Application of near-infrared spectroscopy for discrimination of mental workloads

A. Sassaroli, F. Zheng, M. Coutts, L. H. Hirshfield, A. Girouard, et al.

A. Sassaroli, F. Zheng, M. Coutts, L. H. Hirshfield, A. Girouard, E. T. Solovey, R. J. K. Jacob, Y. Tong, B. deB. Frederick, S. Fantini, "Application of near-infrared spectroscopy for discrimination of mental workloads," Proc. SPIE 7174, Optical Tomography and Spectroscopy of Tissue VIII, 71741H (24 February 2009); doi: 10.1117/12.807737

SPIE.

Event: SPIE BiOS, 2009, San Jose, California, United States

Application of near-infrared spectroscopy for discrimination of mental workloads

A. Sassaroli[a], F. Zheng[a], M. Coutts[a]
L. H. Hirshfield[b], A. Girouard[b], E. T. Solovey[b], R. J. K. Jacob[b],
Y. Tong[c], B. deB. Frederick[c], and S. Fantini[a]

[a]Department of Biomedical Engineering, Tufts University, Medford, MA 02155
[b]Department of Computer Science, Tufts University, Medford, MA 02155
[c]Brain Imaging Center, McLean Hospital, Belmont, MA 02478

ABSTRACT

We show the potential of functional near-infrared spectroscopy for the discrimination of mental workloads during a cognitive task with two different levels of difficulty. Standard data analysis based on filtering and folding average procedures were carried out to locate those source-detector pairs sensitive to the activated cortical regions. On these channels we applied two classification algorithms for the discrimination of mental workloads. Both algorithms showed a high percentage of successful classifications (>80%) on three over a total of four subjects where brain activation was detected. These results are comparable to standard scores found in the field of electroencephalography.
Keywords: functional near-infrared spectroscopy, brain imaging, mental workload, human computer interaction

INTRODUCTION

Current research in human-computer interaction (HCI) explores the measurements of mental activity of computer's users. For example, a typical indirect way to test the design of user interfaces (UI) is the Usability Testing which is a survey carried out at the end of designed experiments. More recently experts in the field of HCI have used more direct and objective methods to measure mental workloads of computer's users. In particular, designers of adaptive user interfaces hope to use an objective measure of mental workloads as a passive input to provide real-time information on the user's state. Electroencephalography (EEG), being one of the first techniques for assessing brain activation, has been used as a direct probe of mental workloads. In EEG, typical classification scores for the discrimination of low against high and medium against high mental workloads can be around 95%, and 80%,[1] respectively. High classification scores (88%) are also possible in EEG for the discrimination of four different levels of mental worklods.[2] However it is well known that EEG is not an imaging technique, therefore due to its poor spatial resolution, it does not provide an accurate mapping of the activated brain regions. Moreover, while the accuracy of EEG to discriminate mental workloads looks promising in a controlled laboratory setting, when it is used for real world applications this accuracy drops significantly. Other drawbacks of EEG are the long preparation time required to place the electrodes (on average 45 minutes) and its high sensitivity to motion and other noise artifacts. More recently functional near infrared spectroscopy (fNIRS) has been used as an alternative tool for the discrimination of mental workloads.[3-5] fNIRS has several advantages with respect to EEG, among which we remind the better spatial resolution (≈ 5 mm), and relative insensitiveness to motion artifacts. In this study, we acquired fNIRS data during a cognitive task characterized by two levels of difficulty. Functional magnetic resonance imaging data were collected on one subject to assess if brain activation was present in the frontal area. We tested five subjects with fNIRS, four of which showed localized activation in the frontal region. Afterwards, on those optical channels showing brain activation, we applied two classification algorithms based on the change of oxy and deoxy-hemoglobin between activation and baseline. The first one is a simple, yet effective, classification algorithm based on the maximum change of oxy- and deoxy-hemoglobin. We also used a more sophisticated classification algorithm, namely dynamic time warping (DTW),[6] which is commonly applied in the field of machine learning. On three subjects we achieved a high classification scores (> 80%) on the activated optical channels. These results show the potential of fNIRS in the field of HCI.

