Cognitive Workload and Learning Assessment During the Implementation of a Next-Generation Air Traffic Control Technology Using Functional Near-Infrared Spectroscopy

Article *in* IEEE Transactions on Human-Machine Systems · August 2014

DOI: 10.1109/THMS.2014.2319822

Cognitive Workload and Learning Assessment During the Implementation of a Next-Generation Air Traffic Control Technology Using Functional Near-Infrared Spectroscopy

Joshua Harrison, Kurtuluş İzzetoğlu, Hasan Ayaz, *Member, IEEE*, Ben Willems, Sehchang Hah, Ulf Ahlstrom, Hyun Woo, Patricia A. Shewokis, Scott C. Bunce, and Banu Onaral, *Fellow, IEEE*

***Abstract*—Neuroimaging technologies, such as functional near-infrared spectroscopy (fNIR), could provide performance metrics directly from brain-based measures to assess safety and performance of operators in high-risk fields. In this paper, we objectively and subjectively examine the cognitive workload of air traffic control specialists utilizing a next-generation conflict resolution advisory. Credible differences were observed between continuously increasing workload levels that were induced by increasing the number of aircraft under control. In higher aircraft counts, a possible saturation in brain activity was realized in the fNIR data. A learning effect was also analyzed across a three-day/nine-session training period. The difference between Day 1 and Day 2 was credible, while there was a noncredible difference between Day 2 and Day 3. The results presented in this paper indicate some advantages in objective measures of cognitive workload assessment with fNIR cortical imaging over the subjective workload assessment keypad.**

***Index Terms*—Air traffic control, functional near-infrared spectroscopy (fNIR), human performance assessment, near-infrared spectroscopy, optical brain imaging, workload.**

I. Introduction

THE deployment of larger and more complex automation systems has led to an increase in the information-processing load and decision-making demands on aviation personnel, including pilots and air traffic control specialists (ATCSs). While skilled operators have demonstrated the ability to sustain a sufficient level of performance as the difficulty of the task increases, eventually increased workload leads to a decrease in performance [1]. In the case of pilots and ATCSs, any decrease in performance can be very dangerous or even deadly. When implementing new technology with the goal of increasing the safety of air travel, it is imperative to avoid overload on the controller and to allow for adequate training before field implementation. Incorrect implementation of the technology could lead to adverse effects on the controller's performance and ultimately, decreased safety. To evaluate the system and assess the training period required to gain expertise with a new technology, accurate and objective assessment of mental workload is imperative. The emerging wearable functional brain activity monitoring technologies can help to evaluate the cognitive status and capacities of the crew in the cockpit, as well as in ground control stations. Such technologies could provide additional performance metrics directly driven from brain-based measures, which would be an important asset in maintaining safe and effective performance. Additionally, objective brain-based measures may help in preventing operator error and allow for timely intervention through predicting a decline in performance that can arise from either work overload or understimulation [2]–[5].

A. Cognitive Workload Assessment

Previous assessment of the operator's cognitive workload used subjective rating techniques, which often required secondary tests that hindered the operator's ability to perform the task. In an effort to independently measure an individual's cognitive workload and avoid convolution of workload measures by secondary tasks, researchers have begun to explore brain-imaging techniques. Changes in the cognitive workload are known to cause a predictable response in neurophysiological and psychophysiological variables [6]. Electroencephalography (EEG) and event-related brain potentials (ERPs) have been notably strong candidates for accurate objective measures of operator's cognitive workload due to their ability to provide direct measures of central nervous system activity. Changes in an EEG's signal, such as increased power in the beta bandwidth, increased theta activity in the frontal lobe, and the suppression of alpha activity, have been strongly associated with increased task difficulty [7]–[9]. Compared with other devices,

Manuscript received March 29, 2013; revised October 5, 2013 and March 23, 2014; accepted April 8, 2014. Date of publication June 5, 2014; current version date July 11, 2014. This paper was recommended by Associate Editor W. A. Chaovalitwongse.

J. Harrison, K. İzzetoğlu, H. Ayaz, and B. Onaral are with the School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA 19104 USA (e-mail: jlh444@drexel.edu; ki25@drexel.edu; ayaz@drexel.edu; banu.onaral@drexel.edu).

B. Willems, S. Hah, U. Ahlstrom, and H. Woo are with the Federal Aviation Administration W. J. Hughes Technical Center, Atlantic City International Airport, Egg Harbor Township, NJ 08234 USA (e-mail: ben.willems@faa.gov; sehchang.hah@faa.gov; ulf.ahlstrom@faa.gov; hyun.woo@faa.gov).

P. A. Shewokis is with the School of Biomedical Engineering, Science and Health Systems and the Nutrition Sciences Department, College of Nursing and Health Professions, Drexel University, Philadelphia, PA 19104 USA (e-mail: pas38@drexel.edu).

S. C. Bunce is with the Penn State Hershey Medical Center and Penn State College of Medicine, Hershey, PA 17033 USA (e-mail: sbunce@hmc.psu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2014.2319822

EEG possesses many superior attributes for mental workload assessment, including excellent temporal resolution; however, EEG's functionality has been hindered from field deployment by low spatial resolution, long setup time, and susceptibility to motion artifacts.

Over the past decade, functional near-infrared (fNIR) spectroscopy has been gaining recognition for its diversity of applications in brain imaging. Two decades ago, Chance *et al*. [10] first observed oxygenation changes related to brain activity in the prefrontal cortex using an fNIR system during a problem-solving task. More recently, our laboratory demonstrated the ability of fNIR to assess the relationship of brain activity to many different human performance assessment scenarios. Research topics have ranged from cognitive workload monitoring of ATCSs/military warship controllers, task complexity, skill acquisition, problem solving, and learning/training assessment [11]–[15].

fNIR spectroscopy is a field-deployable noninvasive optical brain imaging technology that measures cerebral hemodynamics in response to sensory, motor, or cognitive tasks [11], [16], [17]. The most frequently deployed form of fNIR uses light, which is introduced through the scalp, to measure changes in cerebral blood oxygenation. When neurons are activated at varying levels, there is a relative change in cerebral blood flow to match the demand of the neurons, a phenomenon known as neurovascular coupling. Similarly, as neurons are activated at a higher level, there is a local increase in oxygenated hemoglobin (HbO_2) and a decrease in deoxygenated hemoglobin (HbR) indicating increased brain metabolism. With increased brain activity, the required local oxygen supply is generally overestimated, resulting in an increased level of cerebral blood oxygenation [18]. HbO_2 and HbR's distinctive optical properties in the near-infrared light range allow for the relative change in the concentration of these molecules to be measured independently, using optical methods, during increased brain activation [19]. fNIR also possesses many key aspects in reliable cognitive workload assessment. For instance, fNIR compares favorably with other functional imaging methods, including functional magnetic resonance imaging, [20] and demonstrates a dependable test–retest reliability for task-specific brain activation [21], [22]. fNIR also provides increased spatial localization compared with EEG, on the order of 1 cm^2, and is easily integrated with EEG/ERPs for more robust analysis [23]–[25].

B. Learning Evaluation With Brain Imaging

During the implementation of a new system, it is imperative to determine the necessary training period for controllers to gain expertise on the system before implementing the system in a real-life setting. When presented with a new task, an individual will acquire a level of expertise most efficiently by utilizing the most effective mode of practice and/or spending more time engaged in the task. Natural ability contributes to the pace of the individual's improvement and will determine individual differences in each individual's progression toward obtaining expertise in a specific course of training. The literature pertaining to the effect of practice on the functional neuroanatomy of task performance is rather extensive and complex. The practice and the development of expertise have been studied across a range of motor, visuomotor, perceptual and cognitive tasks, and from various research perspectives. To briefly summarize this literature, four main patterns of practice-related activation change can be distinguished [26]. Practice can lead to 1) an increase in activation in the brain areas involved in task performance, 2) a decrease in those areas, or 3) a functional redistribution of brain activity, in which some initial areas of activation decrease, whereas other initial areas of activation increase, and 4) a functional reorganization of brain activity, i.e., the pattern of activation increases and decreases occur in distinct brain areas, as well as the initial areas.

