# Accounting for effects of variation in luminance in pupillometry for field measurements of cognitive workload

Giovanni Pignoni, Sashidharan Komandur and Frode Volden

***Abstract*—Eye-tracking is now above and beyond the sole measurement of visual attention. Amongst the multiple measures it provides, some have been explored as a measure of cognitive workload (CW). One such measure is pupil diameter. Although the relationship between pupil size and CW has been extensively documented, pupil diameter is primarily impacted by luminance variations while the cognitive workload has a relatively minor influence. Therefore luminance variations have to be accounted for, either in the experimental design or in the data processing to avoid the masking of the CW effects. This has meant that the use of pupillometry for the measurement of anything but the pupillary light response, has been restricted to highly controlled lighting conditions in a laboratory.**

Pupillometry in Field Conditions

Cognitive Workload

This study proposes a new method that uses point of view (POV) video in conjunction with a luminance measurement sensor to dynamically estimate the luminance of the visual stimuli. As currently available off the shelf eye trackers are usually not equipped to record luminance variations, a luminance sensor was added to a commercial eye tracker. Eye-tracking gaze data, POV video recording of the operator/observer and a head-mounted (POV) luminance sensor together estimate the expected pupil diameter. This estimate over time is due to sole influence of luminance variations. This expected pupil diameter is used as baseline for the cognitive workload. The method was validated in laboratory conditions with controlled visual stimuli. The method reliably measures induced cognitive workload despite luminance variation.

***Index Terms*—Pupillometry, Workload, Cognitive Workload, Gaze tracking, Signal Processing**
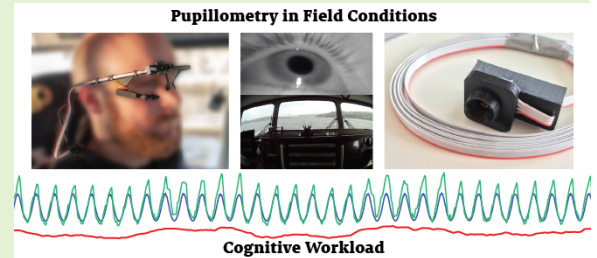
I. Introduction

ACCURATE measurements of cognitive workload can be a valuable resource in the analysis of interaction with safety-critical systems, but it often requires capturing data from operators in challenging and difficult to control, field conditions [1] [2] [3] [4]. The better the estimate of actual cognitive workload, captured in natural operating conditions, the better is the input data for the design of such safety-critical systems [1]. Amongst the available technologies, such as heart rate variability (HRV), brain activity (fMRI, EEG) and eye-tracking, only a few are suitable for use in a field study [5] [6]. As eye-tracking is a widely used tool in design and human factor studies, thanks to its ability to provide multiple measurements with a single device, it would be a natural extension of such studies to analyse pupillometry, a validated metric of cognitive workload [7] [5] [6] [8]. Although eye-tracking devices have evolved in such a way to be portable, wearable, and usable in field conditions, pupillometry still has limited application due to the effect of ambient light variations on pupil dilation [5] [6]. This research aims to develop a method that accounts for this effect and allows wider use of

Giovanni Pignoni and Frode Volden are with the Norwegian University of Science and Technology, Institutt for design, Gjøvik, Norway, Sashidharan Komandur is with the Institute for Energy Technology, Halden, Norway

pupillometry in field studies with focus on the measurement of cognitive workload.

A. The issue of the pupillary light response

The relation between cognitive workload and pupillary responses has been thoroughly documented; observed as back as the the 60s [7] in concurrency with memory-intensive tasks [9] as well as mathematical and language-based tasks [10]. Therefore pupillometry has become a widely accepted and validated measure of CW [5] [6]; still, the conditions within which validity holds are limited by multiple confounding factors, primarily the visual stimuli and ambient illumination [5] [6] [8]. Changes in luminance, as measured from the point of view of the operator, result in an involuntary response from the pupil [8]. As the magnitude of change that light has on the pupil diameter can be ten times the measurable influence of CW, it can easily mask such effects [8], limiting the application of pupillometry beyond laboratory conditions. The only other method capable of providing a luminance independent CW measurement, through eye-tracking, is a closed source and proprietary implementation [11], recently revisited by a third party [12]. This method is based on the wavelet analysis of pupil diameter oscillation and has shown promising results, still the application of this particular technology is limited as

it requires high end eye tracking equipment. A system capable of removing the effect of light from pupillometry would enable high temporal resolution tracking of cognitive workload [13] and would represent an essential step for the implementation of real-time CW monitoring systems. The hardware generally available on wearable eye trackers dictates the interest in using a camera as luminance sensor. Most commercial wearable eye tracker (Tobii, SMI, Pupil Labs, ASL) are equipped with a scene camera (Point of View), but not necessarily include an illuminance (ambient light) or luminance (luminous intensity) sensor.

B. Research Questions

- Can a small video camera be used to characterise the visual stimulus a subject is experiencing (focusing on the visual characteristics affecting the pupillary light response)?
- If so, can this data (the measured visual stimulus) be used to isolate the effect of cognitive workload in field conditions (where luminance varies with little or no control)?

II. BACKGROUND

A. Cognitive workload

Cognitive Workload (CW), in human factors and usability, is defined as a human-centred metric. It results from the interaction between a user, its unique set of characteristics (e.g. psychological state and experience/training), and a task [14] [15].

