Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines

Erin Treacy Solovey*, Audrey Girouard*, Krysta Chauncey*, Leanne M. Hirshfield*, Angelo Sassaroli+, Feng Zheng+, Sergio Fantini+, Robert J. K. Jacob*

*Computer Science Department
Tufts University
Medford, MA 02155, USA

+Biomedical Engineering Department
Tufts University
Medford, MA 02155, USA

{erin.solovey, audrey.girouard, krysta.chauncey, leanne.hirshfield, angelo.sassaroli, feng.zheng, sergio.fantini, robert.jacob}@tufts.edu

ABSTRACT

Because functional near-infrared spectroscopy (fNIRS) eases many of the restrictions of other brain sensors, it has potential to open up new possibilities for HCI research. From our experience using fNIRS technology for HCI, we identify several considerations and provide guidelines for using fNIRS in realistic HCI laboratory settings. We empirically examine 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 study, 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 hope to facilitate further adoption of fNIRS brain sensing technology in HCI research.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

General terms: Human Factors

Keywords: functional near-infrared spectroscopy, fNIRS, brain-computer interface, human cognition, BCI

INTRODUCTION

Brain sensing and imaging techniques, primarily developed for clinical settings, have been powerful tools for understanding brain function as well as for diagnosing brain injuries or disorders. More recently, these devices have found uses outside of hospital and laboratory settings, and human-computer interaction (HCI) researchers have begun to employ them to understand more about the user's cognitive state relative to the task at hand [2, 10]. This has been made possible due to technological advances and lower costs associated with the devices.

However, to be valuable in HCI, the sensors should collect useful information while ideally allowing normal interaction with the computer. In this regard, functional near-infrared spectroscopy (fNIRS) is well-suited for use in HCI (Figure 1) and there have been recent studies using fNIRS in HCI to distinguish game difficulty levels [9], to measure mental workload [13], and for letter drawing [19]. These studies have revealed that fNIRS sensors show potential for opening up new possibilities in HCI research.

From our experience with fNIRS [9, 12, 13], we felt it was important to identify and examine empirically considerations necessary for appropriate use of fNIRS in realistic HCI laboratory settings. Common behaviors such as head and eye movements currently are restricted during fNIRS experiments. Based on the results of our study, we provide guidelines clarifying which behavioral conditions need to be controlled, avoided, or corrected when using fNIRS, and which factors are not problematic. With this information, researchers can better take advantage of fNIRS brain sensing technology.

Figure 1: A participant wearing one fNIRS probe.

Brain Sensing in Human-Computer Interaction

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 [13, 18]. 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.

UIST'09, October 4–7, 2008, Victoria, BC, Canada.

In many traditional brain-computer interfaces, which often were designed for disabled users, the user is trained to control his or her brain activity and this brain signal is used explicitly as the primary input to the system [21]. More 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 [13, 18]. 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.

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.

While we intend to use fNIRS to pick up psychophysiological data, we do not expect that the participant is 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 [14]. Thus, we would like to 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.

FNIRS AND OTHER BRAIN SENSING TECHNOLOGIES

Because most brain imaging and sensing devices were developed for clinical settings, they often have characteristics that make them less suitable for use in realistic HCI settings. For example, although functional magnetic resonance imaging (fMRI) is effective for functional brain imaging, it is susceptible to motion artifacts, and even slight movement (more than 3mm) will corrupt the image. In addition, because of the magnetic field, there can be no metal objects, making computer usage impractical. The most common technology used for brain measurement in HCI is electroencephalography (EEG) because it is non-invasive, portable, and relatively inexpensive compared with other brain imaging devices [18]. Some obstacles with using EEG for HCI are that it is susceptible to artifacts from eye and facial movements, requires gel in the participant's hair, takes some time to set up properly, and is subject to noise from nearby electronic devices.

Recently, fNIRS has been used in HCI because it has many characteristics that make it suitable for use outside of clinical settings [9, 13, 19]. Benefits include ease of use, short setup time, and portability, making it a promising tool for HCI researchers. In addition, there are no technical restrictions for using EEG and fNIRS together [12], and the two technologies could complement one another.

fNIRS provides a measure of blood oxygen concentration, indicative of brain activity when measured on the head [25]. Near-infrared light is sent into the forehead where it probes the tissues of the cortex up to depths of 1-3cm. Oxygenated and deoxygenated hemoglobin (respectively [HbO] and [Hb]) are the main absorbers of light at these wavelengths, and thus the diffusely reflected light that is picked up by the detector correlates with the concentration of oxygen in the blood. 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.

