Designing a passive brain computer interface using real time classification of functional near-infrared spectroscopy

Audrey Girouard*, Erin Treacy Solovey and Robert J.K. Jacob

Department of Computer Science,
Tufts University,
Medford, MA 02155, USA
E-mail: agirou01@cs.tufts.edu
E-mail: etreac01@cs.tufts.edu
E-mail: jacob@cs.tufts.edu
*Corresponding author

Abstract: Passive brain–computer interfaces consider brain activity as an additional source of information, to augment and adapt the interface instead of controlling it. We have developed a software system that allows for real time brain signal analysis and machine learning classification of affective and workload states measured with functional near-infrared spectroscopy (fNIRS) called the online fNIRS analysis and classification (OFAC). Our system reproduces successful offline procedures, adapting them for real-time input to a user interface. Our first evaluation compares a previous offline analysis with our online analysis. While results show an accuracy decrease, they are outweighed by the new ability of interface adaptation. The second study demonstrates OFAC's online features through real-time classification of two tasks, and interface adaptation according to the predicted task. Accuracy averaged over 85%. We have created the first working real time passive BCI using fNIRS, opening the door to build adaptive user interfaces.

Keywords: BCI; brain–computer interface; fNIRS; functional near-infrared spectroscopy; workload; passive BCI; PFC; prefrontal cortex; task classification; adaptive communications.

Reference to this paper should be made as follows: Girouard, A., Solovey, E.T. and Jacob, R.J.K. (2013) 'Designing a passive brain computer interface using real time classification of functional near-infrared spectroscopy', *Int. J. Autonomous and Adaptive Communications Systems*, Vol. 6, No. 1, pp.26–44.

Biographical notes: Audrey Girouard is a PhD Candidate in Computer Science at Tufts University. Specialising in HCI, she is currently interested by passive brain–computer interfaces, studying the use of brain imagery to enhance HCI for all, with the use of functional near-infrared spectroscopy (fNIRS). In 2008, she was a Visiting Researcher in the InSitu group at Université de Paris-Sud, working with Wendy Mackay. She is the Laureate of the Post-Graduate Scholarship from NSERC. She has an MS in Computer Science from Tufts University and a BEng in Software Engineering from École Polytechnique de Montréal.

Erin Treacy Solovey is a Computer Science Doctoral Candidate at Tufts University. She is investigating the use of fNIRS and electroencephalography to sense signals that users naturally give off in order to augment the explicit input provided through standard input devices. She received her bachelor's degree in Computer Science from Harvard University and an MS in Computer Science from Tufts University.

Robert J.K. Jacob is a Professor of Computer Science at Tufts University, where his research interests are new interaction media and techniques and user interface software. He was also a Visiting Professor at the Universite Paris-Sud and at the MIT Media Laboratory and continues collaboration with these groups. Before coming to Tufts, he was in the HCI Lab at the Naval Research Laboratory. He received his PhD from Johns Hopkins University, and he is a Member of the editorial board of Human–Computer Interaction and the ACM TOCHI. He was elected to the ACM CHI Academy in 2007.

1 Introduction

Enhancing user experience is a constant goal for human–computer interaction (HCI) researchers, and the methods to achieve this goal are widespread, from changing the properties of the interface to adapt the task to the user's ability level. Ideally, those modifications should be done automatically, in real time, as to obtain maximum benefit. By sensing different user properties, such as interest, workload, frustration and flow, we can adapt the interface immediately to keep them optimally working. In this research, we focus on mental workload to improve the interface. We investigate ways to obtain workload information the user naturally gives off when using the computer by acquiring brain patterns, to automatically enhance their experience.

Brain–computer interfaces (BCI) are designed to use brain activity as an input for interfaces. Most current work in BCI allows disabled patients to communicate with their environment with the use of electroencephalography (EEG) (Krepki et al., 2007; Wolpaw et al., 2002). This requires a great deal of training from the user, to learn which type of signals to produce, and from the system, to learn which actions yield which signal. The resulting interfaces let the user select a binary choice, move a mouse or type. The interface is often slow to respond, especially in comparison with traditional input technologies (mouse and keyboard).

A new train of thought in the BCI community considers brain activity as an additional source of information, to augment and adapt the interface instead of controlling it directly with the brain. The new methodology focuses on a broader group of users – the general population – for whom current BCIs are impractical because of their slow speed of transfer. Passive BCIs are designed to use brain activity as a new input modality, allowing the adaptation of the interface in real time according to the user's mental state (Cutrell and Tan, 2008). This type of BCI can capture intentional commands, but is best designed for implicit communication (Zander et al., 2010). While most BCI research is done in fields such as psychology and biomedical engineering, the study of passive BCIs could gain from the knowledge and expertise from HCI. HCI studies how to evaluate and improve the connection between human and computer, to create seamless interaction. We hope to achieve a similar interaction using the brain.

Figure 1 The use of fNIRS in typical computer settings (see online version for colours)

We use functional near-infrared spectroscopy (fNIRS) as brain measurement in this research. By measuring the reemission of near-infrared light sent in the brain, this device extrapolates a measure of brain activity from blood flow. More specifically, fNIRS calculates the changes in oxygenated and deoxygenated haemoglobin concentrations. This non-invasive and portable technology offers interesting applications for the field of HCI, as it is relatively impervious to user movement (Solovey et al., 2009) (Figure 1). Researchers have demonstrated fNIRS' ability to measure a number of cognitive and motor brain signals such as mental workload levels, emotions, or motor activity and imagery (Girouard et al., 2009; Hirshfield et al., 2007; Leon-Carrion et al., 2007; Sitaram et al., 2007).

While most work using fNIRS uses offline analyses to evaluate the data collected, the key component of BCI is the ability to perform real time analyses. Many researchers argue that their work could be converted and done in real time (Girouard et al., 2009; Sitaram et al., 2007), yet we found few fNIRS systems in the literature that do (Coyle et al., 2007; Luu and Chau, 2009). The existing systems use simple paradigms, making decisions with a threshold or by comparing the signal from two previous tasks.

