Development of a quantitative evaluation tool of cognitive workload in field studies through eye tracking

Giovanni Pignoni[1] [0000-0003-0730-371X] and Sashidharan Komandur[1]

Department of Design, Norwegian University of Science and Technology, NTNU in Gjøvik, NO-2802, Norway postmottak@ntnu.no
https://www.ntnu.edu/design

Abstract. Eye tracking is mainly employed as mean of tracking visual attention of an observer/operator. Still, eye tracking is also capable of recording a wider variety of data such as traces of mental workload. Pupil diameter have been validated as such measure. Most of the studies that have validated this are in laboratory conditions, where the perceived luminance (measured in candela per square meter) can be controlled. Luminance affects the pupil diameter as well; this means if the pupil diameter varies for an operator/observer in field conditions it cannot be accurately determined if the change in the pupil diameter is due to mental workload alone. Although there are some studies, which have attempted to simultaneously account for the contribution of the change in pupil diameter due to luminance and mental workload, not many have attempted to account for this in field conditions for safety-critical systems such as a helicopter or a maritime ship bridge. In this study as a first step, we define a method to measure luminance while tracking the gaze point. We will record eye-tracking data simultaneously recording the video feed of the field of view of the operator/observer. We will use the video feed to estimate the luminous flux from the point of view of the subject. We will be collecting this data from a helicopter pilot and his co-pilot during an actual operation (e.g. transportation of personnel and carrying a payload for an electrical power provider company in Norway or Sweden). We will also be collecting data from a navigator and his first officer in a high-speed marine craft of the Norwegian navy. We will also be collecting subjective data using paper-based tools such as NASA-TLX in addition to a conventional video recording of the scene of activity and handwritten notes of observation for validation purposes. We will also capture mental workload data from a few other objective sources such as heart rate variability (ECG). We expect to clearly define an approach to separately account for the effect of mental workload independent of the impact of changing light conditions in field situations for safety-critical systems. This includes a mathematical model that we innovate based on other mathematical models that are already available in the literature.

Keywords: Eye Tracking· Pupillometry· Cognitive Workload· Field Study· Maritime Usability

1 Introduction

Pupil dilation is an important metric for assessment of mental workload [13] [19], especially in a safety critical system [?] such as a helicopter (or) ship's bridge or driving [?]. In these environments, the light condition fluctuates dramatically. Since changing light conditions can also impact pupil dilation, it is necessary to separate the effect of the mental workload from the effect of the changing light conditions to be able to utilise it reliably to evaluate the risk profile.

As discovered through the literature review, currently, there is no open and validated method to measure cognitive workload in a field condition trough pupillometry. The only commercially available method is proprietary [24] ad thus closed source, it disconnects the researcher from the ability to adapt the tools to specific research questions and have a deep understanding of the variables at play [15]. Moreover, no methods have been validated for use with low-cost eye trackers, which would enable affordable data gathering, including collaborative studies with multiple eye trackers.

The relation between cognitive workload and pupillary responses has been assessed as back as 1964 [13], here Hess and Polt measured changes in the pupil size of a subject during the resolution of "simple multiplication problems" and observed a link between the pupillary response and the difficulty level. Likewise, D.Kahneman et al. [19] investigated the correlation between task-evoked pupillary diameter and memory intensive tasks; reporting different pupillary responses from the learning and report/recollect phases as well as variations directly related to the task difficulty. These initials results were validated for a variety of "intensive cognitive tasks", including language, writing, listening, speech and the solution of mathematical problems [20]. The psycho-physiological studies on workload and the pupillary response are often limited by three main factors:

- Perceived luminance, as the variable with the more significant influence on the pupil diameter, it can mask the influence of cognitive workload. It is common to keep luminance as controlled conditions to isolate its effect.
- Real-time tracking is only possible tough a limited number of the reviewed ed methods. Online tracking of cognitive workload, [38] and [24] requires high-temporal-resolution, as well as the control of the environmental variables, including the estimate of a continually changing baseline value. Lack of such features limits the study to the evaluation of well defined/separated tasks.
- Open source. The only method that currently provides a solution to both the previous limitation is tied to patented technology, only limited documentation is available on the underlying method [24], and this makes it impossible for independent researchers to reuse, adapt and improve on such systems.

1.1 Research questions

This paper is part of an ongoing, work-in-progress, research and will present only preliminary results to the following research questions:

– How to reliably measure luminance from the POV of a subject using a small calibrated video camera?
– How to calculate the baseline pupil size for a visual stimulus and use it to infer the cognitive state of the subject then?
– How to scale this method in field conditions (where luminance variate in an unpredictable manner)?

1.2 Planned contributions

– Open source code, available in a public repository that can work on files generated by at least one common eye-tracker vendor.
– Thesis report and paper with validation data.

1.3 Existing research

Cognitive workload. Mental workload can be evaluated through a variety of methodologies [4] [33]:

– Subjective-empirical measures of perceived effort as rated by the subject.
– Performance of the subject in a controlled task.
– Physiological indices of cognitive state.

