Ocean Engineering 327 (2025) 120352

Contents lists available at ScienceDirect

Ocean Engineering

journal homepage: www.elsevier.com/locate/oceaneng

Research paper

An investigation of the effect of fatigue on ship engine room operators using *fNIRS*

Steve Symes [a,*], Eddie Blanco-Davis [a], Steve Fairclough [b], Zaili Yang [a], Jin Wang [a], Edward Shaw [a]

[a] School of Maritime Engineering and Technology, Liverpool John Moores University, L3 3AF, UK
[b] School of Psychology and Natural Sciences, Liverpool John Moores University, L3 3AF, UK

A R T I C L E I N F O

Keywords:
fNIRS
Simulator
Fatigue
Human reliability assessment
Operator functional state

A B S T R A C T

80 % of accidents in the maritime sector are due to human error. This could be the result of operator fatigue on top of daily tasks. This paper aims to evaluate the effect of fatigue as a Performance Shaping Factor(PSF) on seafarers. An engine room simulator study was conducted, using a TRANSAS 5000 series simulator, to investigate the influence fatigue has on human performance using a fault detection and correction task.

20 participants were recruited for the investigation; all 20 received training with the engine room simulator. The participants undertook a 30-min scenario where they had to detect and correct a fault. During this interaction, half of the participants experienced simulated increased fatigue. 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 use of fNIRS is the cornerstone of this studies novelty as brain-computer interface (BCI) fNIRS studies are rare in the maritime sector, and the use of BCI-fNIRS on engine room operators to assess their performance has never been done to date. The results indicated increased activation of lateral regions of the DLPFC during fault correction, this trend was significantly enhanced due to the addition of fatigue. 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. This approach can provide guidance on implementing effective risk control measures, automation strategies, and training programs. By improving risk assessment, identifying optimal work-rest schedules, developing targeted training programs and identifying tasks suitable for automation we can create a significant impact on maritime safety. By reducing error rates within the engineering sector, it has the potential to generate significant financial savings. This model can also be applied in other areas such as aviation transportation through the engineering sector. This model could also be tailored to assess the majority of high-profile roles where error would have huge consequences within various sectors.

1. Introduction

This study is an investigation into the effect of Performance Shaping Factors (PSF) on ship engine room operators. More specifically, this part of the investigation looks at fatigue as a PSF.

80 % of maritime incidents reported are linked to human error (National Transport Safety Board (NTSB)) (Bye and Aalberg, 2018) (Fan et al., 2019). In comparison with other sectors, maritime incidents are of a high financial significance (GOV, 2017) (National Transport Safety Board (NTSB)). Despite many efforts in maritime safety studies in general and specifically human error, it has been reported that human error alone costs the maritime sector millions of dollars per annum (European Maritime Safety Agency, 2017) (Bielic et al., 2017). Therefore, there is a research gap that needs more advanced methods to be developed with respect to human error and the seafarers' evolved roles in the development of the maritime industry. Thus, this project provides an in-depth investigation into the duties, training methods and PSFs that negatively affect operators within the engine room of a ship. Starting with the PSF fatigue. Fatigue is a significant safety concern in the maritime industry due to its potential to impair cognitive function and

* Corresponding author.
E-mail addresses: S.W.Symes@ljmu.ac.uk (S. Symes), E.E.BlancoDavis@ljmu.ac.uk (E. Blanco-Davis), S.Fairclough@ljmu.ac.uk (S. Fairclough), Z.Yang@ljmu.ac.uk (Z. Yang), J.Wang@ljmu.ac.uk (J. Wang), e.j.shaw@ljmu.ac.uk (E. Shaw).

https://doi.org/10.1016/j.oceaneng.2025.120352
Received 10 September 2024; Received in revised form 7 January 2025; Accepted 9 January 2025
Available online 18 March 2025

increase the risk of human error. Understanding the impact of fatigue on engine room operators is crucial for developing effective countermeasures and improving overall safety. The investigation can be achieved by using a neuroimaging technique called functional Near-Infrared Spectroscopy (FNIRS). FNIRS allows us to visualise the mental workload of human beings whilst partaking in their daily duties. FNIRS also offers a unique approach to measuring fatigue by providing real-time insights into brain activity. By monitoring changes in oxygenated and deoxygenated haemoglobin, fNIRS can detect the cognitive effects of fatigue, such as decreased attention, impaired decision-making, and reduced reaction time. The use of fNIRS in maritime research can provide valuable insights into the relationship between fatigue and human error. By identifying the specific cognitive processes affected by fatigue, researchers can develop targeted interventions to mitigate its risks and improve operator performance. FNIRS has not been widely used in maritime research, making this study a pioneering effort. By applying this innovative technique, the researchers aim to contribute to a better understanding of fatigue-related human errors in the maritime industry.

It is hypothesised that fatigue has a large detrimental effect on engine room operators. This project will answer the following research questions: 1 – What is the full effect fatigue has on engine room operators compered to standard operations? 2 – Are the effects of fatigue significant enough to cause human error? 3 – How does fatigue quantitatively affect the performance and safety of seafarers in the engine room?

Previous studies have used various Human Reliability Analysis (HRA) methods to analyse human error in the maritime sector (Akyuz et al., 2018) (Jahanshahloo et al., 2006) (Li et al., 2014). HRA is an approach used to identify the potential risk of human error events and to accurately estimate the Human Error Probability (HEP) using experimental data, modelling or expert judgement (Xi et al., 2017). The findings of this study could have significant implications for maritime safety. By identifying the specific factors that contribute to fatigue-related errors, researchers can develop evidence-based recommendations for improving training, workload management, and best practices in the maritime industry. This study aims to provide novel insights into the cognitive effects of fatigue and inform the development of effective countermeasures to improve maritime safety.

1.1. *Maritime human error*

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 (TRANSAS) (Fan et al., 2017) (GOV, 2017) (Verdiere et al., 2018). Another factor is that the majority of HRA studies are conducted with a focus on bridge operations from navigational perspectives (Gautier et al., 2016) leaving engine room errors unaddressed. These statistics warrant a full investigation into human error within the engine room. A full evaluation of the maritime databases (TRANSAS) (European Maritime Safety Agency, 2017) (GOV, 2017) (Takashi et al., 2006) (National Transport Safety Board (NTSB)) (Baker et al., 2018) was conducted in order to obtain the most common errors within the engine room. The accident databases were analysed with respect to incidents relating to the ship engine room only. The accident reports were consulted to see the specific PSFs reported as a contributing factor towards the errors. Reoccurring issues reported from the statistical analysis are multitasking at 20 %, and fatigue at 11 %. The tasks that showed to be the most consistent with human error are ballasting, oil transfer, machine maintenance, fuel system tasks and sea water treatment.

Fatigue, a critical Performance Shaping Factor (PSF), significantly impacts human performance (Bielic et al., 2017), particularly in demanding and safety-critical industries such as maritime operations (Chiarelli et al., 2017). When seafarers are fatigued, their cognitive abilities, such as attention, decision-making, and reaction time, deteriorate (Gevins and Smith, 2005). This can lead to increased errors, accidents, and near-miss incidents (Hitoshi and Kazuki, 2009). To mitigate the risks associated with fatigue, it is crucial to understand its impact on human performance and implement effective countermeasures. This study aims to investigate the neurophysiological effects of fatigue on seafarers' cognitive performance during critical tasks, such as fault detection and correction. By gaining insights into the underlying mechanisms of fatigue-induced performance degradation, we can develop evidence-based strategies to enhance maritime safety. One of the tools used in the analysis of fatigue as a PSF is an engine room simulator.

1.2. *Engine room simulator*

A TRANSAS ERS 5000 TechSim engine room simulator was used to carry out the operator analysis. The simulator (Fig. 1a) closely mimics a real container ship engine room. Utilising a high degree of realism, it allows real-time, real-life exercises to be conducted as they would be in the engine room of a real vessel (TRANSAS).

A scenario with exercises was designed and implemented on the simulator where candidates will participate in the task under the evaluation of the simulator instructor. The instructor sees the effect of the PSF and areas of the scenario where participants experienced significant mental workload outputs from fNIRS. Testing of participants on a simulated scenario allows for the implementation of a scientific human error model of the relationship between Operator Functional State (OFS) and adverse PSF's (Fan et al., 2017).

Candidates participating in the simulated scenario were connected to the neuroimaging device fNIRS. This provides the cornerstone of the project's novelty. Due to the weakness in current HRA methods within the maritime sector and the relative success of the aviation sector's use of fNIRS (Verdiere et al., 2018), it could be said that there is an urgency to implement fNIRS in maritime human error studies.

1.3. *Functional near infra-red spectroscopy*

Due to the relative transparency of human tissue which surrounds the skull (Gautier et al., 2016), infrared light can penetrate said cranial tissue (Takashi et al., 2006), coupled with this, haemoglobin absorbs infrared light (Baker et al., 2018). This facilitates visualisation of relative change in oxygenated and deoxygenated haemoglobin (Chiarelli et al., 2017). It is therefore possible to continuously, non-continuously, and non-invasively monitor the concentration of Oxygenated Haemoglobin (HBO) and De-oxygenated Haemoglobin (HBD) volumes within the human cerebrum (Felix et al., 2013). Fig. 1b shows the detection range of the infrared light.

Neurophysiological activation results in increased cerebral haemoglobin volume due to neurovascular coupling (Gevins and Smith, 2005).

Fig. 1a. Engine room simulator.

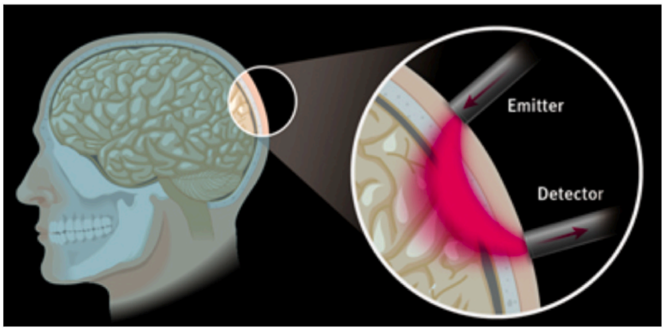


Fig. 1b. fNIRS infra-red detection range (Aghajani et al., 2017).