The majority of studies examining task practice have found decreases in the extent or intensity of activations, particularly in the attention and control areas [26]. This finding is true whether the task is primarily motor (e.g., a golf swing [27]) or primarily cognitive in nature (e.g., the Tower of London problem [28]). Decreases in activation represent a contraction of the neural representation of the stimulus [29] or a more precise functional circuit [30]. This finding provides an important overlap with the literature on expertise. There is considerable evidence that expertise tends to be associated with overall lower brain activity relative to novices, particularly in prefrontal areas (see, e.g., [32]). Both the practice and the development of expertise (the latter of which includes individual differences in ability) typically involve decreased activation across attentional and control areas, freeing these neural resources to attend to other incoming stimuli or task demands. As such, measuring activation in these attentional and control areas relative to task performance can provide an index of the level of expertise. One way to conceptualize this approach is that a relative quantification of the attentional and control resources necessary to perform at a given level can serve as an index of the trainee's neural "reserves," a capacity that can be used to perform effectively under greater situational demands.

This paper introduces research efforts underway to progress fNIR technology toward field applications in aviation, including a study with the Federal Aviation Administration (FAA). In collaboration with FAA's William J. Hughes Technical Center, we explored the impact of the Conflict Resolution Advisory (CRA) on ATCS's behavior and workload over a training course of three days and nine sessions (see Fig. 1).

II. Methods

A. Continuous Wave Functional Near-Infrared Spectroscopy System

Throughout the entire experiment, we collected physiological fNIR recordings and subjective workload assessment ratings. The continuous wave fNIR device utilized for this study, manufactured by fNIR Devices LLC (Potomac, MD, USA; www.fnirdevices.com), was first implemented by Chance *et al.* [31] and further developed at Drexel University, Philadelphia, PA, USA. The device consists of a sensor pad with four light-emitting diode light sources, with peak wavelengths of 730 and 850 nm, 12 detectors, a portable hardware box, and a laptop

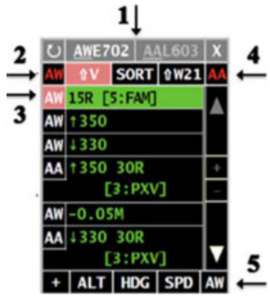

Fig. 1. Representative layout of a search all CRA menu. 1) Advisories are presented for both aircraft involved in the problem: AWE702 ("AW") versus AAL603 ("AA"). The currently selected advisory is for AWE702, and therefore, AWE702 is highlighted. 2) Flight data are available by selecting the aircraft reference (e.g., "AW"). 3) Highest ranked advisory initially turns AWE702 ("AW") 15° to the right. The remaining portion of the advisory ["[5FAM]" is enclosed in brackets, denoting the clearance (a direct to FAM) should be issued at a future time (5 min from now)]. 4) Advisories are initially presented in ranked order, but can be sorted by aircraft by selecting "SORT." The default output mode (e.g., "⇑V" for voice/amendment delivery) can also be changed for each aircraft. 5) Other CRA menus may be assessed for either aircraft, and may be based on a selected trial plan.

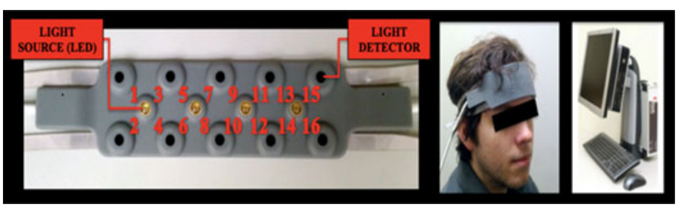

Fig. 2. Design and positioning of a portable fNIR device. (Left) 16-channel portable fNIR sensor pad design with all 16 channels labeled. (Center) Positioning of the fNIR sensor pad on a subject's head. (Right) fNIR data acquisition box and computer system.

computer (see Fig. 2). This source–sensor configuration generates a total of 16 prefrontal cortex measurement locations per reading, with a sampling rate of 2 Hz (see Fig. 2). COBI Studio acquisition software [32] (2010, Drexel University) was used for fNIR data collection.

When near-infrared light with a wavelength of 700–900 nm enters brain tissue, much of the light is scattered, some is absorbed, and a small portion is reflected back to the sensor [33], [34]. Water and tissue do not absorb light very highly in this range; however, deoxy-hemoglobin (HbR) and oxy-hemoglobin (HBO_2) are the main light-absorbing molecules and have distinct spectra within this "optical window," which makes it possible to detect changes in both HbO_2 and HbR concentration through spectroscopy [35], [36]. Utilizing the peak wavelengths that the HbR and HbO_2 chromophores are known to absorb, which are 730 and 850 nm, respectively, it is possible to measure the relative changes of both HbO_2 and HbR and, thus, monitor the brain activity of individuals through fNIR cortical imaging [11], [33], [35]. We utilized a modified Beer Lambert Law, based on these principles and first proposed by Cope *et al.* [33], to quantify the relative changes in the concentration of HbO_2 ($\Delta C_{\mathrm{HbO_2}}$) and HbR ($\Delta C_{\mathrm{HbR}}$) compared with a baseline measurement. From $\Delta C_{\mathrm{HbO_2}}$ and ΔC_{HbR}, we calculated cerebral oxygenation (Oxy) and total blood flow (HbT)

$$\begin{aligned}\mathrm{Oxy} &= \Delta C_{\mathrm{HbO_2}} - \Delta C_{\mathrm{HbR}}\\ \mathrm{HbT} &= \Delta C_{\mathrm{HbO_2}} + \Delta C_{\mathrm{HbR}}.\end{aligned} \tag{1}$$

B. Workload Assessment Keypad

To assess instantaneous perceived workload of the ATCSs, a workload assessment keypad (WAK), with ten numbered buttons, was implemented. The WAK technique is an adaptation of the ATC workload input technique (ATWIT), a technique that was developed to assess instantaneous subjective workload during ATC simulations [37]. ATWIT utilizes a ten-point scale that is anchored in the operational needs of the ATCSs. The low end of the scale (1–2) reflects low workload, where participants can accomplish all their tasks easily with extra time. At input levels 3, 4, and 5, controllers experience increasing levels of moderate workload, where the controller can still finish all tasks; however, the chance of an incomplete task steadily increases with decreasing spare time. At input levels 6, 7, and 8, controllers experience relatively high workload leaving ATCSs time to narrowly finish all essential tasks or leave some unessential tasks unfinished. At input levels 9 and 10 of the WAK scale, participants experience extremely high workload. At this level, ATCSs will most likely only focus on keeping aircraft separated, leaving many essential tasks unfinished.

Prior to each session, participants were reminded of the meaning of each WAK rating and were instructed to indicate their instantaneous workload level by pressing one of ten numbered buttons. The WAK device illuminated every 2 min to prompt participants for the input. The participants were given 20 s to respond while the simulation was in progress. When the ATCSs did not input a response within 20 s, a code for the missing data was recorded.

C. Conflict Resolution Advisory

Given the complexity of the CRA and the concern that controllers may change the way in which they perform their job, consideration has been given to the implementation of the system. ATCSs work in pairs while directing traffic. Within the pair, one controller is referred to as the radar (R-side) controller and the other controller is the data controller (D-side), a position often referred to as the radar assistant. For the implementation of the CRA, three conditions were explored: 1)neither controller had access to the CRA (baseline); 2) D-side only had access to the CRA (D-only); and(3) both controllers had access to the CRA (both). Previous reports indicate that no difference was found between the R-Side and D-Side controllers or the three implementations of the CRA [38]. Given our previous findings, the three CRA conditions and two ATCS sides are analyzed together in this paper.

D. Simulation

1) Simulation Center: The simulation was conducted at the Research Development and Human Factors Laboratory, William

Fig. 3. Air traffic controller's simulation center: Each workstation consisted of a high-resolution (2048 × 2048), radarscope, keyboard, trackball, and direct access keypad.

