SIMULATOR BASED HUMAN PERFORMANCE ASSESSMENT IN A SHIP ENGINE ROOM USING FUNCTIONAL NEAR-INFRARED SPECTROSCOPY

STEPHEN WILLIAM ROBERT SYMES

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the Degree of Doctor of Philosophy

This research programme was carried out in collaboration with Liverpool Logistics, Offshore and Marine Research Institute, the Faculty of Marine Engineering and Technology and the School of Natural Sciences and Psychology

April 2022

Acknowledgments

Throughout my thesis I have received a great deal of assistance and support.

I would like to take this opportunity to acknowledge Liverpool John Moores University for funding the project and their additional support through the SARS-CoV-2 pandemic.

My supervisors Professor Zaili Yang, Professor Steve Fairclough, Dr Eddie Blanco-Davis and Professor Jin Wang have been a huge help as their expertise was invaluable whilst undertaking a project of this magnitude, covering many different subjects.

Senior lecturers; Dr Dan Stancioiu and Dr Frederic Bezombes gave me magnificent guidance when writing mechatronic code.

I must also thank simulator staff; Captain Jonathan Warren and Kamil Kaminski for their guidance on the use of the TRANSAS ship simulator and advice on the correct 'at sea' procedure.

Appreciation is also given to the many participants that gave their valuable time to allow me to conduct my experiments.

I acknowledge and thank my fellow colleagues at Liverpool Logistics, Offshore and Marine Research Institute and the School of Natural Sciences and Psychology for their continued support, suggestions and the tools that I needed in order to successfully complete my thesis.

I would particularly like to single out my supervisor Professor Steve Fairclough, I want to thank you for your excellent tuition and the long hours that you put in to help me understand the Psychology and Neuro-Science aspect of this project.

Abstract

80% of accidents that occur in the maritime sector are due to human error. These errors could be the result of seafarer training coupled with a high mental workload due to the addition of various working conditions.

The aim of this study is to evaluate the effect of various stressors on human performance on engine room operations. To achieve this aim, a simulator study was conducted to investigate the influence of training and working conditions on human performance for the purposes of fault detection and correction in a maritime engine room.

20 participants were recruited for each investigation of performance shaping factors (PSF); for the first test, half received practical training with the engine room software interface, while the other half were provided with paper-based instructions. The remaining tests were conducted with all participants equally practically trained. The participants interacted with a TRANSAS technological simulator series 5000. This is a 1:1 simulation of a ship engine room. The participants took part in a 30-minute scenario where they had to detect and correct a fault with the ballasting system. During this interaction, half of the participants experienced simulated, adverse performance shaping factors, which were distraction, fatigue and an increased workload. The other half were given a standard task.

Functional near-infrared spectroscopy (fNIRS) was utilised to measure neurophysiological activation from the dorsolateral prefrontal cortex (DLPFC).

The results indicated increased activation of lateral regions of the DLPFC during fault correction, this trend was enhanced due to PSF's and training, i.e. participants who received paper-based instructions showed greater activation when conducting the standard task and had an exponential increase in activation when dealing with the addition of an adverse PSF. The results are discussed with respect to the neural efficiency of the operator during high mental workload. From the results of this study a scientific human error model was developed and can be used by the maritime industry to better evaluate and understand human error causation and the effect of PSF on seafarers.

The impact of this study could reduce the frequency of occurrence of human error, reduce the financial impact that human error has on the maritime sector and reduce injuries and fatalities.

TABLE OF CONTENTS

List of figures

List of tables

Glossary

Abbreviation	Definition
AIS	*Automatic Identification System*
ANOVA	*Analysis of Variance*
AUC	*Area Under the Curve*
BCI	*Brain-Computer Interfaces*
BEng	*Bachelor of Engineering*
BWT	*Ballast Water Treatment*
CBSI	*Correction Based Signal Improvement*
CREAM	*Cognitive Reliability and Error Analysis Method*
DIW	*Distracted & Increased Workload*
DLPFC	*Dorsal Lateral Pre-Frontal Cortex*
EEG	*Electroencephalogram*
EMSA	*European Maritime Safety Agency*
ER	*Engine Room*
ERS	*Engine Room Simulator*
FD	*Fault Detection*
FIW	*Fatigued & Increased Workload*
FMEA	*Failure Modes and Effects Analysis*
FNIRS	*Functional Near Infra Red Spectroscopy*
FO	*Fault Occurrence*
FS	*Fault Solution*
FSA	*Formal Safety Assessment*
HBO	*Oxygenated Haemoglobin*
HEP	*Human Error Probability*
HFACS	*Human Factors Analysis and Classification System*
HRA	*Human Reliability Analysis*
Hz	*Hertz*
IEEE	*Institute of Electrical and Electronics Engineers*
IMO	*International Maritime Organisation*
LCS	*Liquid Cargo Screen*
LDA	*Linear Discriminant Analysis*
LNG	*Liquified Natural Gas*
MACD	*Moving Average Convergence Divergence*
MEng	*Master of Engineering*
NASA-TLX	*National Aeronautics and Space Administration Task Load Index*
NIR	*Near Infra Red*
NIRS	*Near Infra Red Spectroscopy*
NRN	*National Rail Networks*
NTSB	*National Transportation Safety Board*
OFS	*Operator Functional State*
PSF	*Performance Shaping Factor*
ROI	*Region Of Interest*
RSS	*Sum of Residuals Squared*
RTMS	*Repetitive Transcranial Magnetic Stimulation*
SPSS	*Statistical Package for the Social Sciences*
STCW	*Standards of Training, Certification and Watchkeeping*
TOPSIS	*Technique for Order Preferences by Similarity to the Ideal Solution*

WL	*Work Load*

Chapter 1. Introduction

1.1 Introductory remarks

This chapter presented herein strives to serve as an outline, aimed at offering the reader a concise context into the research tasks conducted for the successful completion of this PhD study. The following section, background research (1.2), will include a summarised review of the current risks and error statistics involved in maritime operations, followed by a succinct overview of advanced neuroimaging methods and analysis equipment available to tackle the problem. Additionally, a brief description of the current human reliability assessment methods and the practical application of neuroimaging techniques in other studies is presented. Described subsequently is the actual problem that this thesis aims to tackle; and lastly, an elucidation as to the physical composition of this thesis.

1.2 Research Background

Figure 1.1 is an illustration of the process flow for this section of the introduction chapter. The square boxes depict the sub sections involved in the background research conducted to better understand the scope of this study. Each square box has a brief description of the information contained within the subsections for the reader's ease.

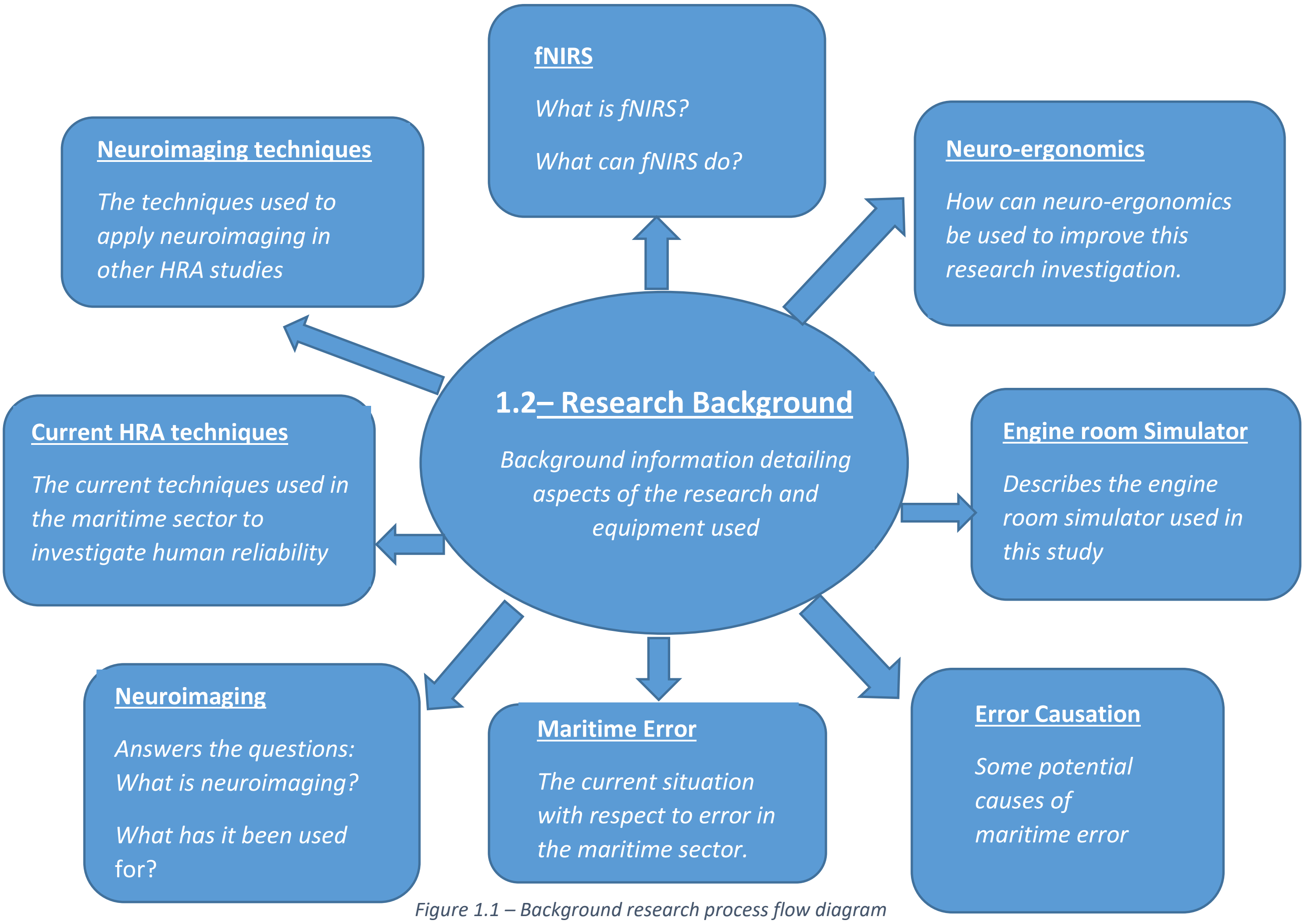


Figure 1.1 – Background research process flow diagram

80% of all maritime incidents reported are directly or indirectly a result of human error [1] [2] [3]. Compared to other sectors, maritime incidents have some of the highest financial significance [1] [4]. It has been reported that human error alone costs the marine sector millions of pounds annually [5] [6]. These statistics show that there is a research gap for current maritime human factor studies. Therefore, there is a research gap that needs to be filled between human error and the maritime industry by reviewing the current literature and evaluating its weakness. Thus, this project provides an in-depth investigation into the duties, training methods and performance shaping factors (PSF) that adversely affect seafarers within the engine room of a ship. This will be done using a neuroimaging technique called functional near-infrared spectroscopy (FNIRS). FNIRS allows us to see the mental workload of human beings whilst conducting their duties. Past studies have used various human reliability analysis (HRA) methods to analyse human error in the maritime sector [7] [8] [9]. HRA is a scientific approach used to identify potential risk of human failure events, and to comprehensively estimate the human error probability (HEP) using experimental data, modelling or expert judgement [10].

1.2.1 Maritime Error

In the maritime shipping industry there are 3 main types of errors that occur; mechanical failure, electrical and human [4]. Of the 80% of accidents resulting from human errors, it is said that 45% of those stem from inefficiently or incorrectly dealing with a fault in the engine room [4] [11] [12] [13]. Another factor is that the majority of HRA studies are conducted with focus on bridge operations from navigational perspectives [14] leaving engine room errors unaddressed. These statistics warrant a full investigation into human error within the engine room. An evaluation of the maritime databases [1] [4] [5] [11] [15] [16] was conducted in order to obtain the most common errors within the engine room. The most common types of human error are influenced by adverse working conditions [17], environmental factors [18] and individual operator issues [19].

1.2.2 Error Causation

There is multiple hypothesis in relation to the cause of human error within the engine room. Interviews with experts in the marine sector conducted by the author, advised that engine room operators are all trained to different levels [20], experienced operators complete tasks more efficiently [21], and experienced operators can cope better with work place factors (for example, distraction) compared to inexperienced operators [22]. For clarity, experts defined experience with relation to the time working at a particular position (1 or 10 years at sea for example). However, experts did acknowledge that some experienced operators as susceptible to ‘cutting corners’ which can lead to error but on average, it was said that experienced operators still outperform inexperienced operators. Various marine accident databases were analysed, to show accidents caused due to human error within the engine room only. The accident reports were then analysed to see if there were any specific duties where human error reoccurred and any PSFs reported as a contributing factor towards the human errors. Reoccurring issues reported from the statistical analysis were; distraction 11%, multitasking 20%, fatigue 10%, engine room temperature 16%, noise and vibration 6%, and time pressure

16%. The tasks that showed to be the most consistent with human error from my own statistical analysis were: ballasting, oil transfer, machine maintenance, fuel system and sea water treatment system.

1.2.2.1 Human error

To provide a more balanced perspective with regards to human error, there are two main definitions taken into account in this thesis. The first is 'sharp end' human error, this corresponds to an active error. For example, a blunt end error is interpreted as a personal error (a surgeon that removes the wrong limb) [23]. The second is referred to as a 'blunt end' human error. This error-type is classed as a latent error. For example, an error due to workplace factors (fatigue due to excessive working hours) [24].

1.2.3 Neuroimaging

Neuroimaging is a modern and novel tool for the investigation of human performance [25]. Neuroimaging is used to evaluate operators' functional state (OFS) whilst performing tasks (experimental or daily duties) [26]. Neuroimaging can be used to look at specific areas of the cerebrum that correspond to various human executive functions, for example hand-eye co-ordination and working memory [27]. It does this by either directly or indirectly imaging the cerebral structure, function or physiology [28]. Neuroimaging has been used in previous studies in the maritime sector, more specifically for bridge operations, as a HRA technique to evaluate human error [3] [14]. One of the neuroimaging techniques used is called functional near-infrared spectroscopy (fNIRS).

1.2.4 Neuro-ergonomics

The branch of neuro-ergonomics considered in this study focuses on human limitations and capabilities, both physical and cognitive [29]. Understanding human limitations allows engineers to develop technologies and work environments so that they are safer, more efficient, and designed with the human operator in mind [30].

The main theory behind neuro-ergonomics is that human factors research and practice considers the results and theories rooted in neuroscience [31]. Modern neuroimaging techniques have the potential to identify maritime operational risks, and measure covert changes in neurophysiology, which may not be apparent in the measurement of performance [32] [33]. Due to the increasing growth of neuro-science, theories of human performance are extended or constrained when considering the results of modern neuroscience [34]. Neuro-ergonomics has potential application for research intended for improving work efficiency, without compromising the unseen wellbeing and mental workload of seafarers within the engine room [30].

The techniques used in this project include; fNIRS which is a non-invasive, user-friendly neuroimaging technique [35]. FNIRS allows for modelling, coupled with machine learning algorithms and interlinked brain-computer interface (BCI) techniques, to provide a means for decoding brain activity [28]. Integration of the above method advances the science of human performance [18], and exploits a potential for addressing questions concerning brain function within a ship engine room environment [36].

The human factors psychologist Peter Hancock conducted a behaviourist analysis of Neuro-ergonomics as a concept. Hancock began his study by looking at neuro-ergonomics critically, with a view from radical behaviourism. He stated that “I am optimistic of punctate successes here, along this line of development in the near future. More understanding in this domain will also help us distinguish between simple, quantifiable processing capacities, and what the human brain actually achieves.” [37]. Hancock concluded his investigation by stating, “neuro-ergonomic designs have proven to epitomize the marriage of pure science and application in the real-world. A greater level of insight into the symphonic productions of the neural orchestra could provide exceptional opportunities to advance human-technology interaction.” Hancock also looked at the use of new techniques for providing neuro-ergonomic signals [37].

A world leader in the field of neuro-ergonomics Raja Parasuraman [37], argues for studying neuroscience in an applied context and developing models of human performance that are grounded in neuroscientific models [36]. He also addresses the fact that sustained attention would very likely result in mental fatigue [30]. The engine room features BCIs that present complex and dynamic visual information such as: monitoring ballast tank volume whilst figuring out flow rates on ballasting tasks [38]. Parasuramen’s research looks into the effects of changing workplace and environmental factors on human performance, and then using neuroimaging and behavioural data to back up his theories. [30]

1.2.5 Functional near-infrared spectroscopy

There is a relative transparency of human tissue which surrounds the skull [33], this falls into the infra-red range [39] and haemoglobin absorbs infra-red light [19]. This allows us to infer relative change in oxygenated and deoxygenated haemoglobin [40]. It is therefore possible to continuously, non-continuously, and non-invasively monitor the concentration of oxygenated and de-oxygenated haemoglobin volumes within the human cerebrum [41]. Figure 1.2 shows the ‘banana shaped’ detection range of the infrared light.

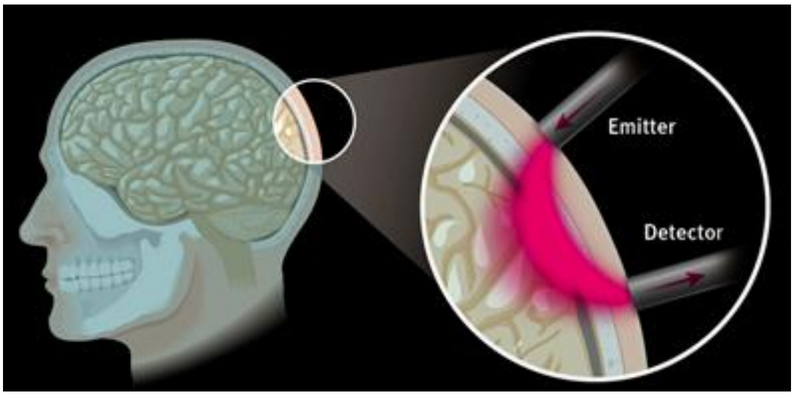


Figure 1.2 - fNIRS infra-red detection range [18]

Neurophysiological activation results in an increase in cerebral haemoglobin volume due to neurovascular coupling [34]. This coupling leads to a change in localised oxygenated and deoxygenated haemoglobin volume [41]. These changes in haemoglobin can be shown using the fNIRS system’s infra-red emitters and detectors [42], which relay the information back to the NIRx software via a desktop computer. It is evident that the higher the oxygenated haemoglobin, the higher the mental workload of the participant [43]. This indicates the areas of the task and the PSF that have the most significant effect on mental workload. Thus, higher the mental workload is associated with increased probability of lower human performance. Therefore, it is assumed throughout this thesis that the higher the mental workload of the participant the more potential there is of human error occurring [44].

For this study, participants will be connected to the fNIRS system via a skullcap containing infra-red light emitters and detectors. Together, the emitters and detectors create a channel.

1.2.6 Engine Room Simulator

A Transas ERS 5000 TechSim engine room simulator will be used to conduct the investigation. The simulator is located on the ground floor of the Liverpool John Moores University (LJMU) Byrom street campus. It closely mimics a real tanker ship engine room. Utilising a high degree of realism allows real-time, real-life exercises to be implemented exactly how they would be in the engine room of a real vessel [38]. Shown in Figure 1.3 is an image of the engine room simulator.

Figure 1.3 - Engine room simulator

The investigation will focus on the 'highest risk' tasks. The 'highest risk tasks' will be determined by accessing the marine accident databases to see which tasks have the highest frequency of occurrence (with respect to human error [See chapter 3]) and the most significant consequences due to the error. Once these tasks have been identified, an exercise will be designed and implemented on the simulator where candidates will participate in the task, (identified using most significance with regards human error within the ship accident databases) under the evaluation of the simulator controller (further information in chapter 4). The controller will see the PSF and areas of the tasks where participants experienced significant mental workload outputs from fNIRS.

Candidates participating in the simulated tasks will be connected to fNIRS as stated earlier. This provides the cornerstone of the project's novelty (to incorporate fNIRS with a marine simulator to investigate human error in ship operations). Due to the weakness in current HRA methods within the maritime sector and the relative success of the aviation sector's use of fNIRS [45], it could be said that there is an urgency to implement fNIRS in maritime human error studies. As stated earlier fNIRS shows the haemoglobin volume within different areas of the cerebrum, this happens via infra-red sensors and detectors strategically placed onto a skull cap worn by the participant performing the task [41]. Monitoring these visualisations allows us to see mental workload of each participant throughout the task [19]. Testing of participants on a simulated scenario will then allow us to implement a scientific human error model of the relationship between operator functional state (OFS) and adverse PSF's [14].

1.2.7 Current human reliability analysis techniques

A few examples of commonly used HRA techniques are; nuclear action reliability assessment (NARA) [10], cognitive reliability and error analysis method (CREAM) [46] and a technique for human event analysis (ATHENA) [47]. These techniques are all similar in approach as they work based on defining the study scope, defining the problematic, defining the PSFs and then calculating the human error probability (HEP) [10]. It will be difficult to use these techniques in a human factors study as fNIRS data has too many complexities to calculate an accurate HEP (for example, 1/1000 chance of human error cannot be determined as the HEP will differ between participants).

The next examples are HRA techniques that have been used in previous studies in conjunction with simulators. Probabilistic cognitive simulator (PROCOS) [48], Information, decision and action in crew context (IDAC) [49] and standardized plant analysis of risk-human reliability analysis (SPAR-H) [39]. These techniques are simulation-based approaches that look at a scenario in normal working conditions before applying human factors. They take into account the interaction of the operator with other crew members and their decisions based on external factors. These techniques are appealing however, they do not incorporate fNIRS. Furthermore, they use HEP to try to predict the root cause of the error which can be heavily scrutinised.

The addition of fNIRS coupled with human factors and psychology techniques accentuates the research gap with current HRA techniques used in maritime engineering and technology. The majority of the current HRA techniques aim to provide a nominal HEP value [47]. Due to the complexities of human performance against various human factors it is no possible to obtain an accurate nominal HEP value [50]. This prevents the use of current maritime HRA techniques without major improvements.

1.2.8 Neuroimaging techniques used in the studies

Many academics such as Dehais et al [45] and Vierdiere et al [51] have found human error to be an interesting field for research. Such research has focused primarily on an analysis of various aircraft operations [52], air traffic control duties [33] and the impact of new flight regulations affecting the safety of personnel [6]. These aforementioned studies used fNIRS to gauge mental workload of pilots at various stages of the tasks. Later, they modelled their findings using a performance classification model in order to deduce the pilot's ability and how each task influences the risk of human error.

Fan et al [3] [14] has conducted human error studies using fNIRS and a ship simulator. These studies used fNIRS again to gauge the mental workload of seafarers on the bridge whilst conducting standard ship operations. In these studies tasks and conditions were manipulated in order to investigate the effect of various performance shaping factors (PSF). The data was then evaluated using a connectivity matrix to analyse the relationship between functional connectivity and seafarer behaviour.

These aforementioned investigations have been conducted on aircraft or solely on ship bridge operations with little consideration of engine room operations. However, it has been documented that engine room operations have a significant impact on the 80% of maritime accidents that result from human error [11] as previously mentioned. Additionally, academics have explored maritime human error incidents with the limitations of expert opinion resulting in speculative data [53]. There have been very few published papers to date investigating ship engine room operators using fNIRS technology. There have been no maritime studies to date that have successfully modelled the relationship between seafarer performance and PSFs as Fan,S et al study only provides stressor-based inter-cerebral interaction without any human error or seafarer performance models. Also, Fan et al's study used a connectivity matrix based solely on the dorsal lateral pre-frontal cortex (DLPFC) which does not take into account the interaction between frontal, temporal, parietal and occipital regions.

1.3 Research questions and scope

The scope of this study is to investigate the effect of stressors on human performance, potentially leading to human errors to occur within a ship engine room. More specifically, the PSFs that cause such errors. Once the PSFs that result in human error have been identified, the effect of each PSF needs to be evaluated. Post evaluation will show the effect of each PSF and OSF and in turn human performance. Given the effect level of each PSF on human performance, a classification performance model of said PSFs will be conducted to show predicted human error probability.

To meet the above scope, the following research questions should be addressed. The research questions below relate to the background work needed in the literature review section (chapter 2 & 3):

1 What are the engine room PSFs and tasks that are associated with the highest levels of human error?

- *What tasks show up the most relating to human error in a ship engine room on maritime accident databases?*
- *What are the most common PSFs on the accident reports associated with human error in a ship engine room?*

2 How to develop an experimental but realistic scenario on a simulator, incorporating PSFs and fNIRS?

- Conduct a neuro-ergonomic examination of engine room problem scenarios.
- What task design methods are used on previous BCI-fNIRS studies?
- What are the rules and regulations set by maritime governing bodies involved in engine room tasks?

3 What techniques are used to investigate the effect of PSFs on OFS using fNIRS and simulators (BCI-fNIRS) in previous studies?

- *What areas of the brain to look at?*
- *fNIRS system settings and software used?*
- *How have tasks been manipulated to investigate PSFs and in what format?*
- *How to statistically analyse fNIRS and the software used?*

4 What methods have been used to model human error using fNIRS and simulators (BCI-fNIRS) in previous studies?

- - *What are the current HRA techniques in maritime engineering and technology and how can they be improved?*
- *What is the research gap with current HRA techniques in the maritime sector?*
- *What modelling techniques have been used in BCI-fNIRS studies?*
- *What is the accuracy of each model compared to one another?*
- *How can the model be changed to suit this research?*
- *How can the model be validated?*

5 How to accurately obtain a classification performance of each PSF?

- What are the optimum features to use?
- Should PSFs be classified against one another or against a standard test?
- What techniques improve classification percentage?

1.4 Research Aims and Objectives

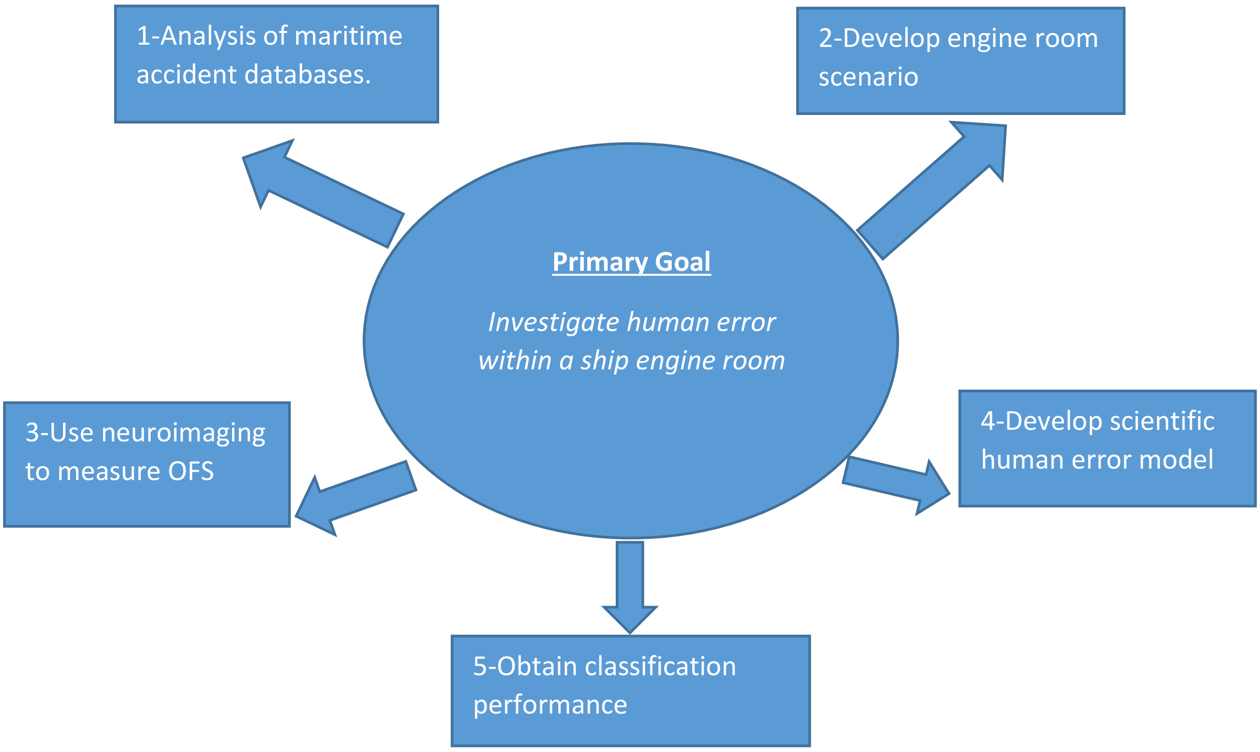


Figure 1.4 - Primary aim and objectives diagram

Figure 1.4 depicts the primary aim of this research is to investigate the cause of human error and the human errors associated with ship engine room operations. More specifically, to evaluate the engine room tasks and the performance shaping factors most associated with human error within a ship engine room.

The novelty of this study is to use fNIRS to investigate PSFs in the maritime industry, more specifically in marine engine room scenarios with operators experiencing an increased workload, a distraction or whilst they are fatigued.

Given the primary aim and novelty of the study outlined above, to achieve this research aim the following objectives are briefly described in Figure 1.4 detailed below. The objectives described below differ from the research questions in the sense that the objectives primary focus is on the empirical work (chapters 5-8) as apposed to the background research done in the literature review stage found in chapters 2 & 3:

1 Analyse the ship accident reports from maritime accident databases to obtain primary data representing the tasks and PSFs most associated with human error in a ship engine room.

2. To develop a simulated scenario and incorporate PSFs into the simulated scenario based on the findings from the maritime accident databases.
3. Use fNIRS to measure the influence of PSFs on operator function state (OFS).
4. Develop a novel scientific human error model to evaluate the relationship between OFS and performance under different PSFs.
5. From the model, obtain classification performance of each PSF.

From a human factors perspective in maritime engine room operations, this study aims at investigating how different performance shaping factors generate an impact on different types of human-related ship engine room incidents. Taking into account the problems arising from traditional HRA studies on human errors, it proposed a novel evaluation of performance shaping factors contributing towards ship engine room incidents. From maritime accident databases, reports of maritime accident investigations and a data-driven neuro-ergonomic assessment of ship engine room operations are taken to form a primary database for this study.

From the performance shaping factors obtained from the above-mentioned maritime accident databases and various literature, the PSFs reported as associated with the incidents contain insufficient or non-existent data allowing for the study of the individual effect of each PSF. This research aims to fill that research gap and obtain evidence of the PSFs that induce the highest levels of neurophysiological activation resulting in the increased mental workload of seafarers to support the hypothesis of this study. Therefore, this study will investigate how ship engine room operations with the addition of PSFs will influence the mental workload of seafarers and evaluate the difference between experienced and inexperienced seafarers.

To gain a full understanding of the neurophysiological activation of the cerebrum and its relationship to human performance, an experimental study is created and implemented for mental workload research. The resultant data supports that fNIRS is a viable neuroimaging technique, which can be used in realistic simulated scenarios and define the role of the DLPFC for seafarers conducting engine room operations and the analysis of their mental workload. FNIRS examines oxygenated and deoxygenated haemoglobin in the DLPFC of ship engine room operators allowing for the understanding of the relationship between PSFs and human performance, which in turn generates insights into risk control options, training and certification.

1.5 Research Methodology and thesis composition

The work propounded in this document is composed of 10 chapters. The following is a concise description of the contents of each chapter followed by a depiction of the process in Figure 1.5:

- Chapter 1 Introduction: this chapter offers an outline of the thesis, including a brief summary of each aspect of the research background, the research gap, and the problem to be tackled. The thesis also defines specific aim and objectives, underlining the novelty and expected outcomes.

- Chapter 2 Literature review: this chapter involves a comprehensive review of the literature available, both with specific regards to the issues presented and more specifically, to assist with achieving objectives 3-8 described above in section 1.4.

- Chapter 3 The significant factors contributing to human error within a ship engine room: this chapter depicts the research conducted into the maritime accident databases and accident reports to obtain the engine room operation and PSFs associated with human error. Furthermore, this tackles objectives 1 and 2 as described above in section 1.4.

- Chapter 4 Methodology for investigating PSFs: this chapter describes the processes used to investigate PSFs effect on OFS, as well as a description of each stage of the chosen scenario used for testing participants. It also defines how the raw fNIRS data will be filtered, processed, analysed and modelled.

- Chapter 5 The effect of distraction on marine engineers whilst conducting ballast water operations: the first part of this chapter is used to test the experimental design and methods described in chapter 4. Based on the outcome of the initial study, a second study is conducted to accurately investigate the PSF ‘distraction’, and its effect on OFS using Statistical Package for the Social Sciences (SPSS).

- Chapter 6 The effect of fatigue on marine engineers whilst conducting ballast water operations: this chapter is a continuation of the techniques used above in chapter 5, except the scenario is adapted to investigate the PSF ‘fatigue’, and its effect of OFS, again using SPSS.

- Chapter 7 The effect of increased workload on marine engineers whilst conducting ballast water operations: this chapter is a further continuation of the techniques used in chapters 5 and 6 with the scenario manipulated further to investigate the PSF ‘Increased workload’ and its effect on OFS. Again, SPSS is used.

- Chapter 8 A comparison of the PSFs; Distraction, Fatigue and increased workload: this chapter uses the FNIRS data gathered from chapters 5,6 and 7 combined with SPSS to investigate the PSFs against one another. This follows on from the investigations conducted in chapters 5,6, and 7, which investigated PSF against a standard scenario.

- Chapter 9 The development and implementation of a scientific human error model: this chapter describes how the fNIRS data was processed, features extracted and cross-validated to develop a classification performance model. Also included in this chapter is the classification performance results for each PSF giving clarification of their level of risk to human performance and human error probability compared to one another.

- Chapter 10 Final conclusion: this chapter defines the research limitations with critical analysis, how the project’s objectives where achieved, contributions to knowledge and how this project could be used for future research.

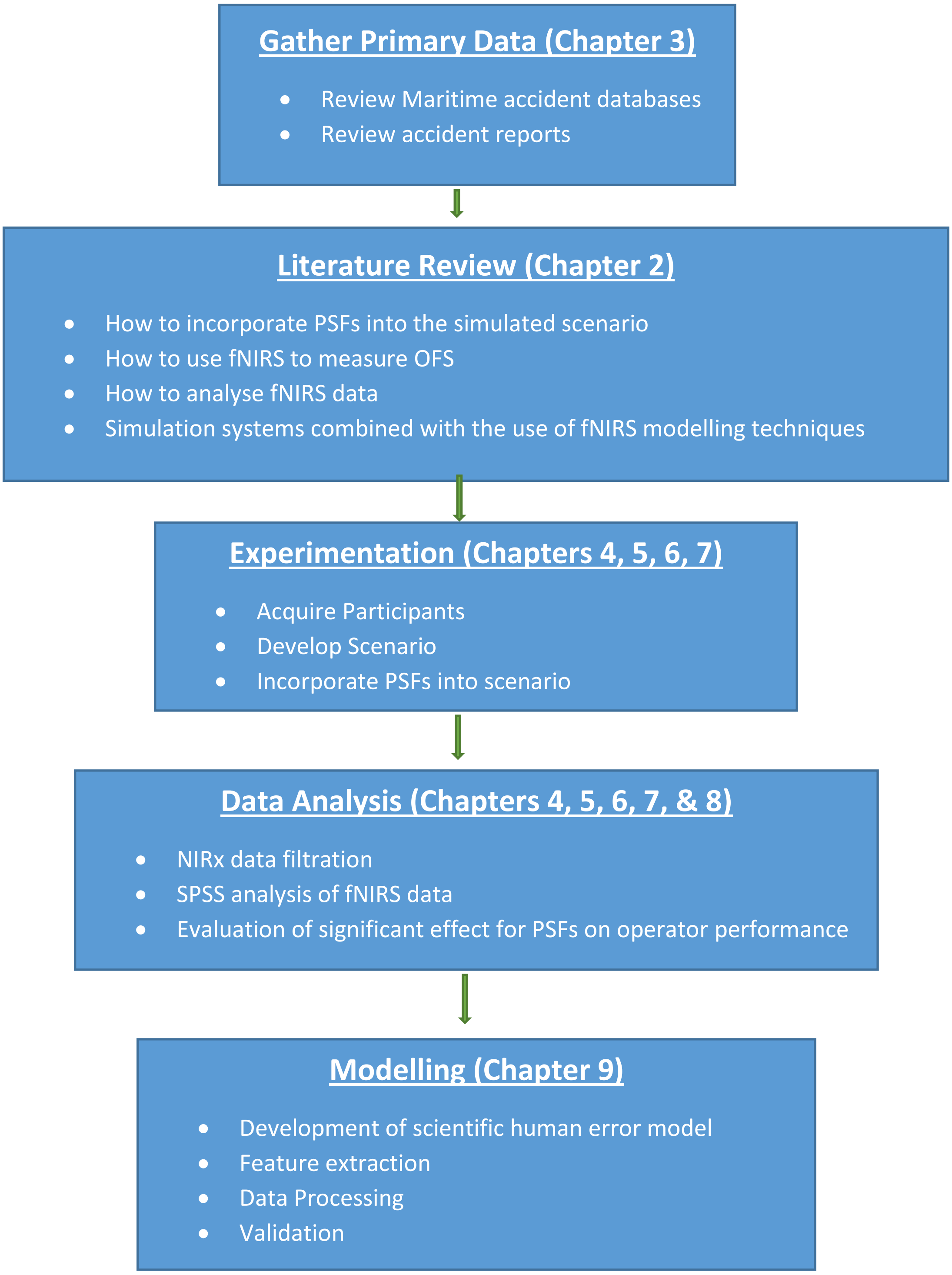


Figure 1.5 - research methodology process flow

1.6 Novelty of the study

The novelty of this study is to use fNIRS to investigate PSFs in the maritime industry, more specifically in marine engine room scenarios with operators experiencing an increased workload, a distraction or whilst they are fatigued. A number of studies have been performed using fNIRS in other sectors. For example; aerospace, psychology, national rail and automotive. However, in the maritime sector there are very few studies that use fNIRS to investigate human error and the few that do can be scrutinized due to the inaccuracies of the virtual reality equipment used. Furthermore, there are no known studies to date that model human error using fNIRS coupled with a ship simulator. Fan et al. 2020 [52] investigate stressors using a bridge simulator and fNIRS but do not model human error, instead functional connectivity is used to see cerebral interaction in increased mental workload tasks. Although the studies mentioned above show some appeal, such studies reveal some complacency with regards to maritime human error from a practical applications perspective as there are zero studies to date that investigate error within a ship engine room.

1.7 Concluding remarks

The following are the most consequential remarks comprised in this chapter, and emphasised as bullet points for the reader's ease:

- The primary aim of this study is to investigate the cause of human error and the human errors associated with ship engine room operations due to the effect of stressors on human performance. More specifically, to evaluate the engine room tasks and the performance shaping factors most associated with human error within a ship engine room.

- The use of fNIRS coupled with ship simulators could serve as a useful tool to bridge the research gap in the maritime sector by better understanding human performance against various human factors commonly experienced during 'day to day' ship engine room operations. This could then lead to the reduction in human error

- The neuroimaging technique fNIRS is a novel technique to the maritime engineering and technology sector and provides this project's novelty. The use of fNIRS can tackle the research gap of high human error levels as a result of inaccuracies in data collection methods, thus a less accurate data analysis, a neglect of a human factors approach to current HRA practices and negligible investigation of a ship engine room department in the maritime sector.

Chapter 2. Literature Review

2.1 Introductory remarks

The chapter that follows will comprise of three main sections. The first section will be looking at ways in which PSFs have been incorporated into simulated scenarios as this will assist with objective 3. This section will compartmentalise the engineering sectors for the reader's ease and to show the prospective literature for each sector. The second section will be to determine the most proficient techniques for using fNIRS to measure OFS, this will assist with objective 4 and 5. The third and final section will be looking at the modelling techniques used, specifically with fNIRS data and simulated scenarios as this will help to achieve objectives 6, 7 and 8. Each section will contain a critical review of the various techniques, additionally with the primary objective of how the current literature can contribute to the existing maritime engineering knowledge.

2.2 Investigation into the incorporation of PSFs into simulated scenarios

Figure 2.1 is an illustration depicting the primary aim of this section. Each aspect reviewed to achieve the primary aim is depicted and briefly summarized in the blue boxes for the ease of the reader.

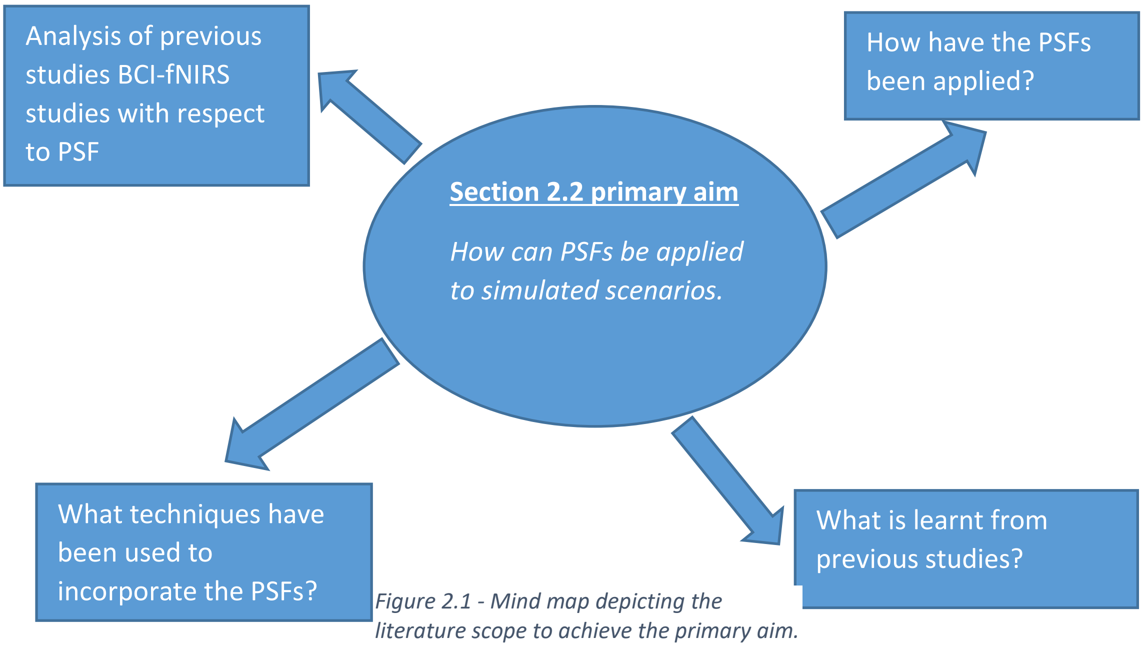


Figure 2.1 - Mind map depicting the literature scope to achieve the primary aim.

2.2.1 Criteria for the literature search

Before beginning the literature search, a scope of the review has to be defined. The research objective three states, "Incorporate PSFs into the simulated scenario based on the findings from the maritime accident databases." This will allow for the understanding of how best to manipulate maritime engine room operations to incorporate PSFs. Moreover, to allow for fNIRS to detect OFS whilst experiencing said PSFs further down the line. Therefore, the focus of this section of the literature review will be investigations testing the ability of train drivers, aircraft pilots, automobile drivers and ship captains. It is already known that the engineering sectors previously mentioned conduct academic and workplace studies for research purposes and to test the ability of their employees. These aforementioned studies have been performed using simulators but there aren't any in the maritime sector to date that use fNIRS as an analysis tool. Some simply witness the participant either passing or failing a task. However, the aerospace sector is the sector that most commonly uses fNIRS in conjunction with aircraft simulators, so the literature search will start in that sector. Peer reviewed journals, conference articles and technical papers are used to conduct this literature review.

The literature search keywords and phrases are: **aircraft, aircraft pilot, national rail, train driver, ship captain, bridge operations, automobile driver, simulator, fNIRS, stressor, PSF** and **human error**.

2.2.2 The aerospace sector.

Gateau et al [54] investigate the pilot's functional state on a flight simulator versus a real aircraft. In this study the pilots are given working memory tasks consisting in repeating series of pre-recorded air traffic control instructions, easy versus hard. From a PSF perspective these tasks could be interpreted as workload or distraction. Dehais et al also conducted another study [55] again, investigating a pilot's inflight working memory using air traffic control commands. This study again used an aircraft simulator in conjunction with fNIRS. In this study Dehais and his team tested pilots by relaying verbal codes with varying difficulty (letter and numbers, from 4 -10 consecutive letters/numbers) and the pilots would be required to relay these alphanumeric codes at a later part of the flight. This study tests pilots' working memory and is also a distraction. In this study Dehais et al did not expect the pilots to be able to memorise the more difficult codes, so wanted to understand their functional state at their working limit. Vierdiere et al [45] investigates automatic pilot against manual landings on an aircraft simulator using fNIRS to evaluate OFS. This study is an evaluation of increased workload as an auto-pilot landing has significantly fewer tasks than a manual landing. Simply increasing the pilot's landing tasks induces a study of a workload volume PSF. Roy et al [56] monitor a pilot's cognitive performance in naturalistic environments, again, using a flight simulator and a light aircraft. In this study EEG combined with fNIRS is used to monitor OFS. This investigation uses simulated traffic patterns followed by an auditory task with varying degrees of difficulty.

2.2.2.1 Discussion of the techniques used in the aerospace sector

The above studies in the aerospace sector show how fNIRS can be used in conjunction with a simulator system. The first study investigated OFS of working memory tasks using a simulator against a real flight scenario. The results of this study showed that the pilots in real flight conditions committed more errors and had a higher anterior pre-frontal cortex activation when compared to those in the flight simulator when completing the working memory tasks [54]. A real versus simulated scenario would be difficult to investigate for this study as an engine room simulator is readily available however, a real ship engine room is not. It could be predicted that the activation of operators would be higher in a 'real life' scenario compared to a simulated environment due to a simulated environment always being an impoverished version of the 'real thing'. A positive to take from this study is the investigation of working memory. They asked pilots to repeat air traffic control instructions of varying difficulty whilst in simulated and real flight. This technique could be incorporated into our study by distracting participants temporarily, requiring them to then remember where they were up to in the task to fully complete the task. A hypothesis would be that under harder distraction tasks the participants would forget where they are up to in the task resulting in either missed sub-tasks or errors.

In Verdiere et al study [45] an investigation was conducted into the differences in activation between a manual versus automated landing scenario on a flight simulator. This is a fairly straightforward test of workload. This study showed the difference in cerebral activation, by way of oxygenation and connectivity features, between the differing landing scenarios. The results of this study were promising due to showing classification performance percentages way above change. This technique is difficult to replicate for our study due to an engine room environment being much less dynamic than a landing scenario. For example, an automated scenario in the engine room would involve participants simply monitoring screens. However, workload investigations could still be incorporated into our study. An example of this would be to increase the number of sub-tasks a participant would need to complete the overall task.

The other three studies [33] [55] [56] are similar investigations of a pilot’s working memory. These studies incorporate air traffic control messages (ATC) into a flight scenario. The results of these studies all show that the majority of errors occurred when the pilots were relaying the high difficulty messages. This could have been predicted prior to the investigation but what could not have been predicted is the level of effect on cerebral activation the varying difficultly of ATC would have. This technique could be incorporated into a maritime engine room study in a similar way as mentioned in the two paragraphs above. The techniques in this study could also be used in our study to investigate a number of PSFs. For example, easy or hard working memory sub-tasks repeated a large number of times could induce fatigue [57] or investigate increased workload [19].

In summary, the aerospace sector has the most literature with regards to fNIRS coupled with simulator systems. However, the techniques seen in this sector are somewhat different to ones that could be used in a ship engine room due to the dynamic nature of an in-flight scenario. Plus, the majority of investigations in the aerospace sector investigate simulator PSFs. The difficulty of an engine room investigation of PSFs on a simulator would be showing a physical consequence of an error or action. Whereas, on a flight simulator the participant could experience a visual and physical effect of error. For example, the plane crashing.

2.2.3 National Rail sector

Kojima et al [58] investigated human error using fNIRS combined with a train driving simulator with the goal of developing driving support systems. The authors of this article hypothesize that due to the monotony of train driving the train driver’s low concentration levels may be a cause of human error. The task involved stopping at three stations along a 2.3km railway line. The stopping time at each station was fifty seconds in order to measure changes in blood flow during a stop. This study can be scrutinised due to; the study involved only 2 participants, neither of the 2 participants had any experience driving trains, the tasks lasted for only 6 minutes, it could be argued that 6 minutes is not long enough to lessen concentration [59]. However, the hypothesis and results of this study have some relevance to ours as it was found predictably that activation occurs during a station stop. Less predictable was the fact that activation lessens whilst driving the train without stops. In national rail operations the station stops are scarcer which could result in lower driver arousal levels [32].

Concentration levels could have been investigated in a better way by increasing the time between stops. This could also be incorporated into an engine room scenario as many of the engine room systems are automated and there are tasks involving long periods of monitoring [60].

In summary, limited articles in the sector involving simulators and fNIRS could be found. The additional literature are revisions of the paper discussed above from the same authors with a similar outcome.

2.2.4 The automotive sector

Huve et al [61] investigated human error using fNIRS coupled with a driving simulator. Their paper looks at the mental state of the car driver whilst undergoing PSFs such as; weather condition, type of road and auto-piloted vehicles vs manual driving. The weather conditions and road type PSFs were incorporated into the brain-computer interface (BCI) using simulator programming. For example, weather conditions like rain would reduce visibility and tyre traction. Road type would test the drivers at differing speeds with varying potential hazards. These PSFs would be difficult to incorporate into an engine room scenario as there is no option to change weather conditions within the engine room simulator programming. However, noise, vibration and temperature could be incorporated into our study by using external hardware and equipment. For example, a portable heater to replicate the adverse temperatures experienced within the engine room [7]. An easier PSF to investigate in the engine room simulator would be automated vs manual operations. In a ‘real life’ scenario the majority of engine room systems are automated [62]. Therefore, requiring participants to simply monitor screens, whilst automated systems are engaged against those operating the systems manually, could be done. Huve et al found that participants in their study had reduced activation whilst the simulator was in ‘auto-pilot’ mode. They hypothesised that this could cause human error if a driver had to quickly react to an incident due to losing concentration because of long periods of auto-pilot. However, when investigating this hypothesis Huve et al did use long periods of automation. Instead, they looked at the levels of neurophysiological activation of auto-pilot vs manual operation, the results of which could be predicted prior to the study. Investigating this using an engine room simulator could be done better by requiring the participants to have a long monitoring of automated system screens task followed by a fault with a system. This could allow for the investigation of concentration levels and mental fatigue.

Yamamoto et al [63] used fNIRS and a driving simulator to investigate a driver’s cerebral activation and behaviour on sighting a road/message sign. For example, change of car’s velocity and acceleration/deceleration and the resulting neurophysiological activation induced by such behaviours. This was evaluated by programming various road signs into the BCI. The results of the study showed higher levels of pre-frontal cortex activation for drivers that reacted physically to the road signs. This could be predicted as a physical reaction to a message would result in a neurophysiological activation of the pre-frontal cortex [18]. Interesting is that some drivers had no reaction or activation to road signs. This was not a human error study so it is difficult to determine which driver had the highest risk of error (the driver with an increased mental workload due to a reaction, or the drivers that had lower levels of activation by not reacting to the signs). These techniques could be translated into our study by distracting participants with questions or messages in the middle of their task. This could show the effect of working memory and executive functions of the participant whilst conducting an engine room task. This would allow us to better understand how a distraction effects cerebral activation.

Tanveer et al [64] investigated the effect of driver-based drowsiness using fNIRS. In the investigation deep neural networks (DNN) were used to classify the drowsy and alert states of participants. The drowsy state was obtained depriving the participants of sleep. The method used to sleep-deprive the participants is not stated. This study shows that fNIRS is a valid method for detecting fatigue and drowsiness, given that the result showed higher levels of activation in sleep deprived participants. The technique of sleep depriving participants could be used in our study to investigate the effect of fatigue and drowsiness. However, the paper by Tanveer et al [64] leave the reader with the question – How were the participants specifically deprived of sleep? Whilst their technique is proved to be valid by the classification accuracy of 99.3% from the convolutional neural network, it will still require further research in order to specifically define 'sleep deprived' participants.

In summary, a lot of the techniques applied to automotive studies can be incorporated into a maritime study. Similar to the other sectors discussed above there is a dynamic element to driving simulator tasks which is not the case for the engine room simulator. A visual environmental change as a consequence of an action is found in driving simulators were as, the engine room simulator could incorporate an auditory and visual alarm [38] to incorporate a dynamic BCI response to a participant's actions.

2.2.5 The maritime sector

Fan et al [14] investigated the effect of human factor concepts (mental workload, attention and fatigue) using fNIRS and a bridge simulator BCI. Similar to the techniques used in other sectors, S Fan et al manipulated the simulation system to induce fatigue by using a long watching task with a large amount of fog surrounding the ship. The workload stressor was induced by having each participant complete read-outs of navigational information via the bridge radio system. These techniques have been made valid by previous literature in other sectors but this paper fails to quantify the results from the PSFs. Therefore, it is difficult to see an accurate effect of said PSFs. However, the paper does show that neurophysiological activation did occur for participants involved in workload and fatigue studies. These techniques could be replicated on a bridge simulator as there are many systems within the engineering room that would require readouts of information at various parts of a task [38] [65]. Also, similar to the officer of the watch task used in the Fan et al paper, engineering staff would have a variety of automated systems that would require long monitoring tasks [21] potentially inducing fatigue.

Fan et al [3] investigated the effect of a seafarer's emotion whilst completing complex and demanding bridge operations. In this study a bridge simulator coupled with an electrocardiogram (EEG) was used. The use of EEG can be scrutinised as studies have shown that EEG induces many anomalies and artifacts due to noise, vibration and movement of the participants [18]. Given that the participants are being tested on a bridge simulator, noise, movement and potentially vibration would be a colossal factor [66]. The emotional response of participants was induced using sound clips of the international effective digitalised sounds (LADS). This result of this study from EEG features were 77.55% using speculative data by participants and a support vector machine classifier (SVM). A result of 77.55% is low however, this may be down to the use of SVM as a classifier. It is difficult to use a SVM classifier due to the parameter settings and choice of functions. To achieve a satisfactory result, a large amount of experimentation is needed to obtain the correct parameter and function settings [67]. Also, speculative data gathering can lead to inaccuracies due to the participant believing they did better or worse in a task than they actually did. A study of the effect of varying emotions could be valid for our study but it is hard to define an emotional effect from sound clips without using a very large number of participants [6].

Fan et al [68] investigated the difference in seafarer experience level during a watchkeeping task using fNIRS. The task involved a 20-minute period of sustained attention to locate a large vessel in the distance and a 10-minute decision making period (whether to manoeuvre the ship away from collision or not). Seafarer experience levels was defined by completing an officer-of-the-watch course using a bridge simulator. The results of this study showed varying levels of activation between experienced and inexperienced participants. It would have been good to see quantification of the results by way of a classifier instead of a connectivity study as connectivity simply shows the areas of the brain that are activated at the same time [52]. Another criticism would be that only the pre-frontal cortex was used in the study. Therefore, the results could have been easily predicted as the connectivity of the pre-frontal cortex during a task involving executive functions is already known [41].

In summary, there is very little literature of fNIRS being using within the maritime sector and all the current literature focuses solely on the bridge. This leaves a research gap as the engine room has not yet been explored.

2.3 Analysis of the use of fNIRS on simulator systems to measure OFS

Figure 2.2 shows the primary goal and the research needed to achieve the goal (blue boxes). Each box contains a question needing to be answered by the literature being reviewed. This gives the reader a guide as to how this section of the chapter will be conducted.

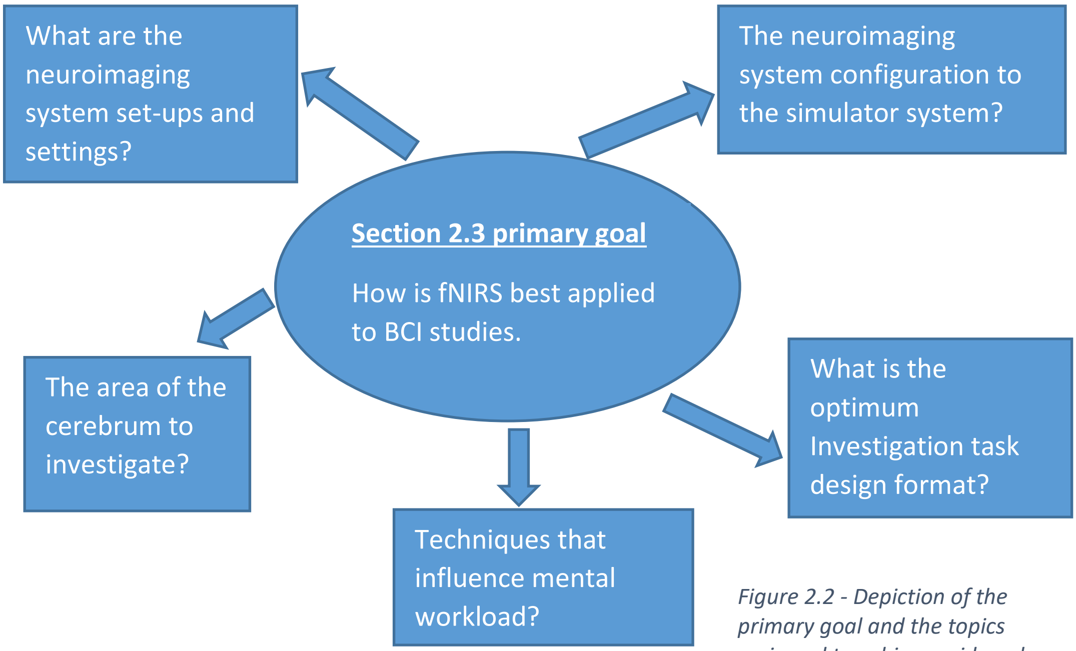


Figure 2.2 - Depiction of the primary goal and the topics reviewed to achieve said goal.

2.3.1 Criteria for the literature search

Before beginning the literature search, a scope of the study had to be defined. Objective three, "Use fNIRS to measure the influence of PSFs on operator function state (OFS)" Therefore, the topics used in this literature search will be; fNIRS to measure OFS using simulators, fNIRS filters and settings to collect accurate data. The main focus of this section of the review being: how fNIRS can be correctly combined with simulator software, fNIRS configuration and settings used and the best methods to filter and format the raw data. Reviewing these topics will answer the question four, "how have fNIRS and simulators been used to investigate the effect of PSFs on OFS in previous studies?"

The literature search keywords are: human error, simulator, mental workload, OFS and fNIRS showed a significant number of papers in the fields of aviation, national rail network and automotive sectors. Peer reviewed journals, conference articles and technical papers are used to conduct this literature review.

2.3.2 The techniques used in previous fNIRS and BCI studies.

The pre-frontal cortex was monitored throughout previous studies using a continuous wave fNIRS system and simulator in the following sectors; Automotive [28], Rail [58] and aviation [31] [45] [69]. Monitoring of the pre-frontal cortex in the aforementioned studies can be further substantiated given that the pre-frontal cortex is the area of the brain that governs executive functions like working memory [44]. Therefore using 7 emitters and 7 detectors all together, creating 15 channels, we can fully evaluate pre-frontal cortex (more specifically the dorsal lateral pre-frontal cortex [DLPFC]) This montage is shown in Figure 2.3.

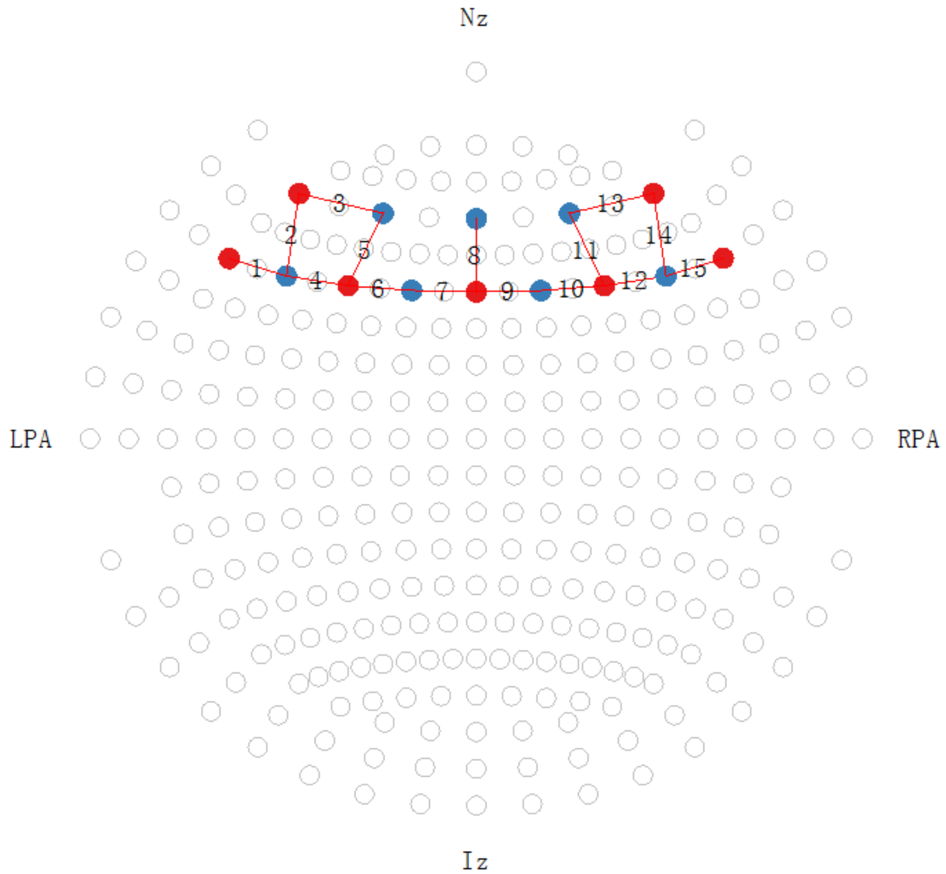


Figure 2.3 - Skull cap montage of optodes

Candidates can then participate in a task (coupled with PSF's) on the engine room simulator, allowing us to evaluate the cause of any increases in mental workload. This is shown by the level of neurophysiological activation [51] (the activation will be higher when the participant is under a higher mental workload [45]). The aforementioned techniques have worked well in past studies. However, to better achieve our first objective it has been shown to be more beneficial to focus on the left side 5 and right side 5 channels and have less focus on the middle DLPFC region [18]. The middle region of the DLPFC is less utilised in practical or working memory tasks [70]. This has been shown in previous studies to cause anomalies in oxygenated haemoglobin data [19] [71].

