REAL-TIME FNIRS BRAIN INPUT FOR ENHANCING INTERACTIVE SYSTEMS

A dissertation

submitted by

Erin Treacy Solovey

In partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Computer Science

TUFTS UNIVERSITY

May, 2012

Advisor: Robert J.K. Jacob

ABSTRACT

Most human-computer interaction (HCI) techniques cannot fully capture the richness of the user's thoughts and intentions when interacting with a computer system. For example, when we communicate with other people, we do not simply use words, but also accompanying cues that give the other person additional insight to our thoughts. At the same time, several physiological changes occur that may or may not be detected by the other person. When we communicate with computers, we also generate these additional signals, but the computer cannot sense such signals, and therefore ignores them. Detecting these signals in real time and incorporating them into the user interface could improve the communication channel between the computer and the human user with little additional effort required of the user. This communication improvement would lead to technology that is more supportive of the user's changing cognitive state. Such improvements in bandwidth are increasingly valuable, as technology becomes more powerful and pervasive, while our cognitive abilities do not change considerably.

In this dissertation, I explore using brain sensor data as a passive, implicit input channel that expands the bandwidth between the human and computer by providing supplemental information about the user. Using a relatively new brain imaging tool called functional near-infrared spectroscopy (fNIRS), we can detect signals within the brain that indicate various cognitive states. This device provides data on brain activity while remaining portable and non-invasive. This research aims to develop tools to make brain sensing more practical for HCI and to demonstrate effective use of this cognitive state information as supplemental input to interactive systems.

First, I explored practical considerations for using fNIRS in HCI research to determine the contexts in which fNIRS realistically could be used. Secondly, in a series of controlled experiments, I explored cognitive multitasking states that could be classified reliably from fNIRS data in offline analysis. Based on these experiments, I created *Brainput,* a system that learns to identify brain activity patterns occurring during multitasking. It provides a continuous, supplemental input stream to an interactive human-robot system, which uses this information in real time to modify its behavior to better support multitasking. Finally, I conducted an experiment to investigate the efficacy of *Brainput* and found improvements in performance and user experience.

ACKNOWLEDGMENTS

There are quite a few people who I would like to acknowledge for their support throughout my work on this dissertation.

Special thanks to:

My advisor, Rob Jacob, for his encouragement, guidance, friendship, and mentorship. I am lucky to have had his support and to have learned so much from his wisdom and invaluable advice.

The Tufts Human-Computer Interaction Lab (past and present). To Audrey Girouard for her collaboration, attention to details, and friendship. To Leanne Hirshfield for getting this project off the ground. To Orit Shaer, Michael Horn, Michael Poor, Jamie Zigelbaum, Jordan Crouser, Evan Peck, Francine Lalooses, Krysta Chauncey, Margarita Parasi, Douglas Weaver, Megan Strait, Ruth Agbaji, Dan Afergan, Hadar Rosenhand, Kelly Moran, Claire Lee, Matthew Knowles, Rebecca Gulotta, James Kebinger, and Kristof Redei for their discussions, suggestions, and collaboration.

The Tufts Computer Science Department staff, faculty, and students. In particular, thanks to Diane Souvaine, Carla Brodley, Roni Khardon, Eric Miller, Lenore Cowan, Umaa Rebbapragada, Eddie Aftandilian, Elena Jakubiak, Amelio Vazquez-Reina, Alexandra Lauric, Jeannine Vangelist, Donna Cirelli, Gail Fitzgerald, and George Preble.

Desney Tan for welcoming me to Microsoft Research, for mentoring me and providing words of wisdom, for sparking new ideas and inspiring me to explore challenging research. Dan Morris, Ed Cutrell, Krzysztof Gajos, Ashish Kapoor, and the rest of

Microsoft Research for brainstorming and discussing research ideas. To the other 2008 Microsoft Research summer interns for their camaraderie and good times.

Matthias Scheutz for his collaboration, advice, and for being part of my thesis committee. Paul Schermerhorn for his collaboration with the robot architecture.

Remco Chang for being on my committee, reading my thesis, for providing valuable advice and feedback.

Sergio Fantini and Angelo Sassaroli for their collaboration and expert tips on fNIRS, along with other researchers in their group: Feng Zheng and Yunjie Tong.

The National Science Foundation and Clare Boothe Luce Foundation for financial support of my research.

My parents, Bob and Barbara, for a lifetime of encouragement and support, and for providing me with such a strong foundation for my life. My sisters, Denise and Rebecca, for their friendship and wonderful memories – we are lucky to have each other. My grandparents, aunts, uncles, cousins, and brother-in-law, for all the fun times. My in-laws, Alex, Simona, Lisa, Rosa, and the rest of the extended Solovey family – I am fortunate to have become part of such a wonderful family.

My daughters, Jaclyn and Kailah, for amazing me all the time, for making me laugh constantly, and for the joy they bring me as they grow.

My husband, Ed, for his unyielding support, encouragement, understanding, friendship and love. For making me smile with his warmth and humor which affects everyone around him. For his loving dedication to his family, for sharing his life with me, and for always standing beside me in this adventure.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 INTRODUCTION

1.1. Motivation

Over the past fifty years, computers have gained power and efficiency, and can now process massive amounts of information at high speeds. Humans, on the other hand, have not witnessed such dramatic improvements. To make humans more effective when they interact with computer systems, we devise novel human-computer interaction techniques. Early systems used punch cards, and later, command line interfaces. Today, the mouse and keyboard are ubiquitous input devices, while graphical displays on monitors are used for transmitting information from the system to the user. However, these techniques do not fully capture the richness of the user's thoughts and intentions when interacting with a computer system.

For example, when we communicate with other people, we do not simply use words, but also accompanying visual and auditory cues that give the other person additional insight to our thoughts. At the same time, several physiological changes occur that may or may not be detected by the other person. When we communicate with computers, we also generate these additional signals, but the computer cannot sense such signals, and therefore completely ignores them. Detecting these signals in real time and incorporating them into the user interface could improve the communication channel between the computer and the human user with little additional effort required of the user. This communication improvement would lead to technology that is more supportive of the user's changing cognitive state. Such improvements are increasingly valuable, as technology becomes more powerful and pervasive, while our cognitive abilities do not change considerably.

In order to automatically infer the user's changing cognitive state in real time, some researchers have explored performance data, interaction history (e.g. keystrokes) or environmental context to assess the user's current state (Fogarty, Hudson, & Lai, 2004; Hudson et al., 2003; Starner, Schiele, & Pentland, 1998), while others use computer vision to detect facial expressions or other behavioral measures. Physiological measures are also emerging as continuous indicators of cognitive state changes (Fairclough, 2009; R. Mandryk, Atkins, & Inkpen, 2006; Nacke, Kalyn, Lough, & Mandryk, 2011). Brain imaging and brain sensing techniques aim to get close to the source by looking at changes in brain activity during task performance (Grimes, Tan, Hudson, Shenoy, & Rao, 2008; Leanne M. Hirshfield et al., 2009) and are becoming realistic tools for HCI research.

Progress in brain imaging has opened the door for promising research on brain-computer interfaces. For example, users without motor control or speech can currently use a virtual keyboard (Kennedy, Bakay, Moore, Adams, & Goldwaithe, 2000) and navigate in their environment (Millán, Renkens, Mouriño, & Gerstner, 2004) using mental motor imagery. Such systems usually are designed with brain activity as the primary, and often only, input to the system. Users concentrate on a certain type of thought (such as imagined hand movement) in order to control the system. This requires concentration, effort, and training, and often seems unnatural. Some require implanted electrodes in the skull (Kennedy, et al., 2000; M. Moore, Kennedy, Mynatt, & Mankoff, 2001; Melody M. Moore & Kennedy, 2000) or long training periods with limited bandwidth (Millán, et al., 2004). While these systems are valuable to paralyzed and locked in patients, they do not provide sufficient gains to healthy users to make the effort required worthwhile.

Here, I take a different approach for brain-computer interfaces that augments traditional input devices such as the mouse and keyboard and that targets a wider group of users. I

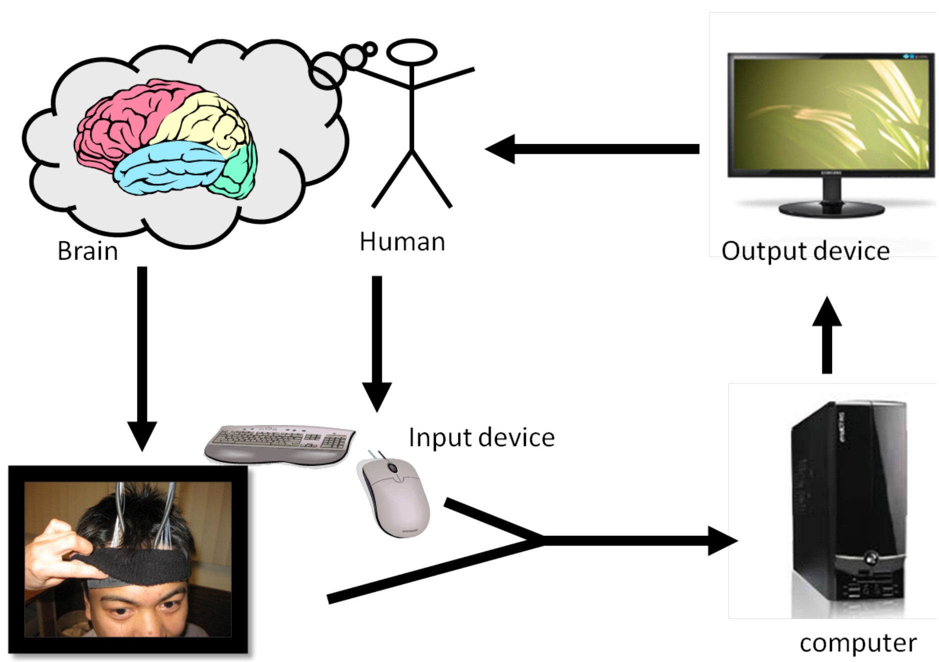


Figure 1-1. FNIRS as a passive, implicit input channel that supplements the primary input to an interactive system.

use brain sensor data as a passive, implicit input channel that expands the bandwidth between the human and computer by providing supplemental information about the user (Figure 1-1). Using a relatively new brain imaging tool called functional near-infrared spectroscopy (fNIRS) (Chance et al., 1988; Villringer, Planck, Hock, Schleinkofer, & Dirnagl, 1993) (Figure 1-2), we can detect signals within the brain that indicate various cognitive states. This device provides data on brain activity while remaining portable and non-invasive. This research aims to develop tools to make brain sensing more practical for HCI and to demonstrate effective use of this cognitive state information as supplemental input to an interactive system.

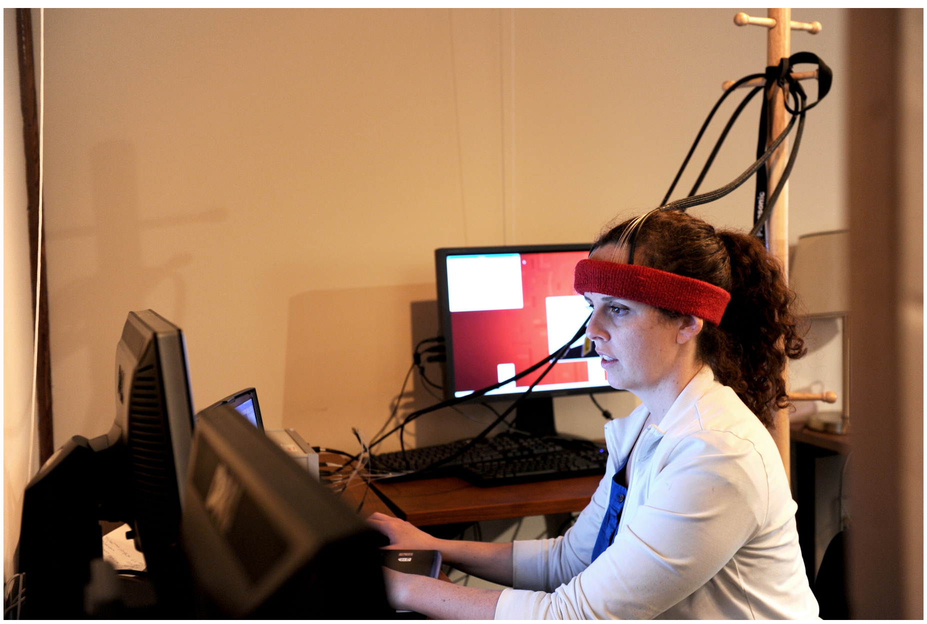


Figure 1-2. Two functional near-infrared spectroscopy sensors are placed under the red headband and non-invasively detect brain activity.

1.2. An Example

In order to explore the potential of fNIRS brain sensing in HCI, I began exploring specific applications where this type of passive, supplemental input may be worthwhile. My goal was to build a working platform for studying these types of systems and to target a key use case for deeper study. One area where such brain-based interfaces would be beneficial is in the support of users who are multitasking, and I investigate the feasibility of measuring cognitive multitasking states with fNIRS in Chapter 4. There is a wide range of contexts that involve information overload, interruptions or multitasking. As a proof-of-concept, in Chapter 5 and Chapter 6, I demonstrate and evaluate a human-robot system that utilizes the fNIRS cognitive multitasking input stream to support the supervision of multiple robots in a team task. Below, I further discuss the use of these domains for exploring fNIRS as a supplemental input stream to an interactive system.

1.2.1. Multitasking Support in Interactive Systems

Multitasking has become an integral part of work environments, even though people are not well-equipped to effectively handle more than one task at a time (Miyata & Norman, 1986b). While multitasking has been shown to be detrimental to performance in individual tasks (Miyata & Norman, 1986b), it can also be beneficial when a secondary task provides additional information for completing the primary task, such as allowing people to integrate information from multiple sources.

Multiple windows, multiple monitors and large displays make it possible for the interface to handle multitasking, and many researchers have investigated how best to support the user who is balancing multiple tasks. Because multitasking can elicit several different cognitive states, the user's needs during multitasking may change over time. However, it is difficult to determine the best way to support the user without understanding the internal cognitive processes occurring during task performance. Recognizing signals generated naturally by the user that differentiate different types of multitasking could lead to higher productivity, better task performance, and improved user experience when the signals are utilized to make the system more responsive to the user's needs.

1.2.2. Human-Robot Interaction

Recent advances in artificial intelligence and robotics have led to the development of autonomous robots that can work closely with human operators to complete tasks. Understanding and improving the interactions during such mixed human-robot team tasks is a key research area in the growing field of human-robot interaction. Many such human-robot team tasks also provide appropriate scenarios for studying adaptive multitasking support, as they inherently involve multitasking: the user is performing a task, while also monitoring the state of the robot(s). Such human-robot team tasks thus may see improved performance with brain-based adaptive interfaces. There has been much work on adaptive robots that change behavior based on the environment or situation. In Chapter 5 and Chapter 6, I demonstrate how we could develop robots that have a greater understanding of the user's cognitive state during multitasking, and that can adapt their behavior to better support the user, based on this supplemental cognitive state information.

1.3. Thesis Statement

In this dissertation, I claim that:

Functional near-infrared spectroscopy, an emerging brain-sensing technology, can infer passive cognitive state and provide real-time input that allows an interactive user interface to adapt its behavior, thus improving user performance and experience compared to a traditional user interface.

To demonstrate this thesis, my research had four phases. First, I explored practical considerations for using fNIRS in HCI research to determine the contexts in which fNIRS realistically could be used. Secondly, in a series of controlled experiments, I explored the cognitive states that could be classified reliably from fNIRS data in offline analysis, focusing on multitasking scenarios. This involved understanding brain activation profiles in the anterior prefrontal cortex, and developing preprocessing, visualization and machine learning techniques for analyzing the fNIRS data. Based on these experiments, I created *Brainput,* a system that learns to identify brain activity patterns occurring during multitasking. It provides a continuous, supplemental input stream to an interactive human-robot system, which uses this information in real time to modify its behavior to better support multitasking. Finally, I conducted an experiment to investigate the efficacy of *Brainput*.

1.4 Thesis Contributions

To support the thesis statement, my interdisciplinary research touches the fields of machine learning, signal processing, brain-computer interfaces, biomedical engineering, human-robot interaction, as well as HCI. It makes several contributions that lay a foundation for future HCI research by overcoming many of the technical challenges and bringing brain sensing for HCI to a point where concrete research and evaluation can be conducted. In particular, with this dissertation, I make the following contributions:

1) **fNIRS Guidelines:** I facilitate further adoption of fNIRS brain sensing in HCI research by providing practical guidelines and considerations for its effective use, based on past experience and experimental evidence.

2) **fNIRS Analysis tools:** I describe visualization, analysis and classification tools for fNIRS that work for offline analysis as well as in real time systems.

3) **Offline fNIRS Multitasking Study:** I show that specific cognitive multitasking states, previously studied with fMRI (which cannot be used in HCI settings), can be detected automatically with fNIRS which is more practical for HCI. I also show that these cognitive multitasking brain processes are detectable across multiple domains and tasks, by moving from a simple letter-based task in previous work to actual HCI-

related tasks that elicit similar states. These processes are almost indistinguishable by examining overt behavior or task performance alone. I explored these cognitive multitasking states because they have direct relevance to many HCI scenarios.

4) **Streaming fNIRS input channel:** I describe *Brainput*, a passive, implicit input channel to an interactive system, based on real-time cognitive multitasking state detection with fNIRS. This system was integrated with a human-robot system. Together, this platform provides the basis for the design and evaluation of future brain-based adaptive user interfaces, with broader applications beyond human-robot team tasks.

5) **System Evaluation and User Study:** I present results of a user study showing that *Brainput* significantly improves several performance metrics, as well as the subjective scores in a dual-task human-robot activity, while requiring no additional effort from the user. This study also confirmed that we can train a machine learning classifier on a set of known tasks and later successfully classify brain activity in unrelated activities that generate similar brain processes. This suggests that implicit brain input as a supplemental input stream has promise both in human-robot interaction and in various other domains and tasks.

6) **Recommendations:** I make recommendations for designing interfaces that can take advantage of a supplementary, implicit input channel such as that coming from fNIRS.

1.5. Thesis Overview

This dissertation is organized as follows:

Chapter 2 provides an overview of related work that lays the foundation for this dissertation. This includes prior work with brain sensing in general, as well as the specifics of functional near-infrared spectroscopy. It also covers prior brain-computer interface research as well as research into the brain processes occurring during multitasking, which is fundamental to this work.

Chapter 3 describes a series of experiments investigating practical aspects of using fNIRS in HCI research, and provides guidelines for its effective use in HCI contexts.

Chapter 4 describes several controlled experiments conducted to determine whether there are patterns in the fNIRS brain signals that could differentiate between various types of multitasking behavior, which would be valuable in HCI research. It also describes tools developed for offline signal processing and analysis of the fNIRS signal.

Chapter 5 contains a description of *Brainput*, a working system that uses fNIRS as a passive, implicit input channel to an interactive human-robot interaction system. To support this, I created tools for real time analysis and classification that can be used in other contexts, and these are explained in this chapter as well.

Chapter 6 details the evaluation experiment of *Brainput,* in which performance data showed improved user performance using this input modality in the human-robot system. It also provides evidence from subjective questionnaires showing that this input modality improved the users' perceived workload and experience.

Chapter 7 summarizes the main contributions of this work and discusses future directions.

Chapter 2 BACKGROUND & RELATED WORK

This chapter lays the foundation for this dissertation by discussing related work in several areas. It begins by discussing the state of the art in brain-computer interface research and then goes into functional near-infrared spectroscopy background. Section 2.1.5. discusses brain sensing in human-robot interaction. Then, I cover related prior research on brain processes occurring during multitasking, which is fundamental to this work. Other related work appears throughout the dissertation when it is closely connected to a particular section.

2.1. Brain-Computer Interface Research

2.1.1. Brain Sensing and Imaging

Non-invasive brain sensing and imaging techniques, primarily developed for clinical settings, have been powerful tools for understanding brain structure and function as well as for diagnosing brain injuries or disorders. Structural imaging techniques, such as computed tomography (CT), generate brain images of the mostly static structure of the brain, as well as brain tumors and injuries. These provide valuable snapshots of the state of the brain, but are not used in brain-computer interfaces, which require measurement of the changing state of the brain due to cognitive activity. Functional imaging detects changes within the brain during various activities, and is used to understand brain function and brain illnesses. Functional magnetic resonance imaging (fMRI) is widely used to generate 3-dimensional images of the brain showing the blood oxygen level dependence (BOLD) effect, which measures changes in volume and oxygenation of the blood. These hemodynamic changes in the brain are an indirect measure of the activity in the brain. Similar to fMRI, functional near-infrared spectroscopy (fNIRS), measures blood oxygen changes, and is discussed in detail below. Positron emission tomography (PET) scans provide 3-dimensional images of blood flow, blood oxygen and metabolic function of cells, but is mainly used for investigating organs for cancers and other diseases. Electroencephalography (EEG) and magnetoencephalography (MEG) provide a more direct measure of neuronal activity by detecting electrical signals generated by neurons firing. For a table comparing these brain sensing technologies for use in HCI, see (Tan & Nijholt, 2010).

Since these tools were designed for use in clinical or laboratory settings, they often require restrictions on the patient or study participant. Most of these restrictions are not reasonable for realistic HCI settings. Besides being expensive, PET, fMRI and MEG require subjects to sit or lay down in unnatural positions and remain essentially motionless (Lee & Tan, 2006). In addition, PET requires ingestion of hazardous material and fMRI exposes subjects to loud noises that may interfere with the study (M Izzetoglu, Izzetoglu, Bunce, Onaral, & Pourrezaei, 2005). Plus, the powerful magnetic field prevents computer usage in both fMRI and MEG. These factors make it impractical to use these techniques in a realistic interactive situation.

Because it is less intrusive, more portable, and less expensive than these other technologies, EEG (Figure 2-1) has seen wide use in BCI research. For example, it has been used to classify tasks (Lee & Tan, 2006), measure cognitive load (Grimes, et al., 2008), support human-aided computer vision (Shenoy & Tan, 2008), as well as limited communication (Keirn & Aunon, 1990; Schalk, McFarland, Hinterberger, Birbaumer, & Wolpaw, 2004; Wolpaw, McFarland, Neat, & Forneris, 1991). However, it can have a significant setup time, and electronic devices in the room can interfere with the signal. It has limited spatial resolution, but high temporal resolution. In addition, most EEG systems require gel to be applied to the scalp, although devices are being developed that use dry electrodes. Because these disadvantages are not prohibitive, EEG has been the main technology used in brain-computer interface research.

Figure 2-1. fNIRS (left) and EEG (right) provide useful cognitive and affective state information while remaining non-invasive and practical for HCI settings.

2.1.2. Functional Near-Infrared Spectroscopy Background

My work focuses on using functional near-infrared spectroscopy (fNIRS) to overcome as well as complement some of the drawbacks of these other brain-imaging systems. However, because it is a novel technique for brain sensing, there have been few studies showing specific measurements with fNIRS and their appropriate use in HCI. The emerging, non-invasive, and lightweight sensors detect changes in oxygenated and deoxygenated blood in a region of the brain by using optical wires to emit near-infrared light (Chance, et al., 1988). The sensors are placed on the forehead and secured with a headband, making them portable, easy to use, and quick to set up—characteristics that make fNIRS suitable for use in realistic HCI settings (Figure 2-1).

