Mental workload assessment using eye-tracking glasses in a simulated maritime scenario

F. Di Nocera\textsuperscript{1}, S. Mastrangelo\textsuperscript{2}, S. Proietti Colonna\textsuperscript{1}, A. Steinhage\textsuperscript{3}, M. Baldauf\textsuperscript{4}, & A. Kataria\textsuperscript{4}

\textsuperscript{1}Sapienza University of Rome, Italy - \textsuperscript{2}Ergoproject s.r.l., Rome, Italy - \textsuperscript{3}Future-Shape GmbH, Germany - \textsuperscript{4}World Maritime University, Malmö, Sweden

Abstract

The primary goal of this study was the assessment of maritime operators’ mental workload in simulation scenarios designed for containing typical traffic situations to be handled by officers during usual routine ship navigation. Taskload was manipulated throughout the sessions to analyze changes in ocular activity during complexity peaks. Specifically, it was tested the viability of implementing a dynamic analysis of eye movements collected through eye tracking glasses (while the operator was freely moving on the bridge) and using the distribution of eye fixations as an indicator of mental workload. Another objective of the present study was to assess the relation between attentional control and the mental workload as perceived by the operators. Results showed that the distribution of eye fixations changed with taskload and that individuals showing high attentional control reported low workload. Furthermore, frequent eye movement transitions have been found between the instruments monitored, suggesting that the information they provide could most usefully be integrated for improving the operators’ performance.

Introduction

According to the International Maritime Organisation (IMO) “Shortcomings in human performance at all levels in the chain of responsibility are a major cause of incidents” (IMO Resolution A.1060(28)). Starting from this consideration and assuming that enhancing human performance should be one of the primary goals of a “safer, more secure and environmentally friendly shipping”, it becomes imperative to find a way to objectively measure it. This quantification should be even more necessary given the growing proliferation of new technologies (e.g. e-Navigation systems, Unified Bridges, Remote monitoring systems, wearable technologies) whose impact should be carefully evaluated before their introduction. At this stage, institutional initiatives on this topic seem to be still insufficient. This is quite surprising if we consider that in the maritime domain any single on board material and equipment is tested from an engineering standpoint to match precise safety requirements (e.g. breaking load, tensile strength). Using a metaphor: we measure any single part of an anchor chain without appropriately considering its weakest
link. In this very case, the weakest link is the interaction between humans, technology, procedures and work environment.

This inconsistency was also noticed by the European Commission that -from 2012 to 2015- consequently funded three Human Factors research projects under the FP7 Programme Transport (SST.2012.4.1-1: Human Element Factors in Shipping Safety):

- FAROS - Human Factors in Risk-based Design Methodology (http://faros-project.eu);
- CASCADe - Model-based Cooperative and Adaptive Ship-based Context-Aware Design (http://www.cascadeproject.eu);
- CyCLaDes - Crew-centered Design and Operations of ships and ship systems. The latter one has been the shell of this study (http://www.cyclades-project.eu).

The present report describes part of the activities carried out in the last project. Mainly, our primary goal was the assessment of Officers Of the Watch (OOW) -who represent the human operators in the socio-technical system of an ocean-going ship-mental workload using eye-movement data collected through wearable eye-tracking glasses. Particularly, we tested the viability of implementing a dynamic analysis of eye movement data collected using eye tracking glasses while the operator was freely moving on the bridge using distribution of eye fixations as an indicator of mental workload.

Moreover, given that the lack of active attentional control can undoubtedly play a role in routine activities carried out on the bridge (where event pacing is slow most of the time, thus facilitating the operator distraction), another objective of the present study was to assess the relation between “absent-mindedness” (Cheyne et al., 2006) and the mental workload as perceived by the operators.

**Eye-tracking in maritime HF/E research**

The use of eye tracking technology in HF/E research and practice has considerably increased over the years. However, eye-tracking studies on the operator functional state carried out in the maritime context are few and cannot compare to others (e.g. aviation and automotive). There are several reasons for this gap. First, the focus on fatigue rather than mental workload didn’t help. Fatigue is an elusive concept to deal with and does not address very well the complexity of the interaction with technology. Moreover, the maritime context has only recently been adopting technological changes that need mental workload to be addressed. Nevertheless, expression of interest towards the use of ocular measures is increasing among researchers working in the maritime domain (Bjørneseth et al., 2012) and some authors have already employed the technique in several simulation studies. Lützhöft and Dukic (2007), for example, attempted an analysis using eye-tracking data in a pilot study aimed at investigating the ocular strategies adopted by experienced and novice navigators looking at the different instruments on the bridge. They reported...
that experienced navigators tended to show a lower ratio of glances per minute (a strategy that is probably related to their knowledge and practice with instrumentation). This effect was interpreted as an indication of lower mental workload, but neither behavioural nor subjective measures of mental load were used in that study.