This coupling leads to a deviation in localised HBO and HBD volume (Felix et al., 2013). These changes in haemoglobin can be shown using the fNIRS system's infra-red emitters and detectors (Hitoshi and Kazuki, 2009), which transmit the information back to the NIRx (This is hardware that stores haemoglobin volume data from each participant) software, interfaced via computer. It is proven that the higher the HBO, the higher the mental workload of the participant (Shimizu et al., 2009). Therefore, this indicates the difference the PSF has on the operator's mental workload and functional state when compared to a standard task. Moreover, the higher the mental workload, the more potential there is for human error occurrence (Parasuraman et al., 2011).

For this study, participants were connected to the fNIRS system via a skullcap containing infrared light emitters and detectors. Together, the emitters and detectors create a channel. These channels can be combined showing readings from a large area, or analysed individually, to show readings from a specific area with respect to the cerebrum (Ayaz et al., 2017).

1.4. The research gap

A thorough literature search was conducted into human error within a ship engine room prior to the study. To date, there are no studies that evaluate the effect of fatigue in a ship engine room in the maritime sector. Moreover, there are very few studies that investigate PSFs within a ship engine room. There are a few studies in the aviation sector that analyse the effects of fatigue (TRANSAS) (Fan et al., 2017). However, these studies are for aircraft pilots and air traffic controllers. The techniques used in these studies have proved to be good assessment techniques of neurophysiological activation, however, there is no definition that the PSF being evaluated is fatigue even though it is assumed. Moreover, the tasks in the study mainly involved working memory and additional duties (autopilot vs manual landing/remembering in flight values) which could be defined as an increase in workload. This study specifically looks at fatigue as a PSF. Also, fatigue is induced in this study using techniques validated by the military (Verdiere et al., 2018).

Engine rooms are complex and often hazardous environments, requiring operators to perform tasks under pressure and with a high degree of concentration. Seafarers often work long hours in rotating shifts, which can disrupt sleep patterns and increase the risk of fatigue. Fatigue can impair cognitive function, leading to errors, reduced situational awareness, and increased risk of accidents. The unique demands of the maritime industry, such as isolation, confinement, and exposure to environmental stressors, can exacerbate the effects of fatigue.

Investigating fatigue in the engine room is crucial to understanding its impact on seafarers' performance and developing effective strategies to mitigate its risks. This research can inform the development of better work schedules, training programs, and technologies to improve safety and efficiency in the maritime industry.

1.5. Project aims and objectives

The main questions proposed from the maritime risk and safety sector are:

1. What are the main factors that are contributing towards human error within a ship engine room?
2. What is the significance of the PSF fatigue on human performance leading to error?

Objectives to achieve the above aims involve:

1. Analysing the maritime accident databases to obtain the main performance shaping factors (PSF) associated with human error within a ship engine room.
2. Developing a simulated engine room scenario incorporating said PSF compared to a standard engine room scenario.
3. Using fNIRS to measure operator functional states (OFS).
4. Modelling data to obtain the classification performance of participants.

2. Method

This section outlines how the PSF and task was identified, the experiment design, and the scenario phases. This section also looks at the optimum area of the cerebrum to evaluate with respect to human performance.

2.1. Identification of the performance shaping factor (PSF)

There are multiple hypothesis in relation to human error causation within the engine room. Meetings with experts in the maritime sector (conducted by the principle investigator), revealed that engine room operators are all trained to different competencies depending on the school/college/university in which they were trained (Meeting with John Carrier, 2017), some operators can cope better with work place factors (for example, fatigue, increased workload or distraction) compared to other operators (Meeting with Jonathan Warren, 2018). A multitude of maritime accident databases were accessed and analysed, to show incidents caused due to human error within the engine room only. The accident reports were then evaluated to see if there were any tasks with a high significance with respect to human error and PSFs reported as a factor contributing towards the human errors. Reoccurring issues reported from the statistical analysis are 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 statistical analysis were: ballasting, oil transfer, machine maintenance, fuel system tasks and sea water treatment system.

The results from the maritime accident databases indicate there are a number of PSFs that significantly influence human error. Fatigue, distraction and multitasking can be easily investigated due to the software capabilities of the engine room simulator allowing for an increased workload or long monitoring/watch keeping, and various alarms to distract operators (TRANSAS). However, engine room temperature and noise/vibration would require the use of external equipment or hardware. Therefore, the investigation increased workload/multitasking, distraction, and fatigue. The first PSF investigated was fatigue, as this scenario can be easily constructed using a long monitoring task. This would also fall in line with the duties experienced by operators on real operations (Hiteshk, 2017a) and adhere to the requirements set about by shipping regulators (Gausdal and Makarova, 2017). The tasks being conducted when human errors occurred showed to be random with no 'stand out' specific scenario. However, the task being conducted which had the highest frequency of human error occurrence is ballasting. Therefore, for the thoroughness of the investigation, ballasting will be used as the experimental scenario.

2.2. Experimental design and participants

Below we will look at the specific area of the human cerebrum that will be analysed during the experiment and the specific task that will be performed by each participant in this study.

2.2.1. The pre-frontal cortex

The dorsal lateral pre-frontal cortex (DLPFC) was the area of the cerebrum monitored throughout previous brain-computer interface (BCI) fNIRS studies using a continuous wave system in the following sectors: Automotive (Solovey et al., 2012), Rail (Kojima et al., 2004), and aviation (Verdiere et al., 2018) (Gautier et al., 2016) (Takeuchi, 2000). Monitoring of the DLPFC in the aforementioned studies can be further substantiated given that this is the area of the brain that governs executive functions like working memory, impulse, attention and cognitive flexibility, etc. (Parasuraman et al., 2011). Therefore, using 7 emitters and 7 detectors, creating 15 channels, DLPFC may be fully analysed. This montage is shown below in Fig. 2a (see Fig. 3a).

Candidates participated in a scenario (coupled with PSF's) on the engine room simulator, permitting the evaluation of the cause of any increases in activation. This is shown by the level of HBO (Thibalut et al., 2015) (the activation will be higher when the participant is under a higher mental workload (Verdiere et al., 2018)). However, to converge towards optimal accuracy, 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 (Aghajani et al., 2017). The middle region of the DLPFC is less utilised in practical or working memory tasks (Bruce et al., 2012). This has been shown in previous studies to cause anomalies in oxygenated haemoglobin data (Toppi et al., 2016) (Baker et al., 2018).

2.3. Workflow of the ballasting scenario

The workflow scenario design was based on the duties that would be carried out by a 2nd engineer (Hiteshk, 2017b) whilst detecting and correcting a fault with a steam powered pump on a ballasting scenario. A 2nd engineer's duties were chosen over the duties of a chief or 4th engineer as a chief engineer's role is a managerial role overseeing the duties completed by the 2nd engineer (Mobility, 2017). The 4th engineer, not common to most ships, is there as a trainee or assistant to a 2nd engineer (Hiteshk, 2017a). The participants will be trained prior to starting the task, as discussed in 2.2.3. A summary of the workflow stages is given below.

Each participant is required to complete five stages of the workflow. These stages include:

The first baseline will be used as a datum reading for each participant. Each participant will have a differing baseline output of HBO. Therefore, in the standard scenario, a 5-min baseline will be taken from the participant monitoring system screens. The fatigued scenario is significantly longer in order to induce fatigue (this is discussed in more detail in 2.2.3). This allows for an analysis of increased activation with respect to the individual participant for each workflow stage.

The second workflow stage is the fault occurrence stage. Unknown to the participants, after the 5-min baseline (or longer for the fatigue scenario) is taken, a system fault will occur, indicated by a red flashing light alarm in the top right-hand corner of the monitor. The participants were tested on the time taken and their ability to navigate to the correct system screens to acknowledge the alarm.

The third stage of the workflow is the fault detection stage. This part of the task requires the participants to navigate to the correct system screens; 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 prompted by the alarm codes.

The fourth stage of the workflow 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.

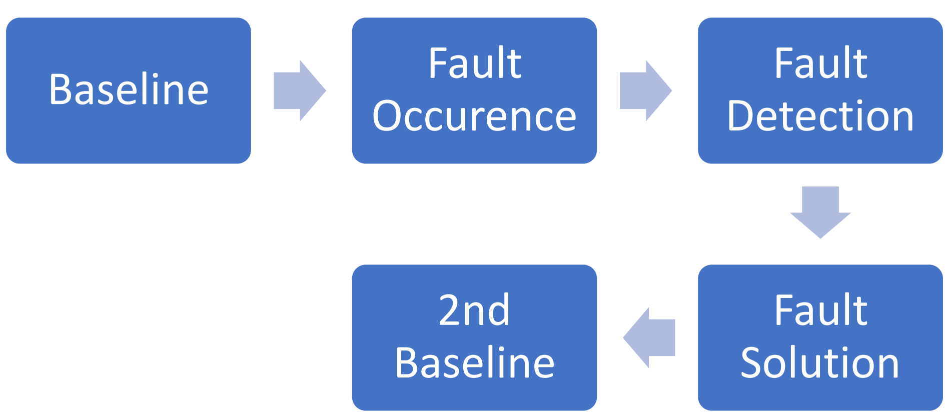
The fifth and final stage is the 2nd baseline. A second baseline is taken from each participant to see how their neurophysiological activation has changed compared to the first baseline. This final stage is the same as the first stage where participants are monitoring system screens.

2.3.1. The ballasting scenario

2.3.1.1. 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 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 standard test group was 28 and the fatigued group 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 master's year.

2.3.1.2. Participant training. All 20 participants had no prior experience with a ship engine room simulator.

The following training methods were conducted based on the engine room regulations set by the International Maritime Organisation (2019). The training course was developed with the assistance of experienced and qualified trainers (Meeting with Jonathan Warren, 2018) (Meeting with Geraint Phylip, 2017) (Meeting with John Carrier, 2017).



