2013

The Use of Functional near Infrared Spectroscopy (FNIRS) to Assess Cognitive Workload of Air Traffic Controllers

Hyun Woo

Ben Willems

Joshua Harrison

Kurtulus Izzetoglu

Sehchang Hah

See next page for additional authors

Authors

Hyun Woo, Ben Willems, Joshua Harrison, Kurtulus Izzetoglu, Sehchang Hah, and Hasan Ayaz

THE USE OF FUNCTIONAL NEAR INFRARED SPECTROSCOPY (FNIRS) TO ASSESS COGNITIVE WORKLOAD OF AIR TRAFFIC CONTROLLERS

Hyun Woo[1], Ben Willems[1], Joshua Harrison[2], Kurtulus Izzetoglu[2], Sehchang Hah[1], & Hasan Ayaz[2]
[1]Federal Aviation Administration
Atlantic City International Airport, NJ
[2]Drexel University
Philadelphia, PA

The performance of a Certified Professional Controller (CPC) can have a critical impact on safety. A specific concern is that a high cognitive load has been associated with performance decrement. Thus, it is important to continuously monitor and accurately assess CPC cognitive load. The purpose of the study was to evaluate the effectiveness of Conflict Resolution Advisory (CRA), automation which provides CPCs with resolutions to avoid conflicts. In this study, we used functional Near Infrared Spectroscopy (fNIRS) to index cognitive workload of 12 CPCs from En Route centers. Results indicate that fNIRS measures were sensitive to air traffic level, but we did not find significant differences across CRA conditions. In addition, we conducted analysis on fNIRS data time-locked to selected events such as clearance commands. With this event-related analysis, we found differences among CRA conditions for different events. The findings indicate that fNIRS can be a potential objective workload measurement tool in the air traffic control domain.

The Federal Aviation Administration (2012) predicts that the amount of air travel will double within two decades. Increased air traffic will lead to increased mental workload for Certified Professional Controllers (CPC). Since performance decrement has been associated with high mental workload (Strayer & Drews, 2006), safety becomes a concern when CPCs' mental workload increases.

It is essential to accurately measure mental workload in order to effectively address safety issues. Mental workload can be indexed using many different methods including secondary task performance measures, subjective measures, and physiological measures. Secondary task measurement and subjective measures can be obtrusive and biased. Physiological measures, however, allow an objective and continuous indexing of mental workload. Neurophysiological measures such as electroencephalogram (EEG) or functional magnetic resonance imaging (fMRI) provide more direct measurements of brain function. Similar to fMRI, functional Near Infrared Spectroscopy (fNIRS) is a hemodynamic measure. fNIRS uses optical signaling to provide biomarkers, such as oxygenated and deoxygenated hemoglobin concentration changes, to index cognitive workload. Unlike fMRI and EEG, fNIRS users' movements are not strictly confined, allowing the device to be a more practical candidate for real-world applications.

In this study, we used fNIRS in an En Route simulation to examine the effectiveness of a decision aid, Conflict Resolution Advisory (CRA). CRA provides controllers with possible resolutions to avoid conflicts in order to reduce their workload during a decision making stage. First, we compared cognitive workload of different air traffic levels. We then compared controller positions and CRA conditions. Finally, we conducted event-related analyses to examine epochs only around decision making timeframes.

Methods

Participants

Twelve En Route CPCs (2 females, mean age 40.7 ± 10.1 years) participated in the study. All participants were active controllers with a mean experience of 14.6 ± 11.3 years and normal or corrected-to-normal vision.

Data Acquisition

We used Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE) to simulate the En Route Automation Modernization (ERAM) system. DESIREE was time-synchronized with other data acquisition tools such as fNIRS and a subjective workload assessment tool. DESIREE simulated two active high altitude sectors, 20 and 22, of Kansas Center (ZKC). In the scenarios, the air traffic steadily increased from 33% to 150% of the Monitor Alert Parameter (MAP) value. DESIREE recorded all user data, including mouse clicks, keyboard inputs, and controller commands.