Optical Tomography and Spectroscopy of Tissue VIII, edited by Bruce J. Tromberg, Arjun G. Yodh,
Mamoru Tamura, Eva M. Sevick-Muraca, Robert R. Alfano, Proc. of SPIE Vol. 7174
71741H · © 2009 SPIE · CCC code: 1605-7422/09/$18 · doi: 10.1117/12.807737

METHODS

a) **Experimental Setup**

The optical instrument was a near-infrared optical spectrometer from ISS, Inc., Champaign, IL (OxiplexTS) comprising two detector channels (A and B) and sixteen laser diodes, eight emitting light at 690 nm and eight at 830 nm. All the laser diodes were coupled to optical fibers that delivered the light on the forehead of the human subjects and that were held by a comfortable optical helmet (Fig. 1). The source and detector fibers (3 m long) were 0.4 and 3 mm in core diameter, respectively. They were arranged on the optical helmet according to two circular arrays with the detectors at the centers and the laser sources at four locations of the perimeters (source-detector distance, 3 cm). In each source location two optical fibers were delivering the light at 690 and 830 nm. The two circular arrays were sampling the tissue in the right and left frontal hemispheres. The acquisition rate of the optical system was set to 6.25 Hz, so that we collected one data point every 160 ms. With the same sampling rate we also monitored arterial oxygen saturation and heart rate by means of a pulse oximeter (Nellcor N-200) and also the respiration by means of a chest gauge (Sleepmate/Newlife Technologies). A schematic of the experimental apparatus together with the optical helmet and fibers is shown in Fig. 1.

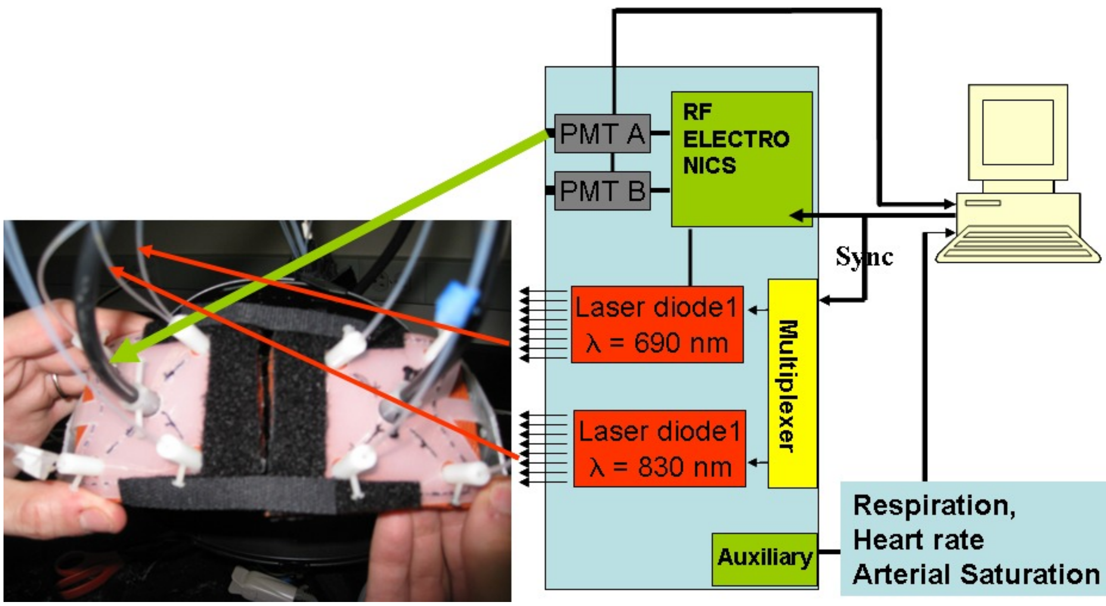


Fig. 1 On the left side the picture shows the optical helmet with the source (transparent) and detector (black) fibers, arranged on two circular arrays. On the right side is shown a schematic of the experimental apparatus.

