Chapter 6
Using fNIRS to Measure Mental Workload in the Real World

Evan M. Peck, Daniel Afergan, Beste F. Yuksel, Francine Lalooses and Robert J. K. Jacob

Abstract In the past decade, functional near-infrared spectroscopy (fNIRS) has seen increasing use as a non-invasive brain sensing technology. Using optical signals to approximate blood-oxygenation levels in localized regions of the brain, the appeal of the fNIRS signal is that it is relatively robust to movement artifacts and comparable to fMRI measures. We provide an overview of research that builds towards the use of fNIRS to monitor user workload in real world environments, and eventually to act as input to biocybernetic systems. While there are still challenges for the use of fNIRS in real world environments, its unique characteristics make it an appealing alternative for monitoring the cognitive processes of a user.

Introduction

As brain sensing technology has become more unobtrusive, portable, and inexpensive, it has become a more viable technology for evaluating everyday work environments. The ability to access physiological parameters that correlate with

E. M. Peck (✉) · D. Afergan · B. F. Yuksel · F. Lalooses · R. J. K. Jacob
Tufts University, Medford, USA
e-mail: evan.peck@tufts.edu

D. Afergan
e-mail: afergan@cs.tufts.edu

B. F. Yuksel
e-mail: beste.yuksel@tufts.edu

F. Lalooses
e-mail: francine.lalooses@tufts.edu

R. J. K. Jacob
e-mail: jacob@cs.tufts.edu

S. H. Fairclough and K. Gilleade (eds.), *Advances in Physiological Computing*, Human–Computer Interaction Series, DOI: 10.1007/978-1-4471-6392-3_6,

brain activity is an appealing option for researchers who are trying to capture cognitive state at its primary source. Monitoring cognitive states such as workload, attention, or emotion provide valuable information that aids in the evaluation of interfaces design, task design, or individual cognitive abilities.

In particular, researchers have sought to capture working memory because it has been labelled as the information bottleneck in the brain, and is necessary for reasoning, anticipation, and planning (Baddeley 1992). However, working memory is also limited by both duration and capacity (Miller 1956). When multiple elements compete for space, there is a loss of information and often a decrease in performance. Thus, preventing users from overloading their working memory, by understanding when and how it is overloaded, is a critical component of supporting analytical thought.

We pursue the use of brain sensing to capture these processes not only because it may lend insight into periods of overload, but because overload can be difficult to observe through other means—increasing load on working memory does not always result in a decrease in performance (Hockey 1997). Performance may be maintained with an increased reliance on extra resources and with extra subjective effort, behavioral, and physiological costs.

In addition, observing load on working memory often hints at the overall mental workload of the user. Wickens (2002) demonstrates how closely the two are connected and how the distinctions are not always so clear with his four-dimensional multiple resources model. Wickens' theory dictates that two tasks that both need the same level of a dimension will yield more conflict than two non-overlapping tasks. Thus, multiple working memory tasks that require the same modalities are likely to yield an overload on that resource. If we can observe working memory (or mental workload) in real-time, then we may be able to improve the user's environment to better support these processes, improving interaction and increasing performance.

To capture measures of mental workload in the brain, most research has focused on the use of electroencephalography (EEG) to monitor the electrical activity of the brain. However, there has also been increasing interest in functional near-infrared spectroscopy (fNIRS) as an alternative brain sensing technique. This is largely due to the fact that fNIRS measures a unique set of physiological parameters (oxygenated and deoxygenated hemoglobin) in a relatively unobtrusive manner. Thus, fNIRS has seen increased use for scenarios where it is necessary to capture cognitive state without applying too many physical restrictions on the user.

In this chapter, we give an overview of why fNIRS is well-suited to observe a user's working memory, specifically, the central executive (Repovš and Baddeley 2006). In particular, we focus on fNIRS work that is geared towards the field of human-computer interaction and ecologically valid evaluations.

To accomplish this goal, we build from a technical description of fNIRS to the use of fNIRS in real world scenarios. First, we describe fNIRS as a brain sensing technology, highlighting its advantages and disadvantages in comparison to other brain sensing techniques. Next, we describe Repovš and Baddeley's (2006) multi-component model of working memory and then dive into the existing neuroscience literature, exploring why the prefrontal cortex is an ideal place to observe increases in working memory. Then, beginning with highly controlled psychology tasks, we highlight experiments that use fNIRS to investigate working memory load in the prefrontal cortex. Finally, we look at the potential of using fNIRS as real-time input to adaptive brain-computer interfaces. We find this last section to be most exciting, as using the brain as input to a computer expands the interaction bandwidth between a user and the computer in unique and powerful ways.

There are still serious challenges for the use of fNIRS, which we will also discuss in this chapter, but it also provides researchers with a unique set of properties that can compliment other physiological signals. We hope to highlight the potential of fNIRS in this chapter, as well as outline its future development.

Introduction to fNIRS

How Does it Work, What Does it Measure?

Functional near infrared spectroscopy (fNIRS) is an optical brain sensing technique developed in the 1990 s that is portable, resistant to movement artifacts (Solovey et al. 2009), and observes similar physiological parameters to functional magnetic resonance imaging (fMRI) (Chance et. al 1998). These characteristics have made it an attractive alternative for researchers seeking to observe the brain in natural working environments.

fNIRS uses near-infrared light to measure concentration and oxygenation of the blood in the tissue at depths of 1–3 cm (Villringer and Chance 1997). Light is sent into the forehead in the near infrared range (650–900 nm), where it is diffusely reflected by the scalp, skull, and brain cortex. At this wavelength, oxygenated and deoxygenated hemoglobin are the primary absorbers of light. A very small percentage of the light sent into the head returns from the cortex to the detector on the fNIRS probe. By measuring the light returned to the detector, researchers are able to calculate the amount of oxygen in the blood, as well as the amount of blood in the tissue.

Biologically, when a region of the brain is active, there is an increase of blood flow to that region (D'Esposito et al. 1999). This increase of blood flow is typically coupled with decreased levels of deoxygenated hemoglobin and increased levels of oxygenated hemoglobin. Thus, fNIRS can be used to measure activity in localized areas of the brain.

To make this calculation, raw data can be transformed into deoxygenated hemoglobin concentrations using the modified Beer-Lambert Law [5]:

$$\Delta A = \varepsilon \times \Delta c \times d \times B \tag{6.1}$$

where ΔA is the change in attenuation of light, ε is the molar absorption coefficient of the absorbing molecules, Δc is the change in the concentration of the absorbing molecules, d is the optical pathlength (i.e., the distance the light travels), and B is the differential pathlength factor. The attenuation of light is measured by how much light is absorbed by oxygenated and deoxygenated hemoglobin (which are the main absorbers of near infrared light at these wavelengths). As the attenuation of light is related to the levels of hemoglobin, given ΔA, we can derive the changes in the levels of oxygenated and deoxygenated hemoglobin (Chance et al. 1998).