Subjective reports such as questionnaires and multidimensional ratings (e.g. the NASA-TLX [12]) are indirect means of evaluation of the perceived workload. They are usually easy and cheap to administer but have several limitations [36]. As post-facto evaluations, relying on the personal impression and memory of the subject, they do not track a change over time and are therefore difficult to use for the identification of specific peaks on cognitive workload. The NASA-TLX questionnaire is a standardised assessment tool of cognitive workload. It employees a "multi-dimensional rating scale" measuring six parameters to give an estimate of the overall task workload: mental, physical, temporal, frustration, performance and effort. Like other forms of self-report, it doesn't record changes in cognitive load over time (multiple questionaries can be used to assist a complex task if divisible in subtasks). Other subjective workload measures are: Multiple Resources Questionnaire (MRQ) [2], Subjective Workload Assessment Technique (SWAT) [32] , Overall Workload Level (OWL) [18] and Integrated Workload Scale (IWS) [29].

Performance-based measures of mental workload indirectly measure the cognitive state of a subject through the execution of a standardised task. Changes (speed, accuracy, response time) in the execution of the secondary task can be interpreted as a difference in cognitive/visual workload. The ISO defined Detection Response Task (DRT) [17] is an example of a performance-based method, T.Cegovnik et al. [40] used a tactile DRT to validate the use of a Tribe eye tracker to assess changes in the cognitive load of the subjects using oculography and pupillometry. The DTR estimates the cognitive load trough response rate and miss-rate of the response task: a stimulus is delivered through a vibrator attached to the subject in a random sequence; the test measures the response time (time needed to press a button attached to the steering wheel in a driving simulator). The use of this class of methodologies is to be planned considering the effect of the controls tasks on the main task as well as the low temporal resolution of the events that can be measured. Moreover, the relationship between cognitive workload and task performance is not linear and follows an inverted U-shape as defined in the Hebb-Yerkes-Dodson Law [40]. Both overload under-load can, therefore, result in decreased performances making the measure potentially unreliable if not paired with subjective data.

Physiological indices indirectly connect a measurable psychophysical parameter to an expected mental workload. Heart rate, respiratory rate, galvanic skin response, brain activity (EEG, fRMITCD), as well as eye activity [3] are parameters that over the years have been used to measure the mental workload. Modern eye tracking has an intrinsic advantage of being unobtrusive, and less impending that most of the other aforementioned techniques and could be a reliable instrument for over time monitoring of the mental workload [4]. Still, psychophysiological measures have several limitations [40] when applied in field conditions. In most of the studies mentioned beforehand, the experimental design included two or more tasks with different levels of difficulty indirectly estimating the workload required for each of the tasks. The physiological and neurological models employed in the psychophysiological methods have to be specially designed and trained to fit a particular task-evoked neural activity. It is, therefore, difficult to compare the results to a generalised measure of workload.

Blinks Eyeblink duration and rate have been identified as an alternative metric for visual workload [1] and [31]. Unfortunately, this metric reliability is limited, as the blink rate can be influenced, in an opposite manner, by both the mental workload and visual workload.

Pupil In the Handbook of Psychophysiology [3, p. 443], J.T Cacioppo et al. defines the pupillary system as a "dually innervated organ". The pupil size is determined by the concurring action of a parasympathetically innervated constricting muscles and sympathetically innervated radial dilator muscles. The parasympathetic activity is dominant, responding to light reflexes, and determine the varying pupil size baseline, the sympathetic activity is instead connected to behavioural and stress contexts and can be used as a psychophysiological parameter of cognitive activity.

Task-evoked pupillary response. Palinko et al. [28] estimated the driver's cognitive load from pupil size measurements finding that it the pupillary response correlates with the measured driving performances, and this as similar studies seems to confirm the reliability of pupillometry as a measure of cognitive workload. However, the analysis is limited to a simulated task with low variability between target luminance. Palinko et al. [28] introduced a pupillometric cognitive load measure for real-time cognitive load changes (every several seconds).

Light and cognitive load effects on pupil diameter. Palinko et al. [27] follows up the previous study with a proof of concept of the possible separation between cognitive and ambient light components of pupil dilatation. The study was conducted using a driving simulator as a controlled environment for both an Aural Vigilance Task an Illumination Task (with different brightness targets) and a combined validation task. The study results show that it should be theoretically possible to model the psychophysical functions of the pupillary response over time to light stimuli and shows the measured trend over time. It also notes how the transitions bright/dark/bright are not equal as different muscle groups are involved in the contraction and dilation movements. The bright light reaction is quick "to protect the retina from overexposure", while the reaction to darkness is slower and gradual. The psychophysical function to predict an expected baseline pupil diameter should, therefore, take into account multiple parameters, current light level, previous light level, the rate of change, as well as age, and target. The study concludes that it is possible to discern the effects of luminance and cognitive load on pupil diameter and that the "proof of concept" predictor works in the limited experimental setting.