b) **Experimental protocol and data analysis**

Five healthy subjects completed eighteen cognitive tasks where they viewed five sides of a rotating cube on a computer screen. The side faces of the cube consisted of sections of different colors, as shown in Fig. 2. In the experiment, the cube presented to the subject had either two, or four sections of different colors, and the number of colored sections defined the level of mental workload. During each task, the subjects were asked to count the total number of sections of each color in the five cube sides (lateral and top sides) and their final answer was recorded for verification. One lateral face of the cube was presented to the subjects for 8 s; then the cube underwent a 90° rotation (which lasted 3 s) and stopped when the next lateral face was facing the subject. The process was repeated three times for a total rotation angle of 270° during which four lateral faces of the cube were shown to the subjects. Four extra seconds were allowed at the end of the last rotation for the examinations of the colored sections on the top face. The period of time during which five cube sides were presented to the subjects, which lasted 45 s, defined a workload period. Each workload was followed by 40 s of rest, for a total trial (task/rest) period of 85 s. Three different workload levels (0, 2, 4), were presented randomly six times to the subjects, for a total measurement duration of about 30 minutes (including one minute of initial baseline). In workload level 0, no cube was displayed on the screen and the subjects were asked to clear their minds and not to engage in any particular mental task.

The continuous wave (CW) intensities of the optical data were processed by calculating an average rest value, in order to apply the modified Beer-Lambert law[7] (MBBL) to translate intensity changes into absorption changes. We assumed a differential path-length factor (DPF) of 6.5 and 5.8, at the wavelengths of 690 and 830 nm, respectively.[8] The absorption changes at 690 and 830 nm were de-trended and filtered with a moving average band-pass filter in the frequency range 0.01-0.10 Hz. Afterwards the changes of oxy-hemoglobin and deoxy-hemoglobin concentrations, Δ[HbO] and Δ[Hb], were calculated from the absorption changes for each source-detector pair. Finally, a folding average procedure was applied to the six repetitions of each workload. We studied the statistical significance of the points in the folding average by using a modified *t* test, namely the Dubey/Armitage-Parmar algorithm, at the level of significance $\alpha = 5\%$. This algorithm offers a heuristic method to control the rate of statistical errors of type I (false positives) when the data points are correlated.[9-10]

The first classification algorithm is based on the maximal amplitude of the oxy-hemoglobin ([HbO]) or deoxy-hemoglobin ([Hb]) concentration changes recorded during each 85 s trial period. In particular the changes of [HbO] and [Hb] were calculated with respect to the initial values at the beginning of each workload. The algorithm is applied as follows: from the temporal trend of Δ[HbO] and Δ[Hb], we pick one trial of one workload (query data) and we determine its maximum change. Then we compare the maximum amplitude of this query trial with that of the remaining seventeen trials. Here we consider a 3-nearest neighbor algorithm (3NN), where we consider a classification to be successful if at least two of the nearest neighbor data points belong to the same workload as the query data point. This procedure is repeated eighteen times, for each of the six trials of the three workloads. In this way, we can quantify the classification success rate as the ratio of number of correct classifications to the total number of classifications. The DTW is a more complex classification algorithm which uses all the points of two temporal trends that are compared. The details of this algorithm can be found in ref. 11. Briefly, as in the previous case we pick one query trial and for all the other seventeen trials we determine the minimum warping path which minimizes the distance between each trial and the query. The rule of successful classification is the same 3NN used for the previous algorithm. For this task, we found that workload 2 required a minimal mental effort, therefore we have performed the classification analysis on workloads 0 and 2 joined together against workload 4.