We have developed a software system that allows for real time fNIRS brain signal analysis and machine learning classification of affective and workload states called the online fNIRS analysis and classification (OFAC) system. This system receives and processes brain signals and event markers, automatically recognises the current cognitive or affective state using a database of previously recoded signals and machine learning techniques, and outputs this state to the interface, allowing for the creation of interfaces that adapt and change in real time according to traditional inputs as well as cognitive activity. OFAC offers the user an additional communication channel based on brain activity, providing multimodal interaction.

Our work aims at reproducing the procedures used offline in previous work, adapting them to be suitable for real time input to a user interface. This paper presents the OFAC system, tests and proves the system's reliability and potential through two studies. Our first evaluation compares a previous offline analysis with our real time analysis. The second study demonstrates the online features of OFAC: its ability to record, process, classify and adapt simple interfaces in real time.

2 Literature review

The most common types of BCI use intentionally generated brain activity as the primary and often only input device. They are called *direct BCIs*, and this is how most researchers define the general term of BCI (e.g. Wolpaw et al., 2002). The original motivation for such BCIs was to provide assistive technology for users with severe physical disabilities, such as paralysed or 'locked in' patients, allowing them to convey messages and commands to the external world (Wolpaw et al., 2002). Direct BCIs often require the user to be trained to generate specific brain states which are interpreted as explicit input. These states are not always related to the specific output action, for instance performing a motor imagery of the left hand to move the cursor up, and the right foot to move it down (Mappus et al., 2009).

This contrasts with *passive BCIs*, which detect brain activity that occurs naturally during task performance for use as an additional input, in conjunction with standard devices such as keyboards and mice (Cutrell and Tan, 2008). While passive BCIs can detect voluntary input, their use (and the user's performance) is maximised when detecting general state signals such as emotions, language production or workload as passive BCIs are designed not to require the user's full attention (Zander et al., 2010).

Less invasive than other brain monitoring techniques, EEG has thus far been the tool of choice for researchers looking to measure brain activity non-invasively (Coffey et al., 2010; Lee and Tan, 2006). Many BCI systems operate in real time, processing EEG data streams and controlling interfaces (see Delorme et al. (2010) and Schlögl et al. (2007) for reviews of existing BCI tools). We learn from these tools and apply their principles to the design of an fNIRS real time system.

2.1 Functional near-infrared spectroscopy

fNIRS is a brain imaging technique alternative that measures haemoglobin concentration and tissue oxygenation in the brain (Chance et al., 1998). It uses light sources placed on the scalp to send near-infrared light. Biological tissues are relatively transparent at these wavelengths, so the light attenuation through tissues is sufficiently low to allow for tissue imaging at depths up to 1–3 cm. Deoxygenated and oxygenated haemoglobin, present in the blood, are the main absorbers of near-infrared light in tissues, and provide relevant markers of hemodynamic and metabolic changes associated with neural activity in the brain. Therefore, fNIRS researchers can estimate hemodynamic changes connected to brain activation by using light detectors sensitive to the diffusively reflected light that has probed the brain cortex (Izzetoglu et al., 2004).

While fNIRS provides high temporal resolution, the slow hemodynamic changes measured occur in a time span of 6–8 sec (Bunce et al., 2006). Its spatial resolution is approximately 5 mm, the area measured being the one right below the sensor. fNIRS is safe, non-invasive, easy to use and relatively impervious to user movement, with a low setup time.

In most fNIRS studies, researchers identify the difference between two states: activation and rest. Activation occurs when participants perform a specific task for a few seconds up to a few minutes, such as mental rotations, arithmetic or language production. Translated to a real time system, this protocol usually leads to binary decisions, where the user is asked to perform an activating task to indicate intent, and to rest otherwise.

The most common outcomes of such binary decisions are direct control of interfaces, active BCIs. For instance, Coyle et al. (2007) presented a real time fNIRS system that allowed participants to select a coloured box by performing a mental rotation when the preferred target was highlighted. In a more complex (but preliminary) interface by Mappus et al. (2009), users drew a line on a two dimensional plane using activated periods to do straight lines, and rest periods to curve the line. In both studies, participants were instructed what brain task to perform in order to use the system. Using a different paradigm than the activation-rest one, Luu and Chau (2009) compared two different activated brain signals to indicate drink preference. To our knowledge, this is the first example of a real time system that distinguished two activation states without specific instructions.

While the studies mentioned are using direct brain input, we believe fNIRS to be better suited for passive BCIs. The relatively slow signal response of fNIRS does not lend itself to be the best technology for rapid communication, direct input, especially when designed for the general public. In addition, researchers use basic techniques such as selecting the signal with the highest amplitude to make the binary decisions. We believe there are more powerful and better suited techniques.

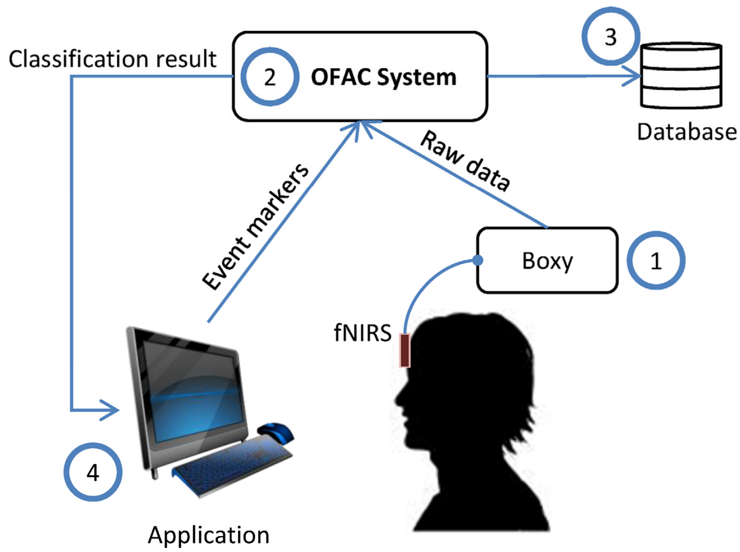
3 OFAC system

We present a new system that uses machine learning to classify a large number of states in real time to obtain the user's affective or cognitive state. OFAC works with fNIRS data in a real time pipeline to feed a user interface, which can in turn adapt to the information.