More recently, Forsman et al. (2012) investigated gaze behaviour from both experienced and novice boat drivers during high-speed navigation at sea. They found that novice drivers tended to look at objects close to themselves to a larger extent than experienced drivers, whereas experienced drivers fixated objects in the far distance. Experienced drivers tend to rely more on environmental cues than novices who fixate more on the navigational aids. Results also showed that novices’ visual search strategies become less flexible as the workload increases. However, also in this case neither subjective nor behavioural measures of mental workload were used, and changes in workload were inferred only on the basis of the differences in difficulty between tasks or parts of a task.

Besides the scarcity of studies, it is worth noting the applied potential to maritime industry of what has been reported so far. Shipyards are increasingly using both simulations and eye-tracking technologies for their bridge consoles’ design/assessment processes. They also consider mental workload evaluation as a relevant part of these processes, even if there is still a lack of consensus on protocols and measures to be used to get meaningful results (see Bjørneseth, 2012).

One technical difficulty with eye-tracking in the maritime domain is that the operator frequently moves from one tool to the other, and only during specific phases of his/her activity sits in front of the bridge. Table-mounted eye-trackers are of little or no use in this context and different strategies for data collection should be employed. One of the aims of the present simulation study is to estimate the usefulness of wearable eye-tracking systems in this field.

Wearable eye-tracking systems shaped as glasses are relatively new devices that have seldom been used in experimental research, whereas there are plenty of marketing-related studies on consumers’ scanning strategies. Tonkin et al. (2011), for example, compared consumers’ visual behavior when searching for an item located on virtual or physical shelves and Gomes et al. (2014) investigated consumers’ attention to labels on beverages. Other studies (e.g. Mormann et al., 2014, Varela et al., 2014 & Wästlund et al., 2015) showed that eye movements between packaged products on a shelf may provide valuable information for making considerations about buying behavior. Nevertheless, the interest towards this type of instruments has been growing and wearable eye-trackers have been used for investigating the gaze patterns of students attending classes (Rosengrant et al., 2012), for evaluating the ease of reading with different page layouts (Chanijani et al., 2015), and for gathering valuable information to improve nurses’ medicine-tray preparation task (Rodriguez et al., 2014). An increasing interest towards this technology is also emerging in HF/E research. For example, Gable et al., (2013) recorded eye movements using eye-tracking glasses while performing a primary
driving task and secondary list-search on a mobile phone and their results showed differences in fixation data between workload conditions. Other studies have investigated the potential of these instruments for developing automated systems in the field of information security (Neupane et al., 2015) and for studying the effect of the curved display on visual performance and user experience (Choi et al., 2015). However, some limitations exist. For example, in a study involving harvester operators in routine field operations, Häggström et al. (2015) reported that large head movements, lightening changes and vibrations may affect the use of these systems in real-world settings. Another issue is what indicators should be used for gathering valuable information from eye-movement data. Several ocular indicators exist and could be utilized in any domain: from the usability of consumer products to operator functional state assessment (see Duchowski, 2007 for a detailed description of standard metrics). However, many indicators require eye movement recordings in controlled settings, allowing little or no movement to the user/operator. Of course, this is a substantial limitation in those settings where the individual is required to move around a room for accomplishing a task.

Based on previous research on the distribution of eye fixations and its relation with mental workload (see Di Nocera, Camilli & Terenzi, 2007; Proietti Colonna et al., 2014) we wanted to test the feasibility of implementing a dynamic analysis of eye movement data collected using eye tracking glasses while the operator was freely moving around the bridge. For this reason, we monitored OOW interacting with a simulated bridge during a typical coastal traffic scenario: navigation through a sea area with established traffic separation scheme, overlaid by Vessel Traffic Service (VTS) monitoring, with medium to dense traffic, including crossing ferries and selected off-shore activities. Taskload was manipulated throughout the sessions to analyse changes in ocular activity during complexity peaks.