For the sake of focus, we will primarily discuss fNIRS probes that are placed on the forehead, measuring brain activity in the anterior prefrontal cortex. This placement offers a significant advantage for researchers: fNIRS measurements can be made without the interference of hair follicles, which can absorb light and disrupt the signal. As a result, probes that monitor activity in the prefrontal cortex are often more unobtrusive, and therefore, more interesting for researchers searching for ecologically sound measurements.

Other Techniques: EEG and fMRI

In the past, various brain sensing technologies have been proposed to observe a user's response to activities in a lab setting. Electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) are two of the most prevalent and both have been successful at measurement and classification of brain activities. fMRI requires a person to lie motionless inside a large, loud chamber in which small movements (those larger than 3 mm) often result in discarded data (i.e. Cahill et al. 2004; Davis et al. 2004). fMRI scanners are also expensive to purchase and maintain, requiring technical staff and purpose-built rooms or buildings.

EEG has recently seen commercial success because it is portable, less invasive, and relatively inexpensive. While EEG has a high temporal resolution, it also has a low spatial resolution, which makes it difficult to pinpoint the origin of neural activity. Although EEG is easier to set up and use than fMRI, many configurations require applying gel into a person's hair to create a conductive contact with the skin. Finally, movement artifacts can be problematic with the use of EEG. Without proper filtering methods, minor movements, such as facial muscles can disrupt incoming data. Despite these limitations, EEG has gained popularity because of its quick temporal response (1 ms), the strong existing body of EEG research, and the availability of well-supported commercial setups.

fNIRS Advantages

While fNIRS preserves some of the core features that make EEG a popular brain sensing technology, most notably its ease of use and portability, fNIRS also has a few unique properties that are worth considering. For example, fNIRS has a short setup time and is generally resistant to movement artifacts. Mouse-clicking, typing, eye movement, and blinking in normal computing environments are acceptable during the use of fNIRS (Solovey et al. 2009). Minor head movement, respiration, and heartbeats can be filtered using known signal processing techniques. Only major head and forehead movement (which could be induced by frowning) are disruptive to the signal.

fNIRS also has a spatial resolution on the order of 0.5–1 cm, and readings have been validated against fMRI (Strangman et al. 2002). Furthermore, fNIRS provides access to hemodynamic and metabolic parameters that are not accessible with EEG (which is sensitive to electrical signals and not to blood flow or tissue oxygenation) and fMRI (which is only sensitive to deoxygenated hemoglobin and not oxygenated hemoglobin).

fNIRS Considerations

It's important to note that there are both caveats and considerations to the use of fNIRS as well. Once a region of the brain becomes active, the biological response to support this increase in activity takes several seconds to reach the cortex (Gore 2003). As a result, changes in blood oxygenation that reflect user state cannot be detected immediately using fNIRS.

The impact of this limitation is two-fold. First, we are more likely to measure signal differences in short-term and long-term cognitive states rather than instantaneous one-time events. Second, because the slow biological response impacts how quickly we can identify or classify user state, it also impacts the design of adaptation mechanisms in biocybernetic systems that respond to fNIRS user state. In a later section, we will discuss some design considerations that can help circumvent this issue.

Because light from fNIRS reaches depths of 1–3 cm, activity in deeper areas of the brain is not directly accessible. Additionally, hair can obstruct light, so sources and detectors must be maneuvered to maintain contact with the skin. Although there are several variants of full-head fNIRS devices (Franceschini et al. 2006), most are noticeably less comfortable than probes designed exclusively for the forehead. As a result, many fNIRS researchers currently investigate activity in the prefrontal cortex (PFC).

Working Memory and the Prefrontal Cortex

Working memory has been defined as the "temporary storage and manipulation of the information necessary for such complex tasks such as language comprehension, learning and reasoning" (Baddeley 1992). It is a complex topic that can be explored at different levels (Repovš and Bresjanac 2006), such as neuroanatomical structure, computational cognitive processes, or the more abstract level of modelling its functional capacities. For the purposes of presenting working memory in the context of using fNIRS, we explain working memory in terms of Repovš and Baddeley's (2006) multi-component model. We then motivate the placement of fNIRS on the forehead by reviewing a sample of the existing literature that has linked the PFC to working memory. In this respect, fMRI studies are particularly illuminating, as they access similar physiological parameters as fNIRS and their results have the potential of being directly replicated.

Working Memory

The model proposed by Repovš and Baddeley (2006) for working memory is made up of multiple components: a central executive, a visuospatial sketchpad, a phonological loop, and an episodic buffer. The phonological loop contains verbal information in the form of a short-term acoustic store and an articulatory rehearsal process. The visuospatial sketchpad is comprised of visual and spatial subsystems with independent storage, maintenance and manipulation processes. The central executive seems to have several different functions which are involved in the manipulation of information within the memory stores. The latest addition to the model, the episodic buffer, holds and integrates information from other working memory components and from long-term memory into scenes or episodes. For an in-depth description and discussion of the multi-component model of working memory please see Repovš and Baddeley (2006).

We propose that the increase in mental workload that the fNIRS measures in the PFC is actually related to an increase in working memory, and more specifically, in the central executive. We present evidence of this by discussing studies linking working memory to the prefrontal cortex.

Prefrontal Cortex

It is largely undisputed that the prefrontal cortex is important for higher-order cognitive functions including working memory (Ramnani and Owen 2004). The dorsolateral prefrontal cortex has been associated with working memory and with high-level organization of the contents of working memory. Bor et al. (2003) presented subjects with a spatial working memory task with structured and unstructured sequences. The structured sequences were easier to remember, however, functional magnetic resonance imaging (fMRI) results showed increased activation in the lateral frontal cortex with structured sequences. The results show that even when working memory demand decreases, the frontal cortex is still active in the organization of the contents of working memory (Bor et al. 2003). Similarly, Bor et al. (2004) showed similar results with verbal working memory using structured and unstructured digit sequences. Both of these studies suggest that the lateral frontal cortex structures high-level information into organized groups which can decrease the load on working memory.

The ventrolateral prefrontal cortex is also associated with working memory and is thought to be linked to the retrieval of a few pieces of information. For example, Jonides et al. (1993) gave participants a memory task by presenting three stimuli concurrently for 200 ms. They then asked participants whether a probe circle displayed 3 s later occupied one of the same locations. Using positron emission tomography (PET), Jonides et al. (1993) found increased activation in the mid-ventrolateral prefrontal cortex in contrast to a simple perception condition. Dove et al. (2001) tested the effects of when participants explicitly intended to remember or retrieve information. Participants looking at pictures of abstract art were instructed either to just look or to explicitly remember similar stimuli for later. The latter condition created increased activation in the mid-ventrolateral cortex (measured with fMRI) (Dove et al. 2001, see also Owen et al. 2005). These findings suggest that the intention to remember may be the cause of the activation in the ventrolateral cortex.