In the aforementioned studies, participants performed a visual identity task, a standard working memory and attention task with varying levels of difficulty. Visual identifying, working memory and attention tasks used in the previously mentioned studies using fNIRS with a simulator are all similar in approach. Therefore, these techniques can be replicated using our TRANSAS engine room simulator. To critique this approach, it could be said that there is no adverse dynamic element when a participant makes a mistake or performs poorly. In the case of a flight simulator, the adverse dynamic element of a poor performance in the worst case would be a simulated crash but this is not the case with simulated engine room tasks.

Papers reviewed outline a similar system, where simulation software is used to conduct an experiment on a participant undergoing real life scenarios in the aviation [45], automotive [42] and Rail [58] Sectors. In each case fNIRS is applied in order to break down the specific periods within each task that show the highest mental workload [41]. The sensitivity of fNIRS is a criticism and acclamation. Previous studies show that a lot of time needs to be spent on pre-processing of fNIRS settlings and post-processing of fNIRS data [55].

Shewokis et al [72], compare EEG to fNIRS and conclude that EEG has many excellent qualities for monitoring mental workload [73], including superior temporal resolution but is limited in its capacity for spatial resolution [54]. In addition, EEG has a long set-up time and susceptibility to motion artifacts are an issue [33]. EEG has a substandard spatial localization when compared to fNIRS [51]. Recent studies have found that the use of EEG with simulator hardware has an accuracy output of below 60% [18] [74] [75] therefore, the use of EEG with simulators is now deemed inadmissible [76]. This validates the use of fNIRS over EEG. One aspect that EEG does have over fNIRS is the ability to detect cerebral change instantly whereas, fNIRS takes approximately 8 seconds [40] but this issue can easily be overcome by analysing data 8 seconds ahead of the task time.

Aviation studies looked at the difference in mental workload focusing on the oxygenation features; average, peak, skewness, variance, slope, kurtosis and area under the curve and connectivity features; Covariance, Pearson's and Spearman's correlation, Wavelet coherence and spectral coherence [45]. Connectivity is using neuroimaging in order to measure and understand functional networks across different areas of the brain [57]. By plotting a scatter graph of two data sets from different areas of the brain against one another, one can see if the two corresponding areas have a positive or negative coherence [19]. A positive coherence will show a connection between the two parts of the brain (for example, when one specific part of the brain is used the other specific part is also activated) [57]. The paper concluded that the best connectivity method to use was wavelet coherence [35]. However, the aforementioned studies looked at the connectivity between the occipital and frontal lobes of the cerebrum [55]. A further benefit would be to look at the connectivity between left and right sides of the DLPFC [14].

Further literature showed that an advantage of a connectivity study is that predictions can be made as to the phase of the task the participant is up to in the workflow [19] and show areas of the brain that interlink in different tasks [57]. Connectivity would be a good tool to use if the objective study was to create a predictive model that shows operator's workflow stage in order to obtain when the participant is likely to be under a high mental workload [55]. This study will be evaluating the effect of PSF on operator performance in order to obtain HEP via classification performance. The only benefit to our project would be to use connectivity features but a connectivity study takes a lot of time and the difference between connectivity features and oxygenation features from previous studies is minimal [45].

Oxygenation features can provide us with quantifiable oxygenated haemoglobin volumes [45]. These quantifiable volumes can be put towards a nominal value of PSF effect which in turn can allow us to accurately evaluate HEP [10]. For example, high oxygenation volumes indicate high mental workload [18]. High mental workload indicates a higher likelihood of error [69]. The PSF can then be compared on this basis indicating the PSF with the highest likelihood to contribute towards human error. Given the knowledge of the PSF, we can then ascertain HEP using frequency of occurrence and consequence [53] in terms of mental workload.

In previous studies two NIRSport systems were used in tandem in order to increase the number of sensors [19] [45]. Each NIRSport system had eight sources and eight detectors. An advantage of using two systems, enables the instructors to look at the frontal and occipital areas of the brain (each covered by 8 sources and 8 detectors resulting in 32 channels) in order to see the connectivity between the two areas simultaneously. The downside to this is the amount of equipment needed and additional weight. There is also the issue of hair density on the back of the participant's head having the potential for reduced optode efficiency [18]. This could also be done on a multimodal fNIRS system instead of having two separate systems running on the same participant [19]. For our project there is no value in looking at frontal and occipital areas of the brain unless a connectivity study is being conducted given the findings above.

In many of the studies reviewed, mental workload manipulation was done by varying the degree of difficulty levels of each task [19] [18] [76]. An example of this is Gateau, Ayaz and Dehais tested 28 pilots undertaking a task using a simulator vs a real aircraft. The group was split evenly into groups of 14. One group on a real aircraft and one in the simulator. The tasks consisted of two different instructions from air traffic control. The first was considered the easy difficulty level and was loading up flight parameters to the flight console. For the easy level the parameters all had the first 2 digits the same (for example speed 140, heading 140, altitude 1400, vertical speed 1400). The hard level consisted of different flight parameter values (for example speed 172, heading 258, altitude 6401, vertical speed -2801).

The task consisted of 20 trials all together with 10 repetitions of each difficulty level. The order of the task difficulty was randomly distributed with two constraints:

- The first 10 trials had five hard and five easy difficulty levels.
- The difficulty level would not be the same for two consecutive trials. [54]

The simulation process provides useful insights into the development of our simulation, however the trials and the groupings will need to be optimised to better reflect on how a simulation based fNIRs assessment should be conducted in a systematic way, including the task selection and trial definitions in order to achieve the project objectives. Varying the difficulty can be done by applying a multitasking aspect similar to that used in (National Aeronautics and Space Administration Task Load Index) NASA TLX tests [77]. To criticize, another aspect to explore would have been environmental changes as this has been shown to be a relevant factor in other sectors [78] [79]. For example, the effect of adverse weather when landing an aircraft.

The varying difficulty levels has been shown in the studies mentioned above to be a good evaluation of operator performance. This can be done similar to Dehais et al (figure 2.4), using a workflow task where the difficulty increases gradually. An engine room fault could consist of visually acknowledging the occurrence of a fault (easy), detecting what caused the fault (medium) and finally solving the problem (hard) [60]. This contributes towards objective 2.

In papers by Ayaz et al [72] and Verdiere et al [45], fNIRS was used to record haemodynamics of the prefrontal cortex. For the pre-frontal cortex the device was equipped with 16 optodes (8 emitters & 8 detectors). The optode separation was approximately 25mm and two different wavelengths were used (730 & 850 nm) as these specific wavelengths were shown to be the optimum wavelengths for recording de-oxygenated and oxygenated blood flow [18] [33] [54]. Each optode was set to record oxygenation variation levels at a frequency of 2Hz. The disadvantage of using 2Hz as a sample rate is that less data is recorded per frame [19]. Therefore, a better rate to use for our project would be 7.8Hz in order to fully show the cerebral fluctuations [80]. This again, helps with the first objective as it gives a guide to fNIRS system settings used and the optode configuration.

Synchronization markers are used in order to indicate when each participant had periods of rest and when certain tasks were undertaken. The advantage of this is the markers clearly show the point at which the participant is in the task, the mental workload of each task and how long each sub-task took to complete [57].

Oxygenation changes for each optode is calculated using the modified Beer Lambert law [51] as stated in 1.2.3 A consistent theme in all literature is that a modified version of the Beer Lambert law is used to calculate changes in concentration of oxygenated and de-oxygenated hemoglobin levels relative to a baseline [45] [54] [73].

In the aforementioned literature the average change in oxygenation is taken as the depended measure. A model was then developed for each participant based on the fNIRS data measured during the tasks. The linear models on the papers mentioned above showed the oxygenation polynomials at points of the tasks and at rest in order to correctly define mental workload at specific points throughout the exercise [39]. The advantage of this is that various sections of data can be separately analysed to different levels of complexity (for example, HRA techniques such as a Bayesian network can be implemented to look at the reasons behind oxygenation spikes at different parts of a workflow) [81].

The consensus is that a ‘workflow style’ framework and task design is used [54] [57][71]. An example of this type of framework is used below (Figure 2.4):

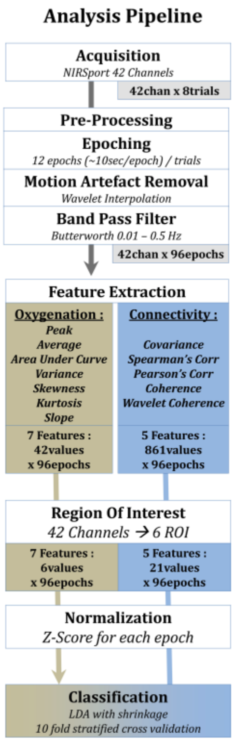


Figure 2.4 - Example framework [45]

An advantage of using this type of framework is that it gives a good structure to the order of the analysis stage. If an error is made, then it is easy to go back to the section prior as opposed to re-starting from the beginning. A disadvantage of the pipeline in Figure 2.2 is that it was developed to test pilots in flight. Therefore, 42 channels, 12 features and 6 regions of interest is more than what is needed for this study. Our study is solely looking to achieve an accurate classification performance to obtain PSF and HEP values.

The evaluation of fNIRS data based on a general liner model is often made more onerous by high inter-subject variability of the haemodynamic response, serial correlations and the presence of motion artifacts [82]. For this reason, a Moving Average Convergence Divergence (MACD) filter is used to remove all the low and high frequency components from the raw fNIRS data [45]. The low order filter has a quasi-linear phase in its bandwidth and is particularly suited for real-time applications [41]. This is another good guide for our objective 1 as it validates the use of a low order filter as opposed to a high or band pass filter. Baker et al, for their raw data analysis, used NIRx filtering of Δ(HbO2) and Δ(hHb) on all individual optodes [19] and this worked well. The data can then be used to show the mental demand of the participants whilst performing said tasks [57]. This is done again as stated in section 1.1.1 using the difference in oxygenation values.

2.3.3 Discussion

There is a difference between the maritime engineering PSF and most other sectors. The main differences being that on board a ship, PSF changes will happen regularly, for example, rough sea, a noisy engine room, a cold engine room or a hot engine room etc. In the aviation sector the PSFs have the most similarities. For example, the number of aircraft to navigate [73] (this is similar to multiple ships in port), the difference between tasks with varying difficulty levels [54], memory based tasks [43] or attention monitoring tasks [32] and the environment itself (weather, temperature, vibration etc). However, it would have been good to see what scenarios forced the pilots to use autopilot or manual functions in Verdiere et al's study (for example, low visibility, high winds, technical problems) as it would have been good to see what PSF condition caused the highest mental workload for pilots.

The analysis of the fNIRS data was done in most other studies via the use of third-party software [19] [32] [33] for example MatLab and homer2 software. This can be used for our project however, the fNIRS system that our school of engineering has invested in comes with its own data analysis and filtration software. It would have been good to see how Verdiere et al would have used NIRx but at that time the latest version of NIRx wasn't available to them.

The framework that is used in Dehais et al's study can be used in a similar way to that on our project. The framework is done in a very similar way to that of a workflow (see figure 2.4). This could be translated into a process framework to go towards objective two and three. However, the framework would have to be changed to suit the engine room. For example, the participants used in Dehais et al's paper are already fully trained and qualified pilots therefore a training element would have to be included in our study. Another factor is the engine room tasks involve a significantly larger number of sub tasks [60] when compared to the task used in Dehais et al's paper. Therefore, the number of epochs and their lengths would have to be changed. Given the previous differences described, the PSFs in our study would have to be incorporated at differing stages of our task or throughout the whole task. Furthermore, it can be assumed that the PSFs of an engine room investigating will differ in nature to those found in flight. For example, an in-flight distraction could be dynamic elements like; weather, air traffic, changes in velocity/groundspeed and other visual obstacles from the environment [33]. Whereas, engine room simulator operators will not have any view of the outside environment or consequences or an invalid action [83]. An example of this is ship listing. Given the previous differences, describing this framework could be doctored and be coupled with the integration PSFs to understand the relationship between OFS and operator performance by changing the elements described above. These techniques could be implemented to most engine room scenarios.

A main feature of the Dehais et al [54] and Verdiere et al [56] papers is the connectivity aspect. Their studies used 42 channels, connectivity features and the region of interest (ROI) covering the majority of the cerebrum. However, the most relevant ROI for our study would be the pre-frontal cortex as this is the area of the cerebrum that involves working memory, executive functions and motor skills [34] associated with engine room tasks [65]. A connectivity study could still be implemented using the prefrontal cortex and 15 channels alone but this would be looking at the right side and left side DLPFC as opposed to multiple regions. However, looking at the PFC alone in a connectivity study could be scrutinized as the resulting connectivity matrix could be predicted due the obvious interaction already known between the left side and the right side PFC [26].

Much of the literature reviewed utilizes a workflow style task design [40] [54] [73]. This could be translated to our project to address the existing human error problems in the maritime industry, by using a workflow to compartmentalize sub-tasks and work through each investigation in a consistent, uniform manner. An example taken from a study by Roy et al [45] is shown below in Figure 2.5. There are four time-windows in place whilst there is an instruction. This task design consists of; a baseline, a question, response and a rest by each candidate. The response windows will obviously all be different for our project given the comparative nature of each study. Utilizing a similar workflow task design will enable us to implement an engine simulator-based human error investigation of OFS incorporating fNIRS. For example, a baseline will be needed for each participant due to the physiology of each human in most cases differing [18]. Then sub-task windows would be needed similar to the ATC message, response and rest windows used in Roy et al's study but for our study it could be used to compartmentalize the whole task to better understand the areas of interest [68].

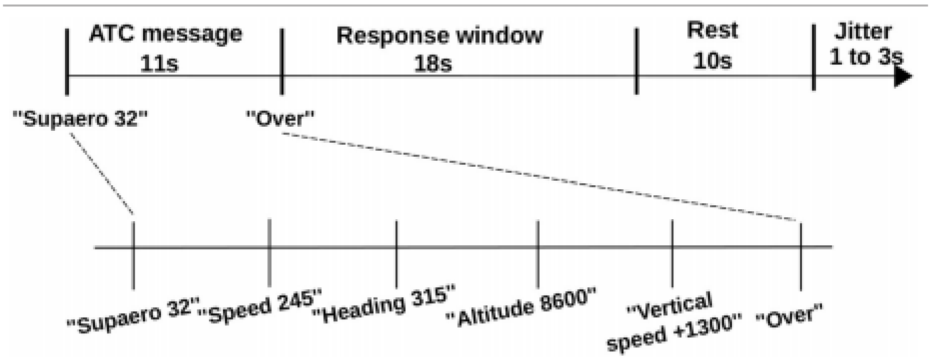


Figure 2.5 - Workflow style task design example [73]

To summarize, there are many drawbacks from the literature reviewed in this section. The research gaps in maritime human reliability assessments are primarily the lack of an accurate and quantifiable way of collecting human data. The majority of current HRA studies rely vastly on speculative data and expert opinion. Therefore, fNIRS will provide a novel and accurate way of providing human error data collection [84].

2.3.4 Comparison of Techniques

The TRANSAS engine room simulator is somewhat different to other simulators in the sense that there is a relatively low visual output of dynamic change due to an incorrect operator action. That being said, the task design and use of fNIRS technology coupled with a simulator can be replicated in a similar way to the less dynamic tasks associated with the work done by Baker et al [19].

The techniques used in each study are somewhat similar in the way that fNIRS technology is implemented however, there was one subtle difference when measuring a baseline. Some studies used a baseline condition of "n-back" tests in order to establish the utility of fNIRS to measure changes in mental workload [73]. This is mainly to test the validity of their methods coupled with the use of fNIRS technology. This differs from the other studies which mainly used a 'relaxed state' or resting baseline [45] [54].

Most studies measured the pre-frontal cortex of the brain [28] but it has been determined that not all parts of the brain involved with certain scenarios could be monitored [73]. This refers to voice commands which would require monitoring of the auditory and parietal cortex. This brings about the conclusion that the experiments for our project need to have visual, written instruction or are given prior to the start of testing.

Comparing the techniques used for gathering and filtering the data is still relevant even though the system for our study has its own fully automated software used for the raw data filtering. Mayers waves and noise filters will still have to be applied to reduce the artefacts from gathered data [84]. The various studies reviewed in this section clarified that mayer waves due to blood pressure and heart rate can be controlled via a band pass filter [51]. More specifically, with a lower band of 0.01Hz [43]. The most efficient upper band pass filter from the studies reviewed showed to be 0.2Hz in order to remove the noise artefacts from various equipment and ambient sounds within the simulator [52]. Rarely, there can also be spikes throughout the data sets due to low nodal contact [85]. These spikes can be removed via interpolation of the data before and after the spike to 'best fit' the missing data removed with the spike [86]. Other techniques from the studies reviewed that can be replicated is the way in which Verdiere et al formatted the task in epochs [54] via mechatronic code written in MatLab or R-Studio, the algorithms used by Aghajani et al, to correct signal correlation anomalies [18], the way the data was formatted by Gateau et al [54], for ease of analysing such large data sets and Gevin's theory for experimentation on human participants with respect to differing human physiology between participants [34].

2.4 Current HRA and modelling techniques.

The illustration below shows the primary goal of this section of the chapter. The surrounding boxes contain the information needed to be known in order to achieve the primary goal. This depiction gives the reader an overview of what to expect from this section.

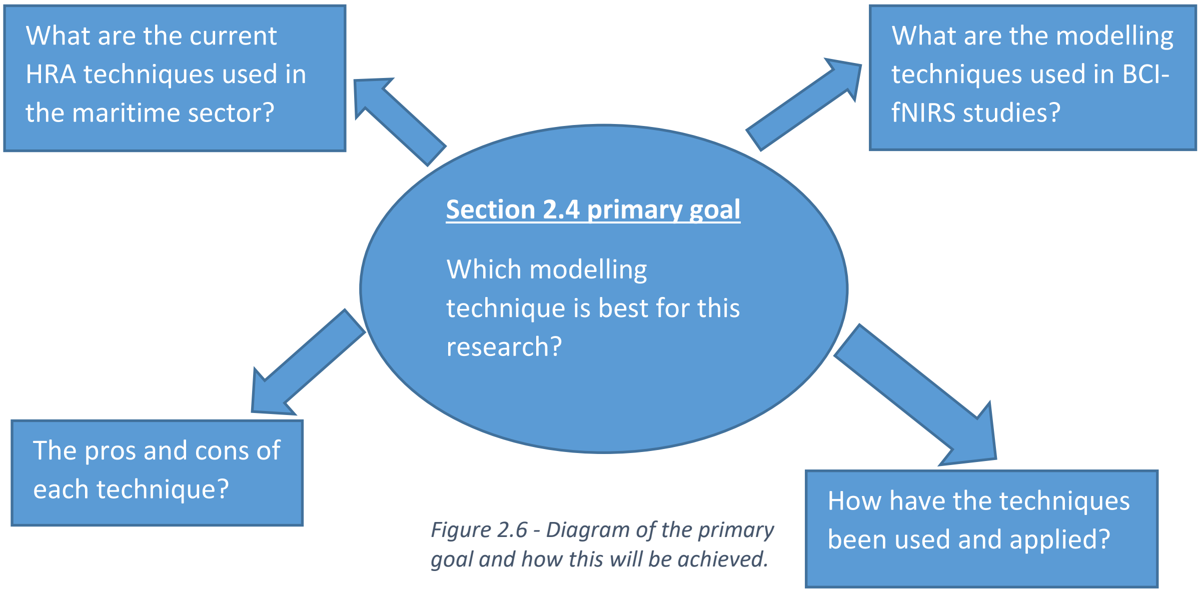


Figure 2.6 - Diagram of the primary goal and how this will be achieved.

2.4.1 Criteria for the literature search

Prior to the literature search, a scope of the study needs to be defined. Objectives three and four state, "Develop a novel scientific human error model to evaluate the relationship between OFS and performance under different PSFs. From the model, obtain classification performance of each PSF". Therefore, the topics at the forefront of this literature search will be; the latest HRA techniques used in maritime engineering and techniques used in other fNIRS and simulator studies to model and classify PSFs. These techniques will then be discussed to see how they could be improved and used in our study. This will answer the research questions three, five and six; "What are the current HRA techniques in maritime engineering and technology and how can they be improved? What methods have been used to model human error using fNIRS and simulation software in previous studies? What techniques can be used and have been used in previous studies to statistically analyse the effect of each PSF on human performance?".

The literature search keywords used are: Human reliability analysis (HRA), Maritime.: fNIRS, Simulators, Modelling techniques. Peer reviewed journals, conference articles and technical papers are used to conduct this literature review.

2.4.2 HRA techniques used in the Maritime sector.

Generally speaking, HRA techniques are used in the maritime sector aim to analyse PSFs with the goal of achieving a nominal human error probability (HEP) [87]. Over the years HRA in the maritime sector has gone through a series of generational changes [88]. In the 1990's first generation models such as THERP, SPAR-H and ASEP were criticized due to the claim that these models lacked a cognitive aspect capable of dealing with the human factors associated with human reliability [89]. As a result of those criticisms a number of new methods emerged; ATHENA, INTEROPS, OMAR and CREAM [88]. These later models represent a more complex BCI methodology taking into account human cognition [90]. These later models have again been scrutinized due to their lack of measurable PSFs, hard to accurately quantify, lack of validated empirical data and relies on speculative data [84]. Therefore, current HRA techniques were either revised, hybridized or developed to incorporate simulation of 'real life' scenarios [91]. For these simulator-based techniques the focus shifted to empirical data validation in order to try to reduce the uncertainties [8].

The latest HRA techniques to be used in the maritime sector coupled with simulators are human error assessment and reduction technique (HEART) [92], A hybridized version of cognitive reliability and error analysis method (CREAM) [46], A revised version of a technique for human event analysis (ATHEANA) [93], probabilistic cognitive simulator (PROCOS) [48], Information, decision and action (IDAC) [94] and Bayesian networks (BN) [3]. These techniques would be difficult to apply to fNIRS data. To better describe the reasons for this a brief description of each technique is needed.

Evan et al's [92] paper states that the heart technique allows tasks to be compared relative to each other and assesses each individual sub-task in turn. Firstly, the generic nature of the task is specified. Then error producing conditions (EPC) are identified. For example, unfamiliarity or low signal to noise ratio. Next, the effect of each EPC is assigned a nominal value between 0 and 1. Finally the human error probability is calculated given the EPC nominal value. CREAM and ATHEANA are similar in approach according to Yang et al [46] and Dsouza et al [93] papers, as they work again, based on defining the scenario, defining the scope of the study, defining the problem tasks, defining the PSFs and then calculating HEP.

The papers by Truccio et al [48] and Chang et al [94] describe how PROCOS has been developed, and more recently how IDAC has hybridized to incorporate the use of simulators to better identify how human factors effect operators' decisions. In the aforementioned literature the techniques were used to look at a scenario in normal working conditions before applying PSF. More specifically, PROCOS used two different flow charts. One reproduces the operator's behavior in normal working conditions and the other looks at their behaviour in an emergency or recovery conditions. Both flow charts are based on the model SHELL-PIPE (SHELL – interaction between equipment, software and operators. PIPE – an evaluation of the operator). There are two possible outcomes 'yes or no' based on a Bernoulli's distribution of 0-1 values of PSF and their influence on each decision block. HEP is obtained based on a success likelihood of each task, defined by the SLI formula (the total number of good events divided by the total number of valid events).

A Bayesian network is a useful technique to analyse an event that has occurred and predicting the likelihood that any one of several possible known causes was the contributing factor [95]. For example, a Bayesian network could represent the probabilistic relationships between a failure event and actions leading up to the failure [96].

2.4.3 Modelling techniques used with fNIRS and simulators.

The literature search in this section will focus on modelling techniques used primarily in human factors studies. Our study involves using fNIRS data extracted from participants engaging with a BCI. Therefore, for this section of the literature search the specific focus will be on models used for BCI studies using fNIRS. The classifier accuracy deemed acceptable for BCI is 70% [85]. Consequently, a brief overview of a number of studies will be conducted, moreover, looking specifically at the performance classification accuracy.

Linear discriminant analysis (LDA) has been used as a performance classifier for fNIRS data on the following BCI studies; [97], [98], [99], [100], [101], [102]. This performance classification modelling technique obtained accuracy levels between 75.6%-90% for fNIRS-based studies using BCI's. For the studies above this classification technique worked well for a variety of feature types. For example, oxygenation and connectivity features. LDA is widely used as a performance classifier of fNIRS-BCI data. From the literature referenced above the classification performance accuracy was consistently above the acceptable accuracy of 70%. LDA also has the flexibility of allowing for a shrinkage parameter chosen based on cross-validation [103]. Shrinkage can be used in situations where the number of input variables greatly exceeds the number of samples, which would be the case for our study, results in the covariance matrix being poorly estimated [104]. Shrinkage would regulate the estimated parameters by putting a restriction in place to ensure that individual covariance matrices shrink toward a common data pooled covariance matrix [104]. LDA as a classifier does have its disadvantages. If the data regularly falls far outside of the means of the classes, then LDA will not work well [103]. LDA is also sensitive to over fit [104],and validation can be problematic which is why the k-fold cross validation method was developed for a small number of samples/features [105].

Support vector machine (SVM) has been used for fNIRS-BCI analysis as a performance classifier in the following studies; [106], [107], [108], [86], [109], [110]. The SVM classification technique obtained accuracy levels between 61.8%-84.9% for BCI studies using fNIRS data. SVM is a good technique to use where the data set contains more features than the number of rows of data [111]. Our study will involve 15 channels and oxygenation features only, which means the number of features will be under 10 [112]. SVM can be used for both regression and classification [113]. This will help with our study as it is both linear regression and a classification problem. The downsides of SVM include: SVM aims to find the largest margin boundary in order to classify the data [114] however, fNIRS data in aerospace sector investigations (which are similar in approach to ours) show the classes to be complex/non-linear and to have minimal boundaries which could result in a low prediction accuracy/performance classification [112]. SVM can convert non-linear, complex data into linear using higher dimensions [111] however, the implementation of this feature is done using kernel functions [113]. In short, the kernel function takes low dimensional input space and transforms it into separable space classes [112]. This is where SVM becomes difficult to use in our study as there are many kernel functions (Linear, Nonlinear, Polynomial, Radial Basis function (RBF) and Sigmoid) and to choose the optimal kernel function takes a lot of time, trial and error and can be challenging [111]. Also, SVM on large data sets takes a lot of training time [114] and the fNIRS data sets in our study will be extremely large. Moreover, SVM is not a probabilistic model so it cannot be used to explain performance classification in terms of probability [113].

K Nearest Neighbour (KNN) has been used as a performance classifier for various fNIRS-BCI studies. These studies include; [102], [115], [116], [117], [118], [119]. The accuracy obtained by KNN for the above studies ranged between 50.7%-90.54. KNN is a model widely used due to its simplicity and easiness to interpret. It does not make any assumptions so can be used in non-linear data sets with conversion [120]. In the above referenced studies KNN has been shown to work well with multiple classes and this model can work on both classification and regression data sets. However, KNN has many disadvantages when used in a study similar to ours as it becomes very time consuming due to the efficiency of the algorithm, as the number of data points increases due to the model needing to store all data points [121], as the data set grows the accuracy of the algorithm declines [122] and as stated earlier our study will contain vast data sets with many participants. The main disadvantage of using KNN in our study is its sensitivity to outliers [122]. It can be assumed from previous fNIRS studies [120] [122] that our data set will contain multiple outliers resulting in a reduced performance classification accuracy. Also, KNN in the studies referenced above proved to have a great variation in performance classification accuracy with some studies falling below 70% (the percentage defined as acceptable with respect to accuracy [120].

Artificial neural network (ANN) is used on a variety of studies as a classifier of performance. ANN has been reviewed in the following fNIRS-BCI studies; [61], [85], [106], [86], [123], [124]. The ANN classification technique obtained an accuracy percentage in the following range 63%-89.35%. Neural networks are used in modern day to advance robotics into simulating the human brain. It does this by recognising correct decisions made in certain scenarios [125]. However, ANN can be used as a classifier for non-linear data sets [126]. There are many advantages of ANN which include: neural networks are flexible in that they can be used for both regression and classification problems [127], works well with numeric data only [128], once trained the prediction are fast [128], ANN can be trained with a varying number of layers and inputs and ANN works best with more data points [126]. Given the above advantages, ANN would be a good fit for our study. However, ANN has the following disadvantages: it is impossible to define how each variable is influencing the dependant variables [127], ANN requires high performance computational equipment which can be very expensive otherwise it would be extremely time consuming with CPUs available for our project and ANN is very dependent on the training data sets which can result in over-fitting and generalization [125]. The application of Neural networks is very much in its infancy with respect to fNIRS-BCI performance classification models as there is limited literature available compared to the other models discussed. This could be a classifier to explore in future studies but currently ANN has a major problem when training. The ANN training algorithms often get stuck in local minima [125] (The point at which the function takes the minimum value is not the actual minimum value (global minima) whereas the global minima is the optimal solution).

After a brief review of the above models, we learn that LDA is best fit for our study due to the model consistently showing accuracy scores above what is deemed acceptable for fNIRS-BCI studies and techniques have been developed to overcome the disadvantages. For example, the

K-fold validation method [102] and using the shrinkage method [104]. Therefore, for the remainder of this section, a more in-depth review of LDA and how it has been used in BCI-fNIRS studies will be conducted.

The paper by Verdiere et al [45] used the technique linear discriminant analysis (LDA) to classify performance based on extracted oxygenation features. In the study, the data is sorted in 12 x 10 second epochs over the 120s task period. LDA is then used to classify the operator's performance for each individual epoch as a resulting percentage against the oxygenation features. This works well as the resultant table from the study showed a comparison of participants (table y axis) against oxygenation features (x axis). This allows the reader to compare each participant and oxygenation feature. However, it would have been good to see how each participant performed with respect to time instead of against one another. For example, time in the x-axis and participants in the y-axis with a resultant classification percentage being of all of the oxygenation features combined. This would show the reader the specific areas of the task with the highest levels of activation. On the other hand, the authors may argue that the areas of the task with the highest levels of activation are predictable and that the task was used to test LDA's evaluation of stressors as opposed to the task itself.

The result obtained from oxygenation features used in LDA was 74.7%. This is a good classification performance score, it is well above chance (56%) and is above the accuracy level defined as acceptable for BCI studies (70%) [61]. This technique is therefore an option for our study, however, it would have to be changed slightly due to the assumption that for our study the times to complete the task will all be different. This will also result in participants being up to a differing part of the task or epoch when compared to one another. Applying LDA to our study could still be done as the paper by Verdiere et al shows us that LDA can classify performance of operators participating in a task but could also be used in the same way to classify PSFs based on operator's performance in a standard task against a task where a PSF is applied. The resultant classification performance would give us the effect of the PSF.

The study by Dehais et al [129] used LDA for inter subject classification in an accident study using EEG on a real flight. The participants were required to respond to auditory alarms to explore the phenomenon known as inattentional deafness. The features used were frequency features (due to the use of EEG as opposed to fNIRS) extracted at 3 second epochs. The authors of the study used a 5-fold cross validation procedure (10 times). LDA was used in this study to differentiate participants who acknowledged alarms and participants who missed the alarms.

36% of alarms were missed therefore an equal number of acknowledged alarms were used (700 missed, 700 hit) to avoid dependency of the classification technique to a specific training/testing test. The classification performance results obtained by LDA was 60.6%. These results showed to be above chance and the authors stated that the "results open promising perspectives" but the overall classification scores are low (70% accuracy is defined as sufficient accuracy for BCI) [93]. This can be explained partly due to the use of EEG. EEG is sensitive to noise and vibrations found in real flight. Therefore, it can be assumed that the data is less accurate due to real-flight conditions.

The LDA model worked well to classify inter subject performance. The two aspects adversely affecting this study have proven to be EEG [18] and the Inter subject approach [88]. Therefore, it would be wise to change these aspects and use fNIRS and instead of focusing on inter-subject, to change the focus to classifying PSFs. This approach should reduce inaccuracies found by the variable differences in human physiology from person to person by instead averaging oxygenation data across all participants in a study of PSFs against a standard test.

2.5 Concluding remarks

The following remarks are the most significant remarks comprised in this chapter. They are emphasised as bullet points for the reader's ease:

- The literature showed that the best and most strategic analysis of a task is to use a workflow style task design. The workflow allows for the ease to follow, compartmentalization of tasks and sub tasks. A workflow style task design has been used successfully in many previous fNIRS-BCI studies. This helps towards objective two.

- The dorsal lateral pre-frontal cortex has been shown to be the optimum area of the cerebrum to collect fNIRS data due to that region being the region involved in working memory, decision making and executive functions relevant to the tasks in this study. This helps towards objective three.

- The best software to use to filter the raw fNIRS data is NIRStar version 16.1 as this has been used in previous studies and is the software that comes ready calibrated to the NIRx system used in this study. Furthermore, to process the data into epochs for ease of analysis R-studio software is used. R studio has also proven on previous studies to analyse and evaluate fNIRS data using mechatronic code. Therefore, R-studio will also be used to model the data collected. This helps towards objective three.

- Current HRA techniques used in maritime engineering are unable to accurately quantify data from fNIRS-BCI studies. This is due to their unambiguous nature and approach to experimental data. Whereas, fNIRS data collected from human participants has many uncertainties. Therefore, human factors classification techniques must be used to quantify the fNIRS-BCI data.

- LDA is the classifier that will be used in this study. LDA has been used in previous studies to great success and has shown to consistently provide performance accuracy percentages above 70% which is deemed the minimum accepted percentage for BCI studies. The disadvantages of LDA are primarily overfitting therefore, the shrinkage technique will be used to overcome this problem resulting in a higher performance classification accuracy. The LDA with shrinkage technique will be performed using R-studio software described above. This helps towards objective four and five.

Chapter 3. The Significant Factors that Contribute to the Increase in Human Error within a Ship Engine Room.

3.1 Introductory remarks

This chapter investigates the performance shaping factors reported as a cause or contributing factor towards human error within a ship engine room. The first section of this chapter will be outlining the criteria for the accident database search. The following section will define the databases used to gather primary data relating to human error within a ship engine room. The third section will contain the primary data gathered in tabular format for the reader's ease. The data tables will include; a list of accident reports used, the source of the report, the accident report code and the PSFs associated with the incident. The final section will then discuss the PSFs with a final summary. The aspects listed will allow us to achieve objective one – *Analyse the ship accident reports from maritime accident data bases to obtain primary data representing the tasks and PSFs most associated with human error in a ship engine room.* This will also allow us to answer the research question – *What are the PSFs associated with the highest levels of human error within the engine room of a ship?*

3.2 Criteria for the database search

Before beginning a search through the various ship accident databases, a scope of the search must be defined. Objective one states - Analyse the ship accident reports from maritime accident databases to obtain primary data representing the tasks and PSFs most associated with human error in a ship engine room. This objective is also linked to the research question - What are the PSFs most associated with the highest levels of human error within the engine room of a ship? Therefore, the most relevant feature of the accident report search is the PSFs reported. The task being conducted at the time the incident occurred is not relevant due to this thesis being an investigation into PSFs and not the tasks. However, if a specific task is shown to have a significant effect with respect to human error, then that task will be simulated and PSFs incorporated. Another important factor is the date the incident occurred. This is important as the older accidents may include vessels with inferior technology compared to recent years. In order to locate the relevant reports, the filters: engine room, human error, date of incident and human reliability will be key features to carry out the search.

The keywords used in the accident report search are as follows: **engine room, human error, human reliability, 2012-2018 (accident date time frame).**

The illustration below for the ease of the reader, shows the process flow as to how this chapter was conducted. The boxes to the left indicate the sub-section goals, while the boxes to the right depict specifics to achieve the section goals.

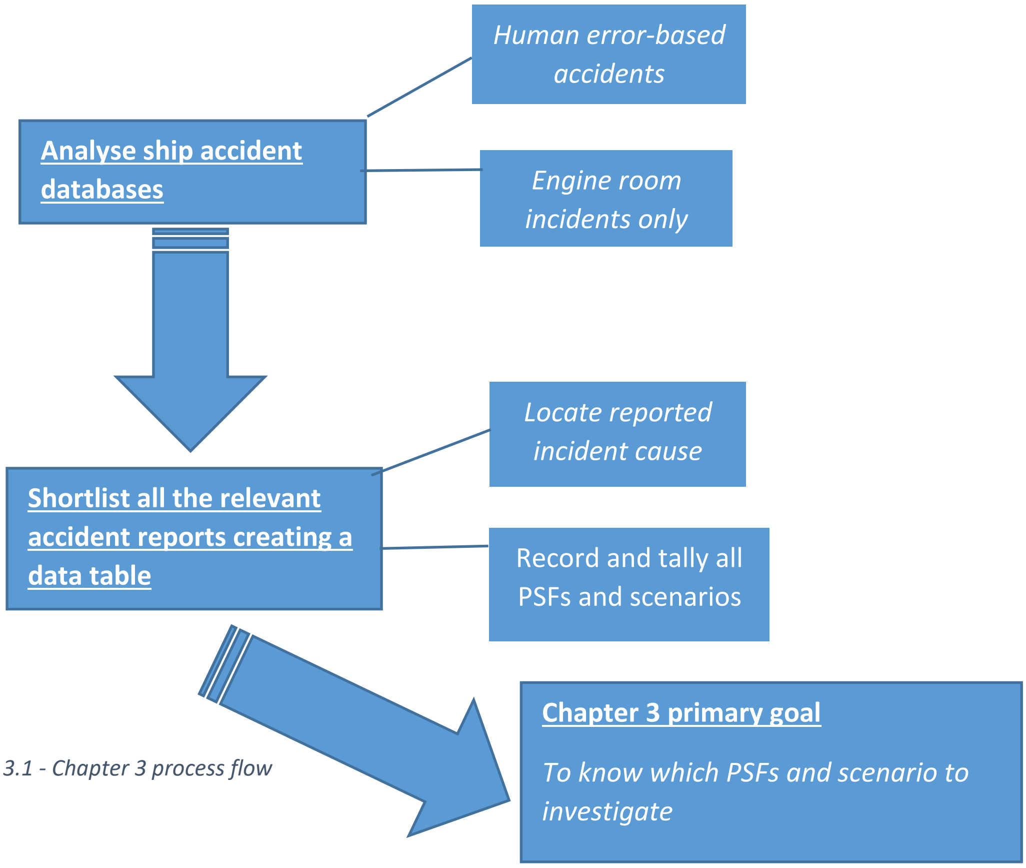


Figure 3.1 - Chapter 3 process flow

3.3 Analysis of ship accident databases

Maritime accident databases and accidents reports describe the operator duties and factors associated with human error that cause significant incidents. The ship accident databases contain all the latest accidents from all around the world. The sources used include Transportation Safety Board (TSB), The Marine Accident Investigation Branch UK (MAIB),The European Maritime Safety Agency (EMSA) and The Nautical Institute MARS database (Mariners' Alerting and Reporting Scheme) involving human error within the engine room. This search will be conducted between the years 2012-2018 as it was deemed more relevant to use the latest incidents and 6 years gave a large enough investigative dataset of 2000+ incidents.

The above databases are searched manually using the filters and keywords discussed in section 3.2. The relevant reports are extracted and analysed, recording incident number or report code and PSFs reported as a cause of the incident. Many papers focus on navigation issues relating to ship bridge. This focus in some human error incidents proved to be the incorrect approach as it was later revealed that the cause was in fact engine room related. On the one hand it shows the novelty of this study and on the other hand, it makes limited contribution to the identification of significant factors from marine engineering perspectives.

The MAIB accident database is accessed through the UK gov website. Using the first search filter of accidents between years 2012-2018 provided 988 reports. A further filter was applied using the words 'engine room'. This narrowed the search to 174 incidents. Further words were added to the search filter (engine room human error). However, the search filter did not perform as required when the additional words are added. Therefore, 174 reports were analysed manually to find engine room incidents relating to human error. 37 incidents were found to be relevant.

The TSB accident database is through the Transportation Safety Board website. The first search filter (2012-2018) was applied and 523 reports were provided. A second search filter of 'engine room' was applied. This provided 36 reports. A final search filter of human error was applied providing 7 reports. The 7 reports were analysed and 6 of the 7 were relevant to our study.

The MARS accident database is accessed through The Nautical Institute website. This database proved to be the easiest to use when searching for relevant incidents as a summary with incident cause and date is displayed without having to download or access the accident report. This allows for quicker manual filtering for relevant reports. Therefore, a date search filter was not used but any accidents prior to 2000 are discarded. There are 853 reports available using the search words ‘engineer error’. Of those reports, 244 occurred between 2012-2018. 244 accidents were analysed using the incident cause description and 46 were found to be relevant to our study.

The EMSA Europe accident database can be accessed through the EMSA website portal. The first and second search filters of accidents between 2012-2018, and only completed investigations were applied. This provided 1,110 reports. A third filter ‘engine room’ was applied narrowing the search to 71 reports. Of the 71 reports 19 were relevant to our study.

The accident databases combined provided 109 accident reports specific to human error within the engine room. Each of the reports was individually analysed with the goal of locating the PSFs reported as a cause of the accident. For example, if an engine room fire or explosion was reported then the report analysis section usually provided information as to the number of engineering staff on duty and the task being conducted at the time. If the accident report stated that the engineers on duty failed to see the fire early enough due to reporting or conducting separate tasks then this would be defined as ‘distraction’. If the report stated that the engineer failed to correctly apply the correct procedure due to having to complete a number of other tasks simultaneously then this would be defined as ‘workload’. Specifically, the search was done looking for engine room incidents which had a human error reported as a catalyst or directly contributing to the incident. Some interpretation and extrapolation had to be applied when categorising the error sections. Fortunately, the same error types were reoccurring and could easily be grouped into fatigue/tiredness, distraction (needing to stop a task to do something else, then go back to the task) and workload (an increase in routine task at once). The other PSFs (fatigue, temperature, time sensitive, visibility, weather, noise, vibration, loss of concentration due to automation) reported are in most cases not left open for interpretation but instead reported unambiguously.

The tasks that have the highest significance in order are: ballasting, bunkering, fuel line management, seawater treatment, system maintenance and repairing faulty electrical systems. However, there was very little difference in frequency of occurrence between the tasks listed. Therefore, the tasks being conducted when the error occurred seemed random and of no relevance to the error that occurred.

Workload was found to be the highest risk PSF due to a frequency of 31% across the four databases. The second is fatigue with 19.4% closely followed by distraction at 14.3%. The remainder of the PSFs reported are low in comparison to the top three and appeared quite randomly. This is depicted in Appendix E.

3.4 The data table

The table (Appendix E) shows references (accident databases and incident number) against the PSF reported as a potential cause of an incident for example, fatigue, distraction, temperature etc. The references are taken from the accident report documents with respect to incident number. The numbers correlate to the number of times a specific PSF has been reported in an accident investigation.

The table shows the PSFs that occur the most frequently, in order are: workload, fatigue, distraction, time dependency, incorrect procedure, noise, temperature, communication, weather, vibration, loss of concentration due to the monotony of automation, visibility and pre-existing injury as shown in the table in Appendix E.

3.5 Performance shaping factors

The consequence of human error within a ship engine room was found to have an average cost of 6.9 million euros annually [5]. 93% of incidents that occurred due to human error within the engine room were reported to be due to an adverse PSFs and not the task. PSF is therefore the most significant factor to investigate.

The main PSFs that contribute to human error are said to be PSFs that increase mental workload [33]. An increase in mental workload increases the probability of human error [44]. The marine accident databases will validate this hypothesis by showing the specific PSFs that appear the most frequently on the accident reports.

The three PSFs that appear the most frequently on the marine accident databases in no particular order are: distraction, workload and fatigue. Khan et al [130] collected speculative data in relation to PSFs that affected 235 experienced seafarers, all of whom work on engine and maintenance departments. The data was collected via a questionnaire. The significant PSFs reported by the 235 seafarers are: fatigue, workload, ship motion (roll and pitch), timescale and stress thus validating the findings from the accident databases. There was no mention of distraction. However, distraction was not an option on the questionnaire and could have been interpreted by the seafarers as other PSFs. For example, pitch and roll could be distracting. There is also the mutual exclusivity issue as some reports had fatigue (no breaks given throughout a shift) in conjunction with an increase in workload (a standard, routine task with additional duties, for example doing the work of another operator as well as your own on a specific exercise (looking after 6 ballast tanks in ballasting/deballasting operations as opposed to 1 or 2). Each PSF would be investigated individually without overlap and modelled at a later stage to evaluate the combined effect.

Fatigue is an important factor relating to human error [59]. Fatigue is defined by reports of tiredness, working over shift, working too many shifts, not taking a break resulting in a human error. Mental and physical fatigue is a factor that appears frequently on the accident databases and is mentioned in human factors journals as a technique used for inducing a high mental workload [40] [57]. As stated above fatigue is mainly brought about due to: monotonous monitoring tasks [29], sleep deprivation [57], long, high workload shifts [66], and monotonous duties [131]. Manipulating increased fatigue would involve a long monitoring task [40]. For example, a long baseline where the candidate will be monitoring a screen. Another way to manipulate fatigue would be to test participants at inconvenient hours making them sleep deprived [57]. Experimenting at inconvenient hours involves difficulties logistically and has ethical issues, therefore a long monitoring exercise before the task begins will be an easier approach to increase the mental fatigue of the candidates.

Distraction is another important factor relating to human error [132]. Distraction is a cause or part cause of many incidents [14]. Literature shows that the main causes of seafarers becoming distracted within the engine room are due to: Questions from colleagues [133], Human Interactions [83], Behavioural/Individual factors (discipline, mind-set, vigilance and sensitivity) [134], On-board Environment (ship motion, weather, noise and vibration) [49]. Tech sim 5000's instructor console provides control over distractions experienced by the participants. The instructor can send various 'on screen' messages to the participant whilst they are performing a task. These messages would include; prompts to read out information from the liquid cargo screen or working out volumetric flow rate (commonly done on a ballasting task). Sent routinely throughout the task, these messages would cause a distraction similar to what a seafarer would experience in a real-life scenario [60].

Workload proved to be a significant factor in previous studies investigating neurophysiological activation [18] [19] [130]. Studies have shown that the more tasks involved, and the higher the difficulty level, the higher the mental workload [36] [54]. Studies show that there is always a point where a human reaches their limits with respect to workload [44]. Increased workload appeared frequently on the accident databases and therefore commonly features in studies as a technique to manipulate increased neurophysiological activation [51]. The literature and databases substantiate increased workload as being an important PSF to investigate when looking at human error causation [54]. The difficulty with the investigation of workload, is the realism of increasing an operator's workload. Methods for increasing operator workload in an engine room include sounding multiple alarms, additional ballast tanks to fill or a separate console where the candidate must routinely press certain keys at a set time (NASA-TLX) [60]. The increased workload scenario needs to be distinctly different from the distraction scenario in order to avoid replication.

Noise, vibration and temperature are environmental factors that also appeared on the accident databases as adverse PSFs. These factors are the main contributor towards discomfort in the engine room [20]. Noise, vibration and temperature have been reported as a main influence of a rushed task within the engine room [21] thus contributing towards an incomplete or badly performed duty [22]. Investigating noise and vibration is going to be difficult in an engine room simulator. Additional equipment running alongside the simulator, or a hydraulic simulator would be needed in order to investigate noise and vibration [51]. A hydraulic simulator is not available and it would be difficult to induce noise and vibration into a task any other way. Therefore, it is not viable for us to investigate noise and vibration as a PSF.

Time sensitive tasks have appeared on the accident databases as a contributor towards an accident in a ship engine room. Performing a task under time constraints, or in an emergency, has proved in previous studies to cause an increased mental workload [66]. This would validate an investigation into time sensitive tasks in an emergency, system fault or failure situation where an alarm would be raised. Emergency tasks that require time dependant action contribute to human error due to a rushed task. For example, a ship listing due to incorrect weight management via ballasting/de-ballasting. Therefore, it would be relevant to the study to include a time dependency for investigation into human error causation [83]. Creating a system fault that has to be dealt with in a time dependent manner would be a way to investigate human error due to time dependency [83]. By indicating to the participants that the risk of error probability goes up with time could influence a participant to increase speed. A criticism of this approach would be the question whether an increased probability of human error will be enough of an incentive to make the candidate rush the task. Another way would be a visual display of a ship listing with time (this would happen in a real scenario if the incorrect ballast tanks continue to fill or empty). The addition of a visual display would be a very impoverished approach to a dynamic outcome of an incorrect action. In a real scenario an operator would have no visual display of the ship listing but would instead feel the movement of the ship. Therefore, it has been decided to use the time dependent fault correction approach.

Automation is a potential cause of human error [6]. Automated systems reduce human interaction considerably and can take over during a system malfunction [135]. When automated systems malfunction, the repairs occur mainly in port [136]. The aid of an automated system is validated as a safety feature [45] as automated systems aid with a variety of tasks within a ship engine room [36]. The use of computer aids takes away the competency of the learnt skills from the seafarer’s training [45]. A problem tends to arise when an automated system fails, as a seafarer may not be equipped to repair an automated system at sea [6]. This results in the seafarers having to perform the task of the automated system manually. This is a PSF that is hard to replicate, as a study of automation against a manual task will result in a monitoring task vs a working task. In this study participants will be trained on the day of the experiment, resulting in not enough time passing to allow candidates to forget learnt skills. It could also be said that the results of an automated vs manual investigation could easily be predicted. Therefore, automation will not be investigated.

Ship motion due to the weather (pitch and roll) is another factor that appeared on the accident databases as a contributing factor towards human error. This aspect in a ‘real life’ scenario would put a stop to the majority of tasks within the engine room until the weather subsides [60]. Another factor would be the difficulty of replicating pitch and roll within the engine room simulator. Therefore, it is unfeasible to investigate pitch and roll as a PSF in this investigation. Moreover, weather had a low frequency of occurrence from the accident databases.

Engine room temperature is a PSF reported to increase the probability of human error [20] [21] [22] due to the heat causing a rushed task [21]. It is important to add that ships do also travel to cold environments but due to safety, a seafarer can wear additional clothing but not remove safety clothing [4]. Therefore, a study on the effects of a cold temperature will be less relevant. Increasing the temperature of a participant’s workstation could be done by external electric heaters positioned to heat up the workspace of the participant. However, maintaining a set temperature and monitoring the temperature throughout the workstation would be difficult and scrutinised. The potential temperature fluctuations and inability to accurately monitor the temperature would result in an inaccurate investigation. Therefore, temperature as a PSF will not be investigated.

Visibility, incorrect procedure, injury, complacency and communication also appeared on the accident databases. These PSFs appeared less frequently and would be hard to manipulate on a simulator. Visibility appeared as a PSF either due to an electrical lighting fault or when maintaining certain equipment in low lighting areas. Lack of knowledge was also reported. This was mainly due to differences in training, using new or unfamiliar equipment and experiencing uncommon problems. Communication was reported fairly regularly. The main issues were the language barrier, differences in training, forgetting to communicate and misinterpretation.

3.6 Concluding remarks

The following remarks are the most significant remarks contained in this chapter. They are emphasised as bullet points for the reader's ease:

- The data table shows that there are certain tasks that appear more regularly than others on the accident databases (Appendix E). However, the tasks being carried out when an accident occurs are fairly random in nature. Ballasting appeared the most on the accident databases but there are a multitude of other tasks that are very close behind. When compared to PSF, a case can be made that the PSFs are far more significant with respect to human error than the task itself.

- The data table shows that the most relevant PSFs to investigate are workload, distraction and fatigue. These PSFs can all be manipulated into a simulated scenario as described in section 3.5. A time dependent study can also be incorporated into each investigation by initiating a system fault. This will prevent participants from taking too much time to complete the task and would replicate a 'real life' scenario more accurately.
- The above two bullet points complete objective one - *Analyse the ship accident reports from maritime accident databases to obtain primary data representing the tasks and PSFs most associated with human error in a ship engine room* and answers the research question - *What are the PSFs most associated with the highest levels of human error within the engine room of a ship?*

Chapter 4. A proposed methodology for a BCI-fNIRS analysis of human error

4.1 Introductory remarks

The following chapter details a proposed methodology for an investigation of PSFs in a ship engine room. Specifically, this chapter outlines how the BCI incorporating fNIRS will be used to investigate OFS. Following this, an assessment methodology will be defined, outlining the workflow stages of the task. The penultimate section will define how the raw fNIRS data will be filtered, sorted, analysed and evaluated. The final section will depict the research frameworks developed for this study. Moreover, a summary will be included describing each phase of the frameworks.

4.2 BCI-fNIRS assessment methodology

Based on the findings from Chapter 3, the PSFs that will be investigated are: distraction, workload and fatigue, with the addition of time sensitivity due to a system fault being applied. Time sensitivity is applied by timing the exercise and advising the participants that time is a factor whilst performing an emergency repair/solution to a fault. If this scenario where to occur during ballasting operations in a 'real life' situation then this would be deemed as time sensitive.

This investigation was designed with reference to similar work conducted in other transport sectors by Ayaz [73], Verdiere [45], Baker [19], and Durantin [33] with the aim to use a workflow style task design to investigate changes in neurophysiological activation, via tasks containing low and high workload, long monitoring and reporting tasks. This study was conducted using a TRANSAS engine room simulator as described earlier in section 1.2.6. A fault will occur during a ballasting task and the cause of the fault will need to be identified and solved by the participant. Ballasting will be used as the scenario for testing due to it being a task that occurred frequently on the ship accident reports analysed in section 3.3. The task specifics will be discussed and defined in more detail at the bottom of this section.

The dorsal lateral pre-frontal cortex (DLPFC) is the specific area of the cerebrum analysed for this investigation. The reason for this is that this area of the brain governs higher cognitive functions, such as switching attention, working memory, maintaining abstract rules, and inhibiting inappropriate responses [137]. The DLPFC was the area of the brain used in similar, successful investigations by Solovey [28], Ayaz [73], Tsunashima [42], and Dehais [33] in other engineering sectors.

This investigation is designed with the aim of testing the neurophysiological activation of the prefrontal cortex of each participant, at each stage of the task, using BCI-fNIRS. This approach will allow us to identify the specific parts of the workflow and PSF that induces the most neurophysiological activation. Thus, allowing us to analyse the extent of the effect of each PSF in comparison to one another and against a standard test. The reason for this is that it has been documented on similar investigations by Aghajani [18], Bu [35], Chiarelli [40], and Hlotova [78] that the level of neurophysiological activation has a relationship with the probability of human error.

The workflow task design was based on the duties that would be carried out by a 2nd engineer [65] whilst detecting and correcting a steam pump fault on a ballasting scenario. A 2nd engineer's duties were chosen over the duties of a chief or 4^{th} engineer as a chief engineer's role is a managerial role overseeing the duties completed by the 2^{nd} engineer [62]. The 4^{th} engineer (not every ship has one) is there as a trainee or assistant to a 2^{nd} engineer [60]. The participants will be trained prior to starting the task to differing levels to replicate seafarer experience. A brief summary of the task stages is given below but participant training and each stage of the task will be discussed in detail in section 5.1.

Each candidate would be required to complete five stages of the workflow. These stages include:

- Baseline
- Fault occurrence
- Fault detection
- Fault solution
- 2^{nd} baseline

The first baseline will be used as a datum for each participant. Each participant will have a slightly different baseline output of HBO. Therefore, a five-minute baseline will be taken from the participant monitoring system screens. This will allow for an analysis of the increase in activation with respect to the individual participant, as opposed to the group.

The second task stage is the fault occurrence stage. Unknown to the participants, after the five-minute baseline is taken a fault will occur. This will be shown by a visual alarm. The participants will be tested on the time taken and their ability to navigate to the correct system screens to acknowledge the alarm.

The third stage of the task is the fault detection stage. This part of the task requires the participants to navigate to the correct system screens to; make a note of the alarm codes, and, based on the alarm codes, locate the fault and the cause of the fault via various system checks again prompted by the alarm codes.

The fourth stage of the task requires the participants to solve the problem. The participants will be required to navigate through various system screens, re-routing the water line, opening and closing valves, switching on and off ballast pumps and completing various system checks along the way.

The fifth and final stage is the 2nd baseline. A second baseline will be taken from each participant to see how their neurophysiological activation has changed compared to the first baseline.

4.3 Data analysis strategy

To quantify the data provided by fNIRS, a modified version of the Beer-Lambert law is used [57] as the Beer-Lambert law alone can only be used on non-scattering data [28]. Therefore, it cannot be applied to biological tissue without modifying the law to allow for light scattering [43]. This is done by the NIRx software as described above in section 2.3. The raw data is filtered using NIRx and exported into R-studio, as a large numerical table of time in frames against oxygenated and de-oxygenated haemoglobin volumes. The first 15 columns are oxygenated haemoglobin results from all 15 channels, the next 15 columns are the de-oxygenated haemoglobin results from all 15 channels. Then a correction-based signal improvement algorithm is applied to the data as described below.

4.3.1 Correction based signal improvement (CBSI)

CBSI is a technique used to improve the fNIRS signal based on negative correlation between oxygenated and deoxygenated haemoglobin dynamics. Improving signal quality and reducing noise, especially noise induced by head motion, is challenging, particularly for real time applications. In a study done of the properties of head motion induced noise, it was found that motion noise causes the measured oxygenated and deoxygenated haemoglobin signals, which are typically strongly negatively correlated, to become more positively correlated [138]. Therefore, the CBSI method was developed to reduce noise based on the principle that the concentration changes of oxygenated and deoxygenated haemoglobin should be negatively correlated [19].

This is done by using the equation;

$$CBSI_n = \frac{HBO_n}{2} - \frac{\delta HBO_n}{\delta HBB_n} * HBB_n \text{. δ representing standard deviation. [41]} \tag{1}$$

For example the equation for the first row on channel 1 would read;

$$CBSI_{Ch1\ row\ 1} = 0.5 * (Ch1\ row\ 1\ HBO) - \frac{\delta\ HBO\ (Column\ 1)}{\delta\ HBB\ (column\ 1)} * (Ch1\ row\ 1\ HBB). \tag{2}$$

HBO = Oxygenated Haemoglobin, HBB = Deoxygenated Haemoglobin,

The CBSI data from each participant was separated into three sets. Channel 1 to 5 took haemoglobin readings from the left side of the dorsal lateral pre-frontal cortex, channels 6 to 10 from the middle and channels 11 to 15 from the right. This is for ease of analysis and allows the observation of the specific parts of the brain in use whilst participating in the task. This is also useful due to the differing functions of the right and left sides of the dorsal lateral prefrontal cortex [139]. The left side is associated with verbal commands/receiving auditory input, word reading, processing information, linear and logical thinking [137]. The right side is associated with visualisation, spatial reasoning, and interpreting information [43].

Using the separated data, the average haemoglobin volumes for each participant were calculated for each stage of the workflow and put into a data table. A second table displays each candidate's time taken to complete each stage of the workflow. This data was exported to a statistical analysis software package.

4.4.2 Statistical package for the social sciences (SPSS)

SPSS is a software package that can read data files from many different formats (for example, R-studio, Excel and MatLab). SPSS then allows the researcher to perform a running inferential statistical analysis such as analysis of variance (ANOVA) with pairwise comparisons. Alternative software packages are available. For example, sequent and STIM but all are closely related and the analyst had previous experience with SBSS hence, this was the software package used.

4.3.3 ANOVA analysis

ANOVA is used to figure out how much of the total variance comes from the variance between the groups of candidates, and how much from the variance within the groups of candidates. This is done by the ratio:

$$F = \frac{Between\ Groups}{Within\ Groups} \tag{3}$$

If a null hypothesis is true, then the F value will be close to 1.0. A large F ratio indicates that the variation among the group means there is more than you would expect to see by chance [121]. This calculation is done with respect to the degrees of freedom. For example:

$$F(b, w) \tag{4}$$

Where; b is the degree of freedom for variance between groups and w is the degree of freedom for variance within groups. Calculated as follows; b = number of groups – 1, w = total number of observations – the number of groups.

The ratio F is based on a significance value P. The value P represents a percentage of potential error in the resulting F value. For example, in psychology any P values of less than 0.05 are deemed acceptable and less than 0.01 are ideal. Therefore, if the P value is less than 0.05 then the analyst can be confident in the results.

In summary, if most of the variation is between groups, then there is probably a significant effect. If most of the variation is within the groups, then there is probably not a significant effect. [80]

SPSS can go on to compute sums and means over columns or rows of data, create tables and charts containing summary statistics for a large group of participants and conduct pairwise comparisons of each stage of a workflow to see if any relate to one another [140]. Pairwise comparisons will be a useful analysis for this study as we have a 5-stage workflow.

The data will then be exported from SPSS back to R-studio for further processing as described below.

4.3.4 R-Studio

The R-studio software package is used to write and implement the relevant code (see Appendix B). The software will be used to sort the data sets into task epochs. The reason for this is that this will enable us to see the effect of PSF and workflow at very specific areas of the task. Then, when modelling the relationship between OFS and PSF using linear discriminant analysis (LDA) we can predict HEP using operator performance classification based on oxygenated haemoglobin volumes provided by fNIRS [45].

4.3.5 Linear Discriminant Analysis

LDA is used as an operator performance classification model. LDA is a viable option as it is proficient at handling cases where the within class outputs are unequal and where performance data is generated randomly [33]. This method will maximize the ratio of variance between classes to variance within classes in any given data set, hence guaranteeing optimal separability [18]. Similar models (for example principal component analysis) change the shape and the location of the original data set when transformed to a different space whereas LDA does not alter the location but tries to provide greater amounts of class separation and draw an accurate decision region between the various classes [70]. LDA also provides a better understanding of the distribution of feature data.

4.4 Thesis Frameworks

Below are two frameworks developed specifically for this thesis. The first depicts the specific stages of the project analysis. The second depicts the BCI-fNIRS framework. Moreover, it shows how fNIRS was used in conjunction with the maritime simulator. The image used in section 4.4.2 is a generic image taken from one task. This would change slightly depending on the PSF under investigation.

A short summary is included below each framework to re-iterate each stage in more detail.

