Implementation of fNIRS for Monitoring Levels of Expertise and Mental Workload

Scott C. Bunce[1,2], Kurtulus Izzetoglu[2], Hasan Ayaz[2], Patricia Shewokis[2,3],
Meltem Izzetoglu[2], Kambiz Pourrezaei[2], and Banu Onaral[2]

[1] Penn State Hershey Medical Center and Penn State College of Medicine
[2] Drexel University School of Biomedical Engineering, Sciences, and Health Systems
[3] College of Nursing and Health Professions, Drexel University

Abstract. An accurate measure of mental workload would help improve operational safety and efficacy in many environments that involve multitasking or sustained vigilance. The current study utilized functional near-infrared spectroscopy (fNIRs) to examine the relationship of the hemodynamic response in dorsolateral prefrontal cortex (DLPFC) as it related to mental workload, level of expertise, and task performance. DLPFC responses were monitored with fNIRs while 8 participants (4 with high practice, 4 novices) completed a quasi-realistic computerized Warship Commander Task with various levels of difficulty. The results show that greater expertise was associated with relatively lower oxygenation (less neural activity) at low to moderate levels of taskload, but higher oxygenation and better performance at high levels of taskload. For novices, oxygenation was higher at moderate levels of taskload, but dropped precipitously at higher levels of taskload, along with performance, consistent with disengaging from the task. Results are interpreted within a “scaffolding-storage" framework.

Keywords: Optical Brain Imaging; functional near infrared spectroscopy; mental workload; expertise; practice, fNIR.

1 Introduction

Mental workload plays a critical role in many complex command and control systems. It is particularly important to understand operator workload in situations where performance failures could result in catastrophic losses (e.g., warship command, Air Traffic Control (ATC)). Accurate assessment of mental workload could help prevent operator error and allow timely intervention by predicting performance decline that can arise from either work overload or from understimulation [1-3]. For some time, investigators have been working on adaptive aiding schemes to facilitate optimal performance in critical mission systems by dynamically matching the momentary mental capabilities of the operator to the task demands [4-5]. To be viable, such a system must improve performance above the levels possible with unaided and fully automated systems [3, 6]. Further, adaptive aiding systems should provide aid only when required [7], as providing an unnecessary intervention can lead to performance errors as readily as not providing aid when it is required [1-3]. An accurate and

D.D. Schmorrow and C.M. Fidopiastis (Eds.): FAC 2011, HCII 2011, LNAI 6780, pp. 13–22, 2011.

reliable measure of the operator's mental workload is a critical component of any such adaptive aiding system [8].

Within the basic methods of measuring mental workload, physiological measures are the most promising for field applications. Neurophysiological and psychophysiological measures are commonly used to index the level of mental demand associated with a given task because they can provide continuous and unobtrusive monitoring that does not contribute to the operator's workload [9-16].

There is considerable evidence that neurophysiological and psychophysiological variables respond to cognitive demand in a predictable manner [17]. Direct measures of central nervous system function such as electroencephalography (EEG) and event-related potentials (ERPs) have been particularly strong candidates for accurate, objective measures of operator workload. Increasing task difficulty, for instance, is known to be associated with EEG changes such as increased power in the beta bandwidth, increased theta activity at frontal sites and the suppression of alpha activity [12, 18-19]. Although EEG has many excellent qualities for monitoring mental workload, including excellent temporal resolution, it is limited in its capacity for spatial resolution.

In this study, we utilized an optical measure of neural activity, functional near infrared spectroscopy (fNIRs), a good candidate for measuring mental workload under field conditions. fNIRS is safe, highly portable, user-friendly and relatively inexpensive, with rapid application times and near-zero run-time costs [20]. The most commonly used form of fNIRs uses light, introduced at the scalp, to measure changes in blood oxygenation as oxy-hemoglobin converts to deoxyhemoglobin during neural activity, ie., the hemodynamic response. Although its temporal resolution is limited to that of the hemodynamic response, fNIRs provides good spatial localization relative to EEG, on the order of 1cm^2, and has the capacity to be integrated with EEG/event-related potentials (ERPs).

