



Aalborg Universitet

AALBORG UNIVERSITY  
DENMARK

## Reliability of oculometrics during a mentally demanding task in young and old adults

Marandi, Ramtin Zargari; Madeleine, Pascal; Omland, Øyvind; Vuillerme, Nicolas; Samani, Afshin

*Published in:*  
IEEE Access

*DOI (link to publication from Publisher):*  
[10.1109/ACCESS.2018.2819211](https://doi.org/10.1109/ACCESS.2018.2819211)

*Creative Commons License*  
Unspecified

*Publication date:*  
2018

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

*Citation for published version (APA):*  
Marandi, R. Z., Madeleine, P., Omland, Ø., Vuillerme, N., & Samani, A. (2018). Reliability of oculometrics during a mentally demanding task in young and old adults. *IEEE Access*, 6, 17500-17517. Advance online publication. <https://doi.org/10.1109/ACCESS.2018.2819211>

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

### Take down policy

If you believe that this document breaches copyright please contact us at [vbn@aub.aau.dk](mailto:vbn@aub.aau.dk) providing details, and we will remove access to the work immediately and investigate your claim.

Received January 15, 2018, accepted March 14, 2018, date of publication March 26, 2018, date of current version April 23, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2819211

# Reliability of Oculometrics During a Mentally Demanding Task in Young and Old Adults

RAMTIN ZARGARI MARANDI<sup>1,2</sup>, PASCAL MADELEINE<sup>1</sup>, ØYVIND OMLAND<sup>4</sup>,  
NICOLAS VUILLERME<sup>1,2,3</sup>, AND AFSHIN SAMANI<sup>1</sup>

<sup>1</sup>Department of Health Science and Technology, Sport Sciences, Aalborg University, 9220 Aalborg, Denmark

<sup>2</sup>Université Grenoble Alpes, 38400 Grenoble, France

<sup>3</sup>Institut Universitaire de France, 75231 Paris, France

<sup>4</sup>Aalborg University Hospital, Clinic of Occupational Medicine, Danish Ramazzini Center, 9000 Aalborg, Denmark

Corresponding author: Afshin Samani (afsamani@hst.aau.dk)

This work was supported by Veluxfonden under Project 00010912.

**ABSTRACT** The sensitivity of oculometrics to the changes in mental load has already been investigated in several studies. However, the reliability of oculometrics remains unexplored, particularly under concurrent mental demands. To address this, we recruited 20 young and 18 elderly healthy adults to perform a functional computer task with three levels of mental load (i.e., low, medium, and high) on two days interspaced by at least seven days. We tracked the participants' eye movements and pupil size during the task and computed the characterizing features of saccades, fixations, blinks, pupillary responses as well as nonlinear dynamics of pupillary responses and gaze trajectories. In addition, we recorded the trace of the computer mouse for performance measurement and acquired subjective ratings of the perceived task load. Among the investigated oculometrics, saccadic peak velocity (SPV) and its rate of changes across saccade amplitude (SVA) were influenced by the level of the task load in both the young and elderly adults, and the effects remained consistent across days. Reliability assessments revealed good to excellent test-retest reliability and acceptable absolute reliability in SPV, SVA, and the duration of fixations and saccades. The perceived load and task performance were monotonically altered with the task load levels. These findings may provide practical implications on mental load quantification in occupational fields.

**INDEX TERMS** Ergonomics, gaze tracking, human computer interaction, human factors.

## I. INTRODUCTION

Mental load is a multifaceted and multidimensional concept with no consensus on its definition. However, according to multiple resource theory [1], it can be referred to the required level of mental resources to meet the performance criteria of a task. Mental load may be mediated by task demands, external support, and past experience [2]. According to the multiple resource theory mental load can be used interchangeably by the term cognitive load concerning induction and measurement [3]. Sustained mental demands of a task may lead to increase in human errors [4].

The tendency of the occupational paradigm in modern societies to mentally demanding jobs along with the population aging calls for technology development concerning mental health risks and productivity [5]. Mental load has an intensified impact on elderlies at work as a decline in cognitive capacity has been reported in the geriatrics literature [6], [7].

Since almost all kinds of human activities involve mental processing, the modelling and quantification of mental load have become important research topics with different aims, e.g., education, mental health, and user interface design, within various interconnected disciplines, e.g., ergonomics, psychology, neuroscience, and industrial engineering [8]–[10]. For more than three decades, this endeavor has led to find psychophysiological metrics for quantification of mental load [11]–[17].