4.4.1 Analysis Framework

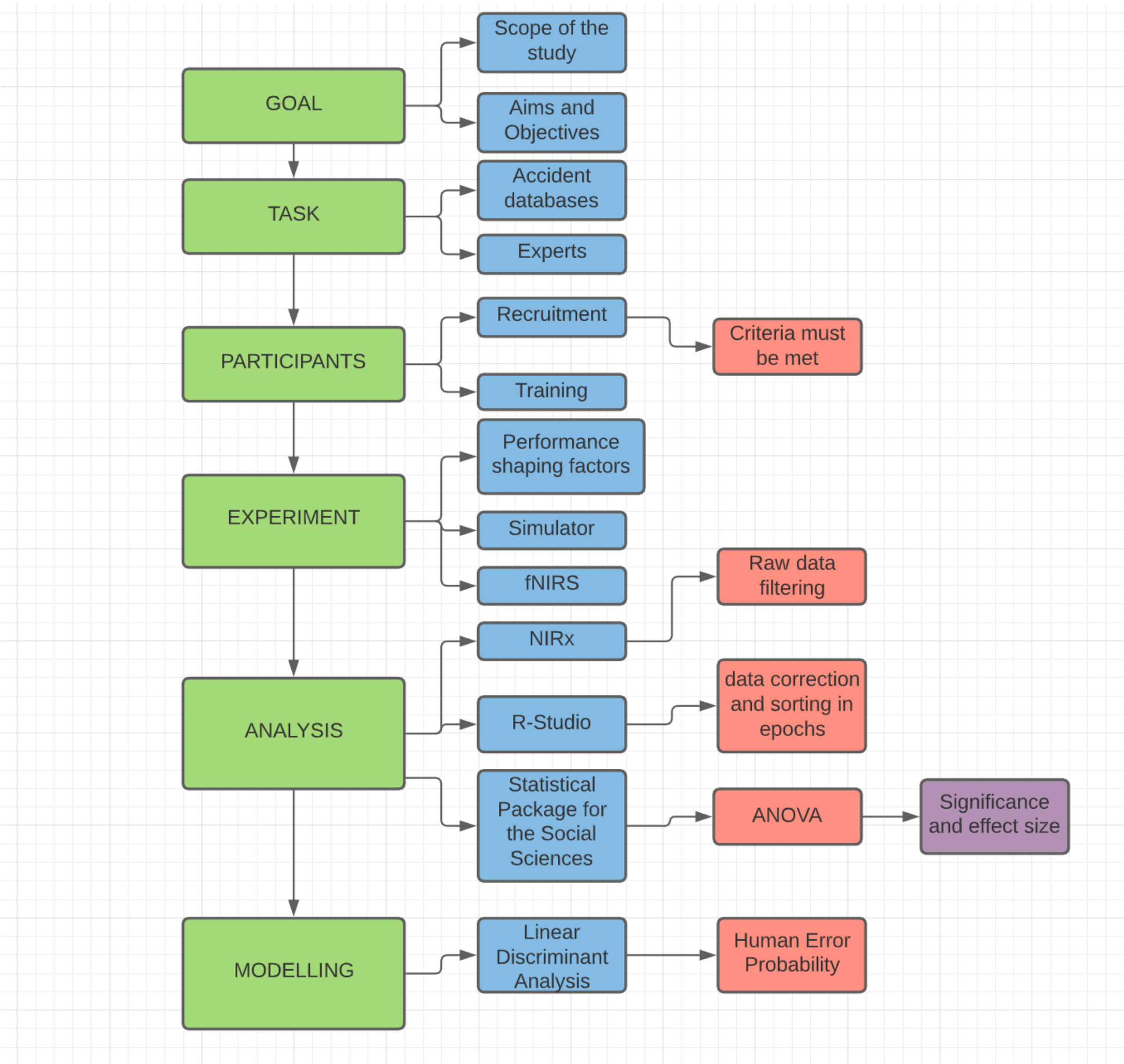


Figure 4.1 - Analysis Framework

The first stage of the framework is the goal. This is the stage were the scope of the study, aims and objectives are defined. This is defined in the introductory stage of the thesis (section 1.2

&1.3). Having this stage in mind throughout the study is an important factor for the framework, as each stage that follows is developed with respect to the research goal.

The second stage of the framework is the task. More specifically, which scenarios and PSFs are the most relevant to investigate for the study. This stage is completed by analysis of the maritime accident databases. More specifically, for this study the analysis was done with respect to human error in the engine room. This is specifically defined in chapter three. It is also important that literature reviews are conducted for this second 'task' stage as knowledge, of how the PSFs and scenarios found are incorporated into a fNIRS based BCI, is needed. This literature review is conducted in sections 2.2 and 2.3.

The third stage of the framework is focusing on the participants. The participants for the study will have to meet certain criteria. For example, specifically for this thesis participants had to have maritime engineering qualifications and knowledge in order to complete the task. Training was given to the participants however, due to time restrictions, prior maritime engineering knowledge was needed. Participant recruitment was done via department specific internal email advertisements and posters. The participant training and required criteria are defined in detail in section 5.1.2 & 5.1.3.

The fourth stage of the framework is the experiment stage. The experiment stage is the stage where data is gathered. The experiments are conducted on the chosen participants in a ship engine room simulator (1.2.6). The experimental stage is where the PSFs and scenarios found are investigated in a practical test using the BCI-fNIRS framework described below in 4.4.2. The specific stages of the experiment, the scenario and experimental design are outlined in chapter 5, section 5.1.4.

The fifth stage of the framework is the data analysis stage. The fNIRS kit used in this study comes with its own 'raw data filtration' software called NIRx. NIRx is used to remove anomalies, artifacts and sort the data into specific, time-related, participant-specific sub-tasks. R-studio is used to write the required mechatronic code to sort the data into epochs and apply CBSI. This is defined in sections 4.4.1 & 4.4.4. Post R-studio analysis, the data needs to be evaluated for any significant effects of the chosen PSFs. This is done by SPSS (4.4.2).

The final stage in the analysis framework is the modelling section. The modelling section of the framework is used to gain a classification performance of the model when applied to PSF data against a standard test. This allows for the understanding of the most significant PSF. The modelling technique used was chosen based on the literature reviewed in section 2.4. The specific chosen model (LDA) is defined in section 4.4.5.

The specific steps for the framework depicted in Figure 4.1 are as follows:

1. The first step is to define and keep in mind the overall goal of the research. This is done by defining the scope of the study (section 1.3), aims and objectives (section 1.4).
2. The second step is to define the relevant scenario to use and PSFs to investigate. This is done by searching through the ship accident databases looking for the PSFs that show up most consistently within human error within the engine room (chapter 3). These findings can then be validated by interviewing experts (experienced seafarers that have previously worked in a ship engine room).
3. Step three is to recruit relevant participants to take part in the experiment. This is done via internal emails and posters.
4. The fourth step is to conduct the experiment. The experiments are done on the TRANSAS simulator whilst participants are connected to the fNIRS system in order to obtain OFS. The experiment is conducted based on, and with reference to, the reviewed literature in section 2.3. However, the experiment design and methods are detailed in section 5.1.
5. The fifth step is to filter, sort and analyse the fNIRS data obtained from the previous experimentation step. This is done using NIRx, R-studio and SPSS software (section 5.2, 5.5-distraction, section 6.2 – increased workload, section 7.2 – fatigue and section 8.1, 8.2, 8.3 – combined PSFs). The software used is summarized in section 4.3 and used in the sections listed above.
6. Step six is to model the PSFs against a standard test and against one another from the fNIRS data, in order to achieve a classification performance percentage. The classification performance percentage will then be used to rank the PSF with the highest likelihood to contribute towards human error, down to the lowest.

4.4.2 BCI-fNIRS Framework

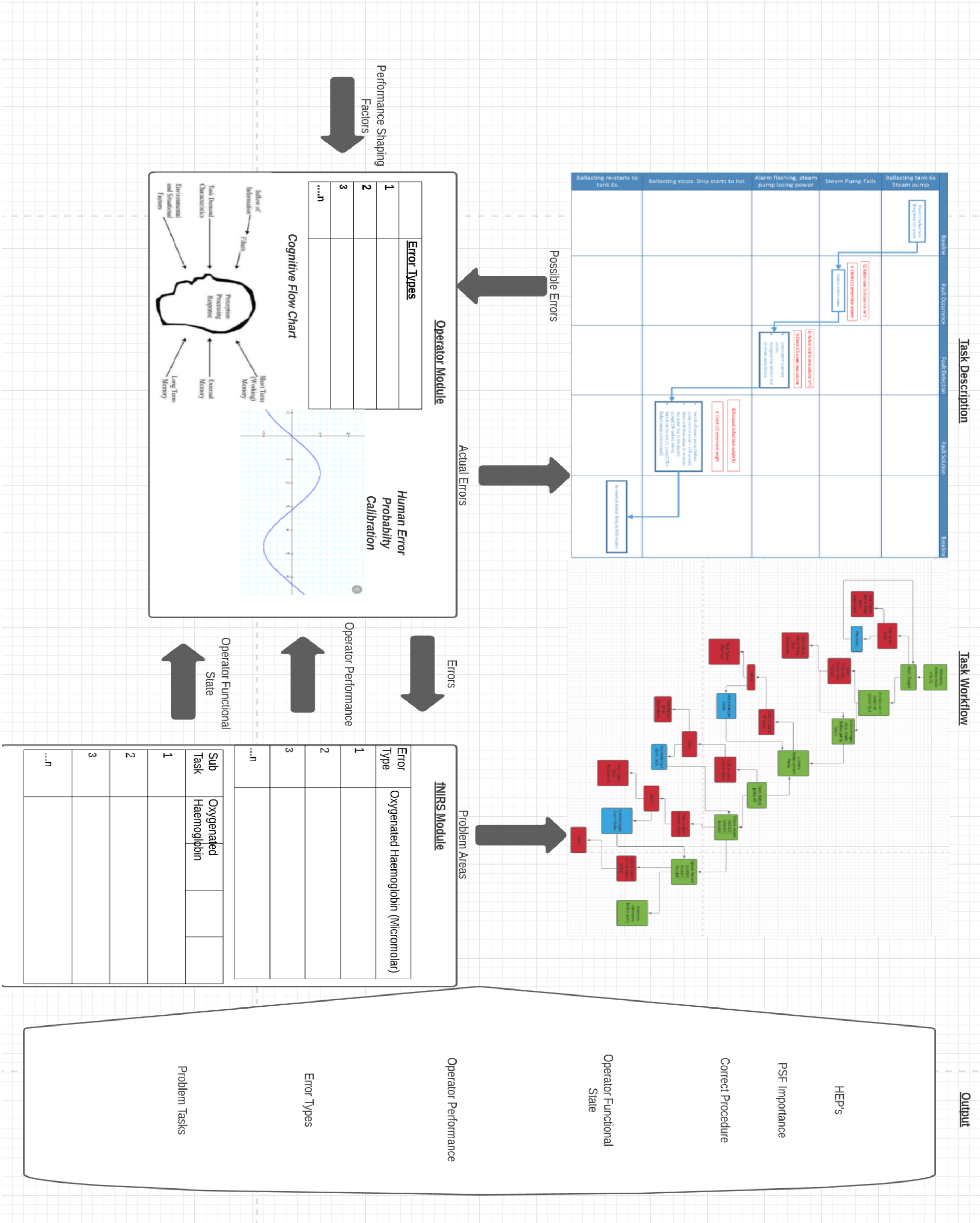


Figure 4.2 - Simulator and Neuroimaging-based HRA framework

The above BCI-fNIRS framework depicts the stages and techniques used to incorporate fNIRS with the TRANSAS engine room simulator to investigate PSFs. This framework was used to test the experiment in order to gather provisional data that could be used to amend the experimental design and then be applied to the 'real' tests.

There are five boxes within this framework; task description, task workflow, operator module, fNIRS module and output:

- The task description box is a box matrix of workflow stage against sub-task. This specific image is taken from a distraction test and would change slightly depending on the PSF being investigated.
- The task workflow box is an event tree or fault tree analysis of the sub-tasks with respect to the workflow (baseline, fault occurrence, fault detection, fault solution and 2nd baseline). This shows the potential route through the task and the consequences of an invalid action.
- The operator module depicts the potential error types with respect to the cognitive flow of each participant. More specifically, the operator module is the part of the BCI-fNIRS framework used to list the error types with respect to the participant's executive functions shown in the cognitive flow chart.
- The error types specific to the participant's executive functions at a specific task time allows for a visual of OFS including artifacts. This is used as a guide towards HEP calibration or classification performance.
- The fNIRS module is used to assess the error types and sub-tasks against the participant's oxygenated haemoglobin. This allows for a better understanding of the sub-tasks and error types to pay closer attention to in the 'real test'.
- The final box is the output. This contains the information that is desired to be extracted from the provisional tests.

These five boxes are all connected. The connections indicate the specific information coming from one part of testing and applied to another. For example, the task description box allows for a visual of error made by participants. These errors are then listed in the operator module and further evaluated within the fNIRS module against oxygenated haemoglobin.

Figure 4.2 is a framework of the BCI-fNIRS aspect from the perspective of the instructor. This allows the instructor to optimise a neuro-ergonomic task design for the 'real' tests. The specific steps for the experiment depicted in Figure 4.2 are as follows:

1. Once the experiment has begun the task description module acts as a prompt for the instructor to apply certain PSFs at set stages of the workflow. For example, the task description module in Figure 4.2 is an example specific to the distraction PSF. The distraction questions will be asked as indicated in this module.
2. The operator module is the second step of the framework. This is where the error types are recorded with respect to the PSF applied to the task. The instructor will apply a marker on each stage of the task where the participant makes an error. This marker is applied via a 'clicker' button attached to the fNIRS hardware. The error types and times are recorded by the same 'clicker' method to ensure that the error type can be associated with the corresponding fNIRS activation signals. This allows for ease of HEP (classification performance percentage).
3. Step three follows on directly from the operator module as the oxygenated haemoglobin values and corresponding error type and sub-task are added to this fNIRS module.
4. These errors and the corresponding mental workload values (from fNIRS haemoglobin values) are then applied to an event tree task workflow module. This enables the instructor to know the potential variables for participants before the sub-task is started. Thus, allowing the instructor to make adjustments to the final scenario to ensure the task is optimised when investigating the specific PSFs.

4.5 Concluding remarks

The most significant PSFs from the ship accident databases proved to be: distraction, fatigue and increased workload. Therefore, these factors will be investigated in this study.

The scenario used to investigate the three PSFs will be ballasting. Ballasting showed up consistently on the accident databases and involves a wide range of different systems and sub-tasks to allow for an easier incorporation of said PSFs.

The data analysis strategy will be to use the following software:

- NIRx – this software comes with the fNIRS kit and is fully calibrated to be used with the specific kit used in this study. NIRx is used to remove various artefacts, filter and 'clean' data to allow for a more accurate and clearer analysis.
- SPSS – this software platform is used to perform an ANOVA study on the data gathered by the fNIRS system. This provides a statistical output of the significant effect from PSFs and workflow stage on each participant.
- R Studio – this programming software is used to write the code to sort the data in epochs and perform the chosen classification model.

LDA was selected as the classification model to be used to evaluate the data. This model has been used in previous studies of a similar nature and consistently provides an accurate classification performance percentage above what is deemed as acceptable in BCI studies.

Chapter 5. The effect of distraction on marine engineers whilst conducting ballast water operations using fNIRs.

5.1 Introductory remarks

The following chapter details the investigation into the PSF distraction. This chapter contains two experiments. The first is a 50/50% split of practically trained and passively trained participants. This is followed by a test where all participants are practically trained.

The first part of the chapter details the experimental design, the participants involved, how the participants are trained and the experiment itself. This is followed by the results from the first test and a discussion. The second part of this chapter contains results and a discussion for the second test.

5.2 Method

5.2.1 Experiment Design

10 of the 20 candidates participated in a study where they were distracted, the other ten had a standard undistracted test. The candidates in the 'distracted' and 'undistracted' groups consisted of an equal share of 10 practically trained and 10 passively trained participants (5 practically trained and distracted, 5 practically trained undistracted, 5 passively trained distracted, and 5 passively trained undistracted). Both distracted and undistracted participants performed the same tasks. The difference was that the distracted candidates had to read out information from the 'Liquid Cargo Screen' (Figure 5.1) on 5 different occasions at set milestones within the task. This replicates what would be done in real operations [60]. The readouts consist of: ballast tank volume percentage, flow rates, max tank volume and tank volumes in m^3.

5.2.2 Experiment Participants

20 candidates were used for this study. All 20 had qualifications to the level of a BEng or higher in marine engineering. 10 of the twenty participants had experience of working at sea in a ship engine room. 4 of the participants were marine engineering PhD students. 3 were ex-navy engineering officers. The average age of the distracted group was 28 and the undistracted 24. 18 were male and 2 were female. The rest were a mixture of post graduate MEng marine engineering students and undergraduate marine students in their masters year.

5.2.3 Participant Training

Candidates were split into 2 even groups of 10. The 10 candidates who had received practical training were put into the 'experienced group'. The remaining 10 candidates who had received passive training were put into the 'inexperienced' group. All candidates had no prior experience of a ship engine room (simulator or sea time).

The following training methods were conducted based on the engine room regulations set by the International Maritime Organisation [141]. The training course was developed with the assistance of experienced and qualified trainers [20] [21] [22].

The passively trained group of candidates were given a 2-hour training session following a customised TRANSAS simulator trainee manual (please see appendix A to see the trainee manual). The training session covered theoretical study of the following areas:

- The liquid cargo handling screen (LCS)
- The alarm system

- The ballast system
- The cargo control room ballast pumps
- The cargo control room ballast system mimic panel

The practically trained group of candidates were given the same 2-hour training session as detailed above, and an hour of practical ballasting tutorials. The tutorials consisted of questions to answer and tasks to complete relating to the areas listed above (please see appendix C for details of the questions and when they were asked). The candidates would be required to navigate the simulator screens answering the tutorial questions and completing the tasks aided by the trainee manual and their training notes. This method was chosen based on the work by Christophe Faisey [142] stating that participants with practical knowledge when compared to participants with solely theoretical knowledge (book learning), usually show a better working memory of the task at hand due to having applied the learned theory practically.

All candidates were authorised to have their own notes taken in the training sessions, coupled with the customised TRANSAS simulator manual whilst participating in the ballasting task. This was decided as engineers would have access to engine room manuals whilst at sea carrying out their duties [143].

5.2.4 The Experiment

5.2.4.1 Baseline

For the first baseline the participants would be expected to monitor the LCS whilst ballasting from pump number two as shown in Figure 5.1 below (ballasting from pump two was set up by the instructor before the task started). The participants would have no active input for the monitoring stage to allow for a five-minute (300s) baseline to be taken.

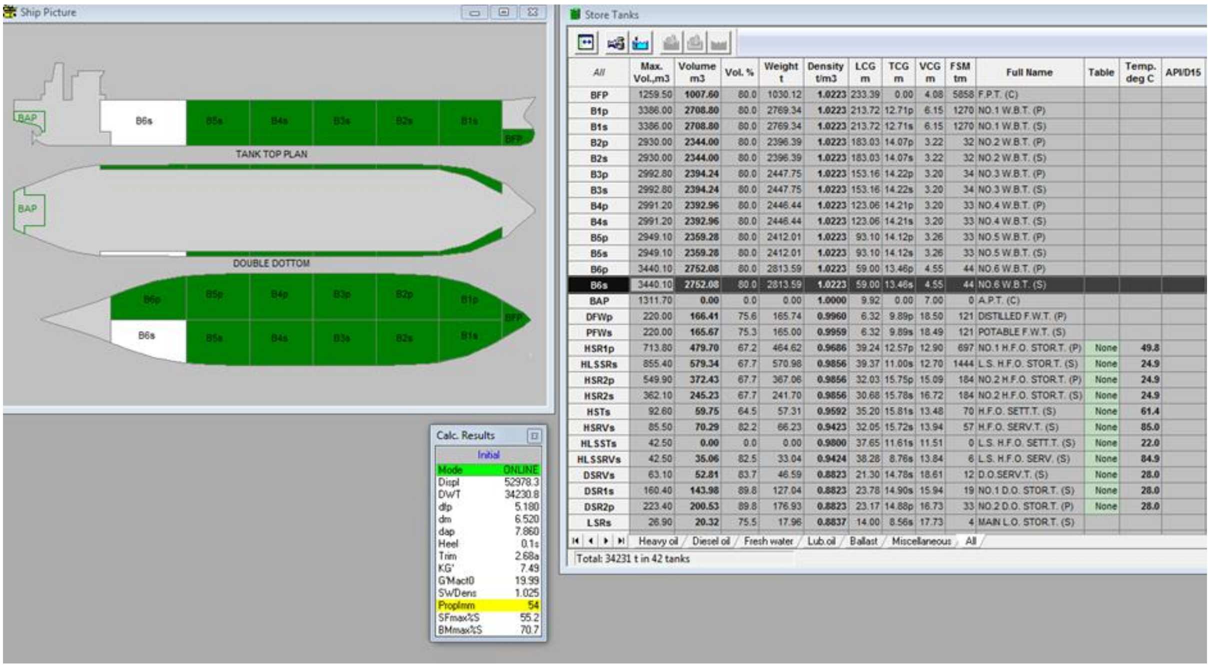


Figure 5.1 - The liquid cargo monitoring screen showing ballast tank readings (e.g. tank volume)

5.2.4.2 Fault Occurrence

This stage of the task took participants between 31 and 46 seconds to complete. For the fault occurrence stage of the workflow, pump number two will fail. The participant must:

a) Orientate to the alarm as shown in Figure 5.2 below. The alarm is solely a visual alarm with no audio.

Figure 5.2 - The ship alarm

(b) navigate to the alarm summary screen (see Figure 5.3 below) to record the details of the alarm.

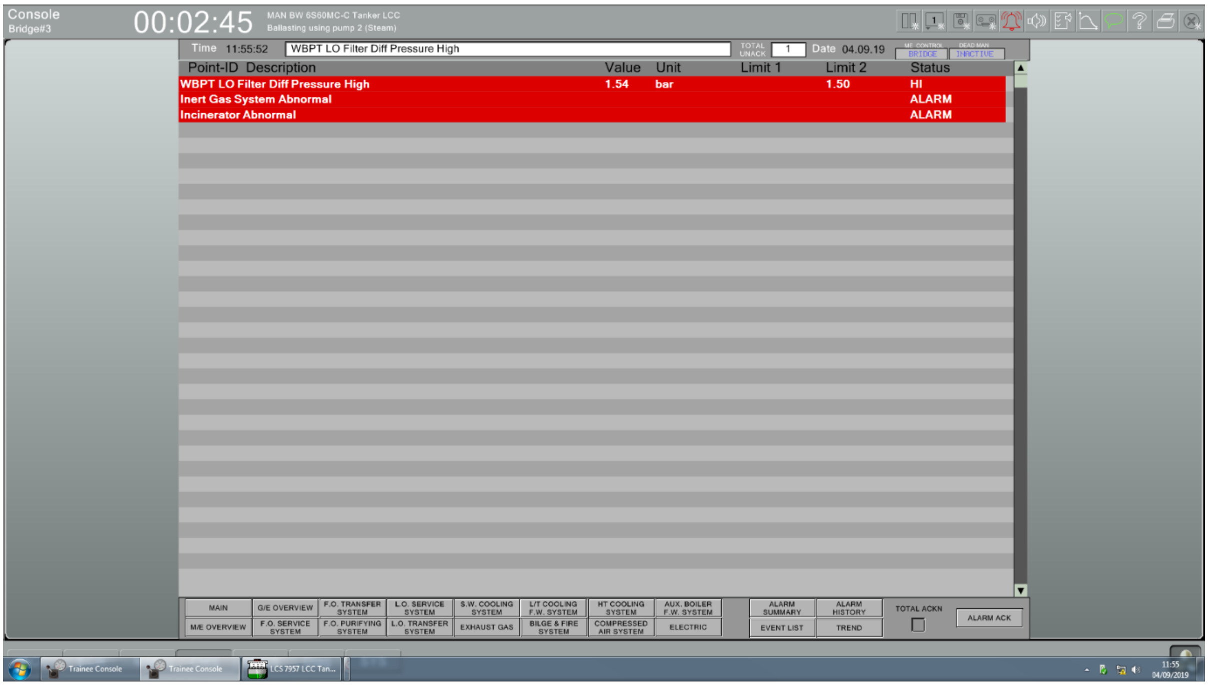


Figure 5.3 - Alarm summary screen

(c) check the ship's log noting any previous faults or maintenance work.

5.2.4.3 Fault Detection

During the fault detection stage, participants must localise the presence of a fault with ballast pump number one. This is achieved by:

(a) navigating back to the LCS screen to check flow rate (Figure 5.1)

(b) navigate to the ballast system screen to check the water line as shown below in Figure 5.4. The participant will be looking to see if there is or isn't an active water flow (the active flow is shown by the illuminated green piping line). If there is an active water flow then this indicates that there is no blockage or leak in the water line indicating that the problem is a fault with the ballast pump.

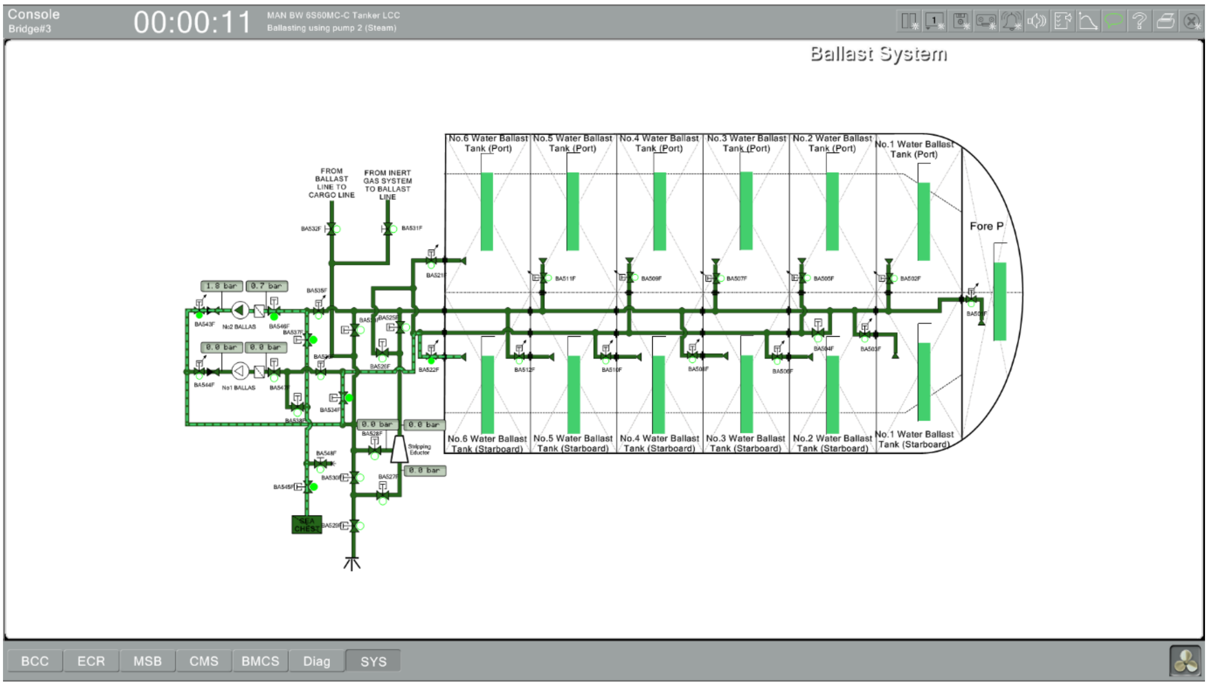


Figure 5.4 - Ballast system screen showing water flow through pump 2

(c) access the cargo control room ballast pumps screen (Figure 5.5) to check the pump pressure gauge (as prompted by the alarm summary screen in Figure 5.3).

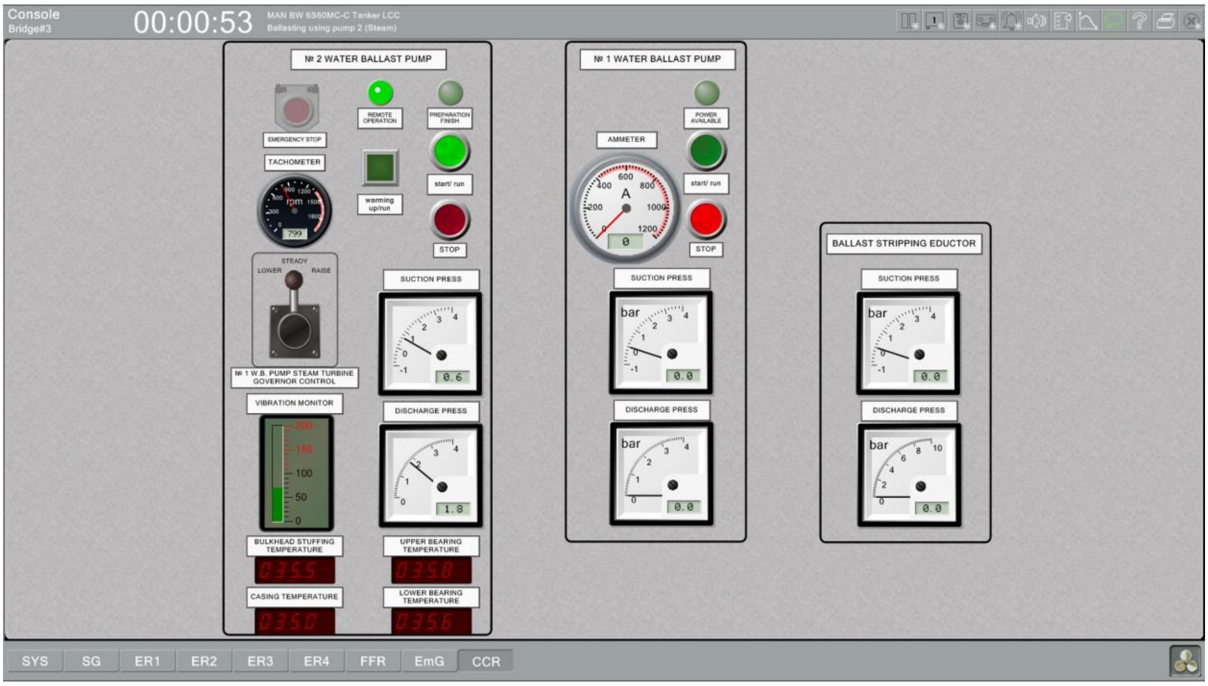


Figure 5.5 - Ballast water pump control panel

5.2.4.4 Fault Solution

The next stage of the workflow requires the participants to determine a solution to correct for the fault. To correct this fault, participants must: (a) navigate to the cargo control room ballast pump screen (Figure 5.4) and switch off pump number two, (b) access the ballast system mimic panel (Figure 5.6 below), (c) open valves BA538F, BA547F and BA544F and close valves BA537F, BA546F and BA543F in order to re-route the water line to ballast pump number one.

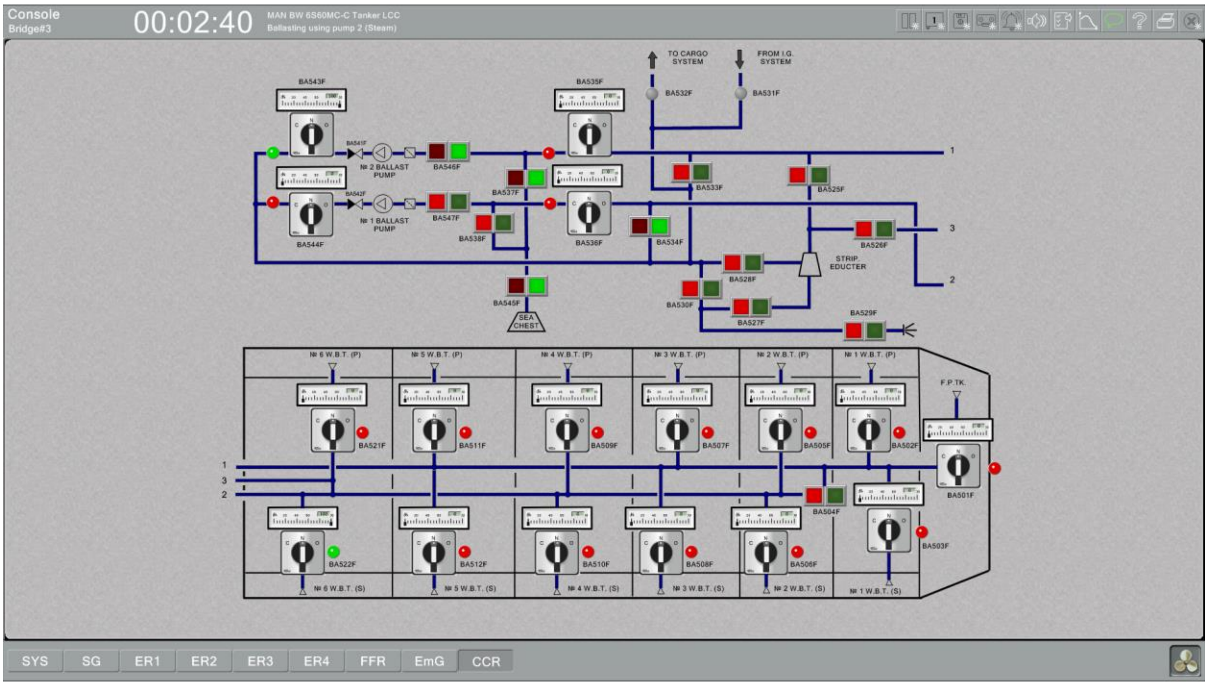


Figure 5.6 - Ballast system mimic panel

(c) access the screen for engine room three (ER3) to power on pump number 2 (the additional task of synchronisation to an additional power generator was performed by the instructor prior to starting the test due to the complexity and the amount of additional time that would be required).

(d) navigate back to the cargo control room ballast pumps screen (Figure 5.5) to check that pump number 2 has power, and switch the pump on.

(e) re-access to the ballast system screen to check that there is a water flow through the new pump as shown in Figure 5.7 below.

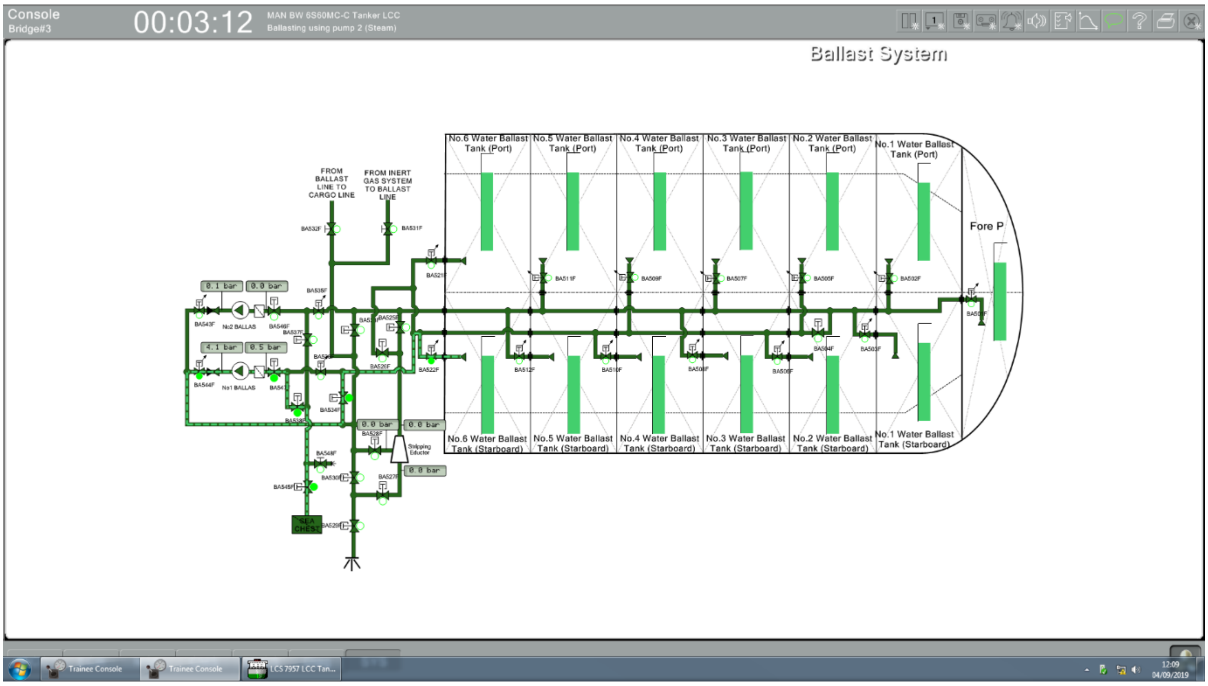


Figure 5.7 - Ballast system showing water flow through pump 1

(f) return to the LCS to identify the new flow rate as shown in Figure 5.1 above.

5.2.4.5 2nd Baseline

The last stage of the workflow requires the participants to continue to monitor the LCS until the tank has filled to the required volume set by the instructor before the task.

5.3 Results

The findings are obtained in a numerical form in this section (5.2) and their implications and practical contributions are given in Section 5.3. It has been decided for the results section to combine workflow stage 2 (Fault occurrence) with stage 3 (Fault detection) for ease of analysis due to the large difference in phase 'size'. The combined phase will be called 'Fault Detection'.

The ANOVA method allowed us to investigate the left, middle and right side of the dorsal lateral pre-frontal cortex, using mean HbO as the dependant variable with a model of 2 differing training groups, a distracted and undistracted group against a 5-stage workflow task.

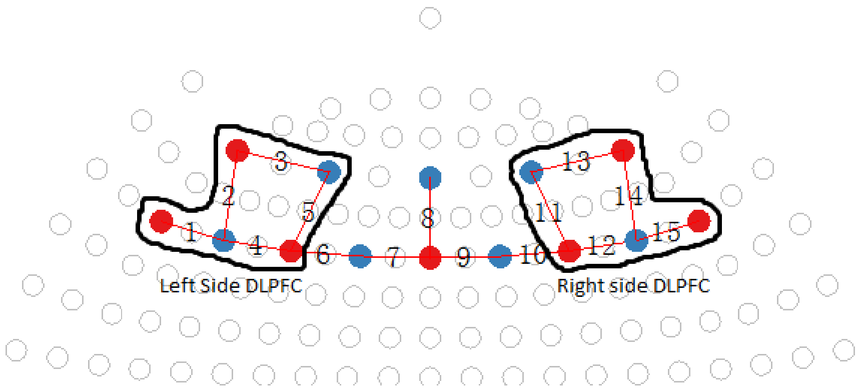


Figure 5.8 - Left (channels 1-5), medial (channels 6-10) and right (channels 11-15) sides of DLPFC

The data from the study were analysed via ANOVA procedures using SPSS v.26. Outliers were identified as any value that deviated more than 3 standard deviations from the cell mean and were omitted from ANOVA testing.

A significant effect was found between distracted and undistracted participants [$F(1,16) = 58.601$, $P<0.01$]. The mean times for participants in the distracted group were significantly slower during fault occurrence, fault detection and fault solution, when compared to those undistracted as shown in Table 5.1 below.

Table 5.1 - Mean time taken for workflow stages with respect to distraction

		Number of participants	Mean time (seconds)	Standard Deviation
Fault Occurrence	Distracted	10	41.5	2.677
	Undistracted	10	32.4	1.647
Fault Detection	Distracted	10	61.3	8.433
	Undistracted	10	53.2	6.339
Fault Solution	Distracted	10	424.7	43.166
	Undistracted	10	366.8	34.185

A significant effect was found between passively trained and practically trained participants [F(1,16) = 57.373, P<0.01. Passively trained participants were significantly slower during fault occurrence, fault detection and fault solution stages of the workflow as shown below in Table 5.2.

Table 5.2 - mean time taken for workflow stages with respect to training level

		Number of participants	Mean time (seconds)	Standard Deviation
Fault Occurrence	Practically Trained	10	36.2	5.35
	Passively Trained	10	37.7	5.1
Fault Detection	Practically Trained	10	51.2	5.051
	Passively Trained	10	63.3	6.395
Fault Solution	Practically Trained	10	364.6	32.878
	Passively Trained	10	426.9	40.709

5.3.1 **HBO Data** - Workflow (left side DLPFC)

A significant effect was found for the workflow stages [F(4,13) = 23.88, P < 0.01, Partial - $\mu^2 = 0.878$].

Pairwise comparisons to a 95% confidence (P<0.05) found significant differences between the workflow stage 4 (fault solution) and all other stages (for example fault solution (FS) and baseline1 P<0.00, FS and fault occurrence P<0.003, FS and fault detection P<0.042, FS and baseline 2 P<0.00). Shown in figure 5.9 below is the mean HBO with respect to the stage of the workflow.

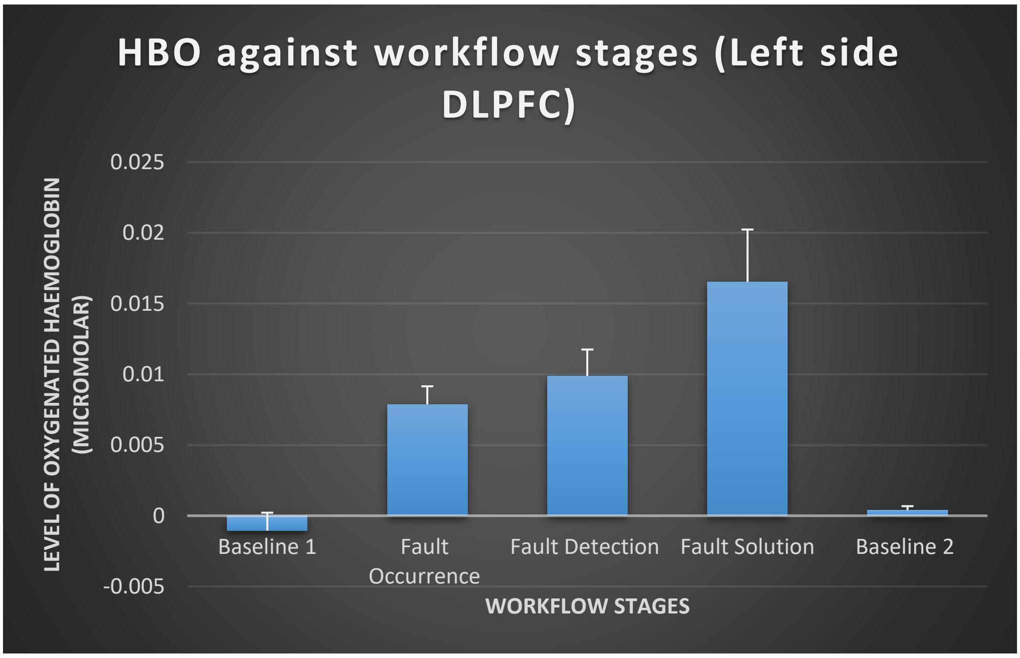


Figure 5.9 - HBO against workflow stages for the left side DLPFC

5.3.2 Practically and Passively trained participants (left side)

There was a significant effect found between practically and passively trained participants: [F (1,16) = 13.98, P < 0.01, Partial - $\mu^2 = 0.651$].

Figure 5.10 shows the significant difference found between passively trained participants', against practically trained participants' HBO levels for the workflow task.

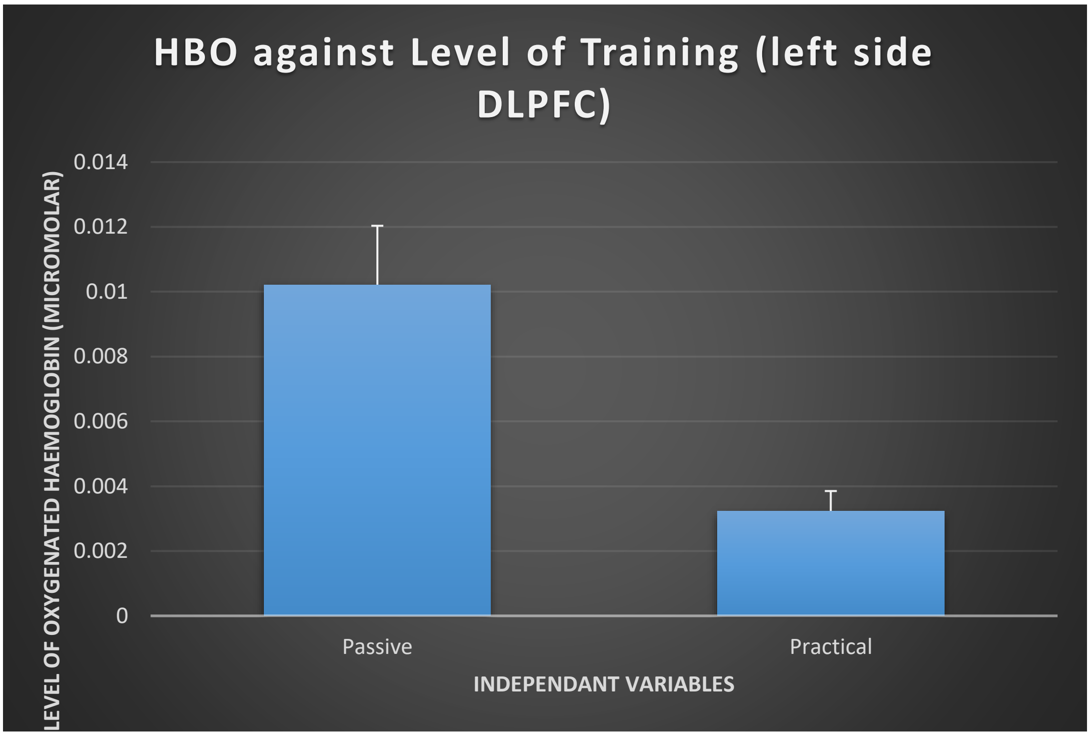


Figure 5.10 - HBO for level of Trainingl

A significant effect can be shown for distraction [F(1,16)=5.12, P < 0.05, Partial - $\mu^2 = 0.529$].

A significant difference can be seen throughout all workflow stages. Figure 5.11 shows distracted compared to undistracted participants.

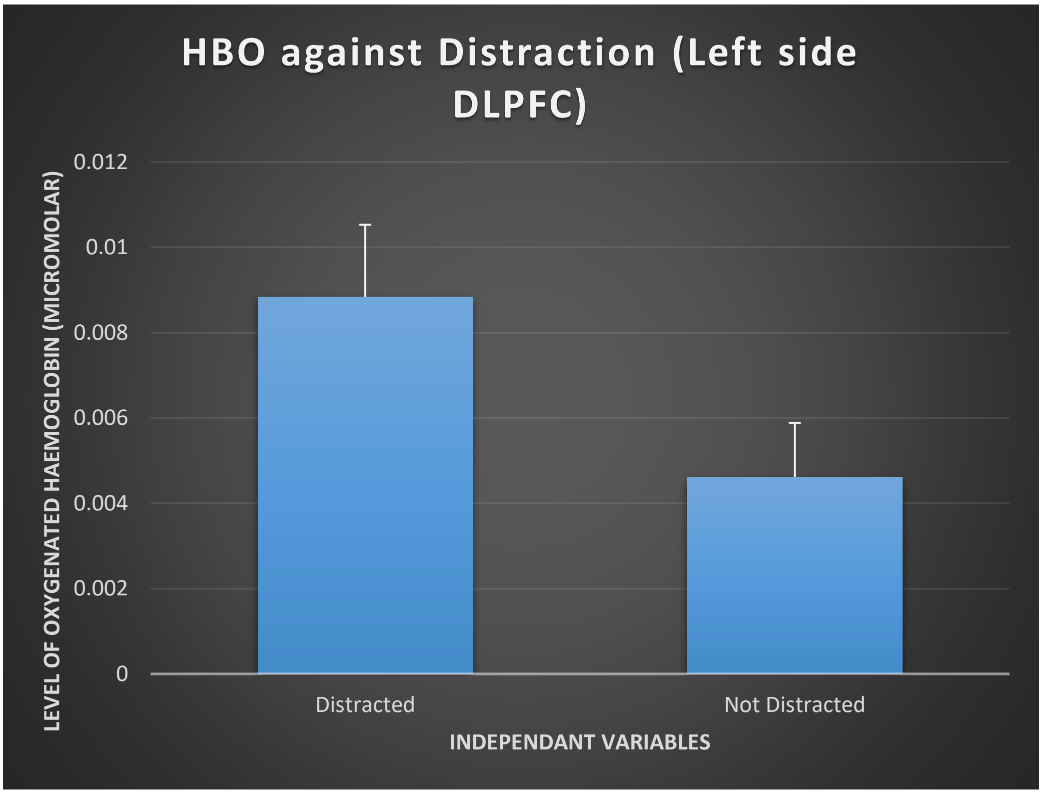


Figure 5.11 - Distracted vs not distracted participants average HBO

5.3.3 Workflow (right side DLPFC)

A significant effect was found for workflow [F(4,13)=14.301, P < 0.01, Partial - $\mu^2 = 0.815$].

Pairwise comparisons found with a 95% confidence (p<0.05), significant differences between workflow stage 4 (fault solution) and all other stages (for example FS and Baseline 1 P<0.001, FS and FO P<0.014, FS and FD P<0.001, FS and Baseline 2 P<0.001). Shown in Figure 5.12 is the HBO with respect to the stage of the workflow.

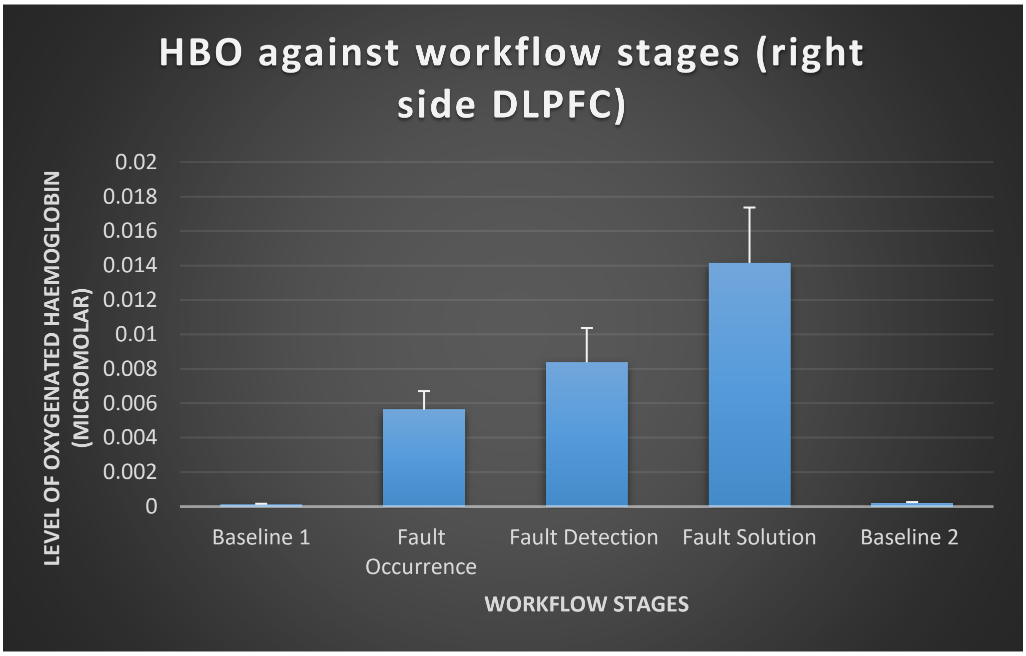


Figure 5.12 - Average HBO against workflow stages for right side DLPFC

5.3.4 Practically and passively trained participants (right side DLPFC)

There were significant effects found for participant training format [F (4,13)=12.440, P < 0.05, Partial - $\mu^2 = 0.538$].

In figure 5.13 the significant difference in HBO between passively and practically trained participants for the right side DLPFC.

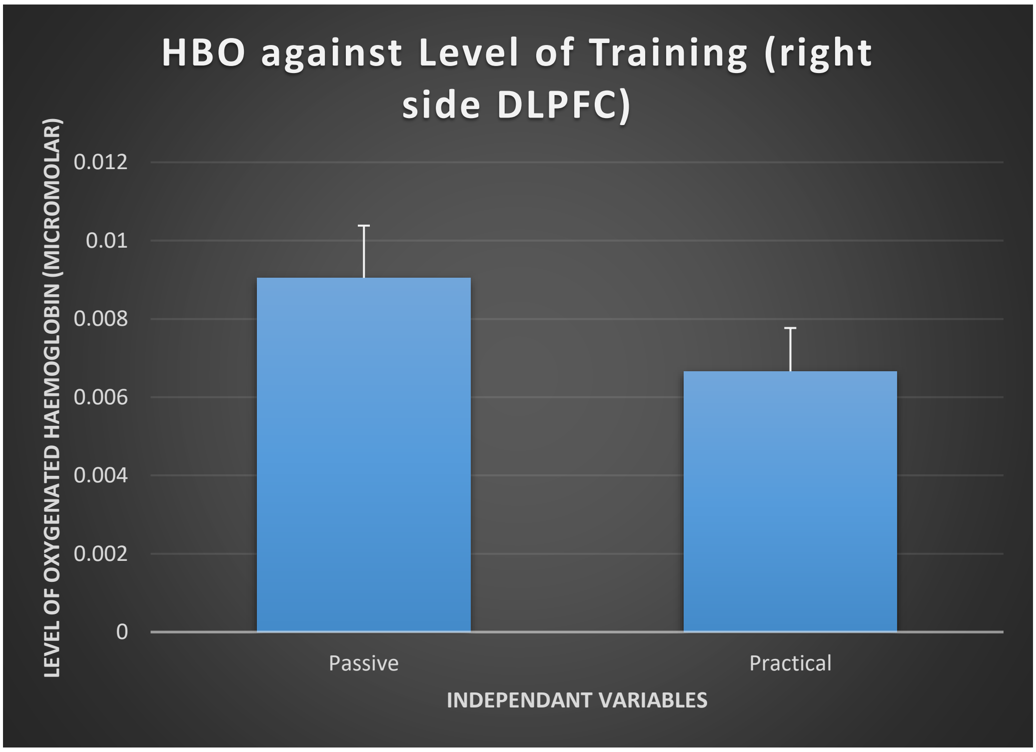


Figure 5.13 - Practically and passively trained participants HBO for the right side DLPFC

5.3.5 Distracted and Undistracted Participants (right side DLPFC)

There were no significant effects found for participant distraction [F (4,13)=0.042, P < 0.05, Partial - $\mu^2 = 0.597$].

Figure 5.14 shows the difference in HBO between distracted and undistracted participants for the right side DLPFC.

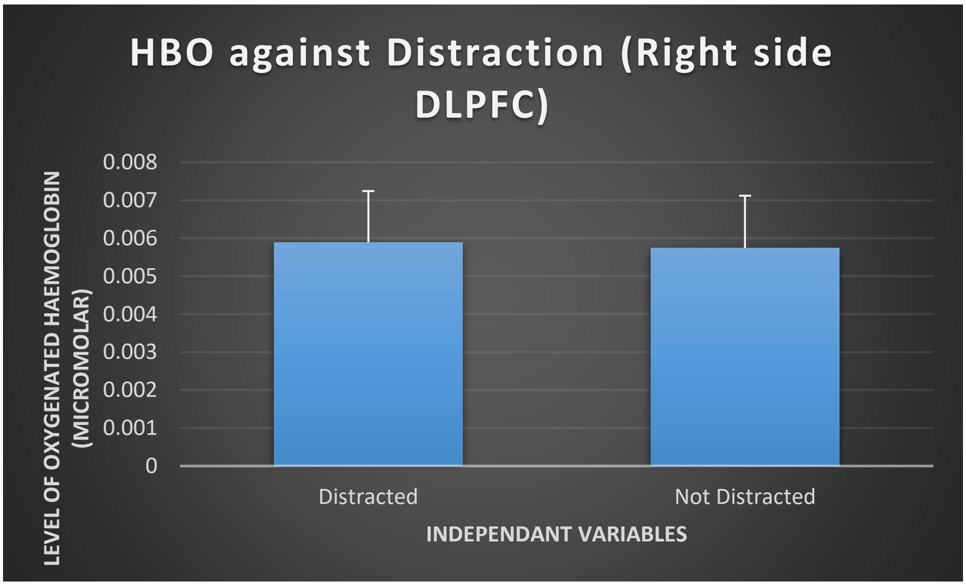


Figure 5.14 - HBO against level of distraction for right side DLPFC

5.4 Discussion

5.4.1 Explanation of findings

5.4.1.1 Workflow

The SPSS software showed that the stage of the workflow had a significant effect on HBO volume. For both, left and right sides of the DLPFC the fault solution stage (workflow stage 4) had by far the most significant effect. This is also shown in figures 5.8, 5.9, 5.10,5.11 and 5.12. This was as expected due to the fault solution stage having the most 'sub-tasks' for each participant to perform and would take the longest time to complete. This is consistent with the work done by Mehler et al [70], Aghajani [18] and Akyuz [133] as they also showed a greater level of significance for the longer more 'sub-task' driven activities. Participants showed predicted activation levels for stage 2, 3 and 4 of the workflow. Stage 2 being the fault occurrence stage where an alarm would indicate a fault. Stage 3 being the fault detection stage where the participant would be required to detect where and what type of fault had occurred. The increase in HBO volume for workflow stages 1 to 2, 3 to 4 then back down to the second baseline (stage 5) as shown in Figures 5.11 and 5.9, gives confidence in the task design for this study. This trend is consistent with the work done by Baker et al [19] and Bauernfiend et al [74] who also showed the same increase as the time and workload increases. The only difference being the effect size (8.9) and (4.1) compared to our 23.8 (left side) and 14 (right side), which could be explained by a difference in task design, training level and participant background.

5.4.1.2 Candidate training

Participants in the passively trained group showed much higher HBO volume when completing the task stages of the workflow (stages 2, 3 & 4) as shown in Figures 5.10 and 5.12. This confirms the original hypothesis in chapter 2, that practically trained participants cope better under a higher mental and physical workload. This is consistent with the work done by Fan et al [3], who also showed a significant difference in HBO volumes between novice and experienced participants on a ship's bridge simulator. Passively trained participants also took longer to complete each task compared to the experienced candidates. Again, confirming what was hypothesised in chapter 2.2 that experienced seafarers are more efficient.

5.4.1.3 Distracted Candidates

Distracted participants took longer to complete each task as shown in Table 5.1. This was expected as distracted participants had additional tasks to complete at the same time as the standard workflow tasks. However, distraction caused a higher average HBO across all candidates (practically and passively trained) when participating in the workflow tasks as shown in the Table 5.1, and Figures 5.11 and 5.13. This was expected and hypothesised in chapter 2. A possible cause for this is that participants had to re-orientate their attention from the task in hand to the reporting task and then re-orientate back to the original task. Figure 5.13 shows that cumulatively, distraction resulted in the second highest average HBO behind passively trained participants. Therefore, distracted participants are showing an increased mental workload when compared to undistracted participants, which is expected. However, the difference between undistracted and distracted average HBO is large therefore, it can be said that efforts (Risk control options) should be implemented to reduce distraction whilst seafarers are participating in an engine room task to reduce the risk of human error. Distraction was also investigated as an adverse human factor by Fan et al [14] and E.T.Solovey [28]. Distraction was shown to have a significant effect for the aforementioned studies but contrary to that, no significant effect for distraction was found in our study. This may have been due to the distraction in our study being a factor coupled with training level. Training level had such a large effect (F=14) that this could have overshadowed the effect of distraction.

5.4.1.4 Comparison of combined human factors against workflow

Figure 5.13 shows that passively trained participants had a higher HBO volume than distracted participants. This indicates that passively trained participants on average found the engineering task more challenging than distracted participants [142]. This shows that there is more of a significant effect from ‘passive type’ training than there is from distraction. This is consistent with the work done by Fan,S [14], where like our study, the effect size was greater for training (F=11.9) than the distraction (F=5). This finding confirms the hypothesis from chapter 2, stating that experienced candidates cope better with workplace factors. Also, from an interview with maritime professionals it was stated in section 1.2.2, “in reality, engine room operators are all trained to different levels depending where they were trained” [21]. Based on that statement and the findings from this study, it can be said that there is a significant link to the way in which seafarers are trained and the occurrence of human error.

The significant outcome from Figure 5.14 is that passively trained but undistracted participants have a higher HBO volume, thus mental workload, than those that are practically trained but distracted. As stated above, this indicates that there is a risk that lack of correct training could be a factor contributing towards human error within a ship engine room.

5.4.2 Limitations and modifications

5.4.2.1 The engine room simulator

The engine room simulator is limited in its capabilities as it is easy to simulate all seafarers' duties within the engine room but the only consequence to an incorrect action is an engine room alarm. The alarm is enough to neuro-physiologically activate each participant but in a 'real-life' scenario, an incorrect action could cause a physical problem [144]. For example; fire, flood, electric shock, the ship to list, the ship to sink, injury or death. The consequences listed would presumably cause an increased amount of neurophysiological activation, which is not shown on this study.

5.4.2.2 The fNIRS system

Due to the sensitivity of the infrared sensors and detectors, a desktop version of the engine room simulator was used. In reality, seafarers would be moving around whilst completing the workflow tasks used for this study but this would have caused great interference and anomalous data. For future experiments, an investigation into how it could be possible to modify the fNIRS system, so it can be used with a portable backpack would give a slightly more realistic investigation.

5.4.2.3 Candidate training

Obtaining 20 candidates with ship engine room experience to different levels, coupled with experience of using the TRANSAS software was very difficult. Therefore, in-house training had to be done for each candidate as stated in 5.1.3. The only way to differentiate experienced and inexperienced was to train the inexperienced candidates passively. Candidates were tested on passive training compared to practical training. It would have been preferable to test 10 participants with 10 years+ ship engine room experience and 10 participants with a few weeks ship engine room experience, all 20 of which trained the same way.

5.5 Suggestions for Second scenario (chapter 6).

It would be beneficial to study in a similar way, the effects of an increased workload. A comparison could be made of distraction against increased workload to see the effect on neurophysiological activity, indicating the higher probability of the two, to contribute towards human error.

