# Trial application of pupillometry for a maritime usability study in field conditions

Giovanni Pignoni [1] [0000−0003−0730−371X]
Odd Sveinung Hareide [2] [0000−0003−0140−3998]
Sashidharan Komandur [1] [0000−0002−9180−4201]
Frode Volden [1] [0000−0002−0000−690X]

[1]  *Norwegian University of Science and Technology, Institutt for design, Gjøvik, Norway*
    *postmottak@ntnu.no https://www.ntnu.edu/design*
[2]  *Norwegian Defence University College, Royal Norwegian Naval Academy, Bergen, Norway*
    *forsvaret@mil.no*

Abstract

Eye-tracking is a tool employed in usability testing. It is primarily intended as a means for tracking the visual attention patterns of an observer on a continuous basis. Eye-tracking can also capture certain physiological data, such as pupil dilation. Pupil diameter is a validated metric of cognitive workload, meaning the pupil dilates with increasing workload.

This research evaluates the fitness, in field conditions, of an eye-tracking based method for accurate measurement of cognitive workload. This implies evaluating the fitness of this tool in changing light conditions such as in coastal navigation. This methodology thus accounts for the effect of light on pupil dilation. This means we are able to account for the effect of only cognitive workload on the pupil dilation even in changing light conditions.

This method was applied as a part of an analysis of a navigational exercise involving the navigator and the navigator's assistant on board a training vessel of the Royal Norwegian Naval Academy. Pupillometry is used alongside egocentric video recordings and Geo-positioning systems (GPS) recordings to allow for multi-faceted evaluation of the activity. Subjective data was recorded as well to evaluate the quality of the eye-tracking data. Subjective data was recorded using NASA-TLX self-report of mental workload, self-report of mental workload (on three levels) using maps and an expert assessment of the mental workload was obtained for the navigational course.

The analysis concluded that pupillometry (through eye-trackers) can have a substantial role in the evaluation of field operations and provide a good and objective estimate of the perceived workload. The eye-tracking technology has substantial limitations, for example sometimes strong infra-red sources of light can impede data collection as such with an eye-tracker, meaning the analysis is labour intensive as it relies on the ability of the operator to filter out low quality data and retain the rest.

1   Introduction

Human mental workload assessment is considered important for the design and operation of safety-critical human-machine systems[2][8]. The list of valuable tools that enable the monitoring of an operators mental workload has grown and is more accessible than ever[8][7][4]. Electrocardiography, heart rate variability, brain activity (fMRI, EEG), eye-tracking (pupillometry, gaze position, saccades velocity) have proven their value in a laboratory. Still, the actual application of these technologies in a field study is limited, either by the necessity of control over confounding variables or limitations in the practical application[8][7][4]. In this regard, modern eye tracking has the advantage of satisfying most of the requirements of field applications: portable and quite unobtrusive, affordable enough to enable tracking of multiple subjects and provides rich data that can be analysed to extract multiple parameters, not only limited to vision. The influence

of ambient illumination has historically limited the application of eye-tracking as it can produce unwanted variability in the pupil size and have a disruptive effect on the tracking technology (principally caused by infrared light) used by many of the commercially available eye-trackers[7][9][13][11].

## 1.1 Pupillometry for a field study

Maritime Navigation can be conducted in a team or by a single person. The Royal Norwegian Navy has presented a methodology, known as the phases of navigation [12], which provides the navigator and the navigation team with a common decision-making strategy. The methodology fits on any type of vessels, but the process is more demanding in confined water and with higher speed. When the speed increases or the spatial environment is more complex, the cognitive workload will increase for the navigator and the navigation team. The importance of not going into cognitive overload is imperative for the navigator and the navigation team, and several maritime accident investigations state loss of Situation Awareness (SA) as one of the factors in their reports [30]. If the navigator and the navigation team is aware of an increase in their cognitive workload, they could be able to implement measures to compensate in order to facilitate safe navigation. This could be to reduce the speed/stop the vessel, or to conduct a better distribution of the tasks at hand in the navigation team. Safe navigation in littoral waters can be a demanding and safety-critical activity. Consequently, the role of the navigator and other critical crew members, as well as the relative Human-Machine Interaction (HMI), has become of ever increasing interest during the design and evaluation of the technology present on ship bridges. Being able to consider and evaluate the human element in the design of the onboard systems is, therefore, necessary to reduce the potential human error and increase safety[11]. As part of the thesis research project [28] and as presented at the HCI 2019 conference [27], an experimental method has been developed to counteract the effect of light on the pupil size and enable pupillometry in a field study. This paper describes the experimental application of the method in a maritime environment.

## 2 Background

### 2.1 Cognitive workload

The cognitive workload is the result of the interaction between a user (his psychological state, experience and personality), conceptualised as the "available resources" and the resources necessary to complete a task [5][14]. Time, accuracy, error rate and fatigue can be considered as a secondary product of workload (e.g. for a given task, lower training results in a higher workload and possibly longer execution time). As cognitive workload is a human-centred metric, user related parameters such as insufficient training, anxiety, fear, fatigue, visual or auditory overload can all have a significant effect on the subjective workload even if the task is kept stable. Common cognitive workload measurement can be categorised as the three different approaches they are based upon [6][29][14]:

– Subjective-empirical measurements of perceived workload, rated by the subject, are usually administered as "paper-and-pencil" questionnaires and multidimen-

sional ratings, (e.g. the NASA-TLX[14], which was selected in this study for its widespread use and relevance). They offer high face validity and high user acceptance as well as being relatively easy to administer. Still, they are limited by relying on memory and self-perception of the participant[31].