The 20 candidates were given a 2-h training session following a customised TRANSAS simulator trainee manual. The training session covered the 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

This was followed by 1 h of practical ballasting tutorials. The tutorials consisted of questions to answer and sub-tasks to complete in relation to the areas listed above. Participants were required to navigate the simulator system screens by answering the tutorial questions, completing the sub-tasks (aided by the trainee manual and their training notes), and generally familiarising themselves with the simulator system. This method was chosen based on the work by Christophe Faisey (Faisy et al., 2016), stating that participants with practical and theoretical 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 permitted to bring their own notes written during the training sessions, along with their personally annotated TRANSAS simulator manual whilst participating in the ballasting scenario. This was decided as engine room operators would have access to engine room manuals whilst at sea carrying out their duties (Kaushik).

2.3.1.3. The experiment

2.3.1.3.1. Baseline. For the first baseline, the participants were expected to monitor the LCS whilst ballasting from pump number two as shown in Fig. 2b (ballasting from pump two was set up by the instructor before the task started). The participants had no active input for the monitoring stage to allow for a 5-min (300s) baseline to be taken.

2.3.1.3.2. Fault occurrence. For the fault occurrence stage of the workflow, pump number two will fail. The participant must:

a) Orientate to the alarm as shown in Fig. 2c. The alarm is visual with no audio.
(b) Navigate to the alarm summary screen (see Fig. 2d) to record the details of the alarm.

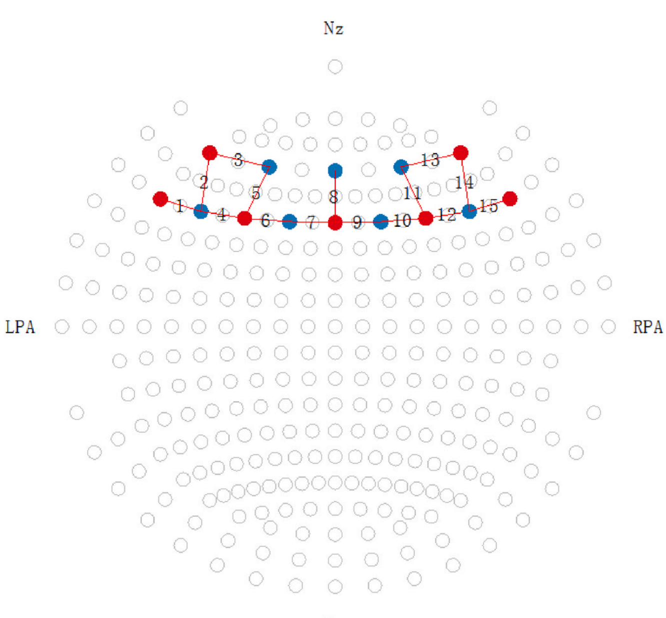


Fig. 2a. Skull cap montage of optodes.

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

2.3.1.3.3. 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 (Fig. 2b).
(b) Navigating to the ballast system screen to check the water line as shown below in Fig. 2e. The participant should 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.
(c) Accessing the cargo control room ballast pump screen (Fig. 2f) to check the pump pressure gauge (as prompted by the alarm summary screen in Fig. 2c).

2.3.1.3.4. Fault solution. The next stage of the workflow requires the participants to determine a solution to correct the fault. To correct this fault, participants must:

(a) navigate to the cargo control room ballast pump screen (Fig. 2f) and switch off pump number two,
(b) access the ballast system mimic panel (Fig. 2g 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.
(d) 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).
(e) Navigate back to the cargo control room ballast pump screen (Fig. 2f) to check that pump number 2 has power and switch the pump on.
(f) Re-access the ballast system screen to check that there is a water flow through the new pump as shown in Fig. 2h below.
(g) Return to the LCS to identify the new flow rate as shown in Fig. 2b.

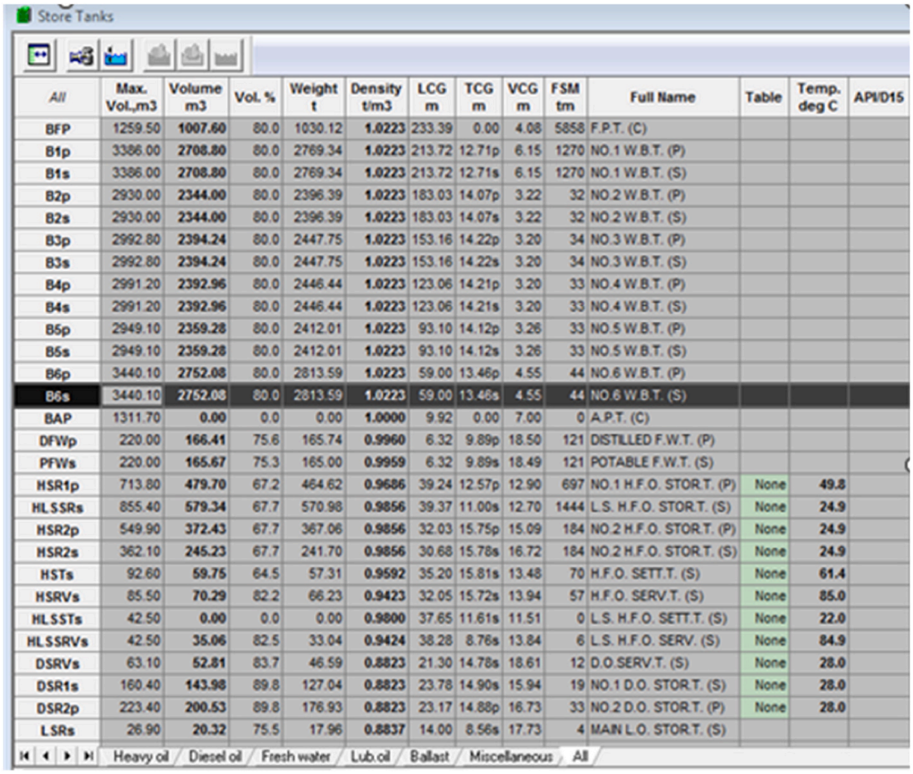
2.3.1.3.5. 2nd baseline. The last stage of the workflow requires the participants to continue to monitor the LCS until the tank has filled to the required volume set by the instructor before the task.

2.3.1.4. The addition of fatigue to the scenario. This experiment replicated the previously mentioned stages for a fault occurrence to the 2nd Baseline. The difference being that all fatigue participants would have a 35-min (2100s) monitoring task instead of the normal 5 min (300s) for the 1st baseline stage of the workflow.

3. Analysis

To quantify the data presented by fNIRS, a modified version of the Beer-Lambert law is used (Bu et al., 2018) as the Beer-Lambert law alone can only be used on non-scattering data (Solovey et al., 2012). Therefore, it cannot be applied to biological tissue without modifying the law to allow for light scattering (Shimizu et al., 2009). The raw data is filtered using the NIRx software and exported into Excel, as a large numerical table of time in frames, measured against oxygenated haemoglobin volumes. Typically, there is an output of 15 columns of oxygenated and deoxygenated haemoglobin results from all 15 channels. A correction-based signal improvement algorithm is applied to the data as described below.

To investigate the impact of fatigue on seafarers' cognitive performance, we employed a rigorous research design combining behavioural and neurophysiological measures. We assessed participants' performance on simulated fault detection and correction tasks, measuring key



All	Max. Vol.,m3	Volume m3	Vol. %	Weight t	Density t/m3	LCG m	TCG m	VCG m	FSM tm	Full Name	Table	Temp. deg C	API/D15
BFP	1259.50	1007.60	80.0	1030.12	1.0223	233.39	0.00	4.08	5858	F.P.T. (C)			
B1p	3386.00	2708.80	80.0	2769.34	1.0223	213.72	12.71p	6.15	1270	NO.1 W.B.T. (P)			
B1s	3386.00	2708.80	80.0	2769.34	1.0223	213.72	12.71s	6.15	1270	NO.1 W.B.T. (S)			
B2p	2930.00	2344.00	80.0	2396.39	1.0223	183.03	14.07p	3.22	32	NO.2 W.B.T. (P)			
B2s	2930.00	2344.00	80.0	2396.39	1.0223	183.03	14.07s	3.22	32	NO.2 W.B.T. (S)			
B3p	2992.80	2394.24	80.0	2447.75	1.0223	153.16	14.22p	3.20	34	NO.3 W.B.T. (P)			
B3s	2992.80	2394.24	80.0	2447.75	1.0223	153.16	14.22s	3.20	34	NO.3 W.B.T. (S)			
B4p	2991.20	2392.96	80.0	2446.44	1.0223	123.06	14.21p	3.20	33	NO.4 W.B.T. (P)			
B4s	2991.20	2392.96	80.0	2446.44	1.0223	123.06	14.21s	3.20	33	NO.4 W.B.T. (S)			
B5p	2949.10	2359.28	80.0	2412.01	1.0223	93.10	14.12p	3.26	33	NO.5 W.B.T. (P)			
B5s	2949.10	2359.28	80.0	2412.01	1.0223	93.10	14.12s	3.26	33	NO.5 W.B.T. (S)			
B6p	3440.10	2752.08	80.0	2813.59	1.0223	59.00	13.46p	4.55	44	NO.6 W.B.T. (P)			
B6s	3440.10	2752.08	80.0	2813.59	1.0223	59.00	13.46s	4.55	44	NO.6 W.B.T. (S)			
BAP	1311.70	0.00	0.0	0.00	1.0000	9.92	0.00	7.00	0	A.P.T. (C)			
DFWp	220.00	166.41	75.6	165.74	0.9960	6.32	9.89p	18.50	121	DISTILLED F.W.T. (P)			
PFWs	220.00	165.67	75.3	165.00	0.9959	6.32	9.89s	18.49	121	POTABLE F.W.T. (S)			
HSR1p	713.80	479.70	67.2	464.62	0.9686	39.24	12.57p	12.90	697	NO.1 H.F.O. STOR.T. (P)	None	49.8	
HLSSRs	855.40	579.34	67.7	570.98	0.9856	39.37	11.00s	12.70	1444	L.S. H.F.O. STOR.T. (S)	None	24.9	
HSR2p	549.90	372.43	67.7	367.06	0.9856	32.03	15.75p	15.09	184	NO.2 H.F.O. STOR.T. (P)	None	24.9	
HSR2s	362.10	245.23	67.7	241.70	0.9856	30.68	15.78s	16.72	184	NO.2 H.F.O. STOR.T. (S)	None	24.9	
HSTs	92.60	59.75	64.5	57.31	0.9592	35.20	15.81s	13.48	70	H.F.O. SETT.T. (S)	None	61.4	
HSRVs	85.50	70.29	82.2	66.23	0.9423	32.05	15.72s	13.94	57	H.F.O. SERV.T. (S)	None	85.0	
HLSSTs	42.50	0.00	0.0	0.00	0.9800	37.65	11.61s	11.51	0	L.S. H.F.O. SETT.T. (S)	None	22.0	
HLSSRVs	42.50	35.06	82.5	33.04	0.9424	38.28	8.76s	13.84	6	L.S. H.F.O. SERV. (S)	None	84.9	
DSRVs	63.10	52.81	83.7	46.59	0.8823	21.30	14.78s	18.61	12	D.O.SERV.T. (S)	None	28.0	
DSR1s	160.40	143.98	89.8	127.04	0.8823	23.78	14.90s	15.94	19	NO.1 D.O. STOR.T. (S)	None	28.0	
DSR2p	223.40	200.53	89.8	176.93	0.8823	23.17	14.88p	16.73	33	NO.2 D.O. STOR.T. (P)	None	28.0	
LSRs	26.90	20.32	75.5	17.96	0.8837	14.00	8.56s	17.73	4	MAIN L.O. STOR.T. (S)			

Fig. 2b. The liquid cargo monitoring screen showing ballast tank readings (e.g. tank volume).