We used Workload Assessment Keypads (WAK) to measure subjective workload during the experiment. Participants were prompted to input an index of their workload on 10-point scale every two minutes, where "1" reflected low workload and "10" reflected high workload. Participants had maximum of 20 seconds to respond, and after 20 seconds, responses were marked as "missed."

We used two continuous wave fNIRS systems developed at Drexel University (Philadelphia, PA), manufactured and supplied by fNIRS Devices LLC (Potomac, MD; www.fNIRSdevices.com) to index prefrontal cortex activity. There were 16 sensors approximately separated by 2.5 cm with approximately 1.25 cm penetration depth. The device collected data at a rate of 2 Hz with Cognitive Optical Brain Imaging (COBI) Studio software (Drexel University) for data collection.

Procedure

The study lasted three weeks with four controllers per week. The first day was devoted to training, the following three days were devoted to experimental sessions, and the last day was devoted to travel. Each experimental session had four participants where two participants were assigned to Sector 20 while the other two were responsible for Sector 22. Within each sector, each participant was assigned to either radar console (R-side) or radar associate position (D-side). Each of the two participants on Sector 20 wore an fNIRS below an eye-tracking device while participants on Sector 22 wore an electroencephalogram (EEG) device. Participants rotated positions according to a predefined schedule. In this paper, we will limit our analysis and discussion to fNIRS data and participants on Sector 20.

We tested three conditions to examine the effects of CRA automation. "No CRA" condition did not have CRA on both R-side and D-side, "CRA on D" condition had CRA implemented only on D-side, and "CRA on B" condition had CRA implemented on both R-side and D-side. There were three runs of same condition for each day of experimental sessions. The order of CRA conditions were counterbalanced each week.

Preprocessing of fNIRS data

We used Matlab (The MathWorks Inc., MA, USA) software to analyze the fNIRS data. We used a sliding motion artifact removal algorithm (SMAR) and linear phase filter with cutoff frequency of 0.14 Hz to filter the data. This helped to remove any detectable motion artifacts attenuate high frequency noise, and compensate for respiration and cardiac cycle effects (Ayaz, Izzetoglu, Shewokis, & Onaral, 2010;

Ayaz et al., 2012). We excluded any saturated channels and channels with less than 75% of valid data points. Finally, we calculated the blood oxygenation change relative to the first two minutes baseline using the modified Beer-Lambert Law.

Results and Discussion

In our analysis, we used the Jeffrey-Zellner-Siow (JZS) Bays factor *t*-tests via a Web-based program (at pcl.missouri.edu) developed by Rouder, Speckman, Sun, & Morey (2009). Compared to the traditional statistical method of the Null Hypothesis Significance Testing (NHST), Bayesian model comparison allows researchers to state evidence for the null hypothesis. For example, in NHST, researchers can reject the null hypothesis, or fail to reject the null hypothesis, but are not allowed to state the evidence for the null hypothesis. However, the Bayesian model compares probability of the null over the alternative where Bayes Factor (BF) above 1 indicates evidence for the null while BF below 1 indicates evidence for the alternative. It gives the anecdotal, substantial, or strong evidence that two conditions are not different. Benefits of Bayes analysis are described in detail in Dienes (2011).

Air Traffic Level

Previous simulation studies have found relationship between traffic volume and oxygenation levels measured by fNIRS (Ayaz et al., 2011). Similarly, we tested the relationship between oxygenation levels and traffic volume. We defined low, medium, and high traffic by 7 to 13, 14 to 20, and 21 to 27 aircraft, respectively. We used the maximum number of aircraft under control by each participant and the oxygenation change of 10 seconds before and 10 seconds after the aircraft count. There was a significant effect for air traffic levels between low and medium traffic, $t(4)=5.61$, $p<.005$, $BF=0.07$, medium and high traffic, $t(4)=5.54$, $p=.005$, $BF=0.07$, and low and high traffic, $t(4)=8.76$, $p<.001$, $BF=.02$. As shown in Figure 1, results still showed significant results when controller position and CRA condition was taken into account.