– Performance based measurements of workload use a controlled task (primary or secondary) and a performance metric to evaluate the workload (error rate, response time or similar). A dual-task performance measure is based on the assumption that the secondary task is executed in a realm of limited resources, used by the primary task in a variable amount and is consequently affected by the variation of the primary's load on the user[19]. Performance-based measurements offer an objective approach to workload measure but do so by mixing the definition of workload to the definition of performance. Contextual variations and the presence of overload or under-load are not well captured by performance data. Moreover, the use of a secondary task can create by itself overload or result in low user acceptance.

– Physiological indices of cognitive state can also be used to measure workload. Psychophysiological techniques are based on a documented relationship between a behavioural phenomenon and a quantifiable activity of the central nervous system. Mental workload has been related to heart rate and heart rate variability, respiratory rate, galvanic skin response, brain activity (fMRI, EEG), as well as eye activity[3][6]. Physiological indices, depending on the setup, can be non-intrusive and provide rich, objective data over time. Difficulty in the control of confounding variables and setup are the main limiting factors to a generalised adoption of physiological indices.

NASA Task Load Index The NASA Task Load Index is a multidimensional rating procedure, developed at the NASA-Ames Research Center, that provides an overall workload score based on a weighted average of ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort and Frustration[16]. The NASA TLX allows data to be obtained in an operational setting, in between tasks or retrospectively (videotaped/regenerated activities)[16].

Pupil size and cognitive workload The pupillary system can be described as a "dually innervated organ"[3], and the pupil size as the result of the opposite action of both the parasympathetically innervated constricting circular muscles and sympathetically innervated radial muscles. The resulting pupil size is predominantly determined by the circular muscles activity in response to the light reflex. This can be interpreted as the pupil size baseline at any given moment. The sympathetic, dilating, activity is not only responsible for slowly adapting the pupil size to a luminance decrement, but it is also affected, in the form of short peaks, by the user's cognitive state, in particular, mental activity, and can thus be used as a psychophysiological parameter[28].

The pupillary response has been observed and linked to workload as far back as 1964[15], with the controlled induction of changes in the pupil size trough a variety of mathematical and memory-intensive tasks[20] as well as writing, listening and speech-based tasks[21]. The pupillary response has been used to estimate the effect of more complex and variable tasks, such as driving (e.g.[26] shows the connection between driving performances and pupil size).

Unified formula for light-adapted pupil size. Watson AB. and Yellott JI. at the NASA Ames Research Center and the University of California[32], reviewed seven historical psychophysical functions of the relation between target luminance (cd/m2) and expected pupil diameter and published a newly developed, unified formula. High-temporal-resolution tracking of cognitive workload compensated for changes in ambient light has been attempted with various levels of success and implementation[33][23]. However, no system is freely available and sufficiently documented for reuse. Building upon the knowledge provided by multiple proofs of concepts[25], Pignoni[28] integrated the Unified formula for light-adapted pupil size[32] with eye-tracking data to estimate the expected pupil diameter for a given visual stimulus and differentiate the changes in pupil diameter related to the cognitive workload from the pupillary light response.

## 3  Methods

This study is enabled by the development of an apparatus able to record the participant's pupil dilation and visual stimulus based on the Pupil Labs EyeTracking Glasses (ETGs), an external luminance sensor as well as the software needed to analyse the raw data, estimate the effect of light on the pupil size and extract a measure of cognitive workload. The data collection was conducted in a field study onboard the Royal Norwegian Navy training vessel, Kvarven (Figure 1). This was done in order to evaluate the use of the apparatus in an actual operational environment (ease of application, flexibility and user acceptance of the apparatus) as well as the quality of the data it could provide. The scope of the field test is not to validate the instrument, partial validation of the instrument has been carried out in laboratory conditions before the field experiment[28]. Other studies (mainly collecting visual attention data from a glasses based eye trackers [12][10][13]) have been conducted on high-speed crafts (both in the form of a simulator and an actual high-speed craft).



Fig.1. The vessel, Kvarven, used for the test session.

Research questions:
– Is the described method reliable in field condition (where luminance variate in an unpredictable manner)?

- Is the described method sufficiently sensitive to record small variations of cognitive workload, as those to be expected in a typical navigation task?
- Does the result correlate with subjective data and can it be used in conjunction with such data?

### 3.1 Eye-Tracking Glasses and pupillometry

The eye-tracking technique and newly developed mean of removing the effect of luminance is extensively described in [28]. The hardware is standardised and commercially available:

- The Pupil Pro eye-tracking glasses[22], equipped with an egocentric video camera ad video tracking of the right eye.
- A single TSL2591[1] external Light sensor mounted on the eye-tracker, alongside the egocentric video camera.
- An Arduino[18] compatible board used to read the TSL2591 and log the data[28].