Fig. 2c. The ship alarm.

performance indicators such as cerebral oxygenation volumes and task completion time.

We employed statistical techniques, such as analysis of variance and linear discriminant analysis, to analyse the data. ANOVA allowed us to compare the performance of fatigued and non-fatigued participants, while LDA helped identify specific neural patterns associated with fatigue.

By combining these methods, we aimed to quantify the impact of fatigue on seafarers' cognitive performance, identify neurophysiological markers of fatigue-related impairment, and develop a comprehensive understanding of the underlying mechanisms. Ultimately, these findings will inform the development of effective countermeasures to mitigate the negative effects of fatigue in the maritime industry.

3.1. *Correction based signal improvement (CBSI)*

CBSI is a technique used to improve the accuracy of the fNIRS signal based on the negative/transverse correlation between oxygenated and deoxygenated haemoglobin dynamics. Improving signal quality and reducing noise artifacts, especially noise induced by head motion, is challenging, particularly for real time applications. In an investigation of the properties of head movement induced noise, it was discovered that motion noise resulted in the measured oxygenated and deoxygenated haemoglobin signals, which are typically highly negatively correlated, to become more positively correlated (Xu et al., 2009). Therefore, the CBSI method was introduced to reduce noise based on the rule that the concentration changes of oxygenated and deoxygenated haemoglobin should have a negative correlation (Baker et al., 2018).

This is done by using the equation:

CBSIn = (HBO/2) - (δHBO/2) * HBB

where δ represents the standard deviation. (1)

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

CBSI(Ch1 row 1) = 0.5 * (Ch1 row 1 HBO) – ((δHBO (column 1)/ δHBB (column 1)) * (Ch1 row 1 HBB) where:

- CBSI (Ch1 row 1) represents the "CBSI" value for the first row of channel 1.
- HBO and HBB are variables or data points related to the channel and row.
- δHBO (column 1) and δHBB (column 1) represent the differences or changes in HBO and HBB values within the first column. (2)

HBO = Oxygenated Haemoglobin, HBB = Deoxygenated Haemoglobin,

The CBSI data from each candidate was separated into three sets. Channels 1 to 5 took haemoglobin readings from the left side of the

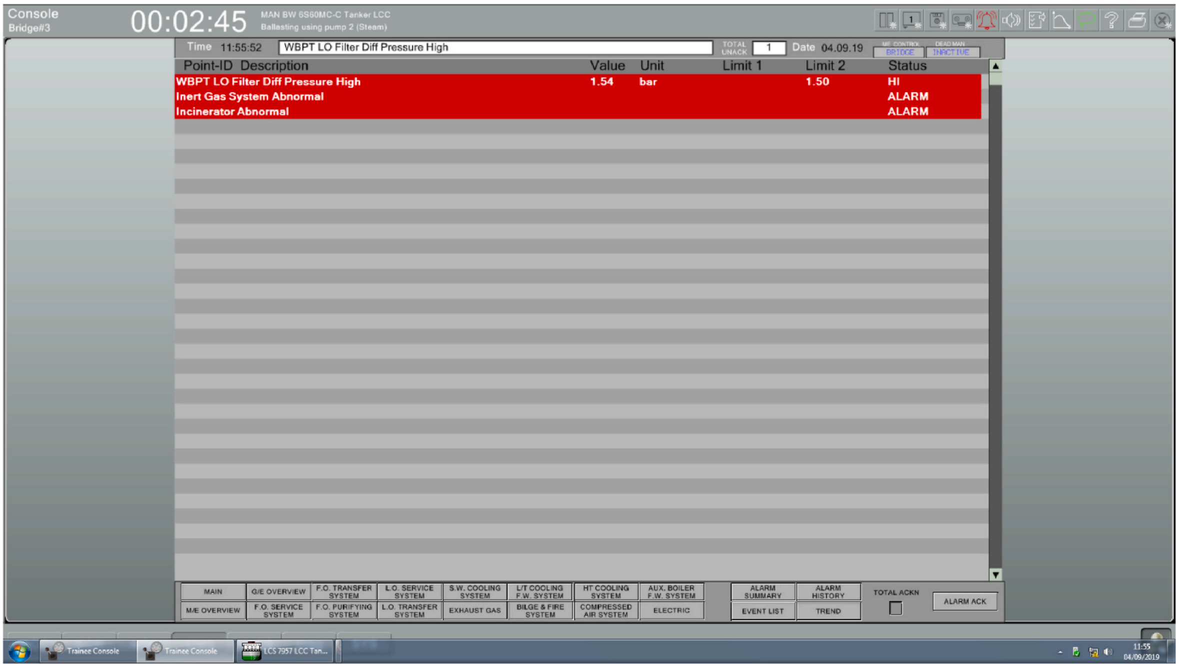


Fig. 2d. Alarm summary screen.

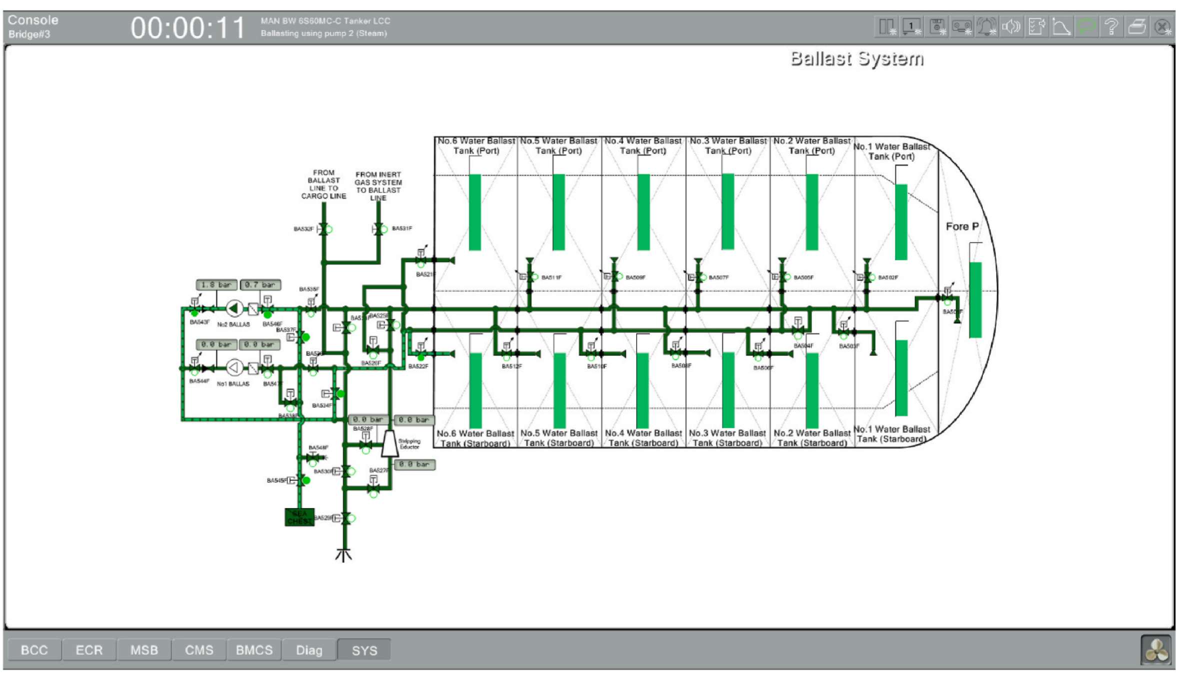


Fig. 2e. Ballast system screen showing water flow through pump 2.

DLPFC, channels 6 to 10 from the middle and channels 11 to 15 from the right. This is for the ease of analysis and allows the visualisation of the specific parts of the DLPFC in use whilst participating in the study. This is also useful due to the varying functions of the right and left sides of the DLPFC (van den Heuvel and Sporns, 2013). The left side is linked with verbal commands/receiving auditory input, word reading, processing information, linear and logical thinking (Heike et al., 2016). The right side is linked with visualisation, spatial reasoning, and interpreting information (Shimizu et al., 2009).