We created a flexible, modular architecture for the OFAC system using Matlab (Figure 2). It allows for advantageous substitution of single modules should another functionality be required, and accepts multiple input signals, such as the combination of fNIRS and EEG. The rest of this paper will describe an fNIRS-only system used in the latter experiments.

OFAC contains three types of modules for data processing: modules to receive and record input data into a database (one for each type of input); to pre-process and to filter data; and to perform machine learning classification and output the brain signal classification to the interface. The current system takes two different types of input: raw brain data, and external markers from the application shown to the user. The raw data (from the fNIRS acquisition software Boxy, ISS Inc.) can include basic markers related to the start and stop of the real fNIRS data when the sensors are correctly in place and the experiment starts, as opposed to uncalibrated data. The external markers could contain behavioural data, for instance, to help with data classification.

Figure 2 FAC system's architecture (see online version for colours)



OFAC is a complex distributed system that can process each module on an individual computer. This architecture imposes minimal requirements for the application software, which can be written in any language on any platform. The only constraint is to have the ability to connect with OFAC and respect a defined communication protocol.

OFAC provides both an *online* and *offline* mode. The latter provides a tool to explore previously recorded data for research purposes, such as evaluating the impact of different filtering methods, or classification algorithms.

The following descriptions explain both the general components of OFAC and the specific implementation used in our studies.

3.1 Data acquisition and storage

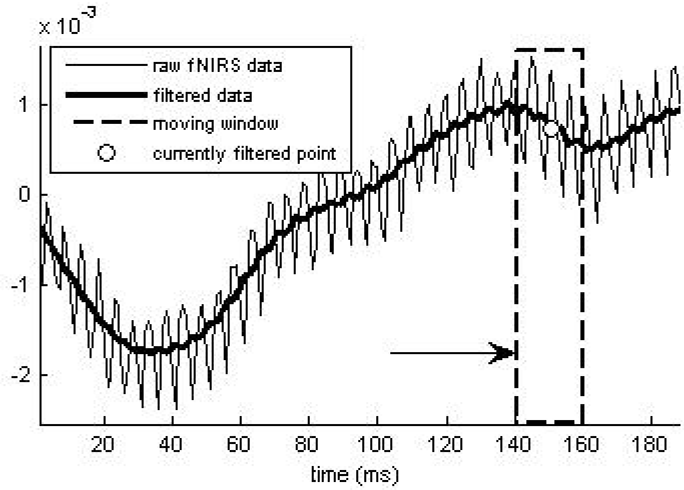
The current system received two data streams: event markers from the application and raw brain signals. It stores the data as it comes in a database, preventing data loss should the system have a major malfunction.

The synchronisation of the system and the different data sources are done with a timestamp of the fNIRS data as soon as it comes in. The application event marker, read immediately after (if any), is time stamped with the raw data time.

3.2 Signal processing

Our work aims at reproducing the procedures used offline in previous work (e.g. Girouard et al., 2009), adapting them for function suitable for real time input to a user interface. We apply a moving average, removing the high frequency noise (Figure 3). The filtered data is converted into oxygenated haemoglobin and deoxygenated haemoglobin concentrations using the modified Beer-Lambert law (Chance et al., 1998).

Figure 3 Moving window of 19 points



3.3 Feature generation and classification

To mimic the procedure used in the previous analysis, we implemented the machine learning technique called sequence classification (Dietterich, 2002). Sequence classification uses each data point of an entire sequence as a feature and applies a label to it. In our case, a sequence is a trial of one task (30 sec of data). Because fNIRS' data is multivariate (multiple sources of light, and 2 haemoglobin types), we classify each channel of data individually first. Then, each classifier votes for the example's label to combine the results of all these classifications. We used a weighted voting technique that takes into account the probability distribution of each example by each classifier.

OFAC calls Weka (Hall et al., 2009) to perform the training and testing of examples created using sequence classification. Using Weka gives OFAC access to a large library of classification algorithms. When enough training data has been accumulated, we first call the classification algorithm to obtain the classifier, and then this classifier is used to test the examples as they come in. To allow flexibility in the analysis, the system can test examples one at a time, or in groups of points.

We evaluate the system's capabilities with two studies. Our first study compares online and offline analyses using a previous study. The second study is a proof-of-concept evaluation that uses the online features of OFAC to classify tasks in real time, and adapts an interface.

4 'Real time' analysis of a previous study

As a first evaluation of the OFAC system, we chose to run the real time analysis on previously recorded data to compare classification results with those obtained offline. To test data from a prior experiment, we used the *offline mode* in the OFAC system which loads previously recorded data and feeds the main loop one line at a time.

We selected a study differentiating levels of a computer game through brain activity, as it showed promising classification results, in particular when comparing rest and play (average accuracy of 94%) (Girouard et al., 2009). Girouard et al. collected data from nine healthy participants as they played two levels of the game Pacman (Namco, Japan) while their brain activity was measured using fNIRS. Participants completed ten sets of two trials (one in each difficulty level) over a 20 min period. In each trial, participants played the game for a period of 30 sec and rested for 30 sec to allow their brain to return to baseline.

They performed two analyses, a classic statistical analysis to establish the differences between conditions, and a more novel task classification to demonstrate the possibility of using this data in a real time adaptive system (though the analysis was done offline). We focus on their offline machine learning analysis, done using sequence classification with k-nearest-neighbour (kNN), $k = 3$, and 10-fold cross validation. The results of both analyses show that we can discriminate well between participants playing or resting (average accuracy of 94.4%, SD 3.7%).

