

Towards an Experimental Design Framework for Evaluation of Dynamic Workload and Situational Awareness in Safety Critical Maritime Settings

Dr. Frøy Birte Bjørneseth
Rolls-Royce Marine AS
Aalesund, Norway
froy.bjorneseth@rolls-royce.com

Sathiya K. Renganayagalu
Aalesund University
College/ Maritime HF Lab
Aalesund, Norway
promison2k7@gmail.com

Dr. Mark D. Dunlop
University of Strathclyde
Glasgow, Scotland
mark.dunlop@strath.ac.uk

Dr. Eva Hornecker
University of Strathclyde
Glasgow, Scotland
Eva.hornecker@strath.ac.uk

Dr. Sashidharan
Komandur
Aalesund University
College
Aalesund, Norway
sash.kom@hials.no

Dynamic positioning (DP) systems are complex systems that challenges the operator's mind during complex offshore DP operations. It is both mentally demanding and the operator has to maintain constant situation awareness to be able to react in time to safety-critical situations. The research design that will be presented in this work-in-progress paper, aims at investigating the variations of high and low mental workload during dynamic positioning operations in the maritime domain using advanced eye-tracking equipment. This type of equipment is utilized to assess the operator's focal areas on the vessel's bridge during the operation and correlate the eye-tracking results with more traditional metrics measuring mental workload, such as heart rate monitoring and NASA-TLX. The experiment has been divided into four different sub-experiments, where the last iteration will compare workload assessments between a current and a new graphical user interface of a DP system. The results from these experiments will give valuable insight in DP operations and provide possibilities of tailoring placement of information from the DP system so that safety can be improved by supporting the operator during operation.

Maritime Domain. Dynamic Positioning. Eye-tracking. Mental workload. Usability. Software Evaluation.

1. INTRODUCTION

The main purpose of this chain of connected experiments is to evaluate the cognitive workload of the dynamic positioning operator of an offshore vessel, to identify focal points of attention during operation and evaluating/comparing two graphical user interfaces of DP systems with each other. This is based on the knowledge concerning the DP operators' work which is in a dynamic and visually challenging environment. This environment demands their attention regularly and they must monitor and process information, and also make decisions under conditions where task load varies across a range of their capabilities.

A Dynamic Positioning system (DP) can be defined as: *A computer controlled system to automatically maintain a ship's (figure 1) position and heading by using her own propellers and thrusters (Bray, 2003).*

Its technology has developed from the first simple systems in the 1960's to today's advanced systems covering single, double and triple redundancy according to the operation's safety critical level.



Figure 1: DP vessel (in courtesy of www.farstad.com)

1.1 Motivation

According to IMCA (2006) it is clear that operator error is the highest contributor of major loss of

positions of DP incidents (IMCA, 2006). Nocera et al. (2007) says:

“Human factors and ergonomics (HF/E) research continues to demonstrate that extreme levels of mental workload decrease an individual’s ability to react to incoming information and increase the likelihood of human error”.

Currently very little research there has been done in the maritime industry to reveal the operator’s real time capabilities and limitations resulting from varying workload levels. In the airline, automotive and power industry they have come much further. A recent example from the power industry by Holzinger et. al (2012), illustrates how it is possible to reduce the complexity of user-interfaces for safety-critical power-plant control systems.

From a DP standpoint concerning the GUI, it is of interest to find a basis of what kind of information is important to present in a DP GUI and identify the information that are not utilized during operation. This gives a base for optimizing the amount of information presented to the user. An identical study will be performed again with the new DP GUI and the results compared. In addition to studying the GUI itself the different foci areas of attention during the operation has never been investigated to such extent. This information can in later stages of development of future DP systems be utilized to place information where the operator intuitively rests the eye during operation.

1.2 Significance of Study

The principal reason for measuring workload is to quantify the mental cost of performing tasks in order to predict the operator and system performance. Mental workload refers to the portion of operator information processing capacity or resources that is actually required to meet system demands. In other words, workload measures are an attempt to characterize the performance of a task relative to the operator’s capability. In DP operations the operators experience fluctuations of mental workload because of task demands that range from high to low. Also occasional unexpected events may occur that generate the highest level of workload.