The capacity for spatial resolution has important implications for the use of fNIRs as a measure of workload. Current models of automaticity related to the development of expertise in certain tasks suggest that there are shifts in the functional neuroanatomy of task performance that support ongoing cognitive efforts. Operator skill and mental work load are generally inversely related [21-22]. This inverse relationship between expertise and the cognitive demand of a given task impacts the accuracy and interpretation of psychophysiological variables as measures of mental workload [23-24]. But as automaticity develops in various tasks, shifts in the functional neuroanatomy of task performance free up attentional resources, largely associated with prefrontal cortices, for other efforts. The development of expertise, or automaticity, can be characterized as freeing up of the limitations on those attentional resources [25-26].

The literature regarding the effect of practice and expertise on the functional anatomy of task performance is extensive and complex. Practice and the development of expertise have been studied across a range of motor, visuomotor, perceptual and cognitive tasks, and from disparate research perspectives. In summary, four main patterns of practice-related activation change can be distinguished [27]. 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 occurs in distinct brain areas as well as the initial areas.

There are significant differences in the brain areas involved across these various tasks. However, when practice is associated with the attainment of automatic or asymptotic performance, as is often the case with the development of expertise, the research consistently finds decreased demand on control or attentional processes and a shift to increased demand on storage and processing in task-specific areas, often referred to as the development of "automaticity." Petersen et al. [28] have referred to this phenomenon as a 'scaffolding-storage' framework.

According to Petersen et al. [28], a set of brain areas (the scaffolding) is used to support or cope with novel demands during unskilled, effortful performance. Practice allows processes or associations that are more efficiently stored and accessed elsewhere to be offloaded to those areas, and the scaffolding network is pruned away. The decreased reliance on the 'scaffolding' attentional and control areas is demonstrated by decreased activation in those areas, while an increase in activation is observed in those areas underlying the task-specific processes. Activations seen earlier in practice therefore involve generic attentional and control areas. The prefrontal cortex (PFC), anterior cingulate cortex (ACC) and posterior parietal cortex (PPC) are the main areas considered to perform the 'scaffolding' role, consistent with theories of PFC function and the involvement of these areas in the distributed working memory system. On the other hand, increases associated with highly practiced performance are primarily seen in task-specific areas such as representational cortex — primary and secondary sensory or motor cortex, or in areas related to the storage of those representations, such as parietal or temporal cortex.

The majority of studies examining task practice have found decreases in the extent or intensity of activations, particularly in the attentional and control areas [27]. This finding is true whether the task is primarily motor, as in a golf swing [29] or primarily cognitive in nature, as in the Tower of London problem [30]. Decreases in activation represent a contraction of the neural representation of the stimulus [31] or a more precise functional circuit [32]. This finding provides an important overlap with the literature on expertise. There is considerable evidence that overall, experts show lower brain activity relative to novices, particularly in prefrontal areas [29]. Both 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 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.

A neuronal measure of expertise must be defined in relationship to behavioral performance. However, at a given level of performance, a neuronal measure of expertise that monitors the attentional and control resources the individual must utilize to maintain that level of performance could be expected to differentiate between relatively lesser and greater expertise. That is, even at 98-100% performance levels, where performance measures cannot differentiate between trainee capacities, some individuals will be performing at close to their peak performance, whereas others will be well below their performance capacity. An assessment of the cortical activity necessary to perform at a given level would indicate the cognitive resources available for more situational demands, consistent with greater expertise.

Below we provide preliminary evidence that fNIR can be used to 1) differentiate among individuals with more or less practice at a given task, and 2) differentiate among individuals who perform better or worse on a task that requires little practice, i.e., an example of neural efficiency. Previously, we have reported that the hemodynamic response over dorsolateral and ventrolateral prefrontal cortex, assessed using fNIRs, was responsive to workload in a realistic command and control task [33]. In this report, we examined the role of expertise (relative levels of practice) on the hemodynamic response over dorsolateral and ventrolateral prefrontal cortex using the same data set.

2 Method

Methods were identical to those reported in Izzetoglu et al. [33]. In brief, participants were 8 healthy adults (3 females), ages 18 – 50 years. Participants had varied experience with the Warship Commander Task (WCT), ranging from 3-4 hours to 300 hours. For the purposes of this study, a median split was used to divide participants into those with high levels of practice versus those with relatively few hours of practice. All participants signed informed consent statements approved by the Human Subjects Institutional Review Board at Drexel University.