Among psychophysiological measurements, ocular metrics have drawn more attention not only for the ease of access, but also for the extensive neural circuitry from different areas inside the brain including the superior colliculus, paramedian pontine reticular formation, premotor nuclei, pretectal nuclei, medial longitudinal fasciculus, and nucleus prepositus hypoglossi, which are involved in controlling the eye movements and cognitive processing [18]–[22]. In light of these findings, many studies on pervasive computing and

human computer interaction have attempted to understand the human behavior based on the interconnection between eye movements and mental processes [23]–[27].

Ocular events can be categorized mainly into saccade, fixation, blink, and pupillary response [28]. Saccades are sudden movements of the eyes to salient areas in the visual field. Fixations are the stillness of the eyes mainly to intake visual information. Blinks are closing and opening of the eyelids mainly to clean and lubricate the surface of the cornea and conjunctiva. Pupillary responses refer to changes in the pupil size mainly to control the input amount of light. These ocular events could also be modulated by mental load variation. For instance, the velocity of saccade was shown in [29] to be correlated with mental load variation. The established interconnection between mental load and eye movements encouraged [30] to develop software to measure mental load based on physiological measurements, e.g., pupil dilations, together with subjective ratings and performance measures in experimental settings. More studies regarding the relationships between ocular events and mental load variation can be found in a review conducted by [31]. Furthermore, the way these links are being applied in real life situations, e.g., [32], is technically promising.

Previous studies on oculomotor behavior often apply commonly known metrics to quantify the characteristics of ocular events. However, novel mathematical techniques provide new insight into the dynamics of eye movements. The characteristics of nonlinear dynamics have provided promising mathematical tools to unravel novel aspects in many physiological time series [33], [34], including pupillary responses [35] and eye gaze trajectory [36]–[38].

An extensive body of research concerning the detection of changes in mental load via eye-tracking technology has addressed mainly the sensitivity of the oculometrics regarding different applications, e.g., in surgery [39]–[41], air-traffic control [42], [43], driving [44], and learning [45]. However, the reliability analysis of the oculomotor system has been less explored. This calls for further studies covering this gap. Table 1 provides a summary of the studies addressing the reliability of oculometrics.

According to most of the studies summarized in Table 1, oculomotor assessment has a potential applicability for clinical evaluations to serve as a cognitive marker [52], [53]. Most of the studies in Table 1 assessed the reliability of oculometrics during standardized tasks, e.g., to diagnose deficits in oculomotor control caused by neurodegenerative diseases [54].

However, there is no equivocality in the results from different studies according to Table 1, e.g., on the reliability of saccade peak velocity. In [56] the effect of visual corrections was stressed, e.g., eye glasses to explain why the oculometrics were less reliable. The effect of age was also emphasized to account for different levels of reliabilities in oculometrics [46].

The effects of age on oculometrics are a fundamental question among research communities. In [57] the effects of age

on oculometrics was studied by recruiting 34 healthy participants, i.e., 17 young and 17 old adults aged 30 and 62 on average, respectively, to perform free viewing and visual tracking tasks. The results revealed that the frequency, amplitude, and peak velocity of saccades were altered by age. In addition, in [58] it was found that the cortical activation levels in old adults (aged 51-70) were higher compared with young adults (aged 21-42) during a pro-saccade task. Saccadic latency and smooth pursuit have been posited to be dependent on age in [59]–[61], whereas age-based changes in other oculometrics, such as saccade peak velocity, have not yet been conclusive [18], [61]. Further, given the demographic changes resulting in more elderly computer workers, the potential age-based difference in oculometrics [62] suggests studying human computer interaction in such cognitive states.

As previously mentioned, many studies have affirmed the relationship between changes in mental load and oculometrics. Based on this, our first hypothesis was that the change in mental load can be reflected in eye movement dynamics during a task with concurrent mental demands. To investigate this hypothesis, we examined the sensitivity of oculometrics to varied levels of mental load during standardized computer work.

Most of the abovementioned studies acknowledged the reliability of oculometrics. Accordingly, our second hypothesis was that the variations in oculometrics due to changes in mental load remain consistent across two separate experimental days. We employed relative and absolute reliability concepts to provide a comprehensive assessment of the reliability of oculometrics.