Discern between mentally and visually workload. M.A. Recarte et al.[30] validated the use of pupillary response as workload index in a field scenario ignoring the effect of illumination changes as the variable was impossible to control. The data they collected shows consistent results across the different driving task (no task, verbal task and visual task) collected during multiple driving session, to such a degree that cannot be explained by the sole different lighting conditions.

Unified formula for light-adapted pupil size. Since the pupil diameter can be deconstructed as the result of multiple concurring factors, in order to correctly differentiate the cognitive workload from the pupillary light response, it can be useful to compute the expected pupil diameter for a given brightness condition and use the resulting value as a baseline value upon which calculate the cognitive-driven component of the measured pupil size. In a recent paper, A.B. Watson et al. [37] (NASA Ames Research Center and University of California) have reviewed seven different published psychophysical functions defining the relation between target luminance (cd/m2) and expected pupil diameter. In the same paper, they also published a newly developed unified formula. The calculated baseline would work in the range of 2 to 8 mm, the reliability of the unified formula have to be tested to ensure that the little variability rage of the pupil size provoked by cognitive workload is preserved ($<$+-1mm)[27]. The unified formula [37] is valid only for a light-adapted condition with stable illuminant and point of view (PoV) as it doesn't account for the adaptation state or the "pupillary unrest" (low-frequency random fluctuation in the range of 0.02 to 2.0 Hz and amplitude within +-0.25 mm).

Independent variables in the unified equation:

- Luminance.
- Age (The maximum pupil diameter, as well as total range, declines as the age grows).
- Field diameter (deg).
- Number of eyes stimulated, the final diameter is dependent on the number of eyes that are adapted to the light condition they defined the "Effective corneal flux density" (the variable controlling the effective pupil diameter) as dependent on the number of eyes (attenuated by a factor of 10 for one eye).

$$F = LaM(e)$$

(F)Flux density as the product of (L)luminance, (a)area, and (M(e))monocular effect.

Wavelet analysis. S.P. Marshall et al. [24] describes a technique to identify the origin of a recorded pupillary response that works independently from the target luminance. The procedure employs wavelet analysis to identify the dilation reflexes of the subject's pupil. She explains how the reflexes can be differentiated as the pupil have different responses to light and psychosensorial stimulus. In a steady light, the pupil shows an irregular pulsation (light reflex) provoked by the interaction of the circular contracting muscles (agonist) and antagonist radial muscles act as the antagonist and are inhibited from dilating the pupil. A cognitive workload provokes a different waveform as both circular and radial muscles dilate the pupil creating a brief peak. This would imply that the cognitive workload is measured as the frequency and intensity of such events and not as a steady dilatation of the pupil (for the duration of the load), but it is unclear how this method would perform in a field condition, with highly variable ambient luminance.

An application of this technology is explained in a study conducted by S.P. Marshall for the US Navy. She applied the patented metric of Index of Cognitive Activity (ICA) [25] thanks to a networked system that is set up to record the cognitive workload for multiple team members during a collaborative task. The study assessed the performance of a three-person team in a simulation system, and the effort to overcome mission-related problems. A similar study, a collaboration between NASA Ames and EyeTracking, Inc. used the ICA and eye metrics to detect the difference between low and high fatigue states [26].

Machine learning for pupillometry. S.M. Wierda et al. [38] and the related work of A. Ferscha et al. [7] represent a different approach to the problem of the indirect assessment of mental workload. As the response time of the pupil to a mental workload event is too slow (several seconds) to be used as a real-time measure, it can be used directly only as an average over time. This makes it suitable to evaluate lengthy tasks that have a reasonably constant load in cognitive workload (at least several seconds). These two studies show a proof of concept of how to obtain an high-temporal-resolution (c.a. 10 Hz) tracking of the cognitive processes through deconvolution. The aim of real-time cognitive workload measurement gains value in the context of the implementation of a real-time feedback loop in the interaction design of a system (e.g. a system able to respond to different cognitive states of the user).

S.M. Wierda et al. [38] fixed the distribution of "attention impulses" every 100ms defining the output's temporal resolution. Employing a model of the "Task-evoked pupil impulse response," it reconstructs the intensity of the attention impulse that provoked the measured pupillary response. A. Ferscha et al. [7] further developed the concept through machine learning for better performances without the need of a fixed temporal resolution of the cognitive impulses. To reduce the effect of incident light A. Ferscha et al. [7] used the average illumination in the subject's field of view analysing the eye tracker camera stream. The technique they used is possibly still insufficient to adapt the technology to a field study with a highly variable illumination. In the described implementation a luminance change more significant than the set threshold would trigger a suspension of the tracking, this state is then maintained until the condition is stable again and a new baseline can be calculated. A similar solution was implemented to filter out blinks. The dynamic baseline is computed through a series of threshold and doesn't adjust for small changes in target luminance.

