Combining Electroencephalograph and Functional Near Infrared Spectroscopy to Explore Users' Mental Workload

Leanne M. Hirshfield*, Krysta Chauncey*, Rebecca Gulotta*, Audrey Girouard*, Erin T. Solovey*, Robert J.K. Jacob*, Angelo Sassaroli^, Sergio Fantini^

*Human Computer Interaction Lab and ^Biomedical Engineering Department
Tufts University
Medford, MA 02155

{Leanne.Hirshfield, Krysta.Chauncey, Rebecca.Gulotta, Audrey.Girouard, Erin.Solovey, Robert.Jacob, Angelo.Sassaroli, Sergio.Fantini}@tufts.edu

Abstract. We discuss the physiological metrics that can be measured with electroencephalography (EEG) and functional near infrared spectroscopy (fNIRs). We address the functional and practical limitations of each device, and technical issues to be mindful of when combining the devices. We also present machine learning methods that can be used on concurrent recordings of EEG and fNIRs data. We discuss an experiment that combines fNIRs and EEG to measure a range of user states that are of interest in HCI. While our fNIRS machine learning results showed promise for the measurement of workload states in HCI, our EEG results indicate that more research must be done in order to combine these two devices in practice.

Keywords: fNIRs, EEG, near infrared spectroscopy, workload

1 Introduction

Current research in human computer interaction (HCI) explores the measurement of computer users' brain activity in an attempt to increase the effectiveness of usability testing and to create adaptive user interfaces. By measuring users' mental states objectively, and in real time, usability experts use information about users' mental workload (WL) as an additional metric during usability studies. Designers of adaptive systems hope to use this information as passive input, providing the system with reliable, real time information about the user's state, so that the system can adapt and make the human-computer interaction as flawless as possible.

Electroencephalography (EEG) and functional near-infrared spectroscopy (fNIRs) are popular, non-invasive, brain imaging techniques. Unlike other brain devices which require subjects to lie in restricted positions (fMRI), or to drink hazardous materials (PET), EEG and fNIRs can non-invasively measure users' brain activity in real working conditions [1]. This makes EEG and fNIRs appropriate choices for brain measurement in HCI. The majority of brain measurement research in HCI uses EEG to measure users' states while a smaller body of research uses fNIRs. Only a handful of researchers have explored the combination of the two devices to measure brain function [2-5]. This is unfortunate, as combining the two devices can provide complementary information about different physiological responses to brain function and compensate for practical and functional limitations of each measurement technology.

EEG and fNIRS measure different physiological responses to mental state changes. EEG measures the electrical potentials caused by neurons firing during brain activity. fNIRs measures blood volume and oxygenation changes, reflecting hemodynamic responses to brain activity. Therefore, combining EEG and fNIRS can provide information about both the neural activations and the subsequent oxygenation and blood flow changes in the brain. Additionally, each device has functional and practical limitations that make it difficult to acquire a range of user states in real world settings. We discuss these benefits and pitfalls in detail in the next section, and we show how combining the devices allows us to compensate for the measurement pitfalls that either device has on its own.

This paper has three primary contributions to the HCI realm. First, the paper is intended as a useful guide for researchers interested in combining fNIRs and EEG. We discuss the physiological metrics that EEG and fNIRs can measure, the functional and practical limitations of each device, and technical issues to be mindful of when combining the devices. Second, we present simple machine learning methods that can be used on concurrent recordings of EEG and fNIRs data. Third, we describe results from an experiment using concurrent recordings of EEG and fNIRs.

The rest of this paper proceeds as follows: First, we discuss background literature about EEG and fNIRs. Next we present an experiment designed to elicit varying levels of working memory load on subjects. We then describe the machine learning techniques that we implemented to analyze the experiment data. After describing the results of our data analysis, we discuss our concluding thoughts and extensions for future work.