In addition, we aimed at investigating a probable interactive effect of age on the sensitivity and reliability of the oculometrics. Thus, we recruited young and old adults to perform our experiment. This enabled us to analyze the sensitivity and reliability of oculometrics separately for each age group. Taking all this into account, this study proposes an experimental framework on the reliability analysis of oculometrics in response to mental load variation in a functional task with concurrent mental demands in both young and old adults.

This paper is organized as follows: Section II presents the methodology of this work, including task description, experimental procedure, and data analysis. In Section III, the results regarding the sensitivity and reliability of the oculometrics at different mental load levels and on different days are described. In Section IV, the results provided in Section III are discussed in detail to provide psychophysiological and clinical interpretations. In section V, final remarks of the current study are provided.

## II. MATERIALS AND METHODS

### A. PARTICIPANTS

The number of participants was determined using the  $\alpha$  level set to 0.05 and the  $\beta$  level to 0.20,  $\rho_0 = 0.4$ ,  $\rho_1 = 0.8$ , and  $n = 2$  [63].  $\rho_0$ ,  $\rho_1$ , and  $n$  denote the minimally acceptable

**TABLE 1. Overview of the studies on the reliability of oculometrics.**

Study (Authors, year)	Participants	Interval (days)	Task	Relevant Oculometrics	Reliability	Interpretation
Bargary et al., 2017 [49]	100 healthy (22±4 y)	19 (median)	Pro- and anti-saccade	Main sequence	ICC <sup>*</sup> = 0.88 R <sup>**</sup> = 0.88	Highly reliable
Vikesdal & Langaas, 2016 [50]	12 healthy (27±6 y)	1-49	Visually guided saccade	Saccade latency	ICC = 0.95 C-α <sup>***</sup> = 0.94	High reliability and good consistency
			Fixations in visually guided saccade	Fixation stability	ICC = 0.62 C-α = 0.75	
Ettinger et al., 2003 [48]	21 healthy (19-44 y)	57 (mean)	smooth pursuit, fixation, pro- and anti-saccade	Pro-saccade latency and gain, and saccade frequency	ICC>0.50 C-α>0.80	High consistency and good reliability
Roy-Byrne et al., 1995 [51]	8 healthy (18-50 y)	28	Fixation,	Fixation frequency	ICC>0.71	Good reliability
			Visually guided saccade,	Saccade latency and velocity	ICC>0.61	
			memory-guided saccade	Saccade latency and velocity	ICC>0.69	
			saccade to moving target	Saccade latency	ICC>0.67	
Meyhöfer et al., 2016 [52]	68 healthy (24±5 y)	7	pro-saccade anti-saccade memory-guided saccade	SPV <sup>1</sup> SPV SPV	ICC = 0.91 ICC = 0.83 ICC = 0.91	Highly reliable
Turner et al., 2017 [53]	65 PD <sup>4</sup> (66±8 y)	30 (mean)	Volitional saccade test	volitional saccade latency	R= 0.70	Good reliability
Versino et al., 1993 [47]	20 healthy (23-59 y)	7	Visually guided saccades	SPV-SA <sup>2</sup> relationship SPV-SCD <sup>3</sup> relationship	ICC ≈ 0.91	Fairly good Excellent
Klein & Fischer, 2005 [46]	327 healthy (9-88 y)	510-630	Pro- and anti-saccade	saccadic reaction times	0.60 ≤ r <sup>+</sup> ≤ 0.97	Good to excellent reliability
	117 healthy (6-18 y)				0.43 ≤ R ≤ 0.66	moderately reliable
Bollen et al., 1993 [55]	58 healthy (51±19 y)	14	Visually guided saccades	SA SPV Main sequence	ICC = 0.19 ICC = 0.10 ICC = 0.34	Low consistency
Stuart et al., 2016 [56]	20 healthy (>50 y), and 14 PD	7	Visually guided saccades while sitting, walking and standing	SA	0.14 ≤ ρ <sup>++</sup> ≤ 0.85	Poor to good relative agreement

<sup>\*</sup>Intraclass Correlation Coefficient (ICC), <sup>\*\*</sup>Pearson's correlation coefficient (R), <sup>\*\*\*</sup>Cronbach's alpha (C-α), <sup>+</sup>Odd-even and split-half correlation coefficient (r), <sup>++</sup>Spearman's correlation coefficient (ρ), <sup>1</sup>Saccade peak velocity (SPV), <sup>2</sup>Saccade amplitude (SA), <sup>3</sup>Saccade duration (SCD), <sup>4</sup>Parkinson Disease (PD).

level of reliability, the expected level of reliability, and the number of test-retest trials, respectively. The α and β levels denote the probability of making a type 1 and type 2 error, respectively. Thus, assuming a drop-out rate of 10%, 18 participants in each age group were required to run this study.