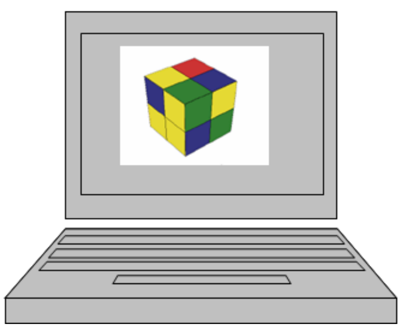


Fig.2 Schematic of computer screen and rotating cube presented to the subjects.

RESULTS

In Fig. 3 is shown the result on the subject who underwent fMRI. As we can see there is a wide region in the frontal area which showed a z score higher than 2.3 ($p < 0.05$). The score was adjusted for multiple comparisons by using Gaussian Random field theory[12] which is implemented in the FMRIB software library (FSL). This experiment confirmed our hypothesis that the cognitive task would activate some regions in the frontal cortex. In Fig. 4 is shown the entire temporal curve of Δ[HbO] measured on one subject, where the gray and the white columns indicate the 45 sec of activation and the 40 sec of rest, respectively, and the number on each column indicate the workload level. From Fig. 4 we can realize that there is a clear change of oxy-hemoglobin (and also deoxy-hemoglobin; trend not shown) during workload 4.

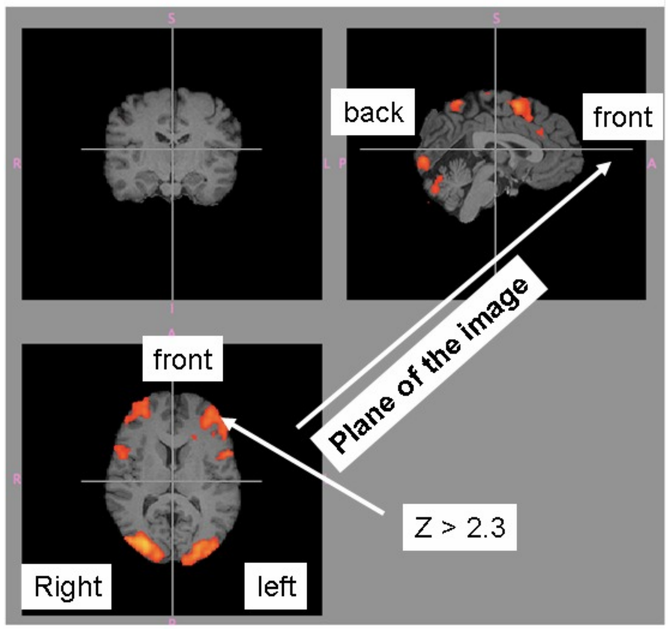


Fig. 3 Activated areas measured on one subject with fMRI. The voxels in bright color showed a z score higher than 2.3, as indicated.