Experimental Task. The WCT (*Pacific Science & Engineering Group, Inc.*) is a quasi-realistic, ship-based navy command and control environment task that requires spatial and verbal working memory and decision-making processes [34]. The version of WCT employed in the current experiment was comprised of two component tasks: Air Warfare Management and the Ship Status (SS) task. Air Warfare Management required the user to monitor “waves” of incoming airplanes, to identify them as friendly or hostile, and to warn and then destroy hostile airplanes using specified rules of engagement. Each wave lasted 75 seconds. The level of cognitive effort in the Air Warfare Management component was manipulated by 1) increasing the number of planes per wave (6, 12, 18 or 24 planes), and 2) changing the proportion of planes whose identity was unknown (hostile or friendly). Airplanes with unknown status require more decision and processing time, and therefore make the task more difficult. The WCT had two levels of difficulty relative to the proportion of unknown planes, *low* in which 33% of the planes were unknown, and *high*, in which 67% of the planes were unknown.

The presence or absence of the Ship Status (SS) Task, a secondary verbal task, was also used to manipulate cognitive workload. The SS task was comprised of auditory messages containing information about the ship and its operation (encoding), and periodic queries, or “recall” about earlier auditory messages. In this auditory task, the participants were required to listen to sporadic auditory messages and memorize various ship status data while answering periodic queries that appeared on the computer screen.

The Air Warfare Management task involves spatial and verbal working memory and decision-making processes, whereas the Ship Status Task is primarily a verbal memory task. When both tasks are operational, the WCT becomes a divided attention task. Although there is no universally accepted definition of cognitive workload, please see St. John et al. [34] for the rationale behind this definition of workload.

Each participant completed four sets of WCT. Each set was comprised of 12 waves, 3 repetitions of each wave size in the order of 6, 18, 12, and 24 planes. The factors of Wavesize (6, 12, 18, or 24 planes), Complexity (*high* versus *low* percentage of unknown airplanes), and *full* versus *divided* attention (secondary SS Task On or Off) were crossed to create a 4 x 2 x 2 repeated-measures design. Performance was assessed using the Percent Game Score, or PctGS, calculated as the percentage of game points that a participant accumulated during any given wave relative to the total game points that were possible for that wave [34].

Analyses. The average rate of change in oxygenation was calculated separately for each 75s wave. Oxygenation values were averaged across the 8 optodes monitoring left and right hemispheres respectively, yielding two values for each wave. Initial analyses showed that hemispheric differences were not significant, so subsequent analyses were collapsed across hemispheres. In these analyses, we were primarily interested in the effects of practice (or level of expertise) on the fNIRs signal as it related to both workload and performance. Analyses were conducted using univariate analyses of variance (MANOVA) in SPSS v.19.

3 Results

Warship Commander Task Performance. Performance data were subjected to a 4 x 2 x 2 factor mixed model ANOVA (with Wavesize and SS Task ON-OFF as within subject factors, and Practice (High, Low) as the between subjects factor). A main effect for Practice (F(1,6) = 10.25, p=.02, *h* =.63), revealed that more expertise was, as expected, associated with greater game score. The interaction between Practice and Wavesize was also significant (F(3,18) = 3.15, p=.05, *h* =.34). As displayed in Figure 1, total game score decreased with increasing task load (Wavesize). However, post-hoc analyses revealed that participants with higher levels of practice maintained relatively greater levels of performance as task demand increased, whereas total game scores dropped off for participants with less practice. A main effect for the secondary SS Task was also found (F(1,6) = 11.45, p=.015, *h* =.89), with the decreased performance when the SS Task was ON (M=87.34, (sd=9.6)) vs. OFF (91.35 (11.9)). There were no other significant effects involving the SS Task.

fNIRs data. fNIRs data were also subjected to a 4 x 2 x 2 mixed model ANOVA (Wavesize, SSTask ON/OFF as within-subjects factors, with Practice (High, Low) as a between-subjects factor). A main effect for Wavesize (F(3,18) = 7.27, p = .002, *h* = .55) was qualified by an interaction between Practice and Wavesize (F(3,18) = 4.02, p =.02, *h* =.40). As depicted in Fig. 2a, low to moderate levels of workload, increasing workload (greater wavesize) resulted in increasing oxygenation. Post hoc analyses revealed that at a moderate level of difficulty (12 planes), participants with greater expertise had lower levels of oxygenation. At 18 planes, the difference in oxygenation between high and low levels of was not significant. However, at the highest level of task difficulty (24 planes), oxygenation was greater among participants with more practice.