2 Background and Relevant Literature

2.1 Electroencephalography

EEG is the most studied non-invasive brain imaging device due to its fine temporal resolution, ease of use, and low set-up cost. EEG uses electrodes placed on the scalp to measure and record the electrical potential caused by neurons firing in the brain during brain activity. These measurements vary predictably in response to changing levels of cognitive stimuli [6]. Additionally, EEG benefits from high temporal resolution, enabling it to measure changes in cognitive activity on the millisecond scale. Therefore, EEG measurements are continuously reflective of a participant's cognitive states [7]. However, there are some significant limitations to EEG signals. EEG has low spatial resolution (about 10 cm) [7] , making it difficult to make precise measurements about the area of the brain being activated. EEG is also susceptible to motion artifacts, such as blinking and movement. These actions create artifacts or noise in the data which is, in some cases, stronger than the signal from the neural activity [7]. Noise is also introduced into the EEG signal from electrical interference and the subjects' breathing and heartbeat. Despite these limitations, EEG is a promising tool for the continuous measurement of cognitive states. Lee and Tan used a simple and inexpensive EEG system (~$1,500) to differentiate between various users' states that have relevance within the HCI domain [8]. They helped to bridge the gap between brain imaging research and the field of HCI, by providing detailed information about the nature of the EEG signal, its potential within the HCI field, preprocessing methods, and machine learning techniques for the EEG signal [8]. In a later paper, Grimes et al provided an overview of the pragmatics and practicality of using EEG for classification of users' working memory load [9]. They classified two working memory (WM) states with up to 99% accuracy, and four WM states with up to 88% accuracy. They discuss possible applications of EEG memory load measurements as additional metrics for usability testing, and as an additional input to adaptive systems [9]. EEG has also been used in more realistic experimental settings that apply to the military domain. For example, 'smart' EEG helmets have been designed to monitor pilots' mental state while in the air [10]. Other experiments have used portable EEG systems that monitor soldiers' mental WL while completing realistic training scenarios [11].

2.2 Functional Near Infrared Spectroscopy

fNIRs has been introduced in the last two decades [12], and it is primarily used in the medical domain, and in research labs where the focus is to validate and re-design the device itself. fNIRs uses optical fibers placed on the scalp or forehead to send light in the wavelength range of 650-850 nm into the head. A small percentage of this light migrates through the scalp, skull and brain cortex and eventually is collected by other optical fibers placed 2-3 cm away from the source fibers. In the near-infrared range the main tissue absorbers are oxy-and deoxy-hemoglobin, therefore any change in the concentration of these two chromophores (as during brain function) is reflected into intensity changes at the detector's sites[12]. The spatial resolution in fNIRs is limited to approximately 5 mm. Researchers have shown that by placing the probes on a subject's forehead, fNIRs provides an accurate measure of activity within the frontal lobe of the brain, which is responsible for many high order cognitive functions, such as memory and problem solving that make up mental WL [1]. More specifically, there is a positive correlation between the increase of oxygenated blood and the increase in cognitive WL [1, 13]. These results are promising since fNIRs is portable, safe, less invasive than other imaging techniques, and has been implemented wirelessly, enabling use in real world settings [1]. However, fNIRs is not without its own limitations. Unlike PET or fMRI, fNIRs can not measure deep brain structures and it is primarily placed on the forehead, as hair can introduce noise into the signal. fNIRs is also limited by its low temporal resolution, as it take several seconds to monitor blood in the brain. Since most research in fNIRs concerns validating the tool itself, the extensive applications conducted with brain imaging techniques such as EEG, have yet to be implemented with fNIRs. Only a handful of researchers have paired the fNIRs data with machine learning techniques [14-17].

2.3 Combining EEG and fNIRs

A few researchers have explored concurrent recordings of EEG and fNIRS [2-5]. These researchers note the functional limitations of each device. By pairing the low spatial and high temporal resolution of EEG with the high spatial and low temporal resolution of fNIRs, it may be possible to overcome limitations of each measurement technology. Not only do the two devices complement each other by improving on one another's measurement pitfalls, but they also measure different physiological markers in the brain, providing further information about a user's mental state than either device could achieve alone. EEG and fNIRs data were concurrently recorded in the DARPA Augmented Cognition Technical Integration Experiment, where participants were instructed to play Warship Commander, a task which involves monitoring a radar screen for airplanes and then responding to their presence [2]. Researchers collected fNIRs and EEG data, in addition to several other types of physical and mental measurements in an attempt to compare sensor technologies [2]. Another instance in which fNIRS and EEG were used concurrently is a study conducted by Salvatori et al [5]. Due to the novel nature of this research, Salvatori's study was primarily interested in the logistics of combining the two devices and emphasis was placed on issues such as data synchronization and sensor placement.