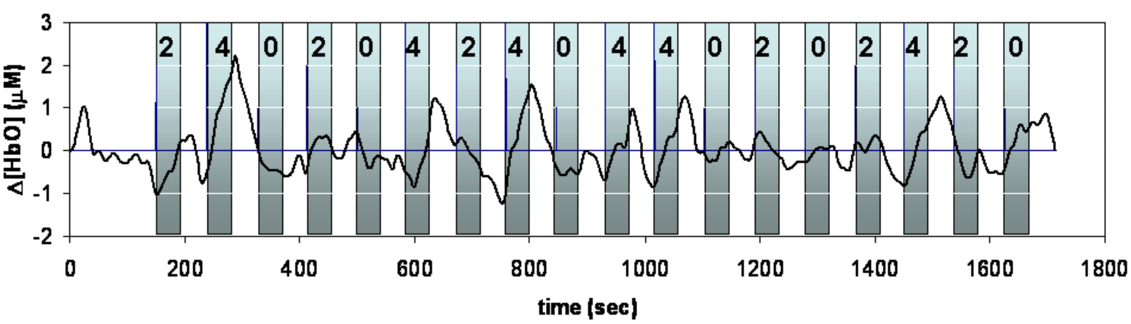
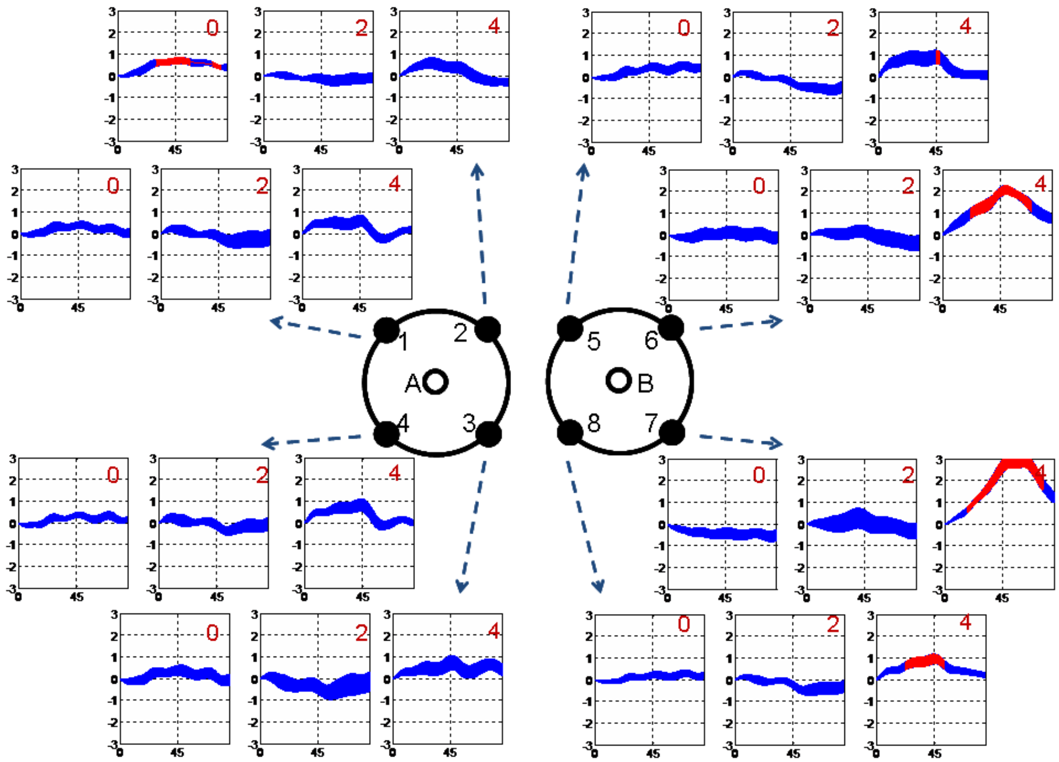
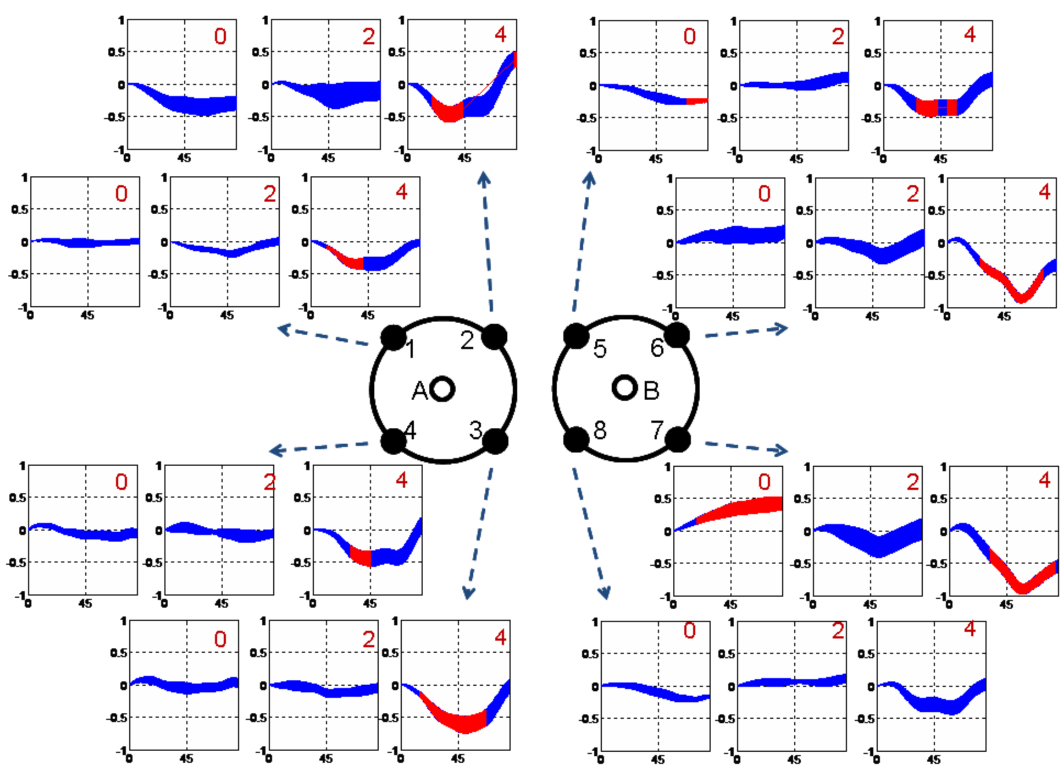