Figure 2-2 shows one of the two fNIRS sensors that would be placed on a person's forehead. On the sensor shown in the photo, there are five possible light sources and one light detector. The light sources send two wavelengths of near-infrared light into the

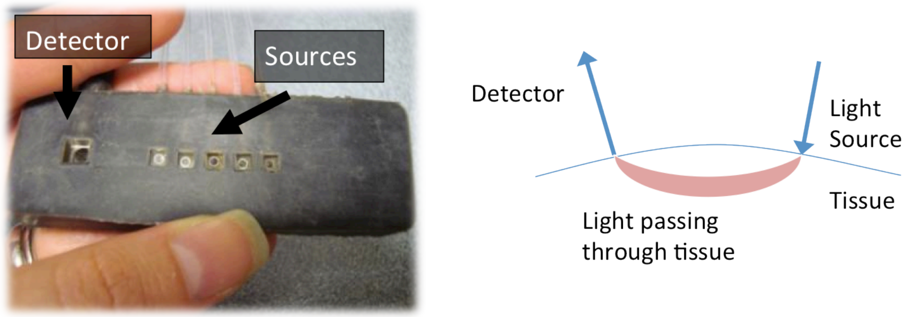


Figure 2-2. Left: One fNIRS sensor. In a typical setup, two sensors are placed on the forehead. The thin clear fibers are attached to the light sources and the black, thicker fiber is attached to the light detector. A headband holds the probes in place. Right: Illustration of path of near-infrared light between the source and detector.

forehead, where it continues through the skin and bone 1-3 cm deep into the cortex. Biological tissues are relatively transparent to these wavelengths, but the oxygenated and deoxygenated hemoglobin are the main absorbers of this light. After the light scatters in the brain, some reaches the light detector. By determining the amount of light picked up by the detector, we can calculate the amount of oxygenated and deoxygenated hemoglobin in the area. Because these hemodynamic and metabolic changes are associated with neural activity in the brain, fNIRS measurements can be used to understand changes in a person's cognitive state while performing tasks.

Like most brain imaging techniques, fNIRS was designed primarily for laboratory and clinical settings. However, it avoids many of the restrictions of other techniques (as will be discussed in more depth in Chapter 3), and therefore has promise for HCI research. However, like EEG, the data can be noisy and less reliable than the more intrusive techniques (e.g. fMRI, MEG, surgically-implanted electrodes), requiring machine learning algorithms that can handle this type of data. Despite this, fNIRS and EEG open new doors for HCI research since they are safe, non-invasive, and portable, yet still provide cognitive state information.

2.1.2.1. Experimental fNIRS Setup

In all studies described in this dissertation, a multichannel frequency domain OxiplexTS from ISS Inc. (Champaign, IL) was used for data acquisition (Figure 2-3). Two probes (Figure 2-2) were placed on the forehead to measure the two hemispheres of the anterior prefrontal cortex. The source-detector distances were 1.5, 2, 2.5, 3cm respectively. Each

Figure 2-3. OxiplexTS by ISS, Inc. is used in all the experiments described in this dissertation.

distance measures a different depth in the cortex. Each source emits two light wavelengths (690nm and 830nm) to pick up and differentiate between oxygenated and deoxygenated hemoglobin. The sampling rate was 6.25Hz.

The basic technology is common to all systems, and the measured signal depends on the location of the probe and the amount of light received. The most common placements are on the motor cortex (Sitaram et al., 2007), and the prefrontal cortex (Ehlis, Bähne, Jacob, Herrmann, & Fallgatter, 2008; Mappus, Venkatesh, Shastry, Israeli, & Jackson, 2009), although other regions have also been explored (Herrmann et al., 2008). The sensors used in this research were designed for the forehead, which is one of the most successful placements because of the absence of hair, which absorbs light and degrades the fNIRS signal. Thus, the anterior prefrontal cortex, which lies behind the forehead, is the main target for fNIRS brain sensing in this dissertation. This area of the brain deals with high-level processing (Ramnani & Owen, 2004), such as working memory, planning, problem solving, memory retrieval and attention. Here, we look specifically at detecting brain activity changes during multitasking as we would like to improve user performance and experience in such difficult situations.

2.1.3. Brain-Computer Interface Approaches

Lee and Tan (Lee & Tan, 2006) describe two approaches to brain-computer interfaces: operant conditioning and pattern recognition. With operant conditioning, the user is trained to control his or her brain signal using feedback from the system. This approach is often used as explicit input to the system. It is most useful when the user is invested in the system, as is the case with disabled users. In the pattern recognition approach, the user does not go through extensive training. Instead, the system uses signal processing and machine learning techniques to learn patterns associated with various cognitive states. This method is most likely used as implicit input to the system, and may be more practical for most HCI settings.

Following the pattern recognition approach, the work in this dissertation makes use of brain activity as an additional input channel, providing hard-to-detect information such as aspects of the cognitive state of the user. In order to do this, a training or calibration session is required for the computer to begin learning about the user's brain patterns. This will be discussed more in Section 5.2 and Section 6.4.

2.1.4. Brain Sensing for HCI

Much prior research on brain-computer interfaces has a primary goal of helping people with severe motor disabilities to interact with their environment by translating their brain activity into specific device control signals. For example, users who are paralyzed or who lack muscle control can currently use BCIs to answer simple questions, control their environment, and conduct word processing (M. M. Moore, 2003).

In HCI contexts, cognitive state information could be valuable to interface designers, both for evaluation of user interfaces as well as for input to interactive systems (Cutrell & Tan, 2008). In evaluation of user interfaces, researchers may use the cognitive state information as an objective, real-time measure to assess and compare user interfaces. When designing interactive systems, the additional information could lead to interfaces that respond carefully to changes in the user's cognitive state.

Until recently, most brain-computer interfaces were designed for disabled users, and employed brain signals as the primary input (Blankertz et al., 2007; Kennedy, et al., 2000; Pfurtscheller, Flotzinger, & Kalcher, 1993; Schalk, et al., 2004; Wolpaw, et al., 1991). While these systems provide this group of users with a valuable communication channel, they likely will not see wider adoption due to the low bandwidth compared to other available methods for non-disabled users.

With lower costs for non-invasive brain sensing, we recently have seen a growing interest in employing brain sensors for a wider audience (for an overview, see (Jackson & Mappus, 2010)). Much of this work has also used brain sensing as explicit input to the system to make selections or control the interface, (e.g. in a game context (Kuikkaniemi et al., 2010; O'Hara, Sellen, & Harper, 2011) or with a multitouch table (Yuksel, Donnerer, Tompkin, & Steed, 2010)), although there have been examples of passive brain sensing to be used either as implicit input or for evaluation of user interfaces (Grimes, et al., 2008; Leanne M. Hirshfield et al., 2011; Leanne M. Hirshfield, et al., 2009; Lee & Tan, 2006). Recently, it has been suggested that untrained users may benefit from systems that use pattern recognition and machine learning to classify signals users naturally give off when using a computer system (Cutrell & Tan, 2008). The system would use brain sensors to automatically discover aspects of the user's cognitive state and use this information as passive or implicit input to a system, augmenting any explicit input from other devices, and increasing the bandwidth from humans to computers.

In Girouard, et al., we brought offline analysis of fNIRS signals into a realtime system with the goal of using it to build passive brain-computer interfaces (Girouard, Solovey, & Jacob, 2010). The work described here goes beyond this work by improving the processing, training and classification algorithms, and building and evaluating a viable new input technique to improve the user performance and experience.

The motivation for using fNIRS and other brain sensors in HCI research is to pick up cognitive state information that is difficult to detect otherwise. It should be noted that some changes in cognitive state may also have physical manifestations. For example, when someone is under stress, his or her breathing patterns may change. It may also be possible to make inferences based on the contents of the computer screen, or on the input to the computer. However, since these can be detected with other methods, we are less interested in picking them up using brain sensors. Instead, we are interested in using brain sensors to detect information that does not have obvious physical manifestations, and that can only be sensed using tools such as fNIRS.

2.1.5. Brain Sensing for Human Robot Interaction

Brain-computer interfaces have previously been incorporated into robot architectures, although these have typically been EEG-based systems. For example, they have been used for controlling mobile robots (Barbosa, Achanccaray, & Meggiolaro, 2010; Escolano, Murguialday, Matuz, Birbaumer, & Minguez, 2010) or an intelligent wheelchair (Perrin, Chavarriaga, Colas, Siegwart, & Millán). fNIRS provides advantages over the more prevalent EEG due to its easy setup and robustness to noise. Past research has proposed the use of fNIRS-based BCIs (Okumura & Zhiwei, 2007; Tsubone, Tsutsui, Muroga, & Wada, 2008). However, these projects focus on using brain data for direct one-way control of robot movement, and are less concerned with the interaction between the human and robot. In addition, the reliability of such active control schemes will vary greatly depending on context, and may be particularly difficult to apply in high-stress or high-load contexts. By using fNIRS to passively identify an operator's cognitive state, we can exploit that very phenomenon to improve interaction efficiency.

Augmenting active communication channels is particularly important in human-robot interaction, because people have a tendency to ascribe human-like abilities of comprehension to autonomous robots (possibly due to their apparent agency, or to their depiction in popular culture) that are, at this point, unrealistic. Hence, any additional information that can help the robot to understand the operator's intentions will be of great value.

When designing autonomy modes for the robot, it is important to understand the user’s cognitive state. Parasuraman et al. (Parasuraman, Sheridan, & Wickens, 2000) propose a framework for supporting user cognition with automation which describes four stages of human information processing (sensory processing, perception, decision making and response selection), each of which can have a different automation level. They also outline criteria for evaluating the user interface by examining both human performance measures and also system performance criteria (automation reliability, costs of action outcomes). This framework provides guidelines but does not prescribe specific adaptive behavior for every system. Instead, each system must be carefully evaluated and iteratively designed to meet the needs of the users.

2.2. Multitasking Background

In Chapter 4 and Chapter 6, I try to identify specific cognitive multitasking states with fNIRS to better support the user. Here, I give background on multitasking and interruptions that lays the foundation for those sections.

Although computers are capable of handling multiple processes simultaneously, people have a difficult time due to high mental workload from increased working memory demands and the overhead of switching context between multiple tasks. Repeated task switching during an activity may lead to completion of the primary task with lower accuracy and longer duration, in addition to increased anxiety and perceived difficulty of the task (Bailey, Konstan, & Carlis, 2001). The challenge is to devise an effective way to measure workload and attention-shifting in a dynamic environment, as well as to identify optimal support for multitasking.

2.2.1. Measuring Mental Workload and Other Cognitive States

Managing mental workload has long been an active topic in HCI research and high mental workload has been identified as a cause of potential errors (Card, Moran, & Newell, 1983). Researchers have shown that different types of subtasks lead to different mental workload levels (Iqbal, Adamczyk, Zheng, & Bailey, 2005). As a measure for mental workload, researchers have proposed pupil dilation (Iqbal, Zheng, & Bailey, 2004) in combination with subjective ratings as this is non-invasive, and allows the user to perform the tasks as the data is processed in real time. Other physiological measures, including skin conductance, respiration, facial muscle tension and blood volume pressure, have also been used to detect cognitive or emotional states to improve machine intelligence (Fairclough, 2009; R. L. Mandryk & Inkpen, 2004; Picard, Vyzas, & Healey, 2001). While adaptive user interfaces may be designed to reduce mental workload, any automation may also result in reduced situation awareness, increased user complacency and skill degradation, and these human performance areas should be evaluated in the system (Parasuraman, et al., 2000).

2.2.2. Task Switching and Measuring Interruptibility

When managing multiple tasks, interruptions are unavoidable. To address this, researchers have developed systems that try to identify the cost associated with interruption based on different inputs, such as desktop activity, environment context (Fogarty, et al., 2004; Hudson, et al., 2003; Starner, et al., 1998), eye tracking (Hornof, Zhang, & Halverson, 2010), or other physiological measures such as heart rate variability and electromyogram (Chen, Hart, & Vertegaal, 2008) and handle interruptions accordingly. They have found interruptions to be less disruptive during lower mental workload (Iqbal & Bailey, 2005; Salvucci & Bogunovich, 2010). Other studies tried placing interruptions near the beginning, middle or end of a task (Czerwinski, Cutrell, & Horvitz, 2000), at task boundaries (Miyata & Norman, 1986a), or between repetitive tasks which were considered as more interruptible (Monk, Boehm-Davis, & Trafton, 2002). It was also shown that interruptions relevant to the main task tend to be less disruptive for the users than irrelevant interruptions (Czerwinski, et al., 2000).

Various interruption schemes may affect performance in different ways; however, there is no universally optimal interruption scheme. Interrupting the user as soon as the need arises, for example, emphasizes task completeness over accuracy, while allowing the user to defer interruptions indefinitely does the opposite (Sasse, Johnson, & Mcfarlane, 1999).

McFarlane (McFarlane, 2002) discusses four distinct methods for coordinating interruption—immediate, negotiated (user selects when to be interrupted), mediated (an intelligent agent selects when to interrupt), and scheduled (interruptions appear at fixed times)—and found that no optimal method existed across users and tasks. Thus, it is crucial that the style of interruption adapts to the task. Systems have been developed that quantify the optimal time to interrupt a user by weighing the value against the cost of interruption (Iqbal, et al., 2005). In addition to determining the optimal time for switching tasks, researchers have tried to determine the best method for reminding users of pending background tasks. Miyata and Norman (Miyata & Norman, 1986b) note that important alerts specifically designed for someone who is deeply engaged in another task would most likely be inappropriate and may even be disruptive in other situations.

2.2.3. Multitasking Scenarios: Branching, Dual Task, Delay

Multitasking behavior involves several high-level brain processes, which vary depending on the types of tasks and the interaction between the tasks. Koechlin et al. (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999) described three distinct, but related multitasking scenarios, which they refer to as *branching*, *dual-task*, and *delay*. These are the foundation for the studies described in Chapter 4.

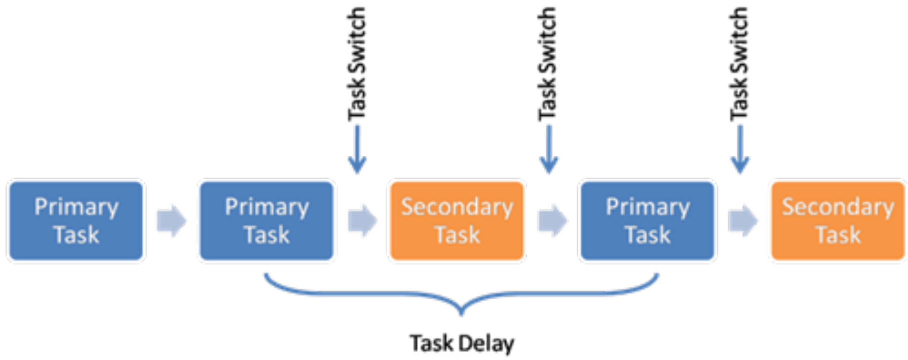


Figure 2-4. Branching: Primary and secondary task both require attentional resources to be allocated, and the primary task goal must be kept in mind over time.

Branching (Figure 2-5, Figure 2-4) is illustrated by the following scenario: *A user is tackling a complex programming task but is interrupted by an incoming email from her boss that is time sensitive.* Thus, the user must "hold in mind goals while exploring and processing secondary goals" (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999). Branching processes are triggered frequently in multitasking environments and pose a challenge to users.

However, some situations may involve frequent task switching without the need to maintain information about the previous task (e.g. *A user is monitoring and responding to high priority software support issues that are logged by clients as well as responding to important emails, and regularly switches between the two tasks*). These tasks are referred to as *dual-task* because there are two tasks that require attentional resources (Figure 2-5). These situations could also utilize adaptive support in the user interface, but the adaptive behavior would be distinct from that of *branching*.

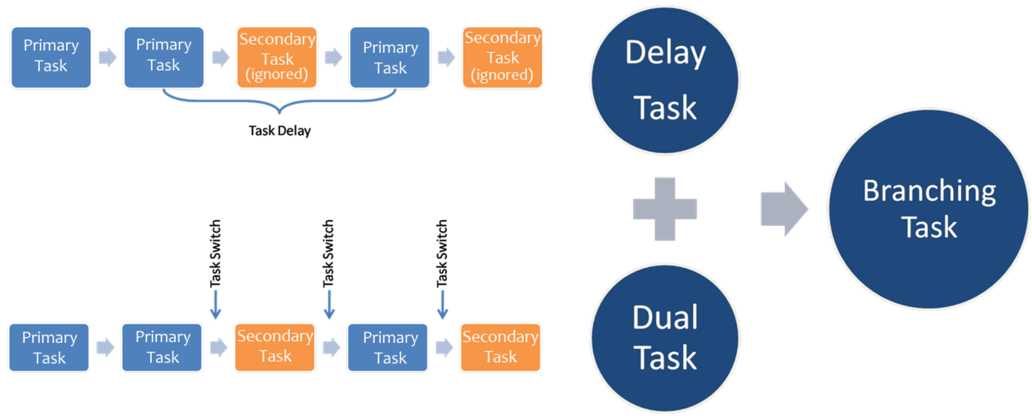


Figure 2-5. In the Delay scenario, the secondary task requires little attention, but the primary task goal is held in working memory. In the Dual-Task scenario, both primary and secondary tasks require attentional resources to be allocated for each task switch, but goals are not held in working memory. Branching has characteristics of both Delay and Dual-Task scenarios (Figure 2-4).

The third multitasking paradigm is illustrated with the following scenario: *A user is tackling a complex programming assignment and at the same time gets instant messages which the user notices, but ignores*. Here, the secondary task is ignored and therefore requires little attentional resources. They refer to this as delay because the secondary task mainly delays response to the primary task (Figure 2-5).

In their experiment, Koechlin et al. demonstrated using functional Magnetic Resonance Imaging (fMRI) that these three multitasking processes have different activation profiles in the prefrontal cortex of the brain, particularly in Brodmann's Areas 8, 9 and 10. Their task involved processing rules based on letters appearing on the screen. Each stimulus was either an uppercase or lowercase letter from the word "tablet." The expected response from the user was different depending on the case of the letter, so switching between uppercase and lowercase letters would be similar to balancing two tasks. There were four conditions in their experiment, each with different rules for responding, designed to trigger specific multitasking behavior (Figure 2-6):

1) *Delay*: Are two consecutive *uppercase* stimuli in immediate succession in the word "TABLET'? Ignore *lowercase*.

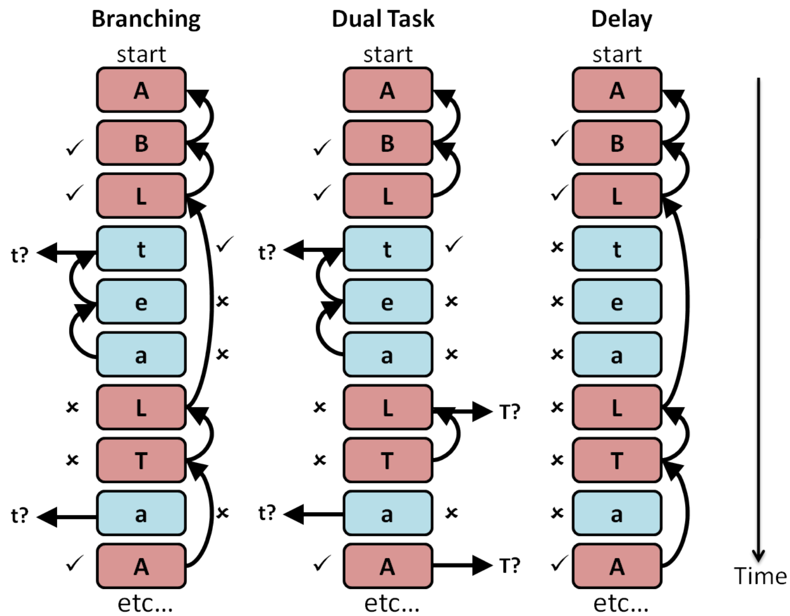


Figure 2-6. Branching, Dual Task and Delay tasks and responses from Koechlin, et al. (1999).

2) *Dual-Task:* Are two consecutive stimuli *of the same case* in immediate succession in the word tablet? When the case changes, is the first letter in the series a 'T' or 't'? 3) *Branching*: For *uppercase* stimuli, respond as in *Delay*. If the letter is *lowercase*, respond as in *Dual Task*.

4) *Control*: Are two consecutive stimuli in immediate succession in "TABLET'? All stimuli were uppercase.

Koechlin et al. (E. Koechlin, Corrado, G., Pietrini, P., & Grafman, J. , 2000) later showed that even during branching, there were distinct activation profiles that varied depending on whether the participant could predict when task switching would occur or whether it was random. The experimental setup was almost identical to the earlier study, except that in all conditions, the *branching* paradigm was used. There were two experimental *branching* conditions (Figure 2-7) and a control:

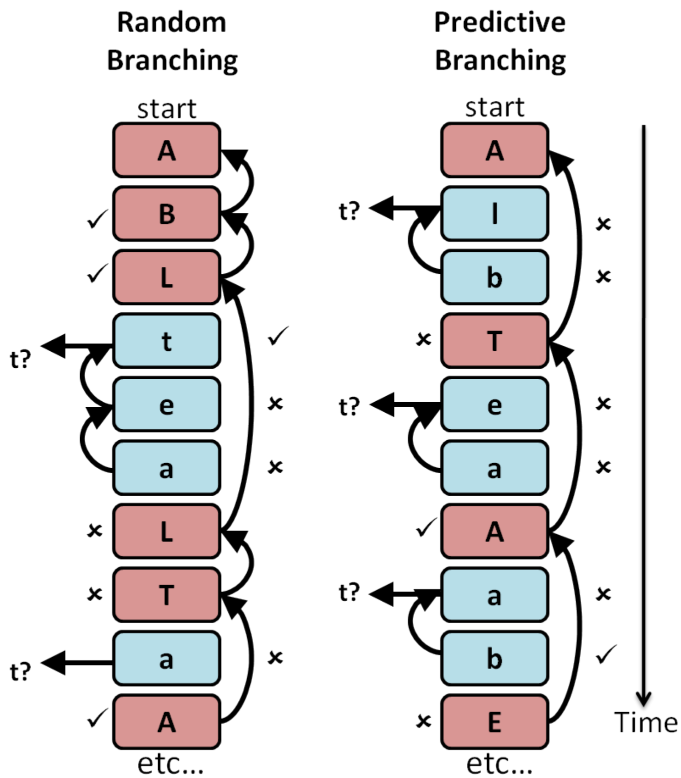


Figure 2-7. Experimental conditions from Koechlin et al. (2000)

1) *Random Branching:* Upper- and lower-case letters were presented pseudorandomly. 2) *Predictive Branching:* Uppercase letters were presented every 3 stimuli. 3) *Control Branching:* The same six-letter sequence (A e t a B t) was shown repeatedly.

The significance of these two experiments lies in the fact that all experimental conditions had the same stimuli and the same possible user responses, so the conditions could not be easily distinguished from one another by simply observing the participant. Using fMRI, however, it became possible to distinguish the conditions based on the distinct mental processes (and thus, distinct blood flow patterns) elicited by each task.