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

3.2. *Statistical package for the Social sciences (SPSS)*

SPSS is a software package that can analyse data files from many varying formats (for example: R-studio, Excel and MatLab). SPSS allows the analyst 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

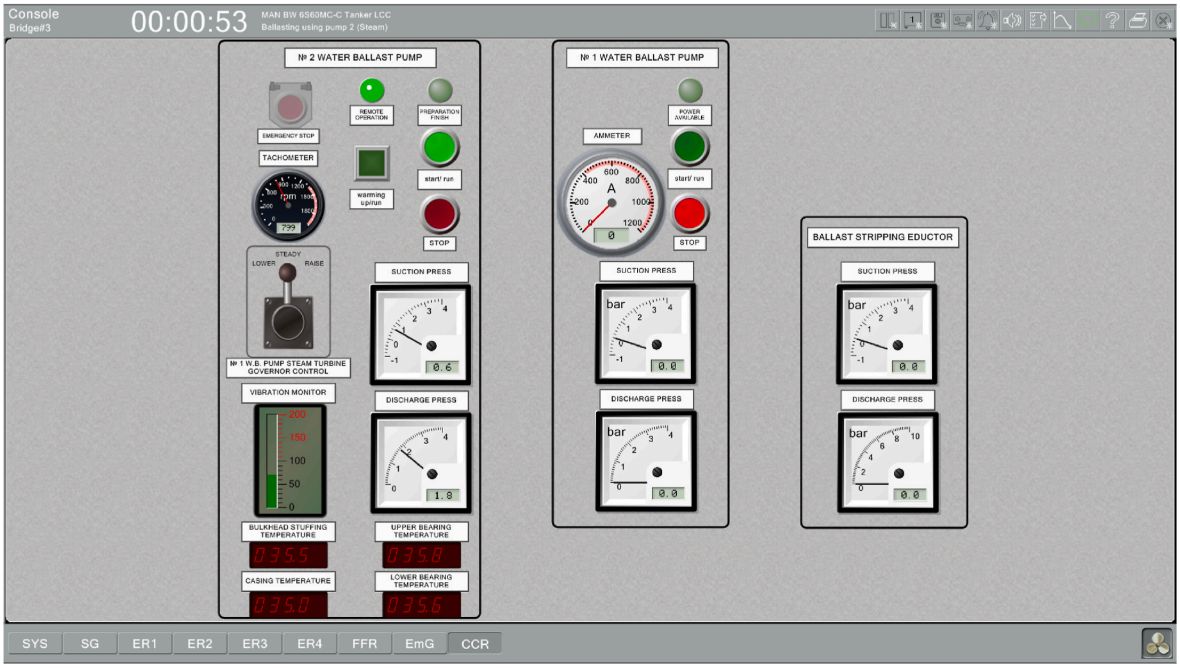


Fig. 2f. Ballast water pump control panel.

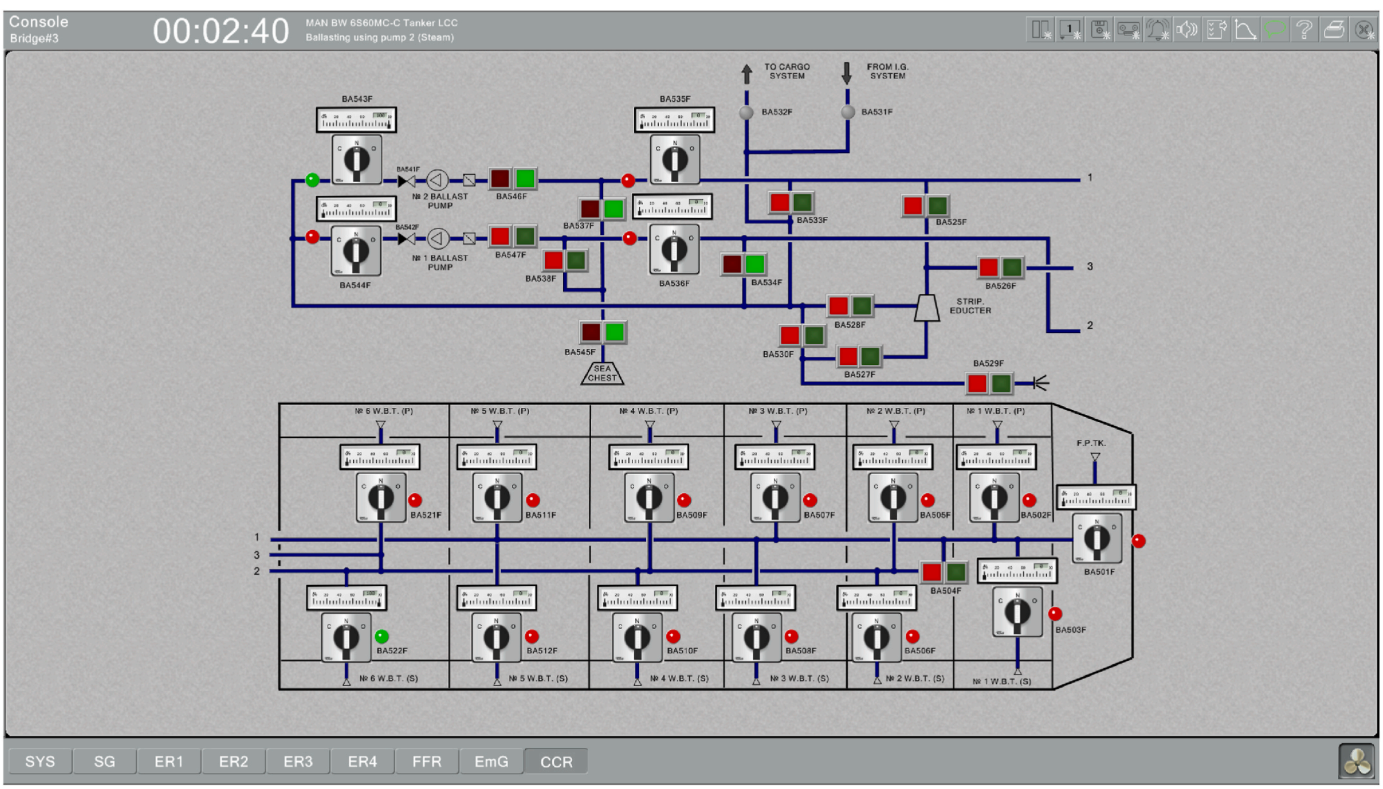


Fig. 2g. Ballast system mimic panel.

are closely related and the author had previous experience with SBSS hence, this was the software package used.

3.3. ANOVA analysis

ANOVA is used to evaluate how much of the total variance comes from the variance between the groups of participants, and how much from the variance within the groups of participants. This is done by implementing the following ratio:

$$\mathbf{F = Between\ Groups\ /\ Within\ Groups} \tag{3}$$

If a null hypothesis is true, then the F value will be close to 1.0. A large F ratio shows that the variation among the group means there is more of a variation than you would expect to observe than by chance

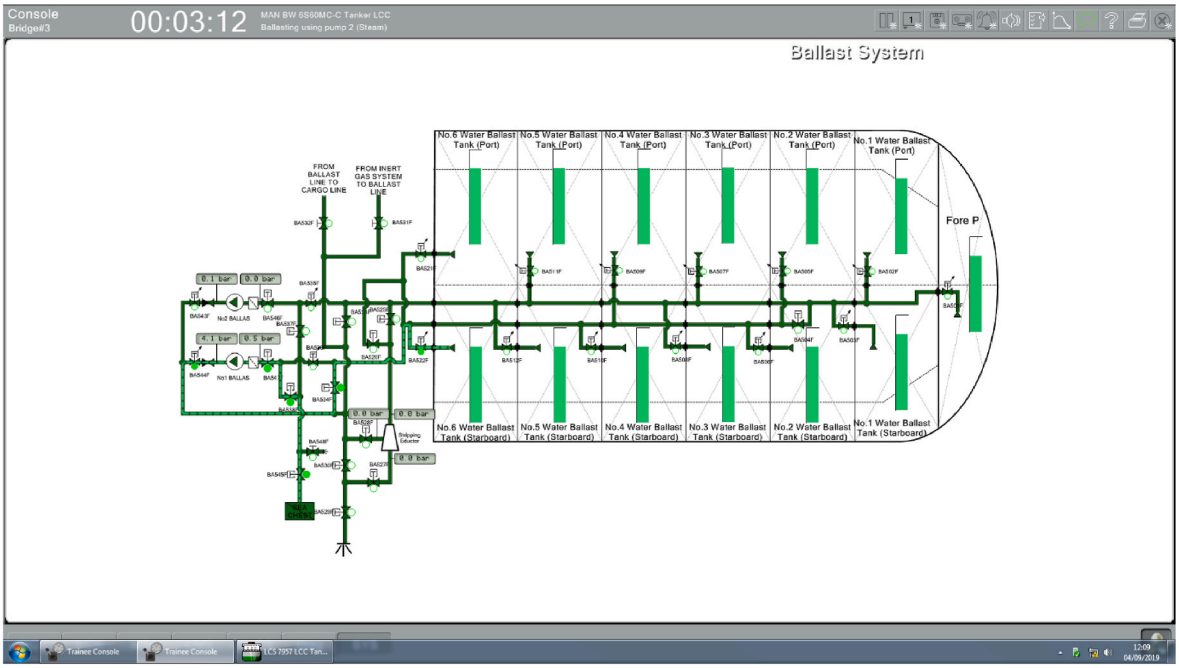


Fig. 2h. Ballast system showing water flow through pump 1.

(Zar, 2010). This calculation is done with respect to the degrees of freedom. For example:

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

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 established from 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 author can be confident in the accuracy of the results.

In summary, if most of the variation is between groups, then there is likely a significant effect. If most of the variation is within the groups, then there is probably not a significant effect (Hocke et al., 2018).

SPSS can go on to evaluate 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 workflow stage to see if any relate to one another (Wenlin and Haiming, 2018). Pairwise comparisons will be a useful evaluation 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.

3.4. R-studio

The R-studio software package is used to write and implement the relevant analysis code. The software sorts the data sets into time-based epochs. This indicates the effect of the PSF and workflow at very specific parts of the task and sub tasks. 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 (Verdiere et al., 2018).

3.5. Linear discriminant analysis

LDA is used as an operator performance classification model. LDA is a feasible option as it is effective at handling cases where the within class outputs are variable and where performance data is generated randomly (Gautier et al., 2016). The LDA method maximizes the ratio of variance between classes to variance within classes in any given data set, hence guaranteeing optimal separability (Aghajani et al., 2017). 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 (Bruce et al., 2012). LDA also provides a better understanding of the distribution of feature data.