One area of the prefrontal cortex that is not so widely understood but is also associated with working memory is the anterior prefrontal cortex (see Ramnani and Owen 2004 for a review). This constitutes Brodmann area (BA) 10 and is very well developed in humans in comparison to other primates. Some of the neuro-imaging evidence points towards functionalities such as processing of internal states and evaluating ones own thoughts and feelings (Christoff and Gabrieli 2000); memory retrieval, retrieval verification and source memory (Tulving 1983; Rugg 1998); prospective memory such as “at time x, do y” (Burgess et al. 2001); cognitive branching or the ability to hold goals while carrying out secondary goals (Koechlin et al. 2000); and relational knowledge (Kroger et al. 2002; Christoff et al. 2001).

A theory put forward by Ramnani and Owen (2004) is that the anterior pre-frontal cortex is activated when a problem requires more than one cognitive process. Hence, the cognitive operations need to be coordinated in the pursuit of a more general goal. Their hypothesis is broadly consistent with much of the previous functional neuroimaging literature described above, and with the theory of the central executive of Repovš and Baddeley’s (2006) model.

Detecting Working Memory with fNIRS

To build towards measuring workload in real world scenarios, we start by using fNIRS to identify it in heavily controlled environments. In particular, we begin by describing the n-back task, a classical workload task in the psychology literature, and review neuroscience work that suggests that fNIRS is capable of detecting activation in the prefrontal cortex.

The N-Back Task

Gevins and Cutillo (1993) created the "n-back" task, a working memory task that has been widely used and quoted in the literature to investigate cognitive working memory processes. The n-back consists of a user being presented with sequentially presented stimuli and answering positively whenever the current stimulus matches the stimulus occurring n positions back in the sequence. The integer n is given to the user at the start and is usually 1, 2 or 3.

The n-back task is demanding and requires on-line storage, monitoring and manipulation of information. The participant must hold information in their working memory whilst incorporating new information and comparing it with the older stimuli. There are different variations of the n-back stimuli: these can be presented verbally (letters and words) or non-verbally (images such as shapes or faces. Another variation is whether the identity of the stimulus has to be remembered or the location of the stimulus (Owen et al. 2005). Below, we show an image of a visuospatial n-back task, labeling the 'yes' responses for both a 1-back and 3-back condition.

The n-back has been shown to produce activation in working memory related cortical regions (Braver et al. 1997; Cohen et al. 1997). Owen et al. (2005) carried out a meta-analysis on 24 functional neuroimaging studies carried out in Talairach space that used the n-back paradigm on healthy subjects. They found that six cortical regions were consistently activated across all studies including the bilateral rostral prefrontal cortex (BA10), the bilateral dorsolateral prefrontal cortex (BA9,46) and the bilateral mid-ventrolateral prefrontal cortex (BA45,47). There were some differences depending on the variation of the n-back task. N-backs that were verbal identity tasks created more activation in the left ventrolateral prefrontal cortex than the non-verbal identity tasks (Owen et al. 2005). Non-verbal location tasks created more activation in the right dorsolateral prefrontal cortex than non-verbal identity tasks. These results suggest that there is functional specialization of working memory systems and that activation of the prefrontal cortex can be identified with n-back tasks to quite a specific degree of accuracy.

N-Back and fNIRS

To validate the use of fNIRS to detect workload in a visual interface HCI task, Peck et al. (2013b) recorded signals in the prefrontal cortex of 16 participants as they interacted with a visuospatial n-back task. In this version of the n-back task, participants were shown a series of slides that have distinct visual patterns and asked whether the pattern in the current slide matched the pattern viewed n slides ago. By increasing the value of n, participants were forced to hold n patterns in their visuospatial short-term memory. Thus, as n increased, the load on their visual short-term memory increased.

Fig. 6.1 *Left* We display one example of an fNIRS probe with four light sources and one detector. *Right* A researcher uses a headband to secutre two probes to a user's forehead

Fig. 6.2 A participant interacts with a computer while wearing fNIRS probes

Participants were given eight trials of 1-back and 3-back conditions (representing low and high workload), with each condition consisting of 20 slides. Trials lasted for 40.7 s and were separated by 12-s rest periods. fNIRS sensors, as shown in Figs. 6.1 and 6.2 were applied to the forehead (prefrontal cortex) and changes in deoxy-Hb were compared in the 1-back and 3-back conditions (Figs. 6.3, 6.4).

In agreement with fMRI studies, Peck found more significant decreases in deoxy-Hb during 3-back trials than with 1-back trials. Figure 6.4 shows the mean change in deoxygenated hemoglobin across all trials and all participants of the 1-back and 3-back task. Additionally, participants reported the 3-back to be more mentally demanding than the 1-back and performance degraded as participants interacted with the 3-back in comparison to the 1-back. These results agree with other behavioral studies of the n-back task and validate the use of fNIRS to record changes in visuospatial working-memory in the prefrontal cortex.

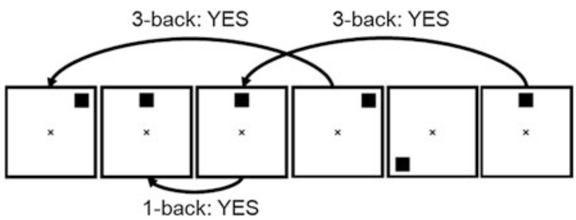

Fig. 6.3 In the visuospatial n-back task, participants view a series of slides and respond whether the current pattern matches the pattern from n slides ago. We show positive answers for both the 1-back and 3-back conditions

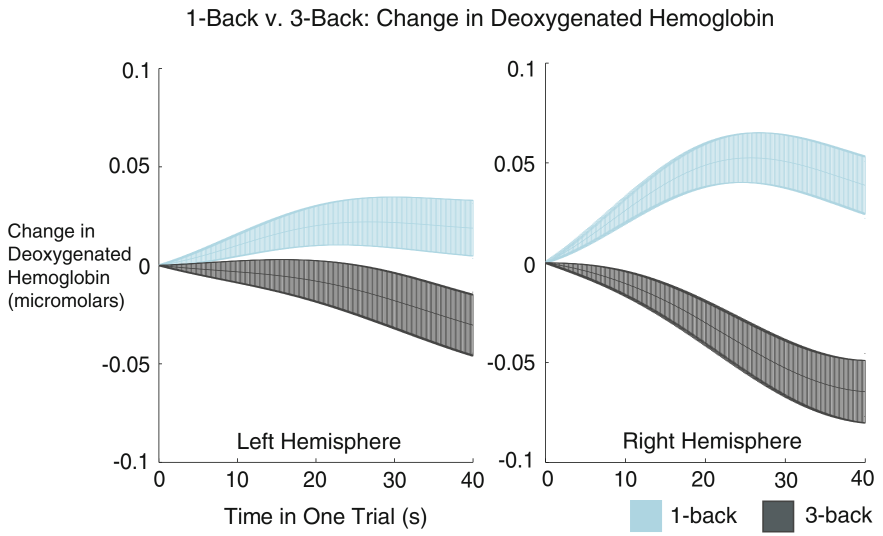

Fig. 6.4 The mean change in fNIRS signal (from a rest state) across all 16 participants in the baseline task since the beginning of the task. The *center line* is the mean for each condition, while the width indicates the standard error. We see a clear separation between the 1-back and 3-back conditions for participants. The more demanding 3-back condition mirrors signals from the graph design that participants believed was more mentally demanding

Detecting Workload in Real World Tasks

So far, we have discussed the potential of fNIRS to detect workload in heavily controlled environments. But can these psychology experiments be generalized to real world environments? As we discussed earlier, fNIRS has the unique advantage of being relatively robust to movement artifacts in a comparison to other brain sensing devices. In this section, we explore techniques that researchers have used to investigate workload in scenarios that are increasingly closer to everyday tasks.

