Distinguishing Difficulty Levels with Non-invasive Brain Activity Measurements

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

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

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

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

ABSTRACT

Passive brain-computer interfaces are designed to use brain activity as an additional input, allowing the adaptation of the interface in real time according to the user's mental state. The goal of the present study is to distinguish between different levels of game difficulty using real-time, non-invasive brain activity measurement with functional near-infrared spectroscopy (fNIRS). The study is designed to lead to adaptive interfaces that respond to the user's brain activity in real time. Nine subjects played two levels of the game Pacman while their brain activity was measured using fNIRS. Statistical analysis and machine learning classification results show that we can discriminate well between subjects playing or resting, and distinguish between the two levels of difficulty with some success. These results show potential for using fNIRS in an adaptive game or user interface. This work is an improvement on previous fNIRS game studies which seldom try to tell apart two levels of brain activity.

Author Keywords

Brain-Computer Interface, human cognition, functional near-infrared spectroscopy, fNIRS, task classification, game, difficulty level

ACM Classification Keywords

H.5.2 User Interfaces: Input devices and strategies; B.4.2 Input/Output Devices: Channels and controllers;

INTRODUCTION

A brain-computer interface (BCI) can be loosely defined as an interface controlled directly or indirectly by brain activity of the user. While most BCI research is designed for direct use with disabled users, we instead focus on passive BCIs [1] for healthy users. Passive BCIs are interfaces that use brain measurements as an additional input, in addition to standard devices such as keyboards and mice.

Unlike much BCI work which uses electroencephalography (EEG), this research uses functional near-infrared spectroscopy (fNIRS), which is non-invasive, portable, relatively impervious to user movement and chosen in part because of its unique sensitivity to changes of oxy- and deoxy-hemoglobin. Figure 1 shows an fNIRS probe. By measuring the reflection of near-infrared light sent into the head, we can extrapolate a measure of brain activity. This tool has been used in biomedical contexts, but little has been done to take advantage of it in a human-computer interaction (HCI) context.

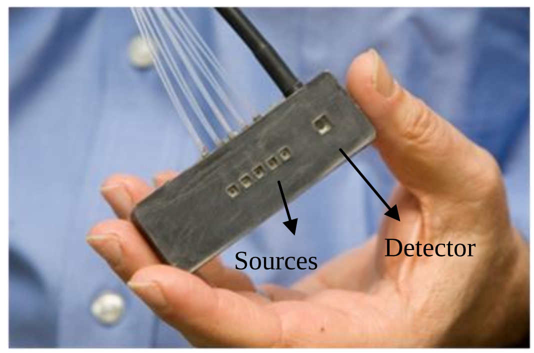

Figure 1. A picture of the right probe. A probe includes a detector (larger square) and four light sources (smaller squares).

The goal of the present study is to distinguish between different levels of game difficulty using functional near-infrared spectroscopy (fNIRS) data collected while subjects played a computer game. The study is designed to ultimately lead to adaptive games and other interactive interfaces that respond to the user's brain activity in real time. Our results show that we could distinguish between the user playing Pacman or being at rest, as well as between two difficulty levels of Pacman.

RELATED WORK

Several fNIRS studies reported a significant variation in hemoglobin concentration in the prefrontal cortex in comparison to resting while playing an arcade game [8], a shooting game, a rhythm action game, a block puzzle and a dice puzzle [7]. Another study showed that one could differentiate between playing and not playing a computer game using fMRI, by comparing three video games, Space Invaders, Othello and Tetris [9]. These studies compare rest versus play, but never more than one level of difficulty. Task load and blood oxygenation have also been shown to be correlated in a non-game environment [5]. Others have measured the brain during game play using EEG and demonstrated the ability to distinguish the user resting, exploring the game environment or playing the video game [6]. Based on these results, we wanted to explore the fNIRS response in the prefrontal cortex during different levels of video game play.

EXPERIMENTAL PROTOCOL

The goal of this study was to differentiate between different levels of a computer game. The arcade game Pacman was selected because of its well known nature and of the ability to change its parameters. The goal is for Pacman to eat as many fruits and enemies as possible, without being killed. Two levels of difficulty were selected through pilot testing, an easy and a hard level, which were hypothesized to be distinguishable both by the participants and by analyzing their brain signal. The easy and hard levels are differentiated by the pace and quantity of enemies.

Nine subjects (4 females) participated in this study (mean age of 24.2 years; std 4.15). All were right-handed, with normal or corrected vision and no history of major head injury. They signed an informed consent approved by the IRB, and were paid $10 for their participation. All knew of the game, and all but one had previously played it. Participants practiced the game for about one minute to familiarize themselves with our version.