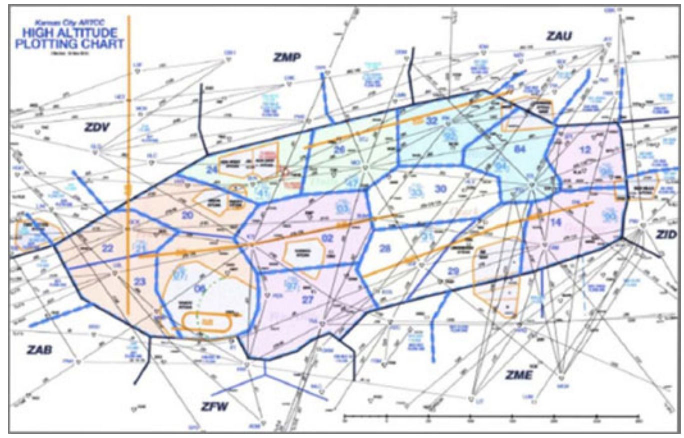

Fig. 4. High altitude sectors of the Kansas Center (ZKC).

J. Hughes Technical Center, Atlantic City International Airport, NJ, USA.

2) Data Collection: In the experiment room, there were two pairs of controller workstations, with fNIR data collected from one pair of controllers (see Fig. 3). The prefrontal cortex of all participants was monitored during the air traffic control simulations with the portable 16-channel continuous wave fNIR system (see Fig. 2). The fNIR device was placed underneath an occulometer device for data collection. Throughout the simulation, the controllers were prompted to give WAK ratings once every 2 min while the scenario was in progress. Similarly, we also collected eye-tracker data on all subjects; however, data quality was poor due to subject discomfort leading to them moving or removing the device, and results will not be presented in this paper. The fNIR recording was synchronized with the traffic scenarios and traffic levels using a custom application implemented to send event markers to the fNIR data acquisition computer via RS232 serial port.

3) Communication Systems: The ATCSs had access to two types of communication system voice (VoiceComm), which allowed for spoken communication, and data (DataComm), which allowed for text-based communication with pilots [39]. We used a VoiceComm system that mimics the operational voice switching and communications system used by ATCSs, allowing for air/ground communication between controllers and simulation pilots, as well as ground/ground communications link between controller participants for intersector communications. The system utilizes Push-To-Talk headsets, as well as handheld and foot-operated switches. To accommodate for previous findings that indicate that cognitive workload levels are different for the different communication systems, 30% of aircraft were equipped with DataComm to standardize communications and simulate possible near-future real airspace communication conditions [11].

4) Traffic Scenario Creation: The Distributed Environment for Simulation, Rapid Engineering, and Experimentation (DESIREE) was used to simulate the en route automation modernization (ERAM) system. The DESIREE emulation of the ERAM system was modified to accommodate the CRA. A target generation facility (TGF) was used to generate radar track and data block information based on stored flight plans. The TGF accepted entries from the simulation pilot workstations and DataComm to control aircraft maneuvers that DESIREE displays on the ATCS's workstations.

The airspace used in the experiment consisted of two active high altitude sectors, i.e., 20 and 22, of the Kansas Center (ZKC; Fig. 4). Traffic scenarios were developed based on samples extracted from the Aircraft Situation Display to Industry (ASDI) feed to ZKC. The traffic was filtered to include only aircraft that crossed a volume of airspace of 300 × 300 nautical miles that included the sectors used in the experiment. The traffic was modified to make the traffic volume steadily increase during each run from a low monitor alert parameter (MAP) to a high MAP. MAP was previously described as the number of aircraft that a sector/airport can accommodate without degraded efficiency during a specific period of time [40], i.e., at 100% MAP volume, the airspace cannot accommodate another aircraft without a decrease in safety and efficiency of the airspace. In Sector 20, a MAP increase from 33% to 100% MAP value results in an aircraft increase from about 6 aircraft in the beginning to 19 aircraft at the end of each session. A similar increase in the MAP value in Sector 20 results in an increase from about 7 to 20 aircraft.

For the initial training day, three sets of 30-min practice scenarios were developed. The participants worked with a low traffic level scenarios first, which started at 33% of the MAP value, built up to about 66% by 15 min, and then remained at that level for the final 15 min of the 30-min scenario. The moderate traffic level training scenario increased from 33% of the MAP value to about 100% for the duration of the scenario. The high traffic training scenarios increased from 33% to 150% of the MAP value. To meet the steady increase of the traffic volume, aircraft were added manually. Realistic call signs, routes, and aircraft types were added to the newly added aircraft by copying the information of the aircraft that traveled to the same route.

A new set of 30-min training scenarios was created for test days. The three training scenarios were created from traffic between 17:45 and 18:45 of September 9, 2011; 18:00 and 19:00 of September 25; and 16:00 and 17:00 of September 26. The traffic levels were modified to make the traffic volume meet

TABLE I
WEEKLY SCHEDULE

Training Day		Testing Days			
Time	**Monday**	**Time**	**Tuesday**	**Wednesday**	**Thursday**
8:00-9:30	Intro	8:00-8:30	Training	Training	Training
9:30-10:00	Intro to Simulator	8:30-9:00	Break	Break	Break
10:00-10:45	Training	9:00-10:00	Test	Test	Test
10:45-11	Break	10:00-10:15	Break	Break	Break
11:00-11:45	Training	10:15-10:45	Training	Training	Training
11:45-12:45	Lunch	10:45-11:15	Break	Break	Break
12:45-13:30	Training	11:15-12:15	Test	Test	Test
13:30-13:45	Break	12:15-13:15	Lunch	Lunch	Lunch
13:45-14:30	Training	13:15-13:45	Training	Training	Training
14:30-15:00	Break	13:45-14:15	Break	Break	Break
15:00-16:00	Training	14:15-15:15	Test	Test	Test
16:00-16:15	Post-Scenario	15:15-15:30	Break	Break	Break
16:15-16:30	Debriefing	15:30-16:30	Debriefing	Debriefing	Debriefing

TABLE II
EXAMPLE RANDOMIZED TEST SCHEDULE ROTATION

Day	Run	CRA Implementation	Position	Sector	
				20 (Eye Tracker & fNIR)	**22 (EEG)**
Tuesday	1	Baseline	R	ATCS 1	ATCS 3
			D	ATCS 2	ATCS 4
Tuesday	2	D-Only	R	ATCS 3	ATCS 4
			D	ATCS 1	ATCS 2
Tuesday	3	Both R&D	R	ATCS 2	ATCS 1
			D	ATCS 3	ATCS 4
Wednesday	4	D-Only	R	ATCS 4	ATCS 2
			D	ATCS 3	ATCS1
Wednesday	5	Both R&D	R	ATCS 1	ATCS 2
			D	ATCS 4	ATCS 3
Wednesday	6	Baseline	R	ATCS 2	ATCS 1
			D	ATCS 4	ATCS 3
Thursday	7	Both R&D	R	ATCS 3	ATCS 1
			D	ATCS 4	ATCS 2
Thursday	8	Baseline	R	ATCS 1	ATCS 3
			D	ATCS 4	ATCS 2
Thursday	9	D-Only	R	ATCS 3	ATCS 2
			D	ATCS 1	ATCS 4

the steady increase from 33% to 100% MAP value during the run. From each of the three scenarios, two additional scenarios were created by changing the call signs of the aircraft in the original scenarios to create a total of nine practice scenarios for test days.

The test scenarios were created from one traffic sample on July 14 between 13:00 and 14:00. Each test scenario lasted for 50 min, in which the traffic volume was increased from 33% to 150% of the MAP value. Eight additional test scenarios were created from changing the call signs in the original scenario. The nine test scenarios were randomly presented in the experiment. All Continental Airline call signs were converted to United Airline call signs because the Continental Airline merged into the United Airline on October 1, 2010, and the Continental Airline call signs changed to the United Airline call signs on November 30, 2011.

5) Simulation Pilots: Four simulation pilots were used for each controlled sector to simulate real ATCS–pilot communication. Each simulation pilot workstation consisted of a computer, keyboard monitor, and communications equipment. The simulation pilot display showed a spatial representation of traffic, a list of assigned aircraft, and a window that displayed incoming DataComm messages.

E. Experimental Protocol

Prior to the study, all participants signed informed consent statements approved by the FAA's Human Subjects Review Board. Twelve certified ATCSs that were previously unfamiliar with ZKC airspace volunteered for the study. The sessions were completed over a period of three weeks with four ATCSs completing the experiment each week.

1) Familiarization With the Airspace: Before beginning the study, participants received one-day extensive training on the airspace, systems, and procedure. During this day, the volunteers participated in five 30-min training scenarios (see Table I). The first training scenario used a low traffic level (33–66% of the MAP value). Other training scenarios included a moderate traffic level (33–100% of the MAP value) and a traffic volume equal to that of the experimental scenario (33–150% of the MAP value).