Participants in the study were asked to watch a computer screen that displayed an alternating white and black checkerboard pattern[5]. Savran et al took concurrent recordings of EEG and fNIRS data while subjects viewed images from the International Affective Picture System (IAPS) [4]. They described the creation of a database to hold EEG and fNIRs data, and they discussed issues in data synchronization and sensor placement. Researchers combining EEG and fNIRs look at the concurrent data separately, and while some initial ideas for concurrent data analysis have been discussed [4], research to date has not taken the next step of actually combining EEG and fNIRS data during data analysis. To the best of our knowledge, we are the first to attempt classification of concurrent recordings of EEG and fNIRs data with machine learning.

3 Experiment

3.1 Equipment

The fNIRs device used in this study is an OxyplexTS (ISS Inc. Champagne, IL) frequency-domain tissue spectrometer with two optical probes. Each probe has a detector and four light sources (Fig. 1c). Each light source emits near infrared light at two separate wavelengths (690nm and 830nm) which are pulsed intermittently in time. This results in 2 probes x 4 light sources x 2 wavelengths = 16 light readings at each timepoint (sampled at 50Hz). Electroencephalograms were collected using 32-channel caps.

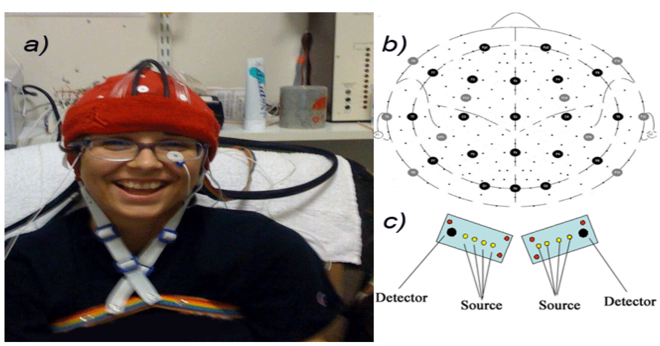

Fig. 1. a) A subject wears fNIRs/EEG setup. b) electrode placement c) fNIRs sensors

Electrodes were arranged according to the International 10-20 system (1b). The EEG was amplified using an SA Bioamplifier (SA Instruments, San Diego Ca.) with a bandpass of .01 and 40 Hz. The computer sampled at a rate of 200 Hz. The fNIRs probes were placed on subjects' forehead using an athletic headband, leaving room for the electrode cap to fit on their heads (1a).

3.2 Experiment Tasks

Our experiment had three conditions (Fig. 2). In each condition, users viewed rows of randomly generated red and blue planes moving down a screen.

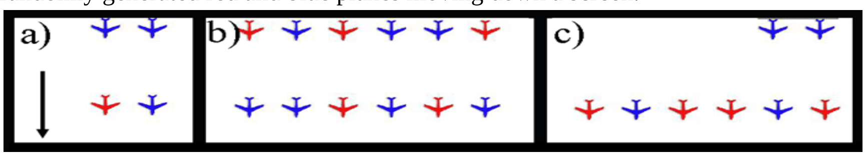

Fig. 2. The three experiment conditions are low WL (a), high WL (b) and random WL (c).

Users kept track of the total number of red and blue planes that they had seen, and after 60 seconds, they were prompted to give the total count. While the figure below shows 2 rows of planes in each condition, in reality, only one row of planes was on the screen at any given time. In the first condition (Figure 2a) subjects made WM updates based on the two planes that were in each row. In the second condition (2b) subjects viewed six planes per row, and in the final condition (2c), subjects viewed randomly generated rows of two or six planes. We based our WL conditions on the fact that set size has been used to manipulate WM for decades. We will refer to these conditions as the low WL (2 planes per row), random WL (2 or 6 planes per row), and high WL (6 planes per row) conditions.

3.3 Task Events and Synchronization

Events were automatically generated and logged each time that one of the 60 second long tasks began or ended. Ideally, the EEG and fNIRs data acquisition would be synchronized in time, and any events recorded throughout the experiment would be logged in the concurrently recorded EEG and fNIRs acquisition systems [4, 5]. Due to hardware limitations, an approximation was made. The EEG and fNIRs acquisition systems were set to start at the same time via two button presses (one for the fNIRs acquisition system and one for the EEG system). Each event was logged directly into the fNIRs acquisition system, which sampled at 50Hz, and added to the EEG data after the experiment ended.