CW measurements can be divided into three major categories [16] [17] [15]:

- Subjective perceived CW (e.g. the NASA-TLX [15]), which rely on various form of self report, are easy and economical to administer but prone to great variability [18].
- Performance based measurements, tracking overload or under-load but still subject to contextual variations [19].
- Physiological indices, able to provide non-intrusive, data-rich and objective measurement of CW over time, such as heart rate variability and eye activity [20] [16] within the limits of confounders and setup/equipment cost.

B. Pupillometry

Eye-tracking is increasingly being considered a promising tool for the measure of cognitive workload [16]. The link between CW and eye tracking is Pupillometry (the measurement of pupil diameter). Three distinct stimuli affect pupil diameter (PD): brightness (pupil light response PLR), near fixation (pupil near response PNR) and cognitive activity, arousal or mental effort, (psychosensory pupil response PPR) [8]. Pupil responses can often be both reflexive and voluntary (modulated by high-level cognition): the pupil constriction that results from a change in light is involuntary and will always be inversely proportional to it. Still, attention on different areas in the field of view can affect the magnitude of such response (e.g. the eye adapt to the area where the attention is directed

to) [8] [21]. The PLR or parasympathetic activity dominates the response (2 to 8 mm) it can, therefore, be considered as the baseline of the small (<1 mm) sympathetic activity connected to behaviour, stress and cognitive activity [8] [22].

C. Task-evoked Pupillary Response

The influence of CW and arousal on the PD has been of interest for the psychology community since the 60s. Multiple studies explored and validated it as a reliable indicator of effort and arousal for a variety of tasks and stimuli such as: arousing images [23] [24], mental calculations of different difficulty [7] and short-term memory load [9] [10]. The conclusions of these early studies still hold, the effects of arousal and cognitive effort are comparable and proportional to the intensity of the stimuli rather than the valence (i.e. mental activity causes pupil dilatation) [8]. The theoretical development that followed in this area has concentrated on the validation of pupillometry for different and specific task manipulations. Correlation of pupillometry and difficulty of manual-visual tasks in Human-Computer Interaction, including reading/comprehension, mathematical reasoning, information search/retrieval and (digital) object manipulation have been explored [25] [26].

D. Unified formula for light-adapted pupil size.

The influence of light on pupillometry (PLR) is recognised and partially accounted for in multiple studies [27], Oskar Palinko and Andrew L. Kun. [28] attempted to isolate the effect of luminance from the pupil dilatation of a user in a driving simulator. It showed a proof of concept of how the pupillary response can be modelled and thus predicted/removed.

The simulation of the PLR response to changes in luminance requires the modelling of two subcomponents of the pupillary response: one regarding the response in the time domain and one regarding the extent of the response.

The response time of the pupil and the speed/shape of such response to a change in light varies significantly in different conditions, transitions from a bright to a dark stimulus do not mirror as different muscle groups are involved in the contraction and dilation movements [27] [22].

The time component is described in details by Sebastiaan Mathôt [8]: The pupil shows 0–0.2s of latency from an increase in luminance; the latency depends on a variety of factors including stimulus intensity and age. After the latency period, the pupil will constrict rapidly to adapt to the increased luminance (0.2 to 1.5s). Once adapted, the pupil remains relatively stable but can un-constrict slightly depending on the light stimulus (colour). The dilatation process triggered by a decrease in luminance is substantially slower (up to 30s) but the majority of the change happens within circa 5s. A high temporal resolution pupillometry based CW measurement can therefore only be achieved if this behaviour is modelled, the resolution is otherwise limited to several seconds.

Andrew B. Watson and John I. Yellott [29] (NASA Ames Research Center and the University of California) have published a review of seven psychophysical functions of target luminance (cd/m^2) and expected PD as well as developing a

unified formula based on the review. This model can be used to compute the expected PD for a given standardised visual stimuli, describing the PLR over time, and could theoretically be used as a baseline value to isolate the CW component of measured PD.

The expected PD described by the equation ranges within 2 to 8 mm in a light-adapted condition to a stable illuminant and fixed point of view (POV).

The unified formula for light-adapted pupil size [29] is based on a standardised variable visual stimulus (luminous circle on dark background) defined trough two parameters that determine the "Effective corneal flux density" F ($cd/m^2 * deg$). F It is the product of $L = luminance$ cd/m^2, $a = fielddiameter$ (degrees of view) and $M(e)$, attenuation factor to compensate for monocular vision (with e as the number of eyes $M(1) = 0.1, M(2) = 1$)1. The pupil diameter "D_{sd}" is computed from the 1995 Stanley and Davies formula [29] 2 and corrected for age y with a linear transformation obtaining the expected pupil diameter PD (mm) [29] 3. The parameter $y0$ is the estimated mean age of the population of observers used by Stanley and Davies [29] and is kept equal to 28.58 years.

$$F = LaM(e) \tag{1}$$

$$D_{sd} = 7.75 - 5.75[\frac{(F/846)0.41}{((F/846)0.41 + 2)}] \tag{2}$$

$$PD = Dsd + (y - y0) * (0.02132 - 0.009562 * Dsd) \tag{3}$$

E. Luminance measurement

A remote (POV) measure of luminance, the photometric measure of luminous intensity per unit emitting area (cd/m^2) is needed to estimate the adapted PD; usually requiring the use of a spectroradiometer.