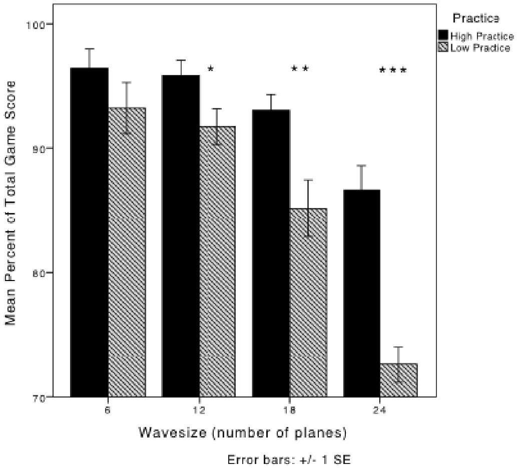

Fig. 1. Percentage of Warship Commander total game score as a function of wavesize (number of planes) and level of expertise. *p < .05; **p < .01; ***p < .001.

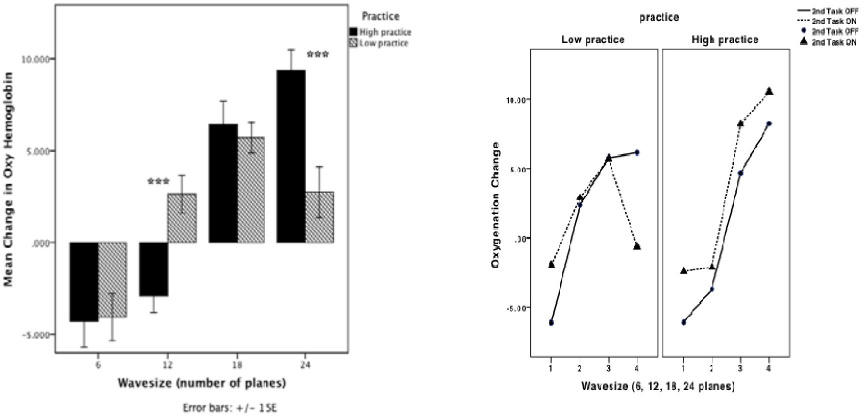

Fig. 2a. fNIRs measures of oxygenation change as a function of wavesize (number of planes) and level of expertise; ***p < .001. **2b.** fNIRs measures of oxygenation change plotted as a function of practice, wavesize, and secondary SStask.

Although there was a main effect for the SSTask (F(1,6) = 6.02, p =.05, *h* = .50), with greater oxygenation when the secondary task was on, the interaction with Practice did not obtain significance (F(3,18) = 2.67, ns. Although this interaction did not obtain significance, a graph of the means helps clarify the nature of the pattern of fNIRs oxygenation measures (Fig 2b). As can be seen in Fig. 3, greater task difficulty was generally associated with greater oxygenation. However, in the highest level of task difficulty, oxygenation in operators with less practice dropped precipitously, as did their game scores.

4 Discussion

The purpose of this study was to examine the impact of practice, or relative levels of expertise, on neurophysiological measures of the hemodynamic response in prefrontal cortex during a complex cognitive task. The complex, but interpretable, pattern of prefrontal cortical response in the WCT is consistent with a 'scaffolding-storage' framework of functional reorganization during the development of expertise [28].

At the lowest level of difficulty (6 planes), neither brain activity nor performance differed much between groups. Moderate levels of objective task difficulty elicited relatively less oxygenation among more experienced operators even as they maintained a high level of task performance relative to novice operators. At the highest levels of objective task difficulty, the oxygenation levels of experienced operators passed that of the novices. Both oxygenation and performance dropped off precipitously for novices at these difficult levels, suggesting a relative disengagement from the task, which was reflected in their performance scores. The more experienced operators maintained high performance scores, suggesting that relatively high levels of neural activity were supporting the more effortful processing necessary to manage the workload. This pattern is consistent with Petersen et al. 's theory that one function of practice and the development of automaticity is to free up attentional processes, i.e., a "cognitive reserve" to be used when task demands increase further, or when unexpected events arise.