It was found that the different training formats (passive and practical) are unrealistic as there was too much of a gap in knowledge between passively trained and practically trained participants. Passively trained participants found the task very difficult, resulting in their need to consult the user manual considerably more regularly than the practically trained candidates, which could result in an increased number of artefacts and anomalous results due to additional head movement.

It has been decided for the 'increased workload' experiment (chapter 6 below) to train every participant actively. This way all participants should be able to complete the task without referencing the manual or prompting by the instructor. Therefore, an additional test was conducted on another 10 candidates. All were practically trained with 5 distracted and 5 not distracted (10 practically trained and distracted, 10 practically trained but not distracted). This enabled us to see the effects of distraction on practically trained participants. The results of this additional test are below in 5.5.

The above study is a test of the validity of the experiment and analysis. The amount of data and aspects to consider for this study are extremely large therefore, analysis results are compartmentalised (for example, right side, left side DLPFC, training level, distraction level etc) purposely to further understand the parts of the analysis with the most relevance. The findings are purposely reported in the style above to understand the interactions between the participant, the simulator and the fNIRS equipment. This will allow for a more concise future analysis, omitting the less relevant findings from the first test above. The studies that follow will be reported differently using human factors methods.

5.6 Results – *Test 2. All candidates given practical training*

The data from the study were analysed via ANOVA procedures using SPSS v.26. Outliers were identified as any value that deviated more than 3 standard deviations from the cell mean and were omitted from ANOVA testing.

5.6.1 Times to complete workflow phases

The time taken to complete each workflow phase was subjected to a 2 (standard/distracted) x 3 (fault occurrence/fault detection/fault solution) ANOVA. As expected, this model revealed a significant main effect for workflow phase [F(2,17) = 12544.14, P<0.01, eta^2 = 0.99], The fault solution phase (M=364.9, s.e = 2.89) took significantly longer than fault detection (M=51.4, s.e=0.59) and the fault occurrence phases (M=36.3, s.e = 0.44). There were also significant main effects for distraction [F(2,17) = 95.982, P<0.01, eta^2 = 0.919, distraction resulted in higher times (M=527.6, s.e 4.19) compared to a standard test (M=452.4, s.e.=3.39). The ANOVA also revealed a significant interaction between factors [F(2,17) = 210.06, p<.01, eta^2 = 0.99]. Post-hoc t-tests revealed that the time to complete the fault solution phase was significantly longer for distracted participants compared to standard test participants [t(17) = 11.97, p<.01] as depicted in figure 5.15; however, there was no effect of distraction on time during the fault occurrence and fault detection phases.

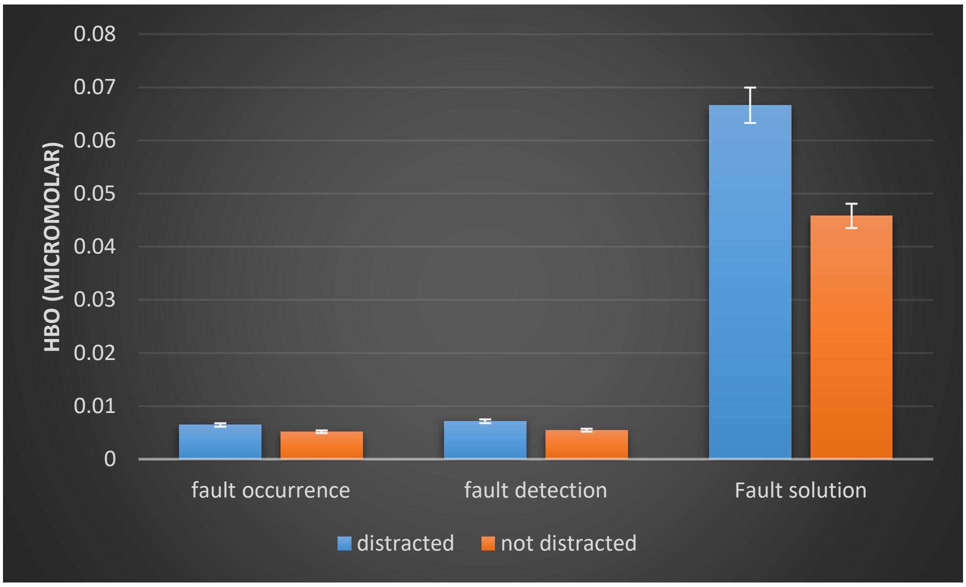


Figure 5.15 - Distraction against a standard test with respect to workflow stage

5.6.2 Analysis of fNIRS data

Average levels of oxygenated haemoglobin (HbO) were estimated using fNIRS for fault occurrence, fault detection and fault solution phases of the workflow. Data from all channels were averaged into three Regions of interest (ROI) corresponding to left, medial, and right regions of the dorsal lateral pre-frontal cortex. All HbO data were subsequently baselined using data gathered during the first phase of the workflow that lasted for 300s, i.e., baselined HbO = HbO during task phase minus HbO during 300s baseline period, hence positive HbO values indicate an increase above the baseline levels.

Activation of the pre-frontal cortex during the fault occurrence and fault detection phases was explored via a 2(Distraction/Standard) x 3 (left, medial and right ROI) ANOVA. The analysis revealed no significant effects for distraction or ROI and no significant interaction.

Activation of the pre-frontal cortex during the fault solution phase was explored again via a 2 (distraction/standard) x 3 (left, medial and right ROI) ANOVA. This analysis revealed significant main effects for distraction [$F(1,18) = 94.7$, $p<.01$, eta^2=0.95] and ROI [$F(1,18) = 37.19$, $p<.01$, eta^2= 0.774]. The ANOVA also revealed a significant interaction between factors [$F(1,18) = 69.17$, $p<0.1$, eta^2=0.95]. The main effect for distraction indicated that the mean HbO was significantly higher during the distraction stressor (M = 0.029, s.e. = .002) compared to a standard test (M = .009, s.e = .002). For ROI, the main effect revealed that mean HbO at medial ROI2 (M = .0014, s.e. = .002) was significantly lower than either left lateral ROI1 (M = .019, s.e. = .002) or the right lateral ROI3 (M = .027, s.e. = .003) ($p<.01$) – See figure 5.16.

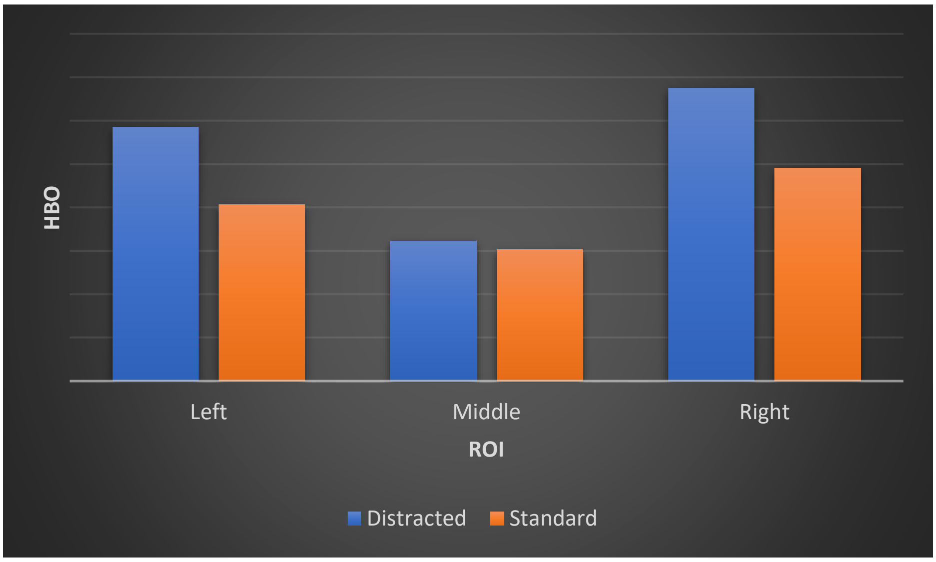


Figure 5.16 - Distraction vs standard test with respect to ROI

In order to explore the interaction, a number of post-hoc t-tests were conducted. These tests revealed that the mean HbO was significantly greater during a distracted test compared to a standard test at left lateral ROI1 [$t(18) = 5.03$, $p<.01$] and right lateral ROI3 [$t(18) = 8.15$, $p<.01$]. There was no significant effect for distraction at medial ROI2.

5.7 Discussion

5.7.1 Workflow

5.7.1.1 Time

The ANOVA showed a significant effect for workflow with respect to time when comparing the fault solution workflow stage with other workflow stages. This can be explained due to the larger number of sub-tasks within the fault solution stage. Table 5.3 shows the differences between the workflow stages.

Table 5.3 - Average time for workflow stages

	Number of Participants	Mean Time (seconds)	Standard Deviation
Fault Occurrence	20	32.4	1.8
Fault Detection	20	53.2	5.72
Fasult Solution	20	366.8	27.74

5.7.1.2 fNIRS data

The ANOVA showed a significant effect for workflow with respect to HBO. Tests of within-subjects effects also showed significance. This can be explained due to the fault solution stage having more sub-tasks and therefore more navigation through the simulator screens. Navigation from screen to screen completing sub-tasks will activate working memory in order to keep up to date with what has been done and what needs to be done. An argument could be made that due to what has been previously stated the fault solution stage of the workflow will be harder to complete, resulting in higher levels of activation. This shows consistency with other studies [17] [18] [35] [66]. The workflow phases involving multiple sub-tasks and longer time scales on average trend towards higher HBO volumes and significant effects, which is to be expected. However, there was a slight difference between our study and the studies referenced above. The sub-tasks in our study are all different whereas, for the other studies there is repetition of sub-tasks multiple times. It can be shown in other studies that repetition of the same tasks results in reduced overall activation compared to differing tasks [76].

5.7.2 Stressor

5.7.2.1 Time

The ANOVA showed a significant interaction for multivariate tests and within-subjects effects between workflow and distraction with respect to time. The mean times for participants in the distracted group were significantly slower during fault occurrence, fault detection and fault solution, when compared to those not distracted as shown in Table 5.4.

Table 5.4 - Average time for workflow stages with respect to Distraction

	Stressor	Number of Participants	Mean Time (Seconds)	Standard Deviation
Fault Occurrence	Distracted	10	33.2	1.05
	Standard Test	10	31.6	2.06
Fault Detection	Distracted	10	58.6	1.42
	Standard Test	10	47.8	1.49
Fault Solution	Distracted	10	392.4	10
	Standard Test	10	341.2	8.18

This can be explained as the technique used to manipulate distraction involves the participant navigating to the liquid cargo screen and reading out the desired output. This takes approximately 10-15 seconds for each question.

5.7.2.2 fNIRS data

The ANOVA showed a significant interaction for workflow and distraction with respect to HBO. This is what was expected as shown in figure 5.17, the mean HBO for participants in the distracted group is higher on average during fault occurrence, fault detection and fault solution workflow stages compared to participants who were not distracted. Again, as previously mentioned above, distracted participants had more tasks to complete. Also, the examiner noted that distracted participants attempted to rush when asked the ‘distraction question’ which would indicate stress [14].

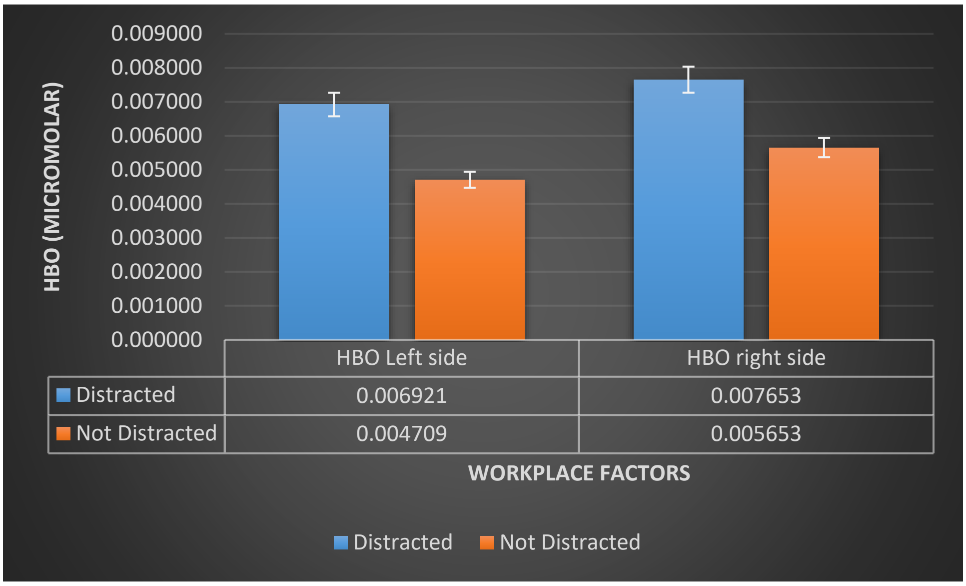


Figure 5.17 - Comparison of HBO for workflow stages with respect to distraction.

5.7.2.3 Region of Interest

The fNIRS equipment showed that the left and right sides DLPFC have a greater sensitivity when detecting oxygenated haemoglobin compared to the middle region for our experiments. This is consistent with other neuroimaging investigations using fNIRS [18] [52] [54] where the left and right sides of the PFC showed fewer anomalies and more usable data for further analysis. This is due to the participants needing the use of the specific functions of the left and right side DLPFC regions when undertaking the tasks, compared to that of the middle region [139]. This justified the omission of the middle region DLPFC from our results.

The ANOVA showed no significant effect for ROI, no significant interaction between ROI and distraction and no significant interaction between ROI, workflow and distraction. This can be explained again due to the small measurements found within left, right and middle HBO volume. However, there is a trend in every test that shows left and right sides of the DLPFC to have a higher (but similar) output of HBO when compared to the middle region.

A significant interaction was found between ROI and workflow. This can be explained due to the large difference in HBO between workflow stages (more specifically the fault solution stage) accentuating the middle region's low sensitivity to our task type.

5.7.3 Validity check

The data gathered via fNIRS and ANOVA analysis was validated within the limits available. It is always preferable to validate work based on what has already been proven. However, due to the novelty of this project it is impossible to test the findings from fNIRS on a simulated environment against real life events at sea. Therefore, validity has been achieved via:

- 10 fold cross validation of fNIRS datasets from all 15 channels.
- Data was separated in epochs for cross validation to prevent ‘double-dipping’.
- Outcomes are checked against R-studio, MatLab and Python software platforms.
- A manual ‘step by step’ linear regression with ANOVA was applied using excel to check the validity of each workflow stage against PSF.

5.8 Further suggestions for the second study (chapter 6).

Having all candidates actively trained helped considerably reduce artefacts and anomalous results brought about by passively trained candidates constantly needing to move their head down to refer to the simulator manual.

The time taken to analyse the results was greatly reduced by setting up the fNIRS equipment filters prior to testing in such a way that they ran throughout the test. Therefore, there was no need to add additional filtering and data processing post experiment apart from our custom filter mentioned in 4.3.

5.9 Concluding remarks

The effect of training type had too large an effect on participants and is deemed an unrealistic level of training to simulate 'real-life' scenarios. Therefore, passive training will no longer be used in further tests. Moreover, a second distraction study was conducted with all participants trained using practical methods.

When compared to a standard test, from the ANOVA, distracted participants showed higher levels of activation. Distraction was shown to also have a significant effect on time taken, resulting in distracted participants taking longer to complete each task.

The workflow was also shown to have a significant effect with respect to time taken to complete each phase. The fault solution stage took participants significantly longer to complete. The fault solution phase also induced higher levels of neurophysiological activation in the DLPFC.

Chapter 6. The effect of increased workload on marine engineers whilst conducting ballast water operations using fNIRS.

6.1 Introductory remarks

This chapter uses a similar task and workflow as described in the previous chapter, the difference being that this chapter looks at the PSF workload. The workload stressor is manipulated in a different way to the distraction stressor. However, the same workflow stages and ballasting scenario are used.

The first section of the chapter defines the changes needed to the experimental design in order to investigate the workload stressor from the distraction stressor. Moreover, the participants used, participant training and the experiment is also defined. Followed, is the results of the experiment for time and HBO. The final section is a discussion of the findings from the workload study. The chapter concludes with a brief suggestions paragraph, detailing suggestions for the next test in the following chapter. These suggestions are based on aspects that could have been improved in the previous chapter.

6.2 Method

6.2.1 Experiment Design

The design of this experiment very closely replicated that of the distraction test. The difference being that all twenty candidates were given an increased workload with ten of the twenty given a distracted test in addition to an increased workload. Another difference being that every participant was actively trained as mentioned in 5.4. All twenty participants performed the same tasks and were distracted in the same way as previously mentioned in 5.1.1. The difference being, all participants with an increased workload test had to fill additional ballast tanks simultaneously. This replicates what would be done in real operations [60].

6.2.2 Experiment Participants

20 candidates were used for this study. All 20 had qualifications to the level of a BEng or higher in marine engineering. 15 of the 20 participants had experience of working at sea or in a ship engine room. 2 of the participants were marine engineering PhD students. 2 were ex-navy engineers. The average age of the increased workload group was 34 and the remaining participants was 27. All 20 participants were male. The rest were a mixture of post graduate MEng marine engineering students and undergraduate marine students in their masters year.

6.2.3 Participant Training

All candidates were trained using the same methods as previously mentioned in 5.1.3. The difference being that all 20 participants were trained actively for this study.

6.2.4 The Experiment

This experiment replicated stages mentioned above in 5.1.4.1 (1st baseline) to 5.1.4.4 (fault solution). The difference being that the 'increased workload', participants would have to monitor multiple tanks (6 in total) for the 1st baseline stage. This is the same for the 2nd baseline stage. The main differences are found within the fault solution stage of the workflow described below.

6.2.5 Fault Solution

All 20 participants had to determine a solution to correct the fault by: (a) navigating to the cargo control room ballast pump screen (Figure 5.3) and switch off steam pump (number 2 pump), (b) access the ballast system mimic panel (Figure 6.1).

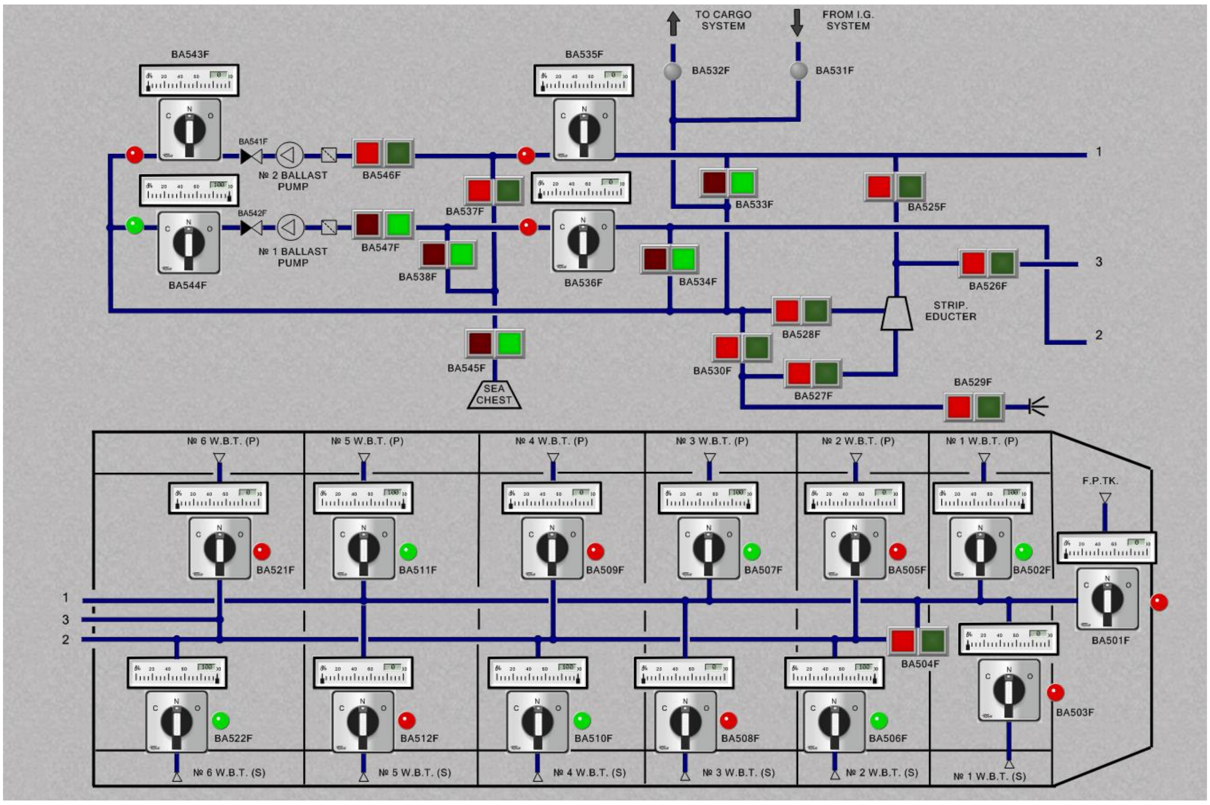


Figure 6.1 - Ballast system mimic panel

(c) calculate the new waterline needed through the electric pump (number one pump), (d) open valves BA538F, BA547F, BA533F, BA511F, BA510F, BA507F, BA506F, BA502F and BA544F and close valves BA537F, BA546F and BA543F in order to re-route the water line to ballast pump number 1, (e) navigate back to the cargo control room ballast pump screen (Figure 5.3) and switch on the electric pump, (f) navigate back to the ballast system screen to ensure the fault is solved and the system is ballasting into the correct tanks as shown in Figure 6.2.

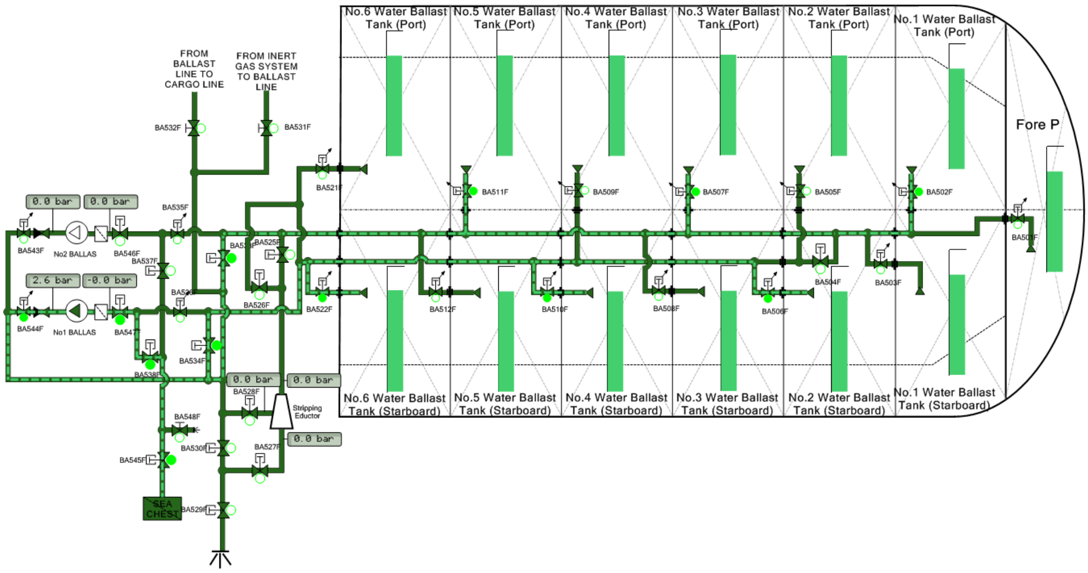


Figure 6.2 - Ballast system screen

Then, (g) navigate back to the liquid cargo screen to monitor the tanks to the required fill volume as depicted below in Figure 6.3.

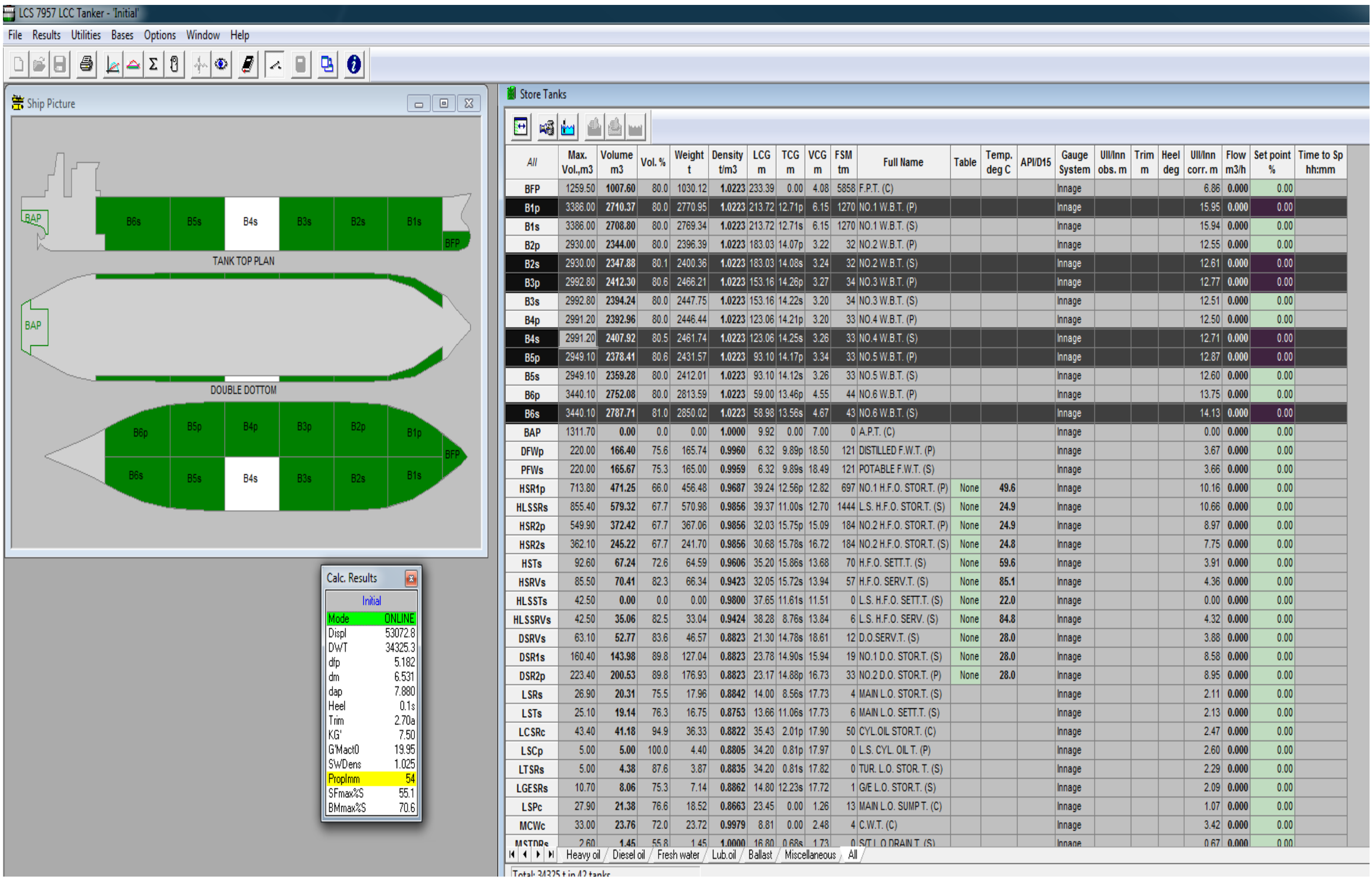


Figure 6.3 - Liquid cargo screen

6.3 Results

The data from the study were analysed via ANOVA procedures using SPSS v.26. Outliers were identified as any value that deviated more than 3 standard deviations from the cell mean and were omitted from ANOVA testing.

6.3.1 Times to complete workflow phases

The time taken to complete each phase of the workflow was subjected to a 2 (standard/high workload) x 2 (fault detection/fault solution) ANOVA. This model revealed significant main effects for workload [$F(1,18) = 301.40$, $p<.01$, $eta^2=0.94$] with high workload resulting in higher times ($M = 269.8s$, s.e. = 2.61) compared to standard workload ($M = 210.30s$, s.e. = 2.43). There was also a significant main effect for workflow phase [$F(1,18) = 9808.98$, $p<.01$, $eta^2=0.99$], which was unsurprising as Fault Solution took significantly longer ($M = 401.81s$, s.e. = 3.23) than fault detection ($M = 78.32s$, s.e. = 0.63). The ANOVA also revealed a significant interaction between both factors [$F(1,18) = 356.19$, $p<.01$, $eta^2=0.95$]. Post-hoc t-tests revealed that the time to complete the fault solution phase was significantly longer during high workload compared to standard workload [$t(18) = -18.42$, $p<.01$]; however, there was no effect of workload on time during the fault detection phase – see Figure 6.4 for illustration.

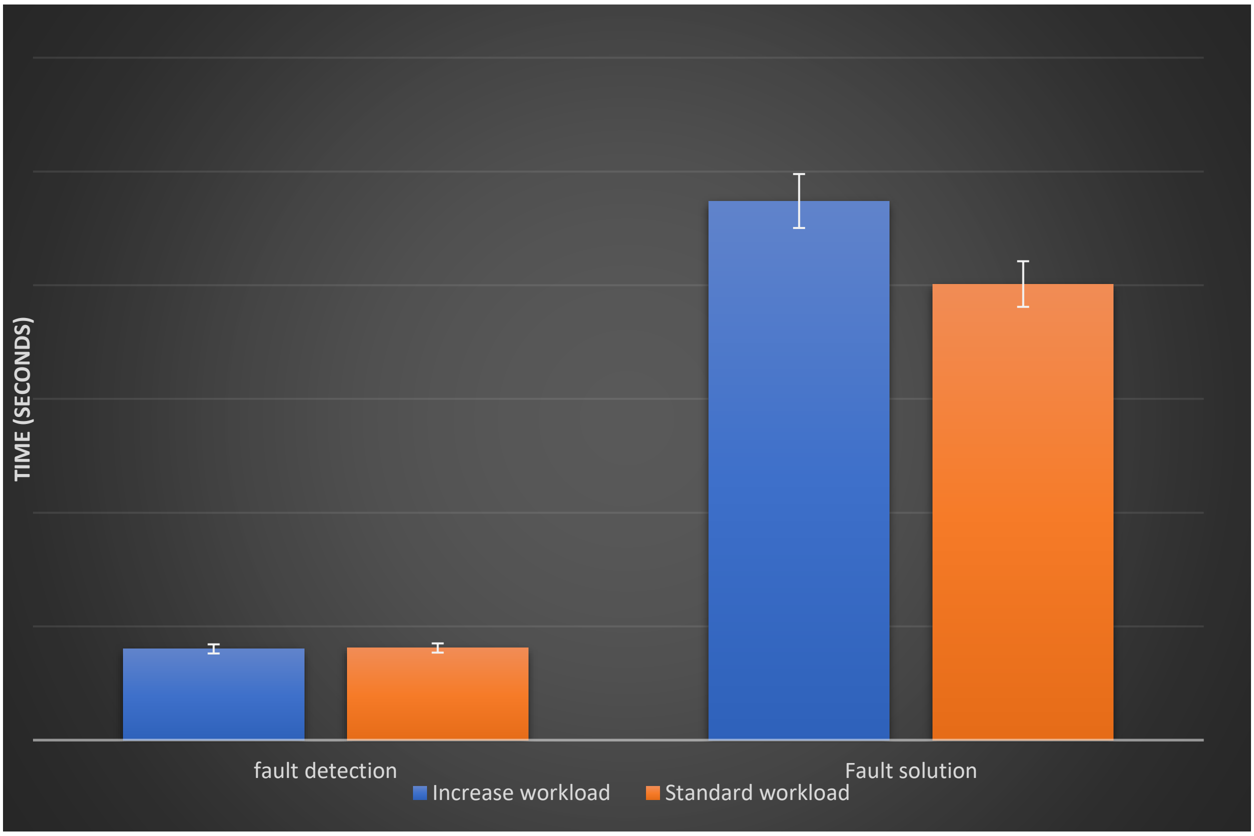


Figure 6.4 - Average times to complete each phase of the workflow for standard and high workload (N=20).

6.3.2 Analyses of fNIRS data

Average levels of oxygenated haemoglobin (HbO) were estimated using fNIRS for fault detection and fault solution phases of the workload flow. Data from all channels were averaged into three Regions of Interest corresponding to left, medial and right regions of the prefrontal cortex. All HbO data were subsequently baselined using data gathered during the first phase of the workload that lasted for 300s, i.e., baselined HbO = Hbo during task phase minus HbO during 300s baseline period, hence positive HbO values indicates an increase above the baseline levels.

Activation of the prefrontal cortex during the fault detection phase was explored via a 2 (standard/high workload) x 3 (left, medial and right ROI) ANOVA. This analysis revealed no significant effect for workload [$F(1,18) = 2.02$, $p=0.17$] or ROI [$F(1,18) = 1.05$, $p=0.33$], and no significant interaction.

Activation of the pre-frontal cortex during the fault solution phase was explored via a 2 (standard/High workload) x 3 (left, medial and right ROI) ANOVA. This analysis revealed significant main effects for workload [F(1,18) = 152.1, p<0.1, eta2=0.894] and ROI [F(1,18) = 40.46, p<0.01, eta2=0.692]. The ANOVA also revealed a significant interaction between both factors [F(1,18) = 82.19, p<0.01, eta2 = 0.906]. The main effect for workload indicated that mean HbO was significantly greater during the high workload condition (M = .052, s.e. = .002) compared to the low workload condition (M = .009, s.e. = .002). For ROI, the main effect revealed that mean HbO at medial ROI2 (M = .006, s.e. = .002) was significantly lower than either left lateral ROI1 (M = .038, s.e. = .005) or the right lateral ROI3 (M = .047, s.e. = .002) (p<.01) – see Figure 6.5.

In order to explore the interaction, a number of post-hoc t-tests were conducted. These tests revealed that mean HbO was significantly greater during high workload compared to low workload at left lateral ROI1 [t(18) = -6.53, p<.01] and right lateral ROI3 [t(18) = -15.54, p<.01], but there was no significant effect of workload at medial ROI2, see Figure 6.5 for descriptive statistics.

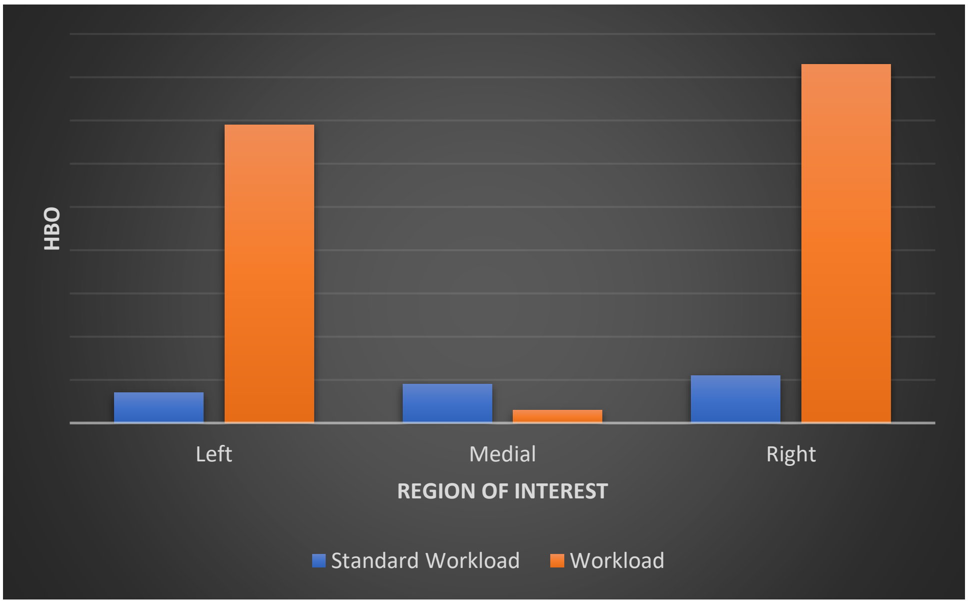


Figure 6.5 - Average HBO for each ROI with respect to standard (blue) and high workload (red).

6.4 Discussion

6.4.1 Time

The significant effect found was expected as participants in the increased workload group had additional tasks to perform resulting in a longer time taken. The mean times for the increased workload group was higher in only the fault solution stage. This was due to the additional tasks being only in the fault solution stage of the workflow. However, it was predicted that participants would take longer for every stage of the workflow as additional checks are needed in stages 2, 3 & 4. These checks are all on the same simulator screen as the standard workload checks. This could result in similar mean times for the fault detection workflow stage.

Post-hoc t-test showed significant differences between all workflow stages. This is expected as the t-tests solely compared the differences in time between workflow stages. There was a difference in time between all workflow stages due to the different activities contained within each stage. As predicted the size of the effects from t-tests showed that the largest effect is always between fault solution and other workflow phases. This is due to the largest mean time difference being for the fault solution stage of the increased workload group.

6.4.2 fNIRS data

A significant effect was found for increased against standard workload participants. This is to be expected as the increased workload participants had significantly more tasks with a greater level of difficulty. Again, like above, the fault solution stage is the workflow stage that contains the greatest difference in task volume in comparison to a standard workload. This explains the significant effect found only in the fault solution workflow stage. Unpredicted was the level of increase between standard and increased workload participants for the fault solution stage of the workflow as shown on the bar chart (Figure 6.4). This shows that the participants in the increased workload group were experiencing a much higher mental workload from the additional tasks. This shows consistencies with Verdiere et al and Dehais et al neuroimaging studies involving an increased workload element [45], [54]. Verdiere's investigation, similar to our study, found that participants undertaking a task with an increased workload (manual landing as opposed to automated) showed a significant effect when compared to that of a standard test (automated landing). Dehais' investigation used a number of air traffic control instructions in order to evaluate working memory whilst in flight using a simulator. Similar to our study, Dehais showed that increased workload resulted in an adverse effect on operator performance and in his case, flight safety.

The tasks in the increased workload group would closely resemble those done by a marine engineer in a ‘real-world’ situation [62].

Maulchy’s test of sphericity showed a significant effect. Therefore, a test of within-subjects effects was done showing a large effect size (F=152.1, t=14). This is to be expected as the individuals go from a mundane monitoring task to a complex sea water line re-routing task. When compared to other studies [3], [42]this effect size found in our study is considerably greater. Fan’s [14] investigation of increased workloads shows an effect size of 4.1 and Kojima’s [58] investigation showed an effect size of 7. This could be expected as our investigation of increased workload involved five more ballast tanks when compared to a standard test. Whereas Fan, S used additional verbal reporting as opposed to a practical approach to increasing workload, which could be said is a less intense workload increase than in our study. Kojima et al used a more similar approach to that in our study as he used additional train signalling, traffic and manoeuvring hence why his effect size is closer to ours. The issue with Kojima’s study is that the operator experience was not stated. Due to their effect size of 11 against ours (14), one can only predict that the difference may have been due to operator experience as the participants for our study had a limited time of training (3 hours) whereas the participants in Kojima’s study could have had months or years.

Post-hoc t-test between increased and standard workload showed a significant effect. This is to be expected given the significance found in the fault solution workflow stage. The size of the effect will be greater if the comparison was made using the fault solution stage instead of the mean from all workflow stages. This is again due to the fault solution stage being the only stage with a significant increase in mental workload for the increased workload group.

A significant interaction was found between workload and workflow. This interaction is again, due to the fault solution stage of the workflow as detailed above.

6.4.3 Region of Interest (ROI)

A significant effect was found for ROI. The left and right regions of the DLPFC were shown to have a much higher mean HBO when compared to the middle region. The reason for this is that the left region of the DLPFC controls remembering goals and any instructions you need to accomplish those goals, speech, comprehension, arithmetic, writing and positive feelings. The right region controls creativity, planning, maintaining focus, spatial ability, artistic, music skills and negative feeling [145]. Whereas the middle region is involved mainly in body regulation, attuned communication, emotional balance, empathy, self-awareness and fear modulation [72]. Therefore, it is to be expected that the middle region would have a lower mean HBO. Interesting is the right region showed a higher mean HBO. This could be due to the participants being new to the simulator system and thus taking a pessimistic approach. This tells us that on average, participants are focusing on creativity and planning rather than working memory. This is consistent with the work done by *Fairclough et al* [52] where they also found the left and right side DLPFC to have the most significance when undertaking a task associated with executive function, working memory and selective attention.

Pairwise comparisons showed a significant effect for ROI. The most significance was found between left – middle and right – middle. This is to be expected when taking into account what has been previously mentioned above. Interesting is that for t-tests the largest effect size was shown for left against the middle region when the bar chart (Figure 6.5) shows that the mean HBO was higher for the right region of the DLPFC. This makes sense as our task involves a large amount of working memory, controlled by the left region.

A significant interaction was found for ROI with respect to workflow. This is predicted as the fault solution stage required a larger number of the right and left region functions when compared to the middle region.

A significant interaction was found for ROI*workflow*distraction. This can be explained as there is a large effect for ROI and workflow individually resulting in a combined significant effect. This is expected due to the left and right regions' interaction with the fault solution stage being larger and exponentially greater for distracted participants. The reason for this could have been due to the left and right regions' functions being needed for our task compared to the middle region as mentioned above. This coupled with the fault solution stage being the most onerous and having more levels of distraction.

A significant effect was found for ROI* increased workload against standard workload. This is to be expected as the mean HBO will be higher for an increased workload for the same task. Interesting is that for a standard workload the right region of the DLPFC showed a higher mean

HBO but for increased workload the left region of the DLPFC showed a significantly higher mean HBO. The reason for this could be that the increased workload task involved 5 additional ballast tanks. This would require participants to remember the tanks being filled, the fill levels and the previous water line used. Working memory being one of the main functions of the left region DLPFC.

Post-hoc t-test showed a significant effect for increased workload for the left and right regions of the DLPFC. The region with the largest effect for the increased workload task was the left region. Again, this can be explained as previously mentioned above and further substantiates the premise that working memory is a prevalent function needed for this task.

6.4.4 Validity check

The data gathered via fNIRS and ANOVA analysis was validated within the limits available. It is always preferable to validate work based on what has already been proven. However, due to the novelty of this project it is impossible to test the findings from fNIRS on a simulated environment against real life events at sea. Therefore, validity has been achieved via:

- 10 fold cross validation of fNIRS datasets from all 15 channels.
- Data was separated in epochs for cross validation to prevent 'double-dipping'.
- Outcomes are checked against R-studio, MatLab and Python software platforms.
- A manual 'step by step' linear regression with ANOVA was applied using excel to check the validity of each workflow stage against PSF.

6.5 Suggestions for third scenario

A trial study was done using participants with no engineering background. The results showed that due to the training given to each participant, the difference between those who have prior engineering experience and those who don’t is negligible. Therefore, the researcher could extend the recruitment search for participants to include those with and without marine engineering experience.

6.6 Concluding remarks

The experiment for this study was changed to include 6 ballast tanks as opposed to a single tank in the previous chapter. This change required participants to perform additional tasks when re-routing the water line and checking the ballast system and LCS resulting in an increased workload.

When compared to a standard test, from the ANOVA, increased workload participants showed higher levels of neurophysiological activation in the DLPFC. This resulted in a significant effect shown for increased workload participants with respect to time taken and HBO. A significant effect was also found for workflow with respect to time and HBO. The fault solution workflow phase induced the highest levels of activation and longest times taken to complete the stage.

The ROIs with the most significance were the left and right (channels 1-5 & channels 11-15) sides of the DLPFC. The middle region (channels 6-10) showed varying levels of sensitivity, anomalies and outliers which had to be omitted due to tolerance. Therefore, the middle region data was deemed invalid/less accurate.

Chapter 7. The effect of fatigue on marine engineers' whilst conducting ballast water operations using fNIRs.

7.1 Introductory remarks

This chapter investigates the effect of fatigue on OFS whilst conducting a fault detection and correction ballasting task. The first section of the chapter defines the differences in experiment design, the participants used, the participant training and the experiment to investigate fatigue compared to distraction and increased workload on the previous two chapters. This is followed by the results of the investigation. The results outline significant effects of time, the fatigue stressor and workflow on OFS using an ANOVA study. The results are then discussed in the final section of the chapter outlining the relevant findings from the investigation.

7.2 Method

7.2.1 Experiment Design

The design of this experiment very closely replicated that of the workload and distraction tests. The difference being that all 20 candidates were given a fatigue test with 10 of the 20 given an increased workload in addition to being fatigued. Again, every participant was actively trained as mentioned in the previous study. The 10 participants in the increased workload group performed the same tasks as stated in 6.1.4. The difference being, all participants with a fatigue test had to sit through an extended monitoring task. This replicates what would be done in real engine room operations [60].

7.2.2 Experiment Participants

20 candidates were used for this study. All 20 had qualifications to the level of a BEng or higher in marine engineering. 18 of the 20 participants had experience of working at sea or in a ship engine room. 2 of the participants are marine engineering business owners also qualified to degree level in marine engineering. The average age of the fatigued group is 23. All 20 participants were male.

7.2.3 Participant Training

All candidates were trained using the same methods as previously mentioned in 5.1.3. All 20 participants were trained actively.

7.2.4 The Experiment

This experiment replicated stages mentioned above in 6.1.4 (fault occurrence) to (2nd Baseline). The difference being that all fatigue participants would have a 35-minute (2100s) monitoring task instead of the normal 5 minutes (300s) for the 1st baseline stage of the workflow. The participants in the fatigue group will have the above mentioned 35-minute (2100s) monitoring task for the 1st baseline stage but followed by a standard task involving a single ballast tank as previously mentioned in section 5.1.4, instead of the six tanks for the increased workload group.

7.3 Results

The data from the study were analysed via ANOVA procedures using SPSS v.26. Outliers were identified as any value that deviated more than 3 standard deviations from the cell mean and were omitted from ANOVA testing.

7.3.1 Times to complete workflow stages

The time taken to complete each phase of the workflow was subjected to a 2 (standard/fatigued) x 2 (fault detection/fault solution) ANOVA. This model revealed significant main effects for fatigue [$F(1,18) = 71.41$, $p<.01$, $eta^2 = 0.99$] with fatigue resulting in higher times ($M = 236.4$s, s.e = 2.41) compared to a standard test ($M = 210.3$s, s.e = 2.43). There was also a significant main effect for workflow phase [$F(1,18) = 7185.51$, $p<.01$, $eta^2 = 0.95$], which could have been predicted as the fault solution stage took significantly longer ($M = 249.6$, s.e = 2.47) to complete compared to the fault detection stage ($M = 48.1$, s.e = 0.49). The ANOVA also revealed a significant interaction between both factors [$F(1,18) = 95.8$, $p<.01$, $eta^2 = 0.93$]. Post-hoc t-tests revealed that the time to complete the fault solution stage was significantly longer whilst fatigued compared to a standard test [$t(18) = 9.15$, $p<.01$]; However, there was no effect of fatigue on time during the fault detection workflow phase – see illustration in Figure 7.1.

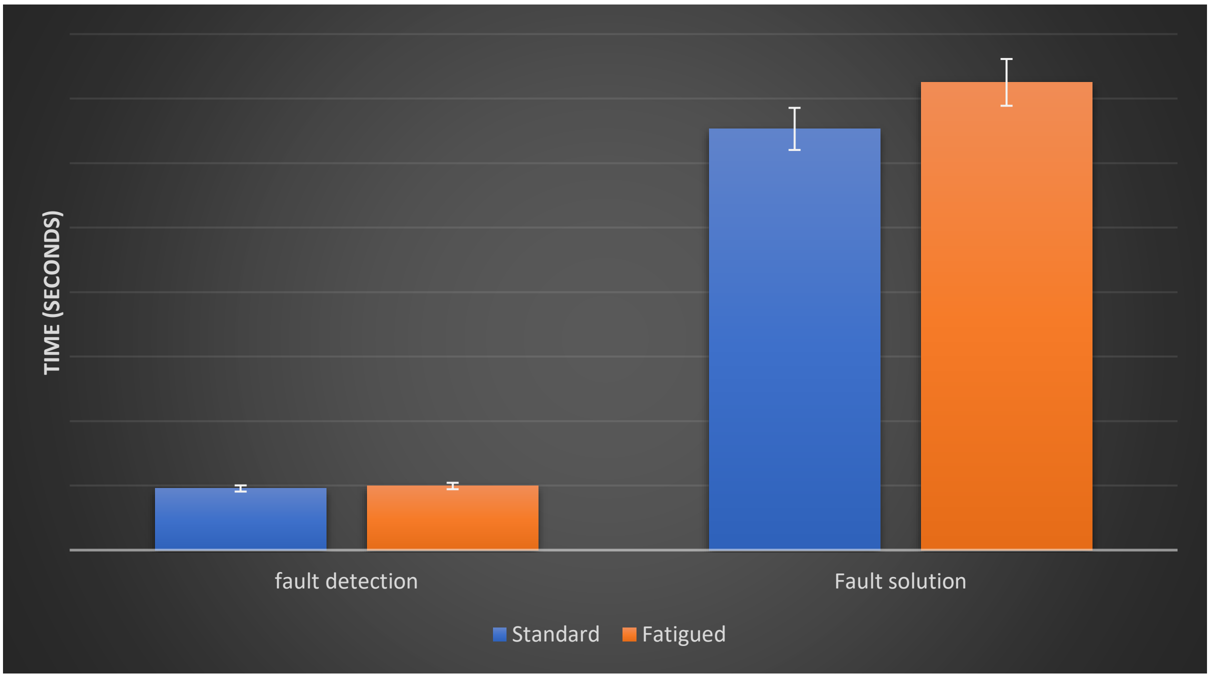


Figure 7.1 - Time taken to complete each workflow stage with respect to fatigue stressor.

7.3.2 Analysis of fnirs data

Average levels of oxygenated haemoglobin (HbO) were estimated using fNIRS for fault detection and fault solution workflow phases. Data from all channels were averaged into three regions of interest corresponding to left, medial and right regions of the prefrontal cortex. All HbO data were subsequently baselined using data gathered during the first phase of the workflow that lasted for 300s, i.e., baselined HbO = HbO during the task phase minus HbO during the 300s baselined period, hence positive HbO values indicates an increase above baseline levels.

Activation of the prefrontal cortex during the fault detection phase was explored via a 2 (fatigue/standard) x 3 (left, medial and right ROI) ANOVA. This analysis revealed no significant effect for fatigue [$F(1,18) = 1.97$, $p=.241$] or ROI [$F(1,18) = 0.93$, $p=.45$], and no significant interaction.

Activation of the prefrontal cortex during the fault solution stage was explored via a 2 (fatigue/standard) x 3 (left, medial and right ROI) ANOVA. The analysis revealed significant main effects for fatigue [$F(1,18) = 61.7$, $p<.01$, eta^2 = 0.913] and ROI [$F(1,18) = 31.7$, $p<.01$, eta^2 = 0.741]. The ANOVA also revealed a significant interaction between both factors [$F(1,18) = 36.92$, $p<.01$, eta^2 = 0.95]. The main effect for fatigue indicated that the mean HbO was significantly greater during the fatigue group (M = .019, s.e = .005) compared to the standard group (M = .009, s.e = .002). For ROI, the main effect revealed that the mean HbO at the medial ROI2 (M = .0008, s.e = .002) was significantly lower than either the left lateral ROI1 (M = 0.016, s.e = .002) or right lateral (M = 0.021, s.e = .002) ($p<.01$) – see Figure 7.2.

In order to explore the interaction, a number of post-hoc t-test were conducted. These tests revealed that the mean HbO was significantly greater with the fatigue group compared to that of a standard test as left lateral ROI1 [$t(18) = 3.91$, $p<.01$] and right lateral ROI3 [$t(18) = 6.39$, $p<.01$], but there was no significant effect of fatigue at the medial ROI2 region, see Figure 7.2 for descriptive statistics.

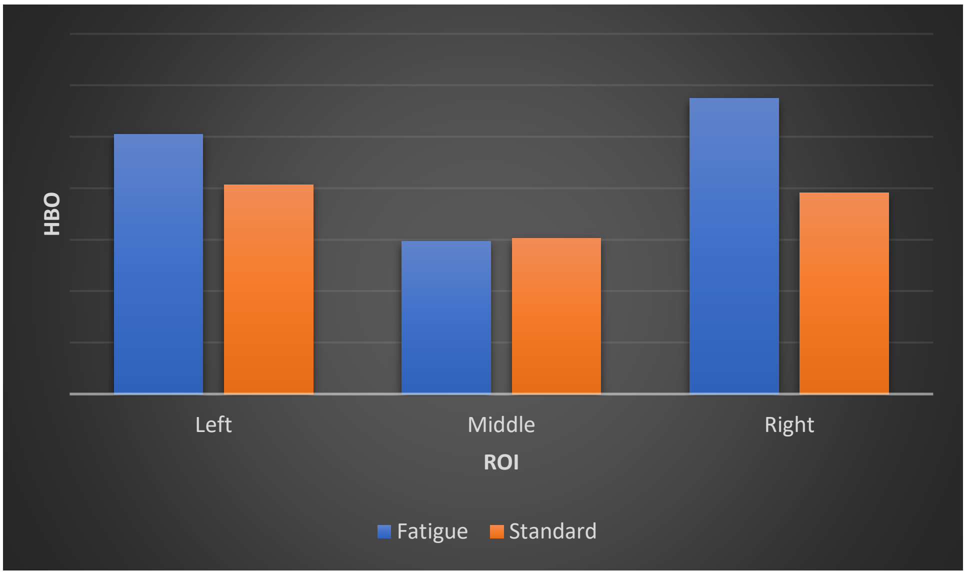


Figure 7.2 - Average HBO for ROI with respect to fatigue stressor.

7.4 Discussion

7.4.1 Time

A significant effect was found for fatigued candidates against those who were not fatigued. Fatigued participants on average took longer to complete each workflow stage. This shows that fatigue has an adverse effect on the time taken for participants to complete the task. This could be predicted as previous studies show that fatigue affects reaction and cognition [34]. Interesting is the size of the effect that fatigue has on time ($F = 71.41$). This is a fairly large effect. A large effect size again proves the proposed hypothesis that fatigue is a PSF that could have an adverse effect on human error. Post-hoc t-tests showed significant effects for all workflow stages. This would be explained by the hypothesis that fatigued participants on average take longer to complete each task. The most significant interaction for fatigued participants from t-tests for the workflow stages are found for the fault solution phase ($t = 9.15$).

A significant interaction was shown between fatigued candidates and specific workflow stage. This could be explained as fatigued participants are on average slower than those who participated in a standard test and the fault solution stage has more sub-tasks compared to other workflow stages. Post-hoc t-tests showed significant effects for all workflow stages. T-tests show that these effects get exponentially larger with time, for both those fatigued and standard groups.

Our results show that fatigue has a greater effect on the fault solution workflow stage. This could be explained due to the fault solution stage being larger and thus taking longer for the participant to complete. This has a 'knock on' effect when fatigue is added due to the participant having to perform more sub-tasks and thus for a longer time whilst in a fatigued state.

Many aspects of this study alone could be predicted. The relevance of this study comes when comparing PSF against one another in the following chapter. For example, it can easily be predicted that the fatigued candidates would take longer to complete each workflow stage but what cannot be predicted is how these candidates will compare to those who are distracted or have an increased workload.

7.4.2 fNIRS data

No significant effect was found for fatigue with respect to the fault detection workflow phase. This was not as predicted as the average of all fatigued participants showed a higher HBO overall. The reason no significant effect was found could have been due to the units of measurement being small or that not all fatigued participants had a higher HBO for the fault detection workflow stage when compared to the participants taking part in a standard test for the same workflow phase. However, a significant effect was found for the fault solution workflow phase. This suggests that fatigued participants only started to struggle compared to the standard test group when the sub-tasks became longer and involved more system navigation. This outcome goes against the findings of Bu et al [57] where fatigued participants showed higher levels of activation from the beginning of the task. The main difference being that in Bu et al study, the majority of the participants were elderly. Therefore, the age difference would have been a factor towards earlier higher HBO volumes due to fatigue. The fatigue investigation for our project was maritime based (done on a ship simulator) whereas, the other fatigue investigations were done across other sectors (aerospace [45] [51] [54], automotive [42] [70], National Rail Networks (NRN) [39] [58] and video Gaming [81]). However, due to the other studies also being 'hands-on' tasks conducted on a simulator they share many similarities with our investigation. Manipulation of the human factor – fatigue was done in a similar way to other studies in other sectors, where a monotonous visual of a readout is monitored by the participant for a long period of time (35 minutes for our study, between 20 minutes to 1 hour for others). The results from the 2 national rail network (NRN) studies showed that fatigue similarly to our study didn't show any significance until the latter stage of the test. However, the aerospace and automotive studies also showed fatigue to have a significant effect on operator performance from the start. This may be due to the impoverished nature of the simulation hardware for our study and NRN studies, in which no real consequences are shown when a task is failed. Whereas an A300 and A320 aircraft simulator (the hardware used in aerospace studies) and an automotive simulator with their hydraulic movements, has a very realistic feel and consequence to failure from the beginning of the test. Another explanation as previously mentioned could have been due to the units of measurement. However, the same units and software were used by Dehais [51] and Verdiere [54] for their study and fatigue was shown to have a significant effect.

Fatigued participants for our study had a higher HBO volume on average, specifically for the fault solution stage. It is shown in the results in section 7.2 that the fault solution stage had a significant difference when compared to the other stages. This resulted in showing fatigued participants to have an inflated average HBO for the whole task. In actual fact, the HBO volumes for all other workflow stages are similar for fatigued and standard test participants. By doing a Critical analysis of the workflow for this task it could be said that if the Fault detection stage were longer, then the fatigue element of the task would have been more prevalent. This is shown by the large difference in HBO volume between fatigued and standard test participants for the longer, fault solution stage. Similarly, this is also shown in the work done by Besikci et al, where fatigue was shown to be more prevalent in the longer tasks even if they were deemed to be easier [59]. Besikci concluded by stating that the fatigue element of their study was manipulated via a combination of verbal reasoning tasks over a long period of time (2 hours). There was then a short break (3-5 minutes) before starting their task. All participants showed similar results for the first 10-12 minutes of the task then differences occurred. By the end of the task (25 minutes) all participants had a significantly higher HBO volume when compared to those who had a standard test.

7.4.3 ROI

A significant effect was found for ROI. Similar to that of the 2 previous experiments in chapter 5 & 6 the left and right sides of the DLPFC showed significantly more HBO when compared to the middle region. The reasons for this are again the same as those in 6.3.3.

7.4.4 Validity check

The data gathered via fNIRS and ANOVA analysis was validated within the limits available. It is always preferable to validate work based on what has already been proven. However, due to the novelty of this project it is impossible to test the findings from fNIRS on a simulated environment against real life events at sea. Therefore, validity has been achieved via:

- 10 fold cross validation of fNIRS datasets from all 15 channels.
- Data was separated in epochs for cross validation to prevent ‘double-dipping’.
- Outcomes are checked against R-studio, MatLab and Python software platforms.
- A manual ‘step by step’ linear regression with ANOVA was applied using excel to check the validity of each workflow stage against PSF.

7.5 Concluding remarks

The experiment was changed to incorporate a 2100s first baseline compared to a 300s baseline in previous tests. This was done to induce a fatigue stressor. The remaining workflow stages involved tasks identical to a standard test.

From the ANOVA, a significant effect was found for fatigue with respect to time taken and HBO. This resulted from higher levels of neurophysiological activation of the DLPFC for fatigued participants. A significant effect was also found for workflow stage with respect to time taken and HBO. Higher levels of neurophysiological activation and time taken were found for the fault solution workflow phase when compared to all other phases.

The ROIs with the most significance were shown to be the left and right (channels 1-5 & 11-15) sides of the DLPFC compared to the middle region (channels 6-10).

Chapter 8. Comparison of 3 main PSFs; Fatigue, Workload & Distraction and Combined PSFs.

8.1 Introductory remarks

This chapter is a continuation of the previous three chapters. Furthermore, this chapter will look at the PSFs in an ANOVA study against one another as opposed to against a standard test.

No significant effects were found from an investigation of fatigue x distraction stressors. Therefore, the results of that study are omitted and instead workload x fatigue and workload x distraction are investigated.

This chapter is broken up into two main sections; workload x fatigue and workload x distraction. Both sections contain time and HBO results followed by a discussion of findings. The final section is a conclusion based on the findings from both ANOVA studies.

8.2 The Effect of Workload x Fatigue.

The data from the study were analysed via ANOVA procedures using SPSS v.26. Outliers were identified as any value that deviated more than 3 standard deviations from the cell mean and were omitted from ANOVA testing.

8.2.1 Times to complete workflow phases

8.2.1.1 Fault Detection Phase

The time taken to complete each phase of the workflow was subjected to a 2 (high fatigue/low fatigue) x 2 (high workload/low workload) x 2 (fault detection/fault solution) ANOVA. This model revealed no significant interaction between both stressors workload*fatigue [$F(3,34) = 2.674$, $p=.111$, $eta^2 = .069$] or significant main effects for workload [$F(1,36) = 1.423$, $p=.241$, $eta^2 = .038$] but revealed significant main effects for fatigue [$F(1,36) = 42.962$, $p<.01$, $eta^2 = .544$] with respect to the fault detection workflow phase. Fatigued participants (M=83.205s, s.e.=.528) had higher times for the fault detection workflow phase when compared to high workload (M=80.315s, s.e.=.528).

8.2.1.2 Fault Solution Phase

This model revealed significant main effects for workload [$F(1,36) = 1251.85$, $p<.01$, $eta^2 = .972$] and fatigue [$F(1,36) = 43.197$, $p<.01$, $eta^2 = .545$] but no significant interactions were found for workload*fatigue [$F(3,34) = .169$, $p = .683$] with respect to the fault solution workflow phase. High workload (M=474.5, s.e =2.45) resulted in higher times for the fault solution workflow stage when compared with fatigued participants (M=424.585, s.e.=2.45).