A digital video camera, although not as accurate as a spectroradiometer, is theoretically capable of capturing luminance of a much larger FOV and take measurements of different parts of the scene at the same time (up to individually every single pixel); multiple attempts have been made to use a camera in this fashion [30] [31]. In practice, the use of digital cameras is limited by multiple factors such as the presence of a Bayer filter, the limited dynamic range of the sensor and the heavy image processing that most commercial cameras do to produce usable images. These factors will introduce non-linearities that can make calibration difficult. D. Wuller [31] attempts to calibrate a camera by partially reversing the image processing. This involves converting the image from gamma-compressed RGB to linear RGB and then to CIE XYZ. The $y(\lambda)$ can then be used as relative luminance (luminance as defined by the luminosity function, reproducing the luminous spectral efficiency of the human eye) but relative to the exposure setting of the camera. This procedure is limited by how much the camera postprocessing deviates from the standard 2.2 gamma (sRGB).

III. METHODS

A Pupil Labs Eye-Tracking Glasses (ETGs) [32] was selected as the base hardware, it was chosen as it is the most affordable wearable eye-tracker and due to the open-source nature of its software.

The "Pupil Capture" software, provided by Pupil Labs, handles the video streams from the ETGs (POV scene camera and eye camera) and performs on-line pupil detection, gaze tracking, calibration and markers tracking in the environment. The eye detector algorithm in "Pupil Capture" fits a geometrical model to the eye video stream and calculates the gaze angle as well as other artefacts such as pupil diameter (PD) and blinks. PD is expressed in pixels as directly measured from the video frames.

The Pupil Headset was equipped with the "high-speed camera" , all the recordings of the scene camera for this study were configured at 1280x720 @ 60fps using a 100 deg. field of view lens.

Pupil Capture (v1.11) doesn't allow fine control over manual exposure of the scene camera, rendering proper calibration of the camera impractical. Likewise, it is not possible to track exposure changes in the automatic mode making it impossible to differentiate between a change in luminance in the scene and automatic change in exposure settings. These software/hardware limitations required the addition of an external sensor for absolute luminance measurement. This is the sensor that we integrated in to the Pupil Labs ETGs. Although the built in camera has limitations (dynamic range and automatic range), it holds numerous advantages. It has the ability to record multiple points (pixels) at the same time. This allows the retrieval of luminance data in any point of the scene while a luminance sensor would require to be always pointed in the direction of the gaze.

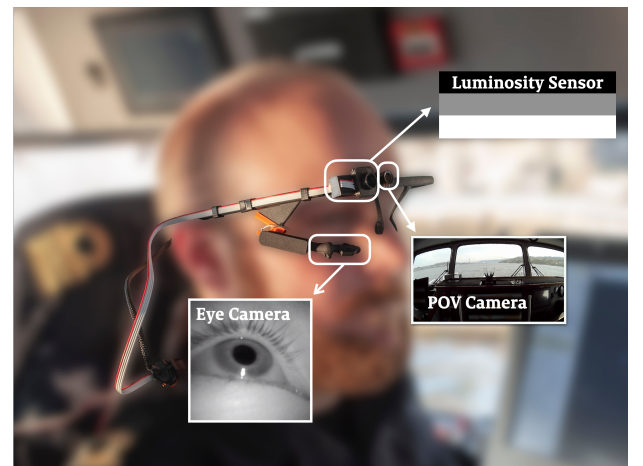### A. Additional Software and Hardware

Fig. 1. The Pupil Headset with installed the TSL2591 luminance sensor module.

The sensor module TSL2591 [33] light-to-digital converter has been added on the frame of the ETGs, mounded alongside the scene camera and used to measure the average luminance in the POV of the subject. Illuminance (lux) is derived from

applying an empirical formula that approximates the human eye response combining data from two photo-diodes, a broadband unit and an infrared unit. The sensor is secured to the eye tracker with a custom bracket and shielded from incident light with a removable hood, the field of view (FOV) of the sensor is limited to circa 60 deg, roughly corresponding to the FOV of the camera at the selected resolution. The Sensor module is connected to an Arduino based data logger, the data can be saved directly on a computer trough usb or on a portable memory card.

The combination of the scene camera and TSL2591 sensor are used to characterise the visual stimuli the subject is experiencing. Thus allowing an estimation of the pupil size as it would change due to the sole effect of the visual stimulus. This estimate is calculated based on the unified formula for light-adapted pupil size [29].

To use the unified formula in field conditions, the visual stimulus (scene) has to be deconstructed into a standardised stimulus that can be fed to the formula. The luminance sensor reading can be used as a measure of average luminance of the entire binocular field of view (200 deg (w) x 135 deg (h)). Thus it approximates a diffuse luminance field scene (e.g. outdoor conditions with vision adapted to the environment). In order to evaluate complex scenes a more refined model has been developed which integrates data from the sensor with POV video (scene) and the gaze position.

The video processing proceeds as follows, for each frame: the luminosity function 4 is applied to the scene to obtain a map of the relative luminance rL (rgb pixel values (R,G,B) converted to relative luminance) as well as the average relative luminance of the entire scene. The area around the gaze is isolated through a "Grab Cut" algorithm, selecting similar pixels around the gaze and the average relative luminance is calculated for the area of interest (AOI). The relative luminance of the AOI is combined with the reading of the external sensor to obtain an absolute luminance measurement of the selected area. This method is based on the paradigm of an area of interest (AOI) based on the gaze point. The AOI is defined as the area of a video frame surrounding the gaze point. The assumption is that the visual adaptation field will roughly correspond to the AOI as the observer adapts to a variable area around the gaze. This should correctly characterise a non uniform visual field where the eye adapts only to a particular area [8]. The luminance information encoded in the video is expressed as RGB values and therefore as a relative luminance 4, [34], [35].