Illuminance measurement using a digital camera. Luminance as a measure needed to dynamically estimate the pupil size of a subject, candela per square meter cd/m2, is the quantity of light radiating from a source. An illuminance meter is an expensive and bulky device, in more than one instance this has resulted in attempts to use a camera as cheaper and more flexible alternative [14] and [39].

A digital sensor is at its core an array of Illuminance sensors. Each pixel measures the number of photons hitting the photoelectric surface. The presence of a Bayer filter for colour photography makes it so that to reconstruct the information form the entire visible spectrum multiple pixels have to be analysed at the same time. Each pixel in the final image has reconstructed values from the neighbouring pixels for all the RGB channels and in itself would be sufficient to reconstruct the illuminance of the scene, in order to reduce noise and gain reliability multiple pixels should be used, the number of pixels used is effectively the field of view of the instrument [14]. The formula proposed by P.D Hiscocks [14] has been optimised for a DSLR camera, not all the parameters are accessible when using an embedded digital video camera.

Parameters:

– Pixel value (0-255 for an 8bit monochrome image).
– Shutter Speed (In a video camera, this is limited by the frame rate, e.g. 1/30s for a 30fps camera) and aperture or focal ratio.
– Iso or film speed.
– Camera Constant (The calibration constant for a specific camera model that has to be determined with a known instrument).

D. Wuller [39] suggest a different model, reversing the colour processing of the camera, first from gamma-compressed RGB to linear RGB and then from linear RGB to CIE XYZ, extrapolating then $y(\lambda)$ as the relative luminance (luminance as defined by the luminosity function, reproducing the spectral luminous efficiency of the human eye). The author notes that to access the linear response of an image sensor the correct inverse gamma has to be applied and that this could deviate from the standard 2.2 (sRGB), the relative luminance can then be converted to luminance through a linear relation specific for a particular sensor/camera settings combination.

Maritime Usability and SA Endsley et al. [5] defines situational awareness (SA) as "the perception of the elements in the environment within a volume of space and time, the comprehension of their meaning and the projection of their status in the near future”. Low SA has been found to be one of the primary sources of human error in safety-critical systems [34]. Real-time monitoring of SA seems possible through the analysis of the subject visual attention aided by a variety of eye tracking data such as:

– Fixation duration: length of fixations (e.g. time spent on a single target without movement).
– Fixation rate: average number of fixations in a unit of time.
– Dwell time: the sum of all the fixation time in a single area of interest.
– Saccadic main sequence: the relation between the saccadic duration and magnitude and between peak velocity (PV) and magnitude [4], as both PV and duration increase with the magnitude.
  • Saccadic duration: the period between two positions of the fovea.
  • Saccadic magnitude: the magnitude of the saccadic movement (angle).
  • Peak saccadic velocity: highest velocity reached during saccades deg/sec.

Use of SAGAT The Situation Awareness Global Assessment Technique (SAGAT), is a global tool developed to assess SA [6]. "A simulation employing a system of interest is frozen at randomly selected times, and operators are queried as to their perceptions of the situation at that time. The system displays are blanked, and the simulation is suspended while subjects quickly answer questions about their current perceptions of the situation. As a global measure, SAGAT includes queries about all operator SA requirements, including Level 1 (perception of data), Level 2 (comprehension of the meaning) and Level 3 (projection of the near future) components. This includes a consideration of system functioning and status, as well as relevant features of the external environment. SAGAT queries allow for detailed information about subject SA to be collected on an element by element basis that can be evaluated against reality, thus providing an objective assessment of operator SA."

Ikuma et al. [16] compared different standard human factors measurement tools: workload ratings (SWAT and NASA-TLX) and Situation Awareness Global Assessment Technique (SAGAT). Eye tracking was also used to analyse the gaze path of the participants during the simulation, "the percentage of time spent looking at different areas of the screen during steady-state periods differed among workload levels". This study only looks at a small number of areas of interest (AOIs) on the interface, to infer the visual attention of the subjects for different areas of the interface. The usability of on-board interfaces on High-Speed Craft (HSC) has been assessed through the application of eye tracking technology [9]. The cognitive workload and SA of the crew of a military HSC in littoral waters were selected as of interest because of the combination of high-speed navigation and the need to navigate outside established routes. With particular interest on the role of the navigator [8] and its use of the onboard interface "Route Monitor Window". Hareide et al. [9] collected data from both field and simulator activities using the Tobii Pro Glasses 2 Eye Tracker. The methodology of the study tried to account for the difference in the environment and datasets between the simulator and field conditions. The Author followed up a mid-life update of the interface [11] [10] , with further validation of the redesigned interface for the primary objective of increased navigator attention dedicated to the "outside" Area of interest opposed to the various interfaces. Eye trackers were in this case used as to indirectly evaluate situational awareness of the navigator through quantisation of the time spent on the interface rather than observing the environment. Hareide et al. [10] apply the concept of dwell time, look-backs and Backtracks to the analysis of AOIs:

- Look-backs (returns, refixation) are saccades landing in an AOIs already visited. The analysis of a look-back can point to a variety of concurring factors: memory failure, confusion on the function of a command/element, the difficulty of content understanding and intrinsic importance of the information present in an AOI.
- Backtracks are calculated on the specific sequence of saccades and are a sudden (inverted gaze direction) rapid eye movement back to a just visited AOI. Confusion or uncertainty, changes in goals, a mismatch between the users' expectation and interface layout.