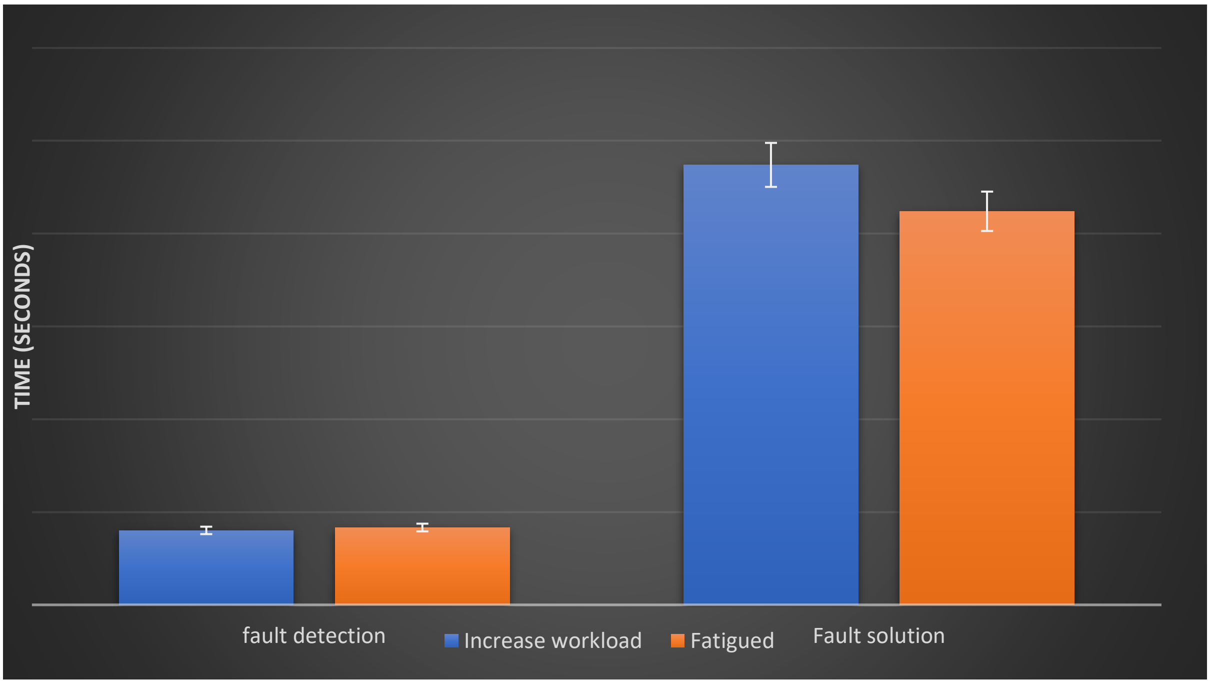


Figure 8.1 - Time taken to complete each workflow phase with respect to fatigue and workload stressors.

8.2.2 Analysis of fNIRS data

Average levels of oxygenated haemoglobin (HbO) were estimated using fNIRS for fault detection and fault solution workflow phases. Data from all channels were averaged into three regions of interest corresponding to left, medial and right regions of the prefrontal cortex. All HbO data were subsequently baselined using data gathered during the first phase of the workflow that lasted for 300s, i.e., baselined HbO = HbO during the task phase minus HbO during the 300s baselined period, hence positive HbO values indicate an increase above baseline levels.

Activation of the prefrontal cortex during the Fault Detection phase was explored via a 2 (high fatigue/low fatigue) x 2 (high workload/low workload) x 3 (Left, medial and right ROI) ANOVA. The analysis revealed a significant effect for workload [$F(3,34) = 2.47$, $p=.08$, eta^2 = .179] but no significant effect for fatigue [$F(3,34) = .610$, $p=.613$] or any significant interaction between both factors workload*fatigue [F3,34) = .628, $p=.602$]. The main effect for workload indicated that mean HbO was significantly greater during the high workload condition (M=.005, s.e.=.001) when compared to fatigue (M=0.0037, s.e.=.001).

Activation of the prefrontal cortex during the fault solution phase was explored via a 2 (low fatigue/high fatigue) x 2 (low workload/high workload) x 3 (left, medial and right ROI) ANOVA. This analysis revealed significant main effects for workload [$F(3,34) = 228.251$, p<.01, eta^2 = .953] but no significant main effects were found for fatigue [F(3,34) = 1.730, p=.179]. [M=0.035, s.e.=.002 (averaged for all ROI)].

A number of post-hoc t-tests were conducted. These tests revealed that the mean HbO was significantly greater during high workload compared to low workload at left lateral ROI1 [t(36) = 6.56, p<.01], medial ROI2 [t(36) = 3, p=.005] and right lateral ROI3 [t(36) = 25.67] as shown in Figure 8.2.

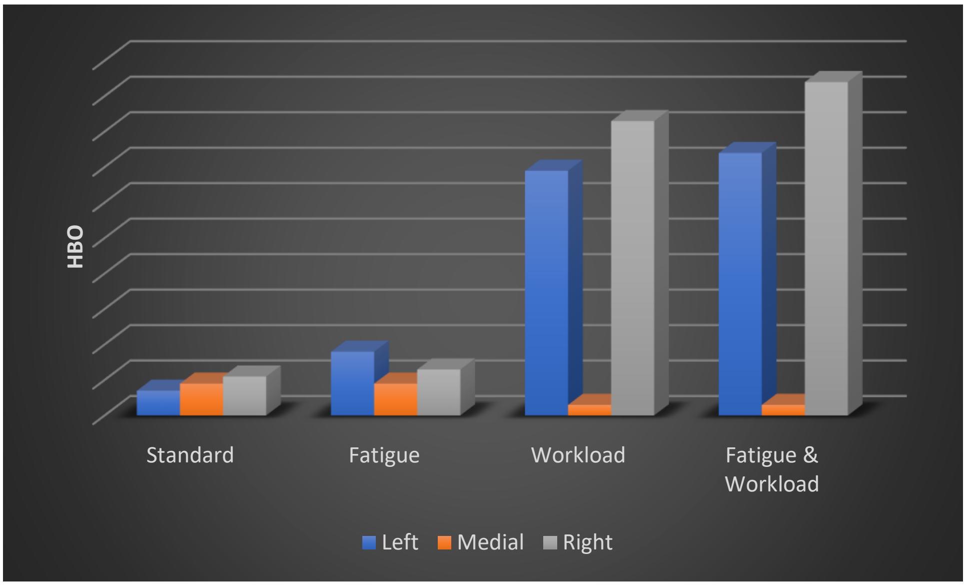


Figure 8.2 - Average HBO of stressors with respect to ROI.

8.3 Discussion

8.3.1 Time

8.3.1.1 The Fault Detection Phase

The study found no significant interaction between increased workflow and fatigue or significant main effects for workload. However, a significant main effect was found for fatigue. This can be explained as the fault detection stage for the increased workload group involved additional ballast tanks to monitor, but it was noted by the examiner that the participants treated the monitoring and detection stages similar to that of the standard workload group. For example, the participants in the increased workload group didn't complete the additional checks as they should have. Instead, they simply followed the procedure for the standard workload group (1 ballast tank). This resulted in this workflow stage showing little increase in time.

8.3.1.2 The Fault Solution Phase

The analysis found significant main effects for fatigue and workload. This can be explained as the fault solution stage for the increased workload group involved significantly more sub-tasks compared to a standard workload test. The fault solution stage is also a much longer workflow phase. This results in already fatigued participants becoming more fatigued due to the extended task time. Again, no significant interaction was found. This is because the low fatigue group are faster to complete the task when compared to the high fatigue group despite the level of workload.

8.3.2 fNIRS data

8.3.2.1 The fault Detection phase

The analysis found no significant interaction between fatigue and workload or significant main effects for fatigue. However, significant main effects were found for workload for the fault detection workflow phase. This tells us that even though no main effect was found for workload with respect to time, the participants were still experiencing increased levels of activation due to the additional ballast tanks. Fatigue was the opposite to this as a significant main effect was found with respect to time but not for level of activation. This tells us that fatigue affects time more than mental workload for the fault detection workflow stage.

8.3.2.2 The Fault Solution Phase

The fault solution workflow stage showed no significant interaction between fatigue and workload or significant main effects for fatigue. However, a significant main effect was found for workload. No significant interaction can be explained as, for the left and right side of the PFC low fatigue participants consistently had lower levels of activation despite the level of workload. The significant main effect found for workload and not for fatigue tells us that participants had an increased mental workload for an increased workload task but not for a fatigued test. This can be interpreted as, participants found increased workload more difficult than high levels of fatigue.

8.3.3 ROI

Region of interest showed some interesting results for this study as there was no significant interaction shown for the left and right regions but a very slight significant interaction was found for the medial region. Higher levels of activation were found for the high fatigue group for low workload and the low fatigue group for an increased workload. This can only be explained by stating that the medial region of the PFC consistently showed anomalous results and lower levels of HBO detection throughout the study. Therefore, it has been decided to omit any findings from the medial region due to inaccuracies in the analysis data.

8.4 The effect of Workload X Distraction

The data from the study were analysed via ANOVA procedures using SPSS v.26. Outliers were identified as any value that deviated more than 3 standard deviations from the cell mean and were omitted from ANOVA testing.

8.4.1 Times to complete workflow phases

8.4.1.1 Fault Detection Phase

The time taken to complete each phase of the workflow was subjected to a 2 (distraction/no distraction) x 2 (high workload/low workload) x 2 (fault detection/fault solution) ANOVA. This model revealed no significant interaction between both stressors workload & distraction or significant main effects for workload but revealed significant main effects for distraction [$F(1,36) = 86.142$, $p<.01$, $eta^2 = .750$] with respect to the fault detection workflow phase. Distracted participants ($M=87.205s$, $s.e.=.568$) had higher times for the fault detection workflow phase when compared to high workload ($M=80.315s$, $s.e.=.528$) as shown in figure 8.3.

8.4.1.2 Fault Solution Phase

This model revealed significant main effects for workload [$F(1,36) = 1251.85$, $p<.01$, $eta^2 = .972$] and distraction [$F(1,36) = 298.197$, $p<.01$, $eta^2 = .945$] but no significant interactions were found for workload*distraction with respect to the fault solution workflow phase. High workload ($M=474.5$, $s.e =2.45$) resulted in higher times for the fault solution workflow stage when compared with distracted participants ($M=446.915$, $s.e.=2.45$) as shown in figure 8.3.

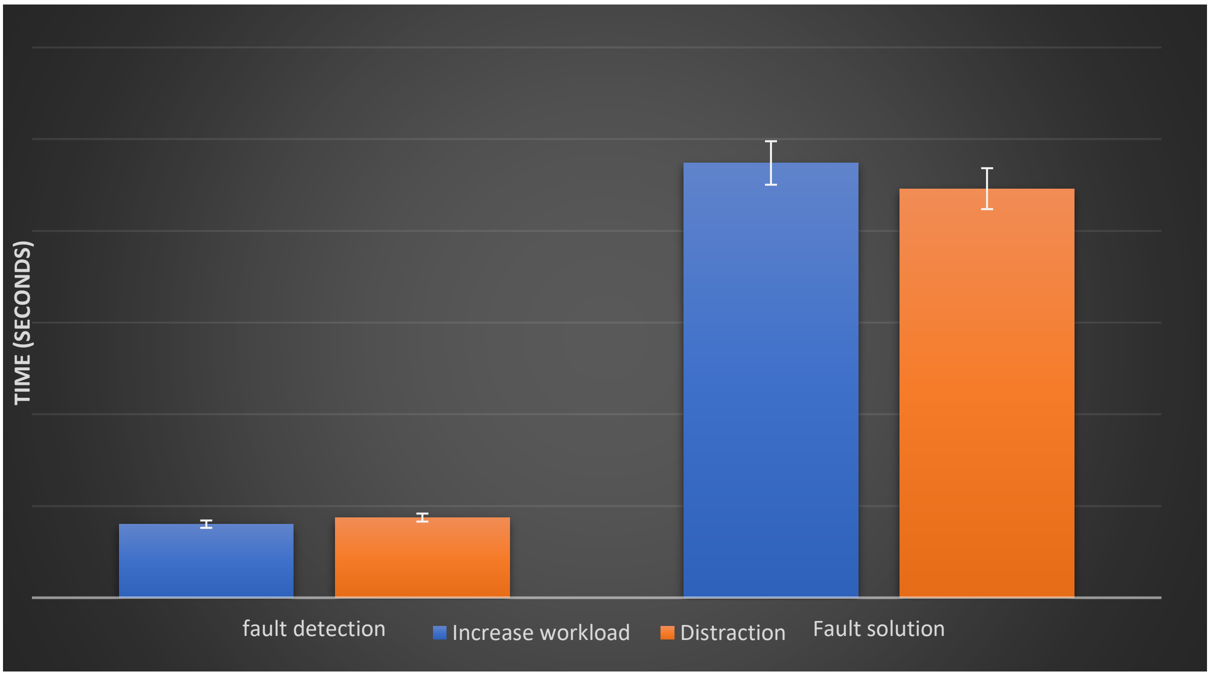


Figure 8.3 - Time taken to complete each workflow phase with respect to distraction and workload stressors.

8.4.2 Analysis of fNIRS data

Average levels of oxygenated haemoglobin (HbO) were estimated using fNIRS for fault detection and fault solution workflow phases. Data from all channels were averaged into three regions of interest corresponding to left, medial and right regions of the prefrontal cortex. All HbO data were subsequently baselined using data gathered during the first phase of the workflow that lasted for 300s, i.e., baselined HbO = HbO during the task phase minus HbO during the 300s baselined period, hence positive HbO values indicates an increase above baseline levels.

Activation of the prefrontal cortex during the fault detection phase was explored via a 2 (distraction/no distraction) x 2 (high workload/low workload) x 3 (left, medial and right ROI) ANOVA. The analysis revealed a significant main effect for workload [F(3,34) = 2.47, p=.08, eta^2 = .179] and for distraction [F(3,34) = 26.0, p=.013, eta2=.744]. No significant interaction was found between both factors workload*distraction [F3,34) = 1.281, p=.322]. The main effect for workload indicated that mean HbO was significantly greater during the high workload condition (M=.005, s.e.=.001) when compared to low workload (M=0.0017, s.e.=.001). The main effect for distraction indicates that HBO was significantly greater during the distracted condition (M=.031, s.e = .002) when compared to not distracted (M=0.0017, s.e = .001).

Activation of the prefrontal cortex during the fault solution phase was explored via a 2 (distracted/not distracted) x 2 (low workload/high workload) x 3 (left, medial and right ROI) ANOVA. This analysis revealed significant main effects for workload [$F(3,34) = 228.251$, $p<.01$, $\eta^2 = .953$] and significant main effects were found for distraction [$F(3,34) = 104.831$, $p<.01$, $\eta^2=.910$]. The significant main effect for workload indicated that the mean HbO was significantly greater during the high workload condition (M=0.064, s.e.=.002) and distracted condition [(M = 0.041, s.e.=.002) (averaged for all ROI)] when compared to the low workload condition [(M=0.035, s.e.=.002) (averaged for all ROI)] and not distracted condition (M=0.035, s.e.=.002).

A number of post-hoc t-tests were conducted. These tests revealed that the mean HbO was significantly greater during high workload compared to low workload at left lateral ROI1 [$t(36) = 6.56$, $p<.01$], medial ROI2 [$t(36) = 3$, $p=.005$] and right lateral ROI3 [$t(36) = 25.67$] and significantly greater during distraction compared to not distracted for the left lateral ROI1 [$t(36) = 3.96$] and right lateral ROI3 [$t(36) = 14.18$] as shown in Figure 8.4.

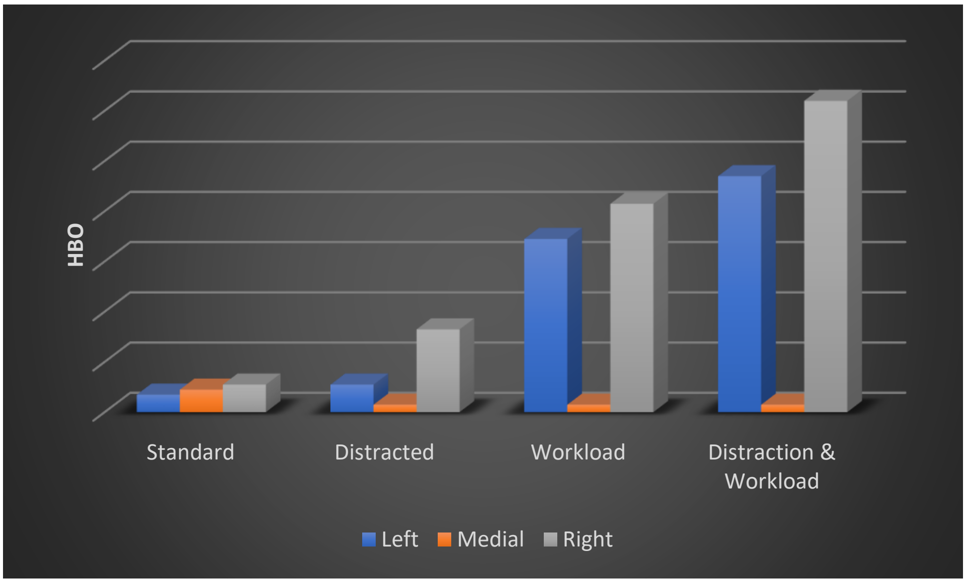


Figure 8.4 - Average HBO of stressors with respect to ROI.

8.5 Discussion

8.5.1 Time

8.5.1.1 The Fault Detection Phase

The study found no significant interaction between increased workflow and distraction or significant main effects for workload. However, a significant main effect was found for distraction. This can be explained as the fault detection stage for the distracted group involved distraction questions which takes additional time to complete.

8.5.1.2 The Fault Solution Phase

The analysis found significant main effects for distraction and workload. This can be explained as the fault solution stage for the distracted group again, involved distraction questions compared to a not distracted test which contained no questions. Again, no significant interaction was found. This is because the undistracted group are faster to complete the task when compare to the distracted group despite the level of workload.

8.5.2 fNIRS data

8.5.2.1 The Fault Detection Phase

The analysis found no significant interaction between distraction and workload. However, significant main effects were found for workload and distraction. This tells us that participants were still experiencing increased levels of activation due to the addition of distraction questions.

8.5.2.2 The Fault Solution Phase

The fault solution workflow stage showed no significant interaction between fatigue and workload. Significant main effect was found for workload and distraction. No significant interaction can be explained as, for the left, medial and right side of the PFC for not distracted participants consistently had lower levels of activation despite the level of workload. The significant main effect found for distraction tells us that participants had an increased mental workload for a distracted task when compared to no distraction.

8.6 Conclusion

The investigations showed that the average mental workload for participants in order from highest to lowest are as follows:

1. Distracted with an Increased workload
2. Fatigued with an Increased workload
3. Increased workload
4. Distracted
5. Fatigued

This tells us that the group with the greatest risk of human error is the distracted with an increased workload group.

The distraction aspect of the exercise was manipulated by asking the participants questions routinely throughout the workflow as stated above in chapter 5. This would be part of the daily duties of an engineer in a real-life scenario. This required participants to stop what they were doing, navigate to the liquid cargo screen, give the answer to the question and then continue with what they were previously doing. This distraction would result in the participants taking extra time on each task and having to remember what they were doing as soon as they had answered the distraction question. This extra time and additional mental workload required to remember the workflow position that the participants were up to before the distraction, would explain why distraction was significant in these tests.

The increased workload element of the study was manipulated by increasing the number of ballast tanks to be filled from 1 tank to 6 tanks. This would be part of the duties of an engineer within a ship engine room in a real-life scenario. This resulted in the participants spending a lot more time on the fault solution workflow stage than for all other PSF. This also required participants to spend more time and thought on the re-routing of the sea water line through the working ballast pump and then into the additional ballast water tanks. This increased time and mental workload would explain why the increased workload group had the highest risk of human error. Increased time and mental workload have been found to increase the risk of human error [18].

The fatigue aspect of the exercise was manipulated by increasing the first baseline monitoring exercise to 35 minutes (2100s) instead of 5 minutes (300s). Long monitoring tasks would be a standard daily duty of a seafarer working in a ship engine room. This resulted in the attention of the participants moving from the monitoring task, to looking around the simulator room and to checking other system screens. The effects of fatigue were only seen when the participant was working on the fault solution stage of the workflow. This would suggest that fatigue affects participants undertaking longer, and more difficult tasks compared to the shorter, easier tasks. Comparing fatigued participants to those taking on a standard test showed us that fatigued participants have an increased risk of human error.

8.7 Concluding remarks

The workload x fatigue study showed significant main effects for fatigue and workload for the fault solution workflow phase but no significant interaction was found with respect to time. The ANOVA showed a significant effect for workload but no effect for fatigue with respect to HBO for the fault detection workflow stage. The ANOVA again showed significant main effects for workload but not for fatigue for the fault solution workflow stage with respect to HBO. This reveals that the main effect for workload was significantly greater during the high workload condition when compared to the fatigue condition.

The workload x distraction study showed significant main effects for workload and distraction with respect to time for the fault solution workflow phase. Significant main effects were found for workload and distraction for all workflow phases with respect to HBO.

Chapter 9. The Development and Implementation of a Scientific Human Error Model

9.1 Introductory remarks

This chapter is used to describe the development and implementation of a scientific human error model. It is used to model the resulting data produced in experimental chapters; 5, 6, 7 and 8.

The first section defines how the fNIRS data is pre-processed, oxygenation features extracted, how classification and cross-validation is performed and how the data is statistically evaluated.

The following section is the result section. This shows the classification performance of each individual PSF against a standard test and against other PSFs with respect to the oxygenation features.

The final section is a discussion of the findings from the results section.

9.2 Data Classification

9.2.1 Data pre-processing

NIRS-analysis aids in converting the raw fNIRS data to optical densities. In this process, the NIRS Star software was used to remove artefacts and apply a band pass filter to the raw data. MatLab was used to implement a wavelet interpolation method for artefact correction as this method showed the greatest signal to noise ratio compared to other artefact removal methods available [54]. A high pass filter (cutoff 0.01Hz – order 3) and a low pass filter (cutoff: 0.5 Hz – order 5) was applied for the band pass filtering stage.

The artefact-free, filtered data was then converted into oxygenated [HbO] and de-oxygenated [HbR] haemoglobin concentrations and extracted into an Excel spreadsheet. The CBSI formula was then applied to the data as described in 4.3.1.

The processed FNIRS data was the extracted from Excel to R studio. R studio was used to write the code needed to separate the data into epochs. The HBO data for the entire workflow for 1 participant consisted of an average of 82,000 frames (~546s). As the task duration (~484s + or – 15, ~546 + or – 33s, ~445 + or – 6.5 for distracted, increased workload and fatigue respectively) slightly differed for each participant with respect to their group, the HBO data was divided into epochs. The fixed number of extracted epochs was based on the shortest task duration, resulting in 52 (~8s) epochs for fatigued and distracted groups and 57 (~9.5s) epochs for the increased workload group. Each epoch was processed independently as this enabled us to show the exact parts of the workflow that had the highest levels of activation.

9.2.1.1 Oxygenation measures

Oxygenation measures were computed using oxygenated [HbO] haemoglobin signals on each epoch. The value x represents the oxygenated haemoglobin [HbO] signal for one epoch (52-57 samples) and one optode. The Six oxygenation measures computed are as follows: Average, Variance, Area Under the Curve, Skewness, Slope and Kurtosis.

The Average, Variance, Skewness and Kurtosis were calculated using the following formula:

$$Average(x) = E(x) \qquad Variance\ (x) = E[(x - E(x))^2] \tag{5}$$

$$Skewness\ (x) = \frac{E[(x-E(x))^3]}{(E[(x-E(x))^2]^{\frac{3}{2}}} \qquad Kurtosis\ (x) = \frac{E[(x-E(x))^4]}{(E[(x-E(x))^2]^2)} \tag{6}$$

The Area Under the Curve (AUC) was calculated using the sum of the absolute values of the signal:

$$Area\ Under\ the\ Curve\ = \sum|x| \tag{7}$$

The slope was calculated using the least-squared linear regression with the polyfit MatLAb function.

9.2.2 Feature Extraction

9.2.2.1 Region of interest

To reduce the amount of data, the 15 channels were condensed to 3 regions of interest (ROI); The left side of the dorsal lateral pre-frontal cortex (DLPFC), the middle DLPFC and the right side DLPFC. The oxygenation features were extracted by averaging all the oxygenation features from the 15 channels included in the 3 regions. This gave us 6 measures for each oxygenation feature per subject and per epoch.

9.2.3 Classification and Cross-Validation

For this study a Linear Discriminant Analysis (LDA) with regularization of the empirical covariance matrix by shrinkage (the shrinkage method) was used as this method has proved to be robust for use with Brain-computer interfaces (BCI) and passive BCI (pBCI) application [45] [73] and also with fNIRS [51] [72].

The Shrinkage Method

The shrinkage method is used to penalize the inclusion of less informative predictors resulting in increased classification accuracy. Based on the prediction equation below in 9.2.3.1, shrinkage finds coefficients $\hat{\beta}$ that minimize the sum of squared residuals (RSS) with the goal of finding coefficients that make predictions as close as possible to the observed responses (make residuals as low as possible).

$$minimize \qquad \boldsymbol{RSS} = \sum_{i=1}^{n}(yi - \hat{y}i)^2\ Where\ (yi - \hat{y}i) = residual \tag{8}$$

Each workplace factor was separated via an intra-subject binary classification (I.E distraction 1, standard test 0). Each participant had to perform 1 of the 6 different workflow tasks (standard, distracted, increased workload, increased workload * distraction, fatigued, fatigued * increased workload). Data was processed to obtain 52 x 8s-9.5s epochs per task for each participant. There were 10 participants for each of the 6 different workflow tasks (10 x 52s epochs = 520 samples per workflow, 3120 samples in total). The prediction of performance for this model was assessed by using a stratified cross validation. As this is deemed a good trade-off between variance estimation and bias [45]. The classifier was trained using 8 of the 10 participants (52 x 8 = 416 samples) and tried to predict the last 2 participants (one of each workplace factor, i.e. 2 x 52 = 104 samples). This method was applied for each of the 3 workplace factors (distraction, increased workload and fatigue) using the same intra-subject binary classification (distraction, increased workload, fatigue [1] ~ standard test [0]) and the average performance was kept. The same method replicated using combinations of workplace factors (distraction*increased workload, fatigue*increased workload [1] ~ standard test [0] and again, the average performance was kept. Regarding the features, 2 different types of comparisons were done. The first was evaluating each of the 6 features separately. Secondly, features were combined to evaluate their potential. They were merged into couples (2 x 2) and the classification performance of each couple was evaluated [54].

9.2.3.1 Further Validation

The R studio and MatLab LDA was further validated by manually inputting code step by step into Excel. This was done on 10 randomly selected participants out of 60, throughout all 3,120 samples. This was done as follows:

- Firstly, the data analysis software tool was installed onto Excel.
- 52 x 8s epochs are manually grouped into the binary categories (workplace factor [distraction, fatigue, increased workload] 1, standard test 0).
- A regression analysis is run using the analysis tool. Input y is the binary group and input x is the HbO values from all 15 channels.
- Using the coefficients output from the regression analysis a 'normalised prediction' can be made. This is done by the sum of the average HbO from each epoch from each channel multiplied by the coefficient of the corresponding channel.

$$\hat{y}_i = \hat{\beta}_{intercept} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \dots\dots. + \hat{\beta}_{15} x_{15} \tag{9}$$

$$where\ \hat{y}_i = prediction, \hat{\beta} = coefficients, x = HbO\ output$$

- The normalised predictions are averaged for binary group 1 and 0.
- A 'cut-off' value can then be obtained by multiplying the number of samples in binary group 0 with the normalised prediction of group 0, the same for group 1, then dividing the sum of the 2 multiplications by the total number of samples in group 0 and 1.

$$Cutt - Off = \frac{(y_{average} of\ 0 * Number\ of\ 0) + (y_{average} of\ \ 1 * Number\ of\ 1)}{Number\ of\ 0 + Number\ of\ 1} \quad (10)$$

$$Where\ y_{average} = average\ prediction\ from\ each\ row$$

- If the normalised prediction is smaller than the cut-off then group 0 is used, otherwise group 1.
- If the Group value given from the bullet point above matches the binary group of the sample, then we have a correct prediction.

9.2.4 Statistical Assessment

9.2.4.1 Subjective workload comparison

Paired sample t-tests were performed in order to compare the average mental workload obtained in HbO for the workplace factors, workflow stage and region of interest amongst participants.

9.2.4.2 Classification performance significance

For a problem like ours involving 2-classes, the theoretical chance level for classification is 100/2=50%. However, this is only correct for an infinite sample size. To assess our classifier's significance (decoding error). The classification error was modelled using a binomial cumulative distribution, as this has shown significance on other studies [43]. The binomial cumulative distribution was calculated as follows:

$$\mathrm{P(Z)} = \sum_{i=z}^{n} \binom{n}{k} \mathrm{x} \left(\frac{1}{c}\right)^{i} \times \left(\frac{c-1}{c}\right)^{n-1} \quad (11)$$

Where P is the is the probability that the correct class is predicted by at least Z times. N is the sample number. C is the number of classes.

The performance classification was assessed by repeating the stratified cross validation for all workplace combinations and then averaging. As previously stated, our classifier was trained using 8/10 participants (416 samples) and tested on the last 104 samples. Using the cumulative binomial distribution, it sets the 5% significance classification threshold at 56.02% chance.

9.2.4.3 Classification performance comparison

In order to compare the classification performance for each feature, a repeated measure ANOVA was used considering FEATURES (or COMBINED FEATURES) within factors.

9.3 Results

The data used to conduct the LDA was taken from the fault solution workflow stage. This was done due to the fault solution stage having the most significance as outlined by the ANOVA results above in chapter 5,6,7 and 8. The stressors are first analysed individually against a standard test, they are then analysed against one another.

9.3.1 Classification of Individual Performance shaping Factors and Individual Features with respect to Chromophore.

9.3.1.1 Distracted participants

Figure 9.1 depicts the classification performance of distracted participants for each of the 6 features using HBO signals. So that we can compare the classification performance with features, a repeated measure ANOVA study was performed. A significant effect was found for feature type on the classification performance for distracted participants [$F(5,234) = 6.96$, $p < 0.01$]. Pairwise comparisons showed significant differences between features for the HBO signal. Specifically, Area Under the Curve and Variance had a significantly better performance than Kurtosis, Slope, Skewness and Average. There was no significant effect shown between Kurtosis, Average, Skewness and Slope. Pairwise comparisons also showed a significant effect for Area Under the Curve (78.25%) and Variance (77.38%) when compared to all other features (=<76.4%).

Every oxygenation feature resulted in an average classification performance above change (>56.09%). Moreover, the minimum classification performance is kurtosis (72.8%) which equates to 16.7% greater than the chance level.

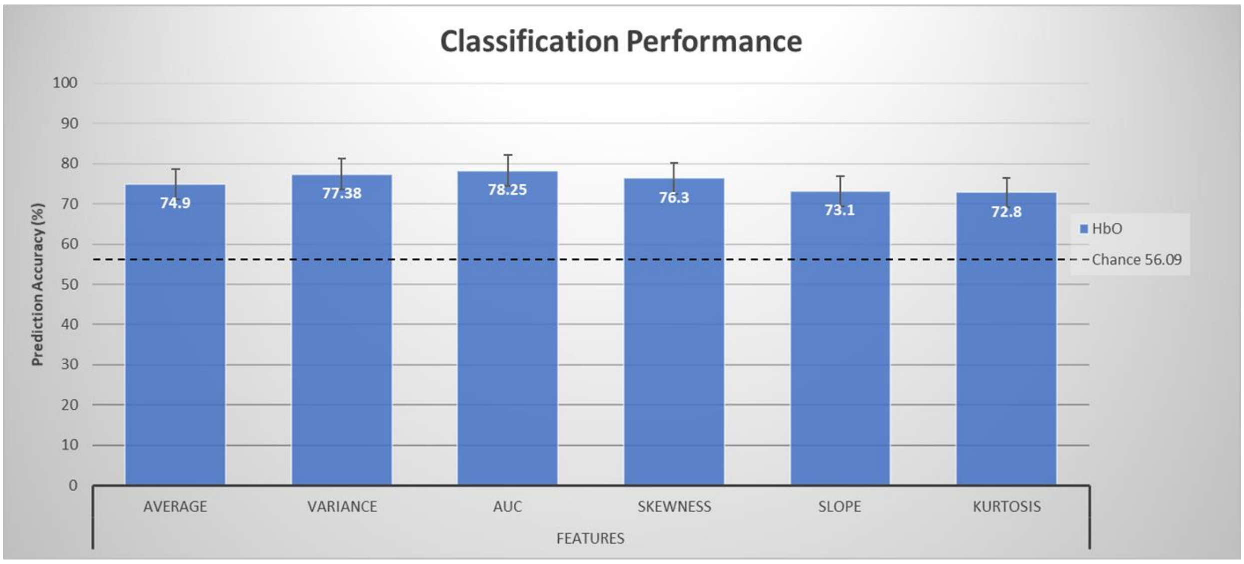


Figure 9.1 - Classification Performance model of distracted participants.

Table 9.1 below shows the classification performance for each feature for each participant (S1-20). This allows us to see specifically how each participant performed (classification performance) with respect to oxygenation features thus, how human performance is affected [45] as a result of distraction. There is a trend shown; AUC, shows the highest classification performance percentage on average for all participants.

Table 9.1 – Distracted versus a standard test classification performance for each epoch with respect to feature type.

Distraction vs Standard						
Participants	**Average**	**Variance**	**AUC**	**Skewness**	**Slope**	**Kurtosis**
S1	75	77	80	76	74	73
S2	73	75	80	77	75	74
S3	77	75	79	77	76	73
S4	74	77	80	76	73	75
S5	76	76	77	75	73	72
S6	75	77	78	77	74	72
S7	74	79	78	76	76	72
S8	77	75	80	77	75	74
S9	76	75	79	75	74	75
S10	75	75	80	77	73	74
S11	78	77	79	76	73	73
S12	74	78	79	75	72	73
S13	74	77	79	77	74	77
S14	74	79	77	75	74	75
S15	75	76	78	75	75	73
S16	75	75	79	76	75	74
S17	78	77	80	75	73	72
S18	74	78	79	75	73	72
S19	75	76	80	77	74	73
S20	75	77	78	74	73	72

9.3.1.2 Fatigued Participants

Figure 9.2 shows the classification performance of fatigued participants for each of the 6 features, using HBO signals. In order to compare the classification performance with features, a repeated measure ANOVA study was performed. A significant effect was found for feature type on the classification performance for fatigued participants [$F(5, 234) = 8.11$, $p < 0.01$]. Pairwise comparisons showed significant differences between features for the HBO signal. Specifically, Area Under the Curve (77.29%) had a slightly better performance than Variance (76.26%) and a significantly better performance than Kurtosis (70.45%), Slope (71.94%), Skewness (74%) and Average (71.9%).

Every oxygenation feature resulted in an average classification performance above chance (>56.09%).

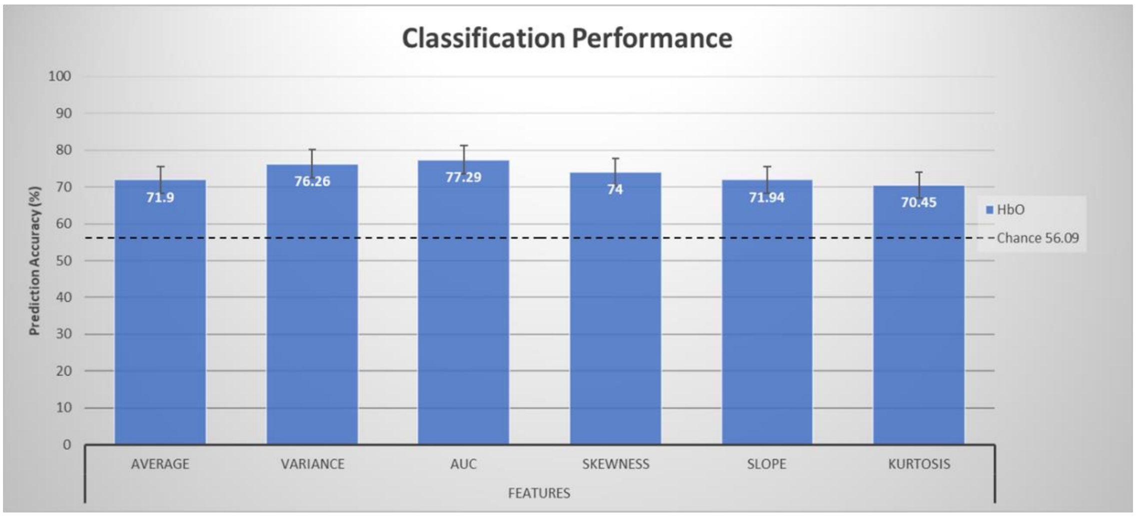


Figure 9.2 - Classification performance of Fatigued participants

Table 9.2 shows the classification performance of each feature for each participant. This allows us to see the influence of fatigue against a standard test by the resultant classification performance. This shows the oxygenation features and their corresponding classification performance. The higher the classification performance, the greater the influence of fatigue [45]. Again, there is a trend between classification performance and the oxygenation feature AUC.

Table 9.2 – Fatigued versus a standard test classification performance for each epoch with respect to feature type.

Fatique vs Standard						
Participants	**Average**	**Variance**	**AUC**	**Skewness**	**Slope**	**Kurtosis**
S1	74	76	77	74	73	70
S2	75	77	75	73	72	71
S3	73	77	74	75	72	70
S4	73	76	77	75	72	71
S5	73	78	77	76	73	70
S6	75	76	78	74	74	69
S7	74	75	79	74	73	71
S8	74	76	77	74	72	72
S9	73	76	76	75	73	72
S10	73	76	79	76	74	72
S11	74	77	78	76	72	70
S12	73	77	78	75	71	70
S13	75	75	77	74	72	71
S14	74	78	78	75	72	71
S15	74	76	76	73	72	72
S16	74	75	79	73	73	69
S17	73	74	78	72	74	69
S18	73	77	78	73	74	71
S19	73	78	77	72	73	71
S20	74	77	77	73	74	70

9.3.1.3 Increased Workload Participants

Figure 9.3 depicts the classification performance of the increased workload participants for each of 6 oxygenation features. In order to compare the classification performance with features, a repeated measure ANOVA study was performed. A significant effect was found for feature type on classification performance for Increased Workload participants [$F(5,234) = 4.17$, $p<0.01$]. Pairwise comparisons showed significant differences between oxygenation features. Area Under the Curve (85.15%) and Variance (83.96%) had the best performance compared to Kurtosis (79.02%), Slope (80%), Skewness (82.9%) and Average (80.1%).

All oxygenation features had a significantly higher classification performance than the chance value (56.09%). Additionally, the lowest classification performance is Kurtosis (79.2%), which is 23.11% above the chance value.

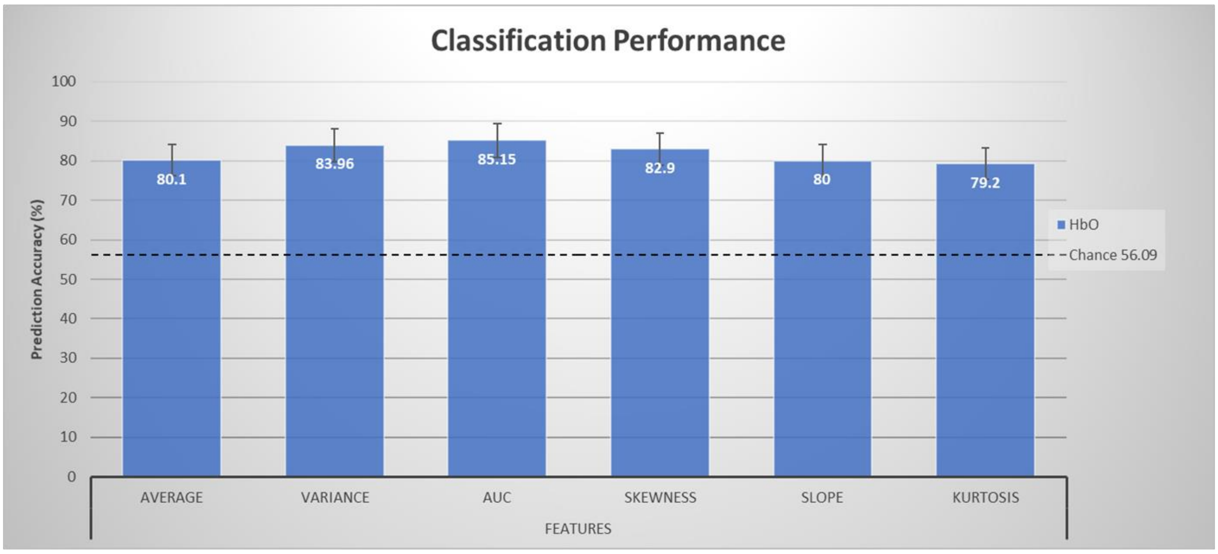


Figure 9.3 - Classification performance of Increased workload participants.

Table 9.3 shows the classification performance of the 6 oxygenation features for each participant throughout the task. Again, the classification performance features tell us the influence of increased workload against those participating in a standard test. [45].

Table 9.3 – Increased workload vs standard workload classification performance for each epoch with respect to feature type.

Increased Workload vs Standard						
Participants	**Average**	**Variance**	**AUC**	**Skewness**	**Slope**	**Kurtosis**
S1	81	85	84	83	81	79
S2	80	84	86	82	80	79
S3	80	85	87	83	81	79
S4	81	84	85	84	80	80
S5	80	84	87	83	80	80
S6	82	83	84	82	82	79
S7	82	84	86	84	81	80
S8	80	85	83	84	82	81
S9	80	85	87	83	80	81
S10	79	84	85	85	81	80
S11	80	85	86	83	82	79
S12	78	83	87	82	82	80
S13	80	85	86	82	81	79
S14	82	86	85	83	80	80
S15	81	83	85	85	80	81
S16	81	83	85	82	80	80
S17	81	84	87	82	81	79
S18	80	82	86	83	80	80
S19	82	83	88	84	81	81
S20	80	84	84	83	80	79

9.3.2 Classification Performance of PSF against PSF with respect to individual oxygenation features and chromophore.

9.3.2.1 Distraction against Fatigue

Figure 9.4 shows the classification performance of distraction against fatigue with respect to individual oxygenation features. As seen below the only classification performance value above chance was AUC (57.9%). Therefore, it can be said that there are no significant findings when testing distraction against fatigue.

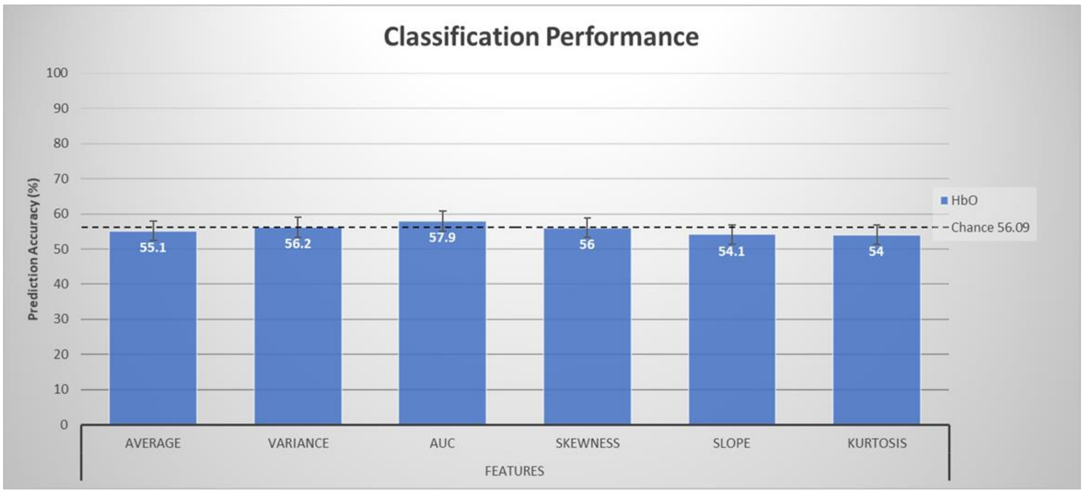


Figure 9.4 - Classification performance of distraction against fatigue with respect to individual oxygenation features.

9.3.2.2 Increased Workload against Distraction

Figure 9.5 shows the classification performance values of increased workload against distraction with respect to oxygenation features. All oxygenation features have a classification performance above chance (56.09%). Kurtosis has the lowest performance percentage of 62.4%, which is 6.3% higher than chance. The highest classification performance came from AUC (68.5%).

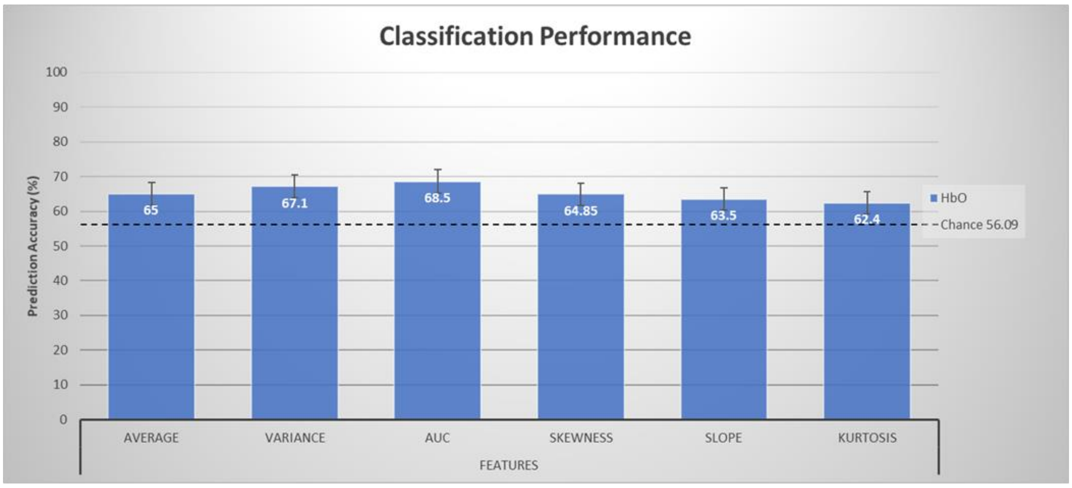


Figure 9.5 - Classification performance of Increased workload against distraction with respect to oxygenation features.

Table 9.4 shows the classification performance values for each participant with respect to oxygenation features. Again, there is a trend that AUC shows the highest performance values throughout the study.

Table 9.4 - Classification performance of individual participants with respect to oxygenation features.

Increased Workload vs Distraction						
Participants	**Average**	**Variance**	**AUC**	**Skewness**	**Slope**	**Kurtosis**
S1	66	68	70	66	63	62
S2	67	67	69	65	64	63
S3	66	68	69	66	65	64
S4	68	69	69	66	64	63
S5	67	67	60	64	66	63
S6	65	65	67	65	65	64
S7	67	66	69	67	64	65
S8	68	66	69	65	63	62
S9	65	67	68	64	64	64
S10	64	65	67	65	63	64
S11	65	64	68	66	63	63
S12	64	68	68	65	64	61
S13	64	69	69	64	64	61
S14	67	69	70	63	63	62
S15	65	68	69	64	64	63
S16	65	67	70	65	64	62
S17	64	68	67	67	63	63
S18	65	69	68	65	65	64
S19	66	69	68	64	63	63
S20	64	69	67	64	64	62

9.3.2.3 Increased workload against Fatigue

Figure 9.6 shows the classification performance percentages with respect to oxygenation features. All oxygenation features show performance percentages above chance (56.09%). The lowest performance shown came from kurtosis (67.4%) which is 11.31% above chance. The highest performance value came from AUC (71.5%).

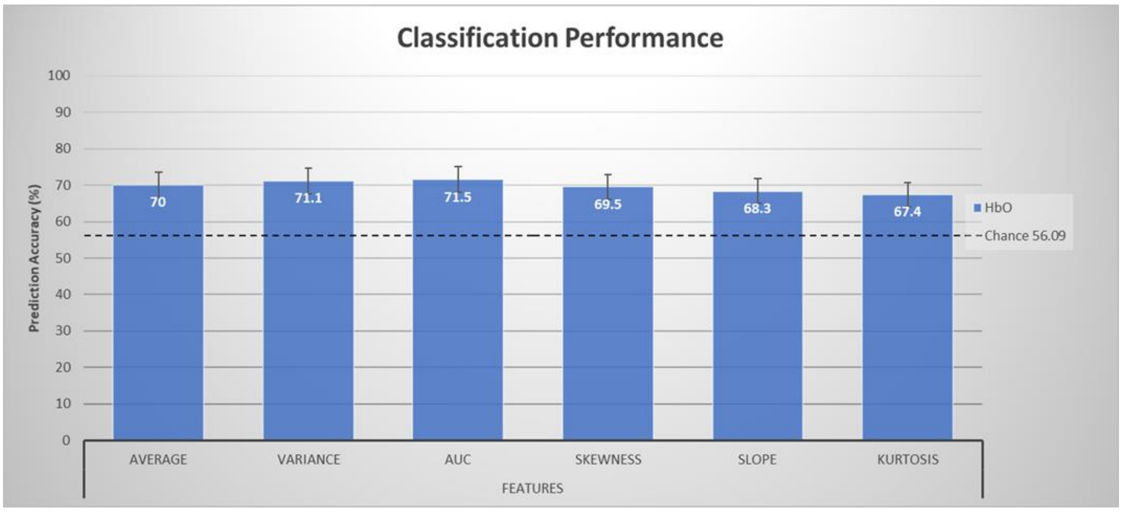


Figure 9.6 - Classification performance of increased workload against fatigue with respect to oxygenation features.

Table 9.5 below shows the classification performance values for individual participants with respect to oxygenation features. AUC and variance showed to have the highest performance percentages throughout the study.

Table 9.5 - Classification performance of individual participants with respect to oxygenation features.

Increased Workload vs Fatigue						
Participants	Average	Variance	AUC	Skewness	Slope	Kurtosis
S1	70	71	72	69	70	67
S2	71	71	73	68	71	66
S3	70	70	72	70	70	68
S4	70	69	71	71	69	68
S5	71	68	72	69	68	67
S6	69	71	73	68	69	66
S7	69	72	71	69	68	65
S8	68	71	70	72	69	68
S9	72	72	71	72	68	69
S10	71	69	73	71	67	69
S11	71	73	72	70	69	68
S12	70	72	71	69	68	67
S13	70	73	70	68	69	68
S14	72	73	71	69	70	68
S15	71	72	72	70	68	69
S16	70	73	71	70	68	67
S17	69	72	71	71	69	67
S18	69	72	72	71	69	68
S19	71	71	70	70	68	66
S20	70	71	71	69	69	68

9.4 Discussion

The main motivation behind this study was to use fNIRS to develop a scientific error model that could be used to assess seafarer performance whilst dealing with adverse PSFs commonly experienced on board a ship. Our subjective measures confirmed that a normal workplace environment is heavily contrasted to that of adverse workplace factors as workplace factors led to significantly higher average oxygenated haemoglobin levels when compared to a standard test.

The overall classification results confirmed that each workplace factor could be discriminated in an engine room simulator. This is substantiated by previous neuro-ergonomics studies showing that fNIRS is well suited for operator mental state monitoring in ecological situations [18] [45] [70] [81].

The best classification accuracy reached 85.15%, taken from AUC oxygenation feature for the increased workload against a standard test study. This result compares favourably with recent studies. For instance, Kevin Verdiere et al [45] obtained a classification performance of 66.9% on 11 subjects using oxygenation features, connectivity features and chromophore concentration. Studies by Hong et al [146], Holper and Wolf [147], Naseer et al [148] obtained classification performance of 75.6% on 10 subjects, 81.3% on 12 subjects and 93% on 7 subjects respectively. On first glance these results compare similarly with ours however, these studies did not consider a continuous but multiple set sub-tasks assessment of specific cognitive activity contrarily to our engine room simulator task which involved different executive and attentional skills. Similar to our study Khan and Hong 2015 [149] also showed that oxygenation features could yield a high accuracy (84.9%) using a driving simulator to monitor fatigue/drowsiness.

The comparison of oxygenation features classification performance revealed that AUC and Variance resulted in significantly higher classification accuracies for all workplace factors (Distraction, Fatigue, Increased Workload). This is similar to the study conducted by *Vierdiere et al* [45] where AUC was found to be the oxygenation feature with the most significance.

It is interesting to note that features present complementary advantages. All oxygenation features are an uncomplicated and low-cost computational measurement to effectuate. This is of considerable advantage as long as passive Brain-Computer Interfaces are concerned. Moreover, the oxygenation features computed in our study can take into account both time and chromophore. Oxygenation features from fNIRS data have been used for some years to evaluate operator performance in the aerospace sector, but up until now haven't been used in the maritime sector to evaluate engine room operators. Therefore, it is difficult to compare results from other maritime studies. Based on the comparisons to other studies above, undertaken using oxygenation features as a classifier we can say that our study had a successful outcome. Our study provides some novel methodological guidance for the implementation of fNIRS based BCI metrics in the maritime industry. To the best of our knowledge, to date, this study is unique, to be the first to benchmark different fNIRS oxygenation metrics and to use them for classification purposes in ecological settings for the benefit of human error assessments. It paves the way forward towards operator mental state estimation in an ecological maritime environment, but some challenges still remain.

Evaluating Fatigue against Distraction showed that most classification performance values were below chance (54-57.9%). This means that LDA could not differentiate between the levels of activation for participants undergoing a fatigued and distracted test. This makes sense as distraction and fatigue had a similar significance when analysed against a standard test.

When comparing PSFs against one another the only significance found was between increased workload against fatigue or distraction. Increased workload against fatigue showed classification performance values of 67.4 - 71.5% and against distraction 62.4 – 68.5%. This shows that increased workload is the PSF that induced a significantly higher activation than other PSFs. On the contrary to the test of fatigue against distraction, the performance values shown from analysis of increased workload against PSFs indicate that distraction has a greater effect on participants. This is shown due to LDA being less able to distinguish (on average) increased workload against distraction (68.5%) compared to fatigue (71.5%). The 3 percentage points is a subtle difference but would further substantiate the larger effect sizes for distraction found in the ANOVA studies in chapter 5 and 6.

9.5 Concluding remarks

The classification performance percentages for PSF against a standard test showed that workload had the highest prediction scores, indicating that workload has the highest significance with respect to human error in a ship engine room. This is followed by distraction and then fatigue with the lowest prediction scores.

The area under the curve oxygenation feature provided the highest average classification percentage scores. Combined features or feature couples were not used due to no significant increase in classification percentages.

The classification performance percentages of PSF against PSF showed that workload was again the most significant PSF, as fatigue against distraction tests provided low classification percentage scores, most of which were below chance.

Chapter 10. Final Conclusion

10.1 Introductory remarks

This is the final chapter of the thesis defining; how the objectives have been achieved, the limitations and critique of the study, contributions to existing knowledge and potential future research.

10.2 Reflections

In chapter 3 the ship accident databases were analysed to find the most significant PSF with relation to human error incidents within a ship engine room. The relevant PSFs were determined by consequence and frequency of occurrence defined in the individual incident reports. These reports allowed for a tabulated formatting of PSFs reported. Once tabulated, the individual PSFs' occurrence frequency could be tallied and the most frequent investigated. Therefore, objective one has been achieved.

Due to the multi-disciplinary nature of this research, a multitude of differing literature topics were reviewed. This literature was reviewed in chapter 2. The information needed from the literature reviewed varied between maritime HRA, psychology techniques, human factors techniques, neuroscience, neuro-ergonomics, Neuro-imaging and simulation studies. There are many differing techniques and settings for the equipment used in BCI-fNIRS studies. Therefore, concise summaries were used detailing the techniques. This chapter provided the foundation knowledge to later achieve objectives two, three and four.

Each participant was connected to fNIRS whilst undergoing a ballasting task. The data provided gave many novel outcomes discussed throughout the thesis. Therefore, the use of fNIRS, as a method to measure operator functional state was a successful approach. This approach was conducted in chapters five, six, seven and eight where objective three was achieved. To reflect on this approach, it could have been beneficial to gather data from additional regions of the cerebrum in order to look at a connectivity matrix.

A ship engine room simulator was used to develop a novel risk-based simulation system to support engine room operations. The tasks performed by the participants would exactly replicate the tasks performed in 'real-world' operations. Using a ballasting scenario, a number of risk-based assessments could be made for monitoring, fault detection, problem solving, working memory and decision making. The TRANSAS system also allowed the examiner to closely watch and record each participant as they navigated through each sub-task. Therefore, the second project objective was successfully achieved.

PSFs were defined using ship accident databases as outlined above in chapter three. These PSFs were implemented into the ballasting scenario and tested against a standard task and against other PSFs. This allowed for the accurate analysis of PSF influence on OFS and in turn human error. Given the results obtained, the influence of PSF on OFS is clearly shown. Therefore, this further substantiates that objective three was successfully achieved.

In chapters five, six, seven and eight, ANOVA is used to evaluate the significant effect of stressors and the workflow used. In chapter nine LDA is used to model the relationship between OFS and PSF to obtain HEP (classification performance-based error probability). Given the results we can clearly see the significant effects and differences in classification performance percentages. Therefore, the model has successfully shown the PSFs most likely to contribute to human error. It also shows that the change of error is greater for certain PSFs when compared to others. Therefore, objectives four and five have been attained.

10.3 Limitations and critical analysis of the research project

Expanding our study to use connectivity measure could be done in the future. This could be of benefit due to the results shown in [45] [51] where connectivity measures resulted in a slightly more accurate classification performance when compared to oxygenation features. An explanation for this would be that the evaluation of workflow-related concentrations (e.g. haemodynamic response) is based on the time-locked event. In the past it has been suggested that these workflow related responses cause a minute increase (<5%) in neural energy depletion when compared to general brain energy consumption [75]. Therefore, by solely focusing on a localized DLPFC haemodynamic response, some of the neural activity is disregarded. There is an argument that states, cognition relies on the activation of multiple cerebral areas as opposed to single site processing units [139], [150], [151],. Therefore, the evaluation of the interaction between cerebral networks could provide a more accurate classification performance coupled with more information on neural dynamics. Especially when attempting to understand the risks associated with real-life engine room factors [152], [153], [154]. Furthermore, some relevant investigations stated that amplitude or frequency correlations amongst low frequency oscillations (<=0.1Hz) are closely associated to cortical processes [155], [156], [157]. Therefore, when a baseline reading is taken from each participant, connectivity features based on frequency or amplitude coupling could give an outlook on the cognitive process with respect to continuous monitoring of neural activity.

The engine room simulator is limited to its capabilities as it is easy to simulate seafarer duties within the engine room but the only consequence to an incorrect action is an engine room alarm. The alarm is enough to neuro physiologically activate each participant but in a real-life scenario, an incorrect action could cause a physical or visual problem. For example, fire, flood, electric shock, ship listing, the ship sinking, injury and even death. The consequences listed would presumably cause an increased neurophysiological activation, which is not shown on this study.

Due to the sensitivity of the infrared sensors and detectors, a desktop version of the engine room simulator was used. In reality, seafarers would be moving around whilst completing the workflow used for this study, but this would have caused great interference, an increased number of artefacts and anomalous data. For future experiments, an investigation into how it could be possible to modify the fNIRS system, so it can be used with a portable backpack would give a more realistic investigation.

A sizable limitation is concerning the fNIRS signal evaluation. The fNIRS signal is the consequence of a global integrant affected by skin blood flow and a particular neuronal component. If our study did not contain the analysis of separate epochs then some algorithms based on spatial filtering and principal component analysis such as the one used by Zhang et al [158] could have been used. Additionally, fNIRS signals can be influenced by several other physiological factors such as blood pressure, respiration, heartbeats and perspiration. It could have been beneficial to record those factors in order to evaluate the way in which they contribute to level of engagement.

Due to the limitations involving the paradigm settings it was not possible to make any conclusion regarding the underlying neurophysiological processes. However, our classification performances were high and satisfactory.

Obtaining candidates with ship engine room experience to different levels, coupled with experience of using the TRANSAS software was very difficult. Therefore, in-house training had to be done for each candidate as stated in chapter 3.1.1. The only way to differentiate experienced and inexperienced was to train the inexperienced candidates passively. Candidates were tested on passive training compared to practical training. It would have been preferable to test participants with 10 years+ ship engine room experience and participants with a few weeks ship engine room experience, all trained the same way. Also, each of the four experiments involved ten participants that performed an equal share of two conditions (one condition each). This is a limited number of trials due to the difficulty in obtaining a larger number of candidates with the required credentials and the time constraints for training and experimentation.

Candidates taking part in the fatigue study would be required to wear the cap for a long period of time (>45 minutes). Therefore, they would experience discomfort and some mild pain from the nodes.

The choice of the four workplace conditions (Fatigue, Increased Workload and Distraction versus A standard test) could have potential confounds such as motor responses that could have an influence on our analysis. However, motor areas were not specifically targeted therefore the influence of such areas is low.

The motivation for our project is to monitor the brain activity of engine room operators when facing realistic workplace factors. The manipulation of such factors in controlled, ecological environments such as on an engine room simulator remains challenging. The first experiment was a datum study but was also used as a way to set the path to more refined protocols to characterize different workplace factors with the view to perform more efficient participant training and monitoring as achieved by Toppi et al [71].

Finally, in critical engine room emergency systems the rate of false negatives (>20%) could not be afforded due to the risk of fatality. Therefore, the performance classification could be improved by combining bimodal EEG and fNIRS pBCI's [159] but it is very unlikely that a classification performance of 100% will ever be reached.

10.4 Contributions to knowledge

This project has contributed considerably towards the knowledge of HRA within the maritime sector. This project is the first to date within the maritime sector to model PSFs against OFS using fNIRS combined with an engine room simulator.

This project is the first within the Maritime sector to obtain a classification performance of PSFs by combining the use of fNIRS and an Engine room simulator.

The use of fNIRS allows for a non-speculative data set of OFS whilst undertaking real engine room tasks. Showing the exact areas of the workflow and sub-tasks that induce higher levels of activation. The result of this gives the maritime sector the knowledge of the problematic areas of engine room tasks allowing for RCOs to be implemented.

The techniques developed in this study can be used in future HRA studies in order to obtain a more accurate outcome.

This study contributes to the current knowledge of the relationship between distraction, fatigue and workload with respect to operator performance via fNIRS providing a detailed outlook on OFS whilst experiencing the PSF listed above. Showing the PSF out of the three most likely associated with mental workload based human error.