In addition, the cognitive states identified in these experiments have direct relevance to many HCI scenarios, particularly when a user is multitasking. Automatically recognizing that the user is experiencing one of these states provides an opportunity to build adaptive systems that support multitasking. For example, by recognizing that most interruptions are quickly ignored, as in the *delay* condition, the system could limit these types of interruptions or reduce their salience as appropriate. Further, if a user is currently experiencing a *branching* situation, the interface could better support maintaining the context of the primary task, whereas during *dual-task* scenarios this would be unnecessary. Finally, distinguishing between *predictive* and *random* scenarios could trigger the system to increase support when the user's tasks become unpredictable.

Using fMRI for brain imaging, Koechlin et al. demonstrated that these three multitasking activities had different signatures in the anterior prefrontal cortex (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999), the area that is best for measuring with fNIRS. This dissertation builds on these results. In Chapter 4, we show that these states could, in fact, be distinguished using fNIRS as well. Then, in Chapter 5 and Chapter 6, we use the known multitasking activities described by Koechlin, et al. as stimuli for creating individual sets of fNIRS training data during multitasking for each user. This training data is used to build a classification model for each individual that is used to later distinguish between multitasking states the user is experience during other tasks and activities.

Chapter 3 Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines[1]

3.1. Introduction

In this dissertation, I explore functional near-infrared spectroscopy (fNIRS) as a potential input to interactive systems because it is safe, non-invasive and relatively portable, but still provides brain activity data. However, because fNIRS was originally developed for use in clinical settings, the typical procedures used with fNIRS called for restrictions that are not actually practical in HCI research settings. To be valuable in HCI, brain sensors

[1] The work in this chapter was originally described in Solovey, et al. "Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines" in the proceedings of the ACM UIST '09 Symposium on User Interface Software and Technology, (2009) p. 157-166. This was joint work with Audrey Girouard.

such as fNIRS should collect useful information while ideally allowing normal interaction with the computer.

In particular, when using fNIRS to pick up brain activity data for HCI, we would not expect the participant to be physically constrained while using the computer. However, in most studies using brain sensors, researchers expend great effort to reduce the noise picked up by the sensors. Typically, participants are asked to remain still, avoid head and facial movement, and use restricted movement when interacting with the computer. In addition, many factors cannot be controlled, so researchers sometimes throw out data that may have been contaminated by environmental or behavioral noise, or they develop complex algorithms for removing the noise from the data. By doing this, the researchers hope to achieve higher quality brain sensor data, and therefore better estimates of cognitive state information.

However, it is not clear that all of these factors contribute to problems in fNIRS data or that these restrictions improve the signal quality. Ideally, for HCI research, the fNIRS signals would be robust enough to be relatively unaffected by other non-mental activity occurring during the participant's task performance. In fact, one of the main benefits of fNIRS is that the equipment imposes very few physical or behavioral restrictions on the participant (Hoshi, 2009).

From our experience conducting a feasibility study with fNIRS (L. M. Hirshfield et al., 2007), we identified several considerations and provide guidelines in this chapter for using fNIRS in realistic HCI laboratory settings. We empirically examined whether typical human behavior (e.g. head and facial movement) or computer interaction (e.g. keyboard and mouse usage) interfere with brain measurement using fNIRS. Based on the results of our studies, we establish which physical behaviors inherent in computer usage interfere with accurate fNIRS sensing of cognitive state information, which can be corrected in data analysis, and which are acceptable. With these findings, we facilitate further adoption of fNIRS brain sensing technology in HCI research and inform the experiments described in the rest of this dissertation.

3.2. fNIRS Considerations

We identify below potential sources of noise and artifacts in the fNIRS signal when used in typical HCI laboratory settings.

3.2.1. fNIRS Considerations: Head Movement

Several fNIRS researchers have brought attention to motion artifacts in fNIRS sensor data, particularly those from head movement (Devaraj, Izzetoglu, Izzetoglu, & Onaral, 2004; Matthews, et al., 2008). Matthews et al. (Matthews, et al., 2008) explains that "motion can cause an increase in blood flow through the scalp, or, more rarely, an increase in blood pressure in the interrogated cerebral regions." In addition, they point out that "orientation of the head can affect the signal due to gravity's effect on the blood." They note that these issues are significant if the head is not restricted, and even more so in an entirely mobile situation. However, other researchers indicate that fNIRS systems can "monitor brain activity of freely moving subjects outside of laboratories" and note that "measurements with less motion restriction in the daily-life environment open new dimensions in neuroimaging studies" (Hoshi, 2009). While fNIRS data may be affected by head movements, this should be contrasted with fMRI where movement over 3mm will blur the image. Because of the lack of consensus in the community, we chose to investigate the artifacts associated with head movements during typical computer usage to determine their effect on fNIRS sensor data in a typical HCI setting. This is described in Experiment 3 below.

3.2.2. fNIRS Considerations: Facial Movement

fNIRS sensors are often placed on the forehead, and as a result, it is possible that facial movements could interfere with accurate measurements. Coyle, Ward, and Markham point out that “slight movements of the optodes on the scalp can cause large changes in the optical signal, due to variations in optical path. It is therefore important to ensure robust coupling of optodes to the subject’s head” (Coyle, et al., 2004). These forehead movements could be caused by talking, smiling, frowning, or by emotional states such as surprise or anger, and many researchers have participants refrain from moving their face, including talking (Chenier & Sawan, 2007). However, as there is little empirical evidence of this phenomenon, we examined it further in Experiment 4 described below. We selected frowning for testing as it would have the largest effect on fNIRS data collected from the forehead.

Eye movements and blinking are known to produce large artifacts in EEG data which leads to the rejection of trials including such an artifact (Izzetoglu, et al., 2004). However, fNIRS is less sensitive to muscle tension and researchers have reported that no artifact is produced in nearby areas of the brain (Izzetoglu, et al., 2004). It would also be unrealistic to prevent eye blinks and movement in HCI settings. Overall, we conclude eye artifacts and blinks should not be problematic for fNIRS, and we do not constrain participants in this study.

3.2.3. fNIRS Considerations: Ambient Light

Because fNIRS is an optical technique, light in the environment could contribute to noise in the data. Coyle, Ward, and Markham advise that stray light should be prevented from reaching the detector (Coyle, et al., 2004). Chenier and Sawan (2007) note that they use a black hat to cover the sensors, permitting the detector to only receive light from the fNIRS light sources.

While this is a concern for researchers currently using raw fNIRS sensors that are still under development, future fNIRS sensors will be embedded in a helmet or hat that properly isolates them from this source of noise. Therefore, we did not further examine how the introduction of light can affect fNIRS data. Instead we just caution that excess light should be kept to a minimum when using fNIRS, or the sensors should be properly covered to filter out the excess light.

3.2.4. fNIRS Considerations: Ambient Noise

During experiments and regular computer usage, one is subjected to different sounds in the environment. Many studies using brain sensors are conducted in sound-proof rooms to prevent these sounds from affecting the sensor data (Morioka, et al., 2008). However, this is not a realistic setting for most HCI research. Therefore, we conducted this study in a setting similar to a normal office. It was mostly quiet, but the room was not soundproof, and there was occasional noise in the hallway, or from heating and air conditioning systems in the building.

3.2.5. fNIRS Considerations: Respiration and Heartbeat

The fNIRS signals picks up artifacts from respiration and heart beat, by definition, as it measures blood flow and oxygenation (Coyle, et al., 2004; Matthews, et al., 2008). These systemic noise sources can be removed using known filtering techniques. For a discussion of the many filtering techniques, see Matthew et al. (Matthews, et al., 2008) and Coyle et al. (Coyle, et al., 2004).

3.2.6. fNIRS Considerations: Muscle movement

In clinical settings, it is reasonable to have participants perform purely cognitive tasks while collecting brain sensor data. This allows researchers to learn about brain function, without any interference from other factors such as muscle movement. However, to move this technology into HCI settings, this constraint would have to be relaxed, or methods for correcting the artifacts must be developed. Fink et al. discussed the difficulty of introducing tasks that have a physical component in most brain imaging devices, explaining that they may "cause artifact (e.g. muscle artifacts in EEG or activation artifacts due to task-related motor activity in fMRI) and consequently reduce the number of reliable (artifact-free) time segments that can be analyzed" (Fink, Benedek, Grabner, Staudt, & Neubauer, 2007). In addition, they note that the test environment of fMRI scanners also makes it difficult for any physical movement.

One of the main benefits of fNIRS is that the setup does not physically constrain participants, allowing them to use external devices such as a keyboard or mouse. In addition, motion artifacts are expected to have less of an effect on the resulting brain sensor data (Girouard et al., 2009). In Experiments 1 and 2 described below, we examine physical motions that are common in HCI settings, typing and mouse clicking, to determine whether they are problematic when using fNIRS.

3.2.7. fNIRS Considerations: Slow Hemodynamic Response

The slow hemodynamic changes measured by fNIRS occur in a time span of 6-8 seconds (S. Bunce, Devaraj, Izzetoglu, Onaral, & Pourrezaei, 2005). This is important when designing interfaces based on fNIRS sensor data, as the interface would have to respond in this time scale. While the possibility of using event-related fNIRS has been explored (Herrmann, et al., 2008), most studies take advantage of the slow response to measure short term cognitive state, instead of instantaneous ones.

	At Rest	Performing Cognitive Task
No artifact present	A. No Artifact + No Cognitive task	B. No Artifact + Cognitive Task
Artifact present	C. Artifact + No Cognitive task	D. Artifact + Cognitive Task

2: Is there a difference between rest and cognitive task?

2.1: When **no artifact** is present, is there a difference between rest and cognitive task?

2.2: When **artifact** is present, is there a differer between rest and cognitive task?

1: Is there a difference between the presence or absence of the artifact?

1.1: When the participant is **at rest,** is there a difference between the presence or absence of the artifact?

1.2: When the participant **performs the cognitive task**, is there a difference between the presence or absence of the artifact?

Figure 3-1: Letters A, B, C, and D show the conditions tested. The numbered questions indicate the comparisons between the conditions done in the analysis.

3.3. Experimental Protocol

Understanding how the potential noise sources described above affect fNIRS data during cognitive tasks is critical for proper use of fNIRS in HCI research. Thus, we devised a study to empirically test whether or not several common behavioral factors interfere with fNIRS measurements. Specifically, we selected typical human behaviors (head and facial movement) and computer interaction (keyboard and mouse usage), to determine whether each of them needs to be controlled, corrected, or avoided at all cost. This will help us determine whether standard interfaces can be used along with fNIRS in real brain-computer interfaces.

We will call each of the examined physical actions artifacts, since they are not the targeted behavior we would like to detect with fNIRS. Using fNIRS, we measured brain activity as these artifacts were introduced while the participant was otherwise at rest, as well as while the participant was performing a cognitive task. We then compared these results to signals generated while the participant was completely at rest with no artifact, as well as to when the participant performed the cognitive task without the artifact. This allowed us to determine whether the artifact had an influence on the signal generated in a rested state, as well as if it has an impact on the signal during activation.

For each artifact, there were four conditions tested: (A) a baseline with no cognitive task or artifact; (B) the cognitive task alone with no artifact; (C) the artifact alone with no cognitive task; and (D) the cognitive task along with an artifact (Figure 3-1).

Our goal in designing the protocol for each artifact was to reproduce realistic occurrences. As these artifacts do not necessarily happen often, we tried to balance conservatism (i.e. highly exaggerated artifact) with optimism (i.e. minute occurrence of artifact), and chose a reasonable exaggeration of the artifact, maximizing the possibility of measuring the artifact if it can be measured, yet keeping the conditions somewhat realistic.

3.3.1 Participants

Ten participants took part in this experiment (mean age = 20.6, std = 2.59, 6 females). All were right-handed, with normal or corrected vision and no history of major head injury. They signed an informed consent approved by the Institutional Review Board of the university, and were compensated for their participation. The experiment is within subject (each participant did all the experiments and conditions), and was counterbalanced to eliminate bias due the order of the experiments, and the conditions.

3.3.2. Apparatus

We used the device described in Section 2.1.2.1. We use the term channel to define a source-detector distance. In previous studies using a similar, linearly arranged probe, researchers have chosen to use data from the furthest two channels only, in order to guarantee that the depth of the measurement reached the cortex (Girouard, et al., 2009; Leanne M. Hirshfield, et al., 2009). While it is likely that the shallower channels pick up systemic responses, or other noise sources, we decided to keep the data from all four source-detector distances measured as they might help separate out artifacts from task activation.

In all the experiments, the participants were at a desk with only a small lamp (60 W) beside the desk turned on, and they were sitting at a distance of roughly 30" from a 19" flat monitor. The room was quiet, but was not soundproof and noise from the hallway outside the laboratory could be heard occasionally. The participants were instructed to keep their eyes fixated on one point on the screen, and to refrain from speaking, frowning or moving their limbs, unless instructed otherwise.

3.3.3. Procedure and Design

There were five different experiments conducted with each participant, all in one session. These corresponded with the four artifacts being studied (keyboard input, mouse input, head movement, and facial movement), plus the tasks without any artifact present. In between each experiment, the participant could take a break. Although the descriptions below are numbered as Experiments 0, 1, 2, 3, 4, the ordering of the experiments was counterbalanced between subjects. The main difference between the experiments was which additional physical artifact, if any, was introduced as the participant performed the two tasks.

3.3.4. Cognitive Task

All five experiments used the same cognitive task. At the beginning of each trial, the participants were shown a 7-digit number on the screen for four seconds. The number then disappeared from the screen, but the participants were instructed to remember it in their head. After 15 seconds, the participants were asked to enter as much of the number as they could remember.

The goal of the cognitive task used in these experiments was to provide a common task that participants would perform in all experiments, which yields a brain signal that could be detected with fNIRS. We choose a simple verbal working memory task because previous fNIRS studies have reported this type of task to produce a clear and consistent brain signal across participants (Ehlis, et al., 2008; Leanne M. Hirshfield, et al., 2009). Many studies have successfully shown discrimination of two (or more) states, and we believe our results will generalize to those as well.

3.4. Experiment 0: No artifacts

This experiment consisted primarily of the cognitive task and rest periods. No additional artifact was introduced. This experiment was used to verify that we could distinguish the fNIRS data while the participant was at rest from the fNIRS data while the participant performed the cognitive task, when no artifact was present.

First, the researcher read instructions to the participants, explaining the two tasks that they would perform in the experiment. Then the participants were presented with a practice trial which included an example of each task in that experiment, so the participants would know what to expect. The participants then relaxed for one minute, so their brains could be measured at a normal, rested state. During this period, as well as all other rest periods, there was a black screen and participants were instructed to focus their eyes on the focal point and relax, clearing their heads of any thoughts. This was followed by ten trials. A trial contained one 15-second condition with the cognitive task, followed by a 15-second rest period to allow the participant's brain to return to a rested state. In

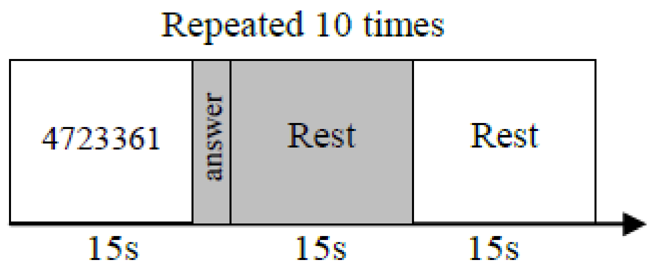


Figure 3-2. Experiment 0 (No artifacts). The white areas represent the two conditions analyzed. The answer period's length was variable.

addition, there was a 15-second condition without the cognitive task in which the participant was essentially at rest (Figure 3-2). These conditions were counterbalanced so that sometimes participants started with the cognitive task, and sometimes they started without the cognitive task.

3.4.1. Preprocessing

The preprocessing step transforms the raw data from the device into hemoglobin values, and smoothes the data to remove any high-frequency noise, as well as heart beat. We chose to filter the data in these experiments because this is a standard step in fNIRS experiments, and the goal was to determine the influence of interaction techniques and artifacts on a typical fNIRS experiment. We applied a simple preprocessing procedure, described in Girouard et al. (Girouard, et al., 2009). We used a non-recursive time-domain band-pass filter, keeping frequencies between 0.01-0.5 Hz (Folley & Park, 2005). The data was then transformed to obtain oxy- ([HbO]) and deoxy-hemoglobin ([Hb]) concentration values, using the modified Beer-Lambert law (Villringer & Chance, 1997). It should be noted that the combination of [HbO] and [Hb] gives a measure of total hemoglobin, which we will refer to as [HbT]. We averaged each trial in two seconds periods, to obtain seven averaged points we call *Time Period*.

3.4.2. Analysis

In this experiment, we wanted to observe whether the cognitive task, on its own, yielded a brain signal that was distinguishable from the signal during a rested state. This result is fundamental to all the other experiments that include the cognitive task. If we were not able to significantly distinguish the cognitive task from rest with no added artifacts, it would have been difficult to distinguish the two when additional noise was introduced into the data.

This dataset and all reported in this chapter were tested for conformity with the ANOVA assumption of normality by creating a normal probability plot, on which normal data produces a straight or nearly straight line, confirming that the ANOVA is an appropriate test of significance.

We did a factorial repeated measures ANOVA on *Cognitive Task* (cognitive task or rest) x *Hemisphere* (left or right) x *Channel* (4) x *Time Period* (7). This would identify differences within each participant, and determine if they are significant across participants. This is Comparison 2.1 in Figure 3-1. We ran this analysis with [HbO], [Hb] and [HbT] data separately. While we did a factorial ANOVA, we are most interested in results that show significant interactions including the *Cognitive Task* factor, since these show significant differences between the signal during the cognitive task and the signal during rest. In this analysis, and all those following, we will only report significant results ($p<0.05$) that are pertinent to current HCI questions.

3.4.3. Results

From these three analyses, the only relevant significant factor found was with [Hb], *Cognitive Task* x *Channel* ($F(3, 27) = 5.670$, $p = 0.031$). This confirms that levels of [Hb] differ between trials where participants performed a cognitive task, and trials where they simply rested, and that this difference in [Hb] levels varied by channel. This positive result allowed us to move forward with the rest of the analysis.

3.5. Experiment 1: Keyboard Input

The keyboard and mouse are the most common input devices for modern computers. We tested keyboard input in Experiment 1 and mouse input in Experiment 2. We hypothesized that keyboard inputs would not be a problem with fNIRS, since most brain activation for motor movement occurs in the motor cortex, an area not probed with our fNIRS sensors. In addition, we did not believe that the physical act of typing would cause the sensors to move out of place or change the blood oxygenation characteristics in the prefrontal cortex.

We decided not to have the participants type specific words because we were only interested in measuring the influence of the typing motions on the signal, instead of any

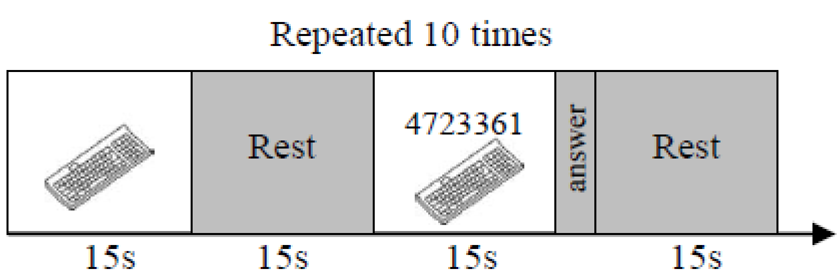


Figure 3-3. Experiment 1 (Keyboard Input). The white areas represent the two conditions analyzed in the experiment.

brain activity associated with composing and typing text. They were instructed to randomly type on the keyboard, using both hands, at a pace resembling their regular typing pace, including space bars occasionally to simulate words. The protocol was analogous to Experiment 0. The main difference is that in both tasks, the participant was also typing randomly as described above (Figure 3-3).

3.5.1. Analysis

To observe the influence of typing on the brain data, we examined the data in several different ways, corresponding with the numbers in Figure 3-1. Comparison 1 determines whether there is a difference between typing and not typing, regardless of whether there was cognitive task. Comparison 1.1 examines whether there is a difference in the fNIRS data between the presence and absence of the typing artifacts when the participant is at rest. Comparison 1.2 determines whether there is a difference between the presence and absence of the typing artifacts when the participant performs the cognitive task. Comparison 2 determines whether there is a difference between doing a cognitive task and no cognitive task, regardless of whether the participant was typing. Comparison 2.2 looks at whether there is a difference between rest and cognitive task when typing artifacts are present. Note that 2.1 was not examined in Experiments 1 to 4, as there are no artifacts present in this condition.

As in Experiment 0, we were most interested in results that showed significant interactions including the *Cognitive Task* factor, since these show significant differences between the signal during the cognitive task and the signal during rest. In addition, we were interested in significant interactions that included the artifact *Typing*, since these show significant differences between when the subject was typing and when the subject was not typing.

Comparison 1, 1.1 and 1.2 used the interaction *Typing* (present or not) x *Hemisphere* (left or right) x *Channel* (4) x *Time Period* (7); Comparison 1.1 uses data from rest tasks; Comparison 1.2 uses data during cognitive tasks; while Comparison 1 uses both datasets. Comparisons 2 and 2.2 used the interaction *Cognitive Task* (cognitive task or rest) x *Hemisphere* (left or right) x *Channel* (4) x *Time Period* (7). Comparison 2.2 used data containing typing while Comparison 2 used data both with and without typing.

Ideally, we would observe the absence of *Typing* as a factor in significant interactions for Comparisons 1, 1.1, and 1.2. For Comparisons 2 and 2.2, ideally we would find *Cognitive Task* as a factor in significant interactions, as this indicates the ability to distinguish the presence or absence of a cognitive task.

For each comparison, we analyze the data for [Hb], [HbO] and [HbT] separately, as was done for Comparison 1 in Experiment 0.

3.5.2. Results

Comparison 1 showed significance for *Typing* x *Time Period* with [HbO] ($F(6, 54)= 3.762$, $p= 0.034$), meaning that with cognitive task and rest tasks combined, we can distinguish typing using the time period. We did not observe any significant interaction that included *Typing* in Comparison 1.1. We can conclude that at rest, there is no significant difference in the fNIRS signal between typing and not typing. We found that for Comparison 1.2, [Hb] data revealed significance with *Typing* x *Hemisphere* x *Channel* ($F(3, 27)= 3.650$, $p= 0.042$). We find *Typing* x *Hemoglobin Type* x *Time Course* to be significant ($F(6, 54)= 6.190$, $p= 0.012$). These results show that when the participant is performing a cognitive task, there is a difference whether the participant is also typing or not, as typing shows up in significant interactions.

In Comparison 2, we found *Cognitive Task* x *Hemisphere* to be significant with [Hb] data ($F(1, 9)= 5.358$, $p= 0.046$. This indicates that when typing and not typing tasks are combined, we can determine whether the participant is performing a cognitive task or not using the right hemisphere. In Comparison 2.2, [Hb] yielded significance with *Cognitive Task* x *Hemisphere* ($F(1, 9)= 5.319$, $p= 0.047$). Comparison 2.2 demonstrates that given typing, we can distinguish whether the participant is also performing a cognitive task or not, specifically using [Hb] and hemisphere.