While the design of the offline analysis is typical and correct, it cannot be directly translated to an online analysis. For instance, the previous machine learning analyses used cross validation, which is not possible in real time as it requires the whole data set. Instead, we used the first sets of trials to be the training data, and tested on the rest as to replicate data that comes in. The filtering algorithm suffers from the same problem, and we chose a method that only requires a partial dataset (moving window). The rest of the pre-processing is identical: we reimplemented the sequence classification technique to work in real time, and used kNN ($k = 3$) for the analysis.

We tested the OFAC system's ability to classify between the participants playing and resting. The original data contained 10 examples of each difficulty level: we combined examples of both levels to obtain 20 examples of play and 20 of rest.

4.1 Analysis and results

We varied the amount of examples of each class in the training dataset to evaluate the effect on accuracy and possibly observe a minimum amount of training required to run real time experiments as users should invest only a minimum amount of time for the training algorithm (Krepki et al., 2007). We compared accuracy using training sets containing the first 2, 4, 6, 8, 10, 12, 14, 16 and 18 examples of each class (play or rest). We also evaluated multiple filtering windows to select a balance between a stronger filter and a minimum delay in the analysis stream, using filtering windows of size 1, 9, 19, 29, 39 and 49.

Before comparing the average values to previous results, we applied a 2-way repeated measures analysis of variance on *size of training set* and *length of filtering window* to see which factors were significant in the analysis. As can be seen in Figure 4, the main effect of size of training set has a significant effect on the accuracy of the classification ($F(8, 64) = 9.500$, $p = 0.000$). The group containing higher accuracies is constituted of training sets of 12 or more examples of each class (12, 14, 16 or 18 examples). The main effect of length of filtering window was neither significant nor was the interaction of these two factors.

We conclude that a minimum training set of 12 examples of each class was necessary to obtain meaningful classification results with this dataset. We also observe that every filtering window tested yielded similar results, and determine that the data is well classified with even a small filtering window which requires a smaller delay in processing.

To compare the current results with those obtained with the offline analysis, we averaged the accuracy results obtained with training set of 12 examples and more, and all filtering windows. Over all participants, the classification accuracy obtained with the real time analysis is 82.0% (SD 17.3%), while the one obtained with the offline analysis is 94.4% (SD 3.7%), a decrease of 12.4%. Figure 5 shows the average accuracies obtained with both analyses, for each participant. We cluster the participants by their difference in accuracies: the first group contains participants with a real time accuracy equal or higher than offline, the second group shows lower accuracy (difference within 15%) and the third group with participants displaying a very low real time accuracy.

The results are promising: with the exception of two participants (s01 and s06), all results show accuracies over 80%, and differences between the two analyses of less than 15%. The real time results from three participants even equalled or surpassed offline accuracy. However, the new analysis is, as predicted, performing worse than the offline analysis. However, we consider that this decrease in performance is outshined by the main advantages of the analysis: classifying in real time, and the ability to reuse this information to adapt the interface. Therefore, we find the real time analysis of a previously recorded dataset a success.

Figure 4 The first 12 examples (or more) in the training set produces a stable average accuracy of approximately 82%

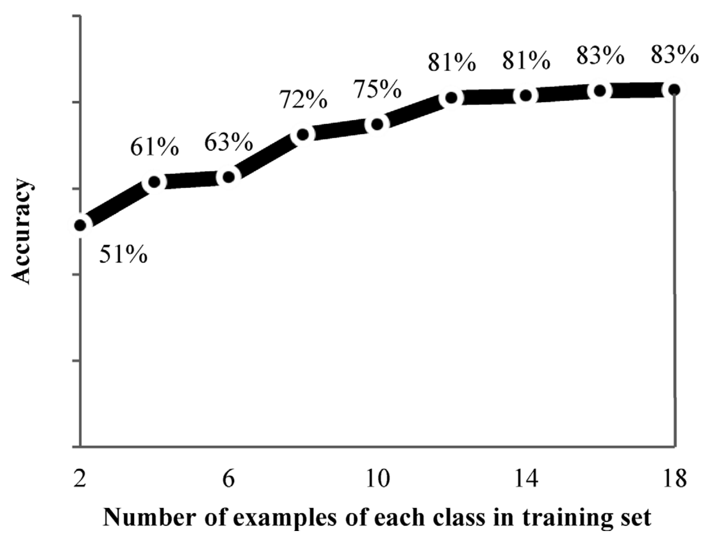
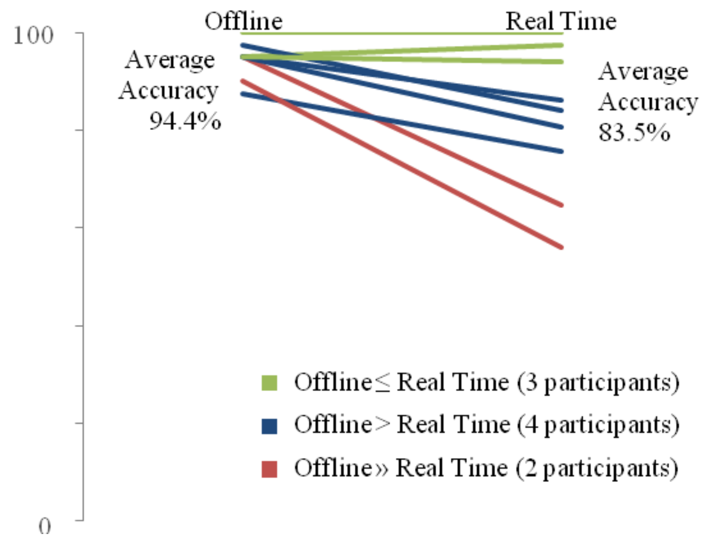


Figure 5 Comparing the real time and offline classification accuracy for each participant (see online version for colours)



The challenges when transforming offline analyses so they are performed in real time resides mostly with the fact that the full dataset is not equally available. In offline analyses, the whole dataset is available at once, while in an online system, we only have the amount of data collected so far, reducing the total information available. This element affects mostly two components of systems, the data filtering and the classification. Most of the differences in accuracy observed between offline and online analysis of the same data should be explained by this.