As we review their work, we gradually move from offline statistical analysis of physiological signals, to real-time automated classification of user state.

Analyzing Changes in Oxy-Hb and Deoxy-Hb

One method for using fNIRS to detect changes in workload is statistical analysis of changes in oxy-Hb and deoxy-Hb. Recall that as brain activity increases, we generally observe increases in oxy-Hb and decreases in deoxy-Hb. By analyzing the changes in these parameters during a user's interaction with a complex task, we can hypothesize the level of activity (and the workload) in the user's prefrontal cortex. Here, we share three examples of studies that have performed offline analysis of changes in oxy-Hb or deoxy-Hb to investigate workload levels.

Ayaz et al. (2012) used this approach to detect the level of workload for participants piloting unmanned air vehicles (UAVs). In this task, participants were asked to sit at workstations and direct simulated air traffic, trying to prevent accidents. The number of UAVs was varied (6, 12, 18) between trials and the mean change in oxy-Hb was calculated over the course of each trial.

As users were forced to keep track of more UAVs, fNIRS detected increased levels of oxy-Hb in the PFC. Ayaz found these changes to be comparable to those observed during interacton with the n-back task—a well-characterized psychological task for increasing working (or short-term) memory load. Increased levels of oxy-Hb also correlated with self-reported NASA-TLX workload measures, further validating the detection of signals that point to workload.

In another translation of fNIRS measures to real world environments, activity was recorded as participants were engaged as part of a human-robot team (Solovey et al. 2011). In this task, participants engaged in a multi-tasking assignment that could not be accomplished by the human nor the robot alone. The study investigated three classifications of multi-tasking—delay, dual-task, and branching. While *branching* required participants to maintain the context of a primary task while exploring a secondary task, users did not have to maintain this context in the *dual-task* condition, and they completely ignored the secondary task in the *delay* condition. Analyzing the mean combined hemoglobin (deoxy-Hb + oxy-Hb) for each participant, the branching condition was found to have higher levels of combined hemoglobin than either the dual-task or the delay conditions. In a second experiment that compared changed in deoxy-Hb in random interruptions with interruptions that could be predicted by the user, Solovey found that random interruptions provoked sharper decreases in deoxy-Hb.

Finally, Peck et al. (2013b) used fNIRS to detect working memory differences in a task that derives exclusively from the visual design of information. To explore the potential of fNIRS in evaluating information visualization, a complex visual task was constructed by mapping a well-studied visual judgment task (Cleveland and McGill 1984) with the n-back task. In Cleveland's original study of visual variables, participants made percentage estimates of elements within either bar

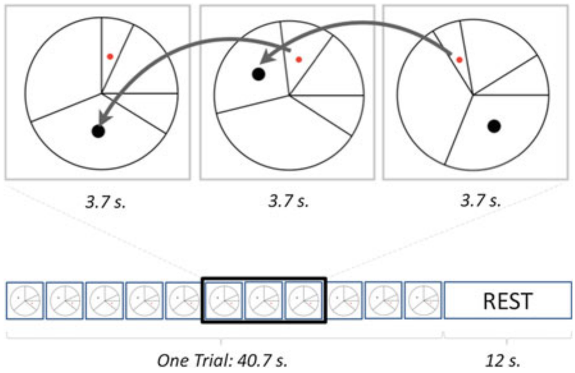

Fig. 6.5 An example of the memory-intensive graphical comparison task used by Peck et al. (2013b). Instead of making percentage comparisons of graph elements on the same chart, participants compared an element in the current chart with an element from the previously seen chart

graphs or pie charts. Their results demonstrated that people can more accurately compare information in bar graphs than in pie charts.

In order to modify this simple visual comparison task to mimic more complex, memory-intensive tasks, participants in Peck's experiment compared a graph element in the current graph (bar graphs or pie charts) with a graph element from the *previous* graph (Fig. 6.5). This manipulation required participants to maintain graph elements in their short-term memory, and fNIRS measurements were intended to capture how each graph design (bar graphs or pie charts) supported this mental process. Rather than test on easy and difficult conditions for each graph type, participants engaged with a single 1-back task for each graph to see whether the visual encoding of information would result in low or high workload (Fig. 6.5).

In addition to recording brain activity with fNIRS, participants completed the NASA Task Load Index (NASA-TLX) after interaction. Results showed that levels of deoxy-Hb differed during interaction with bar graphs and pie charts. However, these differences were not categorical. Instead, they correlated with the visualization technique that participants *believed* was more mentally demanding. Similar to the observations made in other workload-intensive tasks, higher mental demand correlated with larger decreases in deoxy-Hb.

These experiments demonstrate that fNIRS can detect workload in a wide variety of tasks that is not limited to a specific type of stimuli or category of working memory.

Automatic Detection of Workload

If fNIRS is to become a viable tool for analyzing mental state during interaction with an interface, it would be ideal for the analysis of fNIRS signals to move from a manual to an automated process. In this section, we discuss work that has employed the use of predictive models to objectively (and automatically) classify user state.

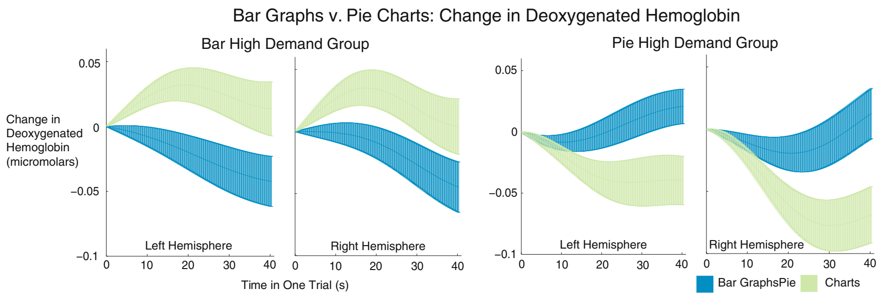

Fig. 6.6 In a study comparing bar graphs (*blue*) and pie charts (*green*), Peck found that a group of participants that subjectively rated bar graphs as more mentally demanding than pie charts (*left*) exhibited reversed fNIRS signals from those who rated pie charts as more mentally demanding than bar graphs (*right*). The plots represent the mean change in deoxygenated hemoglobin (from a rest state) across all trials of each condition. The width of the line represents the standard error at each time point

Classifying Known User State

When the user's intended state is known, some researchers have used predictive models as a statistical method to show that multiple user states is separable. However, the true potential of these models is in the automation of classifying fNIRS signals. This allows evaluators to avoid non-automated analysis—a potentially time-consuming task if fNIRS is to be used as a tool in real user studies. In these circumstances, researchers check the accuracy of their model by using cross-validation techniques.