Design and Procedure

Participants then completed ten sets of two trials (one in each difficulty level) over a twenty minute period. In each trial, participants played the game for a period of thirty seconds and rested for thirty seconds to allow their brain to return to baseline. Conditions within each set were randomized for each subject.

In addition to fNIRS data, we collected performance data– number of times Pacman is killed, and number of apples, cherries and enemies eaten. At the end of the experiment, subjects were asked to rate the overall mental workload of each game level with the NASA Task Load Index (NASA-TLX) [4]. It is a widely used subjective measure of mental workload which we can use to confirm the choice of parameters for the two difficulty levels.

fNIRS Equipment

fNIRS measures changes in hemoglobin concentrations [10]. At the near-infrared range, light can pass through most tissues, allowing them to be probed for depths up to 1-3 centimeters. By measuring the light sent at two wavelengths, we can calculate oxygenated hemoglobin and deoxygenated hemoglobin concentration. fNIRS provides high temporal resolution (in the order of tenths of ms), and a spatial resolution of approximately 5mm. However, it can only measure the outer cortex of the brain. While there are many brain imaging techniques, we believe fNIRS to be a suitable brain sensing technology for HCI research because it is safe, non-invasive, easy to use, and relatively impervious to user movement.

We collected fNIRS data using an OxiplexTS, from ISS, Inc. (Champaign, IL). Our setup is comprised of two probes, each containing one detector and four light sources arranged in a linear array (see Figure 1). Each source emits two wavelengths (690 and 830nm), with a sampling rate of 6.25Hz. The probes were placed in the middle of the forehead. We chose to use the data from the two last channels of the probe only (with source-detector distances of 2.5 and 3cm), because they reach deeper into the cortex. The shallower measures may pick up mainly systemic responses happening in the skin.

The user's right hand was positioned to reach the arrow keys of a standard keyboard comfortably, with the fingers resting on the keys, minimizing all movement of the arm and hand, which can produce motion artifacts in the brain measurements. A pilot test indicated that small finger movements show up only minimally in our data, and is mostly removed with filtering.

ANALYSIS TECHNIQUES AND RESULTS

Subjective Results and Performance Data

We analyzed results from the NASA-TLX data to confirm the parameter choices in the two conditions. We wanted to see whether the users perceived the two conditions as indeed different. Results indicated an average mental workload index of 26 (std 12.9) for the easiest level, and 69 (std 7.9) for the hardest level, on a 100 point scale. These numbers were significant ($p<0.01$), and confirm the hypothesis.

We also examined the performance data. Every type of data collected showed a significant difference between each difficulty level ($p<0.05$).

While some might argue that performance data is sufficient to classify the difficulty level of a game and can be obtained without interference, the goal of this study is to investigate the use of the brain measurements with fNIRS as a new input device. In a more complex problem, performance and brain data coming from fNIRS might not be as related, e.g. if the user is working hard yet performing poorly at some point. The use of the brain signal as input would provide better results in this situation.

Brain Data

Data Preprocessing

To remove motion artifacts, and optical changes due to breathing and heart beat, we applied a folding average filter using a non-recursive time-domain band pass filter, keeping frequencies between 0.01Hz and 0.5Hz. The filtered raw data was then transformed into oxygenated hemoglobin and deoxygenated hemoglobin concentrations (respectively [HbO] and [Hb]), using the modified Beer-Lambert law [10]. Given the assumption that the brain has returned to a baseline state after each rest period following the stimuli, we shift each trial so that the initial value is zero. Finally, we separate each trial according to *Activeness*—whether the user was playing or resting.

Statistical Analysis

For the statistical analysis, we average each trial to get a mean value of [HbO] and [Hb], for each difficulty level, activeness, hemisphere and channel. We then apply a factorial repeated measures analysis of variance (ANOVA) on *Difficulty level* (2) x *Activeness* (2) x *Hemoglobin Type* (2) x *Hemisphere* (2) x *Channel* (2) x *Trial* (10) x *Subject* (9). In addition to the statistical significance, we report the effect size of the interaction (ω^2), which is the magnitude of the observed interaction, and indicates practical significance. An omega-squared measure of 0.1 indicates a small effect, 0.3 shows a medium effect and 0.5 means a large effect [3]. Note that we chose to omit reporting some significant results not pertinent to current questions.

We found that the *Hemoglobin Type* was a significant factor, with a medium effect ($F(1, 8)=6.819$, $p<0.05$, $\omega^2=0.39$). This was expected, because [Hb] and [HbO] are present in different concentrations in the blood.