The author shows how eye tracking data can be used to guide the development of a GUI through the analysis of areas of interest and gaze behaviour, but also notes several limitations in the use of the eye tracker that need to be considered not to influence the behaviour of the user group. This includes the thickness of the eyepiece frame, creating a visible "frame of vision", unwanted reflection and glares on the protective glass, difficulty using the binoculars in conjunction with the trackers and unfavourable lighting conditions.

2 Methods

2.1 Measure of luminance from POV

This is a work in progress. his research starts with the development of the necessary tools. Even though it is theoretically possible to use a video camera as luminance meter, the reliability and accuracy of this technique will depend significantly on the software and hardware. The calibration and validation of the equipment will be done with a know good instrument in the Norwegian Colour and Visual Computing Laboratory. A colour checker will be measured with both the Konica Minolta CS-2000 [23] spectroradiometer and the World Camera mounted on the Pupil lab eye tracker (Pupil Pro). The data can then be used to calibrate the World Camera as a rudimentary luminance meter. A colour checker illuminated by a diffuse light at a variable intensity will be measured through both the Pupil Pro and the spectroradiometer. Different combinations of illumination and exposure settings on the software are required to model the sensor response.

The Pupil lab [21] software in his current version (1.10.20) does not dynamically save the exposure settings during the recording. Libuvc [**?**] (cross-platform library for USB video devices) is used to receive the video stream and communicate with the two cameras. Libuvc supports either getting or setting the exposure value and should allow retrieving the current exposure data during the recording. The lack of support of these functionalities in the Pupil lab software makes it impossible to use automatic exposure as the calibration values would change during the recording in an unpredictable manner. Using a fixed manual exposure is possible but severely limits the maximum dynamic range of the light meter.

Alternatives to the use of the camera, to simulate a scenario in which the aforementioned limitation does not apply, which would be easily reachable with some interest from the developers, is to use an external light meter mounted on the eye-tracker. This would provide a measurement that is not bound to the limited dynamic range of the camera with the drawback of having two disconnected data streams.

Instrumentation The Konica Minolta CS-2000 spectroradiometer [23] is a high precision polychromatortype spectroradiometer, it will allow measurement on a vast range of luminance (0.3 to 500,000 cd/m2) with a $\pm 2\%$ accuracy. The Pupil Pro [21] and [22] World Camera is mounted just above the subject eye, facing outward. The camera offers different combinations of resolution and frame-rate 1920x1080 @30fps, 1280x720 @60fps, 640x480 @120fps covering a FoV of 60 or 100 degrees diagonally (depending on the lens).

2.2 Calculate the baseline pupil size

The measure of luminance from the POV will be connected to the unified formula for light-adapted pupil size developed by [37]. Two elements will need validation:

- the accuracy and precision of the unified formula.
- the different possible methods to convert the input from the camera to the correct input values for the formula.

The [37] unified formula is based on a standard procedure involving a defined stimulus: the observer is shown a bright circle on a dark background. The size (degrees of field of view) and luminance (cd/m2) of the circle determine the corneal flux density (i.e. the product of luminance and subtended area) as defined by [35]: $D = 7.75 - 5.75[(F/846)0.41/((F/846)0.41 + 2)]$ "where D is the pupil diameter (mm), and F is the corneal flux density (cdm-2deg2)".

The model implies that at its core the pupil control mechanism reacts as a 'flux integrator', following an S-shaped curve.

An image from the camera has to be used to evaluate the flux density or to indirectly convert the image into a corresponding standardised stimulus (i.e. a circle on a dark background). The most promising approach is to consider the average luminance on the camera sensor (or external light sensor) as equal to the luminance of a standardised stimulus (bright circle) as wide as the entire field of view (FoV) (120-190°).

This assumption ties the precision of the calculated pupil size to how well the camera FoV matches the user FoV). To test the quality of the model NTNU students and staff (age from 20 to 50) will be recruited for validation in a controlled environment.

Validation The procedure will refer to the methods used by Palinko et al. [27]. It will be divided into three parts:

- No-load - variable light-adapted state. The participants will be sitting in a dark room looking at a selection of projected images; the sitting position will be adjusted to maintain a constant field of view and distance from the screen. The projected images will include standardised stimuli as well as more complex images (e.g. outdoor naturalistic scenery). Each image will be represented for several seconds to let the pupil reach an adapted state (c.a. 15s). Each image will include a focal point ad the participants will be asked to stare at the focal point.
- Load - static light-adapted state. With a constant ambient luminance (e.g. grey image projected), the participants will be asked to perform an Aural Vigilance Task (AVT), as in [27]. The task involves listening to a voice counting from 1 to 18, repeated multiple times. Every 6th number (6, 12, and 18) might contain errors (i.e. another number is replaced to the correct sequence). The participants would have to perform an action such as pressing a button when they detect an error. The task should induce an increased cognitive workload near every 6th number. The location of the error should be randomly selected for each session.
- Load - variable light-adapted state.
  The same AVT task is repeated but with variable standardised images being projected.