There are many possible placements of fNIRS probes, allowing the study of multiple brain regions. The most common placements are on the motor cortex [24], and the prefrontal cortex (PFC) [6, 19], although other regions have also been explored [11]. We built on past experiments and chose to study the anterior prefrontal cortex (aPFC), an active region that deals with high-level processing [23], such as working memory, planning, problem solving, memory retrieval and attention. We believe these signals to be of great potential to HCI, rather than measurements at the motor or visual cortex. Thus, our considerations below are intended for researchers measuring the aPFC, as the impact of the human behavior and typical interactions will vary depending on the measured region of the brain. However, we expect our results to be valid for other experimental setups and contexts that use the prefrontal cortex area.

FNIRS CONSIDERATIONS

With the introduction of any new technology, there are considerations that should be made for its proper use. For this reason, we use our previous 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. In particular, we identify below potential sources of noise and artifacts in the fNIRS signal when used in typical HCI laboratory settings.

Head Movement

Several fNIRS researchers have brought attention to motion artifacts in fNIRS sensor data, particularly those from head movement [5, 20]. Matthews et al. [20] 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" [14]. 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.

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" [4]. 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 [3]. However, as there is little empirical evidence of this phenomenon, we will examine it further in the experiment. 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 [16]. However, fNIRS is less sensitive to muscle tension and researchers have reported that no artifact is produced in nearby areas of the brain [16]. 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.

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 [4]. Chenier and Sawan [3] 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, we feel that future fNIRS sensors will be embedded in a helmet or hat that properly isolates them from this source of noise. Therefore, in this paper, we do 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.

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 [22]. 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.

Respiration and Heartbeat

The fNIRS signals picks up artifacts from respiration and heart beat, by definition, as it measures blood flow and oxygenation [4, 20]. These systemic noise sources can be removed using known filtering techniques. For a discussion of the many filtering techniques, see Matthew et al. [20] and Coyle et al. [4].

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" [7]. 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 [9]. In this study, we examine physical motions that are common in HCI settings, typing and mouse clicking, to determine whether they are problematic when using fNIRS.

Slow Hemodynamic Response

The slow hemodynamic changes measured by fNIRS occur in a time span of 6-8 seconds [1]. This is important when designing interfaces based on fNIRS sensor data, as the interface would have to respond in this time scale. While

	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 difference 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 2: Letters A, B, C, and D show the conditions tested. The numbered questions indicate the comparisons between the conditions done in the analysis.

the possibility of using event-related fNIRS has been explored [11], most studies take advantage of the slow response to measure short term cognitive state, instead of instantaneous ones.

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 as described above: (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 (see Figure 2).

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.

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.

Apparatus

We used a multichannel frequency domain OxiplexTS from ISS Inc. (Champaign, IL) for data acquisition. We used two probes on the forehead to measure the two hemispheres of the anterior prefrontal cortex (see Figure 3). The source-detector distances are 1.5, 2, 2.5, 3cm respectively. Each distance measures a different depth in the cortex. Each source emits two light wavelengths (690nm and 830nm) to pick up and differentiate between [HbO] and [Hb]. The sampling rate was 6.25Hz. 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 [9, 13]. 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.

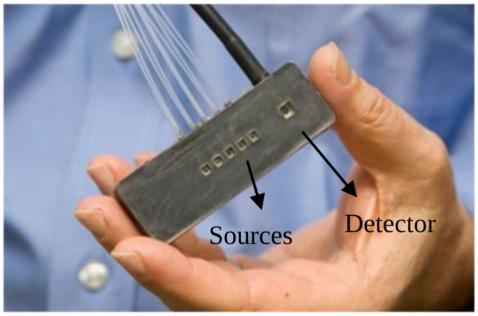

Figure 3: A picture of the right probe. A probe includes a detector and light sources.

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.

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 [6, 13]. Many studies have successfully shown discrimination of two (or more) states, and we believe our results will generalize to those as well.

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.

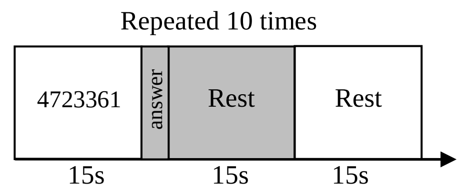

Figure 4: Experiment 0 (No artifacts). The white areas represent the two conditions analyzed. The answer period’s length was variable.

A trial contained one 15s condition with the cognitive task, followed by a 15s rest period to allow the participant’s brain to return to a rested state. In addition, there was a 15s condition without the cognitive task in which the participant was essentially at rest (see Figure 4). These conditions were counterbalanced so that sometimes participants started with the cognitive task, and sometimes they started without the cognitive task.

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. [9]. We used a non-recursive time-domain band-pass filter, keeping frequencies between 0.01-0.5 Hz [8]. The data was then transformed to obtain oxy- ([HbO]) and deoxy-hemoglobin ([Hb]) concentration values, using the modified Beer-Lambert law [25]. 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*.