5 Real time task classification and adaptation

Having shown the classification capabilities of the OFAC system, we designed a simple proof-of-concept experiment to test the true online features of the OFAC system. Our experiment has two goals:

1 to test if we can indeed process and classify in real time

2 to demonstrate a simple interface adaptation.

To achieve these, we will distinguish between two tasks, and change the background music of the interface according to the predicted task.

Given our relative success at classifying activated periods vs. rest periods, we selected tasks that would yield these two different types of signal. The first task consists of showing videos to the participant, while the second task has them play a short game of Tetris. Based on previous work, we expect the game task to activate the anterior prefrontal cortex (aPFC) (Girouard et al., 2009; Saito et al., 2007), and the video task to deactivate it, if the videos are neutral or pleasant, and calm (Leon-Carrion et al., 2007).

With this experimental design, we also hope to replace the typical rest task with an engaging task that will not activate the aPFC. We believe that using a video task instead of rest provides a better, more realistic HCI task than watching a grey screen.

We base our task selection primarily on four studies. León-Carrión et al. (2007) showed that non-arousing videos of neutral or positive valence show little activation in the prefrontal cortex (PFC) in a region encompassing both the aPFC and the dorsolateral PFC, both during and after the stimuli. In addition, Phan et al. (2003) studied the effect of emotional arousal with pictures on the medial PFC with the use of fMRI. The activation found was located in areas too deep for the fNIRS probes to sense, and we do not expect this activation to be measured. Furthermore, pilot studies reproducing the latter work with fNIRS confirmed this statement. Based on those studies, we predict that a video task may be a suitable replacement for the typical rest task.

On the contrary, Girouard et al. (2009) showed activation of the aPFC during a game of Pacman, and suggested the possibility of a higher activation with a more intellectually demanding game such as Tetris, extrapolating their results with those of Siato et al. (2007).

Music is often present when using computers, either related to the current task (such as when playing games), or in the background (Day et al., 2009). Background music can play two different roles in regards to the user's attention: a distracter or an arousal inducer. Day et al. (2009) found that participants were more successful at decision making when the tempo of background music was faster, showing that faster tempo is more beneficial for harder tasks, acting as an inducer. In addition, Wakatsuki et al. (2009) showed using fNIRS no PFC effect for music at low volume, while participants concentrating on music at high volume deactivated their PFC. We deduce from both studies that low volume background music at faster tempo might help the performance of a gaming task, and that a slower tempo should not impact the video task.

5.1 Participants

This study included 10 healthy volunteers (5 females), between the ages of 18 and 32 (mean 25.8, SD 5.8). All participants were right-handed, had normal or corrected-to-normal vision, with no history of brain injury. Informed consent was obtained for all participants. This experiment was approved by the university's internal review board.

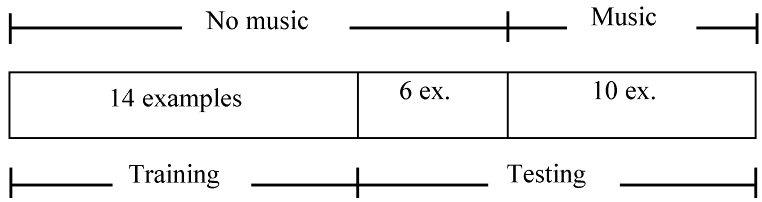
5.2 Protocol and analysis settings

The session contained a total of 30 sets of two tasks (video and Tetris). Ordering of sets was randomised for each participant. Each stimulus was presented after a 3 sec fixation point. A minute of baseline at the beginning of the session allowed the user's brain to get to a rested state.

At the end of the session, participants answered a questionnaire pertaining to their experience with the tasks. Participants also rated the scenes using the same protocol as the one to select the videos.

To achieve our two goals, we divided the experiment into three parts: a training phase and two testing phases – one without music and one with adapting background music (Figure 6). To participants, part one and two are identical (no music).

Figure 6 Experimental protocol and classification periods



While we hypothesised that background music or its adaptation had no impact on the brain data, we tested some of the examples without music. Through pilot participants, we identified that 14 examples of each class was required to obtain significant classification accuracies, both to classify the no music and the music examples. This correlates with results obtained with the previous study. Of the remaining 16 examples, ten were assigned to the music condition (six to the no music condition), because we are more interested showing the adapting interface.

We applied the signal processing method described earlier, with a moving average window of 9. We are using sequence classification with the support vector machine classification algorithm.

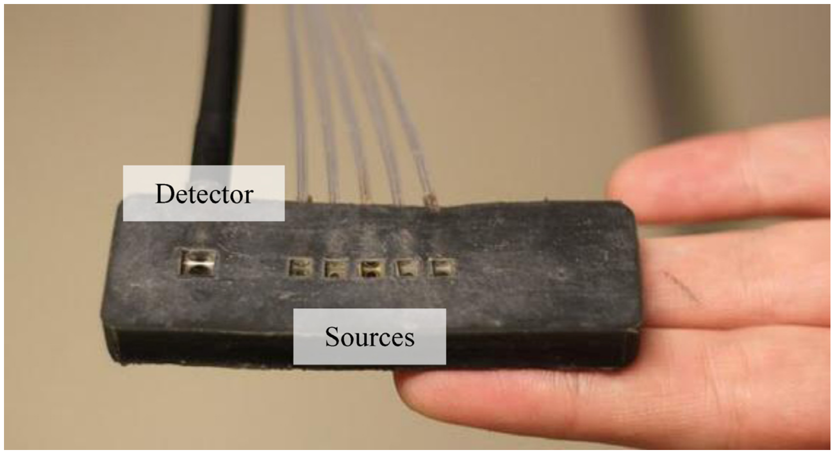
5.3 Stimuli

For the video stimuli, we chose 30 clips that would fit the emotional criterion. The selected clips were 30 sec in length, without sound, and containing mostly nature scenes (beach, forest, trees, streams and clouds). We preselected the scenes by asking an independent group of pilot participants (total of 12 participants) representative of our targeted experimental participants to rate a larger selection of scenes using the self-assessment manikin (SAM) (Bradley and Lang, 1994). SAM measures the dimensions of pleasure (i.e. positive or negative) and arousal (i.e. calming or exciting) using a visual scale from 1 to 9. From those results, we chose a coherent group of scenes that scored below 5 on the arousing scale, and above 3 on the valence scale, using the same scale limits as León-Carrión et al. (2007).