2) Testing Days: Over a three-day test period, the participants completed a total of nine test sessions (see Table I). Each day, three practice and three test sessions were completed using each of the three CRA implementations: a baseline condition where neither radar nor data-side controllers had access to the CRA, a second condition where the data-side controller only had access to the CRA, and the final condition where both radar and data-side controllers had access to the CRA (see Table II). Each training session was performed for 30 min, with no physiological measurements, at the previously described traffic levels and a 30-min break given between the practice and test scenarios (see Table I). The purpose of the training scenarios was to allow controllers to familiarize themselves with the CRA version that would be available during the test scenario. Each test scenario lasted for 50 min, in which the traffic volume was increased from 33% to 150% of the MAP value.

3) Pseudorandom Subject Rotation: Each experimental session consisted of four participants. Two participants were assigned to Sector 20, while the other two were responsible for Sector 22 (see Table II). Within each sector, participants worked in pairs and were assigned to either R-side or D-side. Both participants on Sector 20 wore an fNIR device below an eye-tracking device, while participants on Sector 22 wore an electroencephalogram (EEG) device. Participants were rotated randomly between the two sectors and ATCS side positions according to a predefined schedule. In this paper, we will limit our analysis and discussion to fNIR data and, thus, participants on Sector 20. All training and test blocks were counterbalanced to minimize the order effects across participants.

F. Data Analysis

Raw light intensity recordings were low-pass filtered with a finite impulse response, linear phase filter with a cutoff frequency of 0.1 Hz, and order of 20 to reduce high-frequency noise. Motion artifacts were removed through visual inspection. Using filtered raw fNIR measures, HbO_2, HbR, Oxy, and HbT were calculated using the modified Beer–Lambert Law and (1).

TABLE III
SAMPLE SIZE (n-VALUES) FOR fNIR BLOCKS

	<13 Aircraft	13-15 Aircraft	16-18 Aircraft	19-21 Aircraft	22-24 Aircraft	>24 Aircraft
Day 1	6	6	6	6	6	3
Day 2	10	10	10	10	10	4
Day 3	9	9	9	8	8	6

TABLE IV
SAMPLE SIZE (n-VALUES) FOR WAK BLOCKS

	<13 Aircraft	13-15 Aircraft	16-18 Aircraft	19-21 Aircraft	22-24 Aircraft	>24 Aircraft
Day 1	14	12	15	14	13	10
Day 2	13	14	13	11	13	12
Day 3	14	14	13	13	12	9

The baseline used occurred in the beginning of each session and corresponded to the time period when the ATCSs were controlling less than ten aircraft. Additionally, time synchronization markers were used to divide each session into six "blocks" of different aircraft levels under ATCS control. The "blocks" began after the baseline and were divided as follows: "<13" aircraft, "13–15" aircraft, "16–18" aircraft, "19–21" aircraft, "22–24" aircraft, and ">24" aircraft. Given the nature of the simulation, with the ATCSs having control over the aircraft's route, each block was never of a defined time period; however, the blocks were relatively evenly spaced in time with each block averaging equal lengths. Within each block, HbO_2, HbR, Oxy, and HbT measurements were averaged, respectively. The fNIR results presented in this paper are determined from the average of Oxy data within each of the aircraft level "blocks." Finally, the WAK ratings were averaged within each aircraft block to determine subjective workload ratings.

A total of 27 sessions were completed on Sector 20 by both the R-Side and D-Side controllers for a total of 54 sessions completed. The final fNIR dataset available for statistical analysis included 25 sessions (see Table III). There were seven test sessions in which fNIR data were not properly recorded due to subjects opting out of wearing the device. WAK data were only analyzed for sessions in which fNIR data were properly recorded. Matching WAK data were not input by the subjects in additional four sessions in which fNIR data were properly recorded. The final dataset of WAK data available for statistical analysis included 43 sessions (see Table IV).

A large amountof fNIR data were visually rejected due to under/oversaturated channels and sensor decoupling caused by subject discomfort induced by the combinationeye tracker/fNIR setup. Data were rejected when the device decoupled from the forehead and shifted causing the baseline to be invalid. Decoupling events occurred when the participant inadvertently shifted the eye tracker/fNIR setup or had to remove the device during the run due to discomfort. If more than 60% of a run was valid, then the run could be used for data analysis, i.e, when the decoupling event occurred in the beginning of the run, we were forced to reject the entire run; however, if the decoupling event occurred after 60% of the run was complete, we saved the beginning of the run while only rejecting the end. While all channels were examined during the data analysis, in this paper, the statistical analysis is focused on fNIR "channel 2," which is near to AF7 in the International 10–20 System, located within the left prefrontal cortex (inferior frontal gyrus). Specifically, for "channel 2," a total of ten sessions of fNIR data were visually rejected due to high motion artifact or under/oversaturated channels, and a total of 12 sessions were rejected due to sensor decoupling caused by subject discomfort induced by the combination eye tracker/fNIR setup.

G. Statistical Analysis

For this project, we used the free software R and JAGS [41]–[43] and adapted software code from Kruschke [44]. For the analysis, we used a Bayesian split-plot analysis of variance (BANOVA) model from Kruschke [44], where one factor is a between-subject factor and a second factor is a within-subject factor. The model involves main effects for factors A and B, an interaction effect for A $\times$ B, and a main effect for subject within level of A. In our analysis, the main between-subject factor A was the Day of the session and the within-subject factor B was the aircraft count under control. Although unconventional, we are proposing that the Day factor is more of a between-subject factor than a within-subject factor. Subjects were rotated through all four positions in the simulation room on different days, while fNIR was only collected on two positions per session causing fNIR data to be recorded from different subjects on different days. There were a total of 12 subjects in which fNIR data were available for at least one of the sessions on one of the days. fNIR data were available from some subjects for all sessions in one day and for other subjects in some sessions across multiple days. Moreover, the researchers were blinded to the subjects and sessions in which fNIR data were available from each subject. Given these anomalies, we contend that the randomness of assignments to sessions across days and the blinding of the researchers to specific subject information cause the Day factor to be more similar to a between-subject factor than a within-subject factor. Our future work will conform to specific within-subject design parameters.

The hierarchical prior model used specifies that, at the lowest level, the observed data values (y) are distributed normally around the predicted mean (μ): y[i] $\sim$ dnorm(μ[i],τ), where τ is the precision of the normal (i.e., the reciprocal of the variance). Traditional analysis of variance (ANOVA) models decompose the overall variance across data into variance within and between the nominal predictors. BANOVA models, on the other hand, rest on the assumption that there is a baseline quantity of the predicted variable and that each level of the predictor has a deflection above or below that baseline, with the constraint that all deflections sum to zero. A great benefit of the BANOVA analysis is that it accommodates unbalanced designs with ease, something that traditional split-plot ANOVAs cannot do. The BANOVA analysis does not require each factor to be balanced but instead uses all available data for the analysis. Therefore, for our purposes, the BANOVA was a better choice than a traditional split-plot ANOVA due to missing and rejected data that left us with an unbalanced design.

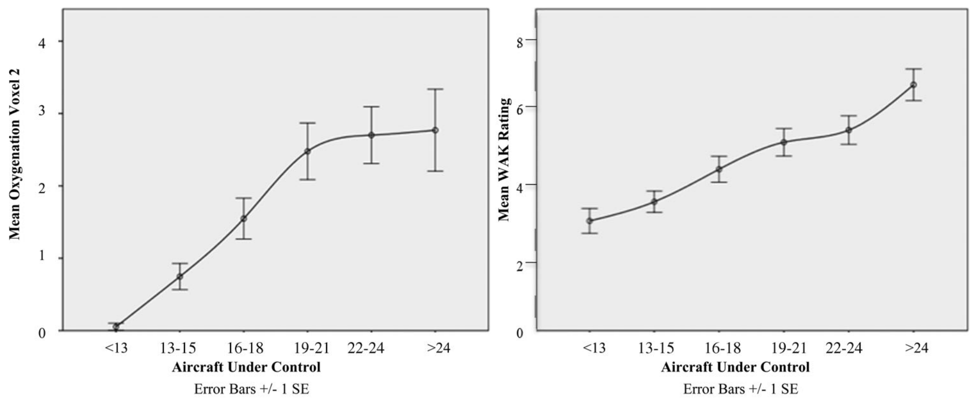

Fig. 5. Increase in workload related to increased number of aircrafts under control as assessed by fNIR channel 2 oxygenation (left) and WAK rating (right). All data are presented as the mean ± standard error.