For example, Luu and Chau (2008) used predictive models to distinguish between fNIRS signals during periods of low and high preference. In their experiment, participants viewed pictures of soft-drinks that they either highly-preferred or did not like at all. After identifying sensors that highly correlated with user preference, they were able to predict the preferred drink with over 80 % accuracy.

Similarly, Moghimi et al. (2012) used fNIRS to measure participants' emotional responses to music, attempting to capture both valence (positive or negative feelings) and arousal. They used linear discriminant analysis (LDA) to build a classifier and found that they could distinguish positive and negative valence with an average accuracy of 71.94 %. They also found they could distinguish between high arousal music and brown noise (low arousal) with an average accuracy of 71.93 %.

Finally, Girouard et al. (2009) measured users as they played a game of Pac-man, interacting with game modes that were both very easy and very difficult. They used a k-nearest neighbor (kNN) algorithm to classify game difficulty levels. While Girouard found that they could distinguish between periods of play and non-play with accuracy levels above 90 %, distinguishing between easy and hard difficulty levels yielded classifications just over 60 %.

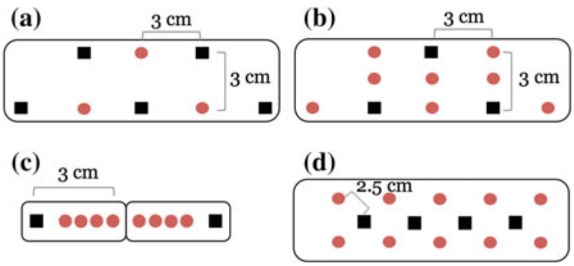

Fig. 6.7 Probes with different numbers/locations of sources and detectors may impact classification accuracy of user state. We show four probe configurations used in studies described in section "Detecting Workload in Real World Tasks". *Red circles* are light sources and *black squares* are detectors. In each case, probes would be centered on the participant's forehead **a** Moghimi et al. (2012), **b** Luu and Chau (2008), **c** Girouard et al. (2009), Hirshfield et al. (2011), Solovey et al. (2011), and Solovey et al. (2012), **d** Ayaz et al. (2012)

As Fig. 6.7 suggests, the discrepancies in classification accuracy in each study may partially stem from the various configurations of fNIRS probes; the first two studies used configurations with significantly more source-detector pairs. This provides two distinct advantages. First, increasing the number of information channels decreases the potential for one noisy channel to adversely impact a model. Second, these configurations provide better coverage of the prefrontal cortex. For example, Moghimi et al. (2012) showed that using information from a source-detector pair on the anatomical midline of the prefrontal cortex yielded the best overall accuracy in their model for capturing emotion. This is information that a smaller probe with a more linear configuration (such as the one used by Girouard) may have difficulty accessing unless it is placed in precisely the right location. However, accuracy is also heavily dependent on analysis methods, and thus far, we have skipped exactly *how* fNIRS is used as input to these models.

Selecting fNIRS Features

When predictive models are created, we need to determine which features of the signal are fed into the models. Choosing too many features with too few training examples may result in the "curse of dimensionality" and low classification rates. Choosing too few features, or incorrect features of the signal, may lead to a set of features that is not truly descriptive of the signal, also resulting in low classification accuracy. Currently, there is no standardized approach to feature extraction. We give four examples from current fNIRS literature:

- Solovey et al. (2012) used the signal value from each time point and each channel over the entire trial as individual features to a support vector machine (SVM).
- Luu et al. (2008) used the average signal value, estimated from a specific channel over a specific time interval within a trial (for example, 15–45 s).

- Hirshfield et al. (2011) extracted the max signal value, min signal value, mean signal value, slope, time to peak, and full width at half maximum.
- Moghimi et al. (2012) used the mean and slope of the signal during each trial. They also used the coefficient of variation, mean difference between signal and noise, and a handful of laterality features.

As these examples suggest, there is yet to be a prevailing consensus about which features of the fNIRS signal potentially result in the highest levels of accuracy. However, there are at least two dominant approaches to feature selection used in current fNIRS literature.

The first is to manually select a set of features based on an expert's knowledge or personal experience. For example, based on the changes of deoxy-Hb in response to load on working memory that we observed in Fig. 6.4, the mean change in deoxy-Hb would appear to be a good indicator of low/high load on working memory. In a finger-tapping task, Cui et al. (2010) show that including both oxy- and deoxy-Hb information to a predictive model improves accuracy. They also found that increasing the number of information channels improves accuracy. Broadly speaking, because the mean change in oxy-Hb is often used in the statistical comparison of fNIRS signals, it stands to reason that this feature is a good starting point for input to a model.

Using this method, context is important. Cui et al. (2010) note that the features they chose were "necessarily dependent" on the classification technique (in their case, support vector machines), and may be dependent on the task. One interesting distinction is that each of the previous examples is event-related—they observe how the fNIRS signals respond following a discrete moment in time. However, to serve as input to biocybernetic systems, it is often desirable for evaluators to view a moment-by-moment picture of workload, introducing new challenges. For example, task starting times, end times, and length may be undefined.

Given these challenges, an alternative approach is to select a large feature space. Then, using the participant's data, automatically determine which features yield the most information for each individual (for example, Luu et al. 2008). While this method often results in higher cross-validation classification rates, there is a danger that we may be building a model that succeeds on a particular dataset rather than one that represents a more general user state. We must take care not to overfit the model to the user's data, making it less flexible and robust for real world environments. In the next section, we describe how research is attempting to construct more generalized models of user state for real-time monitoring.

Classifying Periods of Unknown User State

Unfortunately, in normal user evaluation, researchers often do not know in advance whether a period of interaction *should be* low or high workload, or even when a period of low or high workload may begin. To help solve this problem, Hirshfield et al. (2009) proposed a methodology for using machine-learning techniques to classify user state in these scenarios.