3.5.3. Discussion

Comparison 1.1 confirmed that the sensors are not picking up a difference between the typing task and rest. However, in Comparison 1.2, we found that typing is influenced by the cognitive task. This is also true in general, as typing tasks are usually related to the current task.

Overall, while typing can be picked up when there is a cognitive task present, we can still distinguish the cognitive task itself (Comparison 2.2 and 2). This confirmed our hypothesis and validated that typing is an acceptable interaction when using fNIRS. From this, we can also assume that simple key presses (e.g. using arrow keys) would also be acceptable with fNIRS since it is just a more limited movement than typing with both hands.

3.6. Experiment 2: Mouse Input

We designed a task that tests mouse movement and clicking. We hypothesized that small hand movement such as using the mouse would not interfere with fNIRS signal. The participant was instructed to move a cursor until it was in a yellow box on the screen, and click. The box would then disappear and another one would appear somewhere else. Participants were directed to move at a comfortable pace, not particularly fast or slow, and to repeat the action until the end of the condition. All participants used their right hand to control the mouse. The procedure was identical to Experiment 1, except that the typing was replaced with mouse clicking (Figure 3-4). We analyzed the data using the same comparisons as in Experiment 1, substituting mouse input for keyboard input.

3.6.1. Results

Comparison 1 yielded no significant interactions, indicating that we cannot observe differences between the presence and absence of clicking, when combining data from the cognitive task and rest. In Comparison 1.1, with [Hb], we observe an interaction of *Clicking* x *Channel* ($F(3, 27)= 4.811$, $p= 0.044$). This shows that we can tell whether someone is clicking, depending on the channel with the participant being at rest. In Comparison 1.2, [HbO] data reveals significant interaction with *Clicking* x *Hemisphere* ($F(1, 9)= 9.599$, $p= 0.013$) and *Clicking* x *Hemisphere* x *Time Course* ($F(6, 54)= 4.168$, $p= 0.037$). This indicates the ability to distinguish *Clicking* from no motor activity when the participant is performing a cognitive task, although this effect differs across hemispheres. Finally, we observed significant interactions with *Clicking* x *Hemisphere* with [HbT] ($F(1, 9)= 6.260$, $p= 0.034$) and *Clicking* x *Hemisphere* x *Hemoglobin Type* ($F(1, 9)= 5.222$, $p= 0.048$), which leads to the same conclusion as with [HbO] data only.

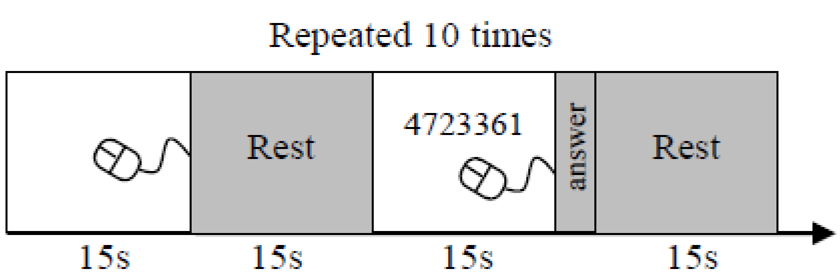


Figure 3-4. Experiment 2 (Mouse Input).

Overall, we can tell whether someone is clicking depending on the Hemisphere.

Comparison 2 yielded no significant interactions, indicating that we cannot distinguish between rest and cognitive task, when the data includes both clicking and not clicking. In Comparison 2.2, we found both *Cognitive Task* x *Hemisphere* x *Hemoglobin Type* ($F(1, 9)= 5.296$, $p= 0.047$) and *Cognitive Task* x *Hemisphere* x *Hemoglobin Type* x *Time Course* ($F(6, 54)= 4.537$, $p= 0.036$) to be significant, indicating that even in data containing clicking, we can tell whether the participant is doing a cognitive task or resting.

3.6.2. Discussion

We found that clicking in this experiment might affect the fNIRS signal we are collecting, as Comparison 1.1 yielded interactions with the factor of *clicking*. This means that when the participant is at rest, there is a difference between the presence and absence of clicking. The difference in activation is not surprising as we did not have a "random clicking" task, but one where subject had to reach targets, which may have activated the anterior prefrontal cortex. However, because Comparison 2.2 still was able to distinguish *Cognitive Task*, the cognitive task of remembering numbers may produce a different signal from clicking.

Hence, results indicate that when we want to observe a cognitive task that contains clicking, we need to have the rest task contain clicking as well, as Comparison 2.2 found significant interactions, but Comparison 2 did not. Overall, we believe that clicking is acceptable if the experiment is controlled, confirming in part our hypothesis.

3.7. Experiment 3: Head Movement

General head movements could affect the fNIRS signal, both because of possible probe movement on the skin, and possible change in blood flow due to the movement itself, as was noted earlier. We hypothesized that head movement could be a problem, as this seems to be reported by many researchers.

Many types of head movements can occur, in all directions. We chose a condition that is representative of common movement while using the computer: we simulated looking down at the keyboard and up at the screen. These movements were done in an intermittent manner, similar to head movements that may occur during normal computer usage, three times per 15s trial.

The procedure was identical to Experiment 1 and 2, except that the typing or mouse clicking was replaced by the head movement (Figure 3-5). We analyzed the data using the same comparisons as in Experiment 1 and 2, substituting head movement for keyboard or mouse input.

3.7.1. Results

We found no significant interactions for Comparison 1, which indicates that it is not possible to distinguish between the presence and absence of head movements when the

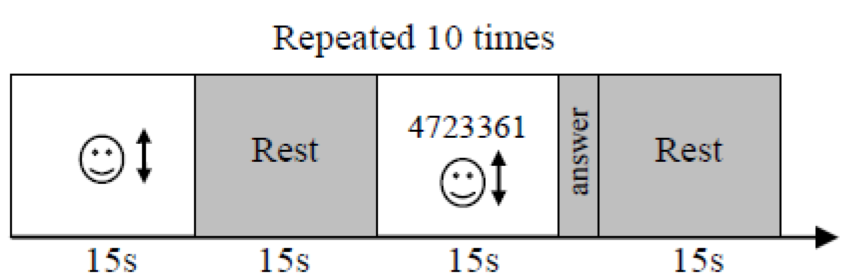


Figure 3-5. Experiment 3 (Head Movement).

cognitive and rest data are combined. There were no significant results for Comparison 1.1, indicating that at rest, there is no significant difference in the signal when the participant is moving his or her head or not. Comparison 1.2 showed that with [Hb] data, we can distinguish *Head Movement* x *Hemisphere* x *Channel* ($F(3, 27)= 5.363$, $p= 0.028$), and we can significantly observe *Head Movement* x *Hemoglobin Type* x *Time Period* ($F(6, 54)= 7.455$, $p= 0.002$), meaning that during the cognitive task, we can tell between the participant moving their head or not.

We found no significant interactions for Comparison 2, meaning that it is not possible to separate the cognitive task from rest when including both data with head movements and data without head movements. In Comparison 2.2, we find that *Cognitive Task* x *Hemoglobin Type* x *Channel* x *Time Period* is significant ($F(18, 162)= 3.915$, $p= 0.048$). With head movements, there is a difference between rest and the cognitive task.

3.7.2. Discussion

Similar to the clicking results, we found that we require the presence of head movements in both the rest and the cognitive task to distinguish it (Comparison 2.2), which leads us to suggest that head movement should be avoided. However, the movements in this experiment were more exaggerated and frequent than regular moving from keyboard to screen: for example, most subjects couldn't see the screen when looking at the keyboard. We suggest that participants minimize major head movements, and instead move their eyes towards the keyboard. We found our initial hypothesis correct, although we believe head movement may be minimized and corrected using filtering techniques.

3.8. Experiment 4: Facial Movement

Forehead facial movement moves the skin located under the probe, which may interfere with the light sent into the brain and its path. We hypothesize that forehead facial movement, e.g. frowning, will have an effect on the data.

In this experiment, participants were prompted to frown for two seconds, every five seconds. Specifically, we asked them to draw the brows together and wrinkle the forehead, as if they were worried, angry, or concentrating.

The procedure was also identical to the other experiments, except that the artifact introduced was facial movement (Figure 3-6). We analyzed the data using the same comparisons as in the other experiments, substituting frowning motion for keyboard or mouse input, or head movement.

3.8.1. Results

Comparison 1 showed significance with [HbO] for *Frowning* x *Channel* ($F(3, 27) = 5.287$, $p = 0.035$). We found significance with *Frowning* x *Channel* with [HbT] ($F(3, 27) = 5.343$, $p = 0.035$), *Frowning* x *Hemoglobin Type* x *Channel* ($F(3, 27) = 4.451$, $p = 0.046$). We see that regardless of whether at rest or doing cognitive task, we can distinguish whether frowning is occurring at some but not all channels, which is consistent with

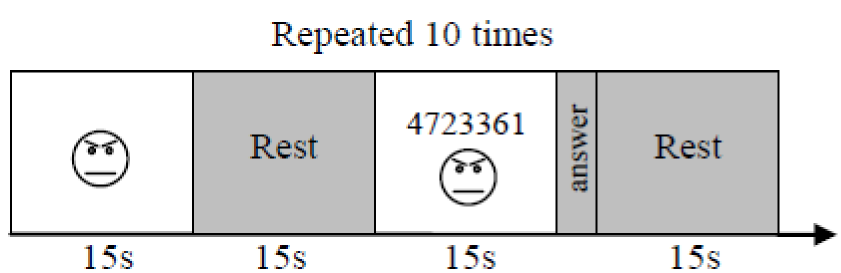


Figure 3-6. Experiment 4 (Facial Movement).

previous results. In Comparison 1.1, we found that [HbO] data showed *Frowning* x *Channel* to be significant ($F(3, 27)= 5.194$, $p= 0.037$), which we also noticed with both types of hemoglobin ($F(3, 27)= 5.191$, $p= 0.037$). When the participant was at rest, we can distinguish whether the participant is frowning or not at some but not all channels. Comparison 1.2 found *Frowning* x *Channel* to be significant for [HbO] data ($F(3, 27)= 4.862$, $p= 0.042$) and with both types of hemoglobin ($F(3, 27)= 4.978$, $p= 0.041$). This indicates that there is a difference in [HbO] levels when participants were frowning or not frowning, and that this difference varied by channel, similarly to Comparison 1.1. Comparison 2 found *Cognitive Task* x *Channel* x *Time Course* to be significant with [HbO] ($F(18, 162)= 3.647$, $p= 0.043$). *Cognitive Task* x *Hemoglobin Type* x *Channel* x *Time Course* was a significant interaction ($F(18, 162)= 4.130$, $p= 0.042$), both indicating that when frowning data is combined with not frowning, we can tell the cognitive task from rest at some but not all channels. Finally, Comparison 2.2 showed no significance for interactions that included *Cognitive Task*, indicating we cannot distinguish the cognitive task from rest when the subject is frowning.

3.8.2. Discussion

We found that frowning data always can be distinguished from non-frowning. We also learned that if all the data includes frowns, then we cannot tell apart the cognitive task from the rest condition. However, we found that if we mix the data that contains frowning and no frowning, we can then discriminate the cognitive task, which shows interesting potential.

Those results indicate clearly that frowning is a problematic artifact, and should be avoided as much as possible. This confirms our hypothesis. However, given that this was an exaggerated movement (3 times in 15s), and that Comparison 2 had good results, we

Table 3-1. Summary of considerations. Legend: ✓ indicates acceptable, C indicates to correct, and ✗ indicates to avoid or control.

Considerations	Result	Referenc e	Correction Methods
Forehead movement	✗	Exp 4	
Major head movement	✗	Exp 3	Use chin rest
Minor head movement	C	Exp 3, (Matthews, Pearlmutter, Ward, Soraghan, & Markham, 2008)	Filter
Respiration and Heartbeat	C	(Matthews, et al., 2008)	Filter
Mouse Clicking	✓	Exp 2	Collect signal during a clicking only task
Typing	✓	Exp 1	
Ambient Light	C	(Chenier & Sawan, 2007; Coyle, Ward, & Markham, 2004)	Wear isolating cap
Hemodynamic Response	✓	(S. C. Bunce, Izzetoglu, Izzetoglu, Onaral, & Pourrezaei, 2006)	Expect 6-8s response
Ambient Noise	C	(Morioka, Yamada, & Komori, 2008)	Minimize external noise
Eye Movement and Blinking	✓	(Izzetoglu, Bunce, Onaral, Pourrezaei, & Chance, 2004)	

can say that if some frowning data found its way into the dataset, it might be possible to still distinguish the cognitive task and the rest task.

3.9. Performance data

In all five experiments, after each cognitive task, participants entered the 7-digit number that they had been remembering. To obtain the error rate of those answers, we compared each digit entered to the original digit, and found the number of digits correctly answered.

A repeated measures ANOVA examining the error rate across artifact types revealed no statistical differences between them ($F(4,36)= 0.637$, $p= 0.526$). This result indicates that each experiment was of similar difficulty.

3.10. Guidelines for fNIRS in HCI

To take advantage of the benefits of fNIRS technology in HCI, researchers should be aware of several considerations, which were identified in this chapter, and summarized in Table 3-1. Our goal was to reveal whether or not several common behavioral factors interfere with fNIRS measurements. We empirically examined whether four physical behaviors inherent in computer usage interfere with accurate fNIRS sensing of cognitive state information. Overall, we found that given specific conditions, we can use typing and clicking in HCI experiments, and that we should avoid or control major head movements and frowns.

Other artifacts, such as minor head movements, heartbeat and respiration may be corrected using filtering. There are many types of filtering algorithms that can help reduce the amount of noise in data (Matthews, et al., 2008). Methods include adaptive finite impulse response (FIR) filtering, Weiner filtering (Devaraj, et al., 2004; M. Izzetoglu, Devaraj, Bunce, & Onaral, 2005), adaptive filtering (Devaraj, et al., 2004) and principal component analysis (Huppert & Boas, 2005; Matthews, et al., 2008; Sitaram, et al., 2007). Matthews et al. (Matthews, et al., 2008) note that FIR can be used in real time if accelerometers are used simultaneously on the head to record head motion. The other methods are mainly offline procedures, making them less practical for real-time systems.

The experimental protocol was designed to reproduce realistic occurrences of artifacts that might be present during typical computer usage in HCI laboratory settings. We purposefully exaggerated the artifacts to make sure they would be measured with fNIRS.

So, we need to keep that in mind as the exaggerated artifacts are less likely to happen than in real experiments. Note that this was run in a typical, quiet office space, and not in a sound proof room like most brain sensing studies.

3.11. Conclusion

In conclusion, we have confirmed that many restrictions such as long setup time, highly restricted position, intolerance to movement, and other limitations, that are inherent to other brain sensing and imaging devices are not factors when using fNIRS. By using the guidelines described above, researchers can have access to the user's cognitive state in realistic HCI laboratory conditions. This is important for adoption in HCI, and these guidelines are followed in the research described in the remaining chapters.

Chapter 4 Sensing Cognitive Multitasking for a Brain-Based Adaptive User Interface[2]

4.1. Introduction

Using the guidelines established in Chapter 3, I began to investigate the cognitive states that we could classify reliably using fNIRS data, focusing on multitasking scenarios. In this chapter, I describe a preliminary study and two experiments using neural data in which we identified four mental processes that may occur during multitasking and have direct relevance to many HCI scenarios. These processes are almost indistinguishable by examining overt behavior (e.g. keystrokes or screen contents) or task performance (e.g. response time, accuracy) alone. However, using fNIRS, we can automatically distinguish

[2] The work in this chapter was originally described in "Sensing Cognitive Multitasking for a Brain-Based Adaptive User Interface" in the proceedings of the ACM CHI'11 Conference on Human Factors in Computing Systems. (E. T. Solovey et al., 2011)

these four states. By detecting specific cognitive states that occur when multitasking, we can build user interfaces that better support task switching, interruption management and multitasking.

This work builds from the experiments described in Koechlin et al. and in Section 2.2.3 with the goal of designing interfaces that recognize these states and behave in appropriate ways to support multitasking. We first conducted a preliminary study to reproduce the results of Koechlin et al. (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999) using fNIRS which is practical for HCI settings unlike fMRI in which slight movement can create motion artifacts and corrupt the image (Erin Treacy Solovey et al., 2009). We then followed with two experiments that look at distinguishing the cognitive multitasking states in other scenarios besides the "tablet" task to investigate whether these are generic cognitive processes, and not simply tied to the particular task used in the earlier study.

4.2. Preliminary Study

The preliminary experiment extends Koechlin et al.'s work (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999) to more realistic HCI settings. The goal was to determine whether we could distinguish between *branching*, *dual-task* and *delay* situations using fNIRS.

As explained in Section 2.2.3, these states were defined as follows:

1) *Branching* occurs when the user must "hold in mind goals while exploring and processing secondary goals" (Koechlin, 1999). Since this is challenging to users, automatically sensing this state would be valuable to HCI and I focus on this state later in this research.

2) *Delay Task* occurs when secondary task is ignored and therefore requires little attentional resources.
3) *Dual Task* entails frequent task switching without the need to maintain information about the previous task (e.g. switching between responding to emails and responding to software support issues being logged)

These states were successfully distinguished using fMRI (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999), but fMRI is not practical in HCI settings. Our hypothesis was that the same could be achieved using fNIRS. Since the sensors are placed on the forehead, they are particularly sensitive to changes in the anterior prefrontal cortex, where Koechlin et al. (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999) showed distinct activation profiles during *delay*, *dual* and *branching* tasks.

Three participants wore fNIRS sensors as they performed the experimental tasks. To trigger the three cognitive states, we used the same experimental paradigm used in (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999).

To determine whether these tasks could be distinguished, we performed leave-one-out cross validation in Weka (Hall et al., 2009) to classify the fNIRS sensor data. In MATLAB, the fNIRS signal was detrended by fitting a polynomial of degree 3 and then a low-pass elliptical filter was used to remove noise in the data. Using support vector machines, we achieved reasonably high accuracy classifying the tasks across the three participants (68.4% mean across three pair-wise classifications, and 52.9% accuracy for three-way classification). This was a small sample of users, and we hope to achieve higher accuracy, but found the results encouraging enough continue in this research direction.

4.3. Multitasking Experiments

From the promising results of the preliminary study, we investigated whether we could detect these three states in other tasks and domains that are more relevant to interactive user interfaces. Our hypothesis was that the cognitive functions elicited in the "tablet" tasks were generic processes that occur during multitasking. Numerous HCI scenarios involve multitasking, and we chose a human-robot team scenario to further explore the detection of cognitive multitasking in user interfaces.

4.3.1. Multitasking in Human Robot Interaction

Human-robot team tasks often involve multitasking, as the user is both performing his or her part of the task, while monitoring the state of the robot(s). Thus, these tasks provide an appropriate example for studying adaptive multitasking support, and may see improved performance with brain-based adaptive interfaces. Thus, the simple word-related task was replaced by a human-robot interaction task that has similar properties.

4.3.2. Experimental Tasks

We conducted two separate experiments which built from the human-robot team task described by Schermerhorn and Scheutz (Schermerhorn & Scheutz, 2009) and adjusted it to include tasks that would induce *delay*, *dual-task* and *branching*, similar to our preliminary study. The tasks involved a human-robot team performing a complex task that could not be accomplished by the human nor the robot alone. The robot and the human had to exchange information in order to accomplish the task. The robot continually updated the human operator with status updates to which the human responded.

In the two separate studies, the participant worked with a robot to investigate rock types on the surface of Mars and had to perform two tasks. The robot presented the participant with status updates, either about a newly found rock or a new location to which it moved. Each rock classification update informed the user of the newly discovered rock's class, which was based on size and ranged from Class 1 to Class 5. Each location update alerted the user of the robot's current location. The spacecraft to which the robot was transmitting could detect the robot's location to the nearest kilometer and assumed the robot was moving in a straight line. Thus, the location updates presented to the user ranged from 0 to 800 meters, in 200 meter increments.

The participant's primary task was to sort rocks, and the secondary task was to monitor the location of the robot. Each time the participant received a status update from the robot (in the form of a pop-up on the screen), s/he had two possible responses: either respond with the left hand by typing "S" to signify *same* or the right hand by typing "N" to signify *new*. After a rock classification, "S" instructed the robot to store the rock in the *same* bin, while "N" instructed the robot to store the rock in a *new* bin. After a location update, "S" instructed the robot to maintain the *same* transmission, while "N" instructed the robot to begin a *new* transmission. The correct response after a particular update varied among the conditions.

4.3.3. Experiment 1: Delay, Dual-Task & Branching

The first experiment contained three conditions, analogous to those in (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999), each with its own rules for the user response (Figure 4-1):

Delay: Do two successive rock classification messages follow in immediate consecutive order? If so, put it in the *same* bin. If not, select a *new* bin. For all location updates, begin a *new* transmission.

Dual-Task: Do two successive messages *of the same type* follow in immediate consecutive order? If so, select the *same* rock bin or maintain the *same* transmission. If the update is of a different type (switch task between rock and location), is the message either a Class 1 rock or a location of 0 meters? If so, select the *same* rock bin or maintain the *same* transmission. In all other cases, place the rock in a *new* bin or begin a *new* transmission.

Branching: For *rock classification* messages, respond as in *Delay*. If the update is a *location*, respond as in *Dual Task.*

4.3.3.1. Participants

This study included 12 healthy volunteers (10 male), between the ages of 18 and 34. Four additional volunteers had participated in the study, but are not included in this analysis because their performance in the tasks was below 70% in more than two trials per condition, indicating that they were not correctly performing the tasks. In addition, data

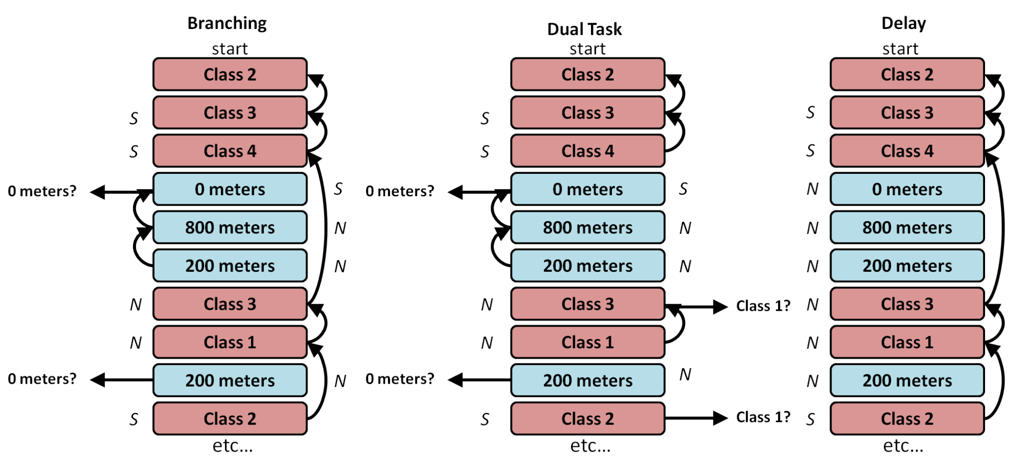


Figure 4-1. Stimuli and responses for conditions in Experiment 1. These conditions are analogous to those in (Koechlin, 1999). (See Figure 2-6).

from another participant is not included due to technical problems with the fNIRS system. All participants were right-handed, had English as their primary language, had no history of brain injury and had normal or corrected-to-normal vision.