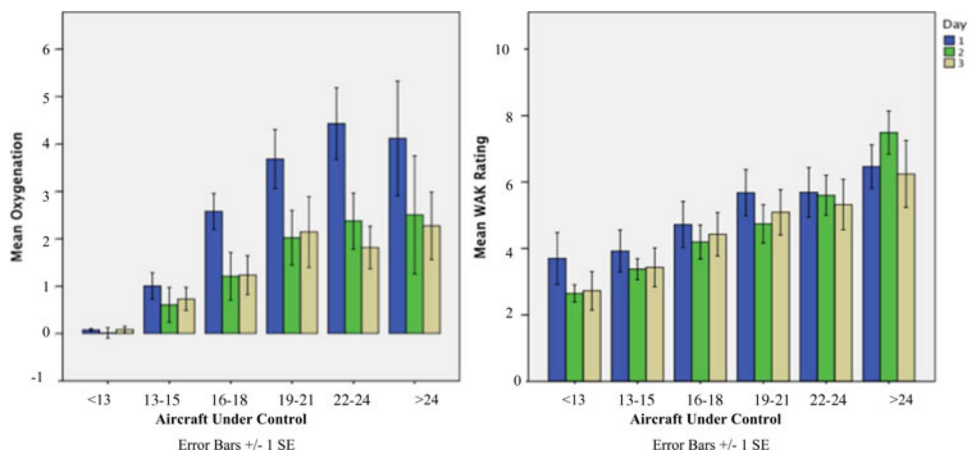

Fig. 6. Learning effect during CRA study assessed by fNIR in channel 2 oxygenation (left) and WAK rating (right). All data are presented as the mean ± standard error.

At the lowest level of the model, we have the observed data values, y, distributed normally around the predicted mean, μ, with precision of the normal τ (i.e., the reciprocal of the variance). Moving up the model ladder, we have the predicted mean (μ), which is the overall baseline (β_0) plus a deflection ($\beta_a x_a$) for factor A (Day), plus a deflection ($\beta_s x_s$) for each subject, plus a deflection ($\beta_b x_b$) for factor B (aircraft count), plus a deflection ($\beta_{axb} x_{axb}$) due to interaction of factors A and B.

In our model, we decided to use uninformative priors because we have no prior research results on constantly increasing workload conditions, and with uninformative priors, it is easy for the data to overcome the priors, and thereby, the data will have a strong influence on the posterior. For the baseline (β_0) prior, we used a broad normal distribution with a mean = 0 and a standard deviation of 100 000. For the prior on the precision τ, we used a uniform (rectangular) distribution with its minimum = 0 and maximum = 1000. There is also a hierarchical prior on each type of deflection so that the variance (i.e., precision, the reciprocal of the variance) of the deflections is estimated. For the priors on the precision (τ_β) for factors A, B, S, and A × B, we chose uninformative gamma distributions with shape (S) and rate (R) parameters with mode = 0.1 and standard deviation = 10. As can be seen in the model, the prior for τ_β is derived by first converting the precision to standard deviation: $\tau_\beta = 1/\sigma_\beta^2$; then, a gamma distribution is used as a prior on σ_β^2.

Therefore, in our JAGS modeling, the ith level of sampling is μ[i] = base + a[aLvl[i]] + s[sLvl[i]] + b[bLvl[i]] + axb[aLvl[i],bLvl[i]]. To sample the data in the BANOVA, we used Markov chain Monte Carlo sampling with 100 000 saved steps [44].

III. Results

fNIR and WAK data were split into six within-subject "aircraft" blocks from each session, as well as three between-subject "Day" blocks (see Tables III and IV). We analyzed the main within-subject effect of aircraft count under control, the between-subject effect of Day, as well as the interaction of Day and aircraft count on the ATCSs workload by objective fNIR recordings and subjective WAK ratings (see Figs. 5 and 6). The presence of a credible difference was determined in the contrasts if the 95% highest density interval (HDI) did not include zero. If, on the other hand, the HDI of a posterior distribution contained the value zero, then a difference of zero between the contrasted posteriors is credible.

TABLE V
OXYGENATION AND WAK AIRCRAFT COUNT COMPARISON RESULTS

	'13-15' v '<13'	'16-18' v '<13-15'	'19-21' v '16-18'	'22-24' v '19-21'	'22-24' v '19-21'
Oxy Mean	0.694*	0.827*	0.91*	0.294	0.28
Oxy 95% HDI	0.181-1.22*	0.31-1.36*	0.383-1.44*	-0.248-0.819	-0.366-0.928
WAK Mean	0.484*	0.812*	0.853*	0.49*	0.696*
WAK 95% HDI	0.114-0.853*	0.441-1.18*	0.484-1.24*	0.113-0.877*	0.287-1.11*

TABLE VI
BAYESIAN CONTRASTS FOR DAY SEPARATED DATA IN fNIR CHANNEL 2 OXYGENATION AND WAK RATINGS

	Day 2 v Day 1	Day 3 v Day 1	Day 3 v Day 2
Oxy Mean	-0.993*	-0.988*	0.006
Oxy 95% HDI	-(1.91-0.079)*	-(1.98-0.0316)*	-0.698-0.726
WAK Mean	-0.364*	-0.612*	-0.248
WAK 95% HDI	-(0.639-0.084)*	-(0.887-0.332)*	-0.519-0.024

A. Analysis of Aircraft Count Effects

For relative oxygenation levels, as measured by fNIR, contrasts of the main within-subject factor, aircraft counts, indicate credible differences between aircraft count blocks (see Table V). Credible differences were found for the contrasts between adjacent blocks "13–15 versus <13," "16–18 versus 13–5," and "19–21 versus 16–18" (see Table V). The positive mean and HDI ranges indicate that as the aircraft count increases oxygenation measures increase. While differences in oxygenation were found between higher aircraft count blocks, the 95% HDIs included zero and, thus, were not credibly different for the comparisons "22–24 versus 19–21" and ">24 versus 22–24" (see Table V). For WAK ratings, credible differences were located between all adjacent blocks with a similar trend of increasing WAK scores associated with higher aircraft counts: "13–15 versus <13," "16–18 versus 13–5," "19–21 versus 16–18," "22–24 versus 19–21," and ">24 versus 22–24" (see Table V).

B. Analysis of Day Effects

In the objective fNIR oxygenation measures, credible differences were established in contrasts of the between-subject effect of Day in contrasts "Day 2 versus Day 1" and "Day 3 versus Day 1" (see Table VI). The negative means and 95% HDIs indicate that the average oxygenation levels for Days 2 and 3 are lower than the oxygenation levels of Day 1. No credible difference was detected between oxygenation levels of Days 2 and 3 (see Table VI). To further localize the oxygenation differences between days, we analyzed the contrasts of the interactions of aircraft levels and days. We found no credible differences between days for aircraft levels of ">13" and "13–15." Differences were viewed between Days 2 and 3 compared with Day 1 at aircraft levels "16–18" and "19–21"; however, the differences were not credible. At aircraft levels of "19–21," a credible difference was located between Day 2 compared with Day 1, $\mu = -1.21$, 95% HDI $= -(2.34–0.103)$, and a difference between Day 3 compared with Day 1 was located that was not credible, $\mu = -1.13$, 95% HDI $= -2.31–0.037$. At aircraft count "22–24," we found a credible difference between Days 1 and Day 2, $\mu = -1.38$, 95% HDI $= -(2.57–0.239)$, as well as Days 1 and Day 3, $\mu = -1.13$, 95% HDI $= -(2.85–0.255)$. Finally, we found no credible difference in oxygenation between days at aircraft levels >24.