$$rL = 0.2126 * R + 0.7152 * G + 0.0722 * B \qquad (4)$$

The absolute luminance L (cd/m^2), required by the unified formula [29] can be retrieved from the video using the external sensor as calibration (as the sensor has similar FOV to the on-board camera). The sensor measurement is divided by the solid angle (steradians, FOV of the sensor) to retrieve the average luminance $avgL$ (cd/m^2) in front of the observer 5.

$$avgL = \frac{Ev(lux)}{2.2sr} \qquad (5)$$

The maximum luminance, threshold of what the camera can measure before clipping with a given exposure, is retrieved comparing the average relative luminance $avgRL$ (average luminance $avgL$ of the entire video frame) to the average absolute luminance measured by the sensor $avgL$, the minimum luminance is assumed as tending to zero.

$$minL = 0 \qquad (6)$$

$$maxL = avgL/avgRL \qquad (7)$$

The average relative luminance from a portion of the video frame representing the area of interest (AOI) obtained trough the "Grab Cut" algorithm 2, $aoiRL$ is converted to absolute luminance (L) with a linear interpolation between the minimum and maximum luminance previously calculated, weighted by the spot relative luminance (0-1) 8.

$$L = (maxL * aoiRL) + (minL * (1 - aoiRL)) \qquad (8)$$

The data processing is visualised in figure 2. The PD measurement is then processed to remove artefacts such as Hippus, camera movement, blinks and general instability of either the 2D or 3D algorithm. The pupil camera was set to the resolution of 400x400px @ 120hz. A Savitzky-Golay [36] low-pass filter is used to remove a significant amount of noise while preserving the shape/height of the waveform peaks.

The measured PD is expressed in pixels and needs to be scaled to mm. The coefficient of the pixel density (pixel per millimetre) of the camera varies for each set up (distance and angle between the camera and the eye). To estimate the scaling coefficient, the measured PD is fitted to the expected PD (i.e. the scaling coefficient is the ratio between the average pixel PD and the average expected PD).

The final output of the data processing is the difference between measured and expected PD: the Δ (change) PD. It is composed of the residual impact of cognitive workload on the pupil plus noise. Due to how the PD has been scaled, a negative Δ PD indicates below-average CW and a positive represents an above-average CW compared to the average of the entire recording.

B. Experiment

An experimental session, conducted in a controlled environment, was organised to evaluate the performance of the ETGs and the overall data processing system.

The experiment is a step by step validation of our algorithm and method. First it was documented that the task difficulty alone resulted in changes in pupil diameter under fixed light conditions. Secondly, we added sinusoidal light variation on top of that, and applied our algorithm to see if changes in CW still could be measured as changes in pupil size.

The experiment included two conditions of interest:

- Control - Variable CW with fixed visual stimuli (luminance).
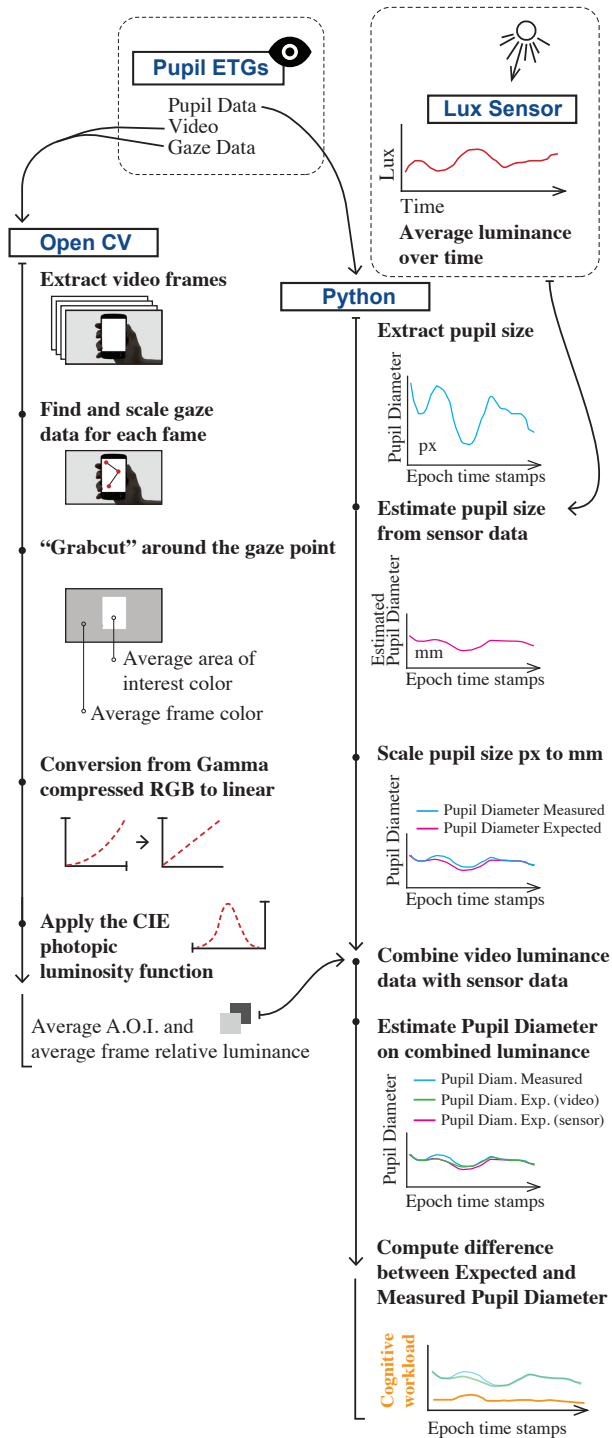- Case - Variable CW with variable visual stimuli.