The activeness of the subject is distinguishable if combined with the channels, with a large effect size, i.e. a significant interaction effect between *Activeness* x *Channel* ($F(1, 8)=27.767$, $p<0.001$, $\omega^2= 0.75$), showing that there is a difference between playing Pacman and resting.

Finally, we observed a significant interaction of *Difficulty Level* x *Activeness* x *Channel* x *Hemoglobin Type*, with a small effect size ($F(1, 8)= 7.645$, $p<0.05$, $\omega^2= 0.18$). This interaction shows that we can significantly distinguish between the activeness of the participant, and the degree of difficulty of the current game, when we take into account the channel and the hemoglobin type. This shows that our initial hypothesis is correct.

Machine Learning Classification

Statistical analysis confirmed our hypothesis that the brain signals in the different conditions were also significantly different. We then wanted to determine whether this signal could be used in an adaptive user interface. To do this, we used machine learning to train a classifier on some of the data and then classified the remaining data.

We chose sequence classification [2] because of its simple nature. Sequence classification applies a label to an entire sequence of data, and uses each data point as a feature. In our case, a sequence is one trial, containing 180 points. We used the same preprocessing as for the statistical analysis, but we use non-zeroed data, as it is more similar to data we would have in a real time brain-computer interface.

Because we have multivariate data (8 recordings for each time point: 2 probes x 2 channels x 2 hemoglobin types), we classify each individually first. To combine the results of all these classifications, each classifier votes for the label of the example. We used a weighted voting technique that takes into account the probability distribution of each example by each classifier.

The classification algorithm used is k-nearest-neighbors (kNN), with k=3. kNN uses the label of the three most similar examples (the closest neighbors) to the example to classify, and assigns a label based on the weighted average of their labels. We used a random 10-fold cross-validation in all classifications.

We attempted three types of classification: (a) *Activeness* (Play versus Rest), (b) *Difficulty level* (Easy versus Hard), and (c) *Two difficulty levels and rest* (Easy versus Hard versus Rest). To accomplish each classification, we selected and/or grouped the trials differently. For *Activeness*, we combined all playing trials into one class, and all resting trials into another to form two classes. For *Difficulty Level*, we compared the easy and hard levels using the play trials only. Finally, in *Two difficulty levels and rest*, we compared three conditions: the play period of the easy level, the play period of the hard level, and all rest periods. Figure 2 shows the average accuracy of each type of classification.

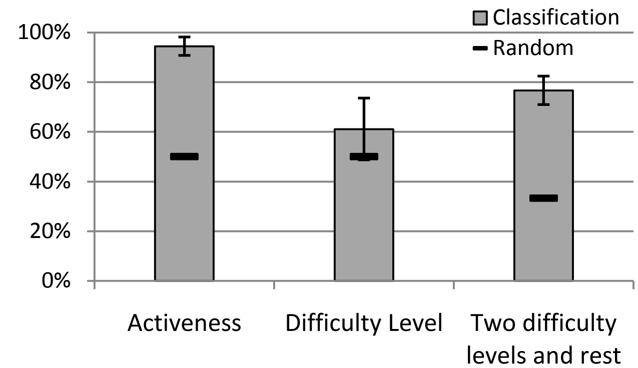

Figure 2. Average accuracy for different classifications, with the standard variation and the random classification accuracy.

DISCUSSION

Brain activation when playing Pacman: play versus rest

Results indicate the presence of a distinct brain signal when playing Pacman, in comparison to the rest periods. The *Activeness* classification in Figure 2 yields an average accuracy of 94.4%. It indicates a noticeable difference between the playing signal, and the resting signal. This corresponds to the results obtained with the statistical analysis, where *Activeness* was a significant factor in multiple interactions. This provides real time measurements that could be used in an adaptive interface. Our results corroborate those of previous studies that showed prefrontal cortex activity related to video games, measured with fNIRS.

Difficulty levels: easy versus hard

The *difficulty level* of the game was shown to be a significant factor in this experiment in both types of analyses. This is corroborated with the fact that users perceived the two levels as being significantly different according to the NASA-TLX. Hence, we can say that there was a significant cognitive difference between the two levels. Previous fNIRS game experiments [7, 8] only analyzed stimuli versus non-stimuli periods (activeness), and not two levels of difficulty, making this result an advance over prior work.