**Method**

**Participants.** Twenty seafarers (2 females, mean age = 34.61 years, std dev = 6.25) participated in this study. Participants’ experience in the maritime domain ranged from 1 to 25 years. All of the participants were qualified navigation officers with Certificates of Competency in accordance with their designated rank.

**Simulator and scenario design.** A comprehensive study on mental workload assessment of OOW was performed using a series of real-time simulation trials. In our trials we studied the overall ship’s navigation process, which includes many sub-tasks all related to ensure the safe and efficient execution of a ship’s voyage from a port of departure to a port of destination. When standardizing minimum requirements for technical equipment for integrated ship navigation, the IMO mentions the following processes and tasks that an integrated navigation system shall contain to support the human operator, the OOW on a ship’s navigational bridge:
dynamic assessment of mental workload of navigators

- Route planning
- Route monitoring
- Collision Avoidance
- Navigation Control (manual and automatic control of ship’s movement)
- Navigation Status (a.o. monitor environmental conditions) and
- Alert Management

All those tasks need to be simultaneously and uninterruptedly performed by the responsible OOW under all possible prevailing circumstances of environmental conditions, characterized by sea area, water depth, visibility, wind (force, direction) and waves (heights, direction), etc. as well as of the ship status (actual draught, loading conditions, speed/course through water and over ground etc.). The overall aim of the OOW is to ensure, that any risk (of a collision, grounding, etc.) is at least at an acceptable level, and any danger of an accident is avoided. For this objective the human operator is obliged to use all his senses (e.g. for visual look out for collision avoidance) but also “all available means” appropriate to the prevailing circumstances (as e.g. Radar/ARPA, Automatic Identification System for collision avoidance or ECDIS or paper charts, position and depth sensors for grounding avoidance).

The Maritime Risk and Safety (MaRiSa) Simulation laboratory at the World Maritime University (WMU) provides a unique combined ship-handling (SHS) and safety and security training (SST) desktop simulator specifically developed for training and research purposes.

With the bi-directional system, complex simulation scenarios can be provided for various research purposes. The entire complex ship is available to replicate the ship-handling process on the navigational bridge (SHS) combined and connected with engine processes from the engine room (ER) and engine control room (ECR). Moreover, emergency management situations and procedures even inside the vessel as e.g. for fire-fighting, water inrush can be integrated into a simulation scenario with the help of the SST simulator. Figure 1 depicts the configuration of the combined SHS and SST simulator at the MaRiSa simulator laboratory at WMU.

As an example, the desktop SHS bridge configuration used for the workload assessment studies corresponds to the SOLAS (International Convention for the Safety of Life at Sea) conformant bridge equipment of a RoRo (Roll-on/roll-off: named according to the way vehicles board and leave the vessel) and RoPAX (RoRo with passenger capacity) ferry. Besides the variable 135° view, it features the elements of the basic navigational equipment including ECDIS, Radar/ARPA with integrated overlay of AIS information, interfaces of GPS; DOLOG (tool that measures vessel’s speed), echo sounder and further devices as well as the handles for rudder, engine and thrusters.
This unique configuration allowed for a wide range of simulation scenarios for

This desktop simulation environment has been proven sufficient during series of simulation experiments e.g. investigating potential effects of newly developed applications in the frame of IMO’s and IALA’s e-navigation initiative and a number of European projects on research and technological development on its impact on behaviour of human operators, like CyClaDes and ACCSEAS (e.g. Baldauf, Benedict & Krüger, 2014; Holder et. al., 2014).

For the purpose of the workload investigations in this research, simulation scenarios have been designed for a series of simulation runs containing typical traffic situations to be handled by navigating officers during usual routine ship navigation. The simulation was run at the MaRiSa facility, providing a typical traffic situation for coastal traffic. In order to create varying workload conditions, challenging situations have been integrated into a typical navigation scenario of 20 min length and requiring careful navigation when coming from an open sea-like area, and then following the approach and passage in the coastal sea area with a traffic separation scheme with a typical medium to dense traffic load situation. Moreover, shore-based VTS monitoring was also added and integrated with the simulation runs, therefore contributing to the realism therein. Through VTS communication, all participants were provided with an update on the local weather conditions and the traffic situation in the area. Furthermore, the VTS doubled up as other communicating entities, such as when participants called other vessels in the vicinity on their Very High-Frequency radio (VHF). The sea area was Northern Öresund (close to Helsingborg – Helsingør).