Given the spatiotemporal pattern of results, it is likely the the fNIRs was measuring the hemodynamic correlates of sustained mental effort, or vigilance [35-36]. As effort dropped off, so did oxygenation. This suggests an important observation relative to the interpretation of the fNIRs signal. Lower oxygenation values may accurately reflect lower mental effort; however, the current results did not differentiate between an engaged but lower level of workload and disengagement from the task. As such, accurate interpretation of the fNIR signal within an adaptive system will likely need to provide some measure of either task difficulty or operator engagement.

One result that is somewhat difficult to interpret is the lack of differentiation across prefrontal cortex. Based on previous fMRI as well as fNIR studies, we expected more localized findings, particularly as activation during vigilance tends to be stronger in the right hemisphere [37-38]. It may be that the nature and complexity of this particular task elicited activation across both hemispheres; this is an area for future research. One limitation of the current study is the small sample size. With a sample size of four participants per cell, some effects were clearly underpowered. On the other hand, the goal of using such measures in adaptive automation systems requires the measure be sensitive for a single individual. It is likely that some factors that increased workload were operating in areas of the brain that were not being monitored in the current study. More information could be gained by monitoring parietal areas as well as frontal cortex, or combining fNIRs with EEG. Alternatively, signal processing algorithms for fNIRs may not be optimized at this time, or the hemodynamic response itself may not be sufficiently sensitive to pick up subtle changes in workload. This is an area for future research.

The current findings are consistent with numerous studies that have examined the neural correlates of the development of expertise in several cognitive domains [27-29]. These studies provide evidence that cortical activation in DLPFC increases with increasing attention/vigilance, as long as performance is held constant. These DLPFC activations reflect the effort of the participant, not the "objective" level of difficulty perse.

This study provides important albeit preliminary information about fNIR measures of DLPFC hemodynamic response and its relationship to mental workload, expertise, and performance, in a complex multitasking environment. Level of expertise does appear to influence the hemodynamic response in dorsolateral/ventrolateral prefrontal cortices, at least for some complex tasks. Since fNIRs technology allows the development of mobile, non-intrusive and miniaturized devices, it has the potential to be deployed in future learning & training environments to personalize the training regimen and/or assess the effort of human operators in critical multitasking environments.

References

1. Hancock, P.A., Parasuraman, R.: Human-Factors and Safety In The Design Of Intelligent Vehicle-Highway Systems (IVHS). J. Saf. Res. 23, 181–198 (1992)
2. Desmond, P.A., Hoyes, T.W.: Workload variation, intrinsic risk and utility in a simulated air traffic control task: evidence for compensatory effects. Safety Science 22, 87–101 (1996)
3. Hancock, P., Verwey, W.: Fatigue, workload and adaptive driver systems. Accident Analysis & Prevention 29, 495–506 (1997)
4. Young, M.S., Stanton, N.A.: Attention and automation: new perspectives on mental underload and performance. Theoretical Issues in Ergonomics Science 3, 178–194 (2002)
5. Young, M.S., Stanton, N.A.: Malleable attentional resources theory: A new explanation for the effects of mental underload on performance. Human Factors: The Journal of the Human Factors and Ergonomics Society 44, 365 (2002)
6. Parasuraman, R., Riley, V.: Humans and automation: Use, misuse, disuse, abuse. Human Factors 39 (1997)
7. Scerbo, M.W.: Theoretical perspectives on adaptive automation. In: Automation and Human Performance: Theory and Applications, pp. 37–63 (1996)
8. Parasuraman, R.: Neuroergonomics: Research and practice. Theoretical Issues in Ergonomics Science 4, 5–20 (2003)
9. Backs, R.W.: Engineering psychophysiology as a discipline: Historical and theoretical aspects. Engineering Psychophysiology: Issues and Applications (2000)
10. Hancock, P.A., Desmond, P.A.: Stress, workload, and fatigue. CRC, Boca Raton (2001)
11. Byrne, E.A., Parasuraman, R.: Psychophysiology and adaptive automation. Biological Psychology 42, 249–268 (1996)
12. 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 (1998)