2. BACKGROUND AND RELATED RESEARCH

The role of the DP operator is mainly influenced by the design of the DP system: the automation in the system and its level of autonomy determines the work that is required from the operator. An operator monitors the DP process continuously during operations. The system is too complex to have a complete overview of what is happening in detail, so it is impossible for the operator to be aware of the exact status of all (sub-) systems all the time. This may cause a lack of situation awareness once

a fault occurs (Tjallema, 2007). The operator will have to search through the system to find relevant details, which provide input for his/her understanding of the fault and its root cause. This search process consumes valuable time in the problem solving process.

One of the primary ways a DP operator (DPO) gathers information is by visually scanning DP control screens and the surrounding environment. Willems et al. (1999) says;

“Visual scanning refers to a systematic and continuous effort to acquire all necessary visual information in order to build and maintain a complete awareness of activities and situations, which may affect the controller’s area of responsibility”.

Generally visual scanning consists two types of eye movements; fixations and saccades. Fixations are the moments when the eyes are relatively stationary, taking in or “encoding” information (Poole, 2006). Saccades are quick eye movements occurring between fixations (Poole, 2006). When the fixations and saccades are combined it produces a scan pattern, which is called a scan path (Ellis, 2009).

In addition to the above mentioned also measuring eye blinks propose interesting possibilities. Eye blinks are closely related to physiological factors including mood state and task demands and reflects the viewer’s attention and tension (Bruneau, 2002). Eye blink rate is inversely proportional to workload and high blink rate may indicate fatigue (Poole, 2006).

2.1 Area of Interest (AOI)

During DP operations the DPO needs a clear view of the aft deck of the vessel where loading and offloading equipment to offshore installations happens. Deck crew are present here and it is important that the DPO has control of the aft deck environment while at the same time communicates with the installation (e.g. an oil rig). However, concurrently the operator must monitor the DP system’s VDUs. The area of visual environment or display screens (VDUs) is a vital feature in this research. It is important to specify the area of interest in order to compile meaningful results. AOIs are task specific and are created to answer questions such as; did the participant look where we expected and what were the properties of their eye movements in the area investigated?

In the DP system the area of interest will generally include displays, control panels and the windows through which operator observes the environment. Specifically the DPs GUI, relevant displays and control panels will be marked as area of interest.

The number of AOIs are limited to only the interested research area and the capability of the eye tracker.

2.2 Correlation between fixations and workload

Eye fixations are sensitive to variations in mental workload (Nocera, 2007). The number of fixations overall is negatively correlated with search efficiency (Jacob and Karn, 2003) and search efficiency is negatively correlated with workload (Goldberg & Kotval, 1998). According to Bruneau (2002) speed, duration and precision of the saccades and the number of glissadic movements will be helpful in examining the fatigue and vigilance level of the operator.

2.3 Subjective workload measurement

In literature, researchers correlate mental workload with performance, physiological and subjective measures (Terenzi, 2008). Performance measures capture the operator's ability to complete the given task. Physiological measures such as heart rate variability and ocular behaviour which will be utilized in this experiment can capture the changes in operator's physiology. Subjective measures capture the operator's subjective assessment of cognitive workload involved in a particular task. NASA-TLX, Subjective Workload Assessment Technique (SWAT) and Instantaneous Self Assessment (ISA) are the rating scales often used by researchers for subjective workload measurement. Farmer et al. (2003) states;

“Subjective workload measures are useful in characterising the perceived demands of the task, which are an important aspect of any system and influence the operator's acceptance of the system”.

Sensitivity to the change in workload is one of the important criteria for selecting a rating scale (Terenzi, 2008). In that sense NASA TLX is more sensitive than the other two measuring scales (Farmer, 2003). The other criteria considered are reliability, validity, sensitivity, diagnosticity, practicality and intrusiveness. NASA TLX has high reliability and has been extensively validated through earlier research (Farmer, 2003). For this research project, NASA TLX has therefore been selected as a source of measuring subjective workload.

Another important aspect worth looking into is detection of emotion concerning user experience. Stickel et.al. (2009) states that;

“Reliable emotion detection in usability tests will help to prevent negative emotions and attitudes in the final products.”