Two groups of subjects were recruited for the study. A young group of twenty participants (nine females, 11 males; age (yrs) M: 23, SD: 3; height (m) M: 1.74,

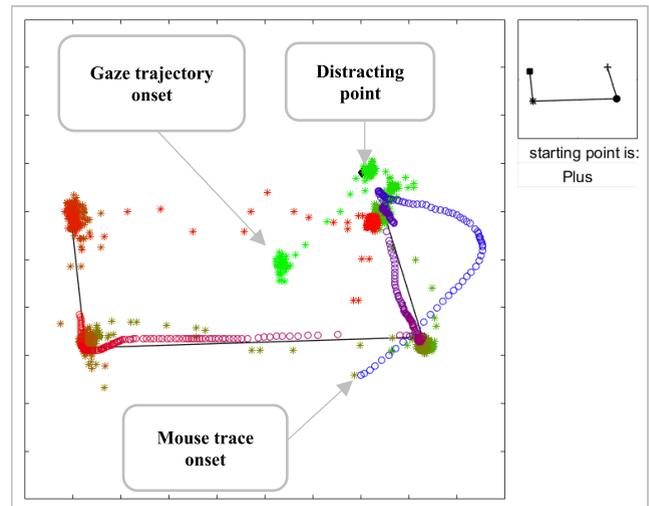
SD: 0.08; body mass (kg) M: 71, SD: 11; and body mass index (kg/m<sup>2</sup>) M: 23.3, SD: 3.3) and an elderly group of eighteen participants (11 females, seven males; age (yrs) M: 58, SD: 7; height (m) M: 1.72, SD: .07; body mass (kg) M: 80, SD: 12; and body mass index (kg/m<sup>2</sup>) M: 26.9, SD: 4.2 kg/m<sup>2</sup> of body mass index). Due to poor gaze recording quality (see Data Recording and Analysis), we had to exclude two participants in the young group from the study. All participants

had normal or corrected-to-normal vision (self-reported and examined by Snellen chart eye acuity test). The participants were all familiar to computer work and used their right hand to work with the computer mouse. The handedness was also confirmed by the Edinburgh Handedness Inventory [64] self-assessment measure, i.e., a mean laterality index (LI) of 74. The participants were asked to abstain from alcohol for 24 h and caffeine for 12 h prior to each experimental session. The participants reported a habitual 6–10 h of sleep per night and slept at least 6 h (mean 7.7 h) before the experimental sessions. The participants were also asked to refrain from smoking or taking any drugs with potential effect on the cognitive performance [65] or eye movements [66] 12 h before each experimental session. Written informed consent was obtained from each participant. The experiment was approved by The North Denmark Region Committee on Health Research Ethics, project number N-20160023, and conducted in accordance with the Declaration of Helsinki.

### B. THE EXPERIMENTAL TASK: FUNCTIONAL COMPUTER WORK

We customized a graphical user interface (GUI) (WAME 1.0) in a MATLAB R2015b environment to implement the computer task framework. The task was designed based on a standard model of computer work which has been used in previous studies [67]–[69] and involved cyclic computer operations in which each cycle was organized in sequential periods, namely memorization (MP), washout (WP), and replication periods (RP). The work panel contained two sections, i.e., a template panel with a textual instruction and a replication panel (Supplementary Materials, Fig 1). The GUI was displayed on a 19-in screen (1280×1024 pixels, refresh rate: 120Hz) located approx. 56 cm in front of the participant. The GUI subtended approx. 27° of the visual angle horizontally and approx. 22° vertically. The replication and template panel subtended approx. 20° and 5° of the visual angle in both directions, respectively. The size of the GUI was fixed and it was positioned in the center of the computer screen. The computer screen height was adjusted such that the ear-eye line was approx. 15 degrees below the horizon when the participants sat upright and stared at the center of the screen [70].