Figure 1. Configuration of the combined Ship-handling and Safety and Security Training Simulator at the MaRiSa-Simulation Lab of World Maritime University, Malmo, Sweden.
**Bridge instrumentation.** During this study, participants interacted with simulated onboard instrumentation. Familiarity with the features of these instruments was assessed using an *ad-hoc* five-point Likert scale. Only 3 participants reported a substantial difference with the type of devices they have previously used. The following is a brief description of the simulated tools and parts of the bridge that were taken into consideration in this study as Areas of Interest (AOIs; figure 2) and for which the transitions of eye movements (between instrument pairs) were examined:

- View of the sea, other ships, land, etc. (OUT)
- Electronic Chart Display & Information System (ECDIS): a computer-based navigation information system that complies with IMO regulations and can be used as an alternative to paper nautical charts.
- Global Positioning System (GPS).
- Automatic radar plotting aid (ARPA) that can calculate the tracked object's course, speed and closest point of approach, thereby knowing if there is a danger of collision with the other ship or landmass.
- Steering and engines controls (CONSOLE).

*Figure 2. Areas Of Interest (AOIs) from up to down and from left to right: OUT, EDCIS, GPS, ARPA, CONSOLE.*
**Design.** The taskload for the acting OOW’s was varied by including crossing ferries and selected off-shore activities (dredging, fishery, maintenance operation). Activity levels were consequently divided into three phases:

- 5 minutes of navigation in easy condition (Easy);
- 10 minutes with increased level of difficulty (Hard);
- 5 minutes of decreased level of complexity (Easy).

The level of complexity was controlled by situations with increasing risk of collisions and additional activities like communication requests via VHF from VTS and/or other targets. The VTS called the participant’s ship upon perceiving the development of a potentially dangerous situation, such as the possibility of the participant entering the opposing traffic lane of the Traffic Separation Scheme to keep clear of buoys and vessels in the area such as crossing ferries and dredgers.

This design allowed to test the sensitivity of the Nearest Neighbour Index (NNI: Di Nocera, Camilli & Terenzi, 2007) to changes in taskload by using the phase change from a low taskload to a high taskload condition (Easy to Hard) and ***vice versa*** (Hard to Easy) as factor in the successive analyses. This was needed because the end of the Hard condition did not necessarily coincide with the end of the evolving situation created by the increased difficulty. The phase (Easy vs. Hard vs. Easy) was instead used as factor for analysing the eye movement transitions between Areas of Interest (AOIs, see figure 2) coinciding with the instrumentation described in the “bridge instrumentation” subsection.

**Measures.** Measures collected included workload indicators derived from the eye-tracking data and subjective reports, as well as a subjective measure of attentional control. Eye movements were collected using the Tobii Eye Tracking Glasses (version 1): a plastic and rubber pair of glasses with lenses made of hot mirror glass and infrared (IR) reflecting coating. IR markers were placed around the displays to compute eye position. The sampling rate was 30Hz. Time series of the NNI values were successively computed using a newer Matlab version of the ASTEF package (Camilli et al., 2008). The index provides a single value that is indicative of the type of spatial distribution of the data by comparing the mean distance between pairs of (nearest) neighbors in the data to that expected by chance (random distribution). The actual mean distances can be smaller (points are clustered; NNI<1), larger (points are regularly distributed; NNI>1), or not different from the expected distances (points are randomly distributed; NNI=1). A total of 20 values (one each minute) was obtained from the raw gaze data recorded. Transitions between instrument pairs (transitions between each AOI and all the others) have been computed using the Tobii Studio software. The perceived level of workload was assessed using the NASA – Task Load indeX (NASA-TLX: Hart & Staveland, 1988). The level of attentional control has been evaluated using the Cognitive Failures Questionnaire (Broadbent et al., 1982) an inventory of everyday errors providing information on the respondent’s attentional control (the higher the score, the lower the attentional control; see Di Nocera, Ferlazzo & D’Olimpio, 2014).

**Procedure.** Participants filled a short ad-hoc questionnaire about the familiarity with the type of instruments that were simulated into the MaRiSa facility and the CFQ.
Prior starting the simulation they underwent a short calibration procedure and were briefed about the route to sail. The NASA-TLX was administered at the end of the simulation.