Fig. 4 Change of [HbO] measured on one subject during the entire cognitive experiment. The gray and the white columns indicate the activation and the rest periods, respectively. The numbers on the gray columns indicate the workload levels.

In Fig. 5 are shown the changes of [HbO] (panel (a)) and [Hb] (panel (b)) calculated by using a folding average procedure on the six trials of workload 0, 2 4, at each source-detector pair for the same subject of Fig. 4. The detector A and B are located in the right and in the left hemisphere, respectively. The x axis of the figures is the elapsed time (sec) of one trial including the rotation of the cube (45 sec) and the following rest (40 sec). The y axis represents the change of [HbO] or [Hb] in μM. Each group of three graphs refer to one particular source-detector pair (broken arrow), and to the three different workloads (as specified on the graphs. The statistically significant points in the folding averages are those shown with a lighter color. The thickness of the curves indicates the standard error on the folding average calculated at each time point. We can see that especially two source-detector pairs on the left hemisphere (namely B-6 and B-7), showed a higher change of [HbO] and [Hb].



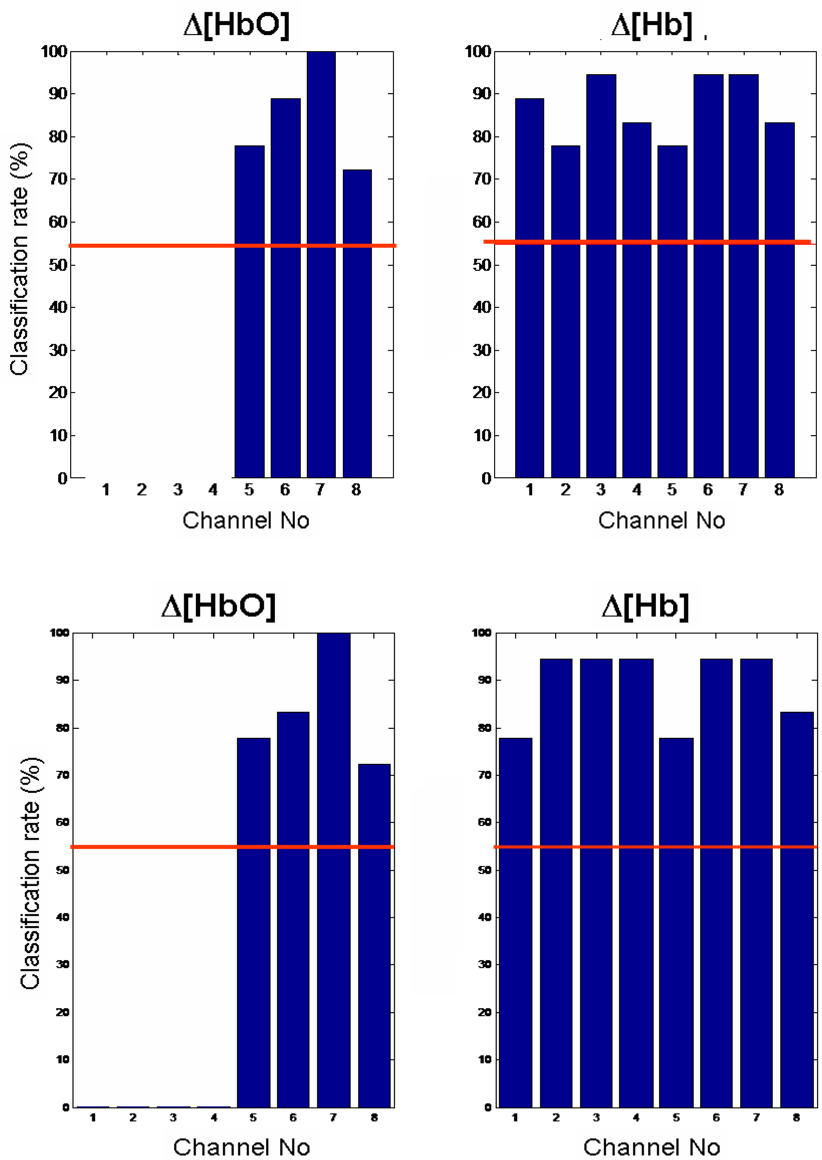
(a)



(b)

Fig.5 Changes of [HbO] (panel (a)) and [Hb] (panel (b)) calculated with a folding average procedure at each source-detector pair and for each workload level (0, 2, 4). The x axis is the elapsed time in one trial, including the workload period (45 sec) and the rest (40 sec). The y axis is the change of hemoglobin concentration in μM. The thickness of the curves is the standard error calculated at each time point. The statistical significant points are indicated with a brighter color. The detector A and B were on the right and left hemisphere, respectively.