2.4 Physiological workload measurement

Heart rate data is not a new workload measurement technique. Many researchers have already used it for measuring both physical and mental workload (Rehmann, 1982). Heart rate (in beats per minute or inter-beat interval (IBI)) has demonstrated evidence of reliability as a measure of operator workload. Heart rate is affected not only by changes in physical workload, but also by mental workload. Heart rate will increase, IBI will decrease, with an increase in workload (Corwin, 1989). Heart rate measures suffer from interaction with respiration, physical work and emotional strain. So this would require unique measures to isolate mental workload contributions (Cain, 2007). However heart rate variability is more reliable and effective than heart rate itself for measuring mental workload (Farmer, 2003). From earlier research by Murai et al. (2004) and Kum et al. (2007), it is clear that operator's mental workload can be effectively measured by heart rate variability.

2.5 Data Interpretation

According to Jakob and Karn (2003) eye tracking data analysis can proceed either top-down – based on cognitive theory or design hypotheses, or bottom-up – based entirely on observation of the data without predefined theories relating eye movements to cognitive activity (Goldberg, Stimson, Lewenstein, Scott & Wichansky, 2002). Below are examples of each of these processes driving data interpretation:

Top-down based on a cognitive theory. Longer fixations on a control element in the interface reflect a participant's difficulty interpreting the proper use of that control.

Top-down based on a design hypothesis. People will look at a banner advertisement on a webpage more frequently if we place it lower on the page. This is the approach selected for this collection of experiments.

Bottom-up. Participants are taking much longer than anticipated making selections on this screen. We wonder where they are looking.

3. EXPERIMENTAL DESIGN

The participants for this experiment are DP operators attending DP training in the facilities of the Rolls-Royce Training Center in Aalesund, Norway. The experiment will stretch over several months to gather enough participants and data. A collection of four tests will be carried out, where the initial test will be a pilot (test 0) followed by three tests (test 1-3) carried out using expert and novice DPOs. See table 2 and 3. Test 0, 1 and 2 there will be one single set of tasks carried out in one GUI environment for all participants. It will be a 2 x 1

design (table 1) where the focus will be to investigate foci areas of attention in general concerning DP operations and cognitive load.

Table 1: Test subject division test 1 and 2

Novice (10 participants)	Current DP GUI
Expert (10 participants)	Current DP GUI

For Test 3 there will be one single set of tasks carried out in two different GUI environments (current and new GUI). This will be a 4 x 2 (see table 2) counterbalanced within-subjects design.

Table 2: Test subject division test 3

Novice 1 (5 participants)	Current, New
Expert 1 (5 participants)	Current, New
Novice 2 (5 participants)	New, Current
Expert 2 (5 participants)	New, Current

The error rate will be monitored by utilising the video close-ups and log-file. The participants' age, gender, experience, if they are right or left handed and official title/education will be registered in the consent.

In order to estimate the cognitive workload effectively both objective and subjective methods will be utilized. The objective methods selected are heart rate monitoring and study of ocular behavior (eye-tracking). The most suitable methods are selected according to the criteria mapped by Cain (2007). The subjective method chosen is to ask the operator to evaluate the workload experienced in performing a task by using NASA TLX (NASA Task Load Index). The TLX assesses mental workload at the end of the scenario or experiment and breaks down mental workload into several components. In this case the form will be filled out after the operation has been carried out.

3.1 Aims, Variables and Hypotheses

We have two main aims for this investigation which are to investigate where on the displays the DPO focuses during operation and to investigate the division of attention between VDUs, equipment (input devices and relevant control panels) and scenery (visualization of aft deck, oil rig and surrounding area).

As the task difficulty (workload) increases performance usually decreases, hence response time and error rate increase. In addition control variability increases and fewer tasks are completed per unit time. We have a set of experimental parameters and hypotheses:

Table 3: Experimental variables

Ind. variables:	Dep. variables:
Graphical User Interface User- Experience	Workload Foci of attention

Hypotheses:

- (i) The DPO's attention will mainly be on the aft deck of the vessel during the operation.
- (ii) The DPO's cognitive load will increase during rig approach.
- (iii) The DPO's cognitive load will increase during drift on/off.
- (iv) The DPO's cognitive load will decrease when utilising new DP GUI design (Test 3)
- (v) The DPO will spend less time focusing on GUI and equipment when utilizing new DP GUI design. (Test 3)

3.2 Metrics

We have four types of metrics which we would like to measure; fixation, saccadic movements, eye blink rate and scanpath. The fifth option is to measure pupillary response. However, in this case we do not hold the equipment to carry out this type of measurements.

3.3 Data collection and analysis

When collecting and analyzing the data it is important to characterize the data in three ways:

- (i) General metric analysis across the DP operating setup
- (ii) Specific analysis of the metrics within the predefined area of interest (AOI)
- (iii) Specific analysis of the metrics within each task performed during the operation (carried out post-experiment during processing of data).