Similar to the objective fNIR measures, contrasts in subjective WAK ratings produced credible differences between Days 2 and 3 compared to Day 1 (see Table VI). The negative mean of the differences indicates that WAK ratings for Days 2 and 3 are credibly smaller than the WAK ratings for Day 1. We also found a difference that was not credible between Days 2 and 3 (see Table VI). To further pinpoint the locus of the difference, we analyzed the interactions between aircraft count and day. Credible differences between days were localized to mostly lower aircraft counts of "<13," "Day 2 versus Day 1," $\mu = -0.86$, 95% HDI $= -(1.46–0.271)$; "Day 3 versus Day 1," $\mu = -1.01$, 95% HDI $= -(1.58–0.437)$, and "13–15," "Day 2 versus Day 1" $\mu = –0637$, 95% HDI $= -(1.22–0.048)$; and "Day 3 versus Day 1," $\mu = -0.66$, 95% HDI $= -(1.23–0.072)$. Other credible differences were located at aircraft level "19–21" in the comparison "Day 2 versus Day 1," $\mu = –0.606$, 95% HDI $= -(1.2–0.021)$ and ">24" aircraft in the comparison "Day 3 versus Day 2," $\mu = -1.26$, 95% HDI $= -(2.06–0.473)$.

IV. DISCUSSION

A. Aircraft Count Differences in Workload Measures

This paper presents the finding that the cognitive workload of ATCSs may be able to be accurately monitored for continuously and incrementally changed task difficulty levels using fNIR, a portable optical brain imaging system. Previous fNIR studies have demonstrated that mental workload can be estimated for controlled conditions in the natural working environment of the operators [11], [45]. However, the cognitive workload had not been analyzed by fNIR for the continuously changing task difficulty levels as in the current study, where ATCSs participated in 50-min sessions to distinguish between cognitive workload levels caused by continuously increased traffic levels. For low aircraft counts, it appears that ATCSs can increase their cognitive function, specifically, working memory, similar to ATCSs performance on the n-back test [11]. However, as the aircraft count increased beyond 100% MAP value, according to the fNIR results, some ATCSs may not have been able to increase their cognitive function to meet the tasks demand. We also observed the same trend when performing a spline fit to the fNIR data (see Fig. 5), revealing that the second derivative changes for the group around the "19–21" aircraft mark, an indication that controllers may not be able to continue to increase their cognitive function to match the task demand. This finding may add objective physiological validity to the MAP rating system previously described by the FAA.

WAK results, on the other hand, indicate that the controllers subjectively experienced a credible increase in workload for every increasing aircraft block described. The spline fit to the WAK results, as depicted in Fig. 6, indicates a rather linear increase in perceived workload along with the linear increase in aircraft count under ATC. Another interesting finding is that fNIR results indicate that the cognitive workload of the ATCSs may increase at a much faster rate than the perceived workload that the ATCSs admit in their subjective WAK ratings. We have observed that some controllers may be reluctant to utilize the entire scale of the WAK and, thus, may have caused the low slope in WAK ratings. When assessing the WAK ratings, it is also important to realize that while the controllers may feel the overall impact of the physical and mental demand of the task, the results presented from channel 2 of the fNIR measurements have previously been shown to assess the working memory areas of the prefrontal cortex [11]. Additionally, due to the complexity of the task, many other fNIR channels in the prefrontal cortex, not reported in this paper, indicate linearly increasing oxygenation levels as aircraft count increases. These findings could indicate the importance of other areas of the prefrontal cortex in completion of the task.

B. Day Differences in Workload Measures/Learning Effect

More interestingly, the results indicate a preliminary finding that we may be able to evaluate the learning of trained ATCSs using a new CRA technology over a course of three days and nine sessions. Previous learning studies, using fNIR, have evaluated the learning of novices performing a completely new task rather than trained individuals learning a new system to complete a task in which they have expertise [11], [46]. Within our results, we have a preliminary indication that we may have successfully evaluated learning of the new CRA with the fNIR measures, as well as subjective WAK ratings. In both fNIR and WAK measures, there were credible decreases comparing Days 2 and 3 with Day 1; however, no credible difference was realized between Days 2 and 3. This finding may indicate that controllers were able to acquire a level of comfort with the CRA after only using the technology for one day with no additional gains on the following two days. While gains in learning level off on Days 2 and 3, it is important to consider that a longer training period may allow for the controllers to fully adapt to the technology and further reductions in workload may arise with more training, as learning is typically realized in a nonlinear fashion either depicted as a power or exponential function [47], [48].

While we found no credible difference in the objective fNIR measures between Days 1, 2, and 3 at low aircraft counts, preliminary results indicate that fNIR may be able to distinguish between cognitive workload levels across different days at higher task load levels, which allowed for the largest level of separation between oxygenation levels. Interestingly, within fNIR measures, the difference between Days 1 and 2 was found to be credible at more aircraft levels than the difference between Days 1 and 3. This finding could potentially be explained by the fact that there were more data for Day 2, $n = 10$, compared with Day 3, $n = 9/8$, allowing the differences to be more credible in Day 2 comparisons. Another possibility for this finding could be attributed to a possible burnout toward the end of the three-day period due to the rigorous testing schedule or the fact that the experimental design was not a true repeated measures design.

Additionally, the credibility of the between-subject day difference disappeared for aircraft level of “>24,” potentially an effect of a lower number of a data points available for this comparison as well. An alternative explanation for the lack of a difference at the aircraft level “>24” would be that some controllers may already be at their peak cognitive activation in this area of their brain for all three days. From Fig. 6, it is evident that we have the largest dispersion in the data for the “>24” aircraft level. For some of the controllers, the fNIR hemodynamic response is increasing from the “22–24” aircraft level to the “>24” aircraft level, while for other controllers, which may have reached their maximum, oxygenation values would not go any higher. Thus, the “>24” aircraft level could indicate a cognitive activation level or a ceiling effect for some controllers.

Similar to the objective fNIR results, the controllers appeared to perceive the task to be less difficult on Days 2 and 3 compared with Day 1. Within the subjective WAK results, the loci of the credible differences were mostly on lower aircraft counts, rather than the higher aircraft counts found in the objective fNIR results. This finding may be explained by a few hypotheses. First, due to the inherent variability in between-subject fNIR measures, the higher oxygenation levels found in higher workload levels may allow for greater resolution between different days, rather than the low oxygenation values. Examining the results from the perspective of the WAK causes one to hypothesize that the WAK ratings are more beneficial when the ATCSs have time to consider their answer as afforded during lower workload levels. Additionally, as previously mentioned, ATCSs are sometimes reluctant to utilize the entire WAK scale. Thus, some ATCSs may be less likely to indicate they were near to their maximum workload level at higher aircraft counts, even if they were, which may have affected the resolution between days at higher workload levels in the WAK ratings. The low aircraft counts may have given ATCSs an opportunity to weigh their workload against the WAK scale with less stress, resulting in less need to use the higher numbers of the WAK scale. Another valid hypothesis when considering this issue is that at lower workload levels compared with higher workload levels, the ATCSs may be more cognizant of their workload and how it is associated with the WAK ratings. Finally, the credibility in the contrasts may have been lost at higher aircraft counts because the increased workload found during the high aircraft counts caused fewer controllers to remember to input WAK ratings as the aircraft count increased, allowing for fewer data points to be available for sampling. Similar to the objective fNIR measures, a larger number of points for sampling and a true repeated measures design may have given the results more credibility during the statistical analysis.

C. Comparison of Workload Measuring Devices

The results presented in this paper indicate some possible advantages in objective measures of cognitive workload assessment obtained with fNIR brain imaging over subjective measures, such as the WAK rating system. Certain aspects, such as the possible inability of the operators to match cognitive activity to cognitive demand, are realized only through fNIR measures. Additionally, fNIR may allow for a more objective measure of the cognitive workload by removing the reluctance of some controllers to use the entire WAK scale and indicate their true workload. While there are many benefits realized in using an objective measure such as fNIR instead of the subjective workload measures, the subjective WAK measures were able to offer better contrast of workload between days for lower aircraft counts in this study. This advantage in the WAK ratings may be lost in a true within-subject measure of learning and additional fNIR data points.