3.4 Experiment Methodology

Four right handed, undergraduate students at Tufts completed this experiment (3 women, 1 male). Subjects were instructed to keep movement to a minimum, and to count the number of red and blue planes that they saw during each 60 second task period. After each task ended, they verbally gave their answers. Subjects rested for 20 seconds between each task. We used a randomized block design with 8 trials, resulting in 24 tasks.

4 Results and Analysis

4.1 Performance Results

We calculated subjects' performance (whether they said the correct number of red/blue planes) on the three experiment conditions. WM load increased as the number of planes per row increased, and subjects' performance on each of the experiment conditions reflected this increase in WL. Subjects had the highest accuracy for the tasks that involved 2 planes per row (low WL). They had the lowest accuracy on tasks with 6 planes per row (high WL), and their accuracy on the random WL condition was always between the accuracy on the high and low WL conditions.

4.2 Data Preprocessing and Machine Learning Analysis

As brain activity differs widely on a person to person basis, we run all analyses separately for each subject. We developed machine learning techniques to classify the EEG and fNIRs data from the experiment. *Our analysis is not intended to compare the two techniques of brain measurement or the machine learning techniques.*

fNIRs: Each experiment lasted about 35 minutes, with fNIRs data recorded every .02 seconds. We recorded 16 channel readings at each timepoint, where we refer to the readings of one source detector pair at one wavelength, as one *channel*. We first normalize the intensity data in each channel by their own baseline values. We then apply a moving average band pass filter to each channel (saving frequency values between .1 and .01 Hz) and we use the modified Beer-Lambert Law[12] to convert our light intensity data to measures of the change in oxygenated hemoglobin (HbO) and deoxygenated hemoglobin (Hb) concentrations in the brain. This results in eight readings of HbO and eight readings of Hb data at each timepoint in the experiment. To choose the best HbO and Hb channels for classification we use the techniques described in [14].We implemented a weighted k-nearest-neighbor classifier (k=3) with a distance metric computed with Symbolic Aggregate Approximation (SAX).

SAX creates a symbolic approximation of time series data, allowing for dimensionality reduction. For a review of SAX, see [18].

EEG: We implemented signal processing and feature generation schemes that have been used with EEG data previously [8]. We split the continuous EEG data into small overlapping windows and we took a Fourier transform of the data in each window. We chose a window size of 2 seconds, with windows overlapping every second. For each window, we compute the magnitude and phase of the signal and the spectral power of the signal in the delta (1-4Hz), theta (4-8Hz), alpha (8-12Hz), beta-low (12-20Hz), beta-high (20-30Hz), and gamma (30-50Hz) frequency bands. We also compute the coherence and cross spectrum between each channel for each frequency band in each window. This results in over 6,200 features for each instance. We use blocked cross validation to select our most relevant attributes for classification. We use an information gain heuristic followed by Weka's *CfsSubsetEval* function to choose the features that best predict the class label in the training data. We then apply a Naïve Bayes classifier to data.

4.3 Classifying Working Memory Load

We attempted to classify each 60 second long period of time when subjects were completing one of the 3 conditions described in Fig 2. We classified these conditions with our fNIRs data, and with our EEG data. Classification results are depicted in Table 1. We looked at our ability to classify low and high WL, low and random WL, random and high WL, and low, random, and high WL. When using just the fNIRs data for analysis, we see that we achieved promising accuracy for each subject. However, with the exception of subject 1, the EEG data yields nearly random accuracy. It is promising that the fNIRs classification yielded accuracy as high as 82% distinguishing two WL classes and up to 50% for distinguishing three WL classes.

Table 1. Classification accuracy on just fNIRs data (unshaded) and just EEG data (shaded columns) for the three conditions in Fig 4. S1 = subject 1, etc.

	random v. low	random v. low	random v .high	random v. high	low v. high	low v. high	low v. random v. high	low v random v. high
s1	57%	63%	64%	75%	61%	69%	45%	42%
s2	53%	56%	53%	69%	51%	69%	34%	46%
s3	50%	82%	52%	69%	49%	75%	33%	50%
s4	52%	78%	50%	65%	52%	75%	34%	46%

4.4 Analysis of Results

There are many possible explanations for the low EEG classification accuracy in this experiment. Our ability to accurately measure users' states depends on a number of factors. Any changes in the placement of the fNIRs and EEG sensors on subjects, changes in the analysis techniques applied to the brain data, or changes in the experiment tasks could result in higher or lower classification accuracy of fNIRs or EEG brain data. It is possible that the chosen task elicited brain activity primarily in the prefrontal cortex, which was dominated by the fNIRs sensors. It is also possible that the fNIRs light sources introduced noise into the EEG signal that was difficult to remove [4]. Therefore, the EEG sensors may not have been able to pick up the resulting brain activity.