Each cycle began by showing a specific pattern composed of a sequence of a few points connected to each other by lines on the template panel. Each point in the pattern was depicted in a different shape, i.e., plus, circle, asterisk, square, diamond, pentagram, and triangle. A short text indicating the shape of the starting point was displayed during the MP when the participant had to memorize the pattern and the starting point (Supplementary Data, Fig 1 (a)). Then the pattern disappeared during WP, with the same duration as the MP, and the participant was asked to fixate on a cross subtended 2° of the visual angle displayed in the center of the replication panel to avoid irrelevant saccades during WP (Supplementary Data, Fig 1 (b)). To avoid any repositioning of the mouse cursor and visual distraction of the participant's attention from the fixation point, the mouse cursor was invisible during the WP.



**FIGURE 1.** A typical example of the gaze trajectory, indicated by “\*”, together with computer mouse trace, indicated by “o”, during the replication period (RP) of a cycle with low mental load. The pattern and the textual cue were not shown during RP; they are depicted here to make the example understandable. The gaze trajectory (visual scanpath) and computer mouse trace are depicted on a color-coded scale from green to red and from blue to red, respectively, to show the time progression. At the beginning of RP, it seems that the participant was visually distracted by the distracting point which was located close to the starting point. The participant then continued by visual scanning all the pattern points and clicking successfully on the first three pattern points.

Subsequently, a scaled version of the pattern points without the connecting lines was displayed on the replication panel, and the participant had a limited time, i.e., RP, to connect the points, in order, to replicate the pattern previously shown during MP (Supplementary Data, Fig 1 (c)). The replication was performed by clicking on the points using a computer mouse in a sequential manner preserving the order of connected points in each pattern.

To indicate that the first click was correctly performed, the displayed point was enlarged by a factor of two. The participant continued by clicking on the next points and whenever he or she clicked on a correct point, a line was drawn connecting the point to the previous one; otherwise, no line appeared. To vary the mental load, specific characteristics of the displayed patterns were manipulated. As such, the number of points and the geometrical complexity of the patterns were changed to induce three levels of mental load, i.e., low, medium and high.

The patterns were picked from an inventory of randomly generated patterns of connected points subjected to two constraints, i.e., the number of points and the relative positioning of the points. The number of points in the patterns was chosen to be four, five and six for low, medium and high mental load levels, respectively. The points were arranged such that the length of the connecting lines was limited within four to eight degrees of visual angle and the pattern was displayed around the center of the panel not to influence the eye movements by the geometrical placement of the points [62]. In order to change the complexity of the patterns, the regularity and

spikiness of the patterns were changed such that the angles between the connecting lines were tightened with increased mental load in accordance with previous studies [71], [72] (Supplementary Data, Fig 1 (d)).

Even though randomly generated, the patterns were kept identical across participants to avoid differences in participant's eye movements due to different patterns. Each participant was assigned a certain time, based on the Methods-Time Measurement (MTM) approach [73], to finish each pattern, i.e., 2.06, 2.34, and 2.62 s for MP (or WP), and 4.11, 5.06, and 6.02 s for RP in low, medium, and high levels of mental load, respectively. A new pattern was displayed on the template panel when the specified time had elapsed. In addition to the points composing the pattern, a distracting point in a different shape from the constituent points appeared on the replication panel. This point had to be ignored by the participants.

### C. EXPERIMENTAL PROCEDURE

The participants took part in two experimental sessions with at least 7 days between sessions to assess the reliability of metrics characterizing ocular events across days. All experimental sessions were conducted between 09.00 and 12.00 h (noon) and 13.00 and 15.00 h to minimize the potential influence of circadian rhythm or diurnal variation. The participants were instructed about the task and practiced it with both low and high load levels until they became accustomed to the task. A 10-min rest was given after the task training. Then, the participants performed three consecutive 5-min task segments with low, medium, and high mental load separated by 1-min rest intervals. The sequence of the task segments was counter-balanced across the participants following a three by three Latin square design. The same order of the sequence was used in the second experimental session. The participants also had to indicate their subjective perception of the mental load by answering the National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaire [64] after the termination of each task segment. On the second day of the experiment, the participants underwent the same procedure without filling in the introductory self-assessment questionnaires.

### D. DATA RECORDING AND ANALYSIS

#### 1) APPARATUS

A video-based monocular eye tracker (Eye-Trac 7, Applied Science Laboratories, Bedford, MA, USA) was utilized to track eye movements at a sampling frequency of 360 Hz during the experimental sessions. Head-mounted systems are known to be the most accurate and versatile eye tracking solutions [28]. Head movements were also tracked in 3D using a motion tracker (Visualeyez II system set up with two VZ4000 trackers, Phoenix Technologies Inc., Canada). The data from the eye tracker and motion tracker were coupled to compensate for head movements and to precisely estimate the point of gaze (built-in functionality of Eye-Trac 7). An example of the trajectory of the eye gaze on the screen

while performing the task is depicted in Fig 1. The calibration of the eye tracker was performed before starting the tasks with a 9-point calibration protocol. The experiments were conducted in an indoor and controlled environment to rule out potential confounding effects due to changes in ambient illumination.