2.3 Measure of luminance from POV

Subjects The subjects for the final session will be recruited as cadets of the Royal Norwegian Naval Academy (RNoNA) and will require access to the training vessel (Kvarven). The crew of a training vessel includes navigator, assistant, helmsman and training instructor. During a study session, the vessel would also include one to three researchers to set up and record the data. Depending on the availability of multiple eye trackers both the navigator and helmsman could participate in the experiment for each session.

Procedure The setup for each session will include:

- Introduction to the research and signature of the informed consent.
- Application of the eye tracker (glasses and recording device).
- Calibration of the eye tracker.
- Reference measurements of ambient illumination.
- The debriefing will include a short interview and the NASA-TLX questionnaire as a further reference of the cognitive workload.

Two researchers should be present on board at any time. Between each session, up to 30 minutes will be required for cleaning of the instrumentation, download of the test data and recharge of the various batteries.

The task aims to highlight different levels of cognitive workload. The navigation task should be repeatable Fig. 1, and it should last less than one hour (not including the setup and debriefing) and should include a mix of low and high workload for the subjects: E.g. Steady navigation - change of course - steady navigation.

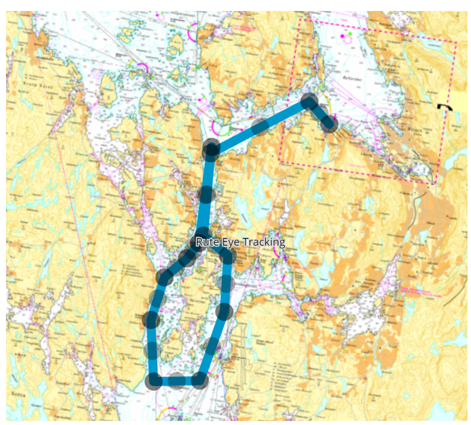

Fig. 1. The course suggested by Odd Sveinung Hareid from Laksevåg (Bergen)

3 Expected Results

3.1 Proof of concept

This is a work in progress, at the time of writing, the research is still in the initial exploratory phase and should be completed by the end of May 2019. The literature review helped in defining a path to follow in order to develop the necessary tools (the pupil baseline calculation), but there are inherent risks and challenges in the generalisation of findings that were initially only meant for controlled condition/laboratory study. A series of non validated tests is being carried out to determine:

- weather the camera-based luminance meter works well enough for a reliable field application (limited dynamic range of the camera, limited bit depth, the difference in the camera FOV compared to the subject FOV).
- Weather pupil baseline calculation is precise to such a degree not to mask the cognitive workload.
- Weather the chosen eye tracker can operate in field conditions.

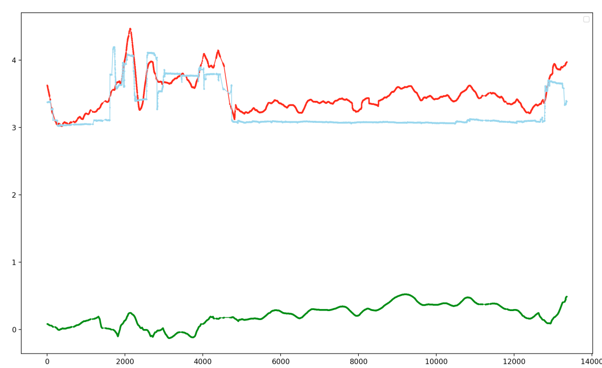

Fig. 2. This sample data output was recorded from a user sitting in front of a laptop in a dark room. The green line is the pupil size (mm) as measured by the eye tracker; the red line is the baseline as calculated from the video data and the blue line is the difference between the two. The blue will ultimately represent the cognitive workload. In this sample, the middle "steady" part has been measured on a subject performing an IQ test. The horizontal axe, time, is expressed in video frames at 30fps.

A series of artefacts have been identified in the sample data collected:

- Pupillary overshoot, the model of the pupil size is specific to adapted state and doesn't account for the pupil natural overshoot that can be observed when a rapid change in luminance occurs Fig. 2.

– Pupillary unrest, in the form of low-frequency random fluctuation in the range of 0.02 to 2.0 Hz and amplitude within +-0.25 mm Fig. 2 and Fig. 3.
– Incorrect measured pupil range; the pupil size is calculated from the video image in pixels to an estimated mm by the 3d model. different calibration of the pupil camera (distance from the eye) can bring the measured range outside the unified formula unified formula [37] range (c.a. more than 2mm and less than 8mm).
– Luminance outside the camera dynamic range, this is visible in the second plot, in this case, an outdoor recording ended inside a building, the completely black image from the video produces an unreliable luminance reading Fig. 3.