1. Choose cognitive benchmark tasks from the psychology literature that are known to induce specific user states. For example, if we are investigating the level of verbal working memory that a visual environment might induce, we might run a participant on a demanding 0-back and 3-back task, representing low and high levels of verbal working memory.
2. Next, we build a machine learning classifier to identify and store a cognitive footprint of the fNIRS signal during each level of the benchmark task. We make the assumption that we have stored an accurate representation of verbal working memory for this particular user.
3. Finally, the participant performs a set of tasks in a more complex environment. We run the fNIRS data from those tasks on the classifier we built in the previous steps. The idea is that we are comparing the fNIRS data from this complex environment to the patterns we identified in the cognitive benchmark tasks. Our machine learning classifier returns whether the signal most closely matches low, medium, or high verbal working memory.

Hirshfield used this methodology to explore the working memory demand in a driving task in which the steering controls were reversed (Hirsfield et al. 2011). In comparison to more natural steering controls, the model classified incoming fNIRS data during the reversed control condition as requiring high working-memory. In the same way, Hirshfield used the Stroop test to detect response inhibition during interaction with an interfaces. We use this general structure as a foundation to move towards adaptive brain-computer interfaces.

Using fNIRS in Biocybernetic Interfaces

One of the primary reasons we focus on algorithms that automatically classify the fNIRS signal is that they enable fNIRS to be used as input to intelligent, adaptive systems. Because fNIRS is lightweight and does not place any unreasonable restrictions on the user, it can feasibly be used to augment many operator stations. In a real-time adaptive system, fNIRS data may be used as an additional implicit or explicit input, relaying information about the user to the computer without any further work for the user.

However, while fNIRS devices typically have a sharp temporal resolution, as we have discussed, the biological signal is sluggish. As a result, there are limitations to the systems that can be constructed. In this section, we discuss adaptive systems that use fNIRS as input, the limitations of such systems, and also identify domains where these systems may thrive.

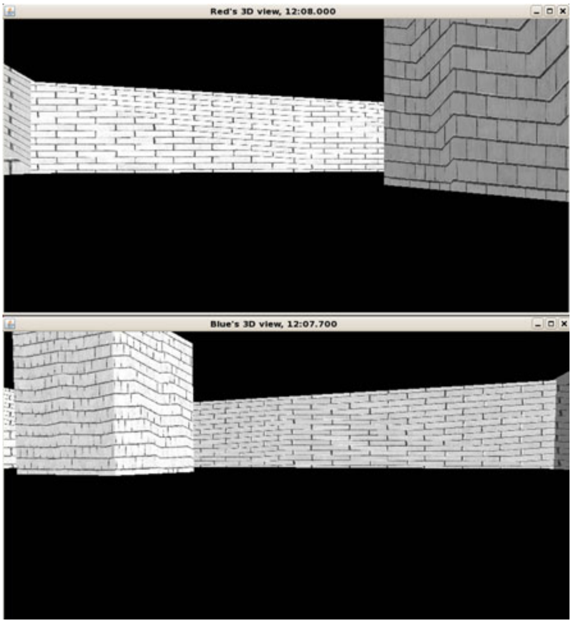

Fig. 6.8 3D View from the robots' perspective in Solovey's multi-tasking navigation environment. Automation was turned on or off based on the user's workload

Calibration and Training

Similar to the training method previously described (Hirshfield et al. 2009), in order to create a successful biocybernetic system, users must first perform a task with a known cognitive effect in order for the system to calibrate to the characteristics of brain patterns of each individual. However, the model must be very cautious since users are often in a different mental state during offline calibration and online feedback (Vidaurre et al. 2010).

Ideally, we strive to minimize training time and maximize classification accuracy. Unfortunately, these two objectives typically compete with each other. Many machine learning algorithms are traditionally designed with the assumption that there are hundreds or thousands of training examples. But in an experimental setting, training the user for hours on end is unreasonable. As a result, researchers typically train users for as long as the ordinary time constraints of a user study allows. However, as more research is done in this field, we may find universal patterns that allow us to circumvent the training period. For example, using fNIRS, Herff et al. (2012) found neural responses to speaking modes to be consistent enough to construct a general classifier of accuracy of 71 %. We suspect that similar general models may be constructed for classification of other user states.

Brainput: A Real-Time fNIRS System

In this section, we give a concrete example of previous work that uses passive fNIRS input to an intelligent system. Solovey et al. (2012) created a system, *Brainput*, which was able to adapt a scenario where an interactive human-robot system changed its state of autonomy based on whether it detected a particular state of multitasking (Fig. 6.8).

To train the system to detect these states, they used a well-validated multi-tasking exercise that had previously been explored using fMRI. Participants were shown either lower-case or upper-case letters of the word 'tablet'. Depending on the case of the letter, participants were instructed to perform different actions, thereby resembling a multi-tasking environment. In a previous study (which we described earlier), Solovey et al. (2011) showed that this task could be used to identify multi-tasking scenarios with fNIRS.

In the testing task, users were instructed to direct two robots through a virtual environment to search for areas with strong transmission strengths, and were told to not let the robots go idle or collide with walls or objects in the environment (Fig. 6.8). fNIRS signals from the participants' prefrontal cortex were collected and classified as one of two user states that described the multi-tasking load associated with navigation. The second robot was autonomously controlled whenever the system detected a state of branching, where the user must hold in mind goals while exploring and processing secondary goals (Solovey et al. 2012). The changes occurred in real-time, allowing the system to dynamically respond to the user's individual, situational needs.

Solovey found that more participants completed the task with fewer collisions in this adaptive condition. However, to demonstrate that adaptation mechanism was indeed reacting to correctly classified fNIRS data, Solovey also introduced a maladaptive condition. This condition caused the system to intentionally perform the *opposite* response that should aid the user. In this maladaptive state, users did worse than in a nonadaptive condition, and had a lower overall average transmission strength than either other condition. This experiment provides a successful example of using fNIRS as input to a real-time system that intelligently adapts to the user.

Potential Domains for Adaptive Interfaces

This method has the capability to be used for many other adaptive interfaces, especially operator stations where users may be in charge of complex tasks, such as driving or controlling aerial vehicles. In driving simulators, oxygenated hemoglobin levels for users increased when they were not using cruise control (Tsunashima and Yanagisawa 2009). In addition, prefrontal activation increased when users were preparing to turn, and there were increases in oxygenation levels during a driving task when users were prompted with directions instead of knowing when to turn (Liu et al. 2012). This opens the door for driving applications that may alleviate the user's workload by delivering relevant information (such as directions) in different modes or at different times depending on the user's state.

Turning to another application area, Bunce et al. (2011) found that oxygenated hemoglobin levels increased in an unmanned aerial vehicle (UAV) commander task when the number of UAVs to identify increase. They also found that practice decreased the overall oxygenation change, with experts maintaining higher performance despite lower prefrontal activation. However, at difficult levels, novice performance and oxygenation fell, indicating that they gave up on the task, while experts showed a strong increase (Bunce et al. 2011). This suggests that neural activation can be used to determine a user's expertise in real-time, as well as when the user disengages from a task. The same trend of decreased activation for experts was found using fNIRS on ground operators performing UAVs approach and landing tasks (Izzetoglu et al. 2011). Using fNIRS to detect expertise during interaction may translate from operators to more general educational scenarios, and an adaptive system could potentially calibrate it's instructional aid based on the sensed expertise level of the user.