The software provided by Pupil Labs is used to perform the eye-tracking recording independently from the luminance logging, it records a video feed from the user's point of view as well as an estimated gaze position on the video and the pupil size. The recorded data, egocentric video recording, gaze position and luminance from the external sensor, is analysed to estimate the variable effect of light on the pupil size, this value is expressed as a baseline pupil diameter (the average dilatation that is expected for a given visual stimulus), sampled for the entire recording. The pupil diameter baseline is used to generate an estimate of cognitive workload, expressed as the residual dilatation of the pupil (the difference between the expected pupil size and the actual pupil size). The data pipeline is described in the implementation chapter of [28].

The use of two Pupil Pro eye-tracking Glasses[22] enabled tracking of two participants for each session. The eye-tracking glasses were configured to work at a sampling frequency of 120 Hz and a resolution of 400x400px. The tsl2591 Lux sensor[1] was used to record the average luminance in front of the participants with a sample rate of 10Hz. The data was saved directly onto two different laptops. Gps coordinates were also recorded through a smartphone. The subjective data relied on pen and paper (NASA TLX [14] and self-report workload maps), the original material is visible in [28].

### 3.2 Royal Norwegian Navy Training Vessels

The training vessel is fifteen metres long, and are capable of speeds up to 40 knots, the vessel speed for the test was set to 25 knots. The integrated Navigation System (INS) is the same as onboard larger vessels (e.g. a Frigate or a Plattform Support Vessel), and the vessels are used for navigation training to prepare the cadets at the Royal Norwegian Naval Academy for their onboard service.

The INS is from a major Original Equipment Manufacturer (OEM), and consists of the traditional setup with Electronic Chart Display and Information System (ECDIS), Radar, and Conning. These three applications can be presented on either of the two Multi-Function Displays (MFDs) in front of the navigation team. The navigation team onboard the Training Vessel consists of three persons:

– The Navigator: Is in charge of safe navigation and is the leader of the navigation team.
– The ECDIS assistant: Provides the navigator with navigational information, which is aligned with Standard Operatin Procedure (SOPs). Conducts navigational tasks for the navigator, such as position fixes, which is aligned with SOPs.
– The helmsman: Is responsible for the wheel and throttle of the vessel. Sets speed and steers course as ordered by the navigator.

### 3.3 Scenario

The data collection was conducted on the 24th of April 2019 in littoral waters west of Bergen, as shown in Figure 2. The participants were recruited from the Royal Norwegian Naval Academy and were graduating students in the operational branch. This implies they have about 300 hours on board the training vessels prior



Fig.2. The route, starting and ending under the Sotra Bridge near the RNoNA harbour, running clockwise around the Bjoryhavn island.

to the data collection. The test session started with a long pretest session to test and adapt the equipment before embarking and briefing the participants inside the vessel. They were familiarised to the route (Figure 2) the stop points and the test procedure. They were given time to read through the instructions regarding the NASA TLX self-report and time to read and sign the consent form.

Five test sessions were performed, requiring three participants for each session: the navigator, sitting in the centre, the ECDIS assistant on the left and the helmsman on the right. Eye-tracking data was recorded for the ECDIS assistant and the navigator only. The field study consisted of five participants, which each conducted one scenario at different positions on board. The participants have been coded with a letter and a number (Nx, Ax). N for the navigator, A for the ECDIS assistant, followed by the sequential number of test rounds (e.g. N2 and A2 were recorded at the same time). Due to technical challenges, data from subjects N1 and A1 had to be discarded and is not part of the data analysis. The test session was divided into two parts, each consisting of on average eighteen minutes of navigation. In between the two halves, the participants were asked to compile the NASA TLX pairwise score for the task [14] and the rating scale. At the end of the course, the participants were asked to compile a second TLX rating scale on a pen and paper form, and the visual self-report by drawing on a map of the route with three markers. Green was

defined as "below average workload", yellow as "average workload" and red/purple as "above-average workload".

## 3.4 Expert Workload Map

To evaluate the use of pupillometry in field conditions, it was deemed necessary to produce comparable measures of workload over the course. The expert map (Figure 3) was created in a workshop with six Subject Matter Experts (SMEs) from the Royal Norwegian Navy. The participants experience range from 10 to 25 years of navigation, primarily with High-Speed Craft (HSC) navigation on vessels up to 50 metres. Each of the SMEs filled out one form to rate the expected cognitive workload throughout the course with regards to three different colour; green (below average workload), yellow (average workload) and red (above-average workload). After each had completed their subjective evaluation, the group discussed discrepancies between them. The discrepancies between the SMEs were presented and discussed, and the group agreed on one overall expert evaluation of the course, which is shown in Figure 3. The experience level will differ between the SMEs and the participants. The expert workload map is a generic workload assessment based on any experience level for the navigation team. The expert team emphasised the importance of reading the map as changes to workload in the different phases of navigation during the given passage, and this is expected to be valid for any experience level. Thus the expert map would apply to the participants in this study.
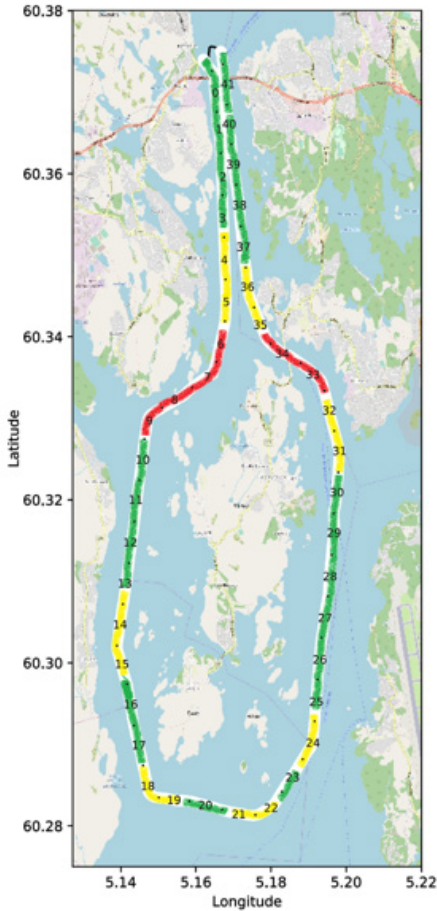