10.5 Future Research

Our developed human error framework using neuroimaging and BCI techniques detailed above could be used in various safety critical systems (for example Container shipping, LNG shipping or the holistic approach to the transportation of goods [port terminals, lorries, trains and shipping]) to gauge operator performance, crew/team performance, operator mental state, the root cause of error, to develop a human error framework or performance classification model specific to each system and a much greater level of human error probability (HEP) accuracy.

Our framework could also be taken a step further. Given that the tasks are separated into epochs, an investigation could be performed on the specific problem areas detailed by the higher LDA prediction or the higher levels of HBO for each individual or group of epochs. From there, the specific problem sub-task and the root cause of the error could be identified. This can be done in 2 ways:

1. There are a number of different ways to complete a sub-task/task. By looking at operator functional state after completing the problem task we can compare the different routes through the task against HBO or LDA outputs in order to see which route shows the highest levels of activation.
2. To analyse the data from specific parts of the sub-task (as opposed to epoch group average) where the participants experience the highest levels of activation, indicating the problem area to an accuracy of a single second.

It would be beneficial to study in a similar way, the effects of performance shaping factors (PSF) using a hydraulic simulator to enable us to mimic real environmental conditions (e.g. noise and vibration and adverse weather conditions [pitch and roll]). A comparison could then be made of the performance shaping factors individually and when combined, to see which PSF or combination of PSF's has the most significant effect on neurophysiological activation, indicating the highest likelihood to contribute towards human error.

A ballasting workflow task was used for this study, but it would be beneficial to study other engine room scenarios. (e.g. bunkering, sea water treatment, electrical systems or machine maintenance). Evaluating the participants' mental workload for each task would indicate the scenario that the participants on average, found the most difficult thus, indicating the tasks with the highest risk of human error.

A connectivity study could be performed on the data obtained to see the neural networks between the left side and right side DLPFC. This would allow us to see the parts of the DLPFC that are activated simultaneously, allowing us to use fNIRS as a prediction method towards an alternative performance classification model or to use connectivity as a method of identifying the areas of a task where participants have the highest levels of activation.

10.6 Recommendations

This project, methodology, model and research findings will be relevant to differing sectors. For example, this method and model could not only be applied to maritime engine room operations but also to bridge operations where it is assumed that deck officers will experience similar PSFs.

This work could also be applied to operator training in various sectors. In the maritime sector this work could be used to evaluate engineer and deck officer training procedures by either:

- Evaluating one training programme against another with respect to OFS post training on a ship simulator. This would tell educators the value of their training programmes against another and how each operator benefits from the training.
- Evaluating the difference in OFS before and after the training programme to see the benefit of the training programme compared to a totally inexperienced operator.
- To evaluate practical against theoretical training methods against OFS. This could influence training techniques with respect to depth of learning and operator competency.
- Training tasks and sub tasks can be broken down into sections and individually evaluated to gauge an understanding of their validity and varying levels of difficulty from a neurophysiological activation perspective.
- Evaluating operator against operator to see who would be best suited to which role. This could achieve an optimal work force of operators.

As stated earlier these techniques can be applied to any simulated scenario using a desktop simulator of tasks or applied to an aircraft, ship, train driving, automotive driving simulator.

Final note - Ethics of the study

Ethical approval of this study was obtained via Liverpool John Moores Universities ethics review panel. A full documented outline of this exact tasks, including environment and equal pay scales (based on an hourly rate) to participants.

References

[1] National Transport Safety Board (NTSB), “Marine Accident Reports,” Open Gov (NTSB), Updated 2017. [Online]. Available: https://www.ntsb.gov/investigations/AccidentReports/Pages/marine.aspx. [Accessed 16th October 2017].

[2] R. Bye and A. Aalberg, “Maritime navigation acidents and risk indicators: An exploratory statistical analysis using AIS data and accident reports,” vol. 176, no. 1, 2018.

[3] S. Fan, E. Blanco-Davis, Z. Jinfen and Z. Yang, “Effects of seafarers’ emotion on human performance using bridge simulation,” *Ocean Engineering,* vol. 1, no. 1, pp. 1-3, 2019.

[4] GOV.UK, “Marine accident investigation branch - reports,” Uk Government, updated 2017. [Online]. Available: https://www.gov.uk/maib-reports. [Accessed 16th October 2017].

[5] European Maritime Safety Agency, “EMSA Accident Investigation,” EMSA, Updated 2017. [Online]. Available: http://www.emsa.europa.eu/fc-default-view/tagged/84-accident-investigation.html. [Accessed November 2017].

[6] T. Bielic, N. Hasanspahic and J. Culin, “Preventing marine accidents caused by technology-induced human error,” vol. 31, no. 1, 2017.

[7] E. Akyuz, M. Celik, I. Akgun and K. Cicek, “Prediction of human error probabilities in a critical marine engineering operation on-board chemical tanker ship: The case of ship bunkering,” vol. 110, no. p102-109, 2018.

[8] G. Jahanshahloo, F. Hosseinzadeh-Lotfi and M. Izadikhah, “Extension of the TOPSIS method for decision making problems with fuzzy data.,” vol. 181, no. 2, 2006.

[9] K. Li, J. Yin, H. Bang, Z. Yang and J. Wang, “Transportmetrica A: Transport Science - Baysian network with quantitive input for maritime risk analysis.,” vol. 10, no. 2, 2014.

[10] Y. Xi, Z. Yang, Q. Fang, W. Chen and J. Wang, “New hybrid approach to human error probability quantification applications in maritime operations,” vol. 1, no. 1, 2017.

[11] Australian Government, “Marine Safety Investigations and reports,” Australian Government, Updated 2017. [Online]. Available: https://www.atsb.gov.au/publications/safety-investigation-reports/?mode=Marine. [Accessed November 2017].

[12] Pomorstvo, “Simulator Instructor - STCW requirements and reality,” vol. 20, no. 2, 2006.

[13] R. Brons, L. Joost and M. Bierens, “Herald of free enterprise,” in *Water-Related Disasters*, Chapter 10.4, Springer Link, 2015, pp. pp 535-585.

[14] S. Fan, X. Yan, J. Wang and Z. Jinfen, A review on human factors in maritime transportation using seafarers' physiological data, Liverpool: Frontiers, 2017.

[15] Government of Canada, “Transportation Safety Board of Canada,” Canadian Government, Updated 2018. [Online]. Available: http://www.bst-tsb.gc.ca/eng/rapports-reports/marine/index.asp. [Accessed October 2017].

[16] The Government of Hong Kong, “The government of hong kong special administrative region,” Hong Kong Government, Updated 2017. [Online]. Available: https://www.mardep.gov.hk/en/publication/ereport.html. [Accessed October 2017].

[17] C. Chauvin, S. Lardjane, G. Morel, J. Clostermann and B. Langard, “Human and Organisational factors in maritime accidents,” Elsevier, Helsinki, 2013.

[18] H. Aghajani, M. Garbey and A. Omurtag, “Measuring Mental Workload with EEG plus fNIRS,” vol. 11, no. 1, 2017.

[19] J. Baker, J. Bruno, A. Gundran, S. Hosseini and A. Reiss, “fNIRS measurement of cortical activation and functional connectivity during a visuospatial working memory task,” vol. 1, no. 1, 2018.

[20] *Meeting with John Carrier,* Liverpool, 2017.

[21] *Meeting with Geraint Phylip-Jones,* Liverpool, 2017.

[22] *Meeting with Jonathan Warren,* Liverpool, 2018.

[23] S. Dekker, Field Guide to Human Error, London: Ashgate Publishing Ltd, 2002.

[24] J. Rasmussen, “Human errors. A taxonomy for describing human malfunction in industrial installations,” *Journal of Occupational Accidents,* vol. 4, no. 2-4, pp. 311-333, 1982.

[25] G. Giorgio and M. K. Stephen, “Neuroimaging,” in *Encyclopedia of the human brain*, science direct, 2002.

[26] J. Fuster M, “The prefrontal cortex (fifth edition),” in *Neurimaging*, science direct, 2015, pp. 296-297.

[27] R. A. Koenigsberg and Y. T. Fong, “Neuroimaging,” in *Textbook of clinical neurology*, Hanover, Germany, science direct, 2007, pp. 130-133.

[28] E. T. Solovey, B. Mehler and B. Reimer, “Brain Sensing with fNIRS in the Car,” Cambridge, MA 02139, 7th October 2012.

[29] W. Lim, Y. Liu, S. Subramaniam, S. Liew, G. Krishnan, O. Sourina, D. Konovessis, H. Ang and L. Wang, “EEG-Based mental workload and stress monitoring of crew members in maritime virtual simulator.,” vol. 10830, no. 1, 2018.

[30] R. Parasuraman, J. Christensen and S. Grafton, Neuroergonomics - Study of the Brain at Work, Fairfax, VA 22030: Elsevier, 2011.

[31] B. Somon, A. Campagne, A. Delorme and B. Berberian, “Performance Monitoring Applied to System Supervision,” vol. 1, no. 1, 2017.

[32] Z. Chua and M. Causse, “Aging Effects on Brain Efficiency in General Aviation Pilots,” vol. 1, no. 1, 2016.

[33] D. Gautier, S. Sebastien, G. Thibault, D. Arnauld and D. Frederic, “Processing Functional Near Infrared Spectroscopy Signla with a Kalman Filter to Assess Working Memory during Simulated Flight.,” vol. 1, no. 1, 2016.

[34] A. Gevins and M. Smith, “Neurophysiologic measures for neuroergonomics,” Las Vegas, Nevada, 2005.

[35] L. Bu, J. Li, F. LI, H. Liu and Z. Li, “Wavelet coherence analysis of cerebral oxygenation signals measured by near-infrared spectroscopy in sailors: an exploratory, experimental study.,” vol. 6, no. 11, 2016.

[36] R. Parasuraman and M. Mustapha, “Automation and Human Performance,” in *Theory and applications*, University of central Florida, Orlando, Lawrence Erlbaum Associates, 1996.

[37] P. A. Hancock, “Neuroergonomics: Where the coretx hits the concrete,” *Frontiers,* vol. 1, no. 1, 2019.

[38] TRANSAS, “ERS 5000 Engine room simulator,” TRANSAS, [Online]. Available: http://www.transas.com/products/simulation/engine-room-and-cargo-handling-simulators/ERS5000#description. [Accessed 1st March 2018].

[39] K. Takashi, T. Hitoshi, S. Tomoki, T. Hiroki and S. Takuji, “Measurement of Train Driver’s Brain Activity by Functional Near-Infrared Spectroscopy (fNIRS),” vol. 27, no. 13-15, 2006.

[40] A. Chiarelli, F. Zappasodi, F. DI Pompeo and A. Marla, “Simultaneous functional near-infrared spectroscopy and ectroencephalography for monitoring of human brain activity and oxygenation: a review,” vol. 4, no. 4, December 2017.

[41] S. Felix, K. Stefan, J. M. Andreas, Z. Raphael, M. P. Juan, W. Ursula and W. Martin, “A review on continuous wave functional near-infrared spectroscopy and imaging,” vol. 7, no. 1, 2013.

[42] T. Hitoshi and Y. Kazuki, “Measurement of brain function of a car driver using functional near infrared spectroscopy,” vol. 1, no. 1, 2009.

[43] T. Shimizu, S. Hirose, H. Obara and K. Yanagisawa, “measurement of pre frontal cortex brain activity attributable to the driving workload and increased attention.,” vol. 1, no. 1, 2009.

[44] R. Parasuraman, J. Christensen and S. Grafton, “Neuroergonomics - Study of the Brain at Work,” vol. 11, no. 1, 2011.

[45] K. Verdiere, N. R. Roy and F. Dehais, “Detecting Pilot's Engagement Using fNIRS Connectivity Features in an Automated vs. Manual LAnding Scenario,” vol. 1, no. 1, 2018.

[46] Z. Yang, S. Bonsall, A. Wall, J. Wang and M. Usman, “A modified CREAM to human reliability quantification in marine engineering,” vol. 1, no. 1, 2012.

[47] W. Ying-Ming, C. Kwai-Sang, K. K. P. Gary and Y. Jian-Bo, “Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean.,” vol. 1, no. 1, 2009.

[48] T. Paolo and C. L. Maria, "A Probabilistic cognitive simulator for HRA studies (PROCOS)," *Reliability Engineering and system safety,* vol. 8, pp. 1117-1130, 2009.

[49] M. Gul, E. Celik and E. Akyuz, "A hybrid risk-based approach for maritime applications: The case of ballast tank maintenance.," vol. 23, no. 6, 2017.

[50] M. P. van den Heuvel and O. Sporns, "Network hubs in the human brain," *Trends Cogn, sci,* vol. . doi: 10.1016/j.tics.2013.09.012, pp. 17,683-696, 2013.

[51] G. Thibalut, D. Gautier, L. Francois, S. Sebastien and D. Frederic, "Real-Time State Estimation in Flight Simulator Using FNIRS," vol. 1, no. 1, 2015.

[52] S. Fairclough, S. Fan, E. Blanco-Davis and Z. Yang, "The Role of the Prefrontal Cortex and Functional Connectivity during Maritime Operations: An fNIRS study," 2020.

[53] Z. Yang, S. Bonsall, A. Wall, J. Wang and M. Usman, "Ocean Engineering - A Modified CREAM to human reliability quantification in marine engineering.," vol. 58, no. pp293-303, 2013.

[54] G. Thibault, A. Hasan and D. Frederic, "In silico vs. Over the Clouds: On-the-Fly mental State Estimation of Aircraft Pilots, Using a Functional Near Infrared Spectroscopy Based Passive-BCI," vol. 1, no. 1, 2018.

[55] G. Thibault, D. Gautier, L. Francois, S. Sebastien and D. Frederic, "Real-Time State Estimation in a Flight Simulator Using fNIRS," PLOS, California, 2015.

[56] R. Roy, F. Dehais, A. Dupres, G. D. Flumeri, K. Verdiere, G. Borghini and F. Babiloni, "Monitoring pilots cognitive fatigue with engagement features in simulated and actual flight conditions using a hybrid fNIRS-EEG passive BCI," in *2018 IEEE international conference on systems, man, and cybernetics (SMC)*, Miyazaki, Japan, 2018.

[57] L. Bu, D. Wang, C. Huo, G. Xu, Z. Li and J. Li, "Effects of poor sleep quality on brain functional connectivity revealed by wavelet-based coherence analysis using NIRS methods in elderly people," vol. 1, no. 1, 2018.

[58] T. Kojima, H. Tsunashima and T. Shiozawa, "Functional brain imaging of a train driver by functional near-infrared spectroscopy (FNIRS)," vol. 1, no. 1, 2004.

[59] B. B. Elif, T. Leyla and A. Ozcan, "The subjective measurement of seafarers' fatique levels and mental symptoms.," vol. 1, no. 1, 2015.

[60] Hiteshk, "Duties of the Chief, 2nd and 4th Engineers," 7th October 2017. [Online]. Available: https://www.marineinsight.com/careers-2/duties-of-2nd-engineer-on-ship/. [Accessed October 2017].

[61] G. Huve, K. Takahashi, M. Hashimoto, V. Kurkova, Y. Manolopoulos, B. Hammer, L. Iliadis and I. Maglogiannis, "fNIRS-Based Brain-Compter Interface Using Deep Neural Networks for Classifying the Mental State of Drivers," in *Artificial Neural Networks and Machine Learning*, Rhodes, Greece, Springer-Verlag Berlin, 2018, pp. 353-362.

[62] Global Mobility, “Marine Chief Engineer in Marine Engineering,” Logistics, June 2017. [Online]. Available: https://www.alchemyrecruitment.com/maritime-job-sectors/crewing/chief-engineer.

[63] K. Yamamoto, H. Takahashi, T. Sugimachi, K. Nakano and Y. Suda, “The study of driver's brain activity and behaviour on DS test using fNIRS,” in *2nd International federation of automatic control (IFAC) conference on cyber physical and human systems (CPHS)*, Miami, FL, 2019.

[64] M. Tanveer, M. Khan, M. Qureshi, N. Naseer and K. Hong, “Enhanced drowsiness detection using deep learning: An fNIRS study,” *IEEE ACCESS,* vol. 7, no. 1, pp. 137920-137929, 2019.

[65] Hiteshk, “Duties of 2nd Engineer on ship,” Marine Careers, 7th October 2017. [Online]. Available: https://www.marineinsight.com/careers-2/duties-of-2nd-engineer-on-ship/.

[66] M. C. Barsan E, “Combined complex maritime simulation scenarios for reducing maritime accidents caused by human error,” vol. 1, no. 1, 2010.

[67] N. Kumar, “Support Vector Machine in machine learning,” Software Architect, Gurgaon, India, 2019.

[68] S. Fan, E. Blanco-Davis, J. Zhang, A. Bury, J. Warren, Z. Yang, X. Yan, J. Wang and S. Fairclough, “The role of the pre-frontal cortex and functional connectivity during maritime operations: An fNIRS Study,” *Brain and Behaviour,* vol. 11, no. 1, 2021.

[69] Y. Takeuchi, “Change in blood volume in the brain during a simulated aircraft landing task.,” vol. 42, no. p60-65, 2000.

[70] M. Bruce, R. Bryan and E. Solovey, “Brain sensing with fNIRS in the car,” 2012.

[71] J. Toppi, G. Borghini, M. Petti, E. J. He, V. De Giusti and B. He, “Investigating cooperative behavior in ecological settings: an eeg hyperscanning study.,” *PLoS ONE,* vol. doi: 10.1371/journal.pone.0154236, p. 11, 2016.

[72] H. Ayaz, P. Shewokis, S. Bunce, K. Izzetoglu, B. Willems and B. Onaral, “Optical brain monitoring for operator training and mental workload assessment.,” vol. 1, no. 1, 2017.

[73] A. Hasan, P. Shewokis, B. Scott, I. Kurtulus, B. Willems and O. Banu, “Optical brain monitoring for operator training and mental workload assessment,” vol. 1, no. 1, 2011.

[74] G. Bauernfeind, A. Kreilinger, A. Kuber, C. Neuper, G. Muller-Putz, T. Kaufmann and V. Kaiser, “Cortical effects of user training in a motor imagery based brain-computer interface measured by fNIRS and EEG,” Vols. 85, part 1, 2014.

[75] M. Raichle and M. Mintun, “Brain work and brain imaging,” *Annurev, Neurosci,* vol. doi: 10.1146/annurev.neuro.29.051605.112819, pp. 29, 449-476, 2006.

[76] A. Unni, K. Ihme, M. Jipp and J. Rieger, “Assessing the drivers's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study,” vol. 11, no. 1, 2017.

[77] S. G. Hart and E. S. Lowell, “Developement of NASA-TLX,” vol. 52, no. 1, 1988.

[78] Y. Hlotova, O. Cats and S. Meijer, “Measuring bus drivers' occupational stress under changing working conditions,” vol. 1, no. 2415, 2014.

[79] R. Burbridge, “Adapting aviation to a changing climate,” vol. 71, no. 1, 2018.

[80] L. Hocke, I. Oni, C. Duszynski, A. Corrigan, B. Frederick and J. Dunn, “Automated processing of fNIRS data-A visual guide to the pitfalls and consequences,” vol. 11, no. 5, 2018.

[81] Y. Li, L. Zhang, K. Long, H. Gong and H. Lei, “Real-time monitoring prefrontal activities during online video game playing by functional near-infrared spectroscopy,” vol. 11, no. 9, 2018.

[82] J. Gemignani, E. Middell, R. Barbour, H. Graber and B. Blankertz, “Improving the analysis of near-spectroscopy data with multivariate classification of hemodynamic patterns,” vol. 15, no. 4, 2018.

[83] E. Akyuz and E. Celik, “The role of human factor in maritime environment risk assessment: A practical application on ballast water treatment (BWT) system in ship,” vol. 24, no. 1, 2018.

[84] S. Symes and S. Fairclough, “Fault detection and correction in the engine room of a ship: An fNIRS study,” in *Human Factors and Ergonomics Society, Europe Chapter Annual Meeting 2019*, Nantes, 2019.

[85] Q. Yang and S. Ge, “Classification of a single channels fNIRS signal for a Brain Computer Interface,” Dalian, 2014.

[86] A. Arif, M. Khan, K. Javed, H. Sajid, S. Rubab, N. Naseer and T. Khan, “Hemodynamic response detection using integrated EEG-fNIRS-VPA for BCI,” vol. 70, no. 1, 2021.

[87] A. Hashimoto, C. Nishizaki and N. Mitomo, “A study on application to marine accident of HRA methods,” in *2015 4th internation conference on informatics, electronics and vision ICIEV 15*, Fukuoka, Japan, 2015.

[88] M. Alvarenga and P. Melo, “A review of the cognitinve basis for HRA,” *Progress in Nuclear engergy,* vol. 107, Nov 2019.

[89] M. R. Grech, T. Horberry and A. Smith, “Human Error in Maritime Operations: Analyses of Accident Reports Using the Leximancer Tool,” vol. 46, no. 19, 2002.

[90] Z. Yang, S. Bonsall and J. Wang, “Expert Systems with Applications - Use of hybrid multiple uncertain attribute decision making techniques in safety management.,” vol. 36, no. 2, 2009.

[91] X. Pan, Y. LIN and C. He, “A review of cognitive methods in human reliability analysis,” *Quality and reliability engineering international,* vol. 33, pp. 1299-1316, 2017.

[92] M. Evan, Y. He, L. Maglaras and H. Janicke, “HEART-IS: A novel technique for evaluating human error related information securioty incidents,” *Computers and security,* vol. 80, pp. 74-89, Jan 2019.

[93] N. Dsouza and L. Lu, “A literature review on human reliability analysis techniques applied for probabilistic risk assessment in the nuclear industry,” in *7th International conference on applied human factors and ergonomics*, Bay-Lake, Florida, 2017.

[94] Y. Chang and A. Mosleh, “Cognitive modelling and dynamic probabilistic simulation of operating crew response to complex ssytem accidents,” *Reliability engineering and system safety,* vol. 92, pp. 997-1013, AUG 2017.

[95] S. Fan, E. Blanco-Davis, Z. Yang, Z. Zhang and X. Yan, “Incorporation of human factors into maritime accident analysis using a data driven bayesian network,” *Reliability engineering and safety systems,* vol. 203, Nov 2020.

[96] R. Patriarca, M. Ramos, N. Paltrinieri, S. Massaiu, F. Costantino, G. Di Gravio and R. Boring, “Human Reliability Analysis: Exploring the interllectual structure of a research field,” *Reliability Engineering & Systems Safety,* vol. 203, pp. 197-214, Nov 202.

[97] M. Bhutta, K. Hong, M. Hong and Y. Kim, “Classification performance analysis of combined fNIRS-polygraph system using different temporal windows,” Busan, South Korea, 2015.

[98] K. Hong, N. Naseer and Y. Kim, “Classification of prefrontal and motor cortex signals for three-class fNIRS-BCI,” vol. 587, no. 87-92, 2015, Feb 5th.

[99] R. Khan, N. Nasser, N. Qureshi, F. Noori, H. Naseer and M. Khan, “fNIRS-based neurorobotic interface for gait rehabilitation,” vol. 15, 2018, Feb 5th.

[100] M. Khan, K. Hong, N. Naseer and M. Bhutta, “Drowsiness detection in dorsolateral-pre-frontal cortex using fNIRS for a passive-BCI,” Busan, Oct 13-16, 2015.

[101] A. Sereshkeh, R. Yousefi, A. Wong, F. Rudzicz and T. Chau, “Development of ternanry hybrid fNIRS-EEG BCI based on imagined speech,” vol. 6, no. 4, 2019.

[102] M. Ghaffar, U. Khan, N. Naseer, N. Rahid and M. Tiwana, “Improved classification accuracy of four class fNIRS-BCI,” Rawalpindi, Pakistan, JUn 25-27, 2020.

[103] W. Cai, G. Guan, R. Pan, X. Zhu and H. Wang, “Network Linear Discriminant analysis,” *Computational statistics and data analysis,* vol. 117, pp. 32-44, 2018.

[104] W. HU, W. Shen and H. K. Zhou, “Matrix Linear Discriminant Analysis,” *Technometrics,* vol. 62, no. 2, pp. 196-205, 2019.

[105] A. Jamaluddin and N. Mahat, “Validation assessment on resampleing method in imbalanced binary classification for LDA,” *Journal of information and communication technology malaysia,* vol. 20, no. 1, pp. 83-102, 2021.

[106] N. Naseer, K. Hong and M. B. Khan, “Comparison of artificial neural network and support vector machine learning classifications for fNIRS-based BCI,” Busan, 2015.

[107] S. GE, Q. Yang, R. Wang, P. LIN, J. GAO, Y. Leng, Y. Yang and H. Wang, “A Brain-Computer based on a few channel EEG-fNIRS bimodal system,” vol. 5, no. 208-218, 2017.

[108] M. Hasan and S. Islam, “Suitibility Investigation of the different classifiers in fNIRS signal classification,” Khulna, Bangladesh, 2020.

[109] N. Naseer, K. Hong, M. Khan, M. Bhutta, A. Soh, H. Selamat, R. Rahman, A. Ishak, S. Ahmad, H. Ramil and A. Faudzi, “Analysis of classification performance of fNIRS signals from prefrontal cortex using various temporal windows,” Kota Kinabalu, Malaysia, 2015.

[110] X. Geng and Z. Li, “Decoding fNIRS based imagined movements associated with speed and force for a brain-computer interface,” vol. 34, no. 4, 2020.

[111] Y. Guo, “Support Vectors Classification method based on matrix exponent boundary fisher projection,” in *16th IEEE internation conference on mechatronics and automation*, Tianjin, 2019.

[112] X. Nguyen, L. Huang and A. Joesph, “Support Vector Machines, data reduction, and approximate Kernal matrices,” in *European conference on principles of data mining and knowledge discovery*, Antwerp, 2008.

[113] Q. Yang and F. LI, “Support vector machine for intrusion detection based on LSI feature selection,” in *6th world congress on intelligent control and automation*, Dalian, 2006.

[114] Z. WU, J. YU, Y. LI and W. Xie, “Fuzzy support vector learning algorithim for mixed attributes data,” in *6th world congress on intelligent control and automation*, Dalian, 2006.

[115] N. Qureshi, F. Noori, A. Abdullah, A. Naseer and N. Naseer, “Comparison of classification performance for fNIRS-BCI system,” Islamabad, 2016.

[116] M. Ghaffar, U. Khan, J. Iqbal, N. Rashid, A. Hamza, W. Quereshi, M. Tiwana and U. Izhar, “Improving classification performance of four class FNIRS-BCI using Mel frequency cepstral coefficients (MFCC),” vol. 112, JAN 2021.

[117] H. LI, A. Gong, L. Zhao, F. Wang, Q. Qian, J. Zhou and Y. Fu, “Identification of gait imagery based on fNIRS and class-dependant sparse representation,” vol. 68, no. 102597, JUL 2021.

[118] C. Shibu, S. Sreedharan, K. Arun and C. Kesavadas, “Comparision of classification performance of handpicked, handcrafted and automated-features for fNIRS-BCI,” Naha, Japan, 202.

[119] U. Asgher, K. Khalil, M. Khan, R. Ahmad, S. Butt, Y. Ayaz, N. Naseer and S. Nazir, “Enhanced accuracy for multiclass mental workload detection using long and short term memory for brain-computer interface,” vol. 14, JUN23 2020.

[120] S. Li, L. Chen and Q. Z. Liu, “Constrained All-k-nearest-neighbour search,” in *International conference on measurement, instrumentation and automation (ICMIA 2012)*, Guangzhou, 2013.

[121] J. H. Zar, “Biostatistical Analysis,” vol. Fifth Edition, 2010.

[122] P. Shui, X. Xia and Y. Zhang, “Sea-Land segmentation in Maritime surveillance radars via K-nearest neighbour classifier,” *IEE transactions on aerospace and electronic systems,* vol. 56, no. 5, pp. 3854-3867, 2020.

[123] T. Trakoolwilaiwan, B. Behboodi, J. Lee, K. Kim and J. Choi, “Convolutional neural network for high accuracy functional near infrared spectroscopy in a brain computer interface: Three-class classification of rest, right, and left motor execution,” vol. 5, no. 1, 2018.

[124] N. Naseer and K. Hong, “fNIRS-based brain-computer interfaces: a review,” vol. 9, no. 3, 2015.

[125] J. XU, “Application of Artificial Neural Network (ANN) for prediction of Maritime Safety,” in *International conference on computing, information and control*, Wuhan, 2011.

[126] M. Gorobetz, A. Levchenkov and L. Ribickis, “Modeling of intelligent agents and neural network controller for maritime transport,” in *3rd International workshop on intelligent technologies in logistics and mechatronic systems*, Paneveys, 2008.

[127] G. Wang, L. Huang and C. Zhang, “Study of artificial neural network model based on fuzzy clustering,” in *6th world congress on intelligent control and automation*, Dailian, 2006.

[128] Z. Yang and E. E. Mehmed, “Artifcial neural networks in freight rate forecasting,” *Maritime Economics & Logistics,* vol. 1, no. 1, pp. 395-404, 2019.

[129] F. Dehais, L. Rida, R. Roy, J. Iversen, T. Mullen and D. Callan, “A pBCI to predict attentional error before it happens in real flight conditions,” in *IEEE 2019*, Bari, 2019.

[130] I. Rabiul, K. Faisal, A. Rouzbeh and G. Vikram, “Human error assessment during maintenance operations of marine systems - what are the effective environmental factors,” vol. 1, no. 1, 2018.

[131] L. Smolarek and J. Soliwoda, “Human fatigue model at maritime transport,” Vols. 1-4, no. 1, 2008.

[132] K. O. Mohd, N. F. Mohd, S. F. Noorul and R. Abdul, “The malaysian Seafarers psychological distraction assessment using a TOPSIS method,” vol. 1, no. 1, 2015.

[133] E. Akyuz, M. Celik, I. Akgun and K. Cicek, “Prediction of human error probabilities in a critical marine engineering operation on-board chemical tanker ship: The case of ship bunkering,” vol. 110, no. p102-109, 2018.

[134] A. Gausdal and J. Makarova, “Trust and Safety onboard,” vol. 16, no. 2, 2017.

[135] Y. Xi, Q. Fang, W. Chen and S. Hu, “Case-based HFACS for collecting, classifying and analzing human error in marine accidents,” vol. 1, no. 1, 2009.

[136] S. Jakovlev, A. Andziulis, A. Daranda, M. Voznak and T. Eglynas, “Research on ship autonomous steering control for short-sea shipping problems,” vol. 32, no. 2, 2017.

[137] S. M. Heike, E. Paul and M. Schulz, “Genomics, Circuits, and Pathways in Clinical Neuropsychiatry - Dorsolateral pre-frontal cortex,” vol. 5, no. 1, 2016.

[138] C. Xu, B. Signe and R. Allan, “Functional Near Infrared Spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics,” vol. 15, no. 2, 2009.

[139] M. P. van den Heuvel and O. Sporns, “Network hubs in the human brain,” *tends cogn, sci,* vol. doi: 10.1016/j.tics.2013.09.012, pp. 17, 683-696, 2013.

[140] L. Wenlin and Z. Haiming, “The Relationship between Principal Component Analysis and Factor Analysis and SPSS Software,” vol. 15, no. 3, 2018.

[141] I. M. Organisation, Training regulations and requirements - At sea - Engine room, London, UK: IMO, 2019.

[142] C. Faisy, C. Davagnar, D. Ladiray, J. Djadi-Prat, M. Esvan, E. Lenain, P. Durieux, J. Leforestier, C. Marlet, M. Seijo and A. Guillou, “Nurse workload and inexperienced medical staff members are associated with seasonal peaks in severe adverse events in the adult medical intensive care unit: A seven-year prospective study,” vol. 62, no. 11, Oct 2016.

[143] M. Kaushik, “How work is done onboard a ship,” Guidelines, March 2017. [Online]. Available: https://www.marineinsight.com/guidelines/how-maintenance-work-is-done-onboard-a-ship/.

[144] L. Monica, L. Margareta, R. Leif and D. Joakim, “Working conditions in the engine department – A qualitative study among engine room personnel on board Swedish merchant ships,” vol. 47, no. 2, 2011.

[145] M. Plichta, A. Gerdes, G. Alpers, W. Harnish, S. Brill, M. Wieser and A. Fallgatter, “NeuroImage - Auditory cortex activation is modulated by emotion: A fNIRS study,” vol. 55, no. 3, 2011.

[146] K.-S. Hong, N. Naseer and Y.-H. Kim, “Classification of prefrontal and motor coretx signals for 3 class fNIRS-BCI.,” *Neurisci and Lett,* vol. 5, no. doi: 10.1016/j.neulet.2014.12.029, pp. 87-92, 2015.

[147] L. Holper and M. Wolf, “. Single-trial classification of motor imagery differing in task complexity: a functional near-infrared spectroscopy study.,” *Neuroengineering and Rehabilitation,* pp. doi: 10.1186/1743-0003-8-34, 2012.

[148] N. Naseer, F. M. Noori, N. K. Qureshi and K.-S. Hong, “Determining optimal feature-combination for lda classification of functional near-infrared spectroscopy signals in brain-computer interface application.,” Frontiers, Human Neuroscience, doi: 10.3389/fnhum.2016.00237., 2016.

[149] M. J. Khan and K.-S. Hong, “. Passive bci based on drowsiness detection: an fNIRS study,” *Biomedical,* vol. doi: 10.1364/BOE.6.004063, pp. 6, 4063-4078, 2015.

[150] M. Siegel, T. H. Donner and A. K. Engel, “spectral fingerprints of large-scale neuronal interactions.,” *Nat. Rev. Neurosci,* vol. 4, no. 1, p. 13.121, 2012.

[151] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun and M. Corbetta, “Dynamic functional connectivity: promise, issues, and interpretations,” *Neuroimage,* vol. doi: 10.1016/j.neuroimage.2013.05.079, pp. 80, 360-378, 2015.

[152] X. Cui, D. M. Bryant and A. L. Reiss, “NIRs-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation.,” *Neuroimage,* vol. doi: 10.1016/j.neuroimage.2011.09.003, pp. 59, 2430-2437, 2012.

[153] D. R. Leff, D. R. James, F. Orihuela-Espina, K.-W. Kwok, L. W. Sun and G. Mylonas, “The impact of expert visual guidance on trainee visual search strategy, visual attention and motor skills,” *Frontiers. Hum. Neurosci,* vol. doi: 10.3389/fnhum.2015.00526, pp. 9, 526, 2015.

[154] E. Isbilir, M. P. Çakir, F. Cummins and H. Ayaz, “Investigating brain-brain interactions of a dyad using fnir hyperscanning during joint sentence,” in *3rd International Symposium on Brain and cognitive Science*, Istanbul, 2016.

[155] M. Lowe, B. Mock and J. Sorenson, “Functional connectivity in single and multislice echoplanar imaging using resting-state fluctuations,” *Neuroimage 7,* vol. doi: 10.1006/nimg.1997.0315, pp. 119-132, 1998.

[156] J. Xiong, L. M. Parsons, J.-H. Gao and P. T. Fox, “Interregionial connectivity to primary motor cortex revealed using mri resting state images,” *Human brain map,* pp. 7, 151-156, 1999.

[157] H. Obrig, M. Neufang, R. Wenzel, M. Kohl, J. Steinbrink and K. Einhäup, “Spontaneous low frequency oscillations of cerebral hemodynamics andc metabolism in human adults,” *Neuroimage,* vol. doi: 10.1006/nimg.2000.0657, pp. 12, 623-639, 2000.

[158] X. Zhang, J. A. Noah and J. Hirsch, “Separation of the global and local components in functional near-infrared spectroscopy signals using principal component spatial filtering.,” *Neurophotonics,* vol. doi: 10.1117/1.NPh.3.1.015004, p. 154, 2016.

[159] S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer and K. Müller, “Enhanced performance by a hybrid nirs–eeg brain computer interface.,” *Neuroimage 59,* vol. doi: 10.1016/j.neuroimage.2011.07.084, pp. 519-529, 2012.

Appendices

Appendix A – TRANSAS Manual

ERS 5000 TechSim

MAN B&W 6S60MC-C Diesel Engine – Tanker LCC (Aframax)

Trainee Manual

Issue Date: December 2018

LIVERPOOL JOHN MOORES UNIVERSITY

Contents

1- Introduction

1.1 - Abbreviations

ACB	Air Circuit Breaker
AHU	Air Handling Unit
AOP	Additional Operator Panel
BMCS	Boiler Monitoring & Control System
BW	Bilge Well
C.F.W. (CFW)	Cooling Fresh Water
C.S.W. (CSW)	Cooling Sea Water
C/R	Control Room
CB	Circuit Breaker
CCC	Cargo Control Console
CCR	Cargo Control Room
CMS	Control Monitoring System
C.O.P.T. (COPT)	Cargo Oil Pump Turbine
DGU	DEIF Generator Unit
DU	Display Unit
E/G (EG) (EmG)	Emergency Generator
E/R (ER)	Engine Room
E/S	Engine Side
ECC	Engine Control Console
ECR	Engine Control Room
EMCY (EM'CY)	Emergency
ESB	Emergency Switch Board
EXH	Exhaust
F.O. (FO)	Fuel Oil
F.W. (FW)	Fresh Water
F/E	Finishing with Engine
FF	Fire Fighting
FP	Feeder Panel
FPP	Fixed Pitch Propeller
G/E (GE)	Generator Engine
GB	Generator Breaker
GPBP	Group Push Button Panel
GSP	Group Starter Panel
H.F.O. (HFO)	Heavy Fuel Oil
H.T. (HT)	High Temperature
HPP	Hydraulic Power Pack
HPU	Hydraulic Power Unit
I.G.G. (IGG)	Inert Gas Generator
J.W. (JW)	Jacket Water
L.O. (LO)	Lube Oil
L.P. (LP)	Low Pressure
L.S. (LS)	Low Sulfur

L.S.H.F.O. (LSHFO)	Low Sulfur Heavy Fuel Oil
L.S.M.D.O. (LSMDO)	Low Sulfur Marine Diesel Oil
L/T (LT, L.T.)	Low Temperature
LAH	Level Alarm High
LAL	Level Alarm Low
LCC	Large Crude (oil) Carrier
LCP	Local Control Panel
LGSP	Local Group Starter Panel
LIAH	Temperature Indicator Level High
LOP	Local Operating Panel
D.O. (MDO)	Marine Diesel Oil
M.G.O. (MGO)	Marine Gas Oil
M.G.P.S.	Marine Growth Prevention System
M/E (ME)	Main Engine

MCD	Main Circuit Diagram
MSB (MSBD, MSWB)	Main Switch Board
O.W.S. (OWS)	Oily Water Separator
P.C.O. (PCO)	Piston Cooling Oil
P/P (PP)	Pump(s)
PAH	Pressure Alarm High
PAL	Pressure Alarm Low
PB	Push Button
PD DB (PDB)	Power Distribution Board
PMS	Power Management System
RCS	Remote Control System
S.W. (SW)	Sea Water
S/G (SG)	Steering Gear / Shaft Generator
ShG	Shaft Generator
S/T	Stern Tube
SC	Sea Chest
STP	Sewage Treatment Plant
T/C (TC)	Turbo compressor
TAH	Temperature Alarm High
TAL	Temperature Alarm Low
T/G (TG	Turbo Generator
TI	Temperature Indication
TIAH	Temperature Indicator Alarm High
TK	Tank
VIT	Variable Injection Timing
W	Water
W.B.P.T. (WBPT)	Water Ballast Pump Turbine

1.2 - The Ship

The simulator is modeling the Propulsion Plant, Electric Power Plant, Control Monitoring System (CMS), auxiliary systems, equipment, units and mechanisms of a general Tanker LCC (Large Crude Oil Carrier). The prototype for development is Aframax Tanker 115,000 DWT.

Ship general characteristics:

Max. continuous rating (MCR)	736 kW at 105 RPM
Normal continuous rating (85% of MCR	364 kW at 101.4 RPM
Length overall	248.92 meters
Breadth, moulded	43.8 meters
Designed draft, moulded	14.925 meters
Service speed	15.5 knots

1.3 - Simulator console button bars

The pages (buttons on the button bar) of the Propulsion console are described in brief below:

- Displays of the page **BCC** model the Bridge control console panels;
- Displays of the page **ECR** model the Engine Control Room control console panels;
- Displays of the page **MSB** model the Electrical Power Plant Main Switchboard control panels;
- Displays of the page **CMS** model Control Monitoring System remote control displays on the ECR desk;
- Displays of the page **BMCS** model Boiler Monitoring & Control System remote control displays on the ECR desk;
- Displays of the page **Diag** model the diagnostic Cylinder Indicator Diagrams of the Main propulsion and Diesel engines of the generators, and combustion process;
- Displays of the page **SYS** contain mimics of the ship's systems. They model manual remote and local control of the respective units and mechanisms.
- The following pages model equipment control in the ship engine rooms:

SG – Steering Gear room page; **ER1** – Engine Room 1 page; **ER2** – Engine Room 2 (Deck 2) page; **ER3** – Engine Room 3 (Upper Deck) page; **ER4** – Engine Room 4 (Deck A) page; **FFR** – Fire Fighting Room page. **EmG** – Emergency Generator room page; **CCR** – Cargo Control Room page.

The displays of a page menu contain mimics of the Local Operating Panels (LOPs) of the units and mechanisms, switchboards (SWBDs), power distribution boards (PDBs), group starter panels (GSP), etc. and the 3-D pictures of the engine rooms where applicable.

The pages of the Virtual Hardware console are:

- **GSP1** – No. 1 Group Starter Panels;
- **GSP2** – No. 2 Group Starter Panels;
- **CCP** – Cargo Control Panel;
- **EG** – EM'CY Generator Engine panel;
- **ESB** – Emergency Switchboard and Shore connection panels;
- **G1** – Generator 1 upper and lower panels;
- **G2** – Generator 2 upper and lower panels;
- **FP1** – MSB No. 1 Feeder AC440V upper and lower panels;
- **FP2** – MSB No. 2 Feeder AC440V upper and lower panels;
- **BUS** – Bus Tie panel;
- **ShG** – MSB Shaft Generator upper and lower panels;
- **Syn** – MSB Synchro upper and lower panels;
- **TG** – MSB Turbo Generator upper and lower panels;
- **ECR** – Engine Control Room panels;
- **220v** – MSB Feeder AC220V upper and lower panels;
- **Full** – unified system diagram (as video wall).

Figure 0-1 - Bottom bar of propulsion console

Figure 0-2 - Bottom bar of the virtual hardware console

1.4 - LCS software

The Loading Control System (LCS) software kit LCS 7957
LCC Tanker - 'Initial' is designed for calculating and controlling the vessel's loading, trim and stability in the course of its operation, as well as for:

- Calculating the vessel loading by entering data on the ship's provisions, liquid cargoes, water ballast and cargoes carried; saving the input loading data in the computer memory; LCS help system allows the trainee to familiarize with the system and its options;
- Estimating an intact vessel trim, stability and longitudinal strength data, comparing it with the allowed values.

Note: The LCS and the task model would always both run on the same computer.

During the execution of an exercise the simulator model communicates with the LCS and sends the information about the volumes and density of the liquids in cargo and ballast tanks. Using this data the LCS calculates the vessel loading, trim and stability parameters. These values are displayed in the LCS dedicated windows and are also displayed on the CMS panels of the simulator.

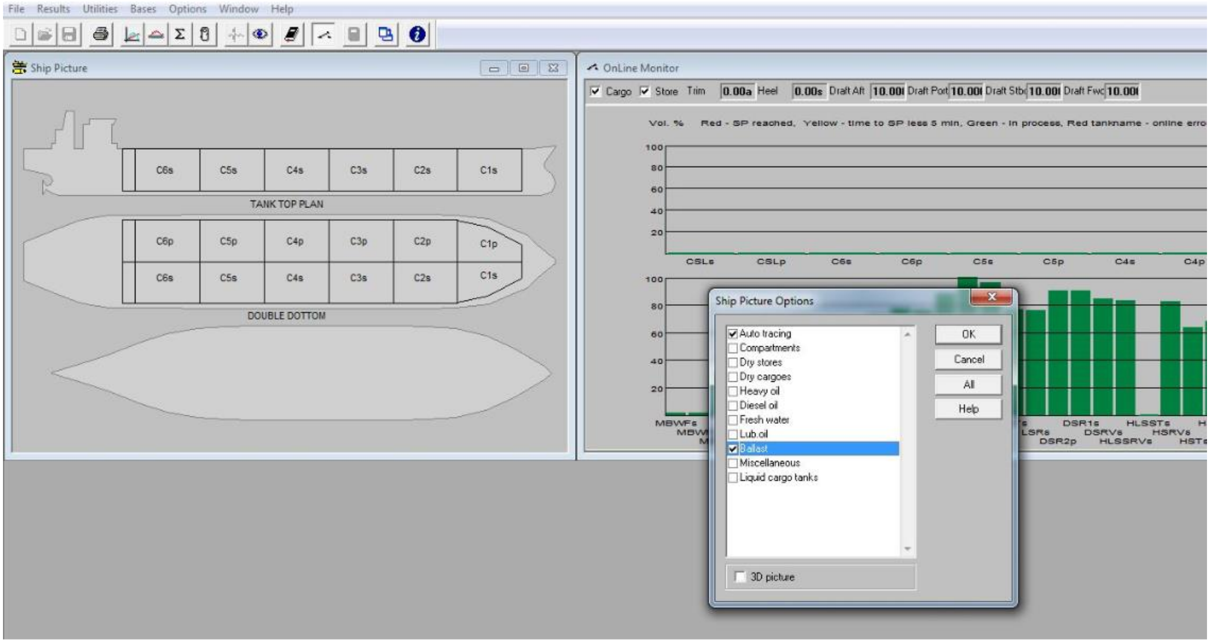
LCS can be switched between Online and Standard modes.

- In Online mode, LCS constantly receives the actual values of Volume, Density and Temperature for cargo tanks.
- In Standard mode the parameters of cargo tanks and store tanks can be edited manually.

Note : When the simulator is running, the LCS must be in Online Mode. Switch to Standard mode only when the exercise is paused.

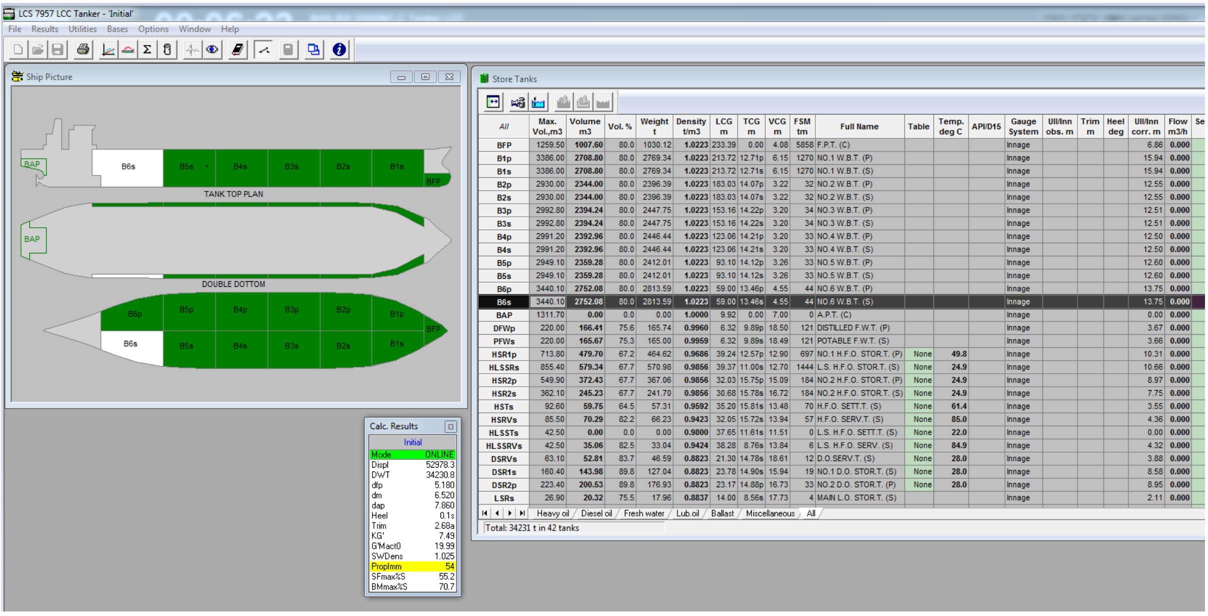
Note : In Online mode, make sure, that the Product column in the Cargo Tanks window is empty for each of cargo tanks. This is required for correct displaying of actual values.

1.4.1 LCS system Monitoring (ballast tanks)

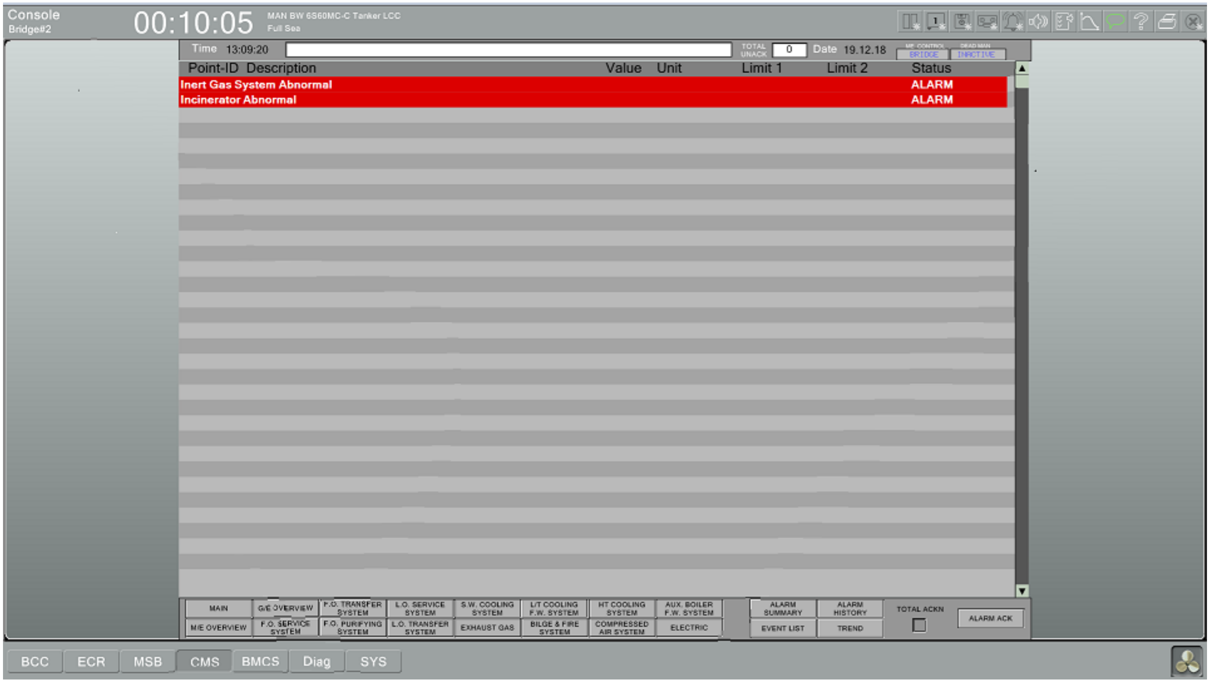


When monitoring the ballast tanks in order to achieve the desired volume please refer to the pictures above and below. To load the correct screen in order to monitor the ballast tank volume please follow the instructions below;

- Right click anywhere on the 'ship picture' as seen above. A 'Ship picture options' box will appear.
- Select the Ballast box and click ok.
- Then select 'options' at the top of the screen, then mode, the standard.
- Then select the tank that you want to monitor (highlighted in green as seen below).
- Re-select 'options', then mode, the online.



1.5 Alarm System

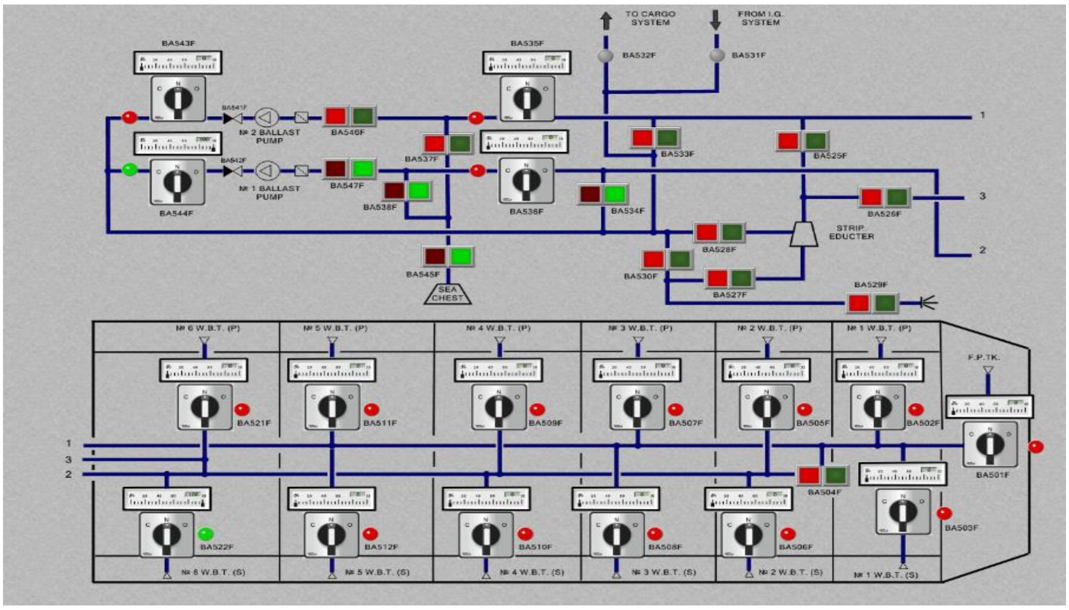


When an alarm is raised a loud siren will be head along with the alarm Icon at the top right hand side of the screen above will flash read (this is the alarm bell icon). When you hear and see the ship alarm you can see the system that needs attention by following the instructions below;

- Click on the CMS icon on the bottom left hand screen (as shown above).
- Click on the alarm summary icon at the bottom of the CMS screen.
- Check the POINT-ID description to locate the source of the problem.

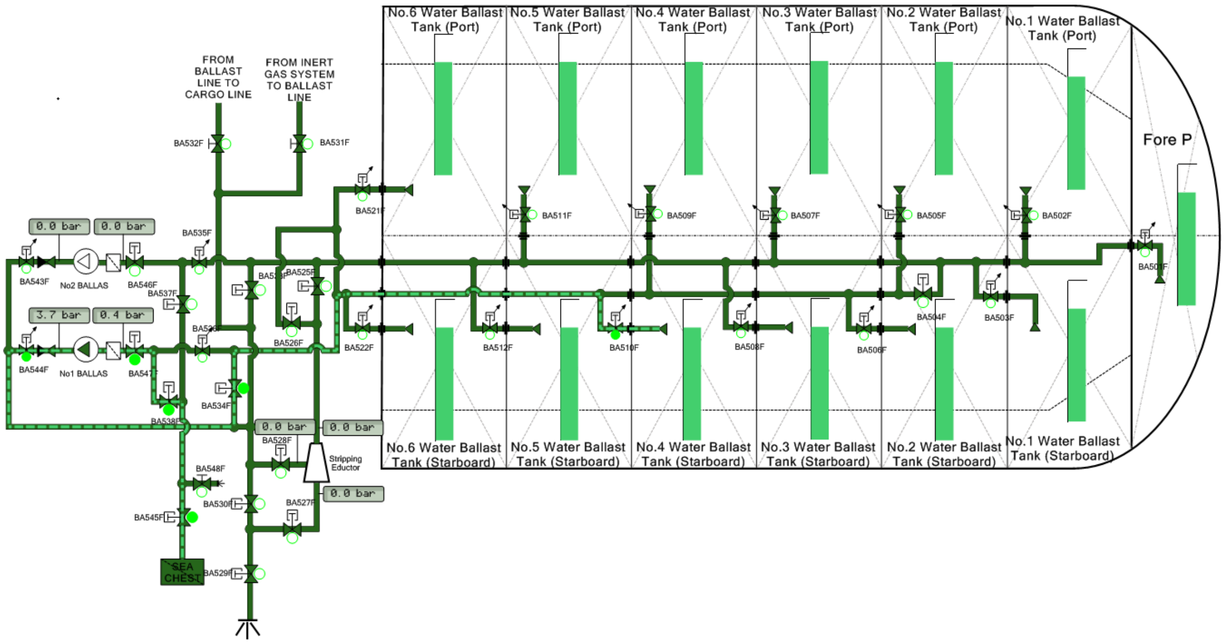
1.6 Ballast System

In order to start ballasting follow the instructions below;



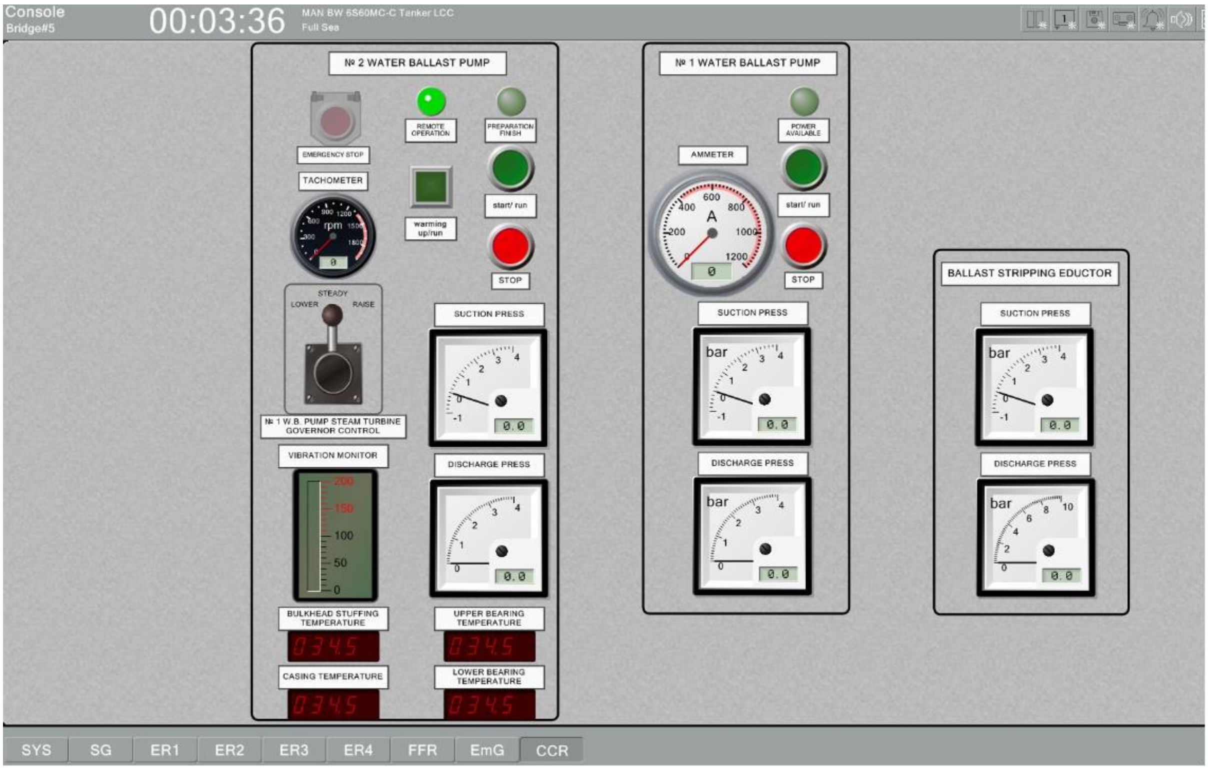
- Select 'CCR' icon and then ballast mimic control panel. It is also advised to have the ballast system on the left hand screen in order to monitor the water flow.

- Open valves on the desired chosen water line by clicking and holding on the green buttons as shown above.
- Open the ball values by clicking and holding open (o), until the value is at 100% open.



When all the valves are open to the desired ballast tank you can turn the pump on by;

- Select the cargo control room icon (CCR) at the bottom of the screen.
- Select Ballast pump control then the below screen will appear.
- Click and hold on the green start button for either pump 1 or pump 2. (You will notice the ammeter value will increase when the pump is on).
- The system will then start ballasting.
- The water flow is shown by the water line turning green and moving in the direction towards the ballast tanks as shown above.