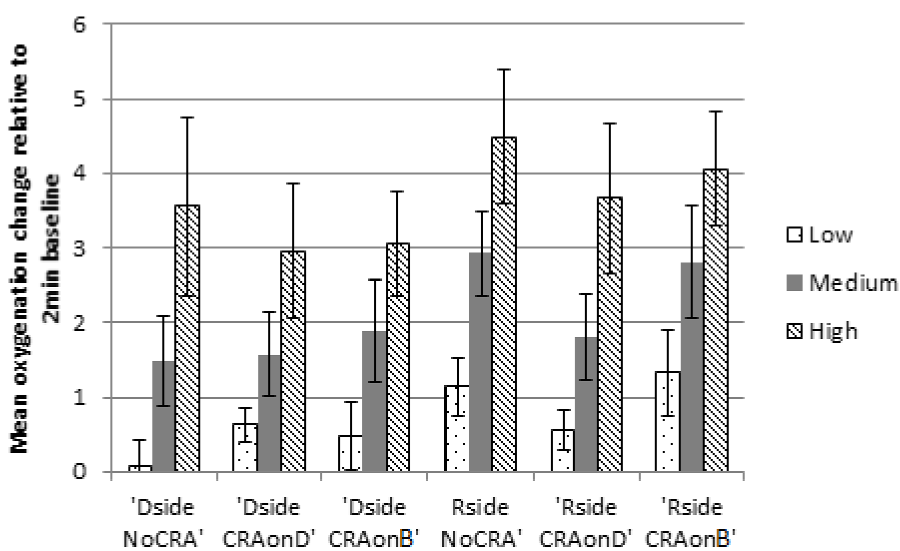

Figure 1. Mean oxygenation change relative to two-minute baseline for three levels of aircraft volume. Error bars represent SEM. With exception of ‘D-side, CRA on D,’ all other conditions show significant oxygenation difference between low vs. medium traffic and low vs. high traffic. Only ‘R-side No CRA’ and ‘R-side CRA on B’ show significant oxygenation differences between medium vs. high traffic.

Controller Position Comparisons

Paired *t*-test and Bayes factor analysis showed anecdotal evidence for no difference between mean oxygenation change for controller positions D-side and R-side for No CRA ($t(5) = 1.01$, p = .36; *BF* = 2.23), CRA on D ($t(7) = 1.55$, p = .17; *BF* = 1.47), and CRA on B ($t(4) = .46$, p = .67; *BF* = 2.93).

CRA Condition Comparisons

Paired *t*-test and Bayes factor analysis showed no support for differences in mean oxygenation change between conditions for both D-side and R-side. Bayes factor (BF) shows substantial evidence for the null ($3.0 < BF < 10.0$) providing evidence that there is no difference between CRA conditions as shown in Table 1.

Table 1.
Summary t-test and Bayesian statistics for CRA condition comparison.

	D-side			R-side		
CRA condition	*No-D*	*No-Both*	*D-Both*	*No-D*	*No-Both*	*D-Both*
p-value	0.68	0.72	0.84	0.67	0.82	0.47
Sample size	7	7	7	7	6	7
t-value	0.44	0.38	0.21	0.45	0.24	0.77
Bayes Factor (*BF*)	3.38	3.46	3.63	3.37	3.39	2.84

We expected to find differences among CRA conditions when we compared CRA conditions for each blocks of aircraft traffic level (low, medium, high). However, we did not find significant difference among condition, and Bayes analysis shows anecdotal evidence for the null ($BF > 1$).