The maximum change in both channels is reached about 10 sec after the end of the workload period. This is consistent with the fact that at the end of the workload period the subjects were recollecting the number of sides of each color they saw during the cube rotation. In this subject the temporal trend of the heart rate showed a clear modulation synchronous with workload 4. In our experience on cognitive studies the respiration and heart rate can be confounds inducing systemic hemodynamic changes in the brain and also in the extracerebral regions. Therefore it is of the utmost importance to include also the study of the heart rate, respiration and other systemic physiological parameters in the analysis of the data.



(a)

(b)

Fig. 6 Results of the classification algorithm based on maximal amplitude change (panel (a)) and DTW (panel (b)). In the left and right columns Δ[HbO] and Δ[Hb], respectively, were used for the classifications. The horizontal lines show the chance level (≈ 54%).

For this case we can reasonably argue that the changes we detected were mainly due to brain activation. Firstly the changes were localized as we can see from Fig. 5, while usually systemic changes are spread evenly in all the regions. Secondly we carried out a normalized Fast Fourier Transform (FFT) analysis on the absorption changes at 690 and 830 nm, revealing that each channel was equally sensitive to the arterial component, where the heart rate should be mainly reflected. In Fig. 6 are shown the results of the two classification algorithms, based on the maximal amplitude change (panel (a)) and on the DTW (panel (b)). The classification algorithms were applied only to those source-detector pairs showing some statistically significant points. In the left and right columns are shown the classifications based on changes of [HbO] and [Hb], respectively. The horizontal lines indicate the chance level (~ 54%), obtained with a simple combinatorial analysis. From Fig. 6 we can see that with both classification algorithms we reached a high level of successful classification (≥ 80%) in those channels where activation was detected. In the other two subjects we found similar localized changes in both oxy and deoxy-hemogobin. On these subjects we also found that [Hb] was a better classifier according to the maximal amplitude method, while [HbO] was a better classifier for the DTW method.