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 paper 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 will identify differences within each participant, and determine if they are significant across participants. This is Comparison 2.1 in Figure 2. 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.

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. We believe we can go forward with the rest of the analysis because of this positive result.

Experiment 1: Keyboard Input

The keyboard and mouse are the most common input devices for modern computers. We tested keyboard input in Exp. 1 and mouse input in Exp. 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 PFC.

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 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.

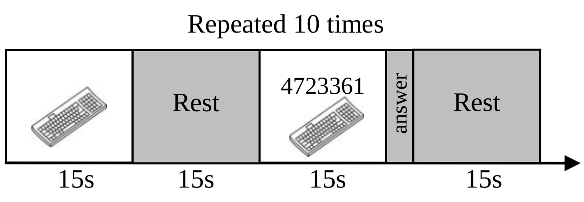

Figure 5: Experiment 1 (Keyboard Input). The white areas represent the two conditions analyzed in the experiment.

The protocol was analogous to Experiment 0. The main difference is that in both tasks, the participant was also typing randomly as described above (see Figure 5).

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 2. 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.

Results. Comparison 1 showed significance for *Typing x Time Course* 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 course. We did not observe any significant interaction that included *Typin*g 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.

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 confirms our hypothesis and validates 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.

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.

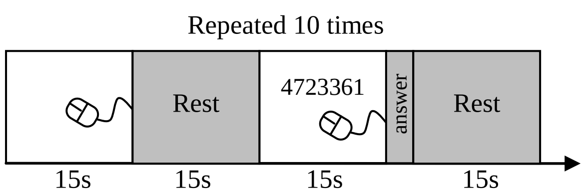

Figure 6: Experiment 2 (Mouse Input).

The procedure was identical to Experiment 1, except that the typing was replaced with mouse clicking (see Figure 6). We analyzed the data using the same comparisons as in Experiment 1, substituting mouse input for keyboard input.

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. 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.

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 aPFC. 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.

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 hypothesize 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 (see Figure 7). We analyzed the data using the same comparisons as in Experiment 1 and 2, substituting head movement for keyboard or mouse input.

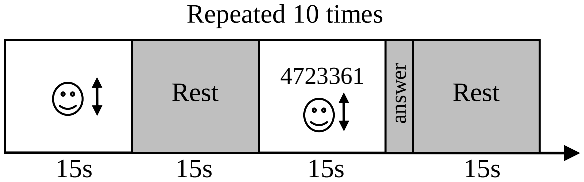

Figure 7: Experiment 3 (Head Movement).

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 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 Course* ($F(6, 54)= 7.455$, $p= 0.002$), 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 Course* is significant ($F(18, 162)= 3.915$, $p= 0.048$). With head movements, there is a difference between rest and the cognitive task.

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.

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 head movement (see Figure 8). We analyzed the data using the same comparisons as in the other experiments, substituting frowning motion for keyboard or mouse input, or head movement.

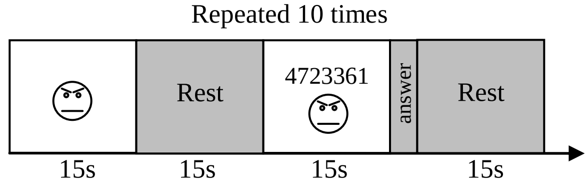

Figure 8. Experiment 4 (Facial Movement).

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 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.

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 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.

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.

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 paper, and summarized in Table 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 [20]. Methods include adaptive finite impulse response (FIR) filtering, Weiner filtering [5, 17], adaptive filtering [5] and principal component analysis [15, 20, 24]. Matthews et al. [20] 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.

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.

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 we recommend fNIRS as a valuable and effective input technology.

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

Considerations	**Result**	**Reference**	**Correction Methods**
Forehead movement	✗	Exp 4	
Major head movement	✗	Exp 3	Use chin rest
Minor head movement	**C**	Exp 3, [20]	Filter
Respiration and Heartbeat	**C**	[4, 20]	Filter
Mouse Clicking	✓	Exp 2	Collect signal during a clicking only task
Typing	✓	Exp 1	
Ambient Light	**C**	[3]	Wear isolating cap
Hemodynamic Response	✓	[1]	Expect 6-8s response
Ambient Noise	**C**	[22]	Minimize external noise
Eye Movement and Blinking	✓	[16]	

ACKNOWLEDGMENTS

The authors would like to thank the HCI research group at Tufts University; Michel Beaudoin-Lafon, Wendy Mackay, and the In|Situ| research group; and Desney Tan and Dan Morris at Microsoft Research. We thank the NSF (Grant Nos. IIS-0713506 and IIS-0414389), the US Air Force Research Laboratory, NSERC, the US Army NSRDEC for support of this research. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of these organizations.