#### 2) PARSING OCULAR EVENTS

We parsed the timeline of each task segment to relevant ocular events, namely, saccades, blinks, and fixations. We employed an adaptive data-driven algorithm [74] to extract ocular events across the entire timeline in each task segment (see Appendix A). Following the steps of the algorithm, the visual angle between consecutive coordinates of the point of gaze through the entire timeline of each task segment was first calculated and then the angular velocity and acceleration of the calculated visual angle were derived following the Savitzky–Golay method [75]. Then, the saccades were extracted following the algorithm described in Appendix A.

Blinking episodes were determined when the pupil was not detected for a 50-700 ms duration [76], [77]. The blinking episodes were also examined by checking the eyelid closure on the eye video frames (30 frames/s). Noise samples were identified and excluded as those with the absence of pupil or corneal reflection without meeting other criteria for an actual blink or a saccade as outlined in appendix A. Excluding the instances of saccades, blinks, and noise samples, the remaining part of the gaze coordinates were considered as fixations if the duration of the detected fixation was  $>30$  ms [28]. Successive fixations were merged into one fixation where  $<5$  intra-fixational noise samples and  $<1^\circ$  of visual dispersion between of the preceding and succeeding fixations occurred [74], [78].

The pupil diameter was preprocessed as a function of time by linear interpolation to estimate the missing samples during blinks. Then, a zero-phase low-pass Butterworth filter of order 3 was applied to remove the noise and artifacts usually occurring prior or after each blink [79]–[81].

#### 3) OCULOMETRICS

Previous studies on task-evoked pupillary responses to mental load variations [79], [82]–[84] have suggested that a transient pupillary response emerges after an increase in the working memory demand. Accordingly, we defined the latency of task-evoked pupillary peak dilation (TPDL) as the time interval between the beginning of the WP in each cycle to the time of the maximum pupil dilation during WP.

Saccade amplitude (SA) ( $^\circ$ ), i.e. saccade magnitude or size, has been operationally defined by [28] as the total distance of the gaze trajectory between two fixation points. The SA was computed by multiplying the saccade duration (SCD) (s) and the average velocity of the movement ( $^\circ/s$ ). The SCD and SA are reported to be sensitive to mental load variation [85], [86]. The saccadic frequency (SF, Hz) or saccade rate reflecting task difficulty [87] was computed as the number of saccades divided by the period of a cycle in s.

Saccade peak velocity (SPV) ( $^{\circ}/s$ ) was computed as the maximum velocity during each saccade. An association between SPV and mental load has been reported in several studies [29], [88], [89]. The SPV has a stereotyped linear relationship with SA; a so-called main sequence [90]. This relationship is expressed as the slope of the line regressing SPV on SA, termed saccadic velocity-amplitude (SVA). The SVA has also been shown to be sensitive to changes in mental load [42], [91]. It is important to note that the linear relationship between SPV and SA holds up to  $20^{\circ}$  [92].

In addition to the aforementioned saccadic metrics, saccade acceleration asymmetry index (SAAI) was computed as the ratio of the difference between peak saccade acceleration and deceleration to their summation [93]. In addition, saccadic curvature (SCR), representing the overall spatial shape of a saccade, was studied since it has been reported to be affected by visual distractions [28], [94]. The SCR ( $^{\circ}$ ) was computed as the Euclidean distance of each gaze point during a saccade from the line connecting the saccade's onset and offset gaze points divided by the number of saccadic samples.

Fixation duration (FD) and fixation frequency (FF) were computed due to their relationships with task demands [87]. The FF (Hz), or fixation rate was computed as the number of fixations divided by a period of a cycle in s.

Blink frequency (BF) and blink duration (BD) were computed due to their sensitivity to mental load [79], [95], [96]. The BF, or blink rate, was computed as the number of blinks within a cycle divided by the duration of the cycle (s).

Based on the timestamps indicating the onset and offset of the task cycles, the detected ocular events were divided into their corresponding cycles. The cycles with less than one correct click or less than five saccades and five fixations (minimum requirement to perform the task) were deemed invalid cycles and excluded from further processing ( $6 \pm 12\%$  of the cycles for each participant and task load level).