Fig. 2. A simplified map of how the data is being processed in order to estimate the cognitive workload.

include instructions. The participants were verbally instructed during the experiment; the information regarding each task was provided in a short briefing session immediately preceding each task. The participants were instructed to focus on the centre of the projector screen and not to close their eyes while performing the tasks. This was necessary in order to avoid variation of the point of view and maintain a constant exposure to the visual stimuli.

1) Visual Stimuli: The Case (Variable Visual Stimuli) condition was delivered by projecting a solid colour on a projection screen placed in a dark room; the projection changes luminance following a sinusoidal (0-1 RGB grayscale) wave at 0.1Hz. The control condition utilises the same setup but keeps the projected visual fixed to a mid-grey value. The luminance is manipulated by changing the RGB values of a solid colour, occupying the entire projected area. The variable visual stimuli ranged between <1 cd/m^2 (RGB black) and 105 cd/m^2 (RGB white), the fixed visual stimuli was set at a constant 20 cd/m^2 (RGB 50% gray). The experimental design includes changes in luminance amplitude sufficient to induce the pupil to dilate/constrict over almost its entire range (maximum range is supposed to be from c.a. 2 to 8mm, measured range was often around 3.5 to 8 mm).

To avoid a "learning curve" or "memory effect", the task could be performed only once by each participant and therefore in a single experimental condition.

The sinusoidal wave of the variable visual stimuli is the same for each task/rest condition and has been selected to be sufficiently slow to let the eye have time to adapt to the light conditions as it changes. The relatively low frequency of 0.1hz has been chosen to avoid discomfort to the participants.

It is assumed that the effect of the visual stimuli (which is sinusoidal) will result in the pupil size following a sinusoidal pattern. In this specific scenario the effect of light can be filtered out from the measured pupil signal using a narrow band-stop filter. A band stop filter is a filter that leaves most of a signal unaltered but will attenuate a specific range of frequencies. The result of the band-stop filter is expected to correspond to the results of the algorithm as the majority of the light effect will have a frequency close to the light change itself. The comparison between the two has been used as an initial validation.

2) Participants: The experiment included Twenty-one (twelve females) participants recruited through convenience sampling from the Norwegian University of Science & Technology (campus in Gjøvik). The median age for females was 25 years (range 23-61 years) and the median age for males was 30 years (range 26-34 years).

3) Independent Variables:

- Task difficulty (Induced CW). The tasks was selected based on what we believe to be of increasingly difficulty. Starting from "counting upwards", believed to be a highly automated and easy task, and ending with the "Fibonacci sequence", which puts an high demand on working memory.
- Stimuli luminance cd/m^2.

4) Dependent Variables:

The variable CW was controlled through a series of mental tasks of increasing difficulty (see table I). The visual stimuli presented to the test-subjects consisted of a focus point and a background, with either constant (control condition), or systematically variable luminance (case condition). The visual stimuli were used solely to manipulate the perceived luminance; it was not part of the task manipulation and did not

- Measured PD, dependent on CW, stimuli luminance and noise (precision of the instrument, hippus).
- expected PD (dependent on the stimuli luminance).
- Δ PD difference between measured and expected PD (cognitive workload).

5) Apparatus: The core instrument utilised in the experiment is the Pupil Pro ETGs (now Pupil Core) [32], the eye camera recorded at 400x400px @120 Hz while the world video was recorded at 1280x720 @60fps. A tsl2591 Lux sensor [33] [37], mounted alongside the world camera, was used for average luminance logging @10Hz.

The experiment is constructed inside the open-source software package "PsychoPy" [38], [39]. The visual stimuli are presented trough the EPSON EB-1776W projector; the project was the only light source in the room. The setup has been characterised with a Konica Minolta CS-2000 Spectroradiometer [40]: measured brightness (luminance) of 105 cd/m^2 and measured ambient contrast ratio (ACR) [41] of 160:1. The projector was mounted 230cm from the screen, the projected area measured 230cm x 150cm, raised 70cm from the floor. The view point was set at 130cm form the screen with a fixed sitting position 44cm high.

6) Procedure: The experiment was conducted at the Norwegian Colour and Visual Computing Laboratory in Gjøvik, within spaces with controlled illumination. The experiment required around twenty-five minutes for each participant.

The participants were expected to perform a series of mathematical tasks. The details are in the Table I. In each "briefing" segment the participants received verbal instruction to perform the task that followed. The timestamp of each step was recorded.

Rest (baseline)	1 min
Briefing	~10sec
Count up 0 to 60	~30 sec
Recover	1 min
Briefing	~10sec
Count down 60 to 0	~30 sec
Recover	1 min
Briefing	~10sec
Count down 91 to 0 every 4	~1.5 min
Recover	1 min
Briefing	~30sec
Fibonacci sequence to over 100	~1.5 min

TABLE I

TASK SEQUENCE.

7) Data Analysis: The eye-tracking data, videos and luminance logging have been processed as described in the implementation section (figure 2) to obtain the estimated CW (Δ PD). A sample of the CW data is visible in figure 3: the expected PD (blue) is removed from the Measured PD (black) to obtain the residual Δ PD (red) that is assumed to be CW plus noise.