For the game task, we used the Tetris Bean implementation of the classic game (Clee, 2002), which uses arrows and keys to control the falling blocks. We limited the game play to 30 sec, to mimic the video's length. We increased the speed of the falling blocks and difficulty of the game to ensure strong brain activation, and provided a preview of the block to come, to add planning to the mental task. Participants practiced a few games to familiarise themselves with this home-made Tetris version before the real experiment.

During the last part of the experiment, the application played quiet, continuous background music that changed according to the predicted task. If the predicted task indicated that the user was watching a relaxing video, the system would play slower tempo music, and it would play faster tempo music for the game task. The transition between the two was gradual. The slow music kept playing between tasks as to not interrupt the flow of the user, and the faster music was overlaid when Tetris was the predicted task. Both pieces of music were classical piano, intended not to cause a strong reaction among participants.

Figure 7 Small flat fNIRS sensors i.e. placed on participant's forehead (see online version for colours)



5.4 Equipment

We acquired fNIRS data using an OxiplexTS, from ISS, Inc. (Champaign, IL). Our setup is comprised of two small flat sensors (see Figure 7), placed in the middle of the forehead. We selected the sources located at 2.5 and 3 cm away from the detector for data collection as they reach deeper into the cortex. Each source emits two wavelengths (690 and 830 nm), with a sampling rate of 6.25 Hz.

5.5 Results

We analysed the results of this proof-of-concept experiment with three questions in mind.

1 Are the classification results obtained meaningful?

2 How did the adaptation affect the user's performance?

3 Did the background music have a positive effect on the tasks performed?

5.6 Behavioural results

Clips used obtained an average valence rating of 5.7 (SD of 1.6), and an arousal rating of 3.4 (SD of 2.3). This confirms that the selected videos were calm and positive, as required.

We performed two *t*-test on the Tetris games performance: first to determine a presence of a learning effect, and second to compare the scores with respect to music adaptation. To observe if a learning effect is present, we compared the first ten trials (trials 1–10) to the second ten trials of Tetris (trials 11–20), both are without music. We found a statistical significance between the two groups ($p < 0.001$), showing a strong learning effect, with higher performance for the second set of games. Participants completed an average of 0.85 lines per game for the first trials, compared to 1.22 lines for the second set of trials. We also compared the second set of trials with no music with the trials with music (trials 21–30), eliminating the first set of games which contain a learning effect. We observe no statistical difference between the performances with music playing, which we attribute to a neutral impact of the background music.

5.7 Classification results

The real time classification shows a high accuracy when distinguishing between our two types of tasks: watching a video or playing a game of Tetris. We present the results by averaging them into three groups of examples:

1 examples with no background music

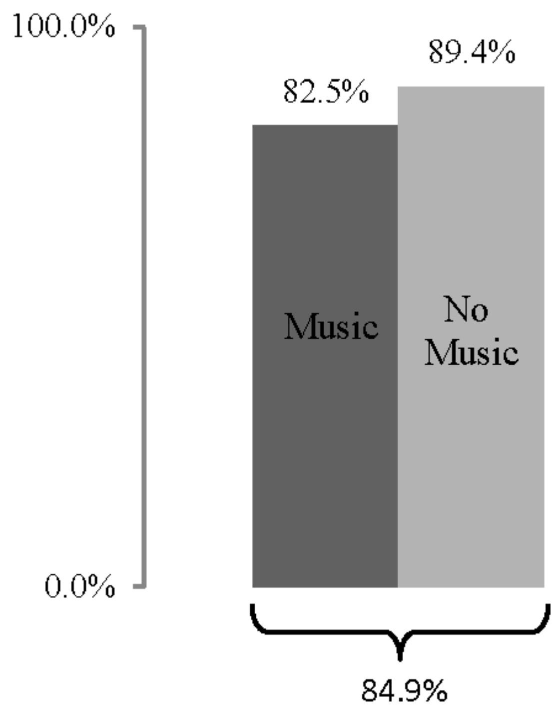
2 examples with background music

3 all examples (a combination of 1 and 2).

Figure 8 illustrates the results. When averaging the examples containing no music, the accuracy is 89.4% (SD 8.8%), while the results with background music averages to 82.5% (SD 8.1%). A *t*-test showed no significance in the difference between the two groups. The overall real time accuracy is 84.9% (SD 6.9%).

Individually, the classification accuracy for all examples varied between 71% and 97%. Three participants rate equal or above 90%, meaning that out of 32 classified examples, the algorithm got less than 3 wrong.

Figure 8 Accuracy results for real time classification of two tasks



5.8 Subjective survey results

Our exit survey revealed that all participants noticed music playing when performing both tasks, but that they only paid attention to it occasionally. Some participants focused on it when the videos came on while others *when the music became more fast-paced*.

Their interest in the music rated 2.9 on average, on a five-point scale, indicated a neutral opinion, which confirms our music choices. They all noticed classical piano, and observed the difference in tempo.

They find the music played on a regular pattern. The slow paced music made them feel relaxed, calm and occasionally bored, while the faster pace music was entertaining and exciting. Two participants noticed the presence of two tracks that were overlaid, and found it distracting. A better integration of two tracks (actually change between them, instead of overlaying them) would have improved their perception of that music. We do not believe, however, that this had a significant impact on their experience.

Most participants noticed that the faster music was associated with the Tetris game. However, the participant with the lowest accuracy during the music examples (65%) noted that 'it did not seem like the changes from one to the other were on a regular pattern', which seems reasonable considering the low accuracy achieved.