4.3.3.2. Design and Procedure

Before the experiment, each participant was given the opportunity to become familiar with each of the three tasks during a practice session without the fNIRS sensors. The conditions were presented in counterbalanced pseudo-random order. Each task was repeated until the participant achieved greater than 80% accuracy in the task. After this accuracy was achieved for all three conditions, the fNIRS sensors were placed on the participant's forehead. The participant was presented with an initial rest screen, which was used to collect a baseline measure of the brain activity at rest. After that, the user had to complete ten 40-second trials for each of the three conditions, which were presented randomly. Between each task, the user was presented with the instructions for the next task, followed by a rest screen.

4.3.3.3. Equipment

We used a multichannel frequency domain OxiplexTS from ISS Inc. for data acquisition, as described in 2.1.2.1.

4.3.3.4. Results

To examine the differences between the three task conditions, we looked at behavioral data collected during the experiment as well as the fNIRS sensor data. In both experiments, any trials where the participant achieved less than 70% accuracy in the task performance were removed in the analysis, since this would indicate that the subject was not actually performing the task correctly.

4.3.3.4.1. Behavioral Results

In the three conditions, the stimuli were essentially the same, as were the possible responses. Thus, it would be difficult for an observer to detect any difference from the screen contents or the subject's behavior alone. Like the sensor data, response time and accuracy measurements can be obtained automatically without interfering with the task so we investigated whether they would vary depending on the condition.

All variables were tested for normal distribution with the Kolmogorov-Smirnov test. For normal distributions, the repeated measurements one-way analysis of variance (ANOVA) with the Tukey post-hoc test for multiple comparisons was used. For non-Gaussian distributions, we used the Friedman (non parametric repeated measurements ANOVA)

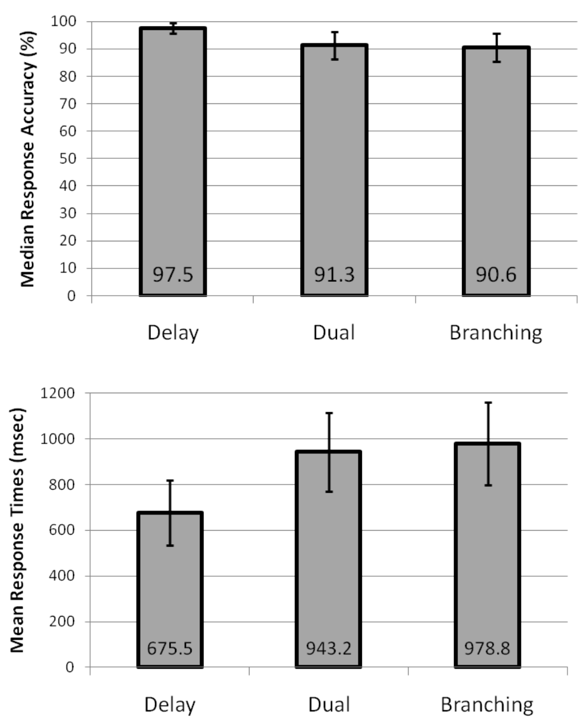


Figure 4-2. Behavioral results for Experiment 1: median accuracy & standard deviation (top); mean response time and standard deviation (bottom).

test. The level of statistical significance was set at 0.05 (Figure 4-2).

Since *dual task* and *branching* behavioral results are similar, the factor was not significant overall, but is in pairwise comparisons. We found statistical significance in response time between *delay* and *dual* ($p < 0.001$), *delay* and *branching* ($p < 0.001$), but not between *dual* and *branching* ($p > 0.05$). Similarly, we found statistical significance in accuracies between *delay* and *dual* ($p < 0.05$), *delay* and *branching* ($p < 0.05$), but not *dual* and *branching* ($p > 0.05$). Also, correlations between accuracy and response time for each task were not statistically significant. We also looked at learning effects based on response time and learning effects based on accuracies as users progressed through the experiment. We did not find a learning effect.

4.3.3.4.2. Analysis of Signal

We wanted to determine whether the hemodynamic response measured by fNIRS has a different signature between the three conditions. For each of the two probes, we selected the fNIRS measurement channels with the greatest source-detector distances (3cm), as these channels are expected to probe deepest in the brain tissue, while the closer channels are more likely to pick up systemic effects and noise.

From each of these channels, we calculated both the change in oxygenated hemoglobin and deoxygenated hemoglobin using the modified Beer-Lambert law (Chance, et al., 1988) after removing noise with a band pass filter. Thus, we used four channels corresponding with changes in oxygenated and deoxygenated hemoglobin on the left and right hemispheres. Figure 4-3 shows the mean across all participants and all trials for the three conditions. This is plotted along with the standard error. Red is delay, green is dual, blue is branching. The figures in the top row show the pattern for oxygenated hemoglobin and the bottom row shows the deoxygenated hemoglobin. The figures on the left are from the sensor on the left side of the head and the figures on the right are from the right.

In order to confirm that there were differences in brain activity during the three conditions, we did an ANOVA comparing condition means within subjects. Since the hemodynamic changes occur over a 5-7 second period, we simplified the signal for analysis by dividing the time series measurement for each trial into seven segments (~5.57 second each) and took the mean over these segments for the four channels. Since there were multiple sensors, factors for the distribution of sensors were included (left/right *hemisphere*), as well as a factor for *hemoglobin type* (oxygenated or deoxygenated) and the *time point*. We used the Greenhouse-Geisser ANOVA values to

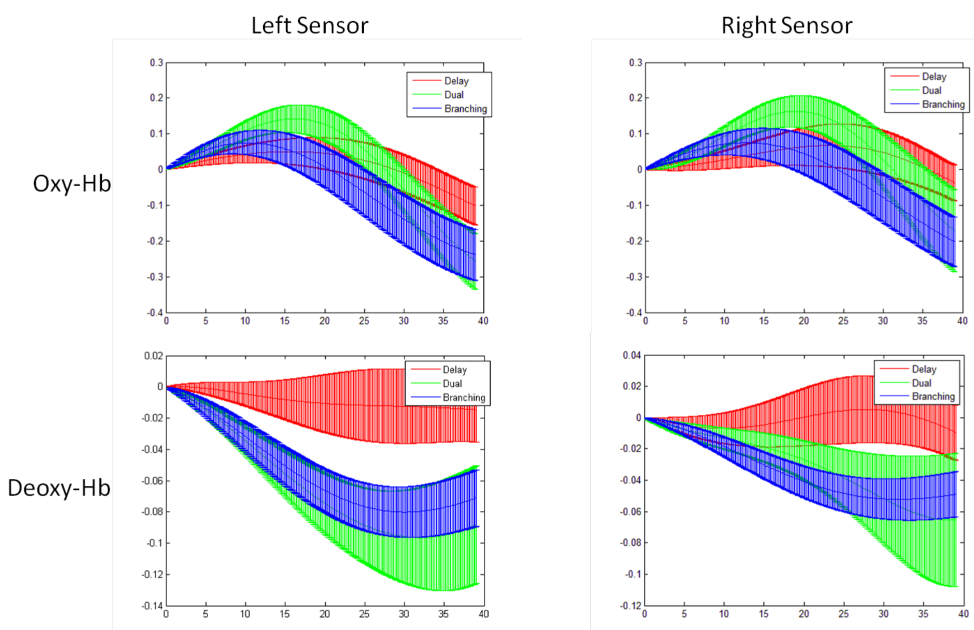


Figure 4-3. Experiment 1: Mean and standard error of fNIRS signal during a trial across all participants and all trials for Branching (Blue), Dual Task (Green) and Delay (Red). The top row displays oxy-hemoglobin signal and the bottom row shows the deoxy-hemoglobin signal. The y-axis shows the change in hemoglobin values in micromolars (μM).

correct for violations in sphericity. We found a main effect of *condition* ($F(2,22)=4.353$, $p=0.029$), indicating that there is significant difference in at least one of the conditions. There were no other significant effects in this analysis.

4.3.4. Experiment 2: Random & Predictive Branching

To follow up on the first study, we conducted a second experiment to determine whether we could distinguish specific variations of the branching task. This experiment had two conditions that were analogous to those in (E. Koechlin, Corrado, G., Pietrini, P., & Grafman, J. , 2000), in which the participant was always following the *branching* rules described in Experiment 1:

Random Branching: Rock classification and location update messages were presented pseudorandomly.

Predictive Branching: Rock classification messages were presented every three stimuli.

Ideally, when using computer systems, the default scenario for a user would be similar to the *predictive* condition, and therefore the user would be able to plan ahead and handle incoming work appropriately. If we could automatically identify that the user is experiencing *random* or unpredictable behavior, there may be appropriate adaptations that the system could make to better support the user, which we are exploring with the adaptive interface platform described below. This experiment investigates whether we can automatically detect the different scenarios using fNIRS.

4.3.4.1. Participants

This study included 12 healthy volunteers (5 male), between the ages of 19 and 32. Three additional volunteers had participated, but are not included in this analysis because their performance in the tasks was below 70% in more than two trials per condition, indicating that they were not correctly performing the tasks. In addition, data from another participant was not included due to technical issues with the fNIRS system.

4.3.4.2. Design, Procedure & Equipment

This experiment used the same procedure and equipment as in Experiment 1. However, in this experiment, there were only two experimental conditions as described above and the participants completed eighteen trials of each condition, which were counterbalanced.

4.3.4.3. Results

4.3.4.3.1. Behavioral Results

As in Experiment 1, we collected response time and accuracy throughout the study to determine whether the conditions elicited different measurements.

All variables were tested for normal distribution with the Kolmogorov-Smirnov test. For normal distributions, a paired t-test was used. For non-Gaussian distributions, we used the Wilcoxon matched-pairs signed-ranks test.

There was no statistically significant difference in response time between *random* (M=998.67, SD=190.02) and *predictive* (M=992.81, SD=213.34) branching, $t(215)=0.53$ ($p>0.05$). There also was no statistically significant difference in accuracy between *random* (M=93.982, SD=8.144) and *predictive* (M=92.824, SD=8.765) branching ($p>0.05$). Also, correlation between accuracy and response time for *random* branching was not statistically significant ($p>0.05$), but there was a statistically significant correlation in the *predictive* branching condition ($p<0.0001$).

4.3.4.3.2. Analysis of Signal

Our goal was to determine whether the hemodynamic response measured by fNIRS has a different signature between the two conditions. Our analysis was the same as in Experiment 1.

We found an interaction between *branching type*, *timepoint*, and *hemoglobin* type ($F(6,66)=3.035$, $p=0.038$). This effect indicates that, the differences in branching type depend on the hemoglobin type (oxygenated or deoxygentated hemoglobin) and on the time point. This can be seen in Figure 4-4. Therefore, it should be possible to distinguish these two conditions if we take these into account. There were no other significant effects

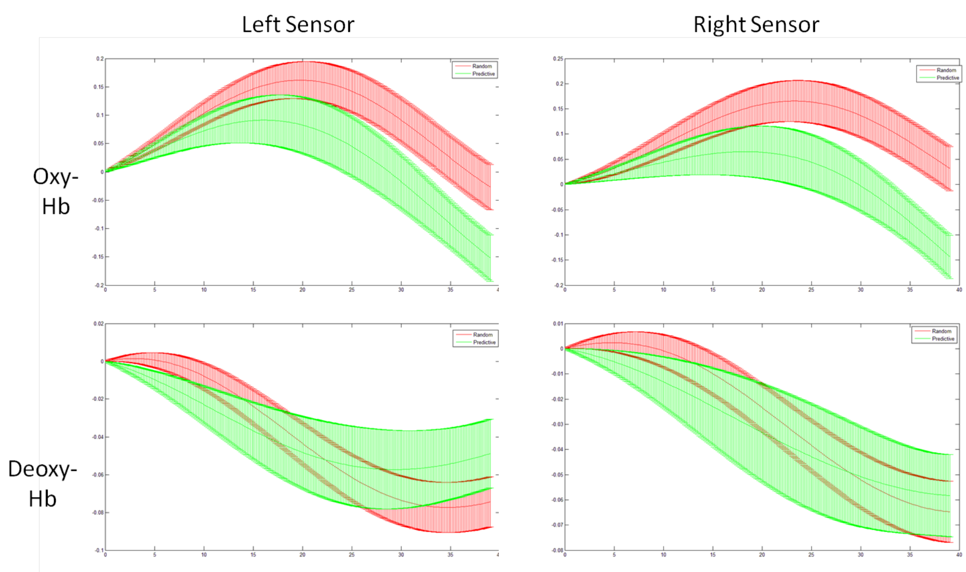


Figure 4-4. Experiment 2: Mean and standard error of fNIRS signal during a trial across all participants and all trials for Random Branching (Red), and Predictive Branching (Green). The top row displays oxy-hemoglobin [HbO] signal and the bottom row shows the deoxy-hemoglobin [Hb] signal. The y-axis shows the change in hemoglobin values in micromolars (µM).

in this analysis.

4.4. Discussion and Conclusion

This section builds a foundation for brain-based adaptive user interfaces by illustrating significant differences in the fNIRS signal in specific multitasking scenarios that could be used in HCI. First, in our preliminary study, we brought research on cognitive activity during multitasking to a system that is practical for HCI by showing that fNIRS sensors could detect states previously studied with fMRI (which cannot be used in HCI settings). In our next two experiments, we further extended this research to HCI by showing that the states elicited in the "tablet" task may be generic processes that occur in more realistic HCI tasks, by using a human-robot scenario. Although all analysis was done offline, we found significant differences in the signals that suggest that a real time classifier may be possible. In Chapter 6, we will return back to these tasks and evaluate a real time, working system that learns to recognize distinct multitasking states and adapts behavior appropriately to better support the user.

Chapter 5 PROOF OF CONCEPT BRAIN-BASED ADAPTIVE USER INTERFACE PLATFORM[3]

In the experiments described in Chapter 4, we verified that there is a significant difference between the cognitive multitasking conditions in the fNIRS signal. Because we can statistically differentiate them, we hypothesize that we can apply machine learning techniques to automatically identify these cognitive multitasking states, in real time, in a user interface. This information could then be used to drive the user interface to better support cognitive multitasking.

As a proof-of-concept, we developed a platform for studying brain-based adaptive user interfaces of this type. The system has two main components: the *Brainput* streaming

[3] Parts of the work in this chapter was originally described in “Sensing Cognitive Multitasking for a Brain-Based Adaptive User Interface” in the proceedings of the ACM CHI’11 Conference on Human Factors in Computing Systems (E. T. Solovey, et al., 2011). The integration with DIARC robot framework was joint work with Paul Schermerhorn and Matthias Scheutz.

fNIRS input channel and the Distributed Integrated Affect, Reflection, Cognition Architecture (DIARC) (Scheutz, Schermerhorn, Kramer, & Anderson, 2007) for human-robot interaction. *Brainput* expands the functionality of our Online fNIRS Analysis and Classification (OFAC) system (Girouard, et al., 2010). When used for real-time streaming input, there are four phases: *baseline*, *calibration*, *modeling,* and *testing.*

5.1. Baseline Phase

Before beginning the training phase, *Brainput* collects a baseline measure. During this period, the user is asked to relax and think of nothing, while focusing on a focal point on a computer screen. This is a standard practice in fNIRS studies and the baseline measure is used for later calculating changes in the oxygenated and deoxygenated hemoglobin values from the baseline values. Usually a baseline measure is collected for 30-60 seconds, and then an average baseline measure is calculated.

The beginning and end of the baseline period are indicated by a marker sent from an application over the serial port. In the experiments described in this dissertation, we present the baseline period using Presentation Software[4] and start and end markers are sent over a serial connection to a MATLAB program that listens for these markers. Once the end marker is received for the baseline period, the fNIRS data from the baseline period is stored for use during the modeling and testing phases.

5.2. Calibration Phase

In order for the *Brainput* system to successfully classify fNIRS data and turn it into meaningful information for a user interface, we must first collect training or calibration

4 http://www.neurobs.com/menu_presentation/menu_features/features_overview

data. A calibration session is required before each use to train a machine learning classifier for the individual that will be using the system, since there can be variation in brain processes across different people. We have the participant complete a set of tasks designed to elicit understood cognitive states. By performing these known tasks repeatedly, we can create a dataset of labeled data. We then build a machine learning model that learns to find specific patterns in this data that indicate one cognitive state or another in future, unlabeled data.

The main architecture of *Brainput* supports various calibration sessions. The only requirement is that markers are sent to the Matlab program indicating the start and end of a trial and the label that should be used for that trial. The *Brainput* system continually reads fNIRS data from the Boxy acquisition software (part of the ISS OxiTS system). In addition, we read a separate stream which is sending start and end markers that indicate when a particular known task starts and ends and which cognitive state is supposed to elicit. As in the baseline phase, we present the known tasks using Presentation Software, and start and end markers are sent over a serial connection to a MATLAB program that listens for these markers. However, this is configurable.

Once *Brainput* receives an end marker indicating the end of one of the training trials, it creates a labeled training example for the machine learning model and stores this until the end of the training session. One training example consists of a sequence of data points, beginning at the start marker and ending with the end marker. There are sixteen channels (2 sensors x 2 wavelengths of light x 4 source-detector distances) coming from the fNIRS system, so we have sixteen sequences that together are a training example.

5.2.1. Calibrating for Multitasking

From the success in Chapter 4 in detecting various multitasking states, both in the simple "tablet" task and in a robot task, we built a training module specifically for calibrating the *Brainput* system to recognize *branching* and *non-branching* states. To do this, we integrated the Presentation program used in Chapter 4 to present the *branching, delay* and *dual task* stimuli trials. We use the original tasks described by Koechlin et al. (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999) as simple calibration tasks to elicit the three multitasking states. After training the system on the fNIRS data generated during these tasks, the system is ready to recognize these states later in other tasks.

5.3. Modeling Phase

Once the training period has ended, as indicated by a marker sent over a serial port, the modeling phase begins. During this phase, the baseline data and training data are used to create a classification model for future brain data. In addition, we generate plots to examine the training data before continuing onto the next phase of the experiment.

5.3.1. Preprocessing

Several preprocessing steps are taken to convert the raw data coming from the Boxy acquisition software into meaningful information.

For each of the sixteen channels, the mean baseline value across each of the samples is taken and plotted. This is used in the both the modeling and testing phases. We then calculate the absorption coefficients for the training set, which uses the baseline measure, differential path length factors for the two wavelengths, and source-detector distances. An elliptic low pass filter with a cutoff frequency of 0.025 Hz, stoppage frequency of 0.03 Hz, max ripple of 3 dB and a stopband attenuation of 50 dB was used to filter the data. We normalized the data by channel using a z-score. For each example, we subtract the value of the first sample point from all points in the sequence to look at the changes from the same starting point. Although all trials lasted for the same length of time, it is possible that the training examples could have slightly different lengths due to differences in sampling. To ensure that all training examples have the same number of features, the examples were shortened to the length of the shortest example in the training set. Once all of these preprocessing steps were performed on the training data, we build a classification model.

5.3.2. Modeling

To build a classification model, we use Weka's (Hall, et al., 2009) sequential minimal optimization (SMO) package for Support Vector Machines which is a Java implementation of John C. Platt's algorithm (Platt, 1999). If the training set is unbalanced, the smaller class is oversampled so that the classes are balanced before training. The *Brainput* system can be modified easily to use any of the other classification algorithms included in the Weka toolkit.

5.3.3. Visualizations

We generate several plots to illustrate the fNIRS data that is generated during the training period. In most cases, these are plotted as two groups of eight subplots. The two groups are associated with the two wavelengths of light (690nm and 830nm). There are two sensors (left and right) and four source-detector distances, leading to sixteen subplots altogether.

5.3.3.1. Raw Data plot

This plot simply shows all of the raw data that was labeled for each channel. Any data collected during breaks or rest periods are removed.

5.3.3.2. Baseline data and means plot

This plots the fNIRS data that was labeled as baseline along with the mean that was calculated.

5.3.3.3. Means and standard error for each class

We plot the mean across all trials of each condition, along with the standard error for all channels. This allows the researcher to observe any obvious differences between the conditions in the training set, which could indicate that the training session was successful. These look similar to Figure 4-3 and Figure 4-4.

5.4. Classification Phase

Once the training and modeling phases are complete, *Brainput* enters classification phase. As fNIRS samples are received from the Boxy acquisition software, they are collected into a sequence of the same length as the training examples. A sliding window is created as new data comes in, and the continuous sequences of data are analyzed. Each sequence is preprocessed in the same manner as the training data (Section 5.2.1) and then sent to the machine learning classifier which classifies the sequence in real time. The classification results are sent over a socket connection. An interactive system can read this input stream and adapt behavior based on the classification.

5.5. Integration with Human-Robot System

As a proof-of-concept, we linked the *Brainput* system with the DIARC human-robot architecture. The DIARC (Scheutz, et al., 2007) is an integrated architecture for working with complex robots that can support various levels of robot autonomy and other adaptive robot behavior. To receive input from the *Brainput* system, we have created a DIARC component to which *Brainput* sends classification results via sockets. DIARC can then use these messages to change the robot's goal structures, allowing the robot to adapt its autonomy and behavior.

5.6 Additional Modes for the Platform

In addition to the normal *online* mode where signals are classified and sent in real time, *Brainput* supports *replay* mode which simulates the analysis and classification of previously recorded data, and is useful in experimenting with various adaptive strategies, similar to the demo mode in OFAC (Girouard, et al., 2010). The DIARC architecture can interface with *physical robots* in an environment, but also has a *simulation mode* that allows for simulated interactions with a robot on a computer screen, along with several different environment configurations (Figure 5-1). In addition, an *fNIRS simulation mode* was created in DIARC, where cognitive state classifications can be entered manually by inputting classification confidence levels for each of the possible states. This allows for debugging and testing of the robot adaptive behaviors, without requiring a human to be physically connected to the fNIRS sensing system.

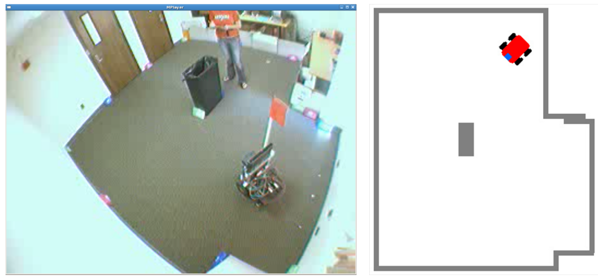


Figure 5-1. The *Brainput* integration with DIARC supports robot navigation (real or simulated), as well as fNIRS cognitive state input (real or simulated).

5.7 Conclusion

The platform we have developed allows us to explore adaptive behavior to find the best strategies for use in interactive systems. By developing a calibration module to recognize multitasking states, we can look specifically at supporting multitasking. Integrating the *Brainput* system with DIARC provides the robot with cognitive state information of the human, affording the robot to adapt its behavior to better support and collaborate with the human operator. Driven by fNIRS cognitive state input, DIARC can adapt various aspects of the human-robot interface, such as level of autonomy of the robot, the frequency, length and style of the robot's status updates, as well as the tone of the robot's voice. We can now begin to develop complex systems that adapt based on fNIRS brain signals and experimentally evaluate them. We will further explore the value of this in Chapter 6.