Appendix B – R-studio code

```
library(MASS)

######Upload data######

dataDistraction <- read.csv(file.choose(), header=TRUE)

dataWorkload <- read.csv(file.choose(), header=TRUE)

dataTraining <- read.csv(file.choose(), header=TRUE)

dataFatigue <- read.csv(file.choose(), header=TRUE)

####LDA analysis on data#####
```

Distraction

```
dataDistraction.lda <- lda( Group ~ Channel.1 + Channel.2 + Channel.3 + Channel.4 + Channel.5 +
Channel.11 + Channel.12 + Channel.13 + Channel.14 + Channel.15,

data=(dataDistraction)



dataDistraction.lda.p <- predict(dataDistraction.lda,
 newdata=dataDistraction[,c(2,3,4,5,6,7,8,9,10,11)])$class
```

Workload

```
dataWorkload.lda <- lda( Group ~ Channel.1 + Channel.2 + Channel.3 + Channel.4 + Channel.5 +
Channel.11 + Channel.12 + Channel.13 + Channel.14 + Channel.15,
                  data=dataWorkload)



dataWorkload.lda.p <- predict(dataWorkload.lda,
                     newdata=dataWorkload[,c(2,3,4,5,6,7,8,9,10,11)]
)$class
```

Training

```
dataTraining.lda <- lda( Group ~ Channel.1 + Channel.2 + Channel.3 + Channel.4 + Channel.5 +
Channel.11 + Channel.12 + Channel.13 + Channel.14 + Channel.15,
                  data=dataTraining)



dataTraining.lda.p <- predict(dataTraining.lda,
                     newdata=dataTraining[,c(2,3,4,5,6,7,8,9,10,11)]
)$class
```

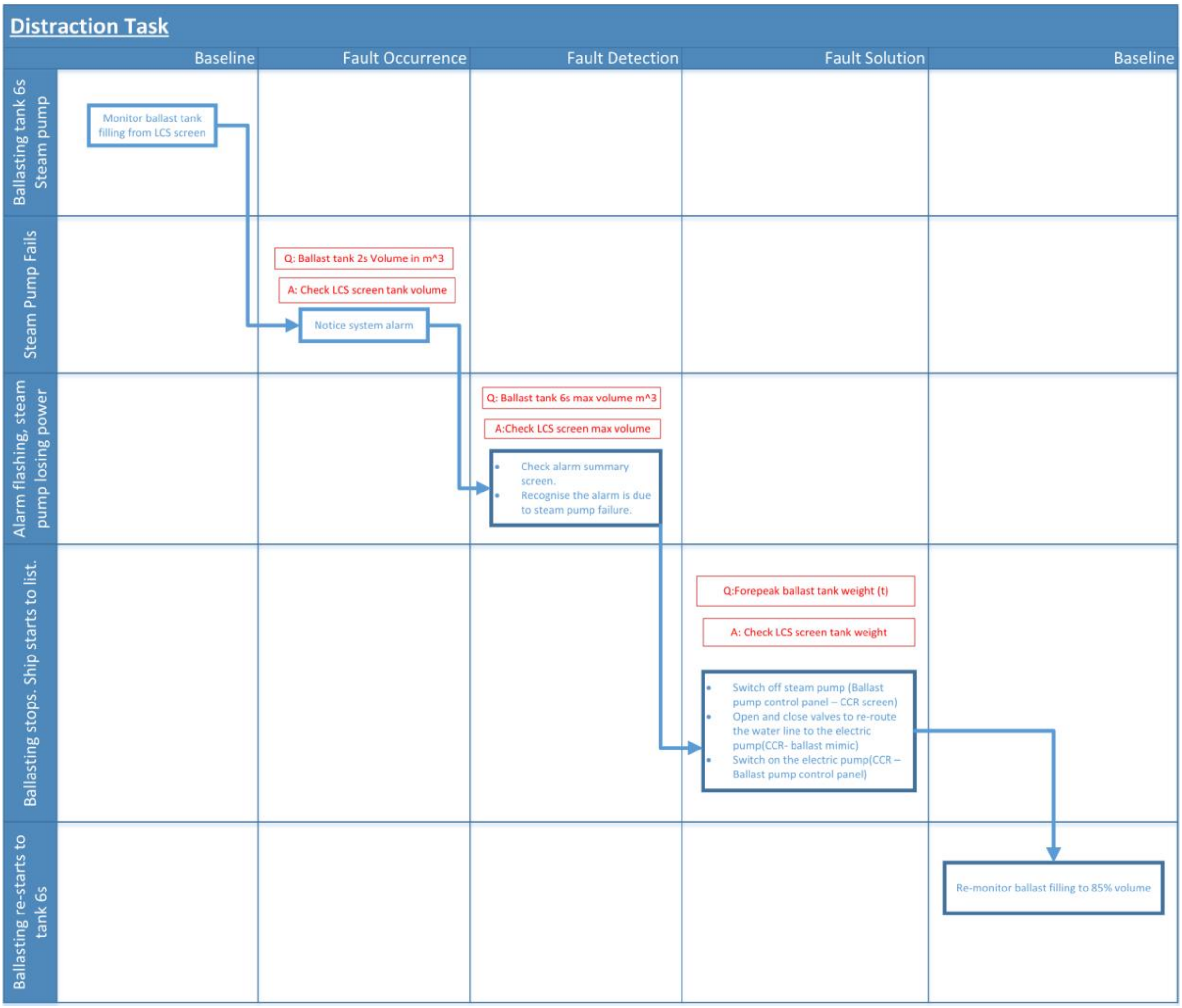
Fatigue

```
dataFatigue.lda <- lda( Group ~ Channel.1 + Channel.2 + Channel.3 + Channel.4 + Channel.5 +
Channel.11 + Channel.12 + Channel.13 + Channel.14 + Channel.15,
                      data=dataFatigue)



dataFatigue.lda.p <- predict(dataFatigue.lda,
                          newdata=dataFatigue[,c(2,3,4,5,6,7,8,9,10,11)]
)$class



###determine how well the model fits#####


###Distraction####

table(dataDistraction.lda.p, dataDistraction[,1])


###Workload###

table(dataWorkload.lda.p, dataWorkload[,1])


###Training###

table(dataTraining.lda.p, dataTraining[,1])


###Fatigue###

Table(dataFatigue.lda.p, dataFatigue[,1])


#####Cross Validate the model#####


###Distractio###

dataDistraction.lda.2 <- lda( Group ~ Channel.1 + Channel.2 + Channel.3 + Channel.4 + Channel.5 +
Channel.11 + Channel.12 + Channel.13 + Channel.14 + Channel.15,
                data=dataDistraction,
```

```
            CV = TRUE)

###Workload###

dataWorkload.lda.2 <- lda( Group ~ Channel.1 + Channel.2 + Channel.3 + Channel.4 + Channel.5 +
Channel.11 + Channel.12 + Channel.13 + Channel.14 + Channel.15,

                  data=dataWorkload,

                  CV = TRUE)

###Training###

dataTraining.lda.2 <- lda( Group ~ Channel.1 + Channel.2 + Channel.3 + Channel.4 + Channel.5 +
Channel.11 + Channel.12 + Channel.13 + Channel.14 + Channel.15,

                  data=dataTraining,

                  CV = TRUE)

###Fatigue###

dataFatigue.lda.2 <- lda( Group ~ Channel.1 + Channel.2 + Channel.3 + Channel.4 + Channel.5 +
Channel.11 + Channel.12 + Channel.13 + Channel.14 + Channel.15,

                  data=dataFatigue,

                  CV = TRUE)


###look at the assigned classes for the observation#####


###Distraction###

table(dataDistraction.lda.2$class, dataDistraction[,1])


###Workload###

table(dataWorkload.lda.2$class, dataWorkload[,1])


###Training###

table(dataTraining.lda.2$class, dataTraining[,1])


###Fatigue###

table(dataFatigue.lda.2$class, dataFatigue[,1])
```

```
######Results######################################################################
######

###Distraction###
##### Predict for A = % accurite #####
##### predict for B = % accurite #####

###Workload###
##### predict for A = % accurite ####
##### predict for B = % accurite ####

###Fatigue###
##### predict for A = % accurite ####
##### predict for B = % accurite ####

###Training###
##### predict for A = % accurite ####
##### predict for B = % accurite ####
```

Appendix C – Distraction task questions and point at which they are asked



Appendix D – Full SPSS Analysis of data from all studies.

The data below includes every analysis conducted for all studies. The majority of which was deemed not relevant/overkill for the final thesis.

Distraction

Time data

The relevant areas of the workflow evaluated in this test are; fault occurrence (1), fault detection (2), and fault solution (3)(x-axis) against time (y-axis).

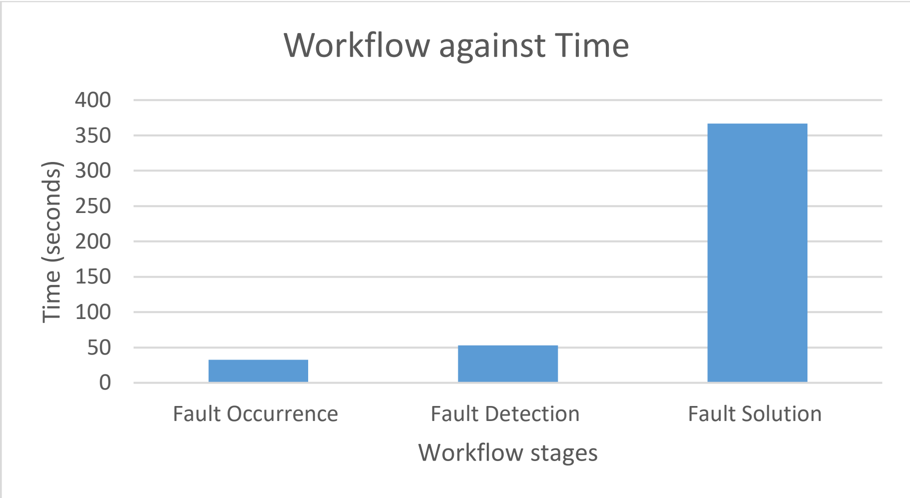


Figure - workflow stages with respect to time

As expected, a significant effect was found between workflow stages[$F(2,17) = 12544.14$, $P<0.01$, eta^2 = 0.99] as depicted in the figure above.

Pairwise comparisons show significant effects between all workflow stages (1=Fault occurrence, 2= fault detection & 3=fault solution) $p<0.01$ as depicted below in table 7. Maulchy's test of sphericity showed a significant effect therefore tests of within subject's effects using Greenhouse-Geisser showed [$F=23957.7$, $p<0.001$, eta^2 = 0.99].

Pairwise Comparisons

Measure: MEASURE_1

(I) workflow	(J) workflow	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-20.775*	.498	.000	-22.089	-19.461
	3	-334.425*	2.069	.000	-339.886	-328.964
2	1	20.775*	.498	.000	19.461	22.089
	3	-313.650*	2.065	.000	-319.099	-308.201
3	1	334.425*	2.069	.000	328.964	339.886
	2	313.650*	2.065	.000	308.201	319.099

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of workflow stages

Due to the significant effects shown in the above pairwise comparisons t-test were conducted in order for us to see the size of the effect. The fault solution stage was used as Level (i) for all tests as this stage has the most significance.

- Fault solution (level i) vs Fault occurrence (level j) t = 334.425/2.069 = 161.63, p<0.001
- Fault solution (level i) vs Fault detection (level j) t = 313.650/2.065 = 151.89, p<0.001

Effect of distraction with respect to time

The ANOVA showed a significant interaction between distraction and workflow with respect to time [F(1,18) = 95.982, P<0.01, eta^2 = 0.919. A significant interaction was found in tests within subjects effects using Greenhouse-Geisser [F=118.8, P<0.001, eta^2 = 0.87].

HBO Data

The original tests showed no significant effects when evaluating data from the middle region of the DLPFC. Therefore, the middle region data was omitted for the second test.

Workflow (left and right side DLPFC)

The ANOVA showed a significant effect between workflow stages[F(3,16) = 16.195, P<0.01, eta^2 = 0.752]. Test within subjects effects showed [F(3,16) = 14.428 P<0.01, eta^2 = 0.650]. This is depicted in the figure below.

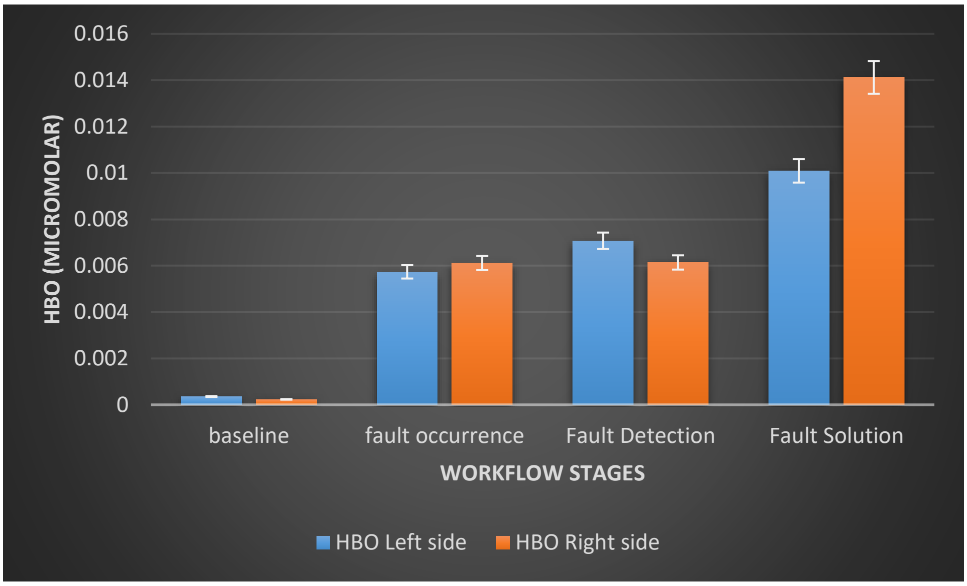


Figure- Workflow stages with respect to HBO

Pairwise comparisons found with a 95% confidence ($P<0.05$) significant main effects between workflow stage 4 (the fault solution stage) and all other stages (baseline $p<0.01$, Fault occurrence $p<0.01$ and Fault detection $p<0.01$) as depicted in the table below:

Pairwise Comparisons

Measure: MEASURE_1

(I) workflow	(J) workflow	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-.004*	.001	.000	-.006	-.002
	3	-.005*	.001	.000	-.007	-.002
	4	-.011*	.001	.000	-.015	-.006
2	1	.004*	.001	.000	.002	.006
	3	-.001	.000	.486	-.002	.001
	4	-.007*	.001	.000	-.010	-.003
3	1	.005*	.001	.000	.002	.007
	2	.001	.000	.486	-.001	.002
	4	-.006*	.001	.001	-.009	-.002
4	1	.011*	.001	.000	.006	.015
	2	.007*	.001	.000	.003	.010
	3	.006*	.001	.001	.002	.009

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table-Pairwise comparisons of workflow stage

Due to the significant effects shown above, paired t-tests were conducted to see the size of the effect. Level (i) remained the fault solution stage, level (j) varied between workflow stages.

- Fault solution (level i) vs Baseline (level j) t = 0.011/0.001 = 11, p<0.001, effect size (r) = 0.7
- Fault Solution (level i) vs Fault Occurrence (level j) t = 0.007/0.001 = 7, p<0.001, r =0.35
- Fault Solution (level i) vs Fault Detection (level j) t = 0.006/0.001 = 6, p=0.001, r = 0.2

Distraction (left and right side DLPFC)

The ANOVA showed no significant interaction for workflow and distraction with respect to HBO[F(3,16) = 0.479, P<0.05, eta^2 = 0.082].

Region of Interest (left, right, middle DLPFC)

The region of interest is the left, middle and right sides of the dorsal lateral pre-frontal cortex. Each region contains 5 channels as depicted in chapter 5.2.

The ANOVA revealed a significant interaction between ROI and workflow [F(6,13)=5.93, p<.01, eta2=.73]. Test within subjects showed no significant effects [F=2.5, p=0.08, eta^2 = 0.123] The main effects for ROI was insignificant [F(2,17)=1.52, p=.25]. This is depicted in the figure below.

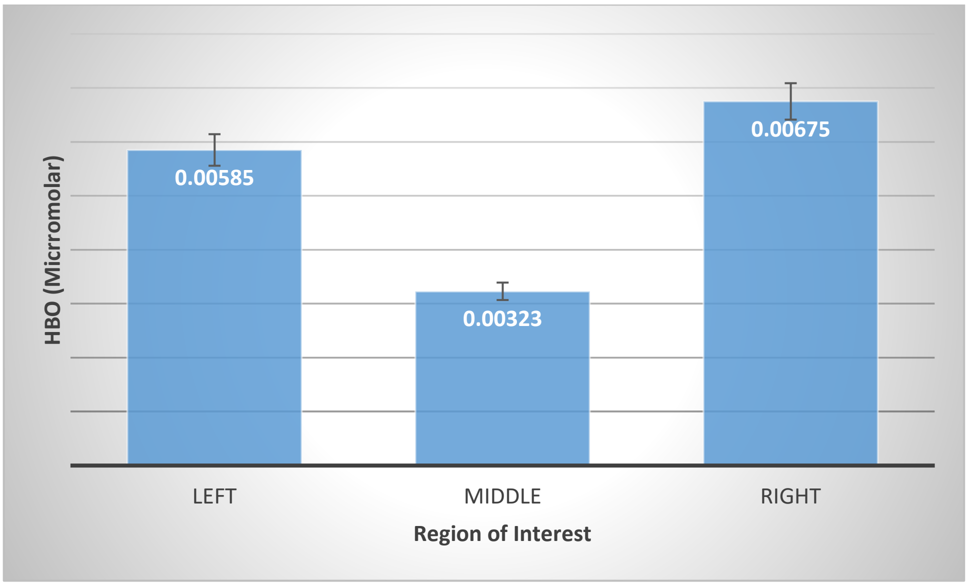


Figure- Region of interest with respect to HBO

Increased Workload

Time data

The relevant areas of the workflow evaluated in the study were; fault occurrence (1), fault detection (2), and fault solution (3)(x-axis) against time (y-axis).

The bar chart below shows all 20 participants partaking an increased workload test. Half are distracted and half are not distracted. From the mean time taken for all participants, the fault solution stage of the workflow took the longest to complete as shown in the figure below.

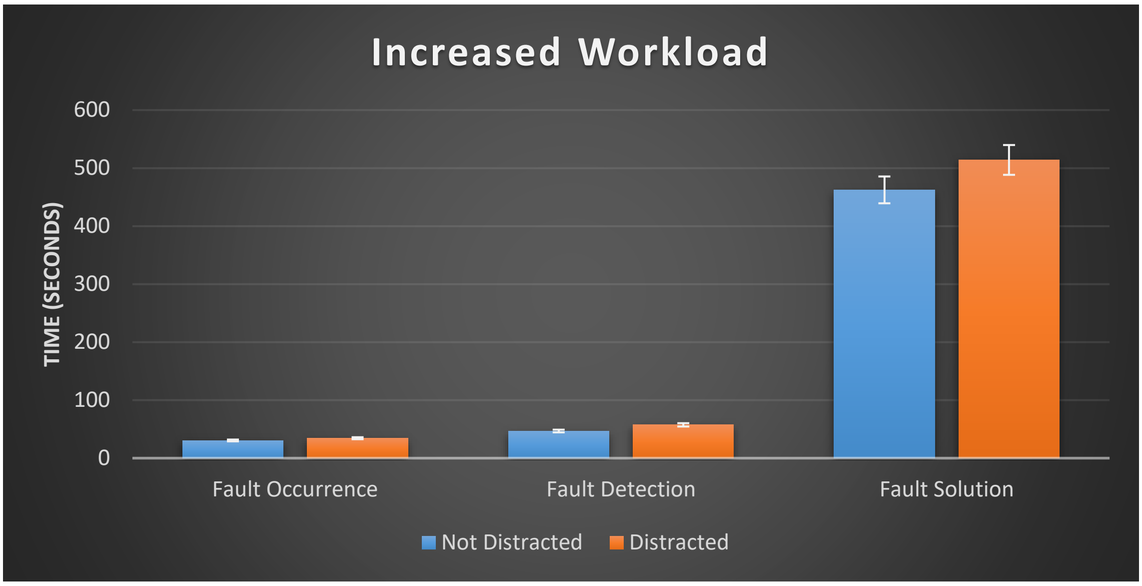


Figure - workflow stage with respect to time.

Below, the figure shows a comparison of the mean time between 10 participants in standard workload group and 10 participants in an increased workload but undistracted group. This allows us to see the effect of increased workload alone.

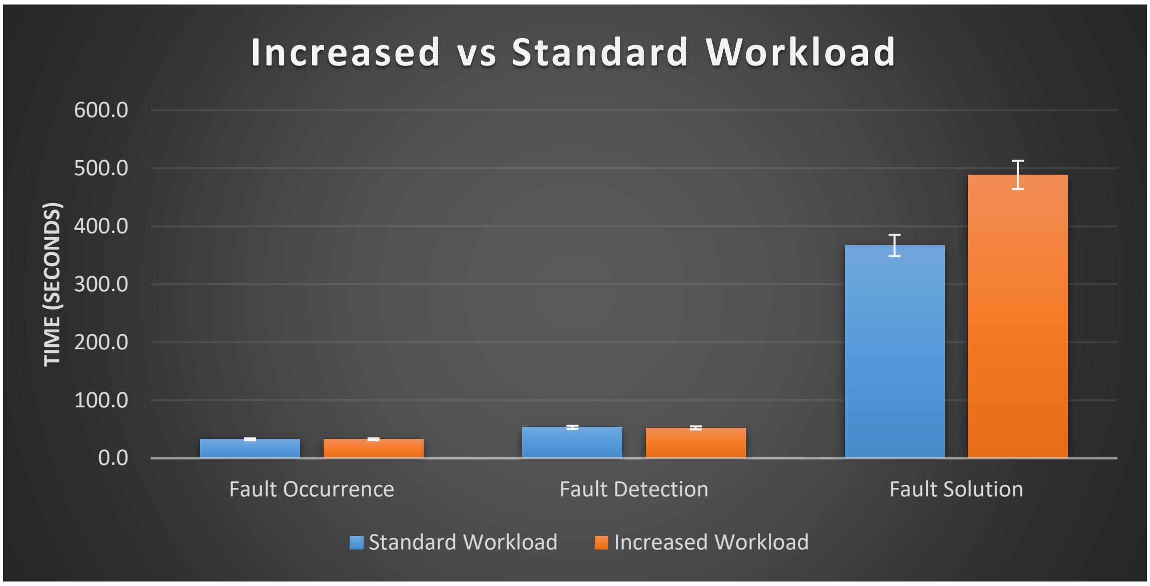


Figure - Time taken for workflow stage with respect to workload.

Effects of Increased workload on distraction with respect to time

A significant effect was found for distracted participants with an increased workload against not distracted participants with an increased workload [$F(2,17) = 8564$, $P<0.01$]. The mean times for participants in the increased workload*distracted group were higher during fault occurrence, fault detection and fault solution, when compared to those participating in the increased workload*not distracted group as shown in the table below.

		Number of Participants	Mean Time (seconds)
Fault Occurrence	Distracted	10	30.51
	Not Distracted	10	34.35
Fault Detection	Distracted	10	57.4
	Not Distracted	10	46.75
Fault Solution	Distracted	10	514.02
	Not Distracted	10	462.4

Table - mean time taken for workflow stage with respect to distraction

Pairwise comparisons of workflow using 20 Increased workload participants (10 distracted, 10 not distracted) in table 12 below show significant effects between all workflow stages (1=Fault occurrence, 2= fault detection & 3=fault solution) p<0.01 as depicted below in table. Maulchy's test of sphericity showed a significant effect therefore tests of within subject's effects using Greenhouse-Geisser showed [F=14884.2, p<0.001, eta^2 = 0.99].

Pairwise Comparisons

Measure: MEASURE_1

(I) workflow	(J) workflow	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-19.645*	.606	.000	-21.245	-18.045
	3	-455.780*	3.580	.000	-465.229	-446.331
2	1	19.645*	.606	.000	18.045	21.245
	3	-436.135*	3.684	.000	-445.859	-426.411
3	1	455.780*	3.580	.000	446.331	465.229
	2	436.135*	3.684	.000	426.411	445.859

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of workflow stage

Due to the significant effects showing in the above pairwise comparisons t-test were conducted in order for us to see the size of the effect. The fault solution stage was used as Level (i) for all tests as this stage has the most significance.

- Fault solution (level i) vs Fault occurrence (level j) t = 455.780/3.580 = 127.31, p<0.001
- Fault solution (level i) vs Fault detection (level j) t = 436.135/3.684 = 118.39, p<0.001

*Effect of Increased workload*distraction*workflow with respect to time*

The ANOVA showed a significant interaction between workload*distraction*workflow with respect to time [F(2,17) = 39.392, P<0.01, eta^2 = 0.99. A significant interaction was found in tests within subject's effects using Greenhouse-Geisser [F=37.439, P<0.001, eta^2 = 0.99].

Effect of Increased workload with respect to time

A significant effect was found for increased workload*not distracted against standard workload*not distracted participants (10 increased workload, 10 standard workload) [F(2,37) = 5801.29, P<0.01, eta^2 = 0.99]. The mean times for participants in the increased workload group were higher during the fault solution stage, when compared to those participating in the standard workload group as shown in the table below.

.

		Number of Participants	Mean Time (seconds)
Fault Occurrence	Standard Workload	20	32.4
	Increased Workload	20	32.43
Fault Detection	Standard Workload	20	53.2
	Increased Workload	20	52.075
Fault Solution	Standard Workload	20	366.8
	Increased Workload	20	488.21

Table - Mean times for workflow stages with respect to workload.

Pairwise comparisons of workflow for 10 increased workload*not distracted and 10 standard workload*not distracted participants in table 14 below show significant effects between all workflow stages (1=Fault occurrence, 2= fault detection & 3=fault solution) p<0.01 as depicted below in table. Maulchy's test of sphericity showed a significant effect therefore tests of within subject's effects using Greenhouse-Geisser showed [F=8404.6, p<0.001, eta^2 = 0.99].

Pairwise Comparisons

Measure: MEASURE_1

(I) workflow	(J) workflow	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-20.210*	.758	.000	-22.109	-18.311
	3	-395.102*	4.433	.000	-406.205	-384.000
2	1	20.210*	.758	.000	18.311	22.109
	3	-374.892*	3.889	.000	-384.632	-365.153
3	1	395.102*	4.433	.000	384.000	406.205
	2	374.892*	3.889	.000	365.153	384.632

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of workflow stages

Due to the significant effects showing in the above pairwise comparisons, t-test were conducted in order for us to see the size of the effect. The fault solution stage was used as Level (i) for all tests as this stage has the most significance.

- Fault solution (level i) vs Fault occurrence (level j) t = 395.102/4.433 = 89.13, p<0.001
- Fault solution (level i) vs Fault detection (level j) t = 374.892/3.889 = 96.4, p<0.001

*Effect of Increased (10 increased workload*not distracted) vs standard workload (10 standard workload*not distracted) *workflow with respect to time*

The ANOVA showed a significant interaction between workload and workflow with respect to time [F(2,37) = 233.782, P<0.01, eta^2 = 0.99. A significant interaction was found in tests within subject's effects using Greenhouse-Geisser [F=210.292, P<0.001, eta^2 = 0.99].

HBO data

*The effect of Increased workload (10 increased workload*not distracted against 10 standard*not distracted) with respect to HBO.*

A significant effect was found for increased workload participants against standard workload participants [F(2,37) = 105.143, P<0.01]. The mean HBO for participants in the Increased workload group were higher only during the fault solution stage, when compared to those participating in the standard workload group as shown in the figure below.

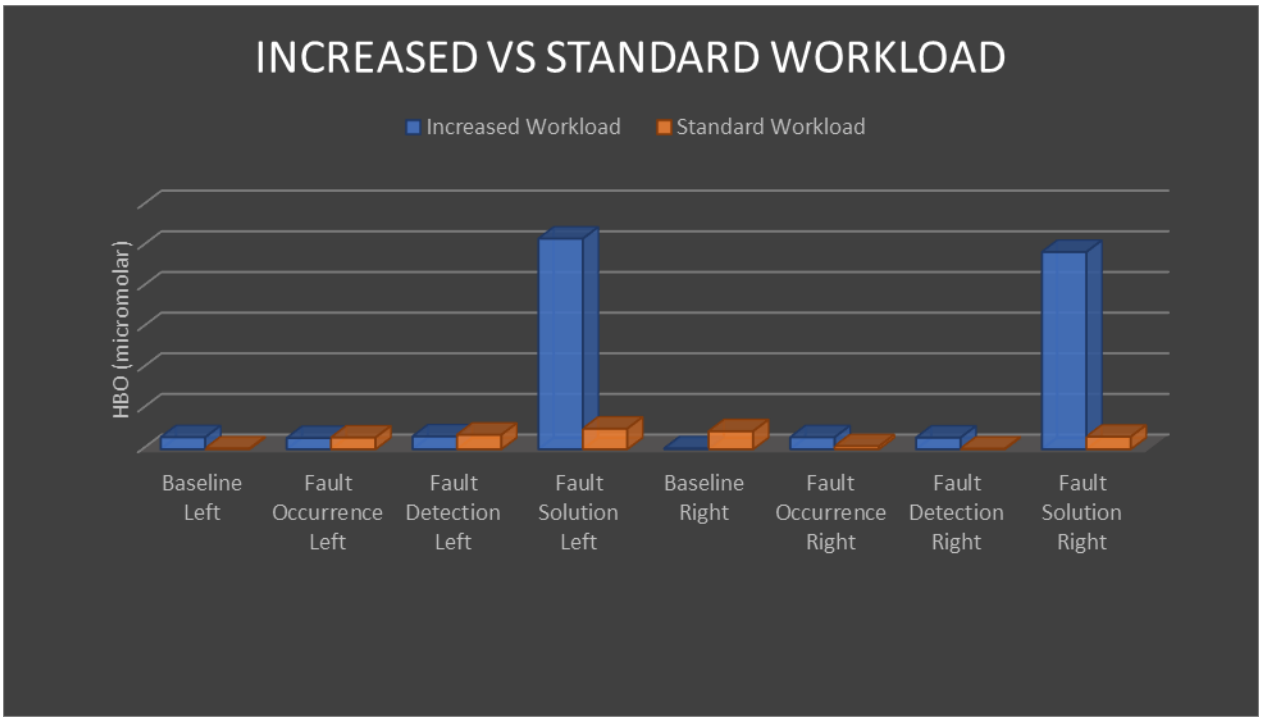


Figure - Workload with respect to workflow stage

The figure below shows the fault solution stage of the workflow for the left and right side DLPFC as this is the area with the most significance.

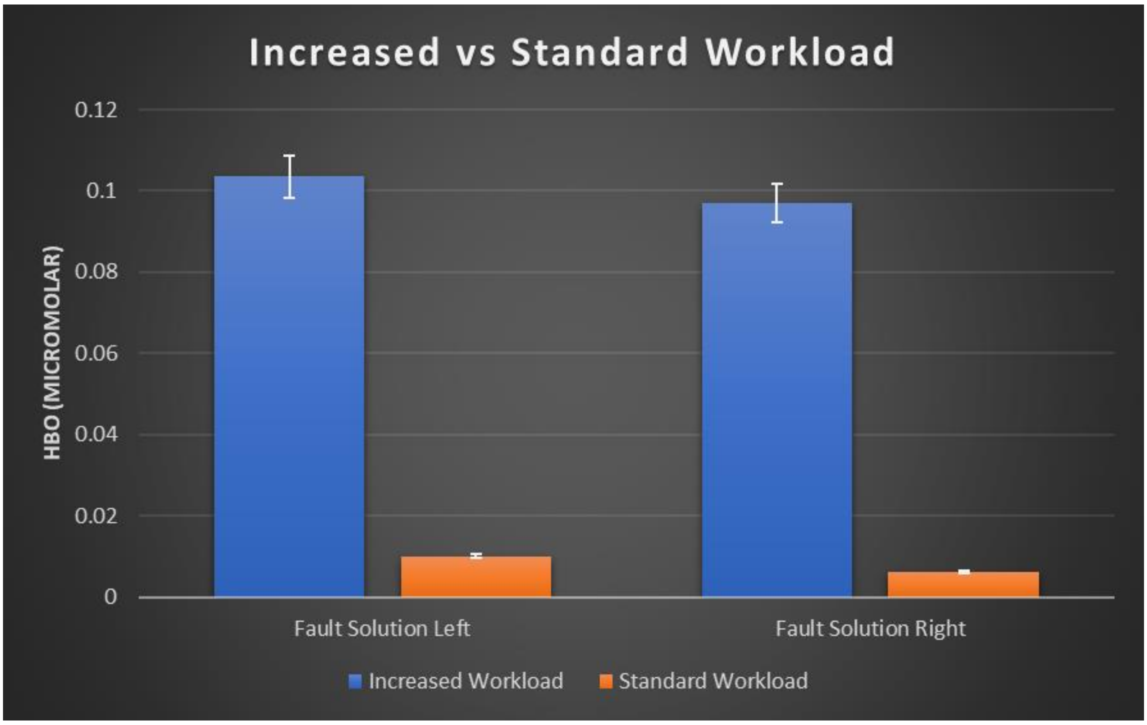


Figure - Workload with respect to the fault solution stage

Pairwise comparisons showed a significant effect for workload shown in the table below.

Pairwise Comparisons

Measure: MEASURE_1

(I) Standard Workload = 1 Increased Workload = 2	(J) Standard Workload = 1 Increased Workload = 2	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-.014*	.001	.000	-.017	-.012
2	1	.014*	.001	.000	.012	.017

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of workload

Due to the significant effects showing in the above pairwise comparisons, t-tests were conducted in order for us to see the size of the effect. Increased workload was used as Level (i) and standard workload level (j).

- Increased Workload (level i) vs Standard workload (level j) t = 0.014/0.001 = 14, p<0.001

*Effect of Increased workload*workflow (10 increased workload*not distracted participants).*

The ANOVA showed a significant interaction between Increased workload and workflow with respect to HBO [F(3,36) = 74.281, P<0.01, eta^2 = 0.99. A significant interaction was found in tests within subject's effects using Greenhouse-Geisser [F=162.177, P<0.001, eta^2 = 0.99].

Pairwise comparisons show significant effects for the fault solution workflow stage p<0.01 as depicted below in the table below.

Pairwise Comparisons

Measure: MEASURE_1

(I) Workflow	(J) Workflow	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.003*	.001	.033	-.005	.000
	3	-.003*	.001	.007	-.006	-.001
	4	-.038*	.002	.000	-.045	-.032
2	1	.003*	.001	.033	.000	.005
	3	-.001	.000	.700	-.002	.000
	4	-.036*	.002	.000	-.041	-.030
3	1	.003*	.001	.007	.001	.006
	2	.001	.000	.700	.000	.002
	4	-.035*	.002	.000	-.041	-.029
4	1	.038*	.002	.000	.032	.045
	2	.036*	.002	.000	.030	.041
	3	.035*	.002	.000	.029	.041

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of workflow stages

Maulchy's test of sphericity showed a significant effect therefore tests of within subject's effects using Greenhouse-Geisser showed [F=262.170, p<0.001, eta^2 = 0.99].

Due to the significant effects shown in the above pairwise comparisons, t-tests were conducted in order for us to see the size of the effect. The fault solution stage was used as Level (i) for all tests as this stage has the only significance.

- Fault solution (level i) vs Baseline (level j) t = 0.038/0.002 = 19, p<0.001
- Fault solution (level i) vs Fault Occurence (level j) t = 0.036/0.002 = 18, p<0.001
- Fault Solution (level i) vs Fault Detection (level j) t = 0.035/0.002 = 17.5, p<0.01

*Effects of distraction on workload (10 increased*distracted, 10 Increased*not distracted) with respect to HBO.*

A significant effect was found for distraction [F(3,16) = 7.617, P<0.01]. The mean HBO for participants in the distracted group were higher during all workflow stages, when compared to those participating in the not distracted group as shown in the figure below. Increased workload is shown by the two bars on the left and standard workload on the right.

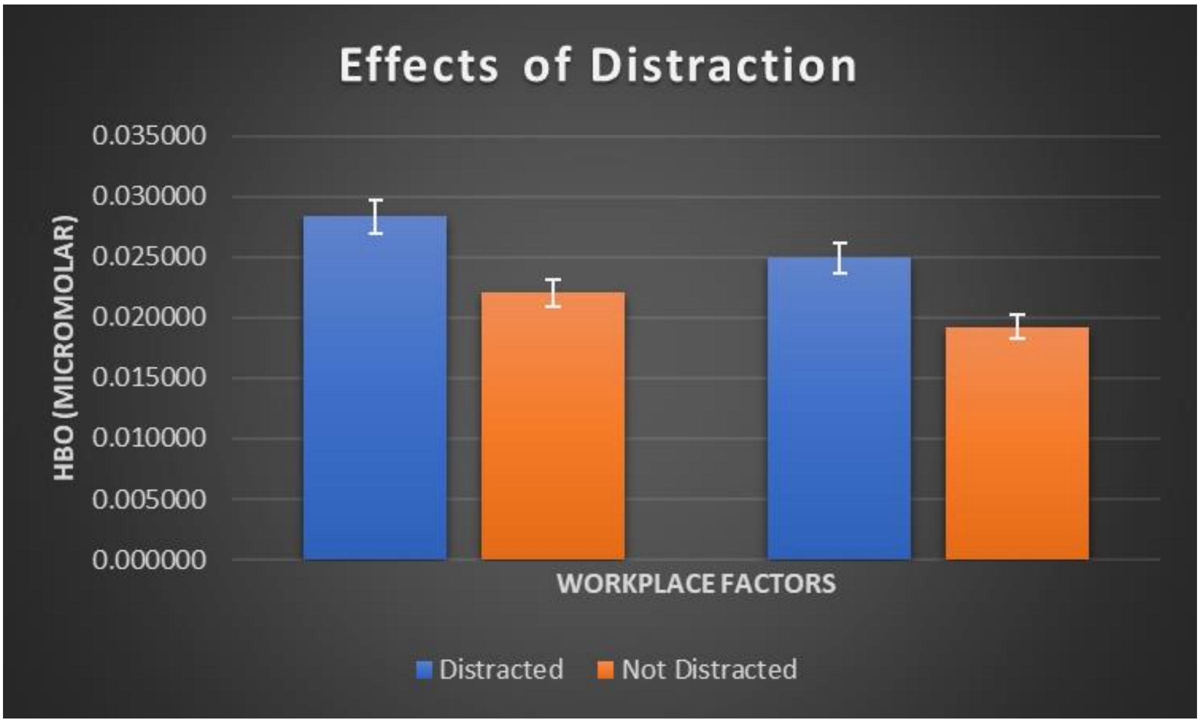


Figure - Effect of distraction

Pairwise comparisons show a significant effect for distraction on workload p<0.01 as depicted below in the table below.

Pairwise Comparisons

Measure: MEASURE_1

(I) Distraction 1 yes 2 no	(J) Distraction 1 yes 2 no	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	.009*	.002	.000	.005	.013
2	1	-.009*	.002	.000	-.013	-.005

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of distraction

Due to the significant effects showing in the above pairwise comparisons, t-tests were conducted in order for us to see the size of the effect. Distraction was used as Level (i) for tests.

- Distracted (level i) vs Not Distracted (level j) t = 0.009/0.002 = 4.5, p<0.001

Effects of Distraction*workload (10 increased*distracted, 10 Increased*not distracted) on workflow stage with respect to HBO.

A significant effect was found for distraction on workload with respect to workflow [F(2,17) = 18.967, P<0.01]. The mean HBO was higher for all workflow stages but is more prominent for the fault solution stage as shown in the figure below.

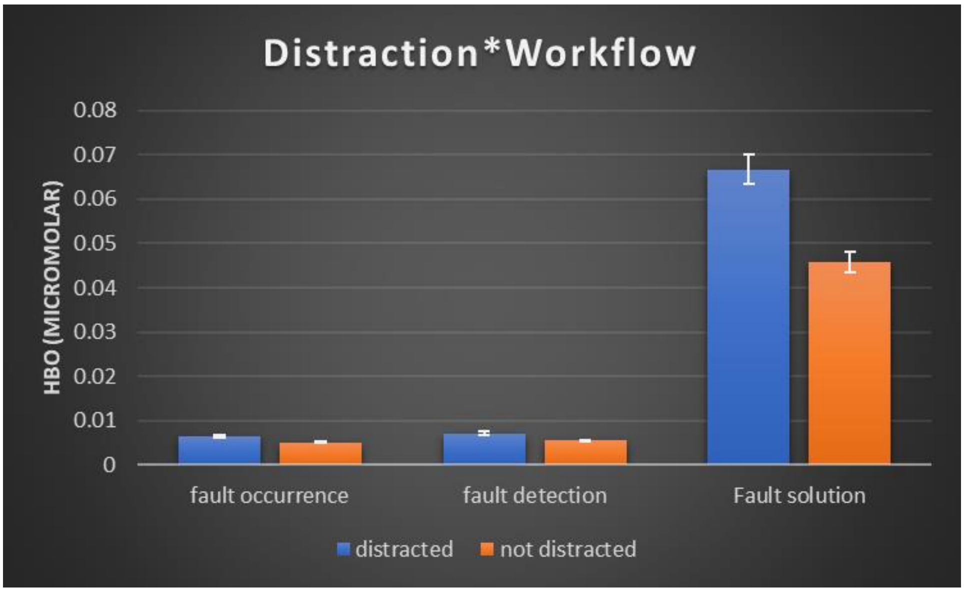


Figure - Effect of distraction with respect to workflow stage

Pairwise comparisons found a significant effect for distraction on workload with respect to workflow P<0.01 as depicted in the table below.

Distraction 1 yes 2 no * workflow

Measure: MEASURE_1

Distraction 1 yes 2 no	workflow	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	.006	.001	.004	.009
	2	.007	.001	.006	.008
	3	.118	.006	.106	.130
2	1	.005	.001	.003	.007
	2	.005	.001	.004	.006
	3	.082	.006	.070	.094

Table - Pairwise comparisons of workflow stage with respect to distraction

Due to the significant effects showing in the above pairwise comparisons, t-tests were conducted in order for us to see the size of the effect. The distraction was used as Level (i) for tests against fault solution for level (j) as this was shown to be significant above.

- Distracted (level i) vs Fault solution (level j) t = 0.118/0.006 = 19.7, p<0.001

Effect of ROI (10 Increased workload and distracted, 10 increased workload and not distracted).

A significant effect was found for ROI [F(2,17) = 247.629, P<0.01]. The mean HBO for left and right regions are higher for all workflow stages, when compared to the middle region as shown in the figure below.

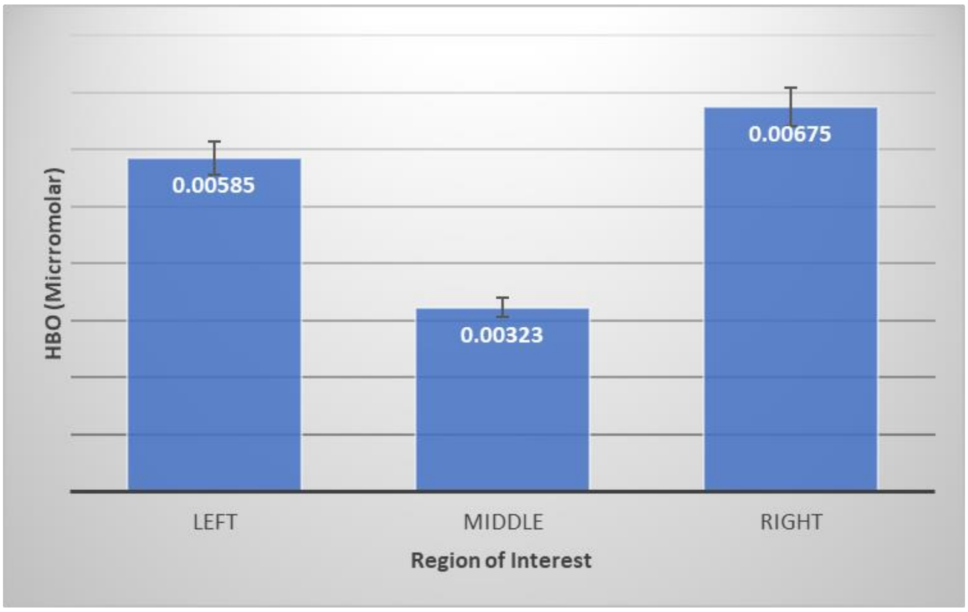


Figure - Region of interest comparison

Maulchy's test of sphericity showed a significant effect (p=0.016) therefore tests of within subjects effects using Greenhouse-Geisser showed [F=99.414, p<0.001, eta^2 = 0.909].

Pairwise comparisons show significant effects for the left and right regions of the DLPFC p<0.01 as depicted below in the table below.

Pairwise Comparisons

Measure: MEASURE_1

(I) ROI	(J) ROI	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	.029*	.002	.000	.024	.034
	3	.003	.003	.904	-.005	.011
2	1	-.029*	.002	.000	-.034	-.024
	3	-.026*	.002	.000	-.031	-.021
3	1	-.003	.003	.904	-.011	.005
	2	.026*	.002	.000	.021	.031

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of ROI

Due to the significant effects showing in the above pairwise comparisons, t-tests were conducted in order for us to see the size of the effect. The left and right regions were used as Level (i) for tests against the middle region as this was shown to be significant above.

- Left region (level i) vs Middle region (level j) t = 0.029/0.002 = 14.5, p<0.001
- Right region (level i) vs Middle region (level j) t = 0.026/0.002 = 13, p<0.001

*Effect of ROI*workflow (10 increased workload and distracted, 10 increased workload and not distracted) with respect to HBO*

The ANOVA showed a significant interaction between ROI and workflow with respect to HBO [F(6,13) = 116.612, P<0.01, eta^2 = 0.99. Maulchy's test of sphericity showed a significant effect therefore, test within subject's effects using Greenhouse-Geisser showed [F=75.761, P<0.001, eta^2 = 0.909].

*Effect of ROI*workflow*distraction (10 increased workload and distracted, 10 increased workload and not distracted) with respect to HBO*

The ANOVA showed a significant interaction between ROI*workflow*distraction with respect to HBO [F(6,13) = 3.998, P=0.017, eta^2 = 0.99. Tests within subject's effects using Greenhouse-Geisser showed [F=3.839, P=0.033, eta^2 = 0.909].

*Effect of ROI*Increased workload (10 Increased workload and not distracted, 10 standard workload and not distracted) with respect to HBO*

A significant effect was found for ROI*increased workload participants [F(2,37) = 56.538, P<0.01]. The mean HBO for left and right regions of the DLPFC was higher for all workflow stages, when compared to the middle region as shown in the figure below.

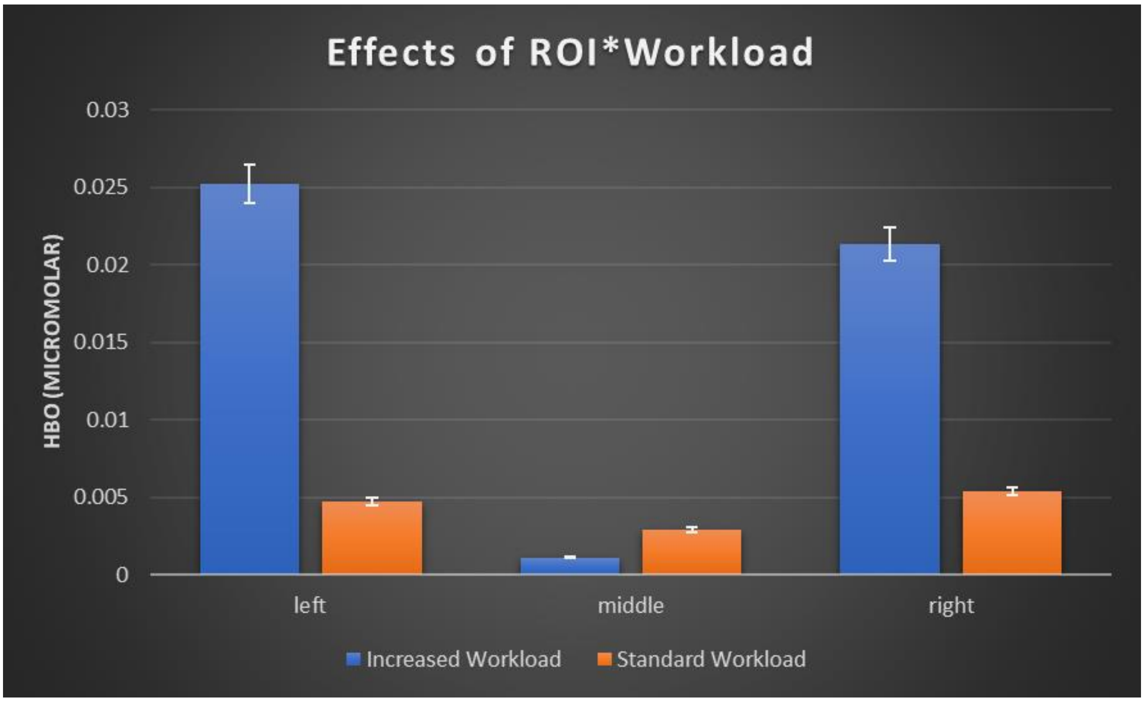


Figure - ROI with respect to workload

Pairwise comparisons showed significant effects for increased workload for the left and right regions as depicted in the table below.

4. Standard Workload = 1 Increased Workload = 2 * ROI

Measure: MEASURE_1

Standard Workload = 1 Increased Workload = 2	ROI	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
1	1	.006	.002	.002	.009
	2	.003	.001	.001	.005
	3	.007	.002	.003	.010
2	1	.030	.002	.027	.034
	2	.001	.001	-.001	.003
	3	.027	.002	.024	.031

Table- Pairwise comparisons of ROI with respect to workload

Due to the significant effects showing in the above pairwise comparisons, t-tests were conducted in order for us to see the size of the effect. The increased workload was used as Level (i) for tests against the left and right region as level (j).

- Increased workload (level i) vs Left region (level j) $t = 0.030/0.002 = 15$, $p<0.001$
- Increased workload (level i) vs Right region (level j) $t = 0.027/0.002 = 13.5$, $p<0.001$

*Effect of ROI*workflow*Increased workload (10 increased workload and not distracted, 10 standard workload and not distracted) with respect to HBO*

A significant effect was found for ROI*Workflow*increased workload participants [$F(6,33) = 46.6$, $P<0.01$]. The mean HBO for participants in the Increased workload group were higher for the left and right regions only during the fault solution stage as shown in the bar chart above in chapter 6.2.2.1.

Pairwise comparisons showed significant effects for standard workload*left and right regions for the fault solution workflow stage and for increased workload*left and right regions for the fault solution workflow stage as shown in the table below.

7. Standard Workload = 1 Increased Workload = 2 * ROI * Workflow

Measure: MEASURE_1

Standard Workload = 1 Increased Workload = 2	ROI	Workflow	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	1	.000	.004	-.007	.007
		2	.006	.001	.004	.008
		3	.007	.001	.005	.010
		4	.010	.005	-.001	.021
	2	1	.001	.001	-.001	.002
		2	.001	.001	.000	.002
		3	.002	.001	.000	.005
		4	.009	.002	.005	.013
	3	1	.000	.000	.000	.001
		2	.006	.001	.004	.008
		3	.006	.001	.004	.008
		4	.014	.006	.003	.026
2	1	1	.006	.004	-.001	.013
		2	.005	.001	.004	.007
		3	.006	.001	.004	.009
		4	.104	.005	.093	.115
	2	1	-1.624E-5	.001	-.001	.001
		2	.000	.001	-.001	.002
		3	.001	.001	-.002	.003
		4	.004	.002	.000	.009
	3	1	.000	.000	.000	.001
		2	.006	.001	.004	.008
		3	.006	.001	.004	.008
		4	.097	.006	.085	.109

Table- Pairwise comparisons of ROI with respect to workflow and workload

Due to the significant effects showing in the above pairwise comparisons, t-tests were conducted in order for us to see the size of the effect. The increased workload and left/right regions are used as Level (i) for tests against fault solution as level (j).

- Increased workload*Left (level i) vs Fault Solution (level j) t = 0.104/0.005 = 20.8, p<0.001
- Increased workload*Right (level i) vs Fault Solution (level j) t = 0.097/0.006 = 16.2, p<0.001

Standard workload*left/right regions are level (i) and fault solution is level (j) below.

- Standard workload*Left (level i) vs Fault Solution (level j) t = 0.010/0.005 = 2, p<0.001
- Standard workload*Right (level i) vs Fault Solution (level j) t = 0.014/0.006 = 2.3, p<0.001

Fatigue

Time data

The relevant areas of the workflow evaluated in the study were; fault occurrence (1), fault detection (2), and fault solution (3)(x-axis) against time (y-axis).

From the mean time taken for all participants, the fault solution stage of the workflow took the longest to complete as shown in the figure below. The bar chart below also shows a comparison of the mean time for participants to complete the task from the fatigued*standard and Fatigued*increased workload groups.

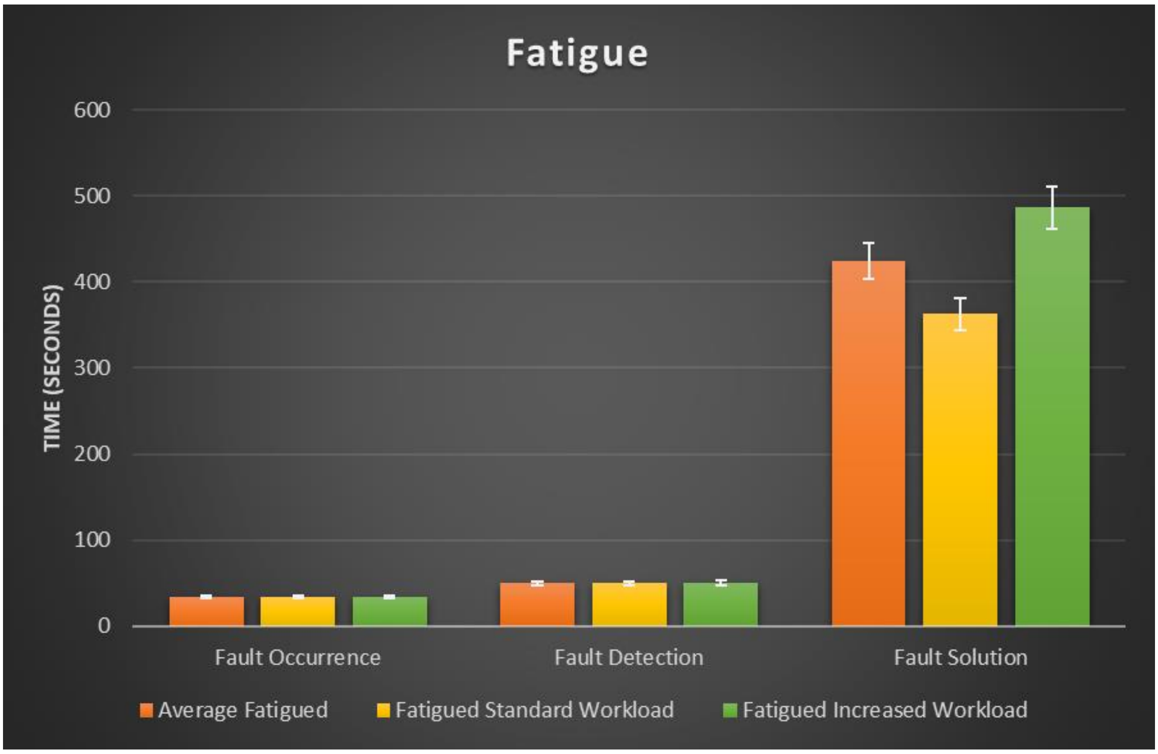


Figure - Fatigue with respect to workflow stage

*Effects of fatigue (10 fatigued*increased workload, 10 Fatigued*standard workload) on workflow with respect to time*

A significant effect was found for Fatigued*Increased workload against fatigued*standard workload participants [F(2,17) = 52815.9, P<0.01]. The mean times for participants in the increased workload group were higher during fault occurrence, fault detection and fault solution, when compared to those participating in the standard workload group as shown in the table below.

		Number of participants	**Mean time (seconds)**
Fault Occurrence	Increased WL	10	33.3
	Standard WL	10	33.3
Fault Detection	Increased WL	10	50.1
	Standard WL	10	49.7
Fault Solution	Increased WL	10	486.6
	Standard WL	10	362.6

Table - Participant averages

Pairwise comparisons show significant effects between all workflow stages (1=Fault occurrence, 2= fault detection & 3=fault solution) p<0.01 as depicted below in the table. Maulchy's test of sphericity showed a significant effect therefore tests of within subject effects using Greenhouse-Geisser showed [F=97363.1, p<0.001, eta^2 = 1.00].

Pairwise Comparisons

Measure: MEASURE_1

(I) workflow	(J) workflow	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-16.525*	.372	.000	-17.507	-15.543
	3	-391.245*	1.176	.000	-394.349	-388.141
2	1	16.525*	.372	.000	15.543	17.507
	3	-374.720*	1.223	.000	-377.947	-371.493
3	1	391.245*	1.176	.000	388.141	394.349
	2	374.720*	1.223	.000	371.493	377.947

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table- Pairwise comparisons of fatigued participants on workflow stages with respect to time.

Due to the significant effects showing in the above pairwise comparisons. T-tests were conducted to see the size of the effect. The fault solution stage was used as Level (i) for all tests as this stage has the most significance.

- Fault solution (level i) vs Fault occurrence (level j) t = 391.245/1.176 = 332.691, p<0.001
- Fault solution (level i) vs Fault detection (level j) t = 374.720/1.223 = 306.394, p<0.001

*Effect of workload*workflow on fatigue (10 fatigued*increased workload, 10 Fatigued*standard workload) with respect to time.*

The ANOVA showed a significant interaction between fatigue *workload on workflow with respect to time [F(2,17) = 1313.6, P<0.01, eta^2 = 0.994. A significant interaction was found in tests within subject effects using Greenhouse-Geisser [F=2542.6, P<0.001, eta^2 = 0.993].

Pairwise comparisons show significant effects between all workflow stages (1=Fault occurrence, 2= fault detection & 3=fault solution) p<0.01 as depicted below in the table below.

3. Increased Workload 1 yes 2 no * workflow

Measure: MEASURE_1

Increased Workload 1 yes 2 no	workflow	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	33.340	.342	32.621	34.059
	2	50.030	.426	49.134	50.926
	3	486.600	1.543	483.359	489.841
2	1	33.340	.342	32.621	34.059
	2	49.700	.426	48.804	50.596
	3	362.570	1.543	359.329	365.811

*Table - Pairwise comparisons of workload*workflow with respect to fatigue*

Due to the significant effects showing in the above pairwise comparisons, t-test were conducted in order to see the size of the effect. Increased workload was used as Level (i) for all tests as this stage has the most significance.

- Increase Workload (level i) vs Fault occurrence (level j) t = 33.34/0.342 = 97.49, p<0.001
- Increased Workload (level i) vs Fault detection (level j) t = 50.03/0.426 = 117.44, p<0.001
- Increased Workload (level i) vs Fault Solution (level j) t = 486.6/1.543 = 315.36, p<0.001

*Effect of Fatigue vs Not Fatigued (10 Fatigue*standard workload, 10 Standard) with respect to time*

A significant effect was found for fatigued against not fatigued participants [F(2,17) = 18589.8, P<0.01, eta^2 = 1.00]. The mean times for participants in the fatigued group were higher during all workflow stages, when compared to those participating in the not fatigued group as shown in the figure below.

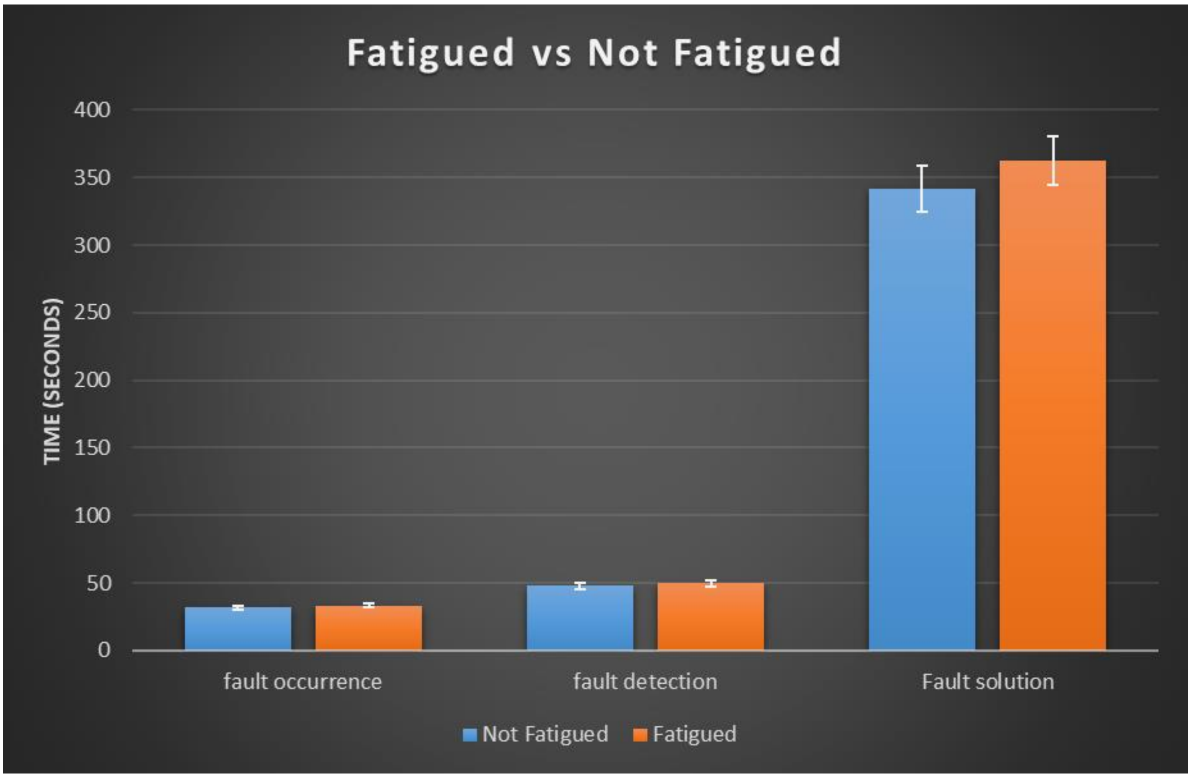


Figure - Workflow stages with respect to fatigue

Maulchy's test of sphericity showed a significant effect therefore tests of within subjects effects using Greenhouse-Geisser showed [F=33487.5, p<0.001, eta^2 = 0.99].

Pairwise comparisons show significant effects between all workflow stages (1=Fault occurrence, 2= fault detection & 3=fault solution) p<0.01 as depicted below in the table below.

Pairwise Comparisons

Measure: MEASURE_1

(I) workflow	(J) workflow	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-16.275*	.519	.000	-17.644	-14.906
	3	-319.430*	1.626	.000	-323.722	-315.138
2	1	16.275*	.519	.000	14.906	17.644
	3	-303.155*	1.699	.000	-307.638	-298.672
3	1	319.430*	1.626	.000	315.138	323.722
	2	303.155*	1.699	.000	298.672	307.638

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of workflow stages

Due to the significant effects showing in the above pairwise comparisons, t-test were conducted to see the size of the effect. The fault solution stage was used as Level (i) for all tests as this stage has the most significance.

- Fault solution (level i) vs Fault occurrence (level j) t = 319.430/1.626 = 196.451, p<0.001
- Fault solution (level i) vs Fault detection (level j) t = 303.155/1.699 = 178.431, p<0.001

*Effect of Fatigued vs Not Fatigued on workflow stage (10 fatigued*standard workload, 10 standard) with respect to time*

The ANOVA showed a significant interaction between fatigue and workflow with respect to time [F(2,17) = 17.15, P<0.01, eta^2 = 0.669. A significant interaction was found in tests within subjects effects using Greenhouse-Geisser [F=32.837, P<0.001, eta^2 = 0.646].