5 Conclusion and Future Work

In this paper, we discussed the physiological metrics that EEG and fNIRs can measure, the functional and practical limitations of each device, and technical issues to be mindful of when combining the devices. We presented machine learning methods that can be used on EEG and fNIRs data. We presented an experiment designed to combine fNIRs and EEG to measure three user states. While the EEG results were low, we demonstrated the ability of fNIRs to classify mental WL states. fNIRs is a relatively new device, and it holds great potential for the HCI domain. Future work will explore factors that may have contributed to the low EEG results. It is possible that EEG and fNIRS can best complement eachother when subjects complete tasks that activate more than just the prefrontal cortex. It is apparent that EEG and fNIRs, when combined, have the potential to acquire complementary information about user states. However, more research is needed to achieve this goal.

6 References

1. Izzetoglu, K., et al., *Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks.* International Journal of Human-Computer Interaction, 2004. **17**(2): p. 211-231.
2. John, M.S., et al., *Overview of the DARPA Augmented Cognition Technical Integration Experiment.* International Journal of Human-Computer Interaction, 2004. **17**(2): p. 131-149.
3. Tong, Y., et al., *Concurrent recordings of electrical evoked potentials and near-infrared responses to brain activation*. 2007, Tufts University.
4. Savran, A., et al. *Emotion Detection in the Loop from Brain Signals and Facial Images*. in *eNTERFACE'06*. 2006. Dubrovnik, Croatia.

5. Salvatori, G., et al. *Combining Near-Infrared Spectroscopy and Electroencephalography to Monitor Brain Function*. in *Instrumentation and Measurement Technology Conference*. 2006. Sorrento: IEEE.
6. Anderson, C. and Z. Sijercic, *Classification of EEG Signals from Four Subjects During Five Mental Tasks.* Solving Engineering Problems with Neural Networks: Proceedings of the Conference on Engineering Applications in Neural Networks, 1996: p. 407--414.
7. Berka, C., et al., *Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset.* International Journal of Human Computer Interaction, 2004. **17**(2): p. 151–170.
8. Lee, J.C. and D.S. Tan. *Using a Low-Cost Electroencephalograph for Task Classification in HCI Research*. in *ACM Symposium on User Interface Software and Technology*. 2006.
9. Grimes, D., et al. *Feasibility and Pragmatics of Classifying Working Memory Load with an Electroencephalograph*. in *CHI 2008 Conference on Human Factors in Computing Systems*. 2008. Florence, Italy.
10. Gevins, A., et al., *Flight helmet EEG system*, in *Final Tech Report AL/CF-SR-1993-0007*. 1993, Sam Technology: San Fracisco, CA.
11. Mathan, S., et al. *Neurophysiological Estimation of Interruptibility: Demonstrating Feasibility in a Field Context.* in *In Proceedings of the 4th International Conference of the Augmented Cognition Society*. 2007. Baltimore, MD.
12. Chance, B., et al., *A novel method for fast imaging of brain function, non-invasively, with light.* Optics Express, 1988. **10**(2): p. 411-423.
13. Parasuraman, R. and D. Caggiano, *Neural and Genetic Assays of Human Mental Workload*, in *Quantifying Human Information Processing*. 2005, Lexington Books.
14. Hirshfield, L.M., et al. *Brain Measurement for Usability Testing and Adaptive Interfaces: An Example of Uncovering Syntactic Workload in the Brain Using Functional Near Infrared Spectroscopy*. in *Conference on Human Factors in Computing Systems: Proceeding of the twenty-seventh annual SIGCHI conference on Human factors in computing systems*. 2009.
15. Sassaroli, A., et al., *Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy.* accepted in the Journal of Innovative Optical Health Sciences, 2009.
16. Son, I.-Y., et al., *Human performance assessment using fNIR.* Proceedings of SPIE The International Society for Optical Engineering, 2005. **5797**: p. 158–169.
17. Sitaram, R., et al., *Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface.* NeuroImage, 2007. **34**(1416).
18. Lin, J., et al. *A Symbolic Representation of Time Series, with Implications for Streaming Algorithms.* in *In proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. 2003. San Diego, CA.