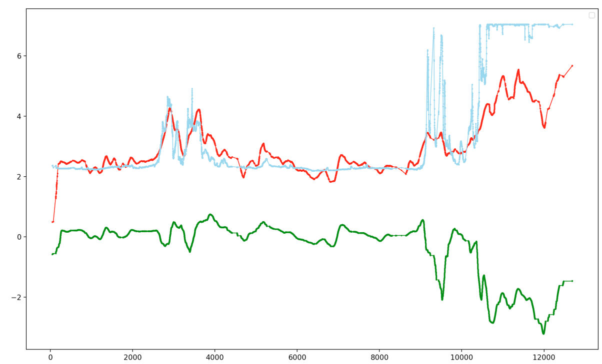

Fig. 3. This outdoor session shows the limitations of the camera if used with a fixed exposure.

During the test multiple data sources will be combined including: Heart rate variability, POV camera, eye tracking camera all will be used to substantiate the measure of cognitive workload. Dates and travel The number of sessions depends on the availability of cadets, five to ten subjects would be a good result. Given the nature of the experiment weather and ambient illumination conditions should be kept constant within a reasonable range. This could require the spread of the study over multiple days.

To account for the limitations of the eye tracker, it would be advisable to plan a portion of the sessions after dawn, (lower the contrast between the user interface inside the cabin and the outside).

4 Conclusion

The development of the necessary tools is in progress and will hopefully end as a refined proof of concept and validation of the method with the intent of attracting the interest of developers to consolidate the application. The validation that will be attempted as part of the research will be by no mean be exhaustive, it is expected that the interest surrounding the measure of the cognitive workload will result in a variety of experiments on the topic, to further explore the benefits and limitations of the developed methods.

The choice of a camera as a luminance meter could severely limit the accuracy of the method but would allow a variety of head-mounted eye trackers to be used for cognitive workload studies without the need of any extra hardware.

References

1. Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., Montanari, R.: Driver workload and eye blink duration. Transportation Research Part F: Traffic Psychology and Behaviour **14**(3), 199–208 (May 2011). https://doi.org/10.1016/j.trf.2010.12.001, http://linkinghub.elsevier.com/retrieve/pii/S136984781000094X
2. Boles, D.B., Adair, L.P.: Validity of the Multiple Resources Questionnaire (MRQ). Proceedings of the Human Factors and Ergonomics Society Annual Meeting **45**(25), 1795–1799 (Oct 2001). https://doi.org/10.1177/154193120104502508, http://journals.sagepub.com/doi/10.1177/154193120104502508
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. Di Stasi, L., Marchitto, M., Antolí, A., Cañas, J.: Saccadic peak velocity as an alternative index of operator attention: A short review. Revue Européenne de Psychologie Appliquée/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
5. Endsley, M.R.: Design and Evaluation for Situation Awareness Enhancement. Proceedings of the Human Factors Society Annual Meeting **32**(2), 97–101 (Oct 1988). https://doi.org/10.1177/154193128803200221, http://journals.sagepub.com/doi/10.1177/154193128803200221
6. Endsley, M.R.: Direct measurement of situation awareness: Validity and use of SAGAT. In: Situation awareness analysis and measurement., pp. 147–173. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US (2000)
7. Ferscha, A., Gollan, B.: Modeling Pupil Dilation as Online Input for Estimation of Cognitive Load in non-laboratory Attention-Aware Systems. COGNITIVE 2016: the Eighth International Conference on Advanced Cognitive Technologies and Applications: March 20-24, 2016, Rome, Italy (2016)
8. Gould, K.S., Røed, B.K., Saus, E.R., Koefoed, V.F., Bridger, R.S., Moen, B.E.: Effects of navigation method on workload and performance in simulated high-speed ship navigation. Applied Ergonomics **40**(1), 103–114 (Jan 2009). https://doi.org/10.1016/j.apergo.2008.01.001, http://linkinghub.elsevier.com/retrieve/pii/S0003687008000094
9. Hareide, O.S., Mjelde, F.V., Glomsvoll, O., Ostnes, R.: Developing a High-Speed Craft Route Monitor Window. vol. 10285, pp. 461–473. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-58625-$0_3 3, http://link.springer.com/10.1007/978-3-319-58625-0_3 3$
10. Hareide, O.S., Ostnes, R.: Maritime Usability Study by Analysing Eye Tracking Data. Journal of Navigation **70**(05), 927–943 (Sep 2017). https://doi.org/10.1017/S0373463317000182, https://www.cambridge.org/core/product/identifier/S0373463317000182/type/journal$_article$
11. 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-$1_2 2, http://link.springer.com/10.1007/978-3-319-91467-1_2 2$