Fig.3. Expert evaluation of the course representing the expected workload of the crew for normal sailing conditions (e.g. only considering the course itself and not other variables such as traffic.). The map express the expected workload in three steps (yellow = average, green = below average and red = above average).

## 3.5 Data Recording and Analysis

The workload data recorded during the five-session includes multiple data sources. This includes pupillometry and eye-tracking data, combined with luminance measurements and video, to produce a psychophysical measure of the relative changes of cognitive workload over time. Multiple subjec-

tive data sources were used: NASA TLX self-reported workload measure, (sampled halfway and at the end of each course), digitised self-report maps, representing the perceived workload over the course and a map of the expected workload (Figure 3) compiled by SMEs from the Royal Norwegian Navy. The NASA TLX final score for each session is a standardised index of cognitive workload and allows to compare the perceived workload of the two halves of each run.

Self-reported workload was scored by the participants in segments. These segments were self-selected at the end of the navigational exercise. This is because the map represented the course as a continuous path. This data was segmented into forty-one 0.27 nautical miles (500 meters) sections and is scored accordingly to colour (1-green, 2-yellow and 3-red), see Figure 7 (bottom) for an example.

The pupillometry data is obtained processing the eye-tracking recording and is paired with GPS tracking. Numerically it is the measurable Δ Pupil Diameter as a consequence of the changes in workload. The Δ Pupil Diameter is used to generate either an objective measure of average workload (mean Δ Pupil Diameter) for each of the forty-one sections and a graphical visualisation of the measured workload over the course for qualitative analysis of workload, see Figure 7 (top) for an example.

A simple linear regression analysis (Figure 4) is used to analyse the relationship between the Δ Pupil Diameter (dependant variable) and the subjective measures (independent variables): Self Reported Workload (Figure 5) and Expert Reference Workload (Figure 6). The tools used to perform the statistical analysis are SPSS [17] and Minitab Express [24]. The variables satisfy the requirements to perform the statistical analysis.

## 4 Results

### Simple Regression: △Pupil Diameter vs Self Reported Workload

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 1 | 0,052669 | 0,0526693 | 12,35 | 0,0011 |
| Error | 40 | 0,170553 | 0,0042638 | | |
| Total | 41 | 0,223222 | | | |

Model Summary

| S | R-sq | R-sq(adj) |
|---|---|---|
| 0,0652980 | 23,59% | 21,68% |

### Simple Regression: △ Pupil Diameter vs Reference Workload

Analysis of Variance

| Source | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Regression | 1 | 0,031557 | 0,0315565 | 6,59 | 0,0141 |
| Error | 40 | 0,191666 | 0,0047916 | | |
| Total | 41 | 0,223222 | | | |

Model Summary

| S | R-sq | R-sq(adj) |
|---|---|---|
| 0,0692217 | 14,14% | 11,99% |

Fig.4. Report of the regression analysis comparing the pupillometry based measure of workload (Δ Pupil diameter) and the subjective measures: Self reported workload (left) and Reference workload (right).