They had a varied perception of the effect of the music on their performance. Half of the participants indicated that the music had no effect, while two said that it hindered, and three mentioned it helped their performance. They indicated that the positive influence was because they *were more relaxed*, and that it gave them *some rhythm*, while negative influence was because *it grabbed [their] attention* or made them *stressed*.

5.9 Discussion

Our results demonstrate the validity and reliability of the OFAC system. Not only did the system operate without fault when processing and classifying the data, but also the results obtained are very encouraging. Using a simple classification algorithm, we have achieved moderate to high accuracies, of up to 97% with some participants. This compares well to Abdelnour and Huppert (2009) who achieved 79% when distinguishing finger tapping using a more complex system. With the OFAC system, we have created the first working real time passive BCI with fNIRS. We encourage other researchers to use our system with their experiments or interfaces with this demonstration.

Our task selection was successful: the distinction between Tetris and video watching was strong. Furthermore, the video task replace the ubiquitous rest task. This conclusion has many design implications for fNIRS' measured tasks, as experiments can move closer to real world scenarios.

While they are not statistically significant, the small difference in accuracy between the music and non-music conditions could be related to brain processes. For instance, the training data was not obtained with background music playing, which might not have caught brain processes differences should they have been present. However, given previous work and the analysis, we attribute the core of the differences to technical issues. Brain signals vary in time, even for the same task, due in part to the presence of a trend, meaning that the further from the training period each classified example is, the less accurate the algorithm is likely to be. Though the current protocol does well to answer our research goals, further studies containing adaptation should contain the adapted element in the training data and be counterbalanced.

We also found that successful adaptation had mostly a neutral and positive impact on the user satisfaction. Those results are inline with Kallinen (2004), who found that music listening prompted higher overall user satisfaction and immersion, less boredom and more pleasure. However, low accuracies lead to higher inconsistencies in the adaptation, which can have negative impact on the performance of the participants.

There are many issues to consider when designing BCIs. Speed and accuracy have a large impact in the design of the integration of such new input into interfaces (Coyle et al., 2007). We believe that user satisfaction is at least as important, especially for indirect, or passive, BCI, as the main goal is not always to increase productivity.

6 Conclusion

This paper describes the OFAC system, our new, real time fNIRS analysis and classification system, and demonstrates through two studies the validity, reliability and potential of the system. Our first evaluation compares a previous offline analysis with our real time analysis. Results show a decrease of 12% in classification accuracy (94%–82%), and that a minimum of 12 examples of each class is required to obtain a stable accuracy. We consider this decrease in performance to be outweighed by the main advantage of the analysis, classifying in real time, and the ability to reuse this information to adapt the interface.

The second study demonstrates the online features of OFAC: its ability to record, process, classify cognitive state signals and adapt simple interfaces in real time. We selected two tasks that would activate and deactivate the PFC, respectively: playing a game of Tetris and showing calm videos. In addition, background music varies according to the predicted task: slower music for relaxing videos, and faster tempo for the game task. We evaluated this system through classification accuracy (average accuracy of 85%), as well as using user satisfaction of the adaptation.

6.1 Future work

Two main avenues of future work can be proposed from this paper: improving the OFAC system, and using it.

First, OFAC could benefit from more sophisticated signal processing and classification techniques. Additional data analysis could further resolve the temporal dynamics of classification efficacy, such as detecting workload changes within the first 2, 5, or 10 sec instead of 30. Automatic determination of the amount of training data required might optimise the training period for every participant and improve accuracy. Online adaptation of the classifier would also have a positive effect of the classification accuracy. Finally, we hope to accumulate a large collection of OFAC components so the system can benefit a large research community.

The main work to be done remains to build adaptive user interfaces using the system, and determining how to evaluate them.

Acknowledgements

The authors would like to thank our colleagues and alumni in the HCI research group at Tufts University, in particular Krysta Chauncey, Evan Peck, Juan Carlos Montemayor Elosua, Doug Weaver, Margarita Parasi, Francine Lalooses, Wyatt Newport, Leanne Hirshfield and Rebecca Gulotta. We thank Sergio Fantini, Angelo Sassaroli, Yunjie Tong from the Biomedical Engineering Department at Tufts for their collaboration; Michel Beaudoin-Lafon, Wendy Mackay and the In|Situ| research group and Desney Tan and Dan Morris at Microsoft Research for their helpful inputs and encouragement. We thank the National Science Foundation (Grant Nos. IIS-0713506 and IIS-0414389), the US Air Force Research Laboratory, the Natural Sciences and Engineering Research Council of Canada, the US Army Natick Soldier Research, Development and Engineering Center for support of this research. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of these organisations.

References

Abdelnour, A.F. and Huppert, T. (2009) 'Real-time imaging of human brain function by near-infrared spectroscopy using an adaptive general linear model', *NeuroImage*, Vol. 46, pp.133–143.

Bradley, M.M. and Lang, P.J. (1994) 'Measuring emotion: the self-assessment manikin and the semantic differential', *Journal of Behavior Therapy and Experimental Psychiatry*, Vol. 25, pp.49–59.

Bunce, S.C., Izzetoglu, M., Izzetoglu, K., Onaral, B. and Pourrezaei, K. (2006) 'Functional near infrared spectroscopy: an emerging neuroimaging modality', *IEEE Engineering in Medicine and Biology Magazine*, Special Issue on Clinical Neuroengineering, Vol. 25, pp.54–62.

Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., Li, C., Murray, T., Ovetsky, Y., Pidikiti, D. and Thomas, R. (1998) 'A novel method for fast imaging of brain function, non-invasively, with light', *Optics Express*, Vol. 2, pp.411–423.

Clee, S. (2002) *Tetris Bean*. Retrieved on 24 January 2010 from http://www.ibm.com/developerworks/java/library/j-tetris/.