Event-Related Analysis

We were interested in how controllers' workload changed before and after they executed a clearance (i.e., command). We hypothesized that oxygenation level would rise before a clearance issuance because controllers would need to exert mental resources in order to make a clearance. For instance, controllers would evaluate the situation, consider different options, make a decision, and execute a clearance. We chose to look at 10 seconds (s) before and after a clearance because previous studies have shown that hemodynamic response evolves over a 10s to 12s period or less (Miezin, Maccotta, Ollinger, Petersen, & Buckner, 2000; Bunce, Izzetoglu, Izzetoglu, Onaral, & Pourrezaei, 2006; Kim, Richter, & Ugurbil, 1997). Ten seconds after each clearance would have lower oxygenation level than 10s before a clearance because a decision was already made. We examined three common commands that controllers use to control traffic: change in altitude, change in heading, and change in speed. However, we did not see any difference between the times before and after the clearance issuances (see Figure 2). This may be because of the complexity of air traffic control. Within 10s, controllers are executing many commands and performing multiple tasks. We only examined 10s before a command for further analysis because of this lack of difference.

Even though we did not see any overall difference in oxygenation levels across different CRA conditions, we expected to see a difference when we examined the epochs around a command. For example, if CRA effectively assists the controllers in making decisions and lowering their workload, the oxygenation level of the conditions where CRA was available would be lower than the No CRA condition. By examining the epochs around a command, all other noise would be averaged out. As shown in Table 2, we found some oxygenation difference among CRA conditions for different commands. For example, there is a decisive evidence for difference between CRA on D and CRA on Both on D-side controllers when giving altitude change command, and between No CRA and CRA on Both on R-side controllers when giving speed change command.

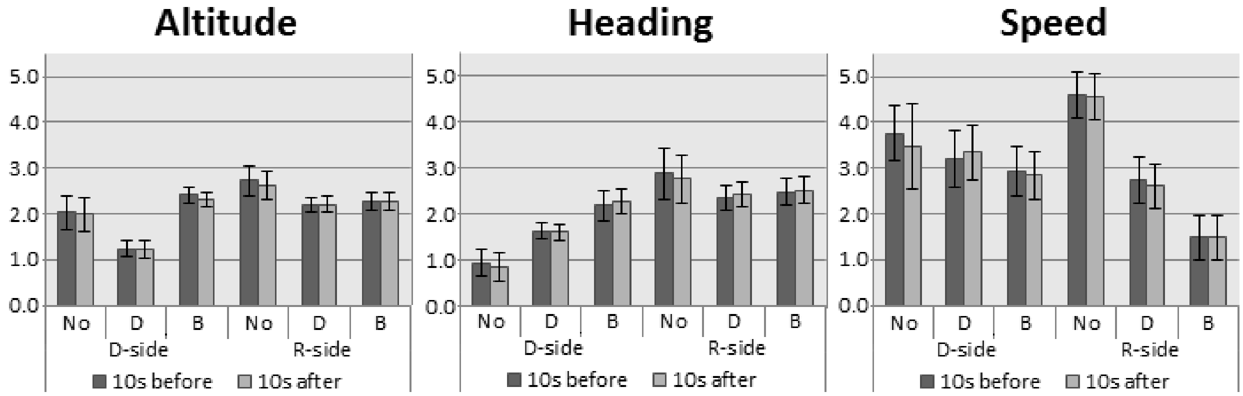

Figure 2. Oxygenation change 10s before and 10s after clearances issued by using the datablock. Error bars represent SEM.

Table 2.
Summary t-test and Bayesian statistics for DB oxygenation 10s before commands

Command	Position	Condition	p-value	df	*t*-value	BF
Altitude	D-side	No-D	0.055	81	1.95	1.41
		No-Both	0.329	81	0.98	5.7+
		D-Both	0.00	106	5.12	0****
	R-side	No-D	0.154	50	1.45	2.66
		No-Both	0.227	50	1.22	3.91+
		D-Both	0.786	78	0.27	8.91+
Heading	D-side	No-D	0.059	23	1.99	0.91
		No-Both	0.007	23	2.98	0.11*
		D-Both	0.141	48	1.5	2.33
	R-side	No-D	0.388	17	0.89	3.35+
		No-Both	0.52	17	0.66	4.06+
		D-Both	0.732	42	0.35	6.08+
Speed	D-side	No-D	0.536	3	0.7	2.35
		No-Both	0.332	3	1.15	1.71
		D-Both	0.724	19	0.36	4.31+
	R-side	No-D	0.015	15	2.75	0.2*
		No-Both	0.00	15	5.22	0****
		D-Both	0.083	15	1.85	0.98