As described below, within each valid cycle of the task, common methods to characterize the oculomotor behavior were applied and the derived oculometrics were averaged across the cycles. Additionally, nonlinear methods were applied to the point of gaze and the pupillary response time series within each cycle to investigate the oculomotor dynamics and its relationship to change in mental load.

#### 4) NONLINEAR DYNAMICS OF PUPILLARY RESPONSE

As a measure of complexity of a univariate time series, sample entropy, SaEn, was calculated for the pupillary response within each valid cycle [97]. To compute the SaEn, a one-dimensional time series,  $x(1), x(2), \dots, x(N)$ , has to be embedded into  $m$ -dimensional embedding vectors portraying the "embedded phase space", i.e.,  $X(i) = [x(i), x(i + \tau), \dots, x(i + (m - 1)\tau)]$  for  $i = 1, \dots, N - (m - 1)\tau$ ; where  $m$  and  $\tau$  are embedding dimension and time lag, respectively. For each embedding vector,  $X(i)$ ,  $C_i^m(r) = \{\text{number of } X(j) \text{ such that } \text{dist}(X(i), X(j)) \leq r \text{ and } i \neq j\}$  can be computed; where  $r$  is the tolerance, a positive real number, and  $\text{dist}$  is a distance function. Then, SaEn can

be calculated as  $\text{SaEn} = -\ln\left(\frac{\Phi^{m+1}(r)}{\Phi^m(r)}\right)$ ; where  $\Phi^m(r) = \frac{\sum_{i=1}^{N-m+1} C_i^m(r)}{N-m+1}$ . As suggested in [97], sample entropy of the pupillary response time series (PSaEn) was computed for each valid cycle with  $m = 2$ ,  $r$  as 0.2 times the SD, and then averaged. PSaEn is a unitless, non-negative value where low values indicate low complexity. The time lag is often set at a delay at which the autocorrelation function of the time series falls by as much as  $1/e$  [98]. The appropriate choice of time lag ensures that the reconstructed dimensions of the phase-space are relatively orthogonal and contain less confounding information regarding the temporal structure in a time series [99]. The auto-correlation function was computed for the entire task timeline and a proper time lag (22 ms) for the pupil signal was determined by taking the grand average of the computed time lags across the entire participant pool and the task segments.

Furthermore, the recurrence behavior of the pupillary response time series in the embedded phase space was quantified using recurrence quantification analysis (RQA) metrics. To compute RQA measures, a distance matrix (DM) was constructed of  $DM_{ij} = \text{dist}(X(i), X(j))$ ; where  $X(i)$  and  $X(j)$  were the embedding vectors as defined above and  $\text{dist}$  corresponded to the Euclidean distance. By comparing the elements of the DM with a constant tolerance threshold, recurring dynamical trajectories of a system were defined as those below the tolerance threshold. If  $DM_{ij}$  was smaller than the threshold, the corresponding element of the recurrence map (RM) was set to "1" (a recurrent point); otherwise, it was set to "0". The threshold was selected such that the percentage of recurrent points stayed within a range of 0.1–2% of the total number of RM elements [100].

The RQA allowed us to compute the following three metrics for the pupillary response:

- (1) The recurrence rate of pupillary response (RRP) as the percentage of recurrent points across the entire timeline of each cycle,
- (2) The determinism of pupillary response (DTP) as the percentage of recurrent points composing a diagonal line in RM, and
- (3) The recurrence map entropy of pupillary response (RMEP) as the Shannon entropy of the distribution of the diagonal lines which were associated with the determination of the pupillary response.

#### 5) NONLINEAR DYNAMICS OF GAZE TRAJECTORY

Since the point of gaze was composed of vertical and horizontal coordinates, it was treated as bivariate time series, and a multivariate version of sample entropy (MSaEn) was applied to analyze the point of gaze time series within each valid cycle. The embedding dimension  $m$  and the distance threshold  $r$  were determined to be 2 and 0.2 times the SD of each coordinate, respectively. The MSaEn estimates were consistent for the data length  $N > 300$  [101]. Time lags of 49 ms were applied based on the method of autocorrelation function [98].

Identical  $m$  and  $r$  were adopted to perform a multidimensional recurrence quantification analysis (MdRQA) [102].