However, the statistically significant interaction that included *Difficulty Level* had a small effect size. Additionally, classifying the difficulty of playing periods yields an average accuracy of 61.1%. This relatively low accuracy indicates that it is difficult with this classifier to differentiate between the two levels, which relate to the small effect size found in the statistical analysis. A comparison of three types of conditions (*Two difficulty levels and rest*) indicates an encouraging average accuracy of 76.7%, explained by the low differentiation between the difficulty levels, and the high separation between the activeness of the subjects. We must note that the difference in brain signal measure is not strong. One explanation may be that the difference in mental processes between each level manifests itself in other brain locations besides the anterior prefrontal cortex.

Overall, the findings indicate the presence of brain activation in the anterior prefrontal cortex when playing Pacman. The machine learning results show the ability of fNIRS data to be classified easily and the potential they can have to be used in an adaptive interface.

FUTURE WORK

Next steps include converting an offline classifier into a real-time algorithm that accepts streaming data for use in an adaptive interactive user interface. Additionally, a probe with more sensors, placed differently, could lead to a stronger signal, as it would pickup brain activity in more locations. Results of the comparison of two different levels could be applied to other games of similar mental demand. The correlation between mental workload and difficulty levels in this experiment indicates we could also apply the current results to general applications that respond to workload or difficulty levels.

There are limitations to using fNIRS in real-time, such as the fact that the metabolic response measured by fNIRS occurs over a few seconds, and the presence of motion artifacts in the data. This has implications for a real time user interface such as the lack of an immediate, perfect response from the system. Using fNIRS as a passive supplemental input will avoid some of these issues since the interface would not be dependent on this signal for its interaction. The interface can be adapted in a subtle matter, when we have a high degree of certainty in the user's cognitive state. In the case of an adaptive Pacman, changing the difficulty level should not be clearly noticeable to the user.

CONCLUSION

In this experiment, we have shown that functional near-infrared spectroscopy can distinguish between the brain at rest and the brain activated when playing a video game, both using statistical analysis and machine learning classification. We also demonstrated that we can differentiate two levels of difficulty. The activation of the different levels of difficulty is correlated with mental workload, measured with NASA-TLX. Hence, we can presume that the difficulty level in this experiment is correlated with mental workload. However, our classification accuracy was low when distinguishing playing easy or hard Pacman. This is a step forward, as previous work only studied the activeness of the user during video games using fNIRS. We believe this work to be a stepping stone to using fNIRS in an adaptive user interface, in this case a passive brain-computer interface. In a real time user interface, we could use fNIRS measurement as an additional input on which to adapt the interface.

ACKNOWLEDGEMENTS

The authors would like to thank our colleagues in the HCI research group at Tufts University; Michel Beaudoin-Lafon, Wendy Mackay, and the In|Situ| research group; and Desney Tan at Microsoft Research for their helpful inputs and encouragement.

We thank the National Science Foundation for support of this research (Grant Nos. IIS-0713506 and IIS-0414389). Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the National Science Foundation. We also thank the Natural Sciences and Engineering Research Council of Canada for financial support.

REFERENCES

1. Cutrell, E. and Tan, D.S., BCI for passive input in HCI. in ACM CHI 2008 Workshop on Brain-Computer Interfaces for HCI and Games, (2007).
2. Dietterich, T.G. Machine Learning for Sequential Data: A Review. in Structural, Syntactic, and Statistical Pattern Recognition, Springer-Verlag, 2002, 15-30.
3. Field, A.P. and Hole, G. How to design and report experiments. Sage publications Ltd., London; Thousand Oaks, Calif., 2003.
4. Hart, S.G. and Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theorical research. in Hancock, P., Meshkati, N. ed. Human Mental Workload, Amsterdam, 1988, 139-183.
5. Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K. and Chance, B. Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. IJHCI, 17 (2). 211-231.
6. Lee, J.C. and Tan, D.S., Using a Low-Cost Electroencephalograph for Task Classification in HCI Research. in Proc. UIST, (2006).
7. Matsuda, G. and Hiraki, K. Sustained decrease in oxygenated hemoglobin during video games in the dorsal prefrontal cortex: A NIRS study of children. NeuroImage, 29 (3). 706-711.
8. Nagamitsu, S., Nagano, M., Tamashita, Y., Takashima, S. and Matsuishi, T. Prefrontal cerebral blood volume patterns while playing video games-A near-infrared spectroscopy study. Brain & Development, 28. 315-321.
9. Saito, K., Mukawa, N. and Saito, M., Brain Activity Comparison of Different-Genre Video Game Players. in ICICIC '07, (2007), 402-406.
10. Villringer, A. and Chance, B. Non-Invasive Optical Spectroscopy and Imaging of Human Brain Function. Trends in Neuroscience, 20. 435-442.