General Linear Model (GLM) Repeated Measures has been used to process the data (variance of the results of a repeated measurement for each subject): the within-subjects factor is the increasing difficulty or task manipulation while the light condition determines the between-subjects factor (variable or fixed) differentiating the two test groups.

IV. RESULTS AND DISCUSSION

The Δ PD is highly influenced by CW manipulation (Task) (F=63.021, $p < 0.01$). Fixed Light and Variable Light conditions present the same reaction of the pupils (Δ PD) to the CW manipulation (Task) with no significant difference between the two conditions; see figure 4. The 0mm point is set as the average pupil change induced by CW and is specific to this series of tasks. The Δ PD indicates a change in CW between tasks and not an absolute value of CW.

It was found that the results correlate well with cognitive workload but not with luminance variation. Therefore the data processing was deemed good enough.
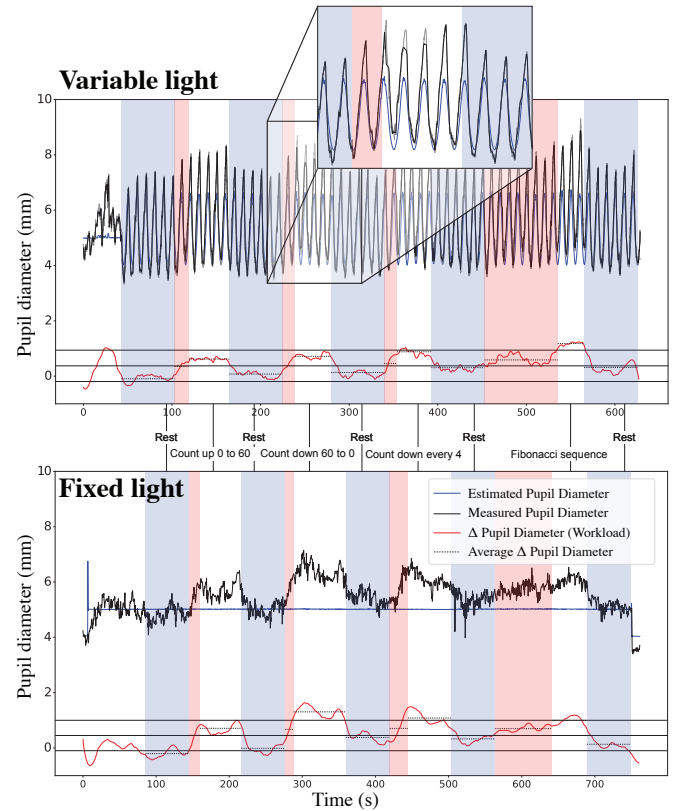

Fig. 3. The figure represents the output data from two participants (top "variable light" and bottom "fixed light"). The main output is the Δ PD (variation of cognitive workload, in red) superimposed to the plus and minus one standard deviation lines.
The Δ PD is obtained, removing the expected PD (blue) from the Measured PD (black). The different background colour indicates the experiment steps (light blue for rest, red for briefing).

It was found that the algorithm isolated the effects of varying light conditions on PD. Therefore the methodology (algorithm and hardware implementation) works well to identify changes in the PD exclusively due to CW, the measure of interest.

It was found that the luminance measurement, taken from the POV of a subject using a small video camera, worked sufficiently well once paired with an external luminance sensor. It was found that the on-board scene camera mounted on the Pupil Labs ETGs did not offer the necessary access to low-level information (exposure) to be used independently for measuring luminance from the POV of a subject.
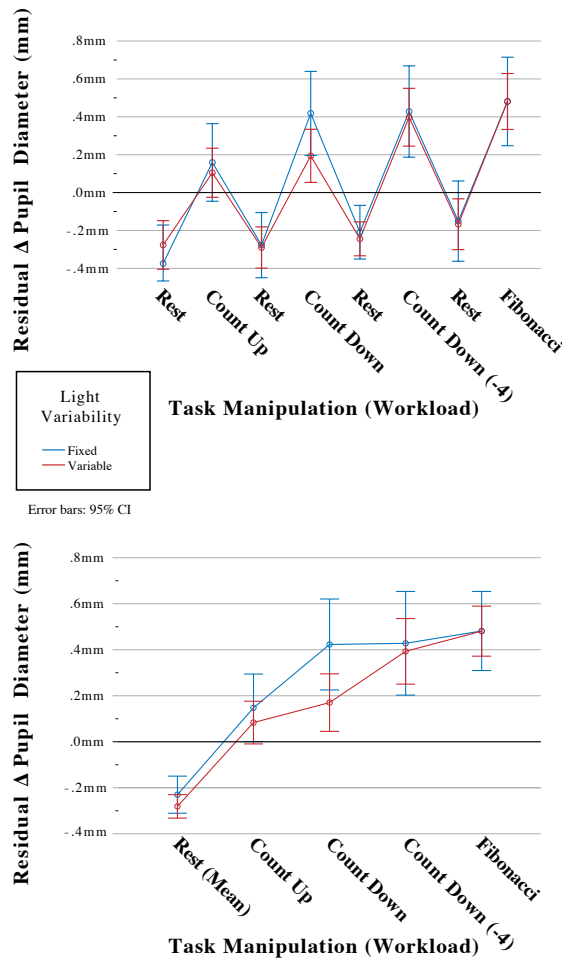
Fig. 4. GLM Repeated Measures data output generated by IBM SPSS [42]. The vertical scale represents the average Δ PD across all the participants. The top plot represents the change in CW with the same sequence experienced by the precipitants (alternating tasks and rest conditions). In the bottom plot, the rest conditions have been aggregated and averaged to show the almost linear increase of CW across the experiment.