Statistical analysis using ANOVA of the data collected using the eye tracker and the NASA TLX will help to estimate the low and high workload operations.

3.4 Subjects

We will utilize in total 40 test subjects. The operators have been divided into expert and novices according to the criteria set by the Nautical Institute and the International Maritime Contractors Association (IMCA). During the training program the DPOs have to go through four phases where the last phase qualifies to get the DPO certificate. A novice has fulfilled the criteria of phase 1 to 3, while an expert has fulfilled the criteria of phase 1 to 4 and has received the certificate.

The apparatus used for the experiment activities will be a Rolls-Royce DP system onboard a realistic 360 degree bridge simulator. Tobii eye-tracking glasses will be utilized to record the ocular activity, while a heart rate monitor will be utilized to measure heart rate variability. Additional equipment is video cameras, notebook/logbook, NASA TLX and consent form. The experiment will last for a duration of maximum 1 hour (excl. training and familiarization).

3.5 Experimental Setup

The operator's cognitive response to various tasks given will be simulated in a controlled environment (simulator). We have little knowledge about the low and high workload levels in a DP operation, so it is important collect and document the workload variations across the operation. This can best be done by going onboard a real vessel and carry out this experiment, however this can propose challenges due to the safety critical nature of the environment. The experiment will be carried out in a simulated environment which is the closest one can get to real in this setting. The layout of the study will be as described below:

Step 1 – formal details and training

Before the experiment commences the DPO will be asked to wear the Tobii eye glasses and sign a consent document. Some basic training will be given, a briefing of what to do according to the experiment protocol and a chance to familiarize with the equipment. The location for this is divided between the instructor station and inside the simulator.

Step 2- experiment

The DPO enters the room of the offshore bridge simulator. The simulator has a realistic full bridge setup (fore and aft bridge) and on the aft bridge a DP operator station where the operator will be seated when the test commences.

- The DPO will start the experiment by taking command of the vessel and transfer control command from fore bridge (utilized in transit to a destination) to aft bridge (utilized for DP operations).
- The DPO will enter DP mode and start the DP operation from a distance of 75 – 100 meters from the oilrig.
- The operator will follow the legal procedures connected to this operation and will before going into DP mode ask the rig for permission to approach the installation and communicate with the rig and the rig's crane operator concerning which side of the platform to stabilize the vessel.
- The DPO will go through the DP checklist.

- When this is settled, the vessel can approach the rig.
- The vessel will be stabilized close to the rig and start loading/offloading pipes, containers and cradles of equipment from the aft deck.
- To slightly increase the DPO's cognitive load, the weather will pick up and a drift on/off situation is happening.
- In addition to the above a failure on the laser/cyscan system will occur. Where only a minor correction is needed by the DPO.
- The DPO must reposition the vessel on the other side of the rig, make sure all reference systems are OK and continue the loading/offloading.
- When finishing the loading/offloading the operator will depart from the platform to a safe distance after communicating with the rig.
- The operation ends when the DPO feels he/she is at a safe/comfortable distance from the rig and transfer control to manual and fore bridge.

Step 3 – evaluation

After the operator has finalized the experiment he/she will fill out a NASA TLX and be interviewed.

With the above scenario in mind the areas of interest (AOI) investigated will be VDUs (showing the DP systems graphical user interface), scenery (aft deck view) and input devices (placed in armrests of chair). These areas will be marked in the Tobii eye-tracking evaluation software and additional data will be collected.

4. PRELIMINARY RESULTS AND FUTURE WORK

In initial studies carried out in our DP vessel simulator, eye-tracking clearly showed differences in performance patterns from what we expected. Earlier results found during observations of DPOs operating a vessel in the North Sea (Anon, 2012), it was possible to estimate from the observations that the operator spent most of his/her time watching the aft deck. However, by using the eye-trackers we could actually see that most of the attention was still on the aft deck, but now we could see clear trends that the operator tended to spend more time than anticipated monitoring the information on the VDUs without interacting with the GUI interfaces. This can indicate that the DP's GUI has more effect on the operator's performance than previously believed.

