S., Blasi, A. and Elwell, C. E. (2010). Illuminating the developing brain: The past, present and future of functional near infrared spectroscopy. *Neuroscience and Biobehavioral Reviews*. 34:3, 269-284.

MacDonald, A.W., Cohen, J.D., Stenger, V.A., and Carter, C.S. (2000). Dissociating the Role of the Dorsolateral Prefrontal and Anterior Cingulate Cortex in Cognitive Control. *Science*. 288:5472, 1835-1838.

Mackey, J., Adamovsky, G., Floyd, B., Tin, P., and Harrivel, A. R. (2012). Improved light injection and detection methods for fNIRS headgear for use in avionics and astronautics. *2nd Biennial fNIRS Meeting,* October 26-28, 2012, London, UK. NASA LEW-19109-1.

Majeed W, Magnuson M, Hasenkamp W, Schwarb H, Schumacher EH, Barsalou L, Keilholz SD (2011): Spatiotemporal dynamics of low frequency BOLD fluctuations in rats and humans. Neuroimage. 54:1140–1150.

Mantini D., Perrucci, M. G., Del Gratta, C., Romani, G. L., and Corbetta, M. (2007). *Electrophysiological signatures of resting state networks in the human brain.* Proceedings of the National Academy of Sciences of the United States of America. 104:32, 13170-13175.

MATLAB, Statistics and Signal Processing Toolboxes, R2012b, The MathWorks, Inc., Natick, MA, United States.

Matthews, F., Pearlmutter, B. A. Ward, T. E. Soraghan, C., and Markham C. (2008). Hemodynamics for braincomputer interfaces. *IEEE Signal Processing Magazine*. 25:1, 87-94.

Mayr, U., Awh, E., and Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*. 6, 450-452.

McKiernan, K.A. Kaufman, J.N., Kucera-Thompson, J., Binder, J.R.. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. Journal of Cognitive Neuroscience, 15:3, 394-408.

Mehnert, J., et al. (2009). Resting state Networks Revealed with Whole-head Near-Infrared Spectroscopy. *15th Annual Meeting of the Organization for Human Brain Mapping*. San Francisco, CA. *NeuroImage*. 47, S163.

Mesquita, R.C., Franceschini, M.A., and Boas, D.A. (2010). Resting state functional connectivity of the whole head with near-infrared spectroscopy. *Biomedical Optics Express*, 1:1, 324-336.

Möller M, Freund M, Greiner C, Schwindt W, Gaus C, Heindel W. (2005). Real time fMRI: a tool for the routine presurgical localisation of the motor cortex. *European Radiology*. 15:2, 292–295.

Moosmann, M., Ritter, P., Krastel, I., Brink, A., Thees, S., Blankenburg, F., Taskin, B., Obrig, H., and Villringer, A. (2003). Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. *Neuroimage*. 20:1, 145-158.

Murphy K., Birn, R. M., Handwerker, D. A., Jones, T. B., and Bandettini, P. A. (2009). The impact of global signal regression on resting state correlations: Are anti-correlated networks introduced? *Neuroimage*. 44:3, 893-905.

Nebel, K., Wiese, H., Stude, P., de Greiff, A., Diener, H. C., and Keidel, M. (2005). On the neural basis of focused and divided attention. *Cognitive Brain Research*. 25:3, 760-776.

Obrig, H. and Villringer, A. (2003). Beyond the visible - Imaging the human brain with light. *Journal of Cerebral* Blood Flow and Metabolism. 23:1, 1-18.

O'Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A., Foxe, J.J., Kelly, S. P. (2009). Uncovering the Neural Signature of Lapsing Attention: Electrophysiological Signals Predict Errors up to 20 s before They Occur. *Journal of Neuroscience*. 29:26, 8604-8611.

Okada, E. and Delpy, D. T. (2003). Near-infrared light propagation in an adult head model. II. Effect of superficial tissue thickness on the sensitivity of the near-infrared spectroscopy signal. *Appl Opt.* 42:16, 2906-2914.

Oken, B. S., Salinsky, M. C., and Elsas, S. M. (2006). Vigilance, alertness, or sustained attention: physiological basis and measurement. *Clinical Neurophysiology.* 117:9, 1885-1901.

Orihuela-Espina F., Leff, D. R., James, D. R., Darzi, A. W., and Yang, G. Z. (2010). Quality control and assurance in functional near infrared spectroscopy (fNIRS) experimentation. *Physics in Medicine and Biology*. 55:13, 3701-3724.