D. Preliminary Nature of Finding

Our findings must be interpreted with caution, as the learning results are evaluated as an incomplete between-subjects model, whereas learning would optimally be analyzed as a repeated measures design. Additionally, the learning effect described could potentially be caused by the ATCSs still learning the unfamiliar ZKC airspace or gaining familiarity with their coworkers rather than the CRA even though a full day of training was given prior to the experiment. Future work, employing an increase in the sample size and a pure repeated measure within-subject evaluation of learning, might help to solidify the learning results presented in this paper. Behavioral data on the ATCSs' success in completing the task across days may also aid in confirming our learning hypothesis.

V. Conclusion

Continuously and objectively monitoring the cognitive workload of ATCSs and other operators, with a portable brain-imaging device, such as fNIR, may allow for an increase in safety of air travel and other high-risk activities by ensuring that the operator does not become overloaded. An accurate objective assessment of the cognitive workload may help prevent operator error and allow for appropriate intervention through predicting probable errors that can arise from work overload [2]–[5]. Additionally, an objective cognitive workload assessment system, such as fNIR, may prove to be a valuable tool in the validation of the array of FAA's NextGen systems, as well as monitoring learning during the implementation of such systems, similar to the CRA presented in this paper.

References

[1] M. R. Endsley and M. D. Rodgers, "Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation.," *Air Traffic Control Quart.*, vol. 6, pp. 21–44, 1998.

[2] P. Hancock and R. Parasuraman, "Human factors and safety in the design of intelligent vehicle–highway systems," *J. Safety Res.*, vol. 23, pp. 181–198, 2008.

[3] P. A. Hancock and W. B. Verwey, "Fatigue, workload and adaptive driver systems," *Accident Anal. Prevention*, vol. 29, pp. 495–506, Jul. 1997.

[4] R. Parasuraman, J. Christensen, and S. Grafton, "Neuroergonomics: The brain in action and at work," *NeuroImage*, vol. 59, pp. 1–3, Jan. 2, 2012.

[5] R. Parasuraman and G. F. Wilson, "Putting the brain to work: Neuroergonomics past, present, and future," *Human Factors*, vol. 50, pp. 468–474, Jun. 2008.

[6] S. H. Fairclough, L. Venables, and A. Tattersall, "The influence of task demand and learning on the psychophysiological response," *Int. J. Psychophysiol.*, vol. 56, pp. 171–184, May 2005.

[7] J. B. Brookings, G. F. Wilson, and C. R. Swain, "Psychophysiological responses to changes in workload during simulated air traffic control," *Biol. Psychol.*, vol. 42, pp. 361–77, Feb. 5 1996.

[8] A. Gevins, M. E. Smith, H. Leong, L. McEvoy, S. Whitfield, R. Du, and G. Rush, "Monitoring working memory load during computer-based tasks with EEG pattern recognition methods," *Human Factors*, vol. 40, pp. 79–91, Mar. 1998.

[9] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis," *Brain Res. Brain Res. Rev.*, vol. 29, pp. 169–195, Apr. 1999.

[10] B. Chance, Z. Zhuang, C. UnAh, C. Alter, and L. Lipton, "Cognition-activated low-frequency modulation of light absorption in human brain," *Proc. Natl. Acad. Sci. USA*, vol. 90, pp. 3770–3774, Apr. 15, 1993.

[11] H. Ayaz, P. A. Shewokis, S. Bunce, K. Izzetoglu, B. Willems, and B. Onaral, "Optical brain monitoring for operator training and mental workload assessment," *NeuroImage*, vol. 59, pp. 36–47, Jan. 2, 2012.

[12] K. Izzetoglu, S. Bunce, B. Onaral, K. Pourrezaei, and B. Chance, "Functional optical brain imaging using near-infrared during cognitive tasks," *Int. J. Human–Comput. Interact.*, vol. 17, pp. 211–227, 2004.

[13] S. Bunce, K. Izzetoglu, and H. Ayaz, "Implementation of fNIRS for monitoring levels of expertise and mental workload," in *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems*. Berlin, Germany: Springer, 2011, pp. 13–22.

[14] H. Ayaz, S. Bunce, P. A. Shewokis, K. Izzetoglu, B. Willems, and B. Onaral, "Using brain activity to predict task performance and operator efficiency," in *Advances in Brain Inspired Cognitive Systems, vol. 7366*, H. Zhang, D. Hussain, D. Liu, and Z. Wang, Eds. Berlin, Germany: Springer, 2012, pp. 147–155.

[15] H. Ayaz, B. Willems, S. Bunce, P. A. Shewokis, K. Izzetoglu, S. Hah, A. Deshmukh, and B. Onaral, "Estimation of cognitive workload during simulated air traffic control using optical brain imaging sensors," in *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems*. Berlin, Germany: Springer, 2011, pp. 549–548.

[16] M. Ferrari and V. Quaresima, "A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application," *NeuroImage*, vol. 63, pp. 921–35, Nov. 1, 2012.

[17] M. Izzetoglu, K. Izzetoglu, S. Bunce, H. Ayaz, A. Devaraj, B. Onaral, and K. Pourrezaei, "Functional near-infrared neuroimaging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 2, pp. 153–159, Jun. 2005.

[18] P. T. Fox, M. E. Raichle, M. A. Mintun, and C. Dence, "Nonoxidative glucose consumption during focal physiologic neural activity," *Science*, vol. 241, pp. 462–464, Jul. 22, 1988.

[19] F. F. Jobsis, "Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters," *Science*, vol. 198, pp. 1264–1267, Dec. 23, 1977.

[20] T. J. Huppert, R. D. Hoge, S. G. Diamond, M. A. Franceschini, and D. A. Boas, "A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans," *NeuroImage*, vol. 29, pp. 368–382, Jan. 15, 2006.

[21] M. M. Plichta, M. J. Herrmann, C. G. Baehne, A. C. Ehlis, M. M. Richter, P. Pauli, and A. J. Fallgatter, "Event-related functional near-infrared spectroscopy (fNIRS) based on craniocerebral correlations: Reproducibility of activation?," *Human Brain Mapping*, vol. 28, pp. 733–741, Aug. 2007.

[22] A. Sassaroli, F. Zheng, L. M. Hirshfield, A. Girouard, E. T. Solovey, R. J. K. Jacob, and S. Fantini, "Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy," *J. Innovative Opt. Health Sci.*, vol. 1, pp. 227–237, Oct. 2008.

[23] G. Strangman, D. A. Boas, and J. P. Sutton, "Non-invasive neuroimaging using near-infrared light," *Biol. Psychiatry*, vol. 52, pp. 679–693, 2002.

[24] A. Villringer, J. Planck, C. Hock, L. Schleinkofer, and U. Dirnagl, "Near infrared spectroscopy (NIRS): A new tool to study hemodynamic changes during activation of brain function in human adults," *Neurosci. Lett.*, vol. 154, pp. 101–104, May 14, 1993.

[25] G. Gratton and M. Fabiani, "Optical imaging of brain function," in *Neuroergonomics: The Brain at Work*, R. Parasuraman and M. Rizzo, Eds. Cambridge, MA, USA: Oxford Univ. Press, 2008, pp. 66–81.

[26] A. M. Kelly and H. Garavan, "Human functional neuroimaging of brain changes associated with practice," *Cerebral Cortex*, vol. 15, pp. 1089–1102, Aug. 2005.

[27] J. G. Milton, S. S. Small, and A. Solodkin, "On the road to automatic: Dynamic aspects in the development of expertise," *J. Clin. Neurophysiol.*, vol. 21, pp. 134–43, May/Jun. 2004.

[28] M. H. Beauchamp, A. Dagher, J. A. Aston, and J. Doyon, "Dynamic functional changes associated with cognitive skill learning of an adapted version of the tower of London task," *NeuroImage*, vol. 20, pp. 1649–1660, Nov. 2003.

[29] R. A. Poldrack, "Imaging brain plasticity: Conceptual and methodological issues—A theoretical review," *NeuroImage*, vol. 12, pp. 1–13, Jul. 2000.

[30] H. Garavan, D. Kelley, A. Rosen, S. M. Rao, and E. A. Stein, "Practice-related functional activation changes in a working memory task," *Microscopy Res. Tech.*, vol. 51, pp. 54–63, Oct. 1, 2000.

[31] B. Chance, E. Anday, S. Nioka, S. Zhou, L. Hong, K. Worden, C. Li, T. Murray, Y. Ovetsky, D. Pidikiti, and R. Thomas, "A novel method for fast imaging of brain function, non-invasively, with light," *Opt. Exp.*, vol. 2, pp. 411–23, May 11, 1998.