Further experimentation, first with a standalone luminance sensor and then combining the sensor with the ETGs camera provided usable data; this has been explored with a variety of empirical tests but not tested in a controlled manner except for the stimuli (presented in a control fashion) described in this paper. The resulting system is still limited by the characteristics of a small camera (distortion, limited dynamic range and vignetting). Yet, it has proven sufficiently precise and reliable to be the foundation of this CW measurement.

It was found that, with the method described in the paper, it was possible to isolate the effect of CW, even in the presence of variable luminance such as one could expect in field conditions.

The Δ PD, attributed to the influence of the variable CW during the experiment, is not significantly affected by the variable visual stimuli. This indicates that the algorithm successfully removed the effect of such stimuli and that Δ PD can be used as a CW measurement in uncontrolled visual

conditions.

A. Limitations

The sample size is small, it would have been preferable to have at least thirty female and thirty male subjects for a total of sixty subjects [43].

There are some inherent limitations of the commercial eye-tracker which we have not separately tried to resolve, although we added other hardware for isolating the PD exclusive to the CW. This may mean that other unknown confounders may impact the measurement of CW, (e.g. Variations in the eye appearance, such as a pronounced "Epicanthic Fold" or a lower contrast between the iris and pupil can significantly reduce the quality of the eye recognition).

The luminance in the laboratory was highly controlled: a single light source projecting in front of the participant, a fixed sitting position and the subjects were instructed to keep their head position. Still, it was not possible to guarantee no movements in the participant's head. Therefore the POV could shift during the experiment, and this can be a potential confounder.

The luminance range was limited by the projector output; as such, it cannot reproduce the range of variability that could be experienced in a field condition. It is therefore unclear how extreme levels of luminance (low or high) can affect the measurement of CW (i.e. if a ceiling effect is present). Only one variation pattern for light conditions was used during the tests; this may not represent all possible states of light conditions one may experience in the field and potentially be one more confounder as well as the possible effect of fluctuating light itself on CW and the eye behaviour.

B. Conclusion

1) Future Work: Future technical development of the system would benefit from the ability to access the exposure data of the camera over time, possibly eliminating the need for an external sensor as well as switching to a different camera model, with better specifications, especially for what concerns dynamic range.

The advantages of processing the POV video have so far been addressed only empirically by comparing different methods on a selection of recording, a systematic test of this part of the system, including not only varying luminance but also varying size/shape and position of the target would therefore be a natural evolution of this study.

Furthermore, it would have been of interest to compare the newly developed CW metric to other objective and validated metrics; something that was unfortunately not possible in this case.

2) Resources: The last version of the software used in this experiment is published as a GitHub Repository `https://github.com/pignoniG/cognitive_analysis_tool` including extensive documentation to get started, assemble the luminance sensor, and analyse the data.

We hope this work will encourage a pragmatic use of technology in which qualitative and quantitative data are used together to represent the complex relations between humans

and interfaces, without being restricted to the artificial boundaries of the laboratory condition.

REFERENCES

[1] Pignoni, Giovanni, Hareide, Odd Sveinung, Komandur, Sashidharan, and Volden, Frode, "Trial application of pupillometry for a maritime usability study in field conditions," *Necesse*, no. 4, Nov. 2019.

[2] O. S. Hareide and R. Ostnes, "Maritime Usability Study by Analysing Eye Tracking Data," *Journal of Navigation*, vol. 70, no. 05, pp. 927–943, Sep. 2017.

[3] Sathiya Kumar Renganayagalu, Sashidharan Komandur, and Robert Rylander, "Maritime simulator training: Eye-trackers to improve training experience," *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014*, Jul. 2014.

[4] O. S. Hareide, "The use of Eye Tracking Technology in Maritime High- Speed Craft Navigation," Thesis for the Degree of Philosophiae Doctor, Norwegian University of Science and Technology Faculty of Engineering, Trondheim, Apr. 2019.

[5] Eric Farmer, Adam Brownson, and QinetiQ, "Review of Workload Measurement, Analysis and Interpretation Methods," EUROPEAN ORGANISATION FOR THE SAFETY OF AIR NAVIGATION, Tech. Rep., 2003, type: dataset.

[6] B. Cain, "A Review of the Mental Workload Literature," NATO RTO, Tech. Rep., 2007.

[7] E. H. Hess and J. M. Polt, "Pupil Size in Relation to Mental Activity during Simple Problem-Solving," *Science*, vol. 143, no. 3611, pp. 1190–1192, Mar. 1964.

[8] S. Mathôt, "Pupillometry: Psychology, Physiology, and Function," *Journal of Cognition*, vol. 1, no. 1, p. 16, Feb. 2018.

[9] D. Kahneman and J. Beatty, "Pupil Diameter and Load on Memory," *Science*, vol. 154, no. 3756, pp. 1583–1585, Dec. 1966.

[10] D. Kahneman, J. Beatty, and I. Pollack, "Perceptual Deficit during a Mental Task," *Science*, vol. 157, no. 3785, pp. 218–219, Jul. 1967.

[11] S. P. Marshall, "Method and apparatus for eye tracking and monitoring pupil dilation to evaluate cognitive activity," USA Patent US6 090 051A, Mar., 1999.

[12] A. T. Duchowski, K. Krejtz, I. Krejtz, C. Biele, A. Niedzielska, P. Kiefer, M. Raubal, and I. Giannopoulos, "The Index of Pupillary Activity: Measuring Cognitive Load *vis-à-vis* Task Difficulty with Pupil Oscillation," in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. Montreal QC, Canada: ACM Press, 2018, pp. 1–13.