Note. *substantial evidence for difference, **strong evidence for difference, ***very strong evidence for difference, ****decisive evidence for difference, +substantial evidence for *no* difference

Conclusion

This paper demonstrated use of fNIRS to index mental workload in a high fidelity simulation of En Route air traffic environment. Although we did not find any statistical significance of mental workload among CRA conditions, we were able to find differences in air traffic level and event-related analysis. A limitation of this study is that we were not able to conduct an analysis for each of the 16 channels due to small sample size and noisy data. In addition, it should be noted that in our study, the fNIRS indexes mental workload only from the prefrontal cortex. It is a possibility that with high workload, resources may be reallocated to different part of the brain which fNIRS cannot measure. Future analysis should include correlates of fNIRS measures with other measures of mental workload such as subjective measurements and performance measurements. In addition, further research on event related analysis with fNIRS can be an interesting method to examine the pattern of oxygenation related with mental workload. Even though not reported here, future fNIRS studies should conduct an analysis by each channel to examine spatial allocation of oxygenation in prefrontal cortex.

Acknowledgements

The authors would like to thank the Human Factors Branch (ANG-E25), the Concept & Systems Integration Branch (ANG-E14), and their contractors. We also would like to thank Sonia Alvidrez and Rob Bastholm for data collection and comments on the manuscript, the Concept Analysis Branch (ANG-C41) for collecting the fNIRS data, and Drexel University for their help on fNIRS data.

References

Ayaz H., Izzetoglu, M., Shewokis, P., & Onaral, B. (2010). Sliding-window motion artifact rejection for functional near-infrared spectroscopy. Proceedings from IEEE EMBS: *32nd Annual International Conference of the IEEE EMBS.* Buenos Aires, Argentina.

Ayaz, H., Willems, B., BUnce, S., Shewokis, P., Izzetoglu, K., Hah, S., …Onaral, B. (2011). Estimation of cognitive workload during simulated air traffic control using optical brain imaging sensors. In Schmorrow, D. & Fidopiastis, C. (Eds.), *Foundations of Augmented Cognition: Directing the Future of Adaptive Systems* (pp. 549-558). Springer Berlin/Heidelberg.

Ayaz, H., Cakir, M., Izzetoglu, K., Curtin, A., Shewokis, P., Bunce, S., & Onaral, B. (2012). Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging. Proceedings from *Aerospace Conference, 2012 IEEE.* Big Sky, MT.

Bunce, S., Izzetoglu, M., Izzetoglu, K., Onaral, B., & Pourrezaei, K. (2006). Functional near-infrared spectroscopy: An emerging neuroimaging modality. *IEEE Engineering in Medicine and Biology Magazine 25*(4). 54-62.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*(3), 274-290.

Federal Aviation Administration. (2012). *FAA aerospace forecast: Fiscal years 2012-2032.* Retrieved from http://www.faa.gov/data_research/aviation/.

Kim, S., Richter, W., & Ugurbil, K. (1997). Limitations of temporal resolution in functional MRI. *Magnetic Resonance in Medicine, 17.* 631-636.

Miezin, F., Maccotta, L., Ollinger, J., Petersen, S., & Buckner, R. (2000). Characterizing the hemodynamic response: Effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *NeuroImage, 11.* 735-759.

Rouder, J., Speckman, P., Sun, D., & Morey, R. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16* (2). 225-237.

Strayer, D., & Drews, F. (2006). Cell-phone-induced driver distraction. *Current Directions in Psychological Science, 16*(3). 128-131.