Pope, A.T., Bogart, E.H., Bartolome, E.S. (1995) Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*. 40, 187-195.

Posner, M.E., (2004). Cognitive Neuroscience of Attention. New York, NY: Guilford Press.

Prado, J., Weissman, D.H. (2011). Heightened interactions between a key default-mode region and a key task-positive region are linked to suboptimal current performance but to enhanced future performance. *NeuroImage*. 56:4, 2276-82.

Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., and Shulman, G.L. (2001). A default mode of brain function. *PNAS*. 98:2, 676-682.

Raichle, M.E. and A.Z. Snyder. (2007). *A default mode of brain function: a brief history of an evolving idea.* Neuroimage. 37:4, 1083-90; discussion 1097-9.

Raley, C., Stripling, R., Kruse, A., Schmorrow, D., Patrey, J. (2004). Augmented Cognition Overview: Improving Information Intake under Stress. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 48:10, 1150-1154.

Sarter, M., Gehring, W.J. and Kozak R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews*. 51:2, 145-160.

Sasai, S., Homae, F., Watanabe, H., Sasaki, A. T., Tanabe, H. C., Sadato, N., Taga, G. (2012). A NIRS–fMRI study of resting state network. NeuroImage. 63:1, 179–193.

Sassaroli, A., Pierro, M., Bergethon, P. R., Fantini, S. (2012). Low-Frequency Spontaneous Oscillations of Cerebral Hemodynamics Investigated With Near-Infrared Spectroscopy: A Review. *IEEE Journal of Selected Topics in Quantum Electronics*, 18:4, 1478 - 1492. doi: 10.1109/JSTQE.2012.2183581.

Sato H., Kiguchi, M., Maki, A., Fuchino, Y., Obata, A., Yoro, T., Koizumi, H. (2006). Within-participant reproducibility of near-infrared spectroscopy signals in sensorimotor activation after 6 months. *Journal of Biomedical Optics*. 11:1, 014021.

Schnell, T., Kwon, Y., Merchant, S, and Etherington, T. (2004). Improved Flight Technical Performance in Flight Decks Equipped with Synthetic Vision Information System Displays. *International Journal of Aviation Psychology*. 14:1, 79-102.

Schroeter, M.L., Kupka, T., Mildner, T., Uludağ, K., Yves von Cramon, D. (2006). Investigating the post-stimulus undershoot of the BOLD signal - A simultaneous fMRI and fNIRS study. *NeuroImage*. 30:2, 349-358.

Schweiger, M., Nissilä, I., Boas, D. A., and Arridge, S. R. (2007). Image reconstruction in optical tomography in the presence of coupling errors. *Applied Optics*. 46:14, 2743-2756.

Sharp D. J., Scott, S. K., Mehta, M. A., and Wise, R. J. (2006). The neural correlates of declining performance with age: Evidence for age-related changes in cognitive control. *Cerebral Cortex*. 16:12, 1739-1749.

Smallwood, J., Brown, K., Baird, B., Schooler, J.W. (2012). Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought. *Brain Research.* 1428, 60-70.

Smilek, D., Eastwood J. D., Reynolds, M. G., Kingstone, A. (2007). Metacognitive errors in change detection: Missing the gap between lab and life. *Consciousness and Cognition.* 16:1, 52-57.

Snow, M., et al., (2006). Service Test and Evaluation Report, Phase IV. *Improving Warfighter Information Intake Under Stress*. *DARPA Augmented Cognition,* contract: NBCHC030031.

Solovey, E., Lalooses, F, Chauncey, K, Weaver, D, Parasi, M, Matthias Scheutz, M, Sassaroli, A, Fantini, S, Schermerhorn, Girouard, PA, Jacob, RJK. (2011). Sensing Cognitive Multitasking for a Brain-Based Adaptive User Interface, *CHI 2011, Session: Brain & Bio-sensor Interactions,* 383-392. Vancouver, BC, Canada.

Soe, N. N., Nakagawa, M. (2008). Chaotic Properties of Hemodynamic Response in Functional Near Infrared Spectroscopic Measurement of Brain Activity. *World Academy of Science, Engineering and Technology*. 2, 2008-03-25

Steinbrink, J., Villringer, A., Kempf, F., Haux, D., Boden, S., and Obrig, H. (2005). Illuminating the BOLD signal: combined fMRI-fNIRS studies. *International School on Magnetic Resonance and Brain Function*. Erice, Italy.