Chapter 6 Evaluation of Brainput [5]

In Chapter 5, I described *Brainput*, a system that processes and classifies fNIRS data in real time and sends classification results to interactive system. In Chapter 4, I presented background experiments that showed that three specific cognitive multitasking states could be distinguished from each other in a non-interactive situation. These results suggest that an interactive system that is aware of the user's changing cognitive state during multitasking would be possible. Here we evaluate the efficacy of the full working system. We demonstrate that we can use non-invasive methods to detect signals coming from the brain that users naturally and effortlessly produce while using a computer system. *Brainput* learns to identify brain activity patterns occurring during multitasking, by building on the experiments in Chapter 4. It then provides a continuous, supplemental

[5] Parts of the work in this chapter was originally described in "*Brainput*: Enhancing Interactive Systems with Streaming fNIRS Brain Input" in the proceedings of the ACM CHI'12 Conference on Human Factors in Computing Systems (E. T. Solovey et al., 2012)

input stream to an interactive human-robot system, which uses this information to modify its behavior to better support multitasking.

To evaluate the effectiveness of using *Brainput* in an adaptive system, we conducted a user study in which participants used the system to complete a human-robot team task that involved multiple robots navigating through an environment. This task required the user to constantly switch context between the two robots, maintaining information about each robot's current location, which we presumed would lead to *branching* states (as described in Chapter 4). The robot adapted its behavior based on whether a *branching* state was identified. We created three adaptation schemes for comparison that are triggered by the brain input stream. Participants completed the robot navigation task three times, each employing a different adaptive behavior as a response to the brain input stream.

6.1. Experimental Task

The main task for the study is a multi-robot version of the task introduced in (Scheutz, Schermerhorn, & Kramer, 2006). Participants remotely supervised two robots (the blue robot and the red robot) that were exploring different areas of a virtual environment. Participants were told that the two robots had collected information that needed to be transmitted back to the control center. The robots could help the participant search for an appropriate transmission location by measuring and reporting the signal strength in its current position. Transmissions were only possible in locations with signal strength of at least 2400 (values ranged from 1300 to 2500, and the single target region in each robot's area covered roughly 1.25 % of the environment). The user had a console to view the environment from each robot's point of view (Figure 6-1) and could issue commands to the robots such as "go straight," or "turn right" and the appropriate robot would follow the commands. They could also ask the robot for the signal strength of the current location by clicking “take a reading” and the robot would report the current signal

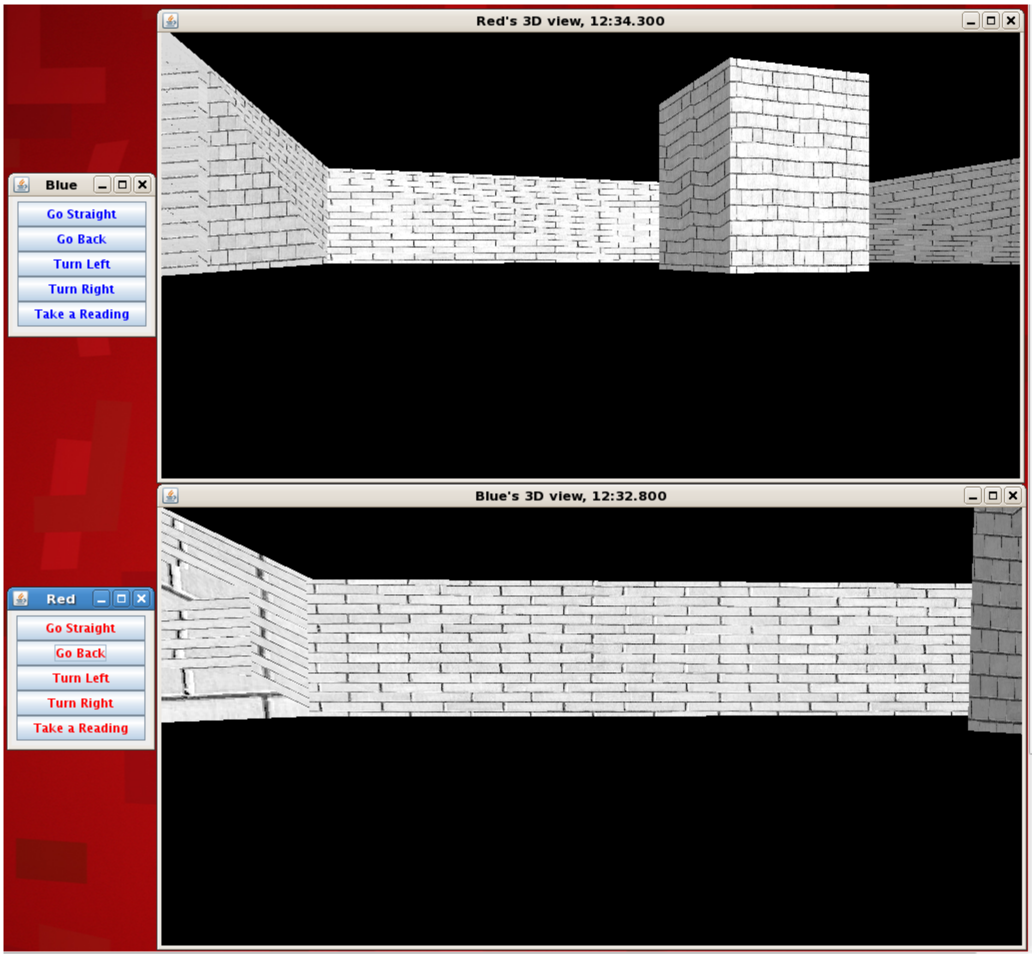


Figure 6-1. 3D view from robots' perspectives in navigation task. There was a separate navigation control for each of the robots, positioned to the left of the robot’s 3D view

strength. This required the robot to stop and also consumed resources so the robot could not be measuring signal strength at all times. In addition to the command interface, the red robot received cognitive load estimates from the fNIRS system.

Participants were told that the task would last for five minutes and that the task was considered a failure if either robot did not find a transmission location in time. The robots moved continuously throughout each task run, except when (a) pausing to measure signal strength, (b) in a collision state with a wall or obstacle, or (c) at the target location.

Participants were instructed to avoid collisions with obstacles and walls, and were advised not to leave either robot idle, as it may go into a hibernation state to save power. These constraints helped to ensure that the participants engaged in multitasking between the two robot consoles and did not focus on finding one robot's transmission point and then moving to the second.

6.2. Participants

This study included eleven participants (three male), between the ages of 18 and 22 (mean 20.7). All participants were right-handed, had no history of brain injury and had normal or corrected-to-normal vision.

6.3. Equipment

We used a multichannel frequency domain OxiplexTS from ISS Inc. (Champaign, IL) described in Section 2.1.2.1.

6.4. Calibration Phase

Before working with the robot, each participant completed a calibration phase to gather fNIRS data in known multitasking exercises (as described in Section 5.2.1). This data was used to build a machine learning model for classifying these cognitive multitasking states. The stimuli were the same as the *delay*, *dual task* and *branching* conditions described in (E. Koechlin, Basso, G., Pietrini, P., Panzer, S. & Grafman, J., 1999) and shown in Chapter 4 to evoke different fNIRS brain activity patterns. By using known tasks that induce known cognitive states, we can train a machine learning classifier to recognize these states later, in new contexts.

Each run of the known multitasking activities lasted approximately forty seconds, and the fNIRS brain signal over this 40-second window became the training example. Once a machine learning model was built from this training data, the Brainput system continuously classified 40-second slices of fNIRS data, in a sliding window that moved with each new sample. The sampling rate for the fNIRS system was 6.25 Hz.

6.5. Conditions

We compared our adaptive system with two alternate conditions in the experiment. One condition simply turned off autonomy as a baseline. The other used the *Brainput* inversely to probe more deeply into whether the brain input had any effect, similar to the experimental design in (Pope, Bogart, & Bartolome, 1995). Thus, there were three conditions in the study, varying only in the adaptive behavior that was triggered in the red robot by the fNIRS brain input:

1) In the *adaptive condition*—our system—the red robot went into autonomy mode whenever a branching state was detected, indicating that the user was tending to multiple tasks and maintaining information about the primary task over time. This allowed the participant to focus on the blue robot. The red robot exited autonomy mode when a non-branching state was detected, requiring the human to give instructions to the robot about where to explore.

2) In the *non-adaptive condition*, the brain input was ignored, and the red robot never acted autonomously.

3) In the *maladaptive condition*, the rule was reversed from the adaptive condition. When a non-branching state was detected, the red robot began working autonomously and stopped when a branching state was detected, waiting for

commands from the participant. The autonomy mode provided the same assistance as in the adaptive condition, and thus should still allow the user to focus on the blue robot. The only difference is the timing of the onset of the autonomy mode. This condition allows us to investigate the effect of the mapping between Brainput and the adaptive behavior.

In autonomy mode (regardless of which condition), the red robot would take over the search task, periodically sensing the signal strength and making appropriate course adjustments to ensure progress toward the target location. Note, however, that even in autonomy mode the robot could be interrupted by the operator (e.g., when asked to take a reading of the signal strength), but would return to the autonomous behavior after completing the requested action.

The blue robot never acted autonomously, as we wanted to ensure that the human operator always had a task to perform. The red robot staying in autonomous mode throughout the entire task would not be ideal as the human needs to be aware of the robot's location and progress to provide corrective feedback (as the robot's search behavior is not optimal), to ensure that the message is transmitted before time is up.

6.6. Experimental Procedure

Before the experiment, each participant completed a practice session without the fNIRS sensors, first with the robot navigation task and then with the multitasking "tablet" exercises. This allowed the participants to familiarize themselves with each of the tasks. For the robot practice session, neither robot was autonomous as we simply wanted the participant to learn how to use the console and see the robot in action. In the multitasking practice exercises, the three distinct multitasking exercises (branching, dual task, and delay) were presented in a counterbalanced order. Each was repeated until the participant achieved greater than 80% accuracy in the task.

After the practice sessions, fNIRS sensors were applied to the forehead and the machine learning calibration session began. The participant completed known multitasking exercises to allow us to build an individual model of fNIRS activity for classification. The participant was presented with an initial rest screen, which was used to collect a one-minute baseline measure of the brain activity at rest. After that, the user had to complete two sets of ten 40-second trials. There were ten trials of *branching*, five of *delay* and five of *dual task,* which were presented randomly. Between each trial, the user was presented with the instructions for the next trial, followed by a 10-second rest screen.

Once the training session was complete, the data was used to build a training set for a machine learning model as described in Section 5.3.2. Any trial where the participant achieved lower than 70% accuracy was not used as this indicated that they were not actually performing the task. For the main experimental task, we were most interested in detecting *branching* states, as the workload level can be quite high with the demand from both context switching and working memory load. The *dual task* and *delay* trials were combined into one category for the machine learning model as non-branching. If the number of branching and nonbranching training examples were not equal, the smaller set was oversampled so that the classes were balanced. The machine learning model was built using Weka's (Hall, et al., 2009) SMO package for Support Vector Machines.

Once the model was built, the participant did one five-minute session of the robot navigation task in each of the three conditions. After each session, the participant filled out a NASA Task Load Index questionnaire (NASA-TLX) to provide their subjective assessment of task load. The first five participants also provided voluntary additional comments about their experience with each of the robots. To formalize this, the second set of six subjects also filled out a questionnaire on their perceptions of the robot in each of the three conditions. During the navigation tasks, the system logged all commands issued, fNIRS multitasking classifications received, and events such as collisions with obstacles.

6.3. Design and Analysis

The study used a within-subjects design. The independent variable is the robot's adaptive condition: *adaptive*, *non-adaptive* and *maladaptive*. All participants performed one five-minute session in each of the conditions. The condition order was counterbalanced. To evaluate whether the multitasking state information was valuable in the navigation task and produced differences between the three conditions, we investigated the following dependent measures: NASA-TLX questionnaire results, the robot perception questionnaire, and task performance from the log files, including number of completed tasks, number of commands issued, number of collisions, and maximum signal strength found.

6.4. Results

6.4.1. Performance Results

We examined several aspects of task performance to see how they were affected by the adaptive condition.

First, we looked at the mean time (in seconds) that the red robot spent in autonomy mode in the *adaptive* condition (M=83.12, SD=35.94) and the *maladaptive* condition (M=88.00, SD = 38.28). During the *non-adaptive* condition, the robot was never in autonomy mode.

There was no significant difference between the two conditions, $t(10) = 0.3737$, $p=0.7164$ (Figure 6-2). Since the time spent in autonomous mode was equivalent in the two conditions, we can attribute differences in performance between the conditions mainly to the timing of the autonomy mode initiated by *Brainput.*

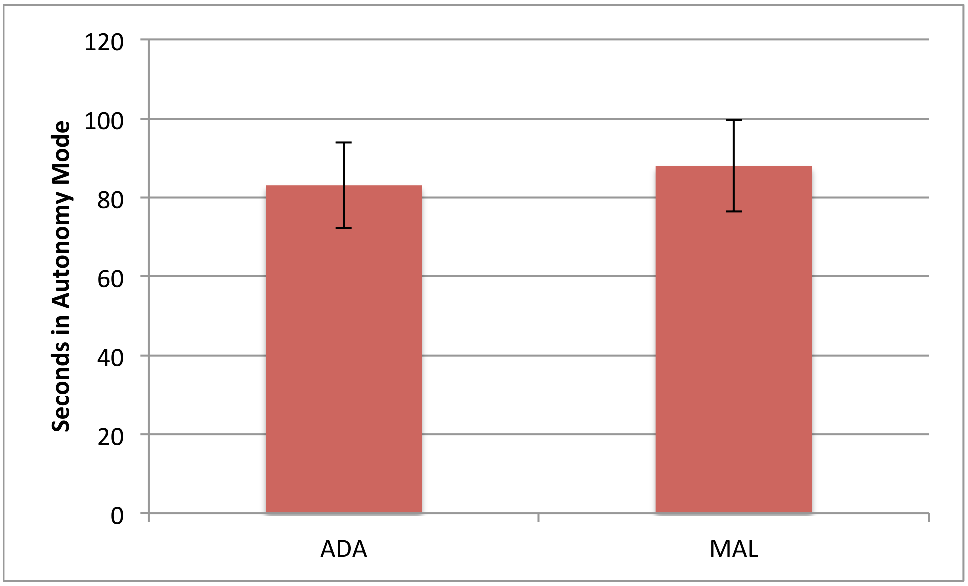


Figure 6-2. Mean and standard error of time spent in autonomy mode across 11 participants.

Since the autonomy mode should help the participant find the transmission location, we expected that we may observe higher task completion in both the *adaptive* and *maladaptive* conditions, over the *non-adaptive* condition where the user always had to control both robots. For the blue robot, we did find this (5, 4, 3 participants, respectively) but the result was not statistically significant (Figure 6-3). For the red robot, we did find a higher completion rate in the *adaptive* condition (9 out of 11 participants) than in the *non-adaptive* condition (5 out of 11 participants), as expected. However, the *maladaptive* condition had a lower completion rate (2 out of 11) than both the *adaptive* and *non-adaptive* conditions (Figure 6-3), indicating that the autonomy is helpful only when it is well-matched to the user's cognitive state. With a Cochran's Q test, we found a significant difference among the three adaptive conditions ($\chi^2(2) = 10.57$, $p < 0.01$). A pairwise comparison using continuity-corrected McNemar's tests with Bonferroni correction revealed that significantly more participants completed the task in the *adaptive* condition than in the *maladaptive* condition ($p<0.1$, $\phi = 0.48$).

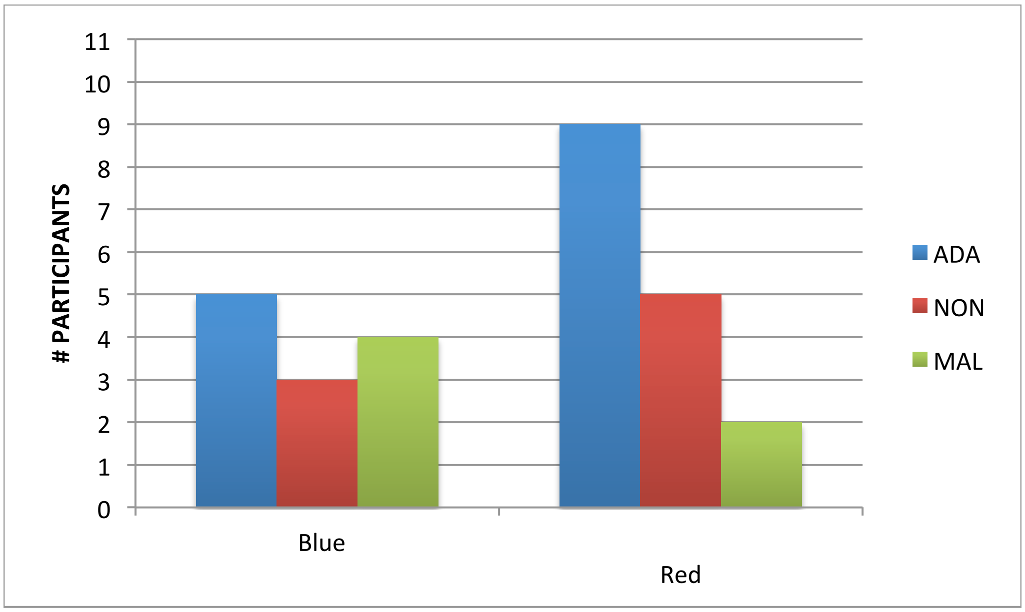


Figure 6-3. Number of participants (out of 11 total) that completed each of the tasks.

To get a more fine-grained look at task completion, we investigated the maximum signal strength found (Figure 6-4). Since the main goal was to find a transmission point above 2400, this could give an indication of how close the participants came to completion. Nonparametric analysis was used since the Shapiro-Wilk normality test showed that the data from each condition was not from a normal distribution. The adaptive condition resulted in the highest median for the maximum signal strength of the red robot (2416.0). The maladaptive condition resulted in the lowest median for the maximum signal strength of the red robot (2108.0). The non-adaptive condition was in the middle (2336.0). A Friedman nonparametric repeated measures ANOVA confirmed that the difference in the medians was statistically significant ($p<0.001$). Dunn's multiple comparisons post-hoc test was conducted and showed a significant difference between the adaptive and maladaptive conditions ($p<0.001$). There was no statistically significant difference in the medians of the maximum signal strength of the blue robot across the three conditions.

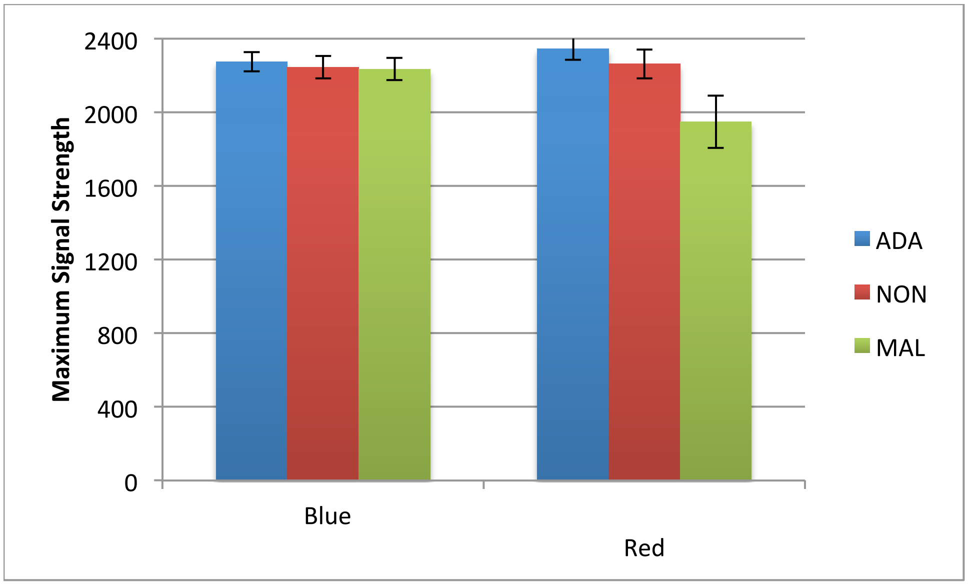
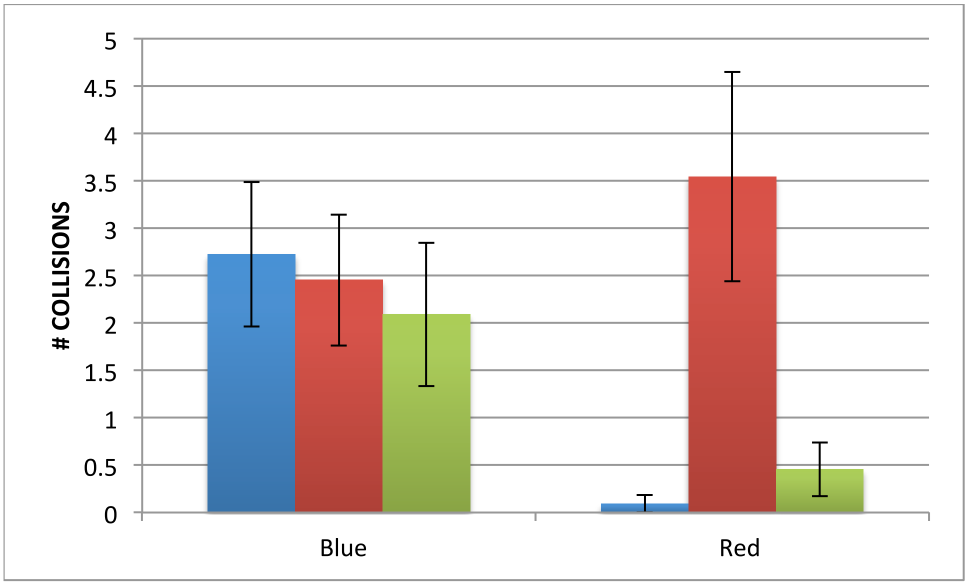


Figure 6-4. Mean and standard error for maximum signal strength found in each condition across 11 participants.

We then looked at the number of collisions in each of the conditions (Figure 6-5). Participants were told to avoid collisions with walls and obstacles, as it would damage the robot. This ensured that both robots were attended to throughout the tasks. The Shapiro-Wilk Normality Test showed that the data from each condition was not taken from a normal distribution. A Friedman nonparametric repeated measures ANOVA was performed to compare the medians. We did not find any statistically significant difference for the blue robot. However, for the red robot, we did find that the non-adaptive condition resulted in a higher number of collisions than the two adaptive conditions ($p=0.005$). Collisions during the non-adaptive condition may indicate a performance degradation in the participant since a non-autonomous robot would likely walk into a wall if ignored, since there was no way to pause or stop the robot from moving.



Figur 6-5. Mean and standard error for the number of collisions with each robot across 11 participants.