REFERENCES

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

2. Chen, D., Hart, J. and Vertegaal, R. Towards a Physiological Model of User Interruptability. in INTERACT'07, 2008, 439-451.
3. Chenier, F. and Sawan, M., A New Brain Imaging Device Based on fNIRS. in BIOCAS 2007, (2007), 1-4.
4. Coyle, S., Ward, T. and Markham, C., Physiological Noise in Near-infrared Spectroscopy: Implications for Optical Brain Computer Interfacing. in Proc. EMBS, (2004).
5. Devaraj, A., Izzetoglu, M., Izzetoglu, K. and Onaral, B. Motion Artifact Removal for fNIR Spectroscopy for Real World Application Areas. Proc. SPIE, 5588. 2004, 224-229.
6. Ehlis, A.-C., Bähne, C.G., Jacob, C.P., Herrmann, M.J. and Fallgatter, A.J. 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). 2008, 1060-1067.
7. Fink, A., Benedek, M., Grabner, R.H., Staudt, B. and Neubauer, A.C. Creativity meets neuroscience: Experimental tasks for the neuroscientific study of creative thinking. Methods, 42 (1). 2007, 68-76.
8. Folley, B.S. and Park, S. Verbal creativity and schizotypal personality in relation to prefrontal hemispheric laterality: A behavioral and near-infrared optical imaging study. Schizophrenia Research, 80 (2-3). 2005, 271-282.
9. Girouard, A., Solovey, E.T., Hirshfield, L.M., Chauncey, K., Sassaroli, A., Fantini, S. and Jacob, R.J.K., Distinguishing Difficulty Levels with Non-invasive Brain Activity Measurements. in INTERACT 2009, (2009), Springer.
10. Grimes, D., Tan, D.S., Hudson, S.E., Shenoy, P. and Rao, R.P.N. Feasibility and pragmatics of classifying working memory load with an electroencephalograph Proc. CHI'08, ACM, 2008.
11. Herrmann, M.J., Huter, T., Plichta, M.M., Ehlis, A.-C., Alpers, G.W., Mühlberger, A. and Fallgatter, A.J. 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). 2008, 28-35.
12. Hirshfield, L., Chauncey, K., Gulotta, R., Girouard, A., Solovey, E.T., Jacob, R., Sassaroli, A. and Fantini, S., Combining Electroencephalograph and Near Infrared Spectroscopy to Explore Users' Mental Workload. in HCI International 2009 (2009).
13. Hirshfield, L.M., Solovey, E.T., Girouard, A., Kebinger, J., Sassaroli, A., Tong, Y., Fantini, S. and Jacob, R.J.K., Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload with Functional Near Infrared Spectroscopy. in Proc. CHI'09, (2009).
14. Hoshi, Y. Near-Infrared Spectroscopy for Studying Higher Cognition. in Neural Correlates of Thinking, 2009, 83-93.
15. Huppert, T. and 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
16. Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K. and Chance, B. Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. IJHCI, 17 (2). 2004, 211-231.
17. Izzetoglu, M., Devaraj, A., Bunce, S. and Onaral, B. Motion Artifact Cancellation in NIR spectroscopy Using Wiener Filtering. IEEE Transactions on Biomedical Engineering, 52 (5). 2005, 934-938.
18. Lee, J.C. and Tan, D.S., Using a Low-Cost Electroencephalograph for Task Classification in HCI Research. in Proc. ACM Symposium on User Interface Software and Technology 2006, (2006).
19. Mappus, R.L., Venkatesh, G.R., Shastry, C., Israeli, A. and Jackson, M.M., An fNIR Based BMI for Letter Construction Using Continuous Control. in Proc. CHI 2009 Extended Abstracts, (2009).
20. Matthews, F., Pearlmutter, B.A., Ward, T.E., Soraghan, C. and Markham, C. Hemodynamics for Brain-Computer Interfaces. Signal Processing Magazine, IEEE, 25 (1). 2008, 87-94.
21. Millán, J.d.R., Renkens, F., Mouriño, J. and Gerstner, W. Brain-actuated interaction. Artif. Intell., 159 (1-2). 2004, 241-259.
22. Morioka, S., Yamada, M. and Komori, T. Frontal Lobe Activity during the Performance of Spatial Tasks: fNIRS Study. J Phys Ther Sci, 20. 2008, 135-139.
23. Ramnani, N. and Owen, A.M. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. Nat Rev Neurosci, 5 (3). 2004, 184.
24. Sitaram, R., Zhang, H., Guan, C., Thulasidas, M., Hoshi, Y., Ishikawa, A., Shimizu, K. and Birbaumer, N. Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. NEUROIMAGE, 34 (4). 2007, 1416-1427.
25. Villringer, A. and Chance, B. Non-Invasive Optical Spectroscopy and Imaging of Human Brain Function. Trends in Neuroscience, 20. 1997, 435-442.