Pairwise comparisons show significant effects between all workflow stages (1=Fault occurrence, 2= fault detection & 3=fault solution) p<0.01 as depicted below in the table below.

3. Fatigued (1 yes, 2 no) * workflow

Measure: MEASURE_1

Fatigued (1 yes, 2 no)	workflow	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	33.340	.525	32.238	34.442
	2	49.700	.502	48.646	50.754
	3	362.570	2.111	358.135	367.005
2	1	31.590	.525	30.488	32.692
	2	47.780	.502	46.726	48.834
	3	341.220	2.111	336.785	345.655

Table - Pairwise comparisons of workflow stage with respect to fatigue

Due to the significant effects showing in the above pairwise comparisons, t-test were conducted to see the size of the effect. The fatigued group was used as Level (i) and workflow stage as level (j)

- Fatigued (level i) vs Fault occurrence (level j) t = 33.34/0.525 = 63.5, p<0.001
- Fatigued (level i) vs Fault detection (level j) t = 49.7/0.502 = 99.0, p<0.001
- Fatigued (level i) vs Fault Solution (level j) t = 362.570/2.111 = 171.8

For the t-tests below, the not fatigued group is used as level (i) and the workflow stage as level (j).

- Not Fatigued (level i) vs Fault occurrence (level j) t = 31.59/0.525 = 60.2, p<0.001
- Not Fatigued (level i) vs Fault detection (level j) t = 47.78/0.502 = 95.2, p<0.001
- Not Fatigued (level i) vs Fault Solution (level j) t = 341.22/2,111 = 161.6, p<0.001

HBO data

*The effect of fatigue (10 Fatigued*standard, 10 Standard) with respect to HBO.*

No significant effect was found for fatigue. However, mean HBO for participants in the fatigued group were higher when compared to those participating in the not fatigued group as shown in the figure below.

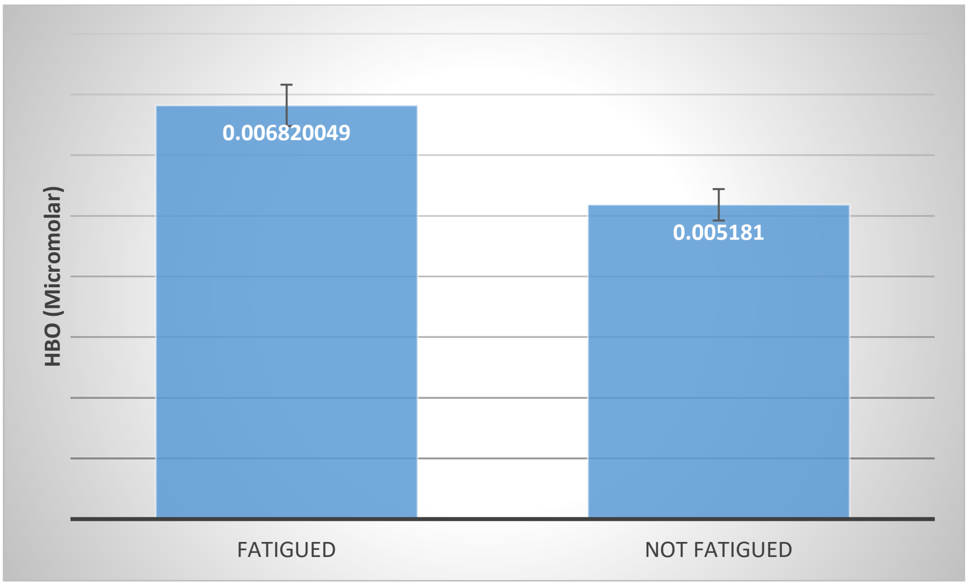


Figure - Effect of Fatigue

*Effect of fatigue (10 fatigued*standard workload, 10 standard workload) on workflow stage with respect to HBO*

The ANOVA showed a significant effect for workflow with respect to fatigue [F(3,16) = 72.506, P<0.01, eta^2 = 0.931 as depicted by the figure below.

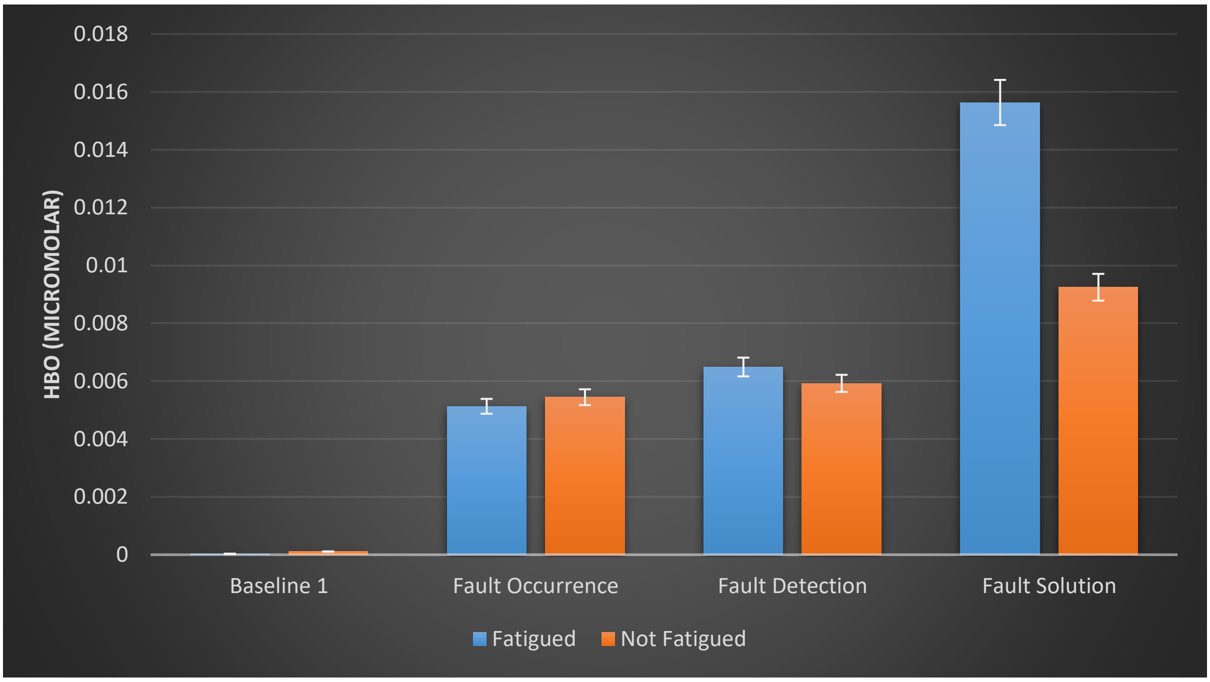


Figure - effects of workflow with respect to fatigue

Maulchy's test of sphericity showed a significant interaction was found in tests within subjects effects using Greenhouse-Geisser [F=29.464, P<0.001, eta^2 = 0.621].

Pairwise comparisons show significant effects for the fault solution workflow stage p<0.01 as depicted below in the table below.

Pairwise Comparisons

Measure: MEASURE_1

(I) Workflow	(J) Workflow	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-.004*	.000	.000	-.005	-.003
	3	-.005*	.000	.000	-.006	-.004
	4	-.011*	.002	.000	-.016	-.007
2	1	.004*	.000	.000	.003	.005
	3	-.001	.000	.142	-.003	.000
	4	-.008*	.002	.001	-.013	-.003
3	1	.005*	.000	.000	.004	.006
	2	.001	.000	.142	.000	.003
	4	-.006*	.002	.011	-.012	-.001
4	1	.011*	.002	.000	.007	.016
	2	.008*	.002	.001	.003	.013
	3	.006*	.002	.011	.001	.012

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of workflow stages with respect to fatigue

Due to the significant effects showing in the above pairwise comparisons, t-tests were conducted in order for us to see the size of the effect. The fault solution stage was used as Level (i) for all tests as this stage has the only significance.

- Fault solution (level i) vs Baseline (level j) t = 0.011/0.002 = 5.5, p<0.001
- Fault solution (level i) vs Fault Occurence (level j) t = 0.008/0.002 = 4, p<0.001
- Fault Solution (level i) vs Fault Detection (level j) t = 0.008/0.002 = 4, p<0.01

*Effects of Fatigue*workload (10 fatigued*increase workload, 10 fatigued*standard workload) with respect to HBO*

A significant effect found for workload whilst fatigued [F(2,17) = 69.896, P<0.01. The mean HBO for participants in the increased workload group are significantly higher, when compared to those participating in the standard workload group as shown in the figure below.

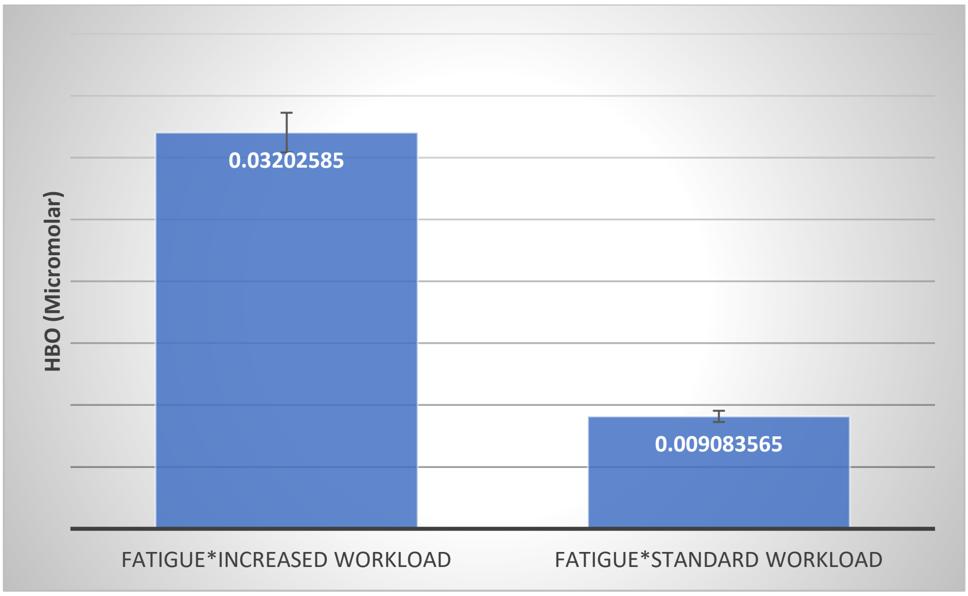


*Figure - Effects of Fatigue*Workload*

Pairwise comparisons show a significant effect for increased workload*fatigued p<0.01 as depicted below in the table below.

Pairwise Comparisons

Measure: MEASURE_1

(I) Standard =1, Increased =2	(J) Standard =1, Increased =2	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-.023*	.003	.000	-.028	-.018
2	1	.023*	.003	.000	.018	.028

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of workload with respect to fatigue.

Due to the significant effects showing in the above pairwise comparisons, t-tests were conducted in order for us to see the size of the effect. Increased workload was used as Level (i).

- Increased workload (level i) vs Standard workload (level j) t = 0.023/0.003 = 7.67, p<0.001

Effects of Fatigue*Workload (10 fatigued*increased workload, 10 fatigued*standard workload) on workflow stage with respect to HBO.

A significant effect was found for fatigue*workload with respect to workflow [F(2,17) = 40.227, P<0.01]. The mean HBO was significantly higher for the fault solution workflow stage as shown in the figure below.

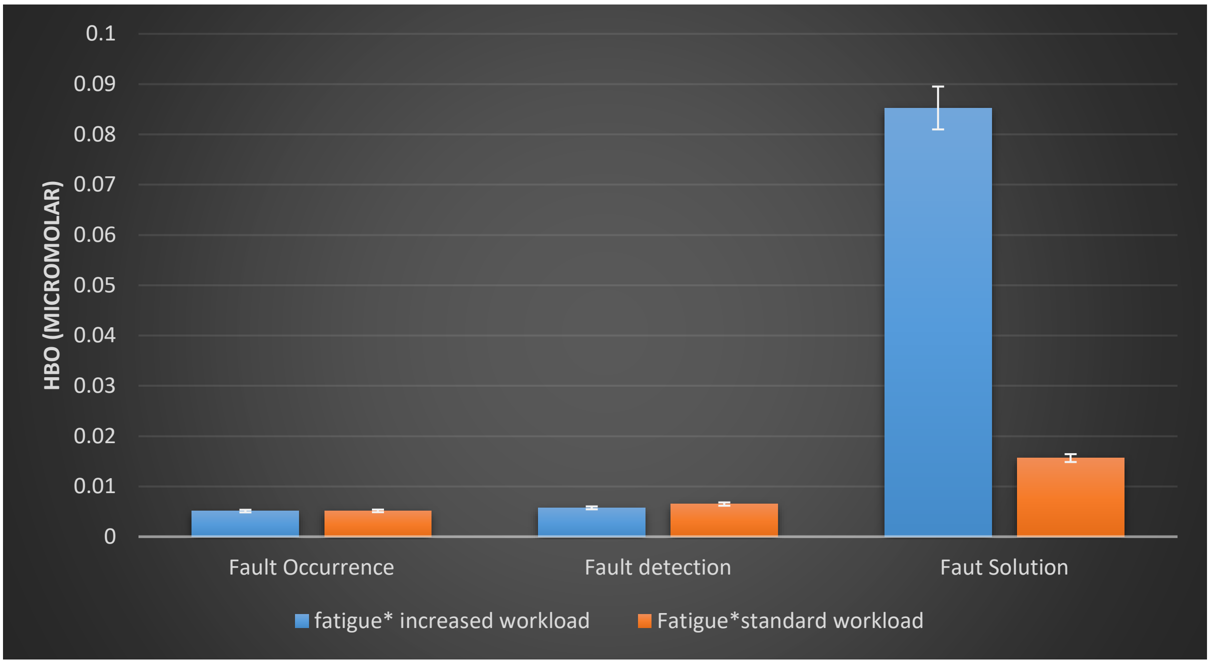


*Figure - Fatigue*Workload with respect to workflow stage*

Pairwise comparisons show a significant effect for fatigue*workload with respect to workflow P<0.01 as depicted in the table below.

5. Standard =1, Increased =2 * Workflow

Measure: MEASURE_1

Standard =1, Increased =2	Workflow	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	.005	.000	.004	.006
	2	.006	.001	.005	.008
	3	.016	.005	.004	.027
2	1	.005	.000	.004	.006
	2	.006	.001	.004	.007
	3	.085	.005	.074	.097

*Table - Pairwise comparisons of fatigue*workload on workflow stage with respect to HBO.*

Due to the significant effects showing in the above pairwise comparisons, t-tests were conducted to see the size of the effect. Increased workload was used as Level (i) for tests against workflow stage for level (j) as this was shown to be significant above.

- Increased Workload (level i) vs Fault Detection (level j) t = 0.006/0.001 = 6, p<0.001
- Increased Workload (level i) vs Fault Solution (level j) t = 0.085/0.005 = 17, p<0.001

*Effect of ROI (10 fatigued*increased workload, 10 Fatigued*standard workload)*
A significant effect was found for ROI [F(6,13) = 9.505, P<0.01]. The mean HBO for left and right regions are higher for all workflow stages, when compared to the middle region as shown in the figure below.

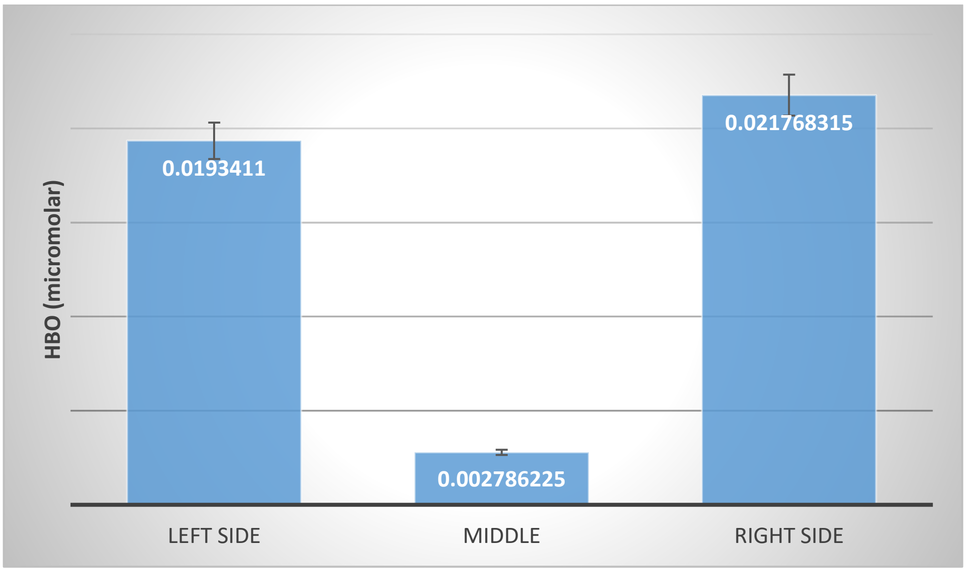


Figure - Effects of ROI

Maulchy's test of sphericity showed a significant effect (p=0.00) therefore tests of within subjects effects using Greenhouse-Geisser was analysed but showed no significant effects].

Pairwise comparisons showed no significant effects for regions of the DLPFC p<0.01.

*Effect of ROI*workflow (10 fatigued*increased workload, 10 fatigued*standard workload) with respect to HBO*
The ANOVA showed a significant interaction between ROI and workflow with respect to HBO [F(6,13) = 18.041, P<0.01, eta^2 = 0.821.

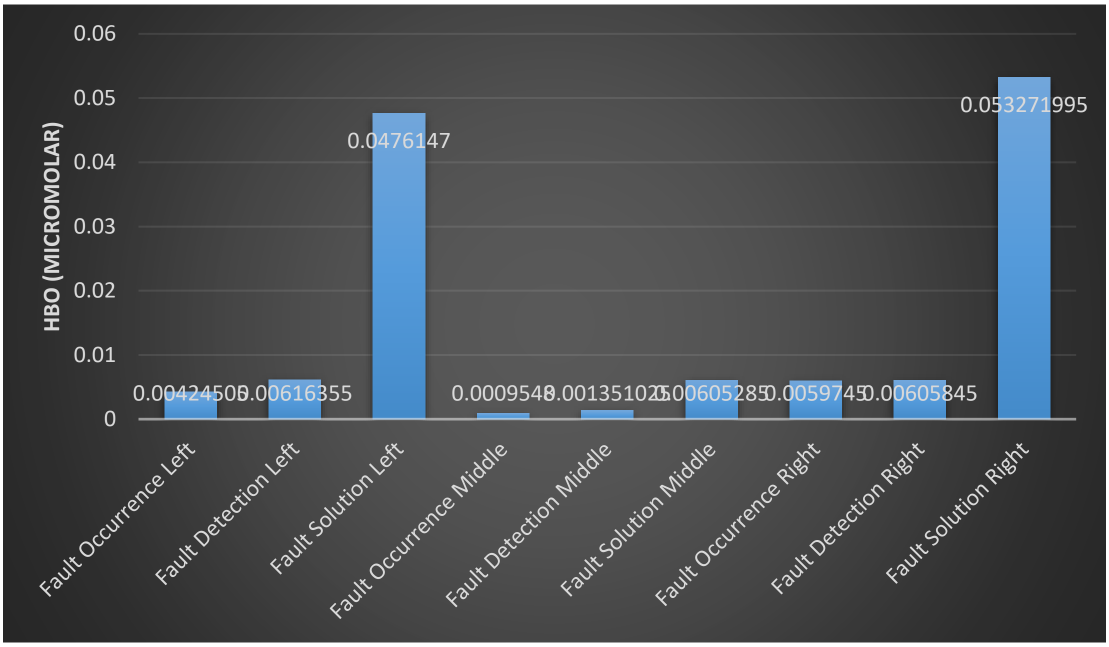


Figure - ROI with respect to workflow stage

Pairwise comparisons showed no significant effect for ROI with respect to workflow.

*Effect of ROI*workflow*Fatigue (10 fatigued*increased workload, 10 fatigued*standard workload) with respect to HBO*

The ANOVA showed no significant interaction between ROI*workflow*fatigue with respect to HBO.

Fatigue vs Distraction vs Increased workload

Time study

The relevant areas of the workflow evaluated in the study were; fault occurrence (1), fault detection (2), and fault solution (3)(x-axis) against time (y-axis).

The figure below shows mean time taken for all participants. The bar chart below also shows a comparison of the mean time for participants to complete the task from the fatigued, distracted, increased workload and standard workload groups.

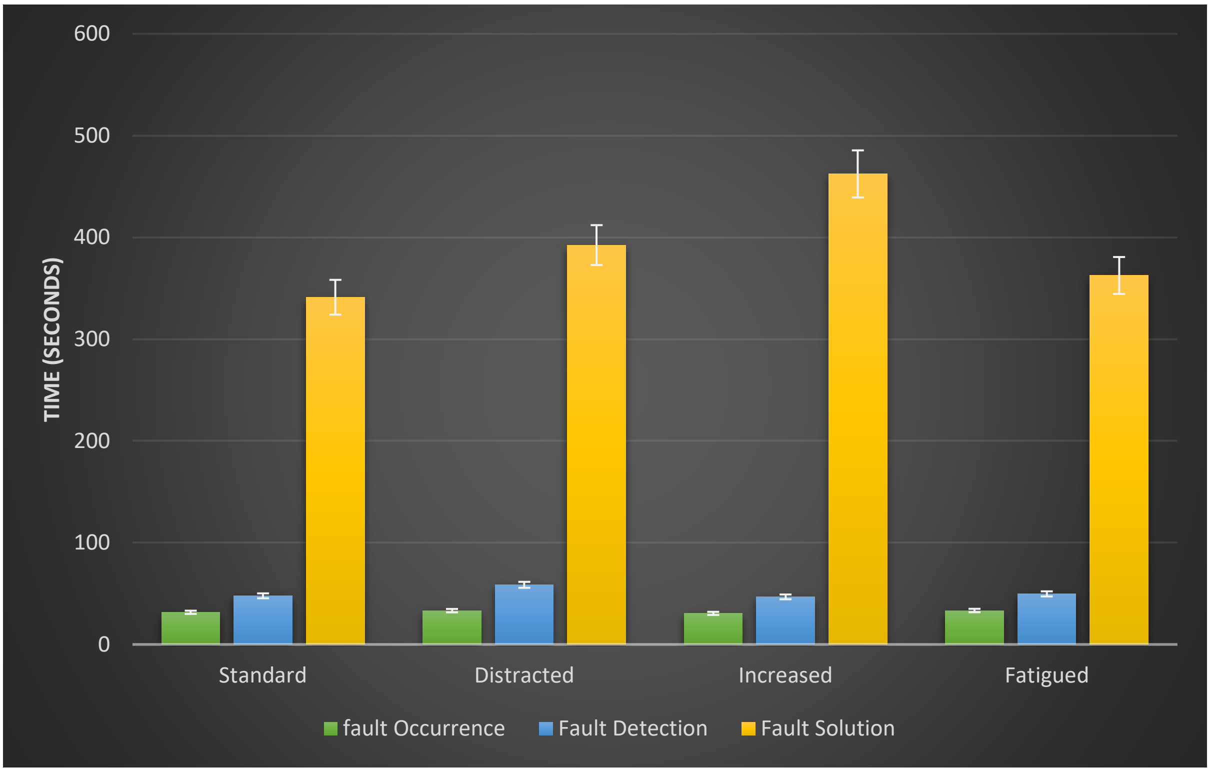


Figure - Comparison of performance shaping factors with respect to workflow stage

The Figure below is a bar chart showing the mean time taken for participants to complete the exercise with respect to PSF.

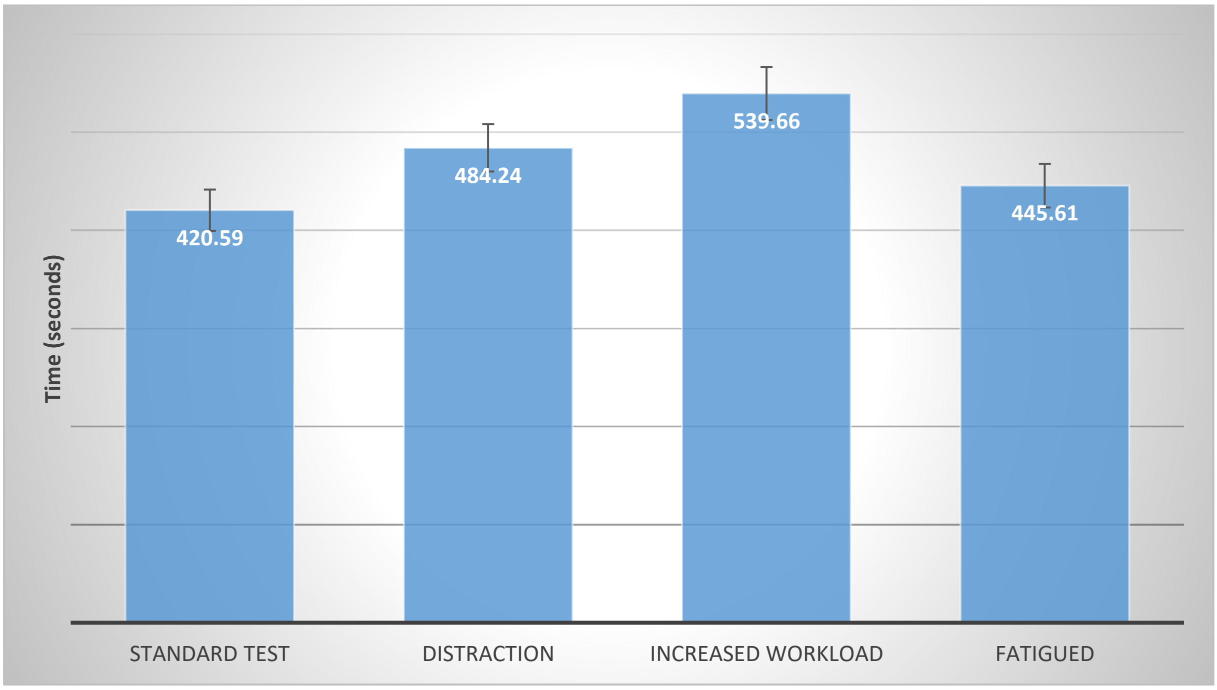


Figure - Comparison of performance shaping factors

The table below shows the workplaces factors and the number of participants in each group against the mean time in seconds for each workflow stage.

	Number of Participants	fault Occurrence (seconds)	Fault Detection (seconds)	Fault Solution (seconds)
Standard	10	31.59	47.78	341.22
Distracted	10	33.22	58.58	392.44
Increased	10	30.51	46.75	462.4
Fatigued	10	33.34	49.7	362.57

Table - performance shaping factors and workflow stage output

The effect of PSF against a Standard test

A significant effect was found for a comparison of PSF [F(2,35) = 20113.431, P<0.01]. The mean times for all groups were higher during the fault solution stage when compared to all other stages.

Pairwise comparisons show significant effects between all performance shaping factors (1=Standard, 2= Distracted, 3=Increased workload & 4=Standard workload) p<0.01 as depicted below in the table.

Pairwise Comparisons

Measure: MEASURE_1

(I) Standard 1 Distracted 2 Increased 3 Fatigued 4	(J) Standard 1 Distracted 2 Increased 3 Fatigued 4	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-21.217*	1.841	.000	-26.358	-16.075
	3	-39.690*	1.841	.000	-44.831	-34.549
	4	-8.340*	1.841	.000	-13.481	-3.199
2	1	21.217*	1.841	.000	16.075	26.358
	3	-18.473*	1.841	.000	-23.615	-13.332
	4	12.877*	1.841	.000	7.735	18.018
3	1	39.690*	1.841	.000	34.549	44.831
	2	18.473*	1.841	.000	13.332	23.615
	4	31.350*	1.841	.000	26.209	36.491
4	1	8.340*	1.841	.000	3.199	13.481
	2	-12.877*	1.841	.000	-18.018	-7.735
	3	-31.350*	1.841	.000	-36.491	-26.209

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of performance shaping factors

Maulchy's test of sphericity showed a significant effect therefore tests of within subject effects using Greenhouse-Geisser showed [F=34963.5, p<0.001, eta^2 = 0.999].

Due to the significant effects shown in the above pairwise comparisons. T-tests were conducted in order for us to see the size of each effect. Increased Workload was used as Level (i) for all tests as this factor has the most significance.

- Increased workload (level i) vs standard test (level j) t = 39.690/1.841 = 21.559, p<0.001
- Increased workload (level i) vs Distraction (level j) t = 18.473/1.841 = 10.034, p<0.001
- Increased workload (level i) vs Fatigued (level j) t = 31.350/1.841 = 17.029, p<0.001

The second group of t-tests show the effect size of each PSF compared to a standard test. For this group of t-tests, a standard test was used as level (j) throughout.

- Distraction (level i) vs Standard test (level j) t = 21.217/1.841 = 11.525, p<0.001
- Increased workload (level i) vs Standard test (level j) t = 39.690/1.841 = 21.559, p<0.001
- Fatigued (level i) vs Standard test (level j) t = 8.340/1.841 = 4.530, p<0.001

The Effects of PSF with respect to workflow

The ANOVA showed a significant effect between PSF with respect to workflow [F(6,72) = 126.381, P<0.01, eta^2 = 0.918.

Pairwise comparisons show significant effects between all PSF p<0.01 as depicted below in the table.

3. Standard 1 Distracted 2 Increased 3 Fatigued 4 * Workflow

Measure: MEASURE_1

Standard 1 Distracted 2 Increased 3 Fatigued 4	Workflow	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
1	1	31.590	.462	30.654	32.526
	2	47.780	.519	46.728	48.832
	3	341.220	3.725	333.666	348.774
2	1	33.220	.462	32.284	34.156
	2	58.580	.519	57.528	59.632
	3	392.440	3.725	384.886	399.994
3	1	30.510	.462	29.574	31.446
	2	46.750	.519	45.698	47.802
	3	462.400	3.725	454.846	469.954
4	1	33.340	.462	32.404	34.276
	2	49.700	.519	48.648	50.752
	3	362.570	3.725	355.016	370.124

Table - Pairwise comparisons of performance shaping factors with respect to workflow.

Maulchy's test of sphericity showed a significant effect therefore tests of within subject effects using Greenhouse-Geisser showed [F=207.998, p<0.001, eta^2 = 0.945].

Due to the significant effects showing in the above pairwise comparisons, t-test were conducted in order to see the size of the effect. PSF were used as Level (i) and Fault solution for level (j) for all tests as this stage has the most significance.

- Distraction (level i) vs Fault Solution (level j) t = 392.440/3.725 = 105.353, p<0.001
- Increased Workload (level i) vs Fault Solution (level j) t = 462.4/3.725 = 124.134, p<0.001
- Fatigue (level i) vs Fault Solution (level j) t = 362.570/3.725 = 97.334, p<0.001

HBO study

The relevant areas of the workflow evaluated in the study were; Baseline (1), fault occurrence (2), fault detection (3), and fault solution (4) (x-axis) against HBO (y-axis). Also evaluated was region of interest; left side and right side. The middle region was omitted due to the lack of usable data in previous tests.

The Figure below shows mean HBO for all participants. The bar chart below also shows a comparison of the mean HBO for participants for the full duration of the task from; the fatigued, distracted, increased workload and standard workload groups. Data is taken from the left and right regions of the DLPFC.

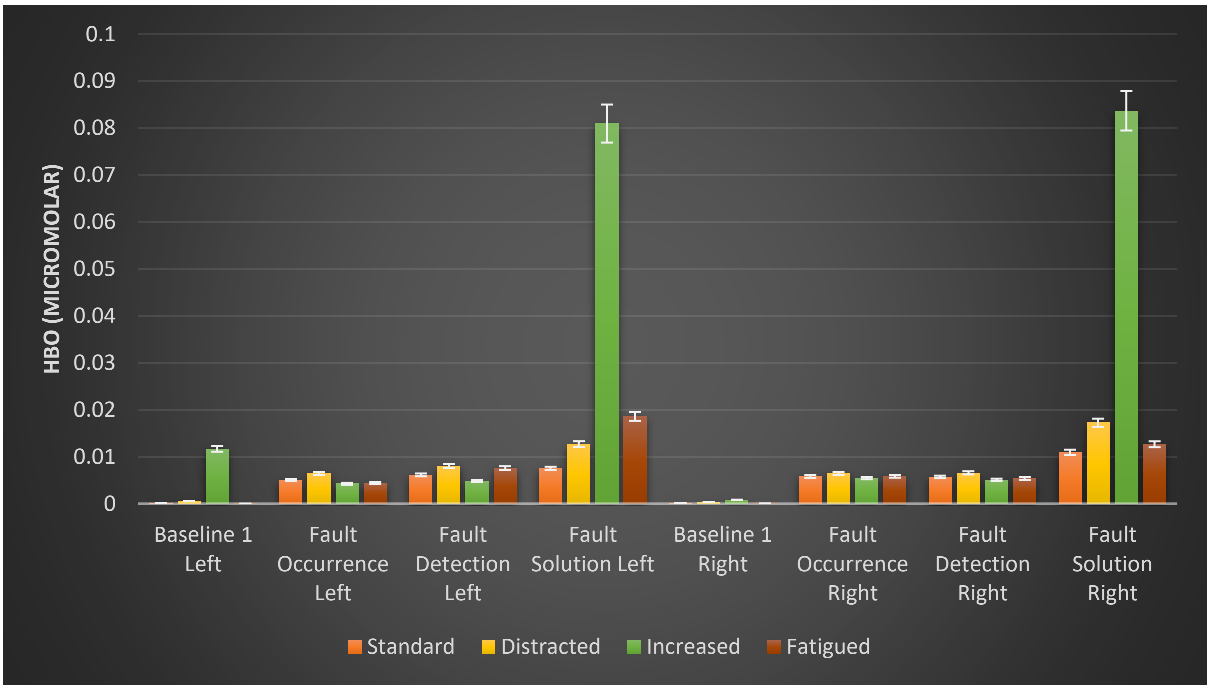


*Figure - Performance shaping factors*workflow with respect to HBO*

Below is the figure showing the mean HBO taken from participants for the full duration of the exercise, against PSF.

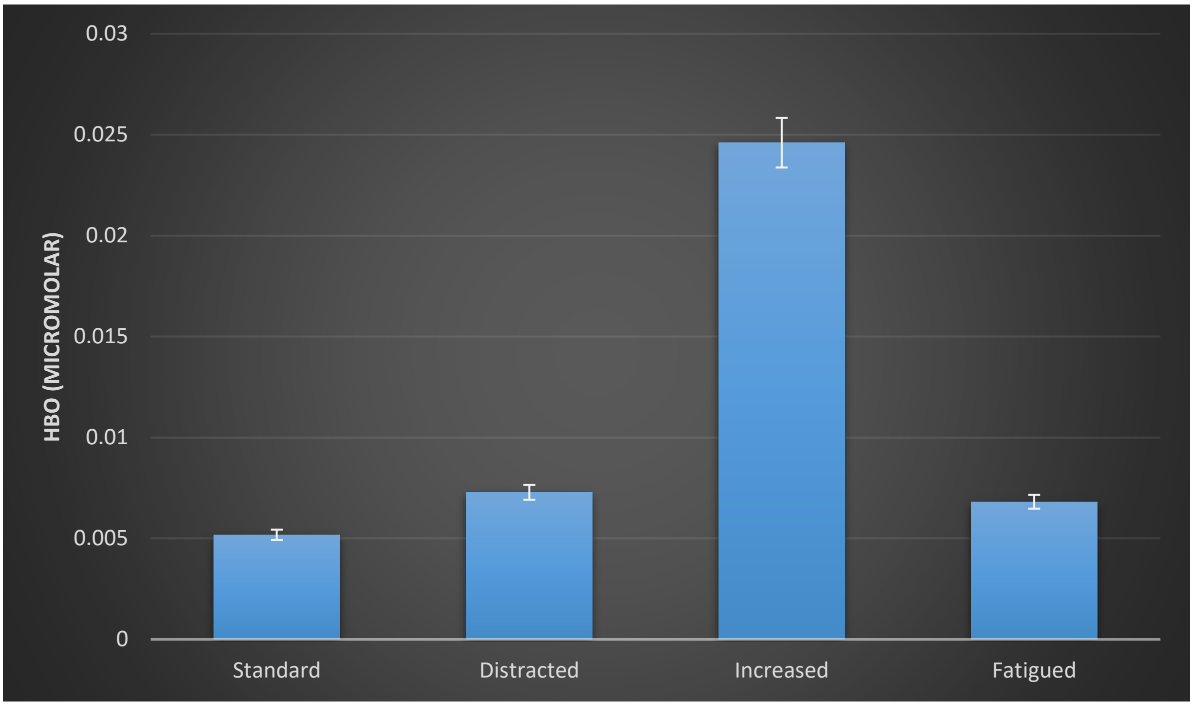


Figure - Comparison of Performance shaping factors

The effect of PSF against a Standard test
A significant effect was found for a comparison of PSF [F(3,35) = 121.617, P<0.01].

Pairwise comparisons show significant effects for Increased workload p<0.01 as depicted below in the table.

Pairwise Comparisons

Measure: MEASURE_1

(I) Standard 1 Distracted 2 Increased 3 Fatigued 4	(J) Standard 1 Distracted 2 Increased 3 Fatigued 4	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-.002	.002	1.000	-.007	.003
	3	-.019*	.002	.000	-.025	-.014
	4	-.002	.002	1.000	-.007	.004
2	1	.002	.002	1.000	-.003	.007
	3	-.017*	.002	.000	-.023	-.012
	4	.000	.002	1.000	-.005	.006
3	1	.019*	.002	.000	.014	.025
	2	.017*	.002	.000	.012	.023
	4	.018*	.002	.000	.013	.023
4	1	.002	.002	1.000	-.004	.007
	2	.000	.002	1.000	-.006	.005
	3	-.018*	.002	.000	-.023	-.013

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of Performance shaping factors.

Maulchy's test of sphericity showed a significant effect therefore tests of within subject effects using Greenhouse-Geisser showed [F=188.706, p<0.001, eta^2 = 0.836].

Due to the significant effects shown in the above pairwise comparisons. T-tests were conducted in order for us to see the size of each effect. Increased Workload was used as Level (i) for all tests as this PSF has the most significance.

- Increased workload (level i) vs standard test (level j) t = 0.019/0.002 = 9.5, p<0.001
- Increased workload (level i) vs Distraction (level j) t = 0.017/0.002 = 8.5, p<0.001
- Increased workload (level i) vs Fatigued (level j) t = 0.018/0.002 = 9, p<0.001

The Effects of PSF with respect to workflow
The ANOVA showed a significant effect between PSF with respect to workflow [F(3,37) = 52.126, P<0.01, eta^2 = 0.823.

Pairwise comparisons show significant effects for Distraction, Increased workload and fatigue mainly for the fault solution stage of the workflow (p<0.01) as depicted below in the table.

5. Standard 1 Distracted 2 Increased 3 Fatigued 4 * workflow

Measure: MEASURE_1

Standard 1 Distracted 2 Increased 3 Fatigued 4	workflow	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	.000	.002	-.005	.005
	2	.005	.001	.003	.008
	3	.006	.001	.003	.009
	4	.009	.003	.003	.015
2	1	.000	.002	-.004	.005
	2	.006	.001	.004	.009
	3	.007	.001	.005	.010
	4	.015	.003	.009	.021
3	1	.006	.002	.001	.011
	2	.005	.001	.003	.007
	3	.005	.001	.002	.008
	4	.082	.003	.076	.089
4	1	2.950E-5	.002	-.005	.005
	2	.005	.001	.003	.007
	3	.006	.001	.004	.009
	4	.016	.003	.010	.022

Table - Pairwise comparisons of Performance shaping factors with respect to workflow

Maulchy's test of sphericity showed a significant effect therefore tests of within subject effects using Greenhouse-Geisser showed [F=77.890, p<0.001, eta^2 = 0.863].

Due to the significant effects showing in the above pairwise comparisons, t-test were conducted in order to see the size of the effect. Workplace factors were used as Level (i) and Fault solution for level (j) for all tests as this stage has the most significance.

- Distraction (level i) vs Fault Solution (level j) t = 0.015/0.003 = 5, p<0.001
- Increased Workload (level i) vs Fault Solution (level j) t = 0.082/0.003 = 27.3r, p<0.001
- Fatigue (level i) vs Fault Solution (level j) t = 0.016/0.003 = 5.3r, p<0.001

Effect of ROI

The ANOVA has shown no significant effects for ROI with respect to PSF or workflow.

Combined PSFs

Time study

The relevant areas of the workflow evaluated in the study were; fault occurrence (1), fault detection (2), and fault solution (3)(x-axis) against time (y-axis).

The Figure below shows mean time taken for all participants. The bar chart below also shows a comparison of the mean time for participants to complete the task from the fatigued*increased workload, distracted*increased workload and standard test groups.

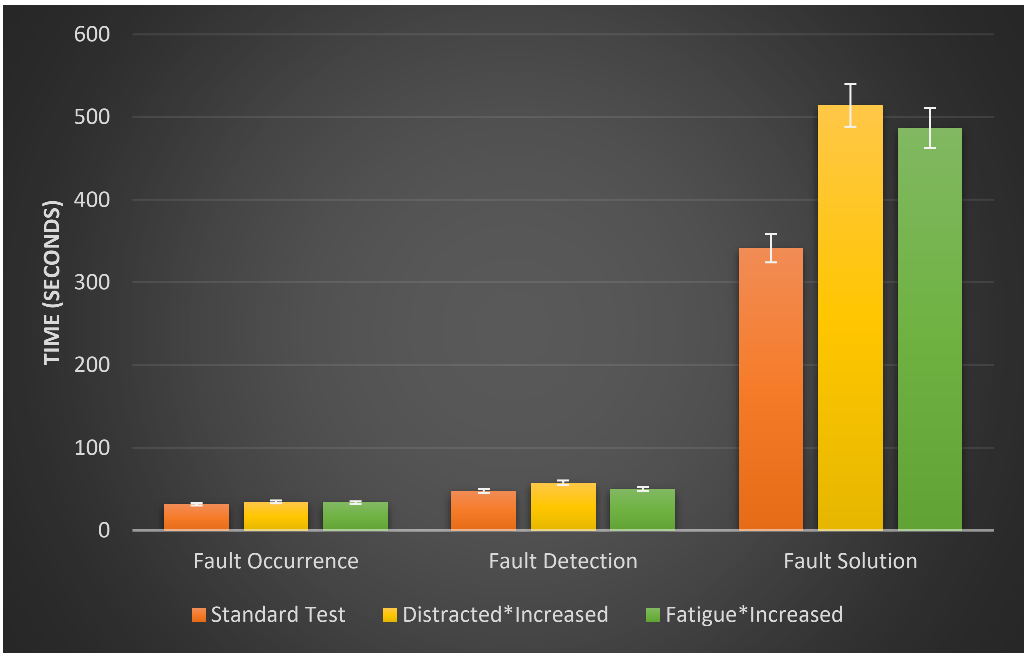


Figure - Combined PSF's with respect to workflow

Below is a bar chart showing the mean time taken for participants to complete the exercise against PSF.

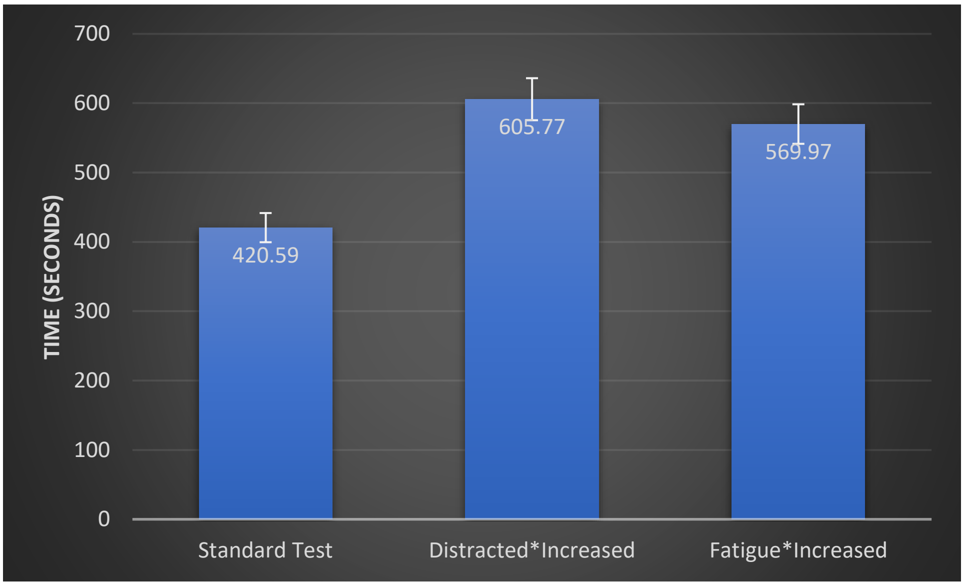


Figure - Comparison of combined PSF's

Below the table shows the PSF and the number of participants in each group against the mean time in seconds of each workflow stage.

	Number of Participants	Fault Occurrence	Fault Detection	Fault Solution
Standard Test	10	31.59	47.78	341.22
Distracted*Increased	10	34.35	57.4	514.02
Fatigue*Increased	10	33.34	50.03	486.6

Table - Mean times of combined PSF's

The effect of combined PSF against a Standard test

A significant effect was found for a combined PSF [F(2,26) = 27068.592, P<0.01.

Pairwise comparisons show significant effects between combined PSF (1=Standard, 2= Distracted*Increased workload, 3=Fatigue*Increased workload) p<0.01 as depicted below in the table.

Pairwise Comparisons

Measure: MEASURE_1

(I) standard 1, D*I 2, F*I 3	(J) standard 1, D*I 2, F*I 3	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-61.727*	1.170	.000	-64.714	-58.740
	3	-49.793*	1.170	.000	-52.780	-46.806
2	1	61.727*	1.170	.000	58.740	64.714
	3	11.933*	1.170	.000	8.946	14.920
3	1	49.793*	1.170	.000	46.806	52.780
	2	-11.933*	1.170	.000	-14.920	-8.946

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of combined PSF's

Maulchy's test of sphericity showed a significant effect therefore tests of within subject effects using Greenhouse-Geisser showed [F=49378.3, p<0.001, eta^2 = 0.999].

Due to the significant effects shown in the above pairwise comparisons. T-tests were conducted to see the size of each effect. A standard test was used as Level (j) for all tests.

- Distracted*Increased workload (level i) vs standard test (level j) t = 61.727/1.170 = 52.758, p<0.001
- Fatigued*Increased workload (level i) vs standard test (level j) t = 49.793/1.170 = 42.558, p<0.001

The Effects of combined PSF with respect to workflow

The ANOVA showed a significant effect between combined PSF with respect to workflow [F(2,27) = 904.021, P<0.01, eta^2 = 0.985.

Pairwise comparisons show significant effects between combined PSF p<0.01 as depicted below in the table.

standard 1, D*I 2, F*I 3	workflow	Mean	Std. Error	95% Confidence Interval Lower Bound	Upper Bound
1	1	31.590	.668	30.219	32.961
	2	47.780	.592	46.565	48.995
	3	341.220	2.812	335.451	346.989
2	1	34.350	.668	32.979	35.721
	2	57.400	.592	56.185	58.615
	3	514.020	2.812	508.251	519.789
3	1	33.340	.668	31.969	34.711
	2	50.030	.592	48.815	51.245
	3	486.600	2.812	480.831	492.369

Table - Pairwise comparisons of combined PSF's with respect to workflow

Due to the significant effects showing in the above pairwise comparisons, t-test were conducted in order to see the size of the effect. PSF were used as Level (i) and Fault solution for level (j) for all tests as this stage has the most significance.

- Standard test (level i) vs Fault Solution (level j) t = 341.220/2.812 = 121.344, p<0.001
- Distracted*Increased Workload (level i) vs Fault Solution (level j) t = 514.020/2.812 = 182.795, p<0.001
- Fatigue*Increased workload (level i) vs Fault Solution (level j) t = 486.6/2.812 = 173.044, p<0.001

HBO study

The relevant areas of the workflow evaluated in the study were; Baseline (1), fault occurrence (2), fault detection (3), and fault solution (4) (x-axis) against HBO (y-axis). Also evaluated was region of interest; left side and right side. The middle region was omitted due to the lack of usable data in previous tests.

The Figure below shows mean HBO for all participants. The bar chart below also shows a comparison of the mean HBO for participants for the full duration of the task from; the fatigued*Increased workload, distracted*Increased workload, and standard test groups. Data is taken from the left and right regions of the DLPFC.

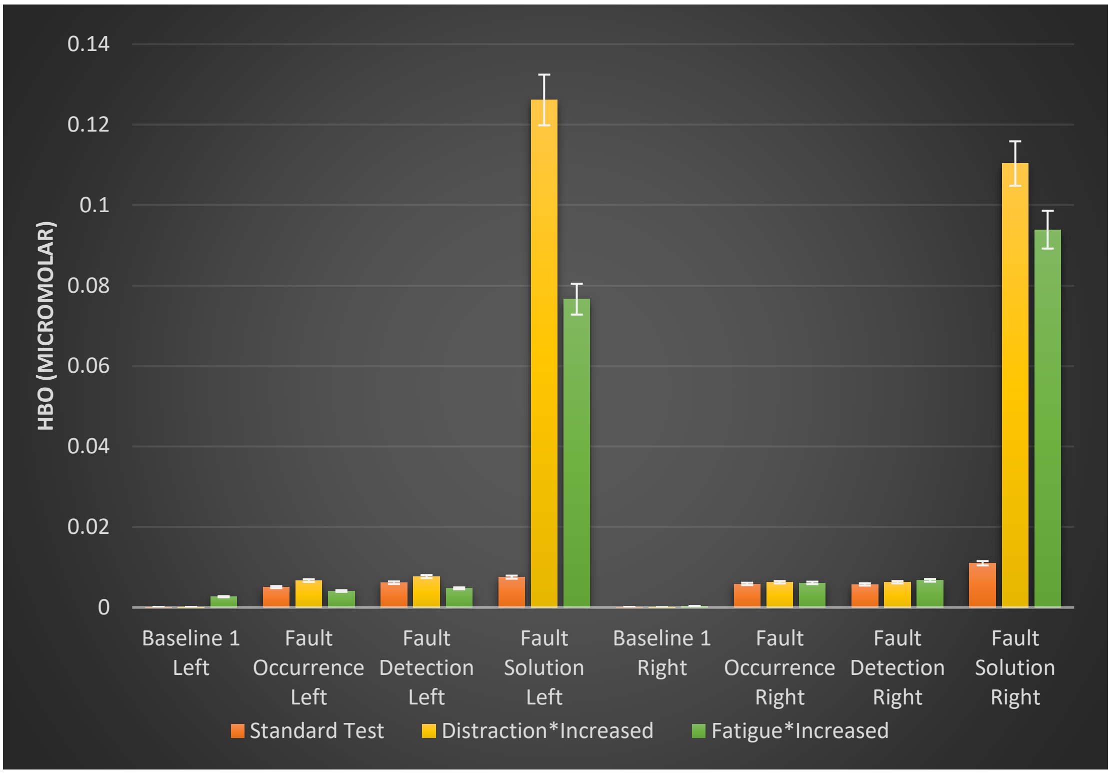


*Figure - Combined PSF's*workflow with respect to HBO*

The figure below is a bar chart showing the mean HBO taken from participants for the full duration of the exercise, against combined PFS.

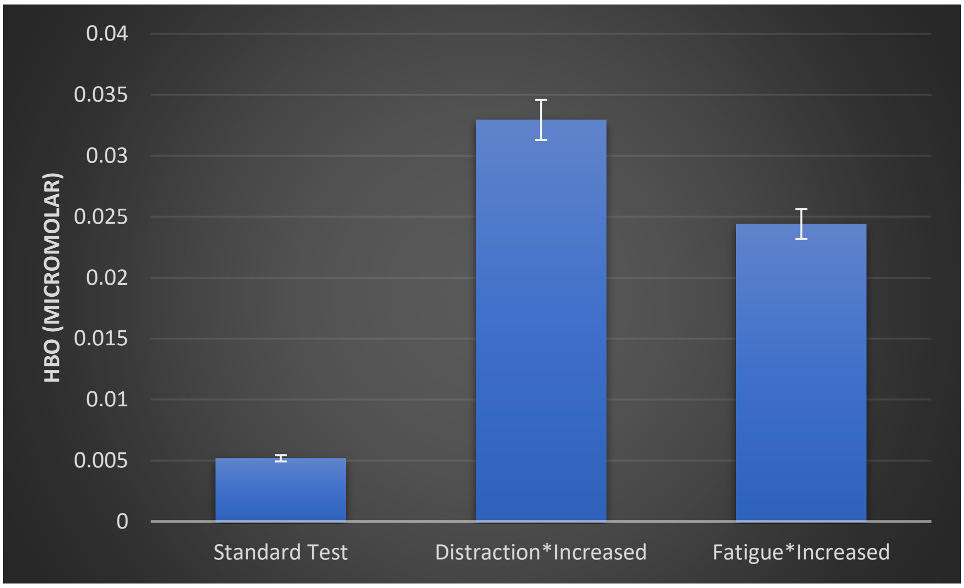


Figure – Comparison of combined PSF's

The effect of combined PFS against a Standard test
A significant effect was found for a combined PFS [F(3,25) = 142.398, P<0.01].

Pairwise comparisons show significant effects for combined PFS p<0.01 as depicted below in the table.

Pairwise Comparisons

Measure: MEASURE_1

(I) Standard 1, INC*Dist 2, Inc*Fat 3	(J) Standard 1, INC*Dist 2, Inc*Fat 3	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b] Lower Bound	Upper Bound
1	2	-.028*	.002	.000	-.034	-.022
	3	-.019*	.002	.000	-.025	-.013
2	1	.028*	.002	.000	.022	.034
	3	.009*	.002	.003	.003	.014
3	1	.019*	.002	.000	.013	.025
	2	-.009*	.002	.003	-.014	-.003

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Table - Pairwise comparisons of combined PSF's

Maulchy's test of sphericity showed a significant effect therefore tests of within subject effects using Greenhouse-Geisser showed [F=405.907, p<0.001, eta^2 = 0.938].

Due to the significant effects shown in the above pairwise comparisons. T-tests were conducted to see the size of each effect. A standard test was used as Level (j) for all tests.

- Distracted*Increased workload (level i) vs standard test (level j) t = 0.028/0.002 = 14, p<0.001

- Fatigue*Increased workload (level i) vs standard test (level j) t = 0.019/0.002 = 9.5, p<0.001

The Effects of combined PSF with respect to workflow
The ANOVA showed a significant effect for combined PSF with respect to workflow [F(3,26) = 69.997, P<0.01, eta^2 = 0.890.

Pairwise comparisons show significant effects for combined PSF, mainly for the fault solution stage of the workflow (p<0.01) as depicted below in the table.

Measure: MEASURE_1

Standard 1, INC*Dist 2, Inc*Fat 3	Workflow	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	1	.000	.000	.000	.001
	2	.005	.001	.004	.007
	3	.006	.001	.005	.007
	4	.009	.006	-.003	.021
2	1	5.843E-5	.000	-.001	.001
	2	.006	.001	.005	.008
	3	.007	.001	.006	.008
	4	.118	.006	.106	.130
3	1	.002	.000	.001	.002
	2	.005	.001	.004	.007
	3	.006	.001	.004	.007
	4	.085	.006	.073	.097

Table - Pairwise comparisons of PSF's with respect to workflow

Maulchy's test of sphericity showed a significant effect therefore tests of within subject effects using Greenhouse-Geisser showed [F=93.246, p<0.001, eta^2 = 0.874].

Due to the significant effects showing in the above pairwise comparisons, t-test were conducted in order to see the size of the effect. PFS was used as Level (i) and Fault solution for level (j) for all tests as this stage has the most significance.

- Standard test (level i) vs Fault Solution (level j) t = 0.009/0.006 =1.5, p<0.001
- Distracted*Increased Workload (level i) vs Fault Solution (level j) t = 0.118/0.006 =19.67, p<0.001
- Fatigue*Increased workload (level i) vs Fault Solution (level j) t = 0.085/0.006 =14.17, p<0.001

Effect of ROI
The ANOVA has shown no significant effects for ROI with respect to combined PSF or workflow.

Appendix E – Ship accident database data

References		**Performance Shaping Factors (PSF)**												
Accident code	**source**	**Fatigue**	**Distraction**	**Workload**	**Weather**	**Noise**	**Vibration**	**Loss of concentration**	**temperature**	**communication**	**procedure**	**time**	**Injury**	**Visibility**
m19co403	MTSB	1	1	1	1				1	1		1		
M18P0014	MTSB		1	1										
M03W0073	MTSB			1		1	1	1						
M01M005	MTSB	1	1						1			1		
M99F0023	MTSB	1										1		
M95N0011	MTSB			1				1					1	
Bitfjord 04/2005	MAIB	1												
SD DEXTEROUS 05/2010	MAIB			1										
THAMES 10/2009	MAIB		1											
HAVEN HAWK 06/2007	MAIB			1										
CORNER BROOK 03/2005	MAIB		1											
FRI STREAM 11/2006	MAIB			1										
KOCATEPE S 10/2006	MAIB	1										1		1

SALINE 02/2009	MAIB		1									1		
MV SONIA 29/2000	MAIB			1										
SUNNA 01/2007	MAIB	1							1		1	1		
HSS STENNA EXPLORE R 05/2003	MAIB			1										
STENNA PIONEER 03/2010	MAIB	1												
VANGUA RD 09/2004	MAIB			1										
WEST EXPRESS 07/2008	MAIB			1						1	1			
RMS ST HELENA 19/2001	MAIB	1						1						
MARIELLA 01/2008	MAIB			1										
NORSEA 16/2003	MAIB			1								1		1
WMS HARLING EN 05/2007	MAIB			1										
BREAKSE A 03/2006	MAIB		1											
CALYPSO 8/2007	MAIB	1							1					
PRIDE OF CANTERB	MAIB			1						1	1			

URY 22/2015														
ISLE OF INISHMO RE 07/2008	MAIB	1										1		
FINLANDI A SEAWAYS 2/2021	MAIB		1	1							1			
SAFMARI NE NUBA 03/2010	MAIB		1											
MORNES 3/2009	MAIB				1	1	1		1			1		
ARCO AVON 17/2006	MAIB	1		1							1			
MSC COLOMBI A 08/2007	MAIB	1												
WIGHT SKY 14/2018	MAIB		1											
MILLENIU M CITY 01/2008	MAIB			1										
HEBRIDES 20/2017	MAIB			1										
CELTICA HAV 1/2019	MAIB			1										
SEA BREEZE 14/2015	MAIB	1												
EDDYSTO NE AND RED	MAIB				1	1	1							

EAGLE 16/2018														
ALEXANDER TVARDOVSKIY 10/2013	MAIB			1										
SVITZER MOIRA 19/2016	MAIB				1	1								
ARROW 8/2021	MAIB			1						1				1
FRI OCEAN 26/2013	MAIB	1		1										
QUEEN VICTORIA 05/2008	MAIB					1	1							
200240	Nautical Institute (MARS)	1		1										
202018	Nautical Institute (MARS)										1	1		
201153	Nautical Institute (MARS)		1	1										
201454	Nautical Institute (MARS)		1	1						1				

200151	Nautical Institute (MARS)	1		1										
200317	Nautical Institute (MARS)		1						1					
200853	Nautical Institute (MARS)	1												
200653	Nautical Institute (MARS)			1							1			
200614	Nautical Institute (MARS)		1											
200908	Nautical Institute (MARS)			1										
201233	Nautical Institute (MARS)	1		1										

200337	Nautical Institute (MARS)	1						1						
200235	Nautical Institute (MARS)	1												
200513	Nautical Institute (MARS)		1	1										
201124	Nautical Institute (MARS)			1										
201146	Nautical Institute (MARS)		1											
2011X61	Nautical Institute (MARS)	1												
201734	Nautical Institute (MARS)			1										

201129	Nautical Institute (MARS)			1										
200962	Nautical Institute (MARS)			1										
200615	Nautical Institute (MARS)	1						1						
201622	Nautical Institute (MARS)		1	1										
201058	Nautical Institute (MARS)			1										
202133	Nautical Institute (MARS)	1												
202148	Nautical Institute (MARS)	1												

201746	Nautical Institute (MARS)		1	1										
201631	Nautical Institute (MARS)			1										
201209	Nautical Institute (MARS)			1										
201114	Nautical Institute (MARS)		1			1	1							
201567	Nautical Institute (MARS)	1								1	1			
201255	Nautical Institute (MARS)			1										1
201246	Nautical Institute (MARS)			1										

201782	Nautical Institute (MARS)		1											
201044	Nautical Institute (MARS)			1						1				
201879	Nautical Institute (MARS)	1		1								1		
201061	Nautical Institute (MARS)			1								1		
201312	Nautical Institute (MARS)	1		1										
201305	Nautical Institute (MARS)	1												
200970	Nautical Institute (MARS)		1		1									

201329	Nautical Institute (MARS)	1						1						
201929	Nautical Institute (MARS)			1										
201425	Nautical Institute (MARS)			1										
2011X02	Nautical Institute (MARS)	1	1											
201007	Nautical Institute (MARS)			1										
2011X06	Nautical Institute (MARS)	1			1									
201662	Nautical Institute (MARS)		1			1	1							

2018/000861	EMSA (Europe)			1										
2018/0001865	EMSA (Europe)		1											
2019/006699	EMSA (Europe)			1										
2018/000926	EMSA (Europe)			1										
2018/0001228	EMSA (Europe)	1												
2018/0001043	EMSA (Europe)			1										
2018/000518	EMSA (Europe)			1									1	
2762/2018	EMSA (Europe)			1										
2524/2018	EMSA (Europe)	1												
2633/2018	EMSA (Europe)		1	1										
2018/000101	EMSA (Europe)			1										
2505/2018	EMSA (Europe)	1									1			
2168/2018	EMSA (Europe)	1									1			

1884/2018	EMSA (Europe)	1	1	1										
2033/2018	EMSA (Europe)		1	1		1			1					
1204/2018	EMSA (Europe)			1										
1146/2018	EMSA (Europe)	1												
1059/2018	EMSA (Europe)	1	1			1								
2019/002177	EMSA (Europe)			1										
		38	28	61	6	9	6	6	7	7	10	12	2	4