Finally, as a measure of effort during the tasks, a repeated-measures analysis of variance was carried out to determine whether the adaptive mode had any effect on the number of commands issued (Figure 6-6). For the red robot, there was a statistically significant main effect of adaptive condition, $F(2,20) = 3.691$, $p = 0.04$, but post-hoc analysis did not reveal any statistically significant results. There was no statistically significant difference for the blue robot, $F(2,20) = 1.153$, $p = 0.34$.

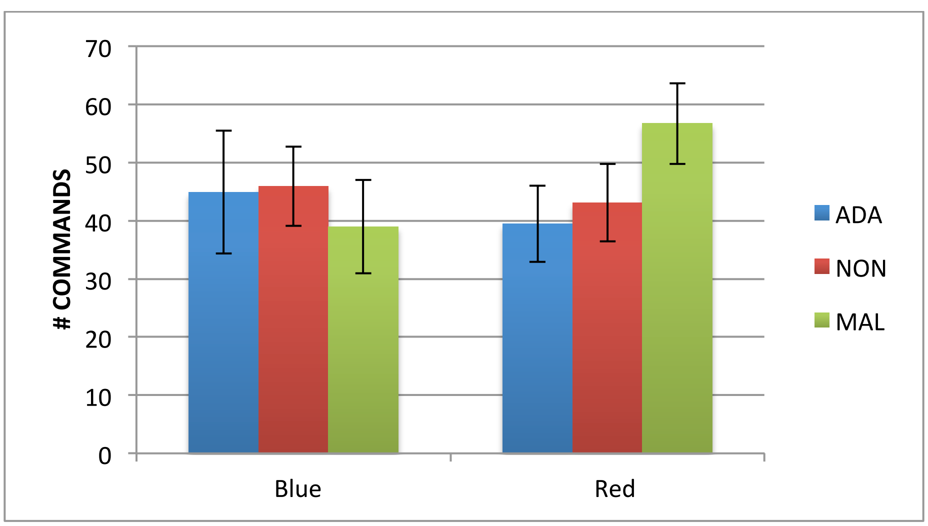


Figure 6-6. Mean and standard error for the number of commands issued with each robot across 11 participants.

6.4.2. NASA-TLX Results

The goal of implementing adaptive behavior is to decrease the user's workload level. To investigate the success of this, we analyze the results of the NASA-TLX questionnaire (Hart & Staveland, 1988). This survey was designed to take into account individual differences in perceptions of workload. The questionnaire asks the participant to rate the workload level of the task in several categories. In addition, the user evaluates pairs of workload categories and indicates the one that contributes most to the workload. This is used to generate a set of weights that are applied to the other workload ratings and that reduce inter-subject variability. The means and standard deviations for the NASA-TLX scores from our study are shown in Figure 6-7. A repeated-measures analysis of variance on the NASA-TLX score showed that there was a statistically significant main effect of adaptive condition, $F(2,20) = 4.65$, $p = 0.02$. A Tukey's pairwise comparison revealed the significant differences between *adaptive* and *maladaptive* ($p < 0.05$). Each dimension of workload (mental demand, physical demand, temporal demand, performance, effort, and frustration) was analyzed separately using Friedman's non-parametric repeated measures ANOVA. The adaptive mode had a significant effect on performance ($p<0.05$) and frustration ($p<0.05$).

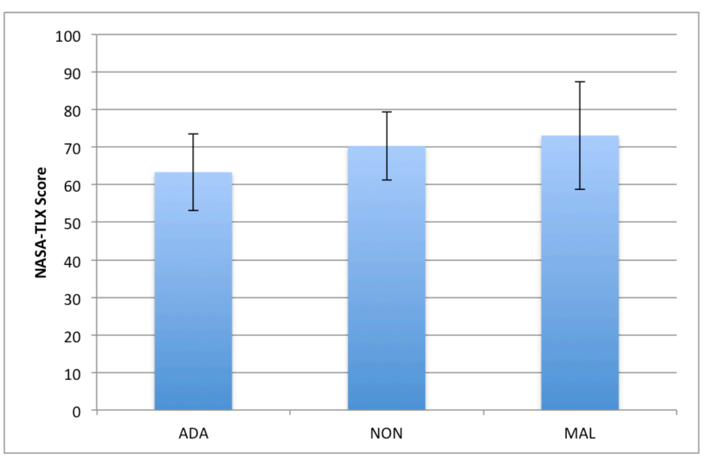


Figure 6-7. Mean and standard deviation in NASA-TLX results. There was a statistically significant main effect of adaptive condition, F(2,20) = 4.65, p = 0.02.

6.4.3. Perceptions of Adaptive Behaviors

The first five participants in the study provided informal comments about the robots and the different conditions. Many commented on the behavior of the red robot. For example, after the *adaptive* condition, one participant said, “Although red robot occasionally disobeyed my commands, for the most part it was cooperative and found the transmission spot. Blue robot was still very cooperative.” The same participant had this comment after completing the *maladaptive* condition, “Blue robot was much more cooperative than the red robot, which frequently disobeyed my commands and would go in its own direction.” From these comments, we can see that the participant found the red robot to be mostly helpful in the *adaptive* condition, but that it disobeyed in the *maladaptive* condition. To capture a clearer picture of the perceptions of the robots, we had the next six participants complete a questionnaire on their perceptions of the robots. The results are illustrated in Figure 6-8.

This is a small sample of users, and so the results of the questionnaire are preliminary. However, some patterns are beginning to emerge. First, we see that the participants seemed to agree that in the *maladaptive* condition, the robots appeared to make their own decisions and that the robots appeared to disobey the user’s commands. There was less agreement on those points for the *adaptive* condition, even though the red robot was autonomous in this condition as well. This indicates that when the robot was autonomous at appropriate times (based on the branching classification from fNIRS), it was less noticeable to the user.

In the *maladaptive* condition, the participants indicated that the robots were more annoying than in the other conditions. This would make sense since they also felt that the robots were disobeying their commands. The lower score for "The robots were cooperative" in the *maladaptive* condition also corresponds with the other statements.

It is interesting to note that the ratings did not show strong differences between the *adaptive* and the *non-adaptive* conditions for these statements: "*The robots acted like members of the team*," "*The robots were annoying*," "*The robots were capable*," and "*The robots were cooperative*." This provides evidence that the user hardly noticed the adaptive behavior when it was consistent with the user's needs. This is consistent with what we found in previous single-robot studies with both real and simulated robots (Schermerhorn & Scheutz, 2011).

6.5. Discussion

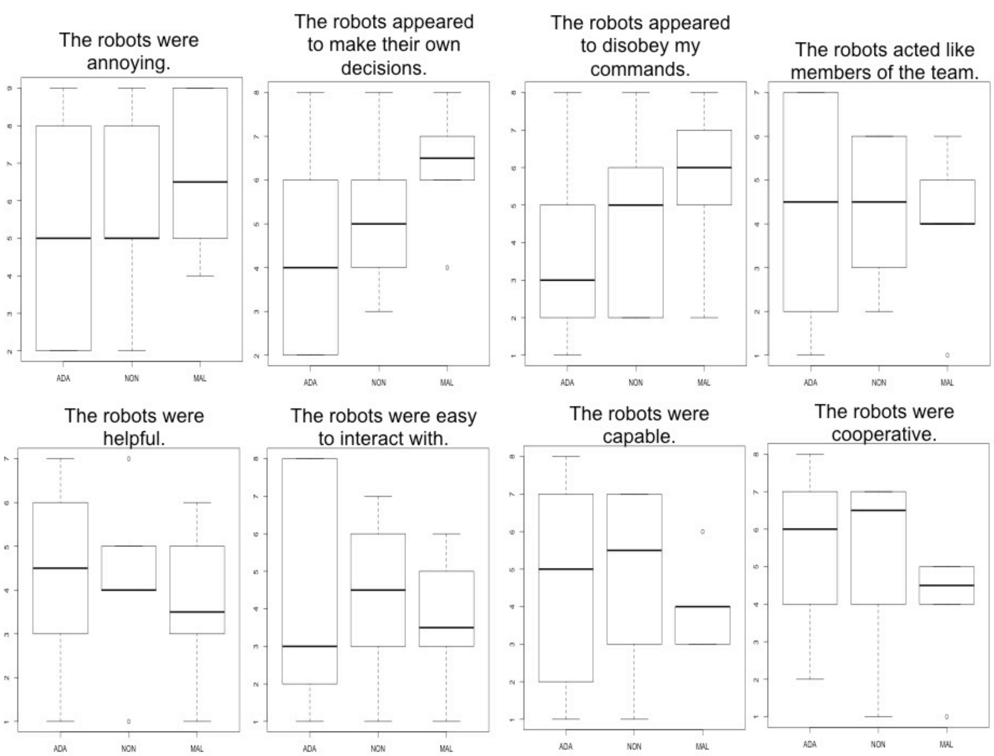


Figure 6-8. Preliminary results from robot perception questionnaire for the three conditions: adaptive (ADA), non-adaptive (NON) and maladaptive (MAL) (N=6). The scale was from 1 (strongly disagree) to 9 (strongly agree).

Overall, our results suggest that *Brainput* provided measureable benefits to the user, with little additional effort required of the user. This study also confirmed that we can train a machine learning classifier on a set of known tasks and later successfully classify brain activity in unrelated activities that elicit similar brain processes.

The NASA-TLX results indicate that the *adaptive* condition had the lowest task load rating and the *maladaptive* condition had the highest, indicating that appropriate adaptive behavior helps to reduce workload, while adverse adaptations can actually make the system perform worse. In addition, the completion rates and maximum signal strength for the red robot were highest in the *adaptive* condition and lowest in the *maladaptive* condition, indicating that the adaptive behavior triggered by the *Brainput* correlates to performance improvements.

One question that could be asked now is: how accurate is the cognitive state classification? However, this is not straightforward to answer because we do not have labeled data in this study. Since each participant may have different strategies and we did not constrain the participant to conduct the task in a specific way, we do not have "ground truth" as to when the participant was actually in a *branching* state and when they were not. However, the primary goal of this system is to better support multitasking, as indicated by better performance or better experience. High classification is a means to the main goal of better performance, which we have obtained.

However, we can make some guesses about the classification accuracy by looking closer at the differences between the *maladaptive* and *adaptive* conditions. The red robot's autonomous behavior in both the *adaptive* and *maladaptive* conditions was the same: it made progress toward the target location. For this reason, we hypothesized that task performance would be improved in both the *adaptive* and *maladaptive* conditions, since both conditions had the aid of an autonomous robot. As noted above (Figure 6-3), this was not the case for the *maladaptive* condition as the number of successful task runs was lowest.

This raises the question of how a properly functioning cooperative teammate could decrease performance. One likely explanation is hinted at by the number of commands issued in each condition (Figure 6-6). Participants issued more commands to the red robot in the *maladaptive* condition than in the other conditions, and the *maladaptive* condition is the only one in which the red robot received more commands than the blue robot. It seems that subjects were less accepting of autonomous behavior occurring during non-branching phases than during branching phases, and expended effort trying to "correct" the robot in that condition. This is reflected also in the subjective assessments: the robots were rated as more annoying and disobedient, and less helpful and cooperative in the *maladaptive* condition. These results demonstrate that basing the autonomy onset on the cognitive multitasking state has a positive impact on subjective task load.

As the name implies, the strategy adopted by the red robot in the *maladaptive* condition is not being proposed as a potentially viable candidate for future robotic architectures. Instead, the *maladaptive* condition is included to serve as a direct contrast to the *adaptive* condition, similar to the comparisons of positive and negative feedback loops in (Pope, et al., 1995).

Comparing the *adaptive* and *non-adaptive* conditions demonstrates that robot autonomy can improve task performance, but that is not surprising, having been shown in prior work (e.g., (Schermerhorn & Scheutz, 2011)). What is unclear, however, is whether the *Brainput*-initiated autonomy transitions correspond to meaningful cognitive state transitions in participants. Periods of autonomy might seem likely to be helpful in a task like this regardless of when they occur, so comparisons between the *adaptive* and *non-adaptive* conditions cannot, by themselves, support claims regarding the legitimacy of *Brainput* classifications. However, contrasting the *adaptive* and *maladaptive* results makes it immediately apparent that *Brainput* has successfully identified a distinction in cognitive states: if *Brainput* were not detecting a genuine difference in cognitive load, one would expect no difference between the *adaptive* and *maladaptive* conditions, and could attribute all of the performance benefits to the proportion of the time spent in autonomy mode. Instead, participants respond significantly differently to autonomy initiated when *Brainput* indicates a branching state than to autonomy initiated when *Brainput* indicates a non-branching state. This constitutes strong evidence that the system is properly categorizing the fNIRS data -- *Brainput* implicitly provides information, distinguishing between times in which autonomous operation is beneficial, and those in which autonomous operation is detrimental to the task.

Chapter 7

CONCLUSIONS

7.1. Summary of Work and Contributions

The ability to capture subtle changes in the user's cognitive state in real time opens up new doors in human-computer interaction research. This information can be used as a continuous input stream to an interactive system, making the system more in sync with the user, and providing appropriate help and support when needed. This dissertation takes concrete steps toward this vision by showing that sensing fNIRS brain data is practical for HCI. I take a different approach for brain-computer interfaces that augments traditional input devices such as the mouse and keyboard and that targets a wider group of users than traditional brain-computer interfaces for disabled users. Brain sensor data is used as a passive, implicit input channel that expands the bandwidth between the human and computer by providing extra information about the user.

The previous chapters served to support my thesis: *The emerging brain sensing technology, functional near-infrared spectroscopy, can be used as streaming input to an interactive user interface, which adapts its behavior based on the brain signal and improves user performance and experience compared to a traditional user interface.*

In support of this thesis, I have made several contributions to HCI research, described below.

7.1.1. Guidelines

With the introduction of any new technology, there are considerations that should be made for its proper use. For this reason, we used our earlier experience with fNIRS as well as a literature review to recognize characteristics specific to fNIRS sensors that are relevant for HCI, and develop paradigms for using fNIRS properly in HCI research. These were described in Chapter 3 and provide guidelines for future researchers exploring fNIRS for HCI.

7.1.2. Analysis tools

Since fNIRS is relatively new, there are not established methods for analyzing the raw data from the device. Thus, I have developed tools that can be used to better understand the data coming from the machine. To classify cognitive states from fNIRS data alone, I implemented noise reduction and machine learning classification algorithms. These work in real time, as data is collected, in order to adapt the system in real time. These techniques will improve analysis of any new fNIRS data.

In Chapter 4, I described preprocessing steps and offline analysis techniques. In Chapter 5, I described the real-time system architecture that includes preprocessing, visualizing, and classification of the fNIRS signal. These tools were valuable in the studies described here, but also can be used for future experiments using fNIRS.

7.1.3. Cognitive State Classification

I show that we can use fNIRS to detect signals that are valuable to HCI. Specific cognitive multitasking states, previously studied with fMRI (which cannot be used in HCI settings), can be detected automatically with fNIRS which is more practical for HCI. I also show that these cognitive multitasking brain processes are detectable across multiple domains and tasks, by moving from a simple letter-based task in previous work to actual HCI-related tasks that elicit similar states. This is valuable because it shows that we can detect it *branching* in new domains, even if we train a classifier using known scenarios. These distinct brain processes are almost indistinguishable by examining overt behavior or task performance alone, and so the fNIRS brain sensors provide otherwise unavailable information. These cognitive multitasking states have direct relevance to many HCI scenarios.

7.1.4. Streaming fNIRS input channel

I describe *Brainput*, a passive, implicit input channel to an interactive system, based on real-time cognitive multitasking state detection with fNIRS. This system was integrated with a human-robot system. Together, this platform provides the basis for the design and evaluation of future brain-based adaptive user interfaces, with broader applications beyond human-robot team tasks.

7.1.5. System Evaluation and User Study

We successfully integrated the *Brainput* system into a robot architecture and demonstrated that it can successfully be used to reduce human workload in interactions with multiple robots. This shows the potential for fNIRS-based brain input in human-robot interaction, and opens the door for studying such interfaces in broader domains and situations. *Brainput* gives the interactive system a valuable additional information channel that can be used to improve team performance without adding to the operator's load. This suggests that implicit brain input as a supplemental input stream has promise both in human-robot interaction and in various other domains and tasks.

7.1.6. Design Implications

Because the brain input is implicit (unlike a mouse or keyboard that the user explicitly uses for input), we do not want to surprise or confuse the user by making unexpected changes to the interface. In addition, the data is often noisy, and is constantly changing. Plus, the machine learning classifications are unlikely to be perfect, leading to unreliable, imperfect cognitive state classification. Therefore, the adaptive interfaces should make subtle, helpful changes to the interface that ideally would not be too disruptive if the user's state was misinterpreted occasionally. For example, in Chapter 6, we used the cognitive multitasking data to change the autonomy level of one of the robots. This does not affect the primary navigation task, and the user can always provide commands to the robot, even during autonomy mode. In our evaluation, it was shown that this adaptive behavior was indeed beneficial to the user. In other contexts, the cognitive state information may be used in other helpful (but not mission-critical) ways such as to change future interactions, to pre-choose defaults, or to change the effect of a click.

7.1.7. Research Approach

Finally, this dissertation presents an approach to exploring fNIRS for HCI. We began by looking to prior work with fMRI to pursue cognitive state detections that may be feasible with the fNIRS sensors worn on the forehead. In controlled experiments, we examined these to determine whether they could be measured accurately using fNIRS. Once a promising set of tasks was established, we built real world systems that take this information as input.

7.2. Future Work

There are several directions for future work originated by this dissertation.

7.2.1. Other Multitasking Scenarios

Multitasking has become integral in many aspects of our lives, so there are opportunities to explore *Brainput* in other tasks and domains. In any activity involving multitasking or information overload, we could expect to see improvements in the user's performance and experience. Some examples of other domains are complex data analytics, air traffic control and management of multiple unmanned vehicles. In addition, we believe that similar brain-based user interfaces may support a wide range of contexts that involve multitasking and interruptions. Moving into entertainment applications, we can imagine augmenting the video game experience in exciting ways by utilizing the subtle signals picked up from the user during game play.

7.2.2. Adaptive Strategies

We demonstrated that the platform can form the basis for brain-based user-interfaces by implementing a simple adaptive scheme based on the experiments. However, to design a successful adaptive user interface many factors could be considered and the automation scheme could be more complex than that illustrated here. Our platform will enable us to conduct evaluations of various adaptive behaviors to determine the appropriate strategy for supporting multitasking by utilizing signals coming implicitly from the brain. We can now begin to develop complex systems that adapt based on fNIRS brain signals and experimentally evaluate them.

To further explore effective adaptive strategies, we also could conduct a follow-up study to determine how a human (as opposed to the robot in our experiment) would adapt behavior to better to support a colleague that was multitasking. This would be valuable information since humans naturally adapt their behavior when interacting with others, and more importantly, we expect this. If computers adapted in similar ways as a person would, their behavior may be more predictable to the human operator. This would not require fNIRS sensors. Instead, we could provide a simulated stream of fNIRS classifications for *branching* and *non-branching.* These could be displayed to a user to indicate the current state for a teammate. It would be valuable to see how the user would adapt his or her behavior when receiving notification that the teammate is in a *branching* state.

7.2.3. Physical Robots

It is likely that interacting with real, physical robots will lead to some differences in operators' overall cognitive states and evaluating *Brainput* with real robots will be important for determining its applicability to real-world human-robot problems.

7.2.4. Additional Cognitive State Sensing

There likely exist other cognitive states that could be exploited to improve human-computer interaction efficiency. In Experiment 1 in Chapter 4, we differentiated between *branching, dual-task* and *delay*. Experiment 2 showed differences between *random* and *predictive* branching. Although we have shown potential of differentiating these states, the system evaluated in Chapter 6 simply distinguished between *branching* and *non-*

branching states. It would be straightforward to expand this to the other states that were discussed in Chapter 4, by creating calibration modules for presenting these states to the user in order to collect training data for the machine learning model. In addition, further studies can be conducted to determine additional cognitive states that could be distinguished reliably for use in an interactive system. Previous studies have been conducted to determine the feasibility of recognizing cognitive workload levels (L. M. Hirshfield, et al., 2007), game difficulty levels (Girouard, et al., 2009), and specific cognitive resources (i.e. verbal working memory) (Leanne M. Hirshfield, et al., 2011; Leanne M. Hirshfield, et al., 2009) with the fNIRS device.

7.2.5. Analysis Improvements

We intend to enhance the machine learning techniques to improve the accuracy of the system. We will also analyze the fNIRS response more deeply in our future work, but our initial goal was show that there was a significant difference between the signals we detected for the different conditions. This allows us to discriminate the conditions, and adapt a user interface when each state is detected.

7.2.6. Noise and Artifact Reduction

In Chapter 3, we investigated sources of noise and artifacts. In the future, it would be worthwhile to take these results a step further, to investigate even more realistic settings with multiple potentially interfering sources of noise. In addition, it would be useful to investigate using machine learning to identify the presence of artifacts in fNIRS data. With a database of undesirable artifacts in fNIRS signals, we could feed data from a new experiment to see whether any of the artifacts are found. This could provide a new and objective way to remove examples contaminated by such artifacts, instead of using visual observation.

7.2.7. Additional Sensors

Our fNIRS machine is made up of sensors that are worn on the forehead and that probe the anterior prefrontal cortex. It would be valuable to collect more data in this area with denser arrays of light sources. We also could look at other areas of the brain if we had sensors designed for this. This would expand the cognitive state data that could be detected for an interactive system. Finally, it would be interesting to integrate the fNIRS system with an EEG system since they provide complementary information.

7.2.8. Disabled Users

Finally, while most of my research has focused on the broader population of healthy users, many of the results would benefit disabled users as well, by providing additional channels of communication in a lightweight manner.

7.3. Closing Remarks

Since *Brainput* has unique characteristics that set it apart from most standard input techniques, I have explored the effective use of the fNIRS brain data in human-computer interaction. From this exploration, I built a system that effectively uses the fNIRS input to adapt an interactive system to better support the user. This is an early step towards computers that can interpret the user's cognitive state and adapt accordingly.

APPENDIX A: PARTICIPANT QUESTIONNAIRE

Thanks for taking part in our study. Please fill out this form. All information will be kept confidential.

What is your subject code? (ask researchers for code) *

Enter age: *

Gender: *

- Male
- Female

Are you: *

- Left handed
- Right handed

Are you a student? *

- Yes
- No

Current Degree you are working on:

- Bachelors
- Masters
- PhD or MS/PhD

Major:

Last degree:

- Bachelor
- Master

PhD

Profession:

Do you have normal vision or corrected-to-normal vision (glasses, or contact lenses)? *

Yes

No

Do you have any history of head injury (head/brain surgery, major concussion, etc)? *

Yes

No

How many hours of sleep did you have last night? *

Compared to your average sleeping time, is this *

A lot less

Less

Equal

More

A lot more

Are you feeling well today? *

Yes

No

Do you have a headache *

Yes

No

Have you had any caffeinated drinks today? *

Yes

No

If yes, how much?

Compared to an average day, is this:

- Less
- Same
- More

Do you have any comments about the study in general?

Appendix B: NASA-TLX Rating Questionnaire

Mental DemandHow much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	○	○	○	High

Physical Demand *How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	○	○	○	High

Temporal Demand *How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	○	○	○	High

Performance *How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

	1	2	3	4	5	6	7	8	9	10	
Good	○	○	○	○	○	○	○	○	○	○	Poor

Effort *How hard did you have to work (mentally and physically) to accomplish your level of performance?