These measures can be applied to other domains as well. fMRI studies have shown that blood oxygen signals in the prefrontal cortex develop during economic choices where users must make gambling evaluations and expected value decisions (Minati et al. 2012) and when determining decision value for different categories of goods for purchase (Grabenhorst and Rolls 2011). Using these activation patterns, Peck et al. (2013a) demonstrated the use of fNIRS as input to information-filtering systems, constructing a movie recommendation engine that gradually personalized movies to the user based on predicted preference values.

These examples illustrate a small subset of application areas for interfaces that adapt according to physiological input. From gaming to education, there are numerous opportunities to cater the computing environment to each individual's cognitive state.

Challenges of Application

It's important to note that building these systems is not without significant challenge. A real-time adaptive system must filter raw data and determine state as quickly as possible in order to respond before the user's state changes and the adaptation becomes obsolete. Preferably, the system should be flexible and robust in that it is able to detect the user's workload in a variety of environments and tasks.

Aside from the difficulties in classification, the relatively sluggish fNIRS signal limits the appropriate adaptive responses of an interface. For example, using fNIRS to directly control a mouse cursor would likely prove to be frustrating for a user, as their actions would be delayed by 5–7 s. In general, we suggest the use of slow, gentle modifications to the system which the user does not (or hardly) perceive. For example, Solovey's (2012) Brainput system increased performance by adjusting the level of automation. Peck et al. (2013a) recommendation system changed *which* information was emphasized to the user. In both cases, individual changes were imperceptible to the user. These studies suggest that manipulating background processes may present an opportunity for designers to positively impact the user without disruptive adaptation mechanisms. Additionally, these changes minimize the impact of misclassification. Given the current challenges in achieving high classification rates, it is important that incorrect guesses do not result in disrupting the user's workflow or mental model of the system.

Finally, while a primary benefit of measuring activity in the prefrontal cortex stems from its involvement in many high-level user states, the general participation of the PFC in cognition is also a challenge. A variety of user states, from economic judgments to emotion to working memory load, have been correlated with the PFC. As we move the use of fNIRS into increasingly complex environments, it will be important to understand how the interaction of these states impacts changes in oxy-Hb and deoxy-Hb, as it may become difficult to identify them individually.

The Future of fNIRS

Although the work that we have discussed suggests a positive trajectory in using fNIRS to monitor user state, we have also been candid with some of the present challenges in using fNIRS in real world environments. Despite these challenges, recent advances hint that fNIRS will soon become even more lightweight and comfortable.

Moving towards an untethered system, a wireless setup of fNIRS has been implemented by a team at Drexel University (Yurtsever et al. 2003). Other research has seen the development of brush optrodes—sensors that are able to navigate through hair and comfortable measure numerous areas of the brain (Wildey et al. 2010). Finally, there has been preliminary work in no-contact fNIRS sensors, or remote sensing of the brain (Sase et al. 2012). Although these are early proof-of-concept studies, they suggest that someday, optical brain sensing may not require probes of any kind. While the equipment in our laboratory is expensive, the basic technology can eventually be implemented at greatly reduced costs since it consists fundamentally of simply a light source and an optical detector. Given these recent advances, we expect that fNIRS will continue to be considered a viable technology for brain sensing in the coming decades.

Conclusion

In this chapter, we reviewed work that has used functional near-infrared spectroscopy to observe changes in user workload during interaction with a computer interface. We began by describing basic, highly-controlled experiments in which fNIRS recorded changes in workload in classical psychology tasks (the n-back). From there, we moved towards tasks that more closely mirror tasks that users face in everyday life. However, the analysis was performed manually and offline, creating a burden on evaluators that is not ideal for industrial user-studies. Finally, we shared work that uses predictive models to automatically identifies the user's state, requiring little intervention from the evaluator.

Turning an eye towards the future, we painted a picture in which these classifications of user state could be used as passive input to real-time adaptive interfaces. We believe that these interfaces have the potential to specially calibrate the user's computing environment to their individual skills and abilities.

Overall, use of fNIRS is not without its challenges. Signal classification is difficult, and systems must push for increased accuracy in increasingly noisy environments. However, fNIRS provides a unique source of input that can be seen as complimentary to other physiological sensors. We believe that the information that we can gain from fNIRS brain sensing can contribute to evaluation in real world environments, as well as personalization in real-time environments.

References

Ayaz H, Shewokis PA, Bunce S, Izzetoglu K, Willems B, Onaral B (2012) Optical brain monitoring for operator training and mental workload assessment. NeuroImage 59(1):36–47

Baddeley AD (1992) Working memory. Science 255(5044):556–559

Bor D, Cumming N, Scott CEL, Owen AM (2004) Prefrontal cortical involvement in verbal encoding strategies. Eur J Neurosci 19:3365–3370

Bor D, Duncan J, Wiseman RJ, Owen AM (2003) Encoding strategies dissociate prefrontal activity from working memory demand. Neuron 37:361–367

Braver TS, Cohen JD, Nystrom LE, Jonides J, Smith EE, Noll DC (1997) A parametric study of prefrontal cortex involvement in human working memory. Neuroimage 5:49–62

Bunce SC, Izzetoglu K, Ayaz H, Shewokis P, Izzetoglu M, Pourrezaei K, Onaral B (2011) Implementation of fNIRS for monitoring levels of expertise and mental workload. In: Foundations of Augmented Cognition. Directing the Future of Adaptive Systems. Springer, Berlin, pp 13–22

Burgess PW, Quayle A, Frith CD (2001) Brain regions involved in prospective memory as determined by positron emission tomography. Neuropsychologia 39:545–555

Cahill L, Uncapher M, Kilpatrick L, Alkire MT, Turner J (2004) Sex-related hemispheric lateralization of amygdala function in emotionally influenced memory: an fMRI investigation. Learn Mem 11(3):261–266

Chance B, Anday E, Nioka S, Zhou S, Hong L, Worden K, Li C et al (1998) A novel method for fast imaging of brain function, non-invasively, with light. Opt Express 2(10):41123

Christoff K, Gabrieli JDE (2000) The frontopolar cortex and human cognition: evidence for a rostrocaudal hierarchical organisation within the human prefrontal cortex. Psychobiology 28:168–186

Christoff K, Prabhakaran V, Dorfman J, Zhao Z, Kroger JK, Holyoak KJ, Gabrieli JD (2001) Rostrolateral prefrontal cortex involvement in relational integration during reasoning. Neuroimage 14(5):1136–1149

Cleveland WS, McGill R (1984) Graphical Perception: Theory, experimentation, and the application to the development of graphical methods. J Am Stat Assoc 387:531–554

Cohen JD, Perlstein WM, Braver TS, Nystrom LE, Noll DC, Jonides J, Smith EE (1997) Temporal dynamics of brain activation during a working memory task. Nature 386:604–608