[32] H. Ayaz, P. A. Shewokis, A. Curtin, M. Izzetoglu, K. Izzetoglu, and B. Onaral, "Using mazesuite and functional near infrared spectroscopy to study learning in spatial navigation," *J. Vis. Exp.*, vol. 56, e3443, 2011.

[33] M. Cope, D. T. Delpy, E. O. Reynolds, S. Wray, J. Wyatt, and P. van der Zee, "Methods of quantitating cerebral near infrared spectroscopy data," *Adv. Exp. Med. Biol.*, vol. 222, pp. 183–189, 1988.

[34] M. Cope and D. T. Delpy, "System for long-term measurement of cerebral blood and tissue oxygenation on newborn infants by near infra-red transillumination," *Med. Biol. Eng. Comput.*, vol. 26, pp. 289–294, May 1988.

[35] S. Bunce, M. Izzetoglu, K. Izzetoglu, B. Onaral, and K. Pourrezaei, "Functional near infrared spectroscopy: An emerging neuroimaging modality," *IEEE Eng. Med. Biol. Mag.*, vol. 25, no. 4, pp. 54–62, Jul./Aug. 2006.

[36] M. Cope, "The application of near infrared spectroscopy to non-invasive monitoring of cerebral oxygenation in the newborn infant," Ph.D. dissertation, Dept. Med. Phys. Bioeng., Univ. College London, London, U.K., 1991.

[37] E. S. Stein, "Air traffic controller workload: An examination of a workload probe," U.S. Dept. of Transportation, Federal Aviation Administration Technical Center, Atlantic City, NJ, USA, 1985.

[38] J. Harrison, K. Izzetoglu, H. Ayaz, B. Willems, S. Hah, H. Woo, P. Shewokis, S. Bunce, and B. Onaral, "Human performance assessment study in aviation using functional near infrared spectroscopy," in *Foundations of Augmented Cognition, vol. 8027*, D. Schmorrow and C. Fidopiastis, Eds. Berlin, Germany: Springer, 2013, pp. 433–442.

[39] B. Willems, S. Hah, and K. Schulz, "En route data communications: Experimental human factors evaluation," Federal Aviation Administration, Atlantic City International Airport, William J. Hughes Technical Center, Atlantic city, NJ, USA, 2010.

[40] FAA, "FAA order 7210.3 facility operation and administration," Federal Aviation Administration, Washington, DC, USA, Tech. Rep. 7210.3, 2007.

[41] M. Plummer, "JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling," presented at the 3rd Int. Workshop Distrib. Statist. Comput., Vienna, Austria, 2003.

[42] M. Plummer, "Rjags: Bayesian graphical models using MCMC," in R. Package, 3–5 ed., 2011.

[43] R. D. C. Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2011.

[44] J. K. Kruschke, *Doing Bayesian Data Analysis: A Tutorial With R and BUGS*. Burlington, MA, USA: Academic, 2011.

[45] A. Girouard, E. T. Solovey, L. M. Hirshfield, K. Chauncey, A. Sassaroli, S. Fantini, and R. J. K. Jacob, "Distinguishing difficulty levels with non-invasive brain activity measurements," in *Proc. Human–Comput. Interact.*, Berlin, Germany, 2009, pp. 440–452.

[46] H. Ayaz, M. P. Cakir, K. Izzetoglu, A. Curtin, P. A. Shewokis, S. C. Bunce, and B. Onaral, "Monitoring expertise development during simulated UAV piloting tasks using optical brain imaging," in *Proc. Aerospace Conf.*, Big Sky, MT, USA, 2012, pp. 1–11.

[47] P. A. Shewokis, H. Ayaz, M. Izzetoglu, S. Bunce, R. J. Gentili, I. Sela, K. Izzetoglu, and B. Onaral, "Brain in the loop: Assessing learning using fNIR in cognitive and motor tasks," in *Foundations of Augmented Cognition. Directing the Future of Adaptive Systems (Lecture Notes in Computer Science and Artificial Intelligence), vol. 6780*. Berlin, Germany: Springer, 2011, pp. 240–249.

[48] C. Speelman and K. Kirsner, *Beyond the Learning Curve: The Construction of Mind*. Oxford, U.K.: Oxford Univ. Press, 2005.

Joshua Harrison received the B.S. degree in biology and chemistry from Dickinson College, Carlisle, PA, USA, in 2011. He is currently working toward the Graduate degree with the School of Biomedical Engineering, Drexel University, Philadelphia, PA, where his research focuses on the human performance in aerospace with a focus on brain imaging using functional near-infrared spectroscopy.

Kurtuluş İzzetoğlu received the Ph.D. degree from the School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA, USA.

He is currently a Research Assistant Professor with the School of Biomedical Engineering, Science and Health Systems, Drexel University. His research interests include functional brain imaging, medical sensor development, and biomedical signal processing.

Hasan Ayaz (S'06–M'11) received the B.Sc. (high Hons.) degree in electrical and electronics engineering from Boğaziçi University, Istanbul, Turkey and the Ph.D. degree in biomedical engineering from Drexel University, Philadelphia, PA, USA, in 2010.

He is currently an Assistant Research Professor with the School of Biomedical Engineering, Science and Health Systems, Drexel University. His research interests include neuroengineering applications of human–computer interaction and neuroergonomics.

Ben Willems received the B.S. degree in mechanical engineering from Eindhoven University of Technology, the Netherlands, in 1988 and the M.A. degree in ergonomics and biomechanics from New York University, NY, USA, in 1993.

He joined the Federal Aviation Administration, Washington, DC, USA, in 1998, where he is currently an Engineering Research Psychologist with the Human Factors Branch. His research interests include the design of a concept air traffic control workstation and an experiment to compare controller behavior between the conventional and redesigned system.

Sehchang Hah received the B.A. degree in psychology from Seoul National University, Seoul, Korea, and the M.A. and Ph.D. degrees in experimental psychology from The Ohio State University, Columbus, OH, USA.

He is currently an Engineering Research Psychologist with the Human Factors Branch, Federal Aviation Administration (FAA), Washington, DC, USA. Since joining the FAA, he has been working on both air traffic and technical operations human factors issues

Ulf Ahlstrom received the Ph.D. degree in psychology from Uppsala University, Uppsala, Sweden, in 1994.

He is currently an Engineering Research Psychologist with the William J. Hughes Technical Center Human Factors Branch, Federal Aviation Administration (FAA), Washington, DC, USA. He has been with the FAA since 1997. His current research interests include broad areas of air traffic control, user-interface design, operator workload, and weather information displays.

Hyun Woo received the B.S. and M.A. degrees in psychology, with a concentration on human factors and applied cognition, from George Mason University, Fairfax, VA, USA.

She was a Research Intern with the Federal Aviation Administration, Washington, DC, USA. She is currently working as an UX Specialist with MetroStar Systems, Reston, VA.

Patricia A. Shewokis received the Ph.D. degree in the psychology of motor behavior from the University of Georgia, Athens, GA, USA, with cognates in biomechanics and research design/statistics.

She is currently a Professor with the College of Nursing and Health Professions, Drexel University, Philadelphia, PA, USA, where she has an appointment in the Department of Nutrition Sciences. She holds a joint appointment with the School of Biomedical Engineering, Science and Health Systems.

Scott C. Bunce received the Doctorate degree in clinical and personality psychology from the University of Michigan, Ann Arbor, MI, USA, where he also completed the Postdoctoral Training with the Department of Psychiatry.

He is currently an Associate Professor of psychiatry with Penn State Hershey Medical Center and the Penn State College of Medicine, Hershey, PA, USA, and holds an adjunct appointment with the School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia, PA.

Banu Onaral (S'76–M'78–SM'89–F'93) received the B.S.E.E. and M.S.E.E. degrees from Boğaziçi University, Istanbul, Turkey, and the Ph.D. degree from the University of Pennsylvania, Philadelphia, PA, USA, in 1978.

She is currently the H. H. Sun Professor and the Director with the School of Biomedical Engineering, Science and Health Systems, Drexel University, Philadelphia. Her research interests include information engineering with special emphasis on complex systems and biomedical signal processing in ultrasound and optics.