13. Gevins, A., Smith, M.E., McEvoy, L., Yu, D.: High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. Cerebral Cortex 7, 374 (1997)
14. Scerbo, M.W., Freeman, F.G., Mikulka, P.J.: A brain-based system for adaptive automation. Theoretical Issues in Ergonomics Science 4, 200–219 (2003)
15. Wilson, G.F., Russell, C.A.: Operator functional state classification using multiple psychophysiological features in an air traffic control task. Human Factors: The Journal of the Human Factors and Ergonomics Society 45, 381 (2003)
16. Wilson, G.F., Russell, C.A.: Real-time assessment of mental workload using psychophysiological measures and artificial neural networks. Human Factors: The Journal of the Human Factors and Ergonomics Society 45, 635 (2003)
17. Fairclough, S.H., Venables, L., Tattersall, A.: The influence of task demand and learning on the psychophysiological response. International Journal of Psychophysiology 56, 171–184 (2005)
18. Gevins, A., Smith, M.E.: Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomics Science 4, 113–131 (2003)
19. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews 29, 169–195 (1999)
20. Bunce, S.C., Izzetoglu, M., Izzetoglu, K., Onaral, B., Pourrezaei, K.: Functional near-infrared spectroscopy. Engineering in Medicine and Biology Magazine 25, 54–62 (2006)
21. Gopher, D., Kimchi, R.: Engineering psychology. Annual Review of Psychology 40, 431–455 (1989)
22. Liu, Y., Wickens, C.D.: Mental workload and cognitive task automaticity: an evaluation of subjective and time estimation metrics. Ergonomics 37, 1843–1854 (1994)
23. O'Donnell, R.D., Workload, F.T.E.: assessment methodology. In: Boff, K.R., T., L.K.A.J.P. (eds.) Handbook of Perception and Human Performance, vol. 2, pp. 42–49. Wiley Interscience, New York (1986)
24. Wierwille, W., Eggemeier, F.T.: Recommendations for mental workload measurement in a test and evaluation environment. Human Factors: The Journal of the Human Factors and Ergonomics Society 35, 263–281 (1993)
25. Schneider, W., Shiffrin, R.M.: Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review 84, 1 (1977)
26. Shiffrin, R.M., Schneider, W.: Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. Psychological Review 84, 127 (1977)
27. Kelly, A., Garavan, H.: Human functional neuroimaging of brain changes associated with practice. Cerebral Cortex 15, 1089 (2005)
28. Petersen, S.E., Van Mier, H., Fiez, J.A., Raichle, M.E.: The effects of practice on the functional anatomy of task performance. Proceedings of the National Academy of Sciences of the United States of America 95, 853 (1998)
29. Milton, J.G., Small, S.S., Solodkin, A.: On the road to automatic: dynamic aspects in the development of expertise. Journal of Clinical Neurophysiology 21, 134 (2004)
30. Beauchamp, M., Dagher, A., Aston, J., Doyon, J.: Dynamic functional changes associated with cognitive skill learning of an adapted version of the Tower of London task. Neuroimage 20, 1649–1660 (2003)
31. Poldrack, R.A.: Imaging brain plasticity: conceptual and methodological issues–a theoretical review. Neuroimage 12, 1–13 (2000)
32. Garavan, H., Kelley, D., Rosen, A., Rao, S.M., Stein, E.A.: Practice-related functional activation changes in a working memory task. Microscopy Research and Technique 51, 54–63 (2000)

33. 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)
34. St John, M., Kobus, D., Morrison, J.: A multi-tasking environment for manipulating and measuring neural correlates of cognitive workload, p. 7. IEEE, Los Alamitos (2002)
35. Parasuraman, R., Caggiano, D.: Neural and genetic assays of human mental workload. Quantifying human information processing (2005)
36. Warm, J.S., Parasuraman, R., Matthews, G.: Vigilance requires hard mental work and is stressful. Human Factors: The Journal of the Human Factors and Ergonomics Society 50, 433 (2008)
37. Parasuraman, R., Warm, J.S., See, J.W.: Brain systems of vigilance. In: Parasuraman, R. (ed.) The Attentive Brain, pp. 221–256. MIT Press, Cambridge (1998)
38. Ayaz, H., Willems, B., Bunce, B., Shewokis, P.A., Izzetoglu, K., Hah, S., Deshmukh, A., Onaral, B.: Cognitive Workload Assessment of Air Traffic Controllers Using Optical Brain Imaging Sensors. In: Marek, T., Karwowski, W., Rice, V. (eds.) Advances in Understanding Human Performance: Neuroergonomics, Human Factors Design, and Special Populations, pp. 21–31. CRC Press Taylor & Francis Group (2010)