12. 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
13. 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
14. Hiscocks, P.D., Eng, P.: Measuring Luminance with a Digital Camera. Syscomp Electronic Design Limited p. 27 (2014)
15. Holmqvist, K., Andersson, R.: Eye tracking: a comprehensive guide to methods and measures (2017), oCLC: 1057373387
16. Ikuma, L.H., Harvey, C., Taylor, C.F., Handal, C.: A guide for assessing control room operator performance using speed and accuracy, perceived workload, situation awareness, and eye tracking. Journal of Loss Prevention in the Process Industries **32**, 454–465 (Nov 2014). https://doi.org/10.1016/j.jlp.2014.11.001, https://linkinghub.elsevier.com/retrieve/pii/S095042301400179X
17. 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)
18. Jung, H.S., Jung, H.S.: Establishment of overall workload assessment technique for various tasks and workplaces. International Journal of Industrial Ergonomics **28**(6), 341–353 (Dec 2001). https://doi.org/10.1016/S0169-8141(01)00040-3, http://linkinghub.elsevier.com/retrieve/pii/S0169814101000403
19. 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
20. 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
21. 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
22. Kassner, M., Patera, W., Bulling, A.: Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction (Apr 2014), http://arxiv.org/abs/1405.0006
23. Ltd, K.M.S.S.P.: CS-2000 SPECTRORADIOMETER. Konica Minolta Sensing Singapore Pte Ltd (Mar 2018), http://sensing.konicaminolta.asia/products/cs-2000-spectroradiometer/
24. 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
25. Marshall, S.P.: What the Eyes Reveal: Measuring the Cognitive Workload of Teams. In: Duffy, V.G. (ed.) Digital Human Modeling, vol. 5620, pp. 265–274. Springer Berlin Heidelberg, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02809-$0_29, http : //link.springer.com/10.1007/978 - 3 - 642 - 02809 - 0_29$

26. Marshall, S.: The Index of Cognitive Activity: measuring cognitive workload. In: Proceedings of the IEEE 7th Conference on Human Factors and Power Plants. pp. 7–5–7–9. IEEE, Scottsdale, AZ, USA (2002). https://doi.org/10.1109/HFPP.2002.1042860, http://ieeexplore.ieee.org/document/1042860/
27. 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
28. 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
29. Pickup, L., Wilson, J.R., Norris, B.J., Mitchell, L., Morrisroe, G.: The Integrated Workload Scale (IWS): A new self-report tool to assess railway signaller workload. Applied Ergonomics **36**(6), 681–693 (Nov 2005). https://doi.org/10.1016/j.apergo.2005.05.004, http://linkinghub.elsevier.com/retrieve/pii/S0003687005000918
30. Recarte, M.A., Nunes, L.M.: Effects of verbal and spatial-imagery tasks on eye fixations while driving. Journal of Experimental Psychology: Applied **6**(1), 31–43 (2000). https://doi.org/10.1037/1076-898X.6.1.31, http://doi.apa.org/getdoi.cfm?doi=10.1037/1076-898X.6.1.31
31. Recarte, M., Pérez, E., Conchillo, , Nunes, L.M.: Mental Workload and Visual Impairment: Differences between Pupil, Blink, and Subjective Rating. The Spanish Journal of Psychology **11**(2), 374–385 (2008). https://doi.org/10.1017/S1138741600004406
32. Rubio, S., Diaz, E., Martin, J., Puente, J.M.: Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. Applied Psychology **53**(1), 61–86 (Jan 2004). https://doi.org/10.1111/j.1464-0597.2004.00161.x, http://doi.wiley.com/10.1111/j.1464-0597.2004.00161.x
33. 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
34. Sharma, C., Bhavsar, P., Srinivasan, B., Srinivasan, R.: Eye gaze movement studies of control room operators: A novel approach to improve process safety. Computers & Chemical Engineering **85**, 43–57 (Feb 2016). https://doi.org/10.1016/j.compchemeng.2015.09.012, https://linkinghub.elsevier.com/retrieve/pii/S0098135415003075
35. Stanley, P.: The effect of field of view size on steady-state pupil diameter. Ophthalmic and Physiological Optics **15**(6), 601–603 (Nov 1995). https://doi.org/10.1016/0275-5408(94)00019-V, http://linkinghub.elsevier.com/retrieve/pii/027554089400019V
36. 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
37. 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

38. 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
39. Wüller, D., Gabele, H.: The usage of digital cameras as luminance meters. p. 65020U. San Jose, CA, USA (Feb 2007). https://doi.org/10.1117/12.703205, http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.703205
40. Čegovnik, T., Stojmenova, K., Jakus, G., Sodnik, J.: An analysis of the suitability of a low-cost eye tracker for assessing the cognitive load of drivers. Applied Ergonomics **68**, 1–11 (Apr 2018). https://doi.org/10.1016/j.apergo.2017.10.011, https://linkinghub.elsevier.com/retrieve/pii/S0003687017302326