The three measures of workload have been compared to attempt cross-validation between them. The regression analysis (Figure 4) between the mean Δ Pupil Diameter (mΔPd) and Expert Reference workload (F-val=6.59 and P-val=0.0141) as well as mean Δ Pupil Diameter and Self Reported Workload (F-val=12.35 and P-val=0.0011) support the validation of this system as a measure of perceived workload.
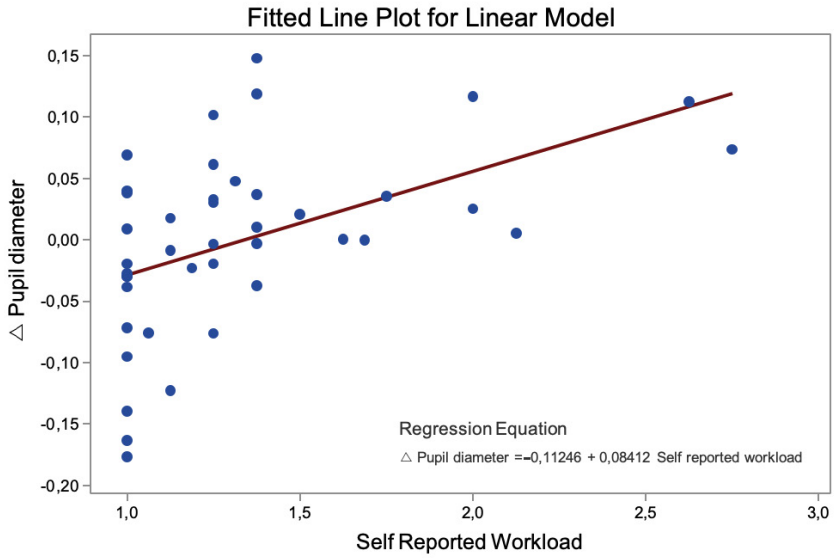
*Fig.5. Plot of the linear regression analysis comparing the pupillometry based measure of workload (Δ Pupil Diameter) to the Self Reported Workload.*
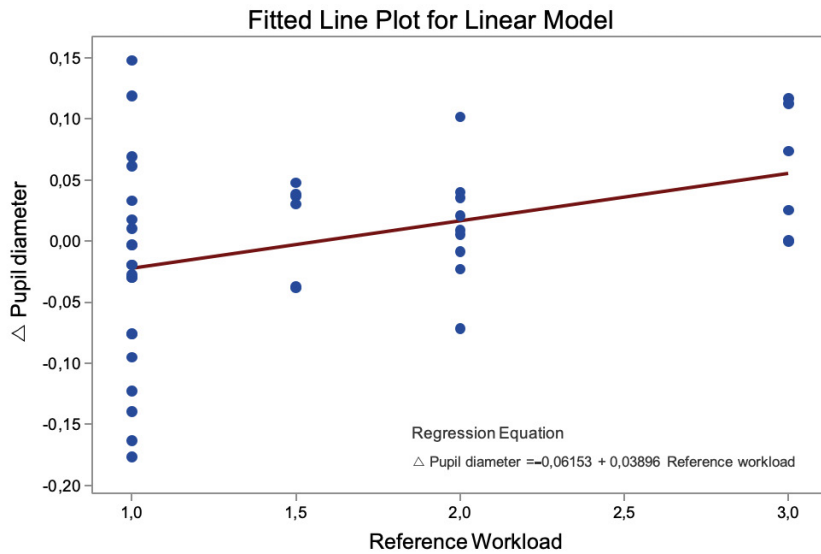


*Fig.6. Plot of the linear regression analysis comparing the pupillometry based measure of workload (Δ Pupil Diameter) to the Expert Reference Workload.*
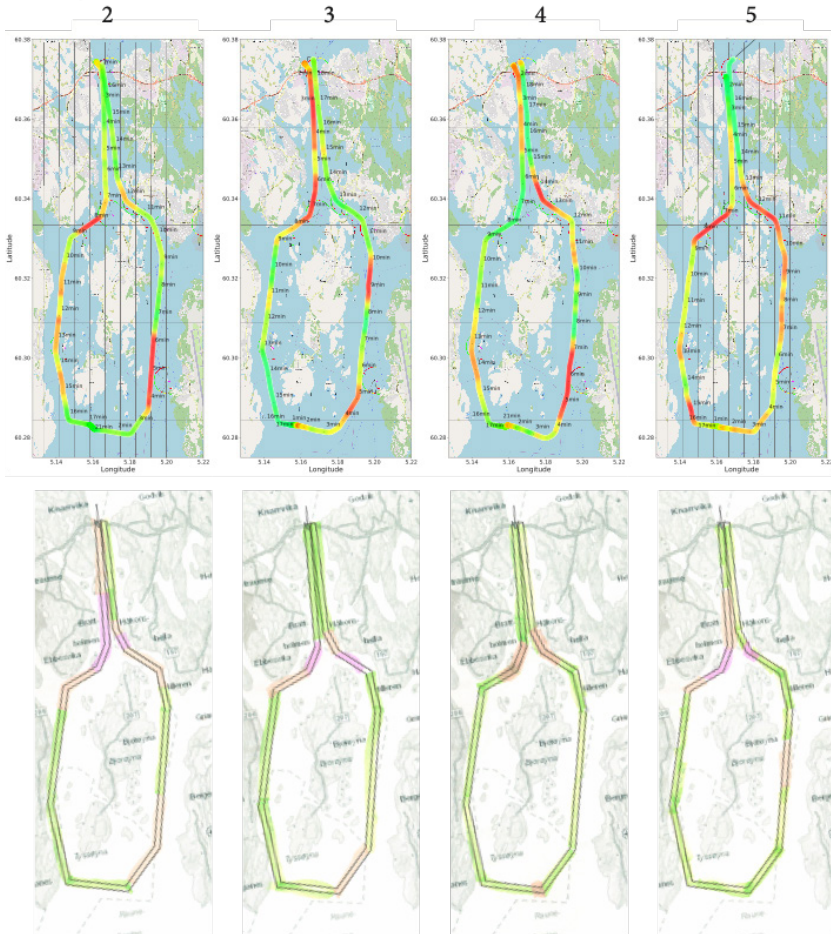
*Fig.7. Output data for qualitative analysis of workload: self report maps (bottom) can be compared to the output of the eye-tracking (top). The self report maps express workload in three steps (yellow = average, green = below average and red = above average). The eye-tracking maps represent workload (mean Δ Pupil Diameter) from -1.5 SD (standard deviations) (green) trough average (yellow) to +1.5 SD (red). The maps for all the participants is available in [28].*

The reliability of this measure is still limited by the presence of noise and the performance of the eye tracker in general. Both the expert evaluation and the self-report map are linearly related to the mean Δ Pupil Diameter. A higher correlation was found between mΔPd and Self Reported Workload than mΔPd and Reference Workload, which can be explained by the abstraction of the expert evaluation, based solely on the course topography and not on the other (unpredictable) confounding variables such as traffic that are instead going to affect the Self-reported workload.