[13] S. M. Wierda, H. van Rijn, N. A. Taatgen, and S. Martens, "Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution," *Proceedings of the National Academy of Sciences*, vol. 109, no. 22, pp. 8456–8460, May 2012.

[14] U. N. R. Council, Ed., *Tactical display for soldiers: human factors considerations*. Washington, D.C: National Academy Press, 1997.

[15] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology*. Elsevier, 1988, vol. 52, pp. 139–183.

[16] L. Di Stasi, M. Marchitto, A. Antolí, and J. Cañas, "Saccadic peak velocity as an alternative index of operator attention: A short review," *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, vol. 63, no. 6, pp. 335–343, Nov. 2013.

[17] C. F. Rusnock and B. J. Borghetti, "Workload profiles: A continuous measure of mental workload," *International Journal of Industrial Ergonomics*, vol. 63, pp. 49–64, Jan. 2018.

[18] P. S. Tsang and V. L. Velazquez, "Diagnosticity and multidimensional subjective workload ratings," *Ergonomics*, vol. 39, no. 3, pp. 358–381, Mar. 1996.

[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] J. T. Cacioppo, L. G. Tassinary, and G. Berntson, Eds., *Handbook of Psychophysiology*, 3rd ed. Cambridge: Cambridge University Press, 2007.

[21] P. Binda and S. O. Murray, "Spatial attention increases the pupillary response to light changes," *Journal of Vision*, vol. 15, no. 2, pp. 1–1, Feb. 2015.

[22] S. R. Steinhauer, G. J. Siegle, R. Condray, and M. Pless, "Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing," *International Journal of Psychophysiology*, vol. 52, no. 1, pp. 77–86, Mar. 2004.

[23] E. H. Hess and J. M. Polt, "Pupil Size as Related to Interest Value of Visual Stimuli," *Science*, vol. 132, no. 3423, pp. 349–350, Aug. 1960.

[24] E. H. Hess, A. L. Seltzer, and J. M. Shlien, "Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study." *Journal of Abnormal Psychology*, vol. 70, no. 3, pp. 165–168, 1965.

[25] S. T. Iqbal, X. S. Zheng, and B. P. Bailey, "Task-evoked pupillary response to mental workload in human-computer interaction," in *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*. Vienna, Austria: ACM Press, 2004, p. 1477.

[26] B. P. Bailey and S. T. Iqbal, "Understanding changes in mental workload during execution of goal-directed tasks and its application for interruption management," *ACM Transactions on Computer-Human Interaction*, vol. 14, no. 4, pp. 1–28, Jan. 2008.

[27] M. A. Recarte and L. M. Nunes, "Effects of verbal and spatial-imagery tasks on eye fixations while driving." *Journal of Experimental Psychology: Applied*, vol. 6, no. 1, pp. 31–43, 2000.

[28] O. Palinko and A. Kun, "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*. Olympic Valley-Lake Tahoe, California, USA>: University of Iowa, 2011, pp. 329–336.

[29] A. B. Watson and J. I. Yellott, "A unified formula for light-adapted pupil size," *Journal of Vision*, vol. 12, no. 10, pp. 12–12, Sep. 2012.

[30] P. D. Hiscocks and P. Eng, "Measuring Luminance with a Digital Camera," *Syscomp Electronic Design Limited*, p. 27, 2014.

[31] D. Wüller and H. Gabele, "The usage of digital cameras as luminance meters," R. A. Martin, J. M. DiCarlo, and N. Sampat, Eds., San Jose, CA, USA, Feb. 2007, p. 65020U.

[32] M. Kassner, W. Patera, and A. Bulling, "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*, ser. UbiComp '14 Adjunct. New York, NY, USA: ACM, 2014, pp. 1151–1160, event-place: Seattle, Washington.

[33] AMS, *TSL25911_Datasheet_EN_v1.pdf ams163.5*. AMS, Apr. 2013.

[34] W3C, *G17: Ensuring that a contrast ratio of at least 7:1 exists between text (and images of text) and background behind the text*. W3C, Mar. 2018.

[35] G. Pignoni, "Effects of ambient illumination on text recognition for UI development," in *2018 Colour and Visual Computing Symposium (CVCS)*. Gjøvik: IEEE, Sep. 2018, pp. 1–6.

[36] A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, Jul. 1964.

[37] Giovanni Pignoni, "Cognitive Analysis Tool," 2019. [Online]. Available: https://github.com/pignoniG/cognitive_analysis_tool

[38] J. W. Peirce, "PsychoPy—Psychophysics software in Python," *Journal of Neuroscience Methods*, vol. 162, no. 1-2, pp. 8–13, 2007.

[39] J. Peirce, *PsychoPy*. Jonathan Peirce, Mar. 2018.

[40] K. M. S. S. P. Ltd, *CS-2000 SPECTRORADIOMETER*. Konica Minolta Sensing Singapore Pte Ltd, Mar. 2018.

[41] H. Chen, G. Tan, and S.-T. Wu, "Ambient contrast ratio of LCDs and OLED displays," *Optics Express*, vol. 25, no. 26, p. 33643, 2017.

[42] C. IBM, "IBM SPSS Statistics," NY, 2017.

[43] R. V. Hogg and E. A. Tanis, *Probability and statistical inference*. New Delhi: Pearson Eduction, 2006, oCLC: 818812168.