Cui X, Bray S, Reiss A (2010) Speeded near infrared spectroscopy (NIRS) response detection. PLoS One 5(11):e15474

Davis MH, Meunier F, Marslen-Wilson WD (2004) Neural responses to morphological, syntactic, and semantic properties of single words: an fMRI study. Brain Lang 89(3):439–449

D'Esposito M, Zarahn E, Aguirre G (1999) Event-related functional MRI: implications for cognitive psychology. Psychol Bull 125(1):155–164

Dove A, Rowe JB, Brett M, Owen AM (2001) Neural correlates of passive and active encoding and retrieval: a 3T fMRI study. Neuroimage 13(Suppl):660

Franceschini MA, Joseph DK, Huppert TJ, Diamond SG, Boas DA (2006) Diffuse optical imaging of the whole head. J Biomed Opt 11(5):054007

Gevins AS, Cutillo BC (1993) Neuroelectric evidence for distributed processing in human working memory. Electroencephalogr Clin Neurophysiol 87:128–143

Girouard A, Solovey E, Hirshfield L, Chauncey K, Sassaroli A, Fantini S, Jacob RJK (2009) Distinguishing difficulty levels with non-invasive brain activity measurements. Interact 2009:440–452

Gore JC (2003) Principles and practice of functional MRI of the human brain. J Clin Investig 112(1):4–9

Grabenhorst F, Rolls ET (2011) Value, pleasure and choice in the ventral prefrontal cortex. Trends Cogn Sci 15(2):5667

Herff C, Heger D, Putze F, Guan C, Schultz T (2012) Cross-subject classification of speaking modes using fNIRS. ICONIP 2012:417–424

Hirshfield LM, Solovey ET, Girouard A, Kebinger J, Jacob RJK, Sassaroli A, Fantini S (2009) Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy. In: CHI 2009

Hirshfield LM, Gulotta R, Hirshfield S, Hincks S, Russell M, Ward R, Williams T, Jacob RJK (2011) This is your brain on interfaces: enhancing usability testing with functional near-infrared spectroscopy. In: CHI 2011

Hockey GRJ (1997) Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol Psychol 45:73–93

Izzetoglu K, Ayaz H, Menda J (2011) Applications of functional near infrared imaging: case study on UAV ground controller. In: Schmorrow DD, Fidopiastis CM (eds) Foundations of augmented cognition. Springer, New York, pp 608–617

Jonides J, Smith EE, Koeppe RA, Awh E, Minoshima S, Mintun MA (1993) Spatial working memory in humans as revealed by PET. Nature 363:623–625

Koechlin E, Corrado G, Pietrini P, Grafman J (2000) Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. Proc Nat Acad Sci. 97(13):7651–7656

Kroger JK, Sabb FW, Fales CL, Bookheimer SY, Cohen MS, Holyoak KJ (2002) Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: a parametric study of relational complexity. Cereb Cortex 12:477–485

Liu T, Saito H, Oi M (2012) Distinctive activation patterns under intrinsically versus extrinsically driven cognitive loads in prefrontal cortex: a near-infrared spectroscopy study using a driving video game. Neuroscience letters, 506(2):220–224

Luu S, Chau T (2008) Decoding subjective preference from single-trial near-infrared spectroscopy signals. J Neural Eng 6:058001

Miller G (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 63(2):8197

Minati L, Grisoli M, Franceschetti S, Epifani F, Granvillano A, Medford N, Harrison N et al (2012) Neural signatures of economic parameters during decision-making: a functional MRI (FMRI), electroencephalography (EEG) and autonomic monitoring study. Brain Topogr 25(1):73–96

Moghimi S, Kushki A, Power S, Guerguerian AM, Chau T (2012) Automatic detection of a prefrontal cortical response to emotionally rated music using multi-channel near-infrared spectroscopy. J Neural Eng 9(2):026022

Owen AM, McMillan KM, Laird AR, Bullmore E (2005) N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. Hum Brain Mapp 25(1):46–59

Peck EM, Afergan D, Jacob RJK (2013a) Investigation of fNIRS brain sensing as input to information filtering systems. In: Augmented human 2013

Peck EM, Yuksel BF, Ottley A, Jacob RJK, Chang R (2013b) Using fNIRS brain sensing to evaluate information visualization interfaces. In: CHI 2013

Ramnani N, Owen AM (2004) Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nat Rev Neurosci 5:184–194

Repovš G, Baddeley A (2006) The multi-component model of working memory: explorations in experimental cognitive psychology. Neuroscience 139:5–21

Repovš G, Bresjanac M (2006) Cognitive neuroscience of working memory: a prologue. Neuroscience 139:1–3

Rugg MD, Fletcher PC, Allan K, Frith CD, Frackowiak RS, Dolan RJ (1998) Neural correlates of memory retrieval during recognition memory and cued recall. Neuroimage 8:262–273

Sase I, Takatsuki A, Seki J, Yanagida T, Seiyama A (2012) Noncontact backscatter-mode near-infrared time-resolved imaging system: preliminary study for functional brain mapping. J Biomed Opt 11(5):054006

Solovey ET, Girouard A, Chauncey K, Hirshfield LM, Sassaroli A, Zheng F, Fantini S, Jacob RJK (2009) Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines. In: UIST 2009

Solovey ET, Lalooses F, Chauncey K, Weaver D, Scheutz M, Sassaroli A, Fantini S, Jacob RJK (2011) Sensing cognitive multitasking for a brain-based adaptive user interface. In: CHI 2011

Solovey ET, Schermerhorn P, Scheutz M, Sassaroli A, Fantini S, Jacob RJK (2012) Brainput: enhancing interactive systems with streaming fNIRS brain input. In: CHI 2012

Strangman G, Culver JP, Thompson JH, Boas DA (2002) A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. NeuroImage 17(2):719731

Tsunashima H, Yanagisawa K (2009) Measurement of brain function of car driver using functional near-infrared spectroscopy (fNIRS). Comput Intell Neurosci 2009:164958

Tulving E (1983) Elements of episodic memory. Clarendon, Oxford

Wickens CD (2002) Multiple resources and performance prediction. Theor Issues Ergon Sci 3:159–177

Wildey C, MacFarlane D, Khan B, Tian F, Liu H, Alexandrakis G (2010) Improved fNIRS using a novel brush optrode. In: Laser science

Vidaurre C, Sannelli C, Muller K-R, Blankertz B (2010) Machine-learning-based coadaptive calibration for brain-computer interfaces. Neural Comput 816:791816

Villringer A, Chance B (1997) Non-invasive optical spectroscopy and imaging of human brain function. Trends Neurosci 20(10):43542

Yurtsever G, Ayaz H, Kepics F, Onaral B (2003) Wireless, continuous wave near infrared spectroscopy system for monitoring brain activity. In: Bioengineering conference, pp 53–53