**Data analysis and results**

Concerning the VTS interventions (barring the standard VHF calls to provide weather and traffic updates to all vessels in the vicinity), nearly 80% of the VHF Calls coincided with the central 10 minutes of the simulation runs that pertained to high levels of complexity. That required the OOW to take note of the information and take requisite action as necessary in the form of exercising caution, maintaining fairway, altering speed, making evasive maneuvers, as the case may be.

Eye tracking data recordings were previously examined for completeness and validity. Six participants with incomplete eye-tracking recordings were excluded from the analysis. NNI values were used as dependent variables in a repeated measures ANOVA design using Phase Change (Easy to Hard vs. Hard to Easy) as factor. Results showed a tendency towards statistical significance (F1,13=3.97, p=.07): NNI values were higher for the passage from Easy to Hard rather than for the passage from Hard to Easy.

![Figure 3. Nearest Neighbour Index by Phase Change.](image-url)
To assess the role of attentional control on the perceived mental workload the Cognitive Failures Questionnaire median score was used to divide participants into two categories according to the median score: High Attentional Control and Low Attentional Control. Total NASA-TLX scores were employed as dependent variables in an ANOVA design using Attentional Control as factor. Results showed a statistically significant main effect ($F_{1,18}=6.38, p<.05$).

The proportion of transitions between all instrument were analysed in an ANOVA design Phase (Easy vs. Hard vs. Easy) x Instrument Pair (transitions from each AOI to the others). Results showed a significant interaction between these two factors ($F_{38,722}=2.12, p<.001$). Post-hoc Duncan testing showed that this interaction was due to the following comparisons:

- **ARPA-CONSOLE, ARPA-ECDIS, ECDIS-ARPA:** In all cases a significantly higher proportion of transitions was found in the second easy condition than the other two phases.
- **ECDIS-OUT and OUT-ECDIS:** In all cases a significantly lower proportion of transitions was found in the second easy condition than the other two phases.
- **ECDIS-CONSOLE and CONSOLE-ECDIS:** In all cases a significantly lower proportion of transitions was found in the first easy condition than the other two phases.
- **ARPA-OUT:** In all cases a significantly higher proportion of transitions was found in the hard condition than the other two phases.

![Figure 4. NASA-TLX raw scores by Attentional Control.](image-url)
Discussion and Conclusions

The use of eye-tracking in the maritime domain is still in its early stage. Nevertheless, the introduction of automated systems (and the technological change in general), as well as the consequent reduction of personnel, are boosting the investigation on mental workload and the quest for a good measure to be employed for gathering information on the on-going functional state of the maritime operator.

Results of the analyses carried out on the eye movement transitions between AOIs contributed to better understand the operators’ behaviour and could support the ongoing discussions about future ship bridge concept layouts. For example, the frequent ARPA-ECDIS-ARPA transitions indicate how these instruments are used comparatively and that the information they provide needs to be integrated by the operator. According to the proximity-compatibility principle (Wickens & Carswell, 1995), the displays relevant to a common task or mental operation (close task or mental proximity) should be rendered close together in perceptual space (close display proximity).

Most importantly, results of this simulation study provide a contribution to the development of a measure of mental workload based on the distribution of eye fixations. Indeed, this was a first attempt to employ the NNI index in such a complex setting and, more important, using eye-tracking glasses. Albeit not statistically significant (p=.07), most participants consistently showed the NNI discrimination pattern in phase change (10 out of 14). Many factors may account for this imperfect outcome: from individual differences to the fixation identification algorithm. Nonetheless, the fact that even in such a complex scenario and with a completely different technology a pattern is consistent is quite encouraging. Future studies will be run for addressing the best way to identify fixations (and consequently their distribution) with eye-tracking glasses.

Interestingly, results obtained from the present study also showed that attentional control (here measured in terms of more or less absent-mindedness) may be involved in operators’ workload perception. This is a quite interesting finding that it is worth to be further explored. Indeed, operators’ distraction may be crucial in those environments as the bridge of a ship, where event pacing is slow most of the time. Operators who tend to be more distracted (and one could argue that this is not much different from what is usually considered a loss of situation awareness) are those who perceive the task as most loading. Likely, this represents a cost of re-engaging a previously disengaged task, which is the cost of (internal) interruptions that has been reported elsewhere (see Foroughi et al., 2014), but not yet studied in the maritime domain.

Acknowledgements

This study was funded by the European Commission through the FP7 project “CyClaDes - Crew-centered Design and Operations of ships and ship systems”.
References