3.5.1. Data classification

3.5.1.1. Data pre-processing. NIRS-analysis aids in the changing of raw fNIRS data to optical densities. In this process, the NIRS Star software is applied to remove artifacts and apply a band pass filter to the raw data. R-Studio was used to apply a wavelet interpolation method for artifact correction as this technique showed the highest signal to noise ratio in comparison to other artifact removal techniques available (Thibault et al., 2018). A high pass filter (cut-off: 0.01Hz – order 3) and a low pass filter (cut-off: 0.5 Hz – order 5) was used for the band pass filtering stage.

The artifact-free, filtered data is then converted into oxygenated (HbO) and de-oxygenated (HbR) haemoglobin concentrations. The data is then extracted and imported into an Excel spreadsheet. The CBSI formula was then applied as per the description in 3.1.

The CBSI processed FNIRS data was then extracted from Excel and imported into R studio. R studio is used to write the mechatronic code needed to sort the datasets into epochs. The HBO datasets for the entire task for one participant consisted of an average of 82,000 frames (~546s). The task duration (~484s + or – 15, ~546 + or – 33s, ~445 + or – 6.5) slightly differed for each candidate. The number of extracted epochs was fixed, based on the task duration. This resulted in 52 (approximately~8s) epochs. Each epoch was analysed independently as this showed the exact parts of the task that had the greatest levels of activation.

3.5.1.2. Oxygenation measures. Oxygenation measures are computed by applying HbO signals on each epoch for each of the 20 participants. The value x represents the HbO signal for one epoch (52 samples) and for one channel. 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:

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

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

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 evaluated by using the least-squared linear regression with the polyfit MATLAB application.

3.5.2. Feature extraction

3.5.2.1. Region of interest. To reduce the quantity of data for processing, the 15 channels were condensed to 3 Regions Of Interest (ROI): The left side of the DLPFC, the middle DLPFC and the right side DLPFC. The oxygenation features were extracted by formulating an average of all the available oxygenation features from the 15 channels included in the three ROIs. This gave us 6 viable measures for each epoch per participant.

3.5.3. Classification and cross-validation

For this study LDA with regularization of the empirical covariance matrix by shrinkage (the shrinkage method) was used, as this technique has proved in previous studies to be robust for use with Brain-Computer Interfaces (BCI) and passive BCI (pBCI) application (Verdiere et al., 2018) (Hasan et al., 2011) and also with fNIRS (Ayaz et al., 2017) (Thibalut et al., 2015).

3.5.3.1. The shrinkage method. The shrinkage method is used to penalize the less informative predictors resulting in a greater classification accuracy. Based on the prediction equation, shrinkage finds coefficients $\widehat{\beta}$ that decrease 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).

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

Workplace factor fatigue was separated via an intra-subject binary classification (e.g fatigued participant 1, standard test participant 0). Each candidate performed 1 of the 2 different workflow scenarios (standard or fatigued). Data was processed to obtain 52 × 8s–9.5s epochs per task for each participant. There were 10 candidates for each of the 2 differing workflow scenarios (10 × 52s epochs = 520 samples per workflow, 3120 samples in total). The performance prediction for this LDA classifier model was achieved by using stratified cross validation. This has previously proven to be a good trade-off between variance estimation and bias (Verdiere et al., 2018). The classifier was trained using 8 of the 10 participants (80/20 split [52 x 8 = 416 samples]) and used to predict datasets from two candidates (one of each workplace factor, i.e., 2 x 52 = 104 sample datasets). This method was applied for each of the two workflow scenarios (standard and fatigued) using the intra-subject binary classification system as mentioned above (fatigue (National Transport Safety Board (NTSB)) ~ standard test [0]) and the average performance of each subject was recorded. With Regards to the oxygenation features, three different types of comparisons were done. The first was evaluating each of the six oxygenation features separately. Secondly, features were combined to evaluate their mutual potential. They were merged into the highest performing couples (2 x 2) and the classification performance of each combined couple was evaluated (Thibault et al., 2018). The third comparison was done by combining all six oxygenation features. The results used in this study's evaluation will be a result of the best accuracy outcome of the three comparisons.

3.5.4. Statistical assessment

3.5.4.1. Subjective workload comparison. Paired sample t-tests were conducted to achieve a comparison of the average mental workload obtained in HbO for the workflow scenarios, the workflow stage and ROI amongst participants.

3.5.4.2. Classification performance significance. For a problem involving two-classes, the theoretical chance level for classification is 100/2 = 50 %. However, this is only accurate for an infinite sample size. To assess our classifier's significance or decoding error, the classification error was evaluated using a binomial cumulative distribution, as this has shown to work well in previous studies (Shimizu et al., 2009). The binomial cumulative distribution was calculated using the following formula:

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

Where P is the probability that an accurate class is predicted by at least 'Z' times, n is the sample number and c is the number of classes.

The performance classification was assessed by repeating the stratified cross validation for standard and fatigued tests and then averaging the results. As previously stated, our classification model was trained using 8/10 (80 %) candidates (416 samples) and tested on 104 samples. Using the cumulative binomial distribution, it sets the 5 % significance classification threshold at a **56.02 %** chance percentage.

3.5.4.3. Classification performance comparison. To compare the classification performance values for each oxygenation feature, a repeated measure ANOVA was used considering FEATURES (or COMBINED FEATURES) within factors.

4. Results

This section outlines the results from the ANOVA and LDA classification study of fatigued participants against standard test participants. Moreover, this section describes in detail the differences between groups nominally.

4.1. ANOVA results

The datasets from this study were analysed by ANOVA procedures using SPSS v.26. Outliers were identified as any value that deviated more than three standard deviations from the mean value and were omitted from ANOVA testing.

4.1.1. Analysis of fNIRS data

Average levels of oxygenated 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 the left, medial and right regions of the prefrontal cortex. All HbO data was 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) – (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

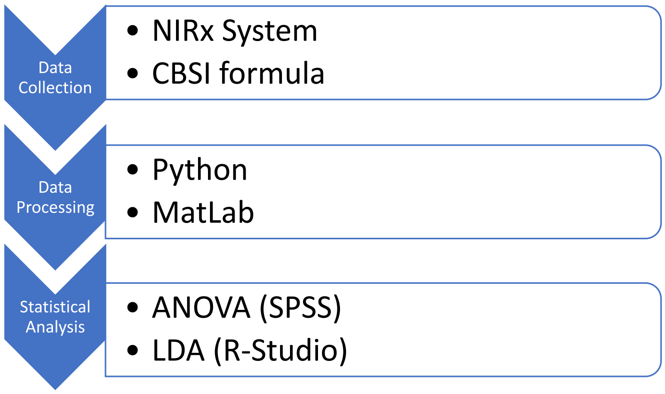


Fig. 3a. Flow chart representing the data processing steps.

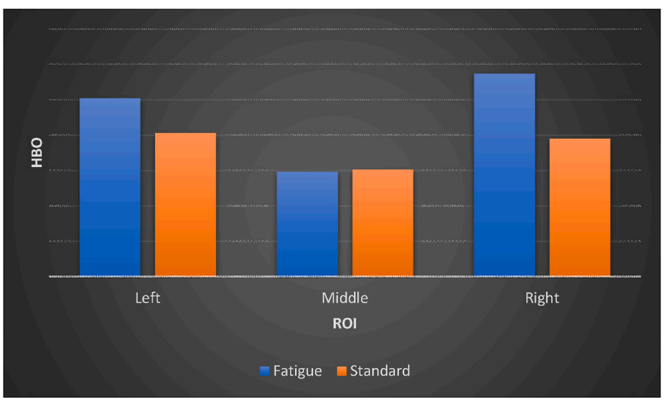


Fig. 3b. Average HBO for ROI with respect to fatigue stressor.

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 ($M = 0.019$, s.e = 0.005) compared to the standard group ($M = 0.009$, s.e = 0.002). For ROI, the main effect revealed that the mean HbO at the medial ROI2 ($M = 0.0008$, s.e = 0.002) was significantly lower than either the left lateral ROI1 ($M = 0.016$, s.e = 0.002) or right lateral ($M = 0.021$, s.e = 0.002) ($p < .01$) – see Fig. 3b below.

To explore the interaction, a number of post-hoc t-tests were conducted. These tests revealed that the mean HbO was significantly greater than 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 Fig. 3b for descriptive statistics.

4.2. LDA results

The datasets used in the LDA classification model were 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. The fatigue stressor is analysed individually against a standard test.

4.2.1. *Classification of fatigue and individual features with respect to chromophore*

Fig. 3c shows the classification performance percentage of fatigued candidates for each of the six oxygenation features, using HBO signals. To compare the classification performance percentages with oxygenation features, a repeated measure ANOVA study was conducted. A significant effect was discovered for oxygenation feature type on the classification performance percentage for fatigued participants [$F(5, 234) = 8.11$, $p < .01$]. Pairwise comparisons showed significant differences between oxygenation features for the HBO signals. Moreover, the 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 (>56.02 %).

Table 1 below shows the classification performance percentage of each oxygenation feature individual for each of the twenty participants. This indicates the effect of fatigue against a standard test by the resultant classification performance percentage. The greater the classification performance percentage, the greater the difference between a standard test and fatigued participants, thus, the greater the effect of fatigue on OFS (Verdiere et al., 2018). Again, similar to what is mentioned previously in 3.2.1, there is a trend between a greater classification performance percentage and the oxygenation feature AUC.

5. Discussion

This section discusses the findings from the ANOVA and LDA results. Furthermore, this section details the specific differences between groups, participants, and oxygenation features.

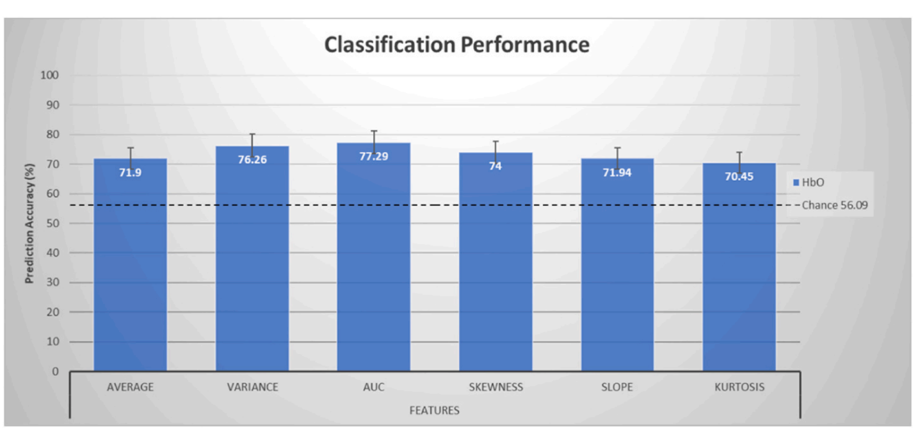


Fig. 3c. Classification performance of Fatigued participants.