Coffey, E.B.J., Brouwer, A-M., Wilschut, E.S. and Van Erp, J.B.F. (2010) 'Brain–machine interfaces in space: using spontaneous rather than intentionally generated brain signals', *Acta Astronautica*, Vol. 67, pp.1–11.

Coyle, S.M., Ward, T.E. and Markham, C.M. (2007) 'Brain–computer interface using a simplified functional near-infrared spectroscopy system', *Journal of Neural Engineering*, Vol. 4, p.219.

Cutrell, E. and Tan, D.S. (2008) 'BCI for passive input in HCI', *ACM CHI'08 Conference on Human Factors in Computing Systems Workshop on Brain–Computer Interfaces for HCI and Games.*

Day, R-F., Lin, C-H., Huang, W-H. and Chuang, S-H. (2009) 'Effects of music tempo and task difficulty on multi-attribute decision-making: an eye-tracking approach', *Computers in Human Behavior*, Vol. 25, pp.130–143.

Delorme, A., Kothe, C., Vankov, A., Bigdely-Shamlo, N., Oostenveld, R., Zander, T. and Makeig, S. (2010) 'MATLAB-based tools for BCI research', in D.S. Tan and A. Nijholt (Eds.) *Brain–Computer Interfaces*. London, UK: Springer, pp.241–259.

Dietterich, T.G. (2002) 'Machine learning for sequential data: a review', *Proceedings of the Joint IAPR International Workshop on Structural, Syntactic, and Statistical Pattern Recognition*, London, UK: Springer-Verlag, pp.15–30.

Girouard, A., Solovey, E., Hirshfield, L., Chauncey, K., Sassaroli, A., Fantini, S. and Jacob, R. (2009) 'Distinguishing difficulty levels with non-invasive brain activity measurements', *Human-Computer Interaction – INTERACT 2009*, pp.440–452.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. (2009) 'The WEKA data mining software: an update', *SIGKDD Explorations*, Vol. 11, pp.10–18.

Hirshfield, L.M., Girouard, A., Solovey, E.T., Jacob, R.J.K., Sassaroli, A., Tong, Y. and Fantini, S. (2007) 'Human-computer interaction and brain measurement using functional near-infrared spectroscopy', *Proceedings of the ACM UIST'07 Symposium on User Interface Software and Technology*, ACM Press, Poster Paper.

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K. and Chance, B. (2004) 'Functional optical brain imaging using near-infrared during cognitive tasks', *Int. J. Human-Computer Interaction*, Vol. 17, pp.211–231.

Kallinen, K. (2004) 'The effects of background music on using a pocket computer in a cafeteria: immersion, emotional responses, and social richness of medium', *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, Vienna, Austria, ACM.

Krepki, R., Curio, G., Blankertz, B. and Muller, K-R. (2007) 'Berlin brain–computer interface – the HCI communication channel for discovery', *Int. J. Human–Computer Studies*, Vol. 65, pp.460–477.

Lee, J.C. and Tan, D.S. (2006) 'Using a low-cost electroencephalograph for task classification in HCI research', *Proceedings of the ACM Symposium on User Interface Software and Technology*, pp.81–90.

Leon-Carrion, J., Martin-Rodriguez, J.F., Damas-Lopez, J., Pourrezai, K., Izzetoglu, K., Barroso, Y., Martin, J.M. and Dominguez-Morales, M.R. (2007) 'A lasting post-stimulus activation on dorsolateral prefrontal cortex is produced when processing valence and arousal in visual affective stimuli', *Neuroscience Letters*, Vol. 422, p.147.

Luu, S. and Chau, T. (2009) 'Decoding subjective preference from single-trial near-infrared spectroscopy signals', *Journal of Neural Engineering*, Vol. 6, p.016003.

Mappus, R.L., Venkatesh, G.R., Shastry, C., Israeli, A. and Jackson, M.M. (2009) 'An fNIR based BMI for letter construction using continuous control', *Proceedings of the CHI 2009 Extended Abstracts*, pp.3571–3576.

Phan, K.L., Taylor, S.F., Welsh, R.C., Decker, L.R., Noll, D.C., Nichols, T.E., Britton, J.C. and Liberzon, I. (2003) 'Activation of the medial prefrontal cortex and extended amygdala by individual ratings of emotional arousal: a fMRI study', *Biological Psychiatry*, Vol. 53, pp.211–215.

Saito, K., Mukawa, N. and Saito, M. (2007) 'Brain activity comparison of different-genre video game players', *Proceedings of ICICIC '07 International Conference on Innovative Computing Information and Control*, pp.402–406.

Schlögl, A., Brunner, C., Scherer, R. and Glatz, A. (2007) 'BioSig: an open-source software library for BCI research', in G. Dornhege, J.D.R. Millan, T. Hinterberger, D.J. Mcfarland and M. Klaus-Robert (Eds.), *Towards Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Press, pp.347–358.

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

Solovey, E.T., Girouard, A., Chauncey, K., Hirshfield, L.M., Sassaroli, A., Zheng, F., Fantini, S. and Jacob, R.J.K. (2009) 'Using fNIRS brain sensing in realistic HCI settings: experiments and guidelines', in A. Press (Ed.), *ACM UIST'09 Symposium on User Interface Software and Technology*, pp.157–166.

Wakatsuki, J., Akashi, K. and Uozumi, T. (2009) 'Changes in concentration of oxygenated hemoglobin in the prefrontal cortex while absorbed in listening to music', *International Conference on Biometrics and Kansei Engineering 2009 (ICBAKE 2009)*, pp.195–200.

Wolpaw, J.R., Birbaumer, N., Mcfarland, D.J., Pfurtscheller, G. and Vaughan, T.M. (2002) 'Brain–computer interfaces for communication and control', *Clinical Neurophysiology*, Vol. 113, pp.767–791.

Zander, T., Kothe, C., Jatzev, S. and Gaertner, M. (2010) 'Enhancing human-computer interaction with input from active and passive brain–computer interfaces', in D.S. Tan and A. Nijholt (Eds.), *Brain–Computer Interfaces.* London, UK: Springer, pp.181–199.