The NASA-TLX scores have a limited temporal resolution can be used as an absolute measure of cognitive workload for the different phases of the experiment. The TLX data shows how the task did not include extreme conditions of overload or under-load. Therefore, the average yellow/orange (in figure 7) can be considered as a comfortable working condition with fluctuations (red/green) representing a normal attention cycle.

The qualitative analysis of the eye-tracking data, (reviewing the video recordings to identify the source of workload), highlights how the precision of the workload data is highly dependant on the quality of the eye-tracking data.

Light reflection in the pupil and incorrect framing can unpredictably affect the tracking either and create false positives. This and other practical limitations of the eye tracker have to be accounted for and expected in field conditions and require manual analysis of the data to be recognised. Good quality eye-tracking data shows a connection between the actions of the participants, the surroundings and the level of workload.

The self-reported workload has proven to be a useful resource to support eye-tracking data and verify the quality of the recordings as well as the Expert Workload maps. The NASA TLX has value, instead, as an absolute measure of workload and helps to contextualise the pupillometry data.

## 5   Discussion

The described method is sufficiently sensitive to record small variations of cognitive workload. Patterns of such measures are visible in the data and correlate with the subjective measurements of workload over the course. Partial validation of the instrument has been carried out in laboratory conditions before the field experiment[28]. Nevertheless, the field experiment indicates the ability of the tool to track small changes in workload over time. The use of the eye-tracker in high-speed craft maritime conditions was reliable. This means it was easy to instrument and set up the eye-tracker for collecting pupil dilation measures. This holds even after the algorithm was corrected to account for changing light conditions. There was only one instance where there was a technical failure which resulted in the loss of data from one navigator and one assistant. The nature of the technical failure is that the video data that was collected from the assistant was corrupted and there was no data collected from the navigator mainly because the apparatus stopped tracking the eye. These errors could have happened even without the modified algorithm. Thus with this method, data can be reliably collected and saved. The use in conjunction with subjective data not only enabled cross-validation for the measurements but also produces useful outputs for either research or training purpose. Examples of such outputs, such as the workload maps, encourages the use of a variety of data sources to produce meaningful data visualisations.

Limitations: The quality of the cognitive workload measurement is directly dependant upon the quality of the eye-tracking data. Consequently inheriting some of the limitations relative to the use of eye-tracking in a filed condition. Incident sunlight has a particularly disruptive effect on the eye-tracking, as reflections of windows and light sources from infrared light becomes visible in the recording and can partially or entirely cover the iris and pupil. When this happens, it is almost

impossible to get a reliable recognition of the pupil size and orientation. For this reason, recordings at sundowns and sunset should be avoided. The size and shape of the participant's eye can have a significant effect on the quality of the recording. It was noted how variations in the eye appearance such as a pronounced "Epicanthic Fold" or a lower contrast between the iris and pupil (dependant on how the iris reflects or absorbs infrared light) can reduce the quality of the eye recognition as the computer vision algorithm struggles to identify the eye's features in the video. Highlighting when the tracking of the pupil has failed during a recording is currently a manual process as the metrics of confidence included in the eye-tracking software is not reliable for what concerns the tracking of pupil size. It was noted that as the data analysis is currently not immediately available, which implies that the results are not available shortly after the end of the recording session, the potential application of the apparatus in a training setting is limited. Nevertheless, optimisation of the algorithm is possible, and it is expected this limitation will be resolved before the next iteration of the study.

Confounding variables: Information regarding participants arousal level was not collected as well as substance consumption (such as caffeine). Environmental factors (e.g. temperature, vibration, weather conditions, etc.) which may also be contributing to the noise in the data as of now. The effect on the workload that could be induced by the different waters the vessel had to navigate during the test was not characterised. More challenging waters, with narrower passages and an increasing amount of turns or spatially complex due to inlets, would potentially create an increased cognitive workload level. Traffic and emerging situations (e.g. if a vessel is sighted and a give way situation occurs) were not recorded, the effect of such occurrences on workload is therefore not tracked. To conclude, the self-reported workload maps are subjected to recall bias as they were compiled by the participants only at the end of each run; each run has a duration of around forty minutes.

The discussion in the SME group highlighted a series of observations on the workload distribution and variability of the course that are of particular interest for the interpretation of the workload analysis. They noted how the cognitive workload should increase when approaching a turn and when the vessel is approaching narrow waters. The cognitive workload is instead expected to decrease when the turn is finished and when the complexity of the waterways decreases (more open waters). The cognitive workload is thus expected to vary with the water confinement. Blue waters should, therefore, be characterised by a general lower cognitive workload than littoral waters. Furthermore, it was mentioned how, when a particular situation occurs (e.g. a give way situation with another vessel), the cognitive workload should increase. This is due to the cognitive process of updating the navigator's situational awareness (SA) with the new environmental information which will affect the future passage of the vessel. The navigator has to notice, comprehend and project the future state of the vessel in order to determine a give way situation. In conjunction with this, the navigator has to conduct an action to change the course or speed of the vessel in order to comply with the collision regulations. An increase or decrease in traffic will influence the cognitive workload of the navigator. This is shown in the south-eastern part of the passage with navigator 2 and 4. When they turn the vessel northbound, they need to update their SA with the introduction of several ships in the environment. Even though the vessels are not in direct conflict

of the passage, the cognitive workload will rise as the navigators are updating their SA. The weather condition will influence the cognitive workload. With demanding weather conditions (darkness, severe wind, rain, fog, snow), there will be an increase in the cognitive workload.