Table 1
Fatigued vs a standard test classification performance 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

5.1. ANOVA

No significant effect was shown for fatigue for the fault detection workflow phase. This was not expected as the mean HBO for fatigued participants showed a greater HBO for all workflow phases combined when compared to standard test participants. An explanation for no significant effect being found could have been due to the units of measurement being such minute increments. Another reason could have been due to the fault detection workflow stage being short. Therefore, there may not have been enough time or subtasks to show the effect of fatigue. However, a significant effect was found for the fault solution workflow phase. This suggests that fatigued participants only began to struggle in comparison to the standard test group, when the sub-tasks became longer and involved a greater amount of system navigation. This outcome is contradictory to the investigation findings of Bu, Lg et al. (Bu et al., 2018) where fatigued candidates were found to have larger levels of neurophysiological activation from the start of their task. The main differing factor between the two investigations is that in Bu et al.'s investigation, the majority of the participants are elderly. Therefore, the age difference would have been a factor in the larger HBO volumes induced by fatigue earlier in the task (Aghajani et al., 2017). The fatigue analysis for our study was maritime-based (done on a ship simulator) whereas, the other fatigue studies were conducted using differing equipment with respect to the other engineering sectors (aerospace (Verdiere et al., 2018) (Thibalut et al., 2015) (Thibault et al., 2018), automotive (Bruce et al., 2012) (Hitoshi and Kazuki, 2009), National Rail Networks (NRN) (Kojima et al., 2004) (Takashi et al., 2006) and video gaming (Li et al., 2018)). However, due to the other studies being similar to this investigation in the sense that they are also 'BCI' studies of PSFs conducted on a simulator, they share many similarities with our investigation. Manipulation of the scenario to induce fatigue in participants was done in a similar way to other studies in other engineering sectors, where a monotonous visual of a readout is monitored by the participant for an elongated period of time (35 min for our study, between 20 min and 1 h for others). The outcome from the two NRN investigations referenced above showed that fatigue, similarly to our investigation didn't find any significance until the latter stages of the test. However, the aerospace and automotive investigations showed fatigue to have a significant effect on OFS and performance from the start of the task. This may be due to the slightly impoverished nature of the simulator in comparison to a 'real life' situation for ours and the NRN investigations, in which no 'real life' consequences are experienced when a task is failed, or an incident occurs. The participant would simply start the task again. Whereas an A300 and A320 aircraft simulator (the hardware used in the aerospace investigations) and an automotive simulator with hydraulically simulated movements, have a very realistic feel and consequence for failure from the beginning of the experiment. Another explanation, as previously mentioned, could have been due to the units of measurement. However, the same units of measurement and software was used by Dehais (Thibalut et al., 2015) and Verdiere (Thibault et al., 2018) for their investigation and fatigue was found to have a significant effect from the start of the test.

Fatigued participants in this investigation had a larger HBO volume on average for all workflow stages, in comparison to standard test participants. The difference in HBO volume increased further for the fault solution workflow stage. It can be seen in the results in section 3.1 that the fault solution stage had a significant difference when compared to the other workflow stages. The resulting differences found in mean HBO volumes between fatigued and standard test participants for the fault solution phase had such a large effect that it resulted in fatigued participants having an overall higher HBO volume for the whole task. Moreover, the HBO volumes for all workflow phases except the fault solution stage are similar for both, fatigued and standard candidates. Looking at the workflow phases critically, it could be said that if the Fault detection phase was longer, then the fatigue element of the test could have had a higher prevalence. This can be visualised by the large difference in HBO volume between fatigued and standard test participants for the longer, and higher workload, fault solution phase. Similarly, this is also found in the investigation done by Besikci et al., where fatigue was found to be more prevalent in the longer tasks even if they were deemed to be easier (Elif et al., 2015). Besikci concluded by stating that the fatigue element of their investigation was induced using verbal reasoning tasks over a long period of time (2 h). There was a short break (3–5 min) before starting their main test. All candidates showed similar results for the first 10–12 min of the test then differences started to occur. By the end of the test (25 min) all candidates had a significantly greater HBO volume when compared to those who partook in a standard test.

5.2. LDA

The main motivation behind this investigation was to use BCI-fNIRS technology to develop a scientific human error model that could be used to assess OFS 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 an artificially induced fatigued scenario. This led to significantly higher average HBO levels for the whole task when compared to a standard test.

The classification performance result confirmed that fatigued participants could be discriminated from standard test participants using BCI-fNIRS technology. This is substantiated by past neuro-ergonomics investigations finding that fNIRS is well suited for operator mental state monitoring in ecological situations (Aghajani et al., 2017) (Bruce et al., 2012) (Verdiere et al., 2018) (Li et al., 2018).

The highest classification performance accuracy percentage reached 77.29 %, taken from the AUC oxygenation feature. This result could be used as a datum against classification performance scores post implementation of training applications or risk control options. This would allow researchers to obtain an optimal risk control option or training intervention. This result compares favourably with other investigations. For instance, Kevin Verdiere et al. (2018) obtained a classification performance percentage of 66.9 % on 11 subjects using combined oxygenation features. Studies by Hong et al. (2015), Holper and Wolf (2012), Naseer et al. (2016) obtained a classification performance percentage of 75.6 %–81 % on 10–12 subjects. At first glance these outcomes compare similarly with ours however, these investigations did not take into consideration 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. Khan and Hong, 2015 (Khan and Hong, 2015) 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 percentage revealed that AUC and Variance resulted in significantly higher classification percentage values. This is similar to the investigation conducted by *Vierdiere* et al. (Verdiere et al., 2018) where AUC was also 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 considerably advantageous when passive BCI is implemented. Moreover, the oxygenation features computed in our investigation can consider both time and chromophore. Oxygenation features from fNIRS data has been used for a long period of time to evaluate operator performance in the aerospace sector, but up until now has not been utilised in the maritime sector for the analysis of engine room operators. Therefore, it is difficult to compare results from other maritime investigations. Based on the comparisons to other investigations above, undertaken using oxygenation features as a classifier we can say that our investigations outcome was a success. 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 HRA. It paves the way forward towards OFS estimation in an ecological maritime environment, but some challenges remain.

5.3. 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.

6. Conclusion

The study successfully demonstrated the potential of fNIRS-based BCI technology for assessing operator fatigue and performance in a maritime context.

To be critical, the results from this study being, that fatigue causes an increased activation in participants could have been predicted prior to this study. However, the level of activation that fatigue caused the participants could not have been predicted prior to this study. Also, the activation levels can be compared to that from other PSFs, resulting in a quantifiable value of risk associated with the PSFs. When compared to one another, this would allow decision makers to see the PSFs that contribute detrimentally to human performance thus, have the highest prevalence of risk.

The study successfully demonstrated the potential of fNIRS-based BCI technology for assessing operator fatigue and performance in a maritime context. Key findings revealed a significant impact of fatigue on task performance during complex phases, suggesting the effectiveness of fNIRS-based BCI for monitoring operator mental state. The study achieved a high classification accuracy (Felix et al., 2013) in discriminating fatigued participants from standard test participants, highlighting the potential of this technology for real-world applications. Specific oxygenation features were identified as promising indicators of fatigue, contributing to the development of a scientific human error model for maritime environments [61]. This research provides a foundation for future studies and applications in operational safety and performance improvement.

By leveraging fNIRS-based BCI technology, the maritime industry can gain valuable insights into operator fatigue and develop strategies to mitigate its risks [61]. Future research could explore the application of this technology in other high-risk environments, such as aviation and healthcare, to improve safety and efficiency.

While the study is a valuable contribution to understanding the impact of fatigue on seafarers, it's important to note that it does have limitations. A sample size of 20 participants, while sufficient for initial investigation, may not be representative of the entire seafarer population. It would have been preferable to have a larger sample size as a larger sample size would increase statistical power and improve the generalisability of the findings.

While the study attempted to simulate fatigue, it's difficult to perfectly replicate real-world fatigue conditions, which can vary significantly in terms of duration, intensity, and individual factors. It would have been preferable to explore the impact of different fatigue levels and durations on human performance, but time restraints would not allow this.

The simulator environment, while controlled, may not fully capture the complexities of real-world maritime operations, which involve various stressors and distractions. A simulator would always be a somewhat impoverished approach to real world applications.

The study focused on a specific fault detection and correction task. Real-world maritime operations involve a multitude of tasks and decision-making processes, which could be influenced by fatigue in different ways. However, such an in-depth analysis would have been too time consuming.

While fNIRS is a valuable tool for measuring brain activity, it has limitations, such as sensitivity to noise and movement artifacts. The artifacts are corrected using NIRx software coupled with MatLab and Python. However, the raw data processing is very time consuming and should be improved. The specific interpretation of fNIRS data can be challenging, and further research is needed to fully establish robust correlations between brain activity and cognitive performance.

The findings may not be directly applicable to all seafarers and differing vessels, as individual differences in fatigue tolerance and cognitive abilities can vary. Also, further research is needed to explore the impact of fatigue on different demographic groups and experience levels.

6.1. Recommendations for future research

- To Investigate the impact of task complexity and duration on fatigue-related performance changes.
- Explore the potential of fNIRS-based BCI for real-time monitoring and intervention in other maritime operations. For example, bunkering operations, port piloting operations or other adverse working conditions (hot/cold temperatures, noise & Vibration, distractions etc).
- Conduct larger-scale studies to further validate the findings and establish norms for fatigue assessments in maritime settings.
- Future studies could incorporate additional physiological and self-reported measures. For example; heart rate variability (HRV), electroencephalogram (EEG), and salivary cortisol levels. The aforementioned measures could provide objective evidence of fatigue. Additionally, participants could complete self-reported fatigue scales at regular intervals to capture subjective perceptions of fatigue. By combining these measures with the existing task-based simulation, the study may be able to provide a more comprehensive and convincing assessment of the impact of fatigue on human error in the engine room.