The future work and results of this experiment will gain insight in the collaboration between the operator and the equipment on an aft bridge of an

offshore vessel. The desirable result after finalizing test 3 would be to improve the interface sufficiently so that the operator would spend more time watching the aft deck and less time monitoring and scanning the information displayed in the DP GUI. This would imply that only the most needed information is visible. This would also indicate that the DPOs cognitive load has been lowered as he/she can focus on the operation rather than the supporting equipment surrounding the operator.



Figure 2: Simulator setting with operator using eye-tracking glasses and IR markers for marking AOIs

4. REFERENCES

Bjørneseth, F.B., Dunlop, M.D., Hornecker, E. (2012). *Assessing the effectiveness of direct gesture interaction for a safety critical maritime application*. International journal of Human Computer Studies, Article in press.

Bray, D. (2003). *Dynamic Positioning-2nd Edition*, Oilfield Publications Ltd, London, UK

Bruneau, D., Sasse, M. A., & McCarthy, J.D. (2002). *The eyes never lie: The use of eye tracking data in HCI research*, CHI 2002, Workshop on physiological computing, Minnesota, USA, 20-25th of April, ACM Press

Cain, B. (2007). *A review of the mental workload literature*, Technical Report, Defence Research and Development Canada Toronto

Corwin, W. H., Sandry-Garza, D.L., Biferno, M. H., Boucek, G.P. (1989). *Assessment of crew mental workload measurement methods, techniques and procedure*, Guidelines for the Use of Workload Assessment Techniques in Aircraft Certification, Vol.1.

Di Nocera, F. Camilli, M., Terenzi, M. (2007). *A Random Glance at the Flight Deck: Pilots' Scanning Strategies and the Real-Time Assessment of Mental Workload*, Journal of Cognitive Engineering and Decision Making, Vol.1, 271 -285.

Ellis, K. (2009). *Eye Tracking Metrics for Workload Estimation in flight Deck Operation*, Thesis, University of Iowa

Goldberg, J. H. and Kotval, X. P. (1998), *Eye Movement-Based Evaluation of the Computer Interface*. In Kumar, S. K. (Ed.), *Advances in Occupational Ergonomics and Safety*, Amsterdam: IOS Press, 529-532.

Holzinger, A., Popova, E., Peischl, B., Ziefle, M. (2012) *On complexity reduction of user interfaces for safety-critical systems*. International Conference on Availability, Security, and Reliability, IFIP Cross Domain Conference, LNCS 7465(2012), 108-122

IMCA (2006), *Station Keeping Incidents reported for 2006*. International Marine Contractors Association.

Jacob, R. J.K., Karn, K.S. (2003) *Eye tracking in human-computer interaction and usability research: Ready to deliver the promises*. The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. Elsevier Science. Amsterdam. 573-605.

Kum, S., Furusho, M., Duru, O., Satir, T. (2007). *Mental Workload of the VTS Operators by Utilising Heart Rate*, International Journal on Marine Navigation and Safety of Sea Transportation.

Murai, K., Hayash, Y., Nagata, N., Inokuchi, S. (2004). *The mental workload of a ship's navigator using heart rate variability*, Interactive Technology and Smart Education, Vol. 1 Iss: 2, 127 – 133.

Poole, A., Ball, L.J. (2005). *Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects*. Chapter in C. Chaoui(Ed.): *Encyclopedia of HCI*. Pennsylvania: Idea Group, Inc.

Rehmann, J.T. (1982) *Cockpit display of Traffic information and the measurement of pilot workload: An Annotated bibliography*, US department of transportation.

Stickel, C., Ebner, M., Steinbach-Nordmann, S., Searle, G., Holzinger, A. (2009). *Emotion Detection: Application of the Valence Arousal Space for Rapid Biological Usability Testing to Enhance Universal Access*. Universal Access in Human-Computer Interaction, Part I: Addressing Diversity, HCI International, LNCS 5614 (615-624). Berlin, Heidelberg, New York: Springer.

Terenzi, M. (2008): *Assessment of a remote eye tracker system in a real time simulator*, European organization for the safety of air navigation

Tjallema, A., Van der Nat, C. (2007): *The Road to Eliminating Operator Related Dynamic Positioning Incidents*, Dynamic positioning conference. Vol. 2.

Willems, B., Allen, R.C., Stein, E.S. (1999) *Air Traffic Control Specialist Visual Scanning II: Task Load, Visual Noise, and Intrusions Into Controlled Airspace*, Technical Note (DOT/FAA/CT-TN99/23), FAA.