Steinstrater O., S.J., Frank A., Jansen A., and Knecht S., (2011). Transcranial magnetic stimulation and language: how variable are our stimulation sites? *Neuroimage*. 13, S257.

Stins, J.F., van Leeuwen W.M.A., and de Geus, E.J.C. (2005). The Multi-Source Interference Task: The Effect of Randomization. *Journal of Clinical and Experimental Neuropsychology*. 27:6, 711-717.

Strangman, G., Culver, J.P., Thompson, J.H., and Boas, D.A. (2002a). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. *NeuroImage*. 17:2, 719-731.

Strangman, G., Boas D.A., and Sutton, J.P. (2002b). Non-invasive neuroimaging using near-infrared light. *Biological Psychiatry*. 52:7, 679-693.

Strangman, G., Franceschini, M.A., and Boas, D.A. (2003). Factors affecting the accuracy of near-infrared spectroscopy concentration calculations for focal changes in oxygenation parameters. *NeuroImage*. 18:4, 865-879.

Strayer D.L., Watson, J.M., and Drews, F.A. (2011). *Cognitive Distraction While Multitasking in the Automobile. In Brian Ross, editor: The Psychology of Learning and Motivation*, Burlington: Academic Press.

Tachtsidis, I., Leung, T. S., Chopra, A., Koh, P. H., Reid, C. B., and Elwell, C. E. (2009). False positives in functional nearinfrared topography. In: Liss, P., Hansell, P., Bruley, D.F., Harrison, D.K. (Eds.), *Oxygen Transport to Tissue XXX, Advances in Experimental Medicine and Biology*. (pp. 307–314). New York, NY: Springer.

Thompson, G. J., Magnuson, M. E., Merritt, M. D., Schwarb, H., Pan, W. J., McKinley, A., Tripp, L. D., Schumacher, E. H., Keilholz, S. D. (2013). Short-Time Windows of Correlation Between Large-Scale Functional Brain Networks Predict Vigilance Intraindividually and Interindividually. *Human Brain Mapping.* 34:3280–3298.

Tuchin, V. (2007). *Tissue Optics* (2nd ed.). Bellingham, WA: SPIE Press.

Uddin L. Q., Kelly, A. M., Biswal, B. B., Castellanos, F. X., and Milham, M. P. (2009). Functional Connectivity of Default Mode Network Components: Correlation, Anticorrelation, and Causality. *Human Brain Mapping.* 30:2, 625-637.

Vander, et al. (2000). *Human Physiology: The Mechanism of Body Function* (8th ed.). New York City, NY: The McGraw-Hill Companies.

Vohn R., Fimm, B., Weber, J., Schnitker, R., Thron, A., Spijkers, W., Willmes, K., and Sturm, W. (2007). *Management of attentional resources in within-modal and cross-modal divided attention tasks: An fMRI study*. Human Brain Mapping. 28**(**12): 1267-1275.

Weissman, D.H., Roberts, K.C., Visscher, K.M. and Woldorff, M.G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience*. 9:7, 971-978.

Welch, G., Bishop, G. (2006). An Introduction to the Kalman Filter. available at: http://www.cs.unc.edu/~welch/kalman/kalmanIntro.html

Weng S. J., Wiggins J. L, Peltier S. J, Carrasco M., Risi S., Lord C., Monk C. S. (2010). Alterations of resting state functional connectivity in the default network in adolescents with autism spectrum disorders. *Brain Research*. 1313**,** 202-214.

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(1): 148–156. doi:10.1016/j.neuroimage.2009.03.058.

Wojciulik, E. & Kanwisher, N. (1999). The Generality of Parietal Involvement in Visual Attention. *Neuron.* 23, 747–764.

Yerkes, R. M., and Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. Journal of Comparative Neurology and Psychology. 18: 459–482. doi:10.1002/cne.920180503

Zhang, Q., Brow, E., Strangman, G. (2007a). Adaptive Filtering for global interference cancellation and real-time recovery of evoked brain activity: a Monte Carlo simulation study. *Journal of Biomedical Optics*. 12:4, 044014-1 - 044014-12.

Zhang, Q., Brown, E., Strangman, G. (2007b). Adaptive filtering to reduce global interference in evoked brain activity detection: a human subject case study. *Journal of Biomedical Optics*. 12:6, 064009 -1 - 064009-12.

Zhang, Q., Strangman, G. E., Ganis, G. (2009). Adaptive filtering to reduce global interference in non-invasive NIRS measures of brain activation: How well and when does it work? *Neuroimage*. 45:,788-794.