CONCLUSIONS

In this work we have shown the possibility of fNIRS to discriminate mental workloads during cognitive tasks. These results are promising for future applications in the field of HCI. It is somehow surprising that a simple classification algorithm, like the one based on maximal amplitude change, has similar performance to the DTW (>80%). However we stress that the classification algorithms should be applied only to those channels were a clear change between activation and rest is found. The preliminary fNIRS data analysis is therefore fundamental in this phase. We have also pointed out the importance of monitoring the physiological systemic parameters (heart rate, respiration, arterial saturation etc.) since they can be confounds during cognitive experiments. In future work we are planning to use the FSL software library used in fMRI for the analysis of fNIRS data, therefore we will apply the general linear model which can include and fit for all the signals recorded during the experiments (including the confounds). The other direction of research will address other machine learning algorithm which can yield higher scores than the DTW.

Acknowledgments:
This research is supported by NSF Award IIS-0713506, NIH R21-DA021817, and NIH CA95885.

REFERENCES

[1] Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., Rush, G., "Monitoring working memory load during computer-based tasks with EEG pattern recognition methods," Human Factors **40**, 79-91 (1998).
[2] Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., and Rao, R. P. 2008, "Feasibility and pragmatics of classifying working memory load with an electroencephalograph," Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems. CHI '08. ACM, New York, NY, 835-844 (2008).
[3] Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., Chance, B., "Functional Optical Brain Imaging Using Near-Infrared during cognitive tasks," International Journal of Human-Computer Interaction **17**, 211-227 (2004).
[4] Son, I. Y., Guhe, M., Gray, W. D., Yazici, B., Schoelles M. J., "Human Performance assessment using fNIR," SPIE Proceedings, **5797**, 158-169 (2005).
[5] Sassaroli, A., Zheng, F., Hirshfield, L. M., Girouard, A., Solovey, E. T., Jacob, R. J. K., Fantini, S., "Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy," Journal of Innovative Optical Health Sciences, **1**, 227-237 (2008).
[6] Sakoe, H., and Chiba, S., "Dynamic programming algorithm optimization for spoken word recognition," IEEE Transactions on Acoustics, Speech and Signal Processing, **26**, 43- 49 (1978)
[7] Sassaroli, A., Fantini, S., "Comment on the modified Beer-Lambert law," Phys. Med. Biol. **49**, N255 (2004).
[8] Duncan, A., Meek, J. H., Clemence, M., Elwell, C. E., Tyszczuk, L., Cope, M., Delpy, D.,"Optical pathlength measurements on adult head, calf and forearm and the head of newborn infants using phase resolved optical spectroscopy," Phys. Med. Biol. **40**, 295 (1995).

[9] Sankoh, A.J., Huque, M.F., Dubey, S.S., “Some comments on frequently used multiple endpoint adjustment methods in clinical trials,” Statistics in Medicine,” **16**, 2529-2542 (1997).
[10] Sassaroli, A., Tong, Y., Benes, C., Fantini, S., “Data analysis and statistical tests for near-infrared functional studies of the brain,” Proceedings. SPIE **6850**, 685008 (2008).
[11] Salvador, S., Chan, P., “FastDTW: Toward accurate dynamic time warping in linear time and space,” KDD workshop on Mining Temporal and Sequential Data,” 70-80 (2004).
[12] Adler, R. J., Taylor, J. E., “Random fields and geometry,” Springer (2008).

Contact information: angelo.sassaroli@tufts.edu