In a manner similar to the extraction of nonlinear metrics from the pupillary response by means of univariate RQA, the following two metrics were extracted from the sequences of point of gaze using MdrQA: the percentage of recurrent points (RRG) and the percentage of diagonally adjacent recurrent points (DTG).

## 6) PERFORMANCE METRICS

We computed an overall performance (OP) to assess how successful each participant performed the tasks. The OP quantifies how accurate and fast the participant performed the task. Fig 1 gives an insight into the motor coordination between the eye and hand while performing the task.

The times elapsing in each cycle to correctly click on the pattern points were obtained to address the dexterity of the participants. A parameter called mean reaction time (MRT) was defined to account for the participant's clicking speed. In each cycle, MRT was computed for each cycle in three possible cases, in which all, some, or no pattern points were clicked correctly (1). If all the points in a pattern were clicked correctly, the time interval (TI) between correct clicks (CC) plus the first click time with respect to the cycle onset time were computed and averaged on the number of correct clicks in each cycle. In the second case, in which some of the points in a pattern were clicked correctly, the remaining time of the RP (RTRP), the time between the last correct click to the end of RP, was added to the summation of the interval between correct clicks. The duration of RP in a cycle without any correct clicks was calculated as the MRT of that cycle.

$$MRT = \begin{cases} \frac{\sum_{i=1}^{No. of CC} TI_i}{No. of CC}, & \text{Completed Pattern} \\ \frac{\sum_{i=1}^{No. of CC} TI_i + RTRP}{No. of CC + 1}, & \text{Partially completed pattern} \\ RP, & \text{No correct clicks} \end{cases} \quad (1)$$

The MRT was normalized by dividing it by the minimum of MRT across all participants (i.e., 0.5) to assign a maximum achievable performance. In addition to OP, the mouse click positions during RP were obtained to compute a selective attention (SelA) metric (2) in accordance with a concept introduced by [103] and [104]. SelA is defined as the ability to keep a set of actions in the face of distracting or competing stimuli; hence:

$$SelA = \frac{CC}{IC + PP + DC} \quad (2)$$

The SelA for each cycle can become one in the highest attentive case when all pattern points (PP) are correctly clicked and no incorrect clicks (IC) are performed, including clicks on the distracting point (DC) as penalty terms, and descended in case of low attentiveness. The OP for each cycle was defined as the ratio of the SelA and MRT. Thus, the OP theoretically is a positive value with zero for the lowest performance and one for the highest performance. The calculated OP was averaged across the cycles of the task.

## 7) SUBJECTIVE MEASURES

A computerized version of NASA-TLX [105] was utilized to assess the workload of the tasks based on six subscales: mental, physical and temporal demands, own performance, effort and frustration level. Each subscale was scored by the participant on a scale of 0 to 100. Then, the subscale ratings were weighted according to the participant's perception of the subscale contribution to the task load.

## E. STATISTICAL ANALYSIS

The oculometrics were assessed for normality using Kolmogorov–Smirnov tests. For each of the participant groups (young and elderly), a full-factorial repeated-measures analysis of variance was used to examine the effects of changes in load levels (i.e., low, medium and high) and experimental sessions (day 1 and day 2) as within-subject factors on the oculometrics, the performance metric (OP) and NASA-TLX scores. A Greenhouse-Geisser correction was applied if the assumption of sphericity was not met. The measure of effect size, partial eta-squared,  $\eta_p^2$ , was also calculated. Pairwise comparisons between the levels of task load were performed using a Bonferroni adjustment.

Relative and absolute reliability of the oculometrics sensitive to changes in task level were also assessed [106]. Two-way mixed single measures ICC(3,1), absolute agreement, were considered to describe the relative reliability [107]. As suggested by Becser, Sand, and Zwart, (1998) in the assessment of oculomotor variables' reliability [48], ICC > 0.75 and ICC > 0.40 indicate excellent and good reliability, respectively. The absolute reliability was described by limits of agreement (LOA) accompanied by Bland-Altman plots. Bland-Altman plots were constructed for each level of task load and inspected visually for consistency of agreement. The LOA% was calculated as  $100 \times (1.96 \times SD(diff) / grandmean)$  according to [106]. A LOA% value of up to  $\pm 30$  can be regarded as an acceptable reliability [109]. To test for heteroscedasticity, the absolute difference was plotted against the individual means in the data for significantly affected oculometrics and the existence of any trend of difference against the means was inspected. All statistical analyses were performed in SPSS 24.0 except for the