## 6    Conclusion

The developed system is promising as a tool for use in maritime field conditions, but the general limitations of the eye-tracking technology still apply: the set up of the eye tracker and designing the experiment are time-consuming, but critical to producing good quality workload data. It provides a higher temporal resolution recording compared to pen and paper methods but cannot measure workload on an absolute scale (only records changes in workload).

An online and objective measure of workload could find relevant application in a variety of fields. During design/testing/validation the data produced by this tool can be used as design input, alongside other measures of workload (subjective or psychophysical). In the context of training and human performance assessment, this tool can be a valuable feedback tool both for instructors and trainees.

Self-reported Workload and Expert Reference Workload both correlate to the result of the pupillometry. The use of three different methods to measure workload helps to support the eye-tracking data and verify the quality of the recordings. The NASA TLX scores can still be used in parallel as a tool to evaluate the cognitive workload as an absolute measure over larger portions of the task.

### 6.1  Future work

To resolve most of the limitations related to the outdoor field conditions and control the variability in the course (traffic and weather) a new session of tests should be scheduled using a more controlled condition, and the most obvious solution would be to use a navigation simulator. Nevertheless, it was important to test the fitness of this technology in a field setting first. As a next step, we can expand the applicability of our method. Thus we would like to understand the variability in the measures of workload for a wide variety of coastal navigation conditions which is best recreated in a simulator. This assumes a recreation of a wide variety of changing light conditions.

## References

1.  AMS: TSL25911 Datasheet. AMS (Apr 2013), https://cdn-shop.adafruit.com/datasheets/TSL25911_Datasheet_EN_v1.pdf
2.  Bjørneseth, F.B., Renganayagalu, S.K., Komandur, S., Dunlop, M., Hornecker, E.: Towards an experimental design framework for evaluation of dynamic workload and situational awareness in safety critical maritime settings. https://doi.org/10.14236/ewic/HCI2012.43
3.  Cacioppo, J.T., Tassinary, L.G., Berntson, G. (eds.): Handbook of Psychophysiology. Cambridge University Press, Cambridge, 3 edn. (2007). https://doi.org/10.1017/CBO9780511546396, http://ebooks.cambridge.org/ref/id/CBO9780511546396

4. Cain, B.: A review of the mental workload literature
5. Council, U.N.R. (ed.): Tactical display for soldiers: human factors considerations. National Academy Press, Washington, D.C (1997)
6. Di Stasi, L., Marchitto, M., Antol, A., Caas, J.: Saccadic peak velocity as an alternative index of operator attention: A short review. Revue Europenne de Psychologie Applique/European Review of Applied Psychology **63**(6), 335–343 (Nov 2013). https://doi.org/10.1016/j.erap.2013.09.001, https://linkinghub. elsevier.com/ retrieve/pii/S1162908813000741
7. Eric Farmer, Adam Brownson, QinetiQ: Review of workload measurement, analysis and interpretation methods. https://doi.org/10.1163/1570-6664_iyb_ SIM_org_39214 https://referenceworks.brillonline.com/entries/ international-year-book-and-statesmens-who-s-who/*-SIM_org_39214, type: dataset
8. Gao, Q., Wang, Y., Song, F., Li, Z., Dong, X.: Mental workload measurement for emergency operating procedures in digital nuclear power plants **56**(7), 1070–1085. https://doi.org/10.1080/00140139.2013.790483, https://doi.org/10. 1080/00140139.2013.790483
9. Groen, M., Noyes, J.: Using eye tracking to evaluate usability of user interfaces: Is it warranted? IFAC Proceedings Volumes **43**(13), 489–493 (2010). https:// doi.org/10.3182/20100831-4-FR-2021.00086, https://linkinghub. elsevier.com/ retrieve/pii/S1474667015325817
10. Hareide, O.S., Mjelde, F.V., Glomsvoll, O., Ostnes, R.: Developing a HighSpeed Craft Route Monitor Window. vol. 10285, pp. 461–473. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-58625-033, http://link.springer.com/10.1007/978-3-319-58625-0_33
11. Hareide, O.S., Ostnes, R.: Maritime usability study by analysing eye tracking data. Journal of Navigation **70**(5), 927–943 (2017). https://doi.org/10.1017/ S0373463317000182, https://www.cambridge.org/ core/product/identifier/ S0373463317000182/type/journal_article
12. Hareide, O.S., Ostnes, R.: Scan pattern for the maritime navigator. TransNav **11**(1), 39–47 (2017). https://doi.org/10.12716/1001.11.01.03, http:// www. transnav.eu/Article__Hareide,41,696.html
13. Hareide, O.S., Ostnes, R.: Validation of a Maritime Usability Study with Eye Tracking Data. In: Schmorrow, D.D., Fidopiastis, C.M. (eds.) Augmented Cognition: Users and Contexts, vol. 10916, pp. 273–292. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319- 91467-122, http://link.springer.com/10.1007/978-3-319-91467-1_22
14. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Advances in Psychology, vol. 52, pp. 139–183. Elsevier (1988). https://doi.org/10.1016/S0166-4115(08) 62386-9, http://linkinghub.elsevier.com/retrieve/pii/S0166411508623869
15. Hess, E.H., Polt, J.M.: Pupil Size in Relation to Mental Activity during Simple Problem-Solving. Science **143**(3611), 1190–1192 (Mar 1964). https://doi.org/ 10.1126/science.143.3611.1190, http://www.sciencemag.org/ cgi/doi/10.1126/ science.143.3611.1190
16. Human Performance Research Group, N.A.R.C.: TASK LOAD INDEX NASATLX Paper and Pencil Pakage, https://ntrs.nasa.gov/archive/nasa/casi. ntrs.nasa.gov/20000021488.pdf