By addressing these recommendations, future research can further refine the application of fNIRS-based BCI technology in maritime environments, enhancing safety and efficiency.

CRediT authorship contribution statement

Steve Symes: Writing – review & editing. **Eddie Blanco-Davis:** Supervision. **Steve Fairclough:** Supervision. **Zaili Yang:** Supervision. **Jin Wang:** Supervision. **Edward Shaw:** Investigation.

Declaration of competing interest

The authors declare the following personal relationships which may be considered as potential competing interests: Dr Steve Symes reports a relationship with Liverpool John Moores University School of Engineering that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Aghajani, H., Garbey, M., Omurtag, A., 2017. Measuring Mental Workload with EEG Plus fNIRS, vol. 11, 1.
Akyuz, E., Celik, M., Akgun, I., Cicek, K., 2018. Prediction of Human Error Probabilities in a Critical Marine Engineering Operation On-Board Chemical Tanker Ship: the Case of Ship Bunkering, vol. 110, pp. 102–109.
Ayaz, H., Shewokis, P., Bunce, S., Izzetoglu, K., Willems, B., Onaral, B., 2017. Optical brain monitoring for operator training and mental workload assessment, 1 (1).
Baker, J., Bruno, J., Gundran, A., Hosseini, S., Reiss, A., 2018. fNIRS measurement of cortical activation and functional connectivity during a visuospatial working memory task, 1 (1).
Bielic, T., Hasanspahic, N., Culin, J., 2017. Preventing marine accidents caused by technology-induced human error, 31 (1).
Bruce, M., Bryan, R., Solovey, E., 2012. Brain Sensing with fNIRS in the Car.
Bu, L., Wang, D., Huo, C., Xu, G., Li, Z., Li, J., 2018. Effects of Poor Sleep Quality on Brain Functional Connectivity Revealed by Wavelet-Based Coherence Analysis Using NIRS Methods in Elderly People, vol. 1, 1.
Bye, R., Aalberg, A., 2018. Maritime navigation acidents and risk indicators: an exploratory statistical analysis using AIS data and accident reports, 176 (1).
Chiarelli, A., Zappasodi, F., DI Pompeo, F., Marla, A., 2017. Simultaneous functional near-infrared spectroscopy and ectroencephalography for monitoring of human brain activity and oxygenation: a review, 4 (4).
Elif, B.B., Leyla, T., Ozcan, A., 2015. The subjective measurement of seafarers' fatique levels and mental symptoms, 1 (1).
European Maritime Safety Agency, 2017. EMSA Accident Investigation. EMSA [Online]. Available: http://www.emsa.europa.eu/fc-default-view/tagged/84-accident-investigation.html. November 2017.
Faisy, C., Davagnar, C., Ladiray, D., Djadi-Prat, J., Esvan, M., Lenain, E., Durieux, P., Leforestier, J., Marlet, C., Seijo, M., Guillou, A., 2016. 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, 11.
Fan, S., Yan, X., Wang, J., Jinfen, Z., 2017. A Review on Human Factors in Maritime Transportation Using Seafarers' Physiological Data. Frontiers, Liverpool.
Fan, S., Blanco-Davis, E., Jinfen, Z., Yang, Z., 2019. Effects of seafarers' emotion on human performance using bridge simulation. Ocean Eng. 1 (1), 1–3.
Felix, S., Stefan, K., Andreas, J.M., Raphael, Z., Juan, M.P., Ursula, W., Martin, W., 2013. A review on continuous wave functional near-infrared spectroscopy and imaging, 7 (1).
Gausdal, A., Makarova, J., 2017. Trust and safety onboard, 16 (2).
Gautier, D., Sebastien, S., Thibault, G., Arnauld, D., Frederic, D., 2016. Processing functional near infrared spectroscopy signla with a kalman filter to assess working memory during simulated flight, 1 (1).
Gevins, A., Smith, M., 2005. Neurophysiologic Measures for Neuroergonomics. Las Vegas, Nevada.
Gov, U.K., 2017. Marine Accident Investigation Branch - Reports. Uk Government [Online]. Available: https://www.gov.uk/maib-reports. (Accessed 16 October 2017).
Hasan, A., Shewokis, P., Scott, B., Kurtulus, I., Willems, B., Banu, O., 2011. Optical brain monitoring for operator training and mental workload assessment, 1 (1).
Heike, S.M., Paul, E., Schulz, M., 2016. Genomics, circuits, and pathways in clinical neuropsychiatry - dorsolateral pre-frontal cortex, 5 (1).
Hiteshk, 2017a. Duties of the chief, 2nd and 4th engineers. https://www.marineinsight.com/careers-2/duties-of-2nd-engineer-on-ship/. October 2017.
Hiteshk, 2017b. Duties of 2nd engineer on ship. Marine Careers. https://www.marineinsight.com/careers-2/duties-of-2nd-engineer-on-ship/.
Hitoshi, T., Kazuki, Y., 2009. Measurement of brain function of a car driver using functional near infrared spectroscopy, 1 (1).
Hocke, L., Oni, I., Duszynski, C., Corrigan, A., Frederick, B., Dunn, J., 2018. Automated processing of fNIRS data-A visual guide to the pitfalls and consequences, 11 (5).
Holper, L., Wolf, M., 2012. ". Single-trial classification of motor imagery differing in task complexity: a functional near-infrared spectroscopy study.,". Neuroengineering and Rehabilitation. https://doi.org/10.1186/1743-0003-8-34.
Hong, K.-S., Naseer, N., Kim, Y.-H., 2015. Classification of prefrontal and motor coretx signals for 3 class fNIRS-BCI. Neurisci and Lett 5, 87–92. https://doi.org/10.1016/j.neulet.2014.12.029.
Jahanshahloo, G., Hosseinzadeh-Lotfi, F., Izadikhah, M., 2006. Extension of the TOPSIS method for decision making problems with fuzzy data, 181 (2).
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/.
Khan, M.J., Hong, K.-S., 2015. ". Passive bci based on drowsiness detection: an fNIRS study,". Biomedica (6), 4063–4078. https://doi.org/10.1364/BOE.6.004063.
Kojima, T., Tsunashima, H., Shiozawa, T., 2004. Functional Brain Imaging of a Train Driver by Functional Near-Infrared Spectroscopy (FNIRS), vol. 1, 1.
Li, K., Yin, J., Bang, H., Yang, Z., Wang, J., 2014. Transportmetrica A: Transport Science - baysian network with quantitive input for maritime risk analysis, 10 (2).
Li, Y., Zhang, L., Long, K., Gong, H., Lei, H., 2018. Real-time monitoring prefrontal activities during online video game playing by functional near-infrared spectroscopy, 11 (9).
Meeting with Geraint Phylip-Jones, 2017. Liverpool.
Meeting with John Carrier, Liverpool, 2017.
Meeting with Jonathan Warren, Liverpool, 2018.

Mobility, Global, 2017. Marine chief engineer in marine engineering. Logistics [Online]. Available: https://www.alchemyrecruitment.com/maritime-job-sectors/crewing/chief-engineer.

Naseer, N., Noori, F.M., Qureshi, N.K., Hong, K.-S., 2016. Determining optimal feature-combination for lda classification of functional near-infrared spectroscopy signals in brain-computer interface application. Front. Hum. Neurosci. https://doi.org/10.3389/fnhum.2016.00237.

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 16 October 2017).

Organisation, I.M., 2019. Training Regulations and Requirements - at Sea - Engine Room. IMO, London, UK.

Parasuraman, R., Christensen, J., Grafton, S., 2011. Neuroergonomics - study of the brain at work, 11 (1).

Shimizu, T., Hirose, S., Obara, H., Yanagisawa, K., 2009. Measurement of pre frontal cortex brain activity attributable to the driving workload and increased attention, 1 (1).

Solovey, E.T., Mehler, B., Reimer, B., 2012. Brain Sensing with fNIRS in the Car. Cambridge, MA 02139, 7th October.

Takashi, K., Hitoshi, T., Tomoki, S., Hiroki, T., Takuji, S., 2006. Measurement of Train Driver's Brain Activity by Functional Near-Infrared Spectroscopy (fNIRS), vol. 27, pp. 13–15.

Takeuchi, Y., 2000. Change in Blood Volume in the Brain during a Simulated Aircraft Landing Task, vol. 42, pp. 60–65.

Thibalut, G., Gautier, D., Francois, L., Sebastien, S., Frederic, D., 2015. Real-Time State Estimation in Flight Simulator Using FNIRS, vol. 1, 1.

Thibault, G., Hasan, A., Frederic, D., 2018. "In silico vs. Over the clouds: on-the-fly mental state estimation of aircraft pilots. Using a Functional Near Infrared Spectroscopy Based Passive-BCI," 1 (1).

Toppi, J., Borghini, G., Petti, M., He, E.J., De Giusti, V., He, B., 2016. Investigating cooperative behavior in ecological settings: an eeg hyperscanning study. PLoS One 11. https://doi.org/10.1371/journal.pone.0154236.

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].

van den Heuvel, M.P., Sporns, O., 2013. Network hubs in the human brain. Tends Cogn. Sci. (17), 683–696. https://doi.org/10.1016/j.tics.2013.09.012.

Verdiere, K., Roy, N.R., Dehais, F., 2018. "Detecting pilot's engagement using fNIRS connectivity features in an automated vs. Manual LAnding Scenario 1 (1).

Wenlin, L., Haiming, Z., 2018. The relationship between principal component analysis and factor analysis and SPSS software, 15 (3).

Xi, Y., Yang, Z., Fang, Q., Chen, W., Wang, J., 2017. New hybrid approach to human error probability quantification applications in maritime operations, 1 (1).

Xu, C., Signe, B., Allan, R., 2009. Functional Near Infrared Spectroscopy (NIRS) Signal Improvement Based on Negative Correlation between Oxygenated and Deoxygenated Hemoglobin Dynamics, vol. 15, 2.

Zar, J.H., 2010. Biostatistical Analysis, fifth ed.