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	○	○	○	High

Frustration *How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

	1	2	3	4	5	6	7	8	9	10	
Low	○	○	○	○	○	○	○	○	○	○	High

Appendix C: NASA-TLX Weights

Click on the factor that represents the more important contributor to workload for the task *

- Physical Demand: How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
- Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Click on the factor that represents the more important contributor to workload for the task *

- Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
- Temporal Demand: How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Click on the factor that represents the more important contributor to workload for the task *

- Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?
- Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Click on the factor that represents the more important contributor to workload for the task *

- Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?
- Mental Demand: How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Click on the factor that represents the more important contributor to workload for the task *

- Temporal Demand: How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
- Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Click on the factor that represents the more important contributor to workload for the task *

- Physical Demand: How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
- Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Click on the factor that represents the more important contributor to workload for the task *

- Mental Demand: How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?
- Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Click on the factor that represents the more important contributor to workload for the task *

- Physical Demand: How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
- Temporal Demand: How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Click on the factor that represents the more important contributor to workload for the task *

- Temporal Demand: How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Mental Demand: How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Click on the factor that represents the more important contributor to workload for the task *

Mental Demand: How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical Demand: How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Click on the factor that represents the more important contributor to workload for the task *

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Physical Demand: How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Click on the factor that represents the more important contributor to workload for the task *

Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Click on the factor that represents the more important contributor to workload for the task *

Temporal Demand: How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Effort: How hard did you have to work (mentally and physically) to accomplish your level of performance?

Click on the factor that represents the more important contributor to workload for the task *

Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Frustration: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Click on the factor that represents the more important contributor to workload for the task *

Performance: How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Mental Demand: How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Appendix D: Robot Perception Questionnaire

Please indicate whether you agree or disagree with the following statements.

The robots were helpful. *

	1	2	3	4	5	6	7	8	9	
Strongly Disagree	○	○	○	○	○	○	○	○	○	Strongly Agree

The robots were capable. *

	1	2	3	4	5	6	7	8	9	
Strongly Disagree	○	○	○	○	○	○	○	○	○	Strongly Agree

The robots appeared to make their own decisions. *

	1	2	3	4	5	6	7	8	9	
Strongly Disagree	○	○	○	○	○	○	○	○	○	Strongly Agree

The robots appeared to disobey my commands. *

	1	2	3	4	5	6	7	8	9	
Strongly Disagree	○	○	○	○	○	○	○	○	○	Strongly Agree

The robots were cooperative. *

	1	2	3	4	5	6	7	8	9	
Strongly Disagree	○	○	○	○	○	○	○	○	○	Strongly Agree

The robots acted like members of the team. *

	1	2	3	4	5	6	7	8	9	
Strongly Disagree	○	○	○	○	○	○	○	○	○	Strongly Agree

The robots were easy to interact with. *

	1	2	3	4	5	6	7	8	9	
Strongly Disagree	○	○	○	○	○	○	○	○	○	Strongly Agree

The robots were annoying. *

	1	2	3	4	5	6	7	8	9	
Strongly Disagree	○	○	○	○	○	○	○	○	○	Strongly Agree

Appendix E: Instructions for Human-Robot Navigation Task

You will be remotely supervising two robots that are exploring different areas. The two robots have collected information that needs to be transmitted back to the control center. Your job is to supervise the two robots which will be exploring at the same time.

The robots' job is to help you find an appropriate transmission location. Transmissions are only possible in locations with signal strength of at least 2400.

Since only the robot can determine the signal strength and only in its current position, you will have to instruct each robot using this console to explore the surface to find an appropriate transmission location.

For example, you can tell the robots "go straight," or "turn right" and they will follow your orders. You should avoid collisions with obstacles and walls. You do not want to leave a robot idle, as it may go into hibernation to save power and slow you down.

Any messages from the robot will be displayed next to the robot view. The messages from the red robot will come in red text and the messages from the blue robot will be in blue text.

You can ask the robot for the signal strength (e.g., "take a reading") and it will tell you the current signal strength. This does require the robot to stop and also takes up resources so the robot cannot be measuring signal strength at all times. Signal strengths vary from 0-3000. You need a strength of at least 2400 to transmit the data.

The task will last for five minutes total. The robots will periodically inform you of the remaining time.

The task will be considered a failure if either robot does not find a transmission location in time (i.e., before the 5 min. are up), so plan accordingly. You will now have the opportunity to see the robot and go through a training phase. I'll start by demonstrating some commands and then I'll give you a chance to interact freely with the robot.

Since we are interested in evaluating different robotic control systems for team tasks, we will ask you to perform the navigation task three times. The blocks will use different robot control systems that have, however, similar functionality. In some blocks, one of the robots may occasionally start working autonomously, without your command. You can always interrupt the robot by issuing your own commands. After the experiment, we will ask you a few questions about your view of the robot's performance in each of the three blocks.

BIBLIOGRAPHY

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2001). The Effects of Interruptions on Task Performance, Annoyance, and Anxiety in the User Interface. *Proc. of INTERACT*.

Barbosa, A. O. G., Achanccaray, D. R., & Meggiolaro, M. A. (2010). Activation of a mobile robot through a brain computer interface. *Proc. of IEEE International Conference on Robotics and Automation (ICRA).*

Blankertz, B., Dornhege, G., Krauledat, M., Kunzmann, V., Losch, F., Curio, G., & Müller, K. (2007). The Berlin Brain-Computer Interface: Machine Learning-Based Detection of User Specific Brain States. In G. Dornhege, J. d. R. Millán, T. Hinterberger, D. J. McFarland & K. Müller (Eds.), *Toward Brain-Computer Interfacing*: MIT Press.

Bunce, S., Devaraj, A., Izzetoglu, M., Onaral, B., & Pourrezaei, K. (2005). Detecting deception in the brain: a functional near-infrared spectroscopy study of neural correlates of intentional deception *Proceedings of SPIE The International Society for Optical Engineering, 5769*, 24-32.

Bunce, S. C., Izzetoglu, M., Izzetoglu, K., Onaral, B., & Pourrezaei, K. (2006). Functional Near Infrared Spectroscopy: An Emerging Neuroimaging Modality. *IEEE Engineering in Medicine and Biology Magazine, Special issue on Clinical Neuroengineering, 25*(4), 54 - 62.

Card, S., Moran, T., & Newell, A. (1983). *The Psychology of Human Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., Li, C., Murray, T., Ovetsky, Y., Pidikiti, D., Thomas, R. (1988). A novel method for fast imaging of brain function, non-invasively, with light. *Optics Express, 10*(2), 411-423.

Chen, D., Hart, J., & Vertegaal, R. (2008). Towards a Physiological Model of User Interruptability. *Proc of INTERACT'07* (pp. 439-451).

Chenier, F., & Sawan, M. (2007). A New Brain Imaging Device Based on fNIRS. *Proc. of Biomedical Circuits and Systems Conference*, *BIOCAS 2007*. IEEE.

Coyle, S., Ward, T., & Markham, C. (2004). Physiological noise in near-infrared spectroscopy: implications for optical brain computer interfacing. *Proc. of IEEE Engineering in Medicine and Biology Society IEMBS '04.*

Cutrell, E., & Tan, D. (2008). BCI for passive input in HCI. *Proc. of CHI 2008 Workshop on Brain-Computer Interfaces for HCI and Games.*

Czerwinski, M., Cutrell, E., & Horvitz, E. (2000). Instant Messaging: Effects of Relevance and Timing. *Proc. of Human-Computer Interaction.*

Devaraj, A., Izzetoglu, M., Izzetoglu, K., & Onaral, B. (2004). Motion Artifact Removal for fNIR Spectroscopy for Real World Application Areas. *Proceedings of the SPIE International Society for Optical Engineering, 5588*, 224-229.

Ehlis, A.-C., Bähne, C. G., Jacob, C. P., Herrmann, M. J., & Fallgatter, A. J. (2008). Reduced lateral prefrontal activation in adult patients with attention-deficit/hyperactivity disorder (ADHD) during a working memory task: A functional near-infrared spectroscopy (fNIRS) study. *Journal of Psychiatric Research, 42*(13), 1060-1067.

Escolano, C., Murguialday, A. R., Matuz, T., Birbaumer, N., & Minguez, J. (2010). A telepresence robotic system operated with a P300-based brain-computer interface: Initial tests with ALS patients. *Proc. of IEEE Engineering in Medicine and Biology Society (EMBC).*

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interact. Comput., 21*(1-2), 133-145.

Fink, A., Benedek, M., Grabner, R. H., Staudt, B., & Neubauer, A. C. (2007). Creativity meets neuroscience: Experimental tasks for the neuroscientific study of creative thinking. *Methods, 42*(1), 68-76.

Fogarty, J., Hudson, S. E., & Lai, J. (2004). Examining the robustness of sensor-based statistical models of human interruptibility. *Proc. ACM Conference on Human Factors in Computing Systems'04.*

Folley, B. S., & Park, S. (2005). Verbal creativity and schizotypal personality in relation to prefrontal hemispheric laterality: A behavioral and near-infrared optical imaging study. *Schizophrenia Research, 80*(2-3), 271-282.

Girouard, A., Solovey, E. T., Hirshfield, L. M., Chauncey, K., Sassaroli, A., Fantini, S., & Jacob, R. J. K. (2009). Distinguishing Difficulty Levels with Non-invasive Brain Activity Measurements. *Proc. of INTERACT 2009.*

Girouard, A., Solovey, E. T., & Jacob, J. K. (2010). Designing a Passive Brain Computer Interface using Real Time Classification of Functional Near-Infrared Spectroscopy. *Intl J of Autonomous and Adaptive Communications Systems* (2010).

Grimes, D., Tan, D. S., Hudson, S., Shenoy, P., & Rao, R. (2008). Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph. *Proc. ACM Conference on Human Factors in Computing Systems.*

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explor. Newsl., 11*(1), 10-18.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theorical research. In P. Hancock, Meshkati, N. (Ed.), *Human Mental Workload* (pp. pp 139 - 183). Amsterdam.

Herrmann, M. J., Huter, T., Plichta, M. M., Ehlis, A.-C., Alpers, G. W., Mühlberger, A., & Fallgatter, A. J. (2008). Enhancement of activity of the primary visual cortex during processing of emotional stimuli as measured with event-related functional near-infrared spectroscopy and event-related potentials. *Human Brain Mapping, 29*(1), 28-35.

Hirshfield, L. M., Girouard, A., Solovey, E. T., Jacob, R. J. K., Sassaroli, A., Tong, Y., & Fantini, S. (2007). Human-Computer Interaction and Brain Measurement Using Functional Near-Infrared Spectroscopy. *Poster Paper Proc. ACM UIST 2007 Symposium on User Interface Software and Technology.*

Hirshfield, L. M., Gulotta, R., Hirshfield, S., Hincks, S., Russell, M., Ward, R., Williams, T., Jacob, R. (2011). This is your brain on interfaces: enhancing usability testing with functional near-infrared spectroscopy. *Proc. ACM Conference on Human Factors in Computing Systems'11.*

Hirshfield, L. M., Solovey, E. T., Girouard, A., Kebinger, J., Sassaroli, A., Tong, Y., . . . Jacob, R. J. K. (2009). Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload with Functional Near Infrared Spectroscopy. *Proc CHI.*

Hornof, A. J., Zhang, Y., & Halverson, T. (2010). Knowing where and when to look in a time-critical multimodal dual task. *Proc CHI '10.*

Hoshi, Y. (2009). Near-Infrared Spectroscopy for Studying Higher Cognition *Neural Correlates of Thinking* (pp. 83-93).

Hudson, S., Fogarty, J., Atkeson, C., Avrahami, D., Forlizzi, J., Kiesler, S., Lee, J., Yang, J. (2003). Predicting human interruptibility with sensors: a Wizard of Oz feasibility study. *Proc CHI '03.*

Huppert, T., & Boas, D. A. (2005). Hemodynamic Evoked Response NIRS data analysis GUI Program User's Guide, 2009, from http://www.nmr.mgh.harvard.edu/PMI/resources/homer/HomER_UsersGuide_05-07-28.pdf

Iqbal, S. T., Adamczyk, P. D., Zheng, X. S., & Bailey, B. P. (2005). Towards an index of opportunity: understanding changes in mental workload during task execution. *Proc CHI '05.*

Iqbal, S. T., & Bailey, B. P. (2005). Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption. *Proc. ACM Conference on Human Factors in Computing Systems '05 extended abstracts.*

Iqbal, S. T., Zheng, X. S., & Bailey, B. P. (2004). Task-evoked pupillary response to mental workload in human-computer interaction. *Proc. ACM Conference on Human Factors in Computing Systems '04 extended abstracts.*

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

Izzetoglu, M., Devaraj, A., Bunce, S., & Onaral, B. (2005). Motion artifact cancellation in NIR spectroscopy using Wiener filtering. *Biomedical Engineering, IEEE Transactions on, 52*(5), 934-938.

Izzetoglu, M., Izzetoglu, K., Bunce, S., Onaral, B., & Pourrezaei, K. (2005). Functional Near-Infrared Neuroimaging. *IEEE Trans. on Neural Systems and Rehabilitation Engineering, 13*(2), 153-159.

Jackson, M. M., & Mappus, R. (2010). Applications for Brain-Computer Interfaces. In D. Tan & A. Nijholt (Eds.), *Brain Computer Interfaces: Applying our Minds to Human-Computer Interaction* (pp. 89-104): Springer.

Keirn, Z. A., & Aunon, J. I. (1990). A new mode of communication between man and his surroundings. *Biomedical Engineering, IEEE Transactions on, 37*(12), 1209.

Kennedy, P. R., Bakay, R. A. E., Moore, M. M., Adams, K., & Goldwaithe, J. (2000). Direct control of a computer from the human central nervous system. *IEEE Transactions on Rehabilitation Engineering, 8*(2), 198.

Koechlin, E., Basso, G., Pietrini, P., Panzer, S. & Grafman, J. (1999). The role of the anterior prefrontal cortex in human cognition. *Nature, 399,* 148-151.

Koechlin, E., Corrado, G., Pietrini, P., & Grafman, J. (2000). Dissociating the role of the medial and lateral anterior prefrontal cortex in human planning. *PNAS, 97*, 7651-7656.

Kuikkaniemi, K., Laitinen, T., Turpeinen, M., Saari, T., Kosunen, I., & Ravaja, N. (2010). The influence of implicit and explicit biofeedback in first-person shooter games. *Proc CHI.*

Lee, J. C., & Tan, D. (2006). Using a Low-Cost Electroencephalograph for Task Classification in HCI Research. *Proc. ACM Symposium on User Interface Software and Technology 2006.*

Mandryk, R., Atkins, M., & Inkpen, K. (2006). A continuous and objective evaluation of emotional experience with interactive play environments. *Proceedings of the SIGCHI conference on Human factors in computing systems*.

Mandryk, R. L., & Inkpen, K. M. (2004). Physiological indicators for the evaluation of co-located collaborative play. *Proc. CSCW '04.*

Mappus, R. L., Venkatesh, G. R., Shastry, C., Israeli, A., & Jackson, M. M. (2009). An fNIR Based BMI for Letter Construction Using Continuous Control. *Proc. ACM Conference on Human Factors in Computing Systems 2009 Extended Abstracts.*

Matthews, F., Pearlmutter, B. A., Ward, T. E., Soraghan, C., & Markham, C. (2008). Hemodynamics for Brain-Computer Interfaces. *Signal Processing Magazine, IEEE, 25*(1), 87-94.

McFarlane, D. (2002). Comparison of four primary methods for coordinating the interruption of people in human-computer interaction. *Hum.-Comput. Interact., 17*(1), 63-139.

Millán, J. d. R., Renkens, F., Mouriño, J., & Gerstner, W. (2004). Brain-actuated interaction. *Artif. Intell., 159*(1-2), 241-259.

Miyata, Y., & Norman, D. (1986a). The Control of Multiple Activities. In D. Norman & S. W. Draper (Eds.), *User Centered System Design: New Perspectives on* Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates.

Miyata, Y., & Norman, D. (1986b). Psychological Issues in Support of Multiple Activities. In D. A. Norman & S. W. Draper (Eds.), *User Centered System Design: New Perspectives on Human-Computer Interaction* (pp. 265-284). Hillsdale, NJ: Lawrence Erlbaum Associates.

Monk, C., A. , Boehm-Davis, D., A. , & Trafton, J. G. (2002). The attentional costs of interrupting task performance at various stages. *Proc. HFES Annual Mtg.*

Moore, M., Kennedy, P., Mynatt, E., & Mankoff, J. (2001). Nudge and Shove: Frequency Thresholding for Navigation in Direct Brain-computer Interfaces *Proc. ACM CHI 2001 Human Factors in Computing Systems Conference Extended Abstracts* (pp. 361-362): ACM Press.

Moore, M. M. (2003). Real-world applications for brain-computer interface technology. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11*(2), 162.

Moore, M. M., & Kennedy, P. R. (2000). Human factors issues in the neural signals direct brain-computer interfaces. *Proceedings of the fourth international ACM conference on Assistive technologies*.

Morioka, S., Yamada, M., & Komori, T. (2008). Frontal Lobe Activity during the Performance of Spatial Tasks: fNIRS Study. *J Phys Ther Sci, 20*, 135-139.

Nacke, L. E., Kalyn, M., Lough, C., & Mandryk, R. L. (2011). Biofeedback game design: using direct and indirect physiological control to enhance game interaction. *Proc. ACM Conference on Human Factors in Computing Systems'11*.

O'Hara, K., Sellen, A., & Harper, R. (2011). Embodiment in brain-computer interaction. *Proc. ACM Conference on Human Factors in Computing Systems 2011,* Vancouver, BC, Canada.

Okumura, M., & Zhiwei, L. (2007, 15-18 Dec. 2007). On NIRS-based brain-robot interface. *Proc. IEEE International Conference on Robotics and Biomimetics, ROBIO'07.*

Parasuraman, R., Sheridan, T., & Wickens, C. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A, 30*(3), 286-297.

Perrin, X., Chavarriaga, R., Colas, F., Siegwart, R., & Millán, J. d. R. Brain-coupled interaction for semi-autonomous navigation of an assistive robot. *Robot. Auton. Syst., 58*(12), 1246-1255. doi: 10.1016/j.robot.2010.05.010

Pfurtscheller, G., Flotzinger, D., & Kalcher, J. (1993). Brain-computer interface: a new communication device for handicapped persons. *J. Microcomput. Appl., 16*(3), 293-299.

Picard, R. W., Vyzas, E., & Healey, J. (2001). Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. *IEEE Trans. Pattern Anal. Mach. Intell., 23*(10), 1175-1191.

Platt, J. (1999). Fast training of support vector machines using sequential minimal optimization *Advances in kernel methods: support vector learning* (pp. 185-208): MIT Press.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology, 40*(1-2), 187-195.

Ramnani, N., & Owen, A. M. (2004). Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat Rev Neurosci, 5*(3), 184.

Salvucci, D. D., & Bogunovich, P. (2010). Multitasking and monotasking: the effects of mental workload on deferred task interruptions. *Proc CHI.*

Sasse, A., Johnson, C., & Mcfarlane, D., C. (1999). Coordinating the Interruption of People in Human-Computer Interaction. *Proc. INTERACT.*

Schalk, G., McFarland, D. J., Hinterberger, T., Birbaumer, N., & Wolpaw, J. R. (2004). BCI2000: a general-purpose brain-computer interface (BCI) system. *Biomedical Engineering, IEEE Transactions on, 51*(6), 1034.

Schermerhorn, P., & Scheutz, M. (2009). Dynamic robot autonomy: investigating the effects of robot decision-making in a human-robot team task. *Proc ICMI.*

Schermerhorn, P., & Scheutz, M. (2011). Disentangling the Effects of Robot Affect, Embodiment, and Autonomy on Human Team Members in a Mixed-Initiative Task. *Proc. 2011 International Conference on Advances in Computer-Human Interactions.*

Scheutz, M., Schermerhorn, P., & Kramer, J. (2006). The utility of affect expression in natural language interactions in joint human-robot tasks. *Proc HRI.*

Scheutz, M., Schermerhorn, P., Kramer, J., & Anderson, D. (2007). First steps toward natural human-like HRI. *Auton. Robots, 22*(4), 411-423.

Shenoy, P., & Tan, D. S. (2008). Human-aided computing: utilizing implicit human processing to classify images. *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, Florence, Italy.

Sitaram, R., Zhang, H., Guan, C., Thulasidas, M., Hoshi, Y., Ishikawa, A., Shimizu, K., Birbaumer, N. (2007). Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *NEUROIMAGE, 34*(4), 1416-1427.

Solovey, E. T., Chauncey, K., Lalooses, F., Parasi, M., Weaver, D., Scheutz, M., Schermerhorn, P., Sassaroli, A., Fantini, S., Girouard, A., Jacob, R. J. K. (2011). Sensing Cognitive Multitasking for a Brain-Based Adaptive User Interface. *Proc. ACM Conference on Human Factors in Computing Systems'11.*

Solovey, E. T., Girouard, A., Chauncey, K., Hirshfield, L. M., Sassaroli, A., Zheng, F., Fantini, S., Jacob, R. J. K. (2009). Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines. *Proc ACM SYMPOSIUM ON USER INTERFACE SOFTWARE AND TECHNOLOGY'09.*

Solovey, E. T., Schermerhorn, P., Scheutz, M., Sassaroli, A., Fantini, S., & Jacob, R. J. K. (2012). Brainput: Enhancing Interactive Systems with Streaming fNIRS Brain Input. Proc. *ACM CHI'12 Conference on Human Factors in Computing Systems.*

Starner, T., Schiele, B., & Pentland, A. (1998). Visual Contextual Awareness in Wearable Computing. *Proc ISWC.*

Tan, D., & Nijholt, A. (2010). Brain-Computer Interfaces and Human-Computer Interaction. In D. S. Tan & A. Nijholt (Eds.), *Brain-Computer Interaction:*

Applying our Minds to Human-Computer Interaction. (pp. 3-19): Springer London.

Tsubone, T., Tsutsui, K., Muroga, T., & Wada, Y. (2008). Estimation of Force Motor Command to Control Robot by NIRS-Based BCI. In I. Masumi, D. Kenji, M. Hiroyuki & Y. Takeshi (Eds.), *Neural Information Processing* (pp. 969-978): Springer-Verlag.

Villringer, A., & Chance, B. (1997). Non-Invasive Optical Spectroscopy and Imaging of Human Brain Function. *Trends in Neuroscience, 20*, 435-442.

Villringer, A., Planck, J., Hock, C., Schleinkofer, L., & Dirnagl, U. (1993). Near Infrared Spectroscopy (NIRS): A New Tool to Study Hemodynamic Changes During Activation of Brain Function in Human Adults. *Neurosci. Lett., 154*, 101-104.

Wolpaw, J. R., McFarland, D. J., Neat, G. W., & Forneris, C. A. (1991). An EEG-based brain-computer interface for cursor control. *Electroencephalography and Clinical Neurophysiology, 78*(3), 252.

Yuksel, B. F., Donnerer, M., Tompkin, J., & Steed, A. (2010). A novel brain-computer interface using a multi-touch surface. *Proc CHI.*