17. IBM, C.: IBM SPSS Statistics (2017)
18. Industries, A.: Adafruit Industries, Unique & fun DIY electronics and kits (2019), https://www.adafruit.com
19. ISO: Road vehicles Transport information and control systems Detection-Response Task (DRT) for assessing attentional effects of cognitive load in driving. ISO 17488:2016. (2016)
20. Kahneman, D., Beatty, J.: Pupil Diameter and Load on Memory. Science **154**(3756), 1583–1585 (Dec 1966). https://doi.org/10.1126/science.154.3756.1583, http://www.sciencemag.org/cgi/doi/10.1126/science.154.3756.1583
21. Kahneman, D., Beatty, J., Pollack, I.: Perceptual Deficit during a Mental Task. Science **157**(3785), 218–219 (Jul 1967). https://doi.org/10.1126/science.157.3785.218, http://www.sciencemag.org/cgi/doi/10.1126/science.157.3785.218
22. Kassner, M., Patera, W., Bulling, A.: Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction. In: Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. pp. 1151–1160. UbiComp '14 Adjunct, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2638728.2641695, http://doi.acm.org/10.1145/2638728.2641695
23. Marshall, S.P.: Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity (Mar 1999), https://patents.google.com/patent/US6090051A/en
24. Minitab, L.: Minitab statistical process improvement tools (2019)
25. Palinko, O., Kun, A.: Exploring the Influence of Light and Cognitive Load on Pupil Diameter in Driving Simulator Studies. In: Proceedings of the 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design : driving assessment 2011. pp. 329–336. University of Iowa, Olympic Valley-Lake Tahoe, California, USA> (2011). https://doi.org/10.17077/drivingassessment.1416, http://ir.uiowa.edu/drivingassessment/2011/papers/48
26. Palinko, O., Kun, A.L., Shyrokov, A., Heeman, P.: Estimating cognitive load using remote eye tracking in a driving simulator. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10. p. 141. ACM Press, Austin, Texas (2010). https://doi.org/10.1145/1743666.1743701, http://portal. acm.org/citation.cfm?doid=1743666.1743701
27. Pignoni, G., Komandur, S.: Development of a Quantitative Evaluation Tool of Cognitive Workload in Field Studies Through Eye Tracking. In: Harris, D.(ed.) Engineering Psychology and Cognitive Ergonomics, vol. 11571, pp. 106–122. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-22507-09, http://link.springer.com/10.1007/978-3-030-22507-0_9
28. Pignoni,G.: Development of a quantitative evaluation tool of cognitive workload in field studies through eye tracking, http://hdl.handle net/11250/2617732
29. Rusnock, C.F., Borghetti, B.J.: Workload profiles: A continuous measure of mental workload. International Journal of Industrial Ergonomics **63**, 49–64 (Jan 2018). https://doi.org/10.1016/j.ergon.2016.09.003, https://linkinghub.elsevier.com/retrieve/pii/S0169814116301287

30. Sandhland, H., Oltedal, H., Eid, J.: Situation awareness in bridge operations a study of collisions between attendant vessels and offshore facilities in the north sea. Safety Science **79**, 277–285 (2015). https://doi.org/10.1016/j.ssci.2015.06.021, https://linkinghub.elsevier.com/retrieve/pii/S0925753515001630

31. Tsang, P.S., Velazquez, V.L.: Diagnosticity and multidimensional subjective workload ratings. Ergonomics **39**(3), 358–381 (Mar 1996). https://doi.org/10.1080/00140139608964470, http://www.tandfonline.com/ doi/abs/10.1080/00140139608964470

32. Watson, A.B., Yellott, J.I.: A unified formula for light-adapted pupil size. Journal of Vision **12**(10), 12–12 (Sep 2012). https://doi.org/10.1167/12.10.12, http://www. journalofvision.org/lookup/doi/10.1167/12.10.12

33. Wierda, S.M., van Rijn, H., Taatgen, N.A., Martens, S.: Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. Proceedings of the National Academy of Sciences **109**(22), 8456–8460 (May 2012). https://doi.org/10.1073/pnas.1201858109, http://www.pnas.org/ cgi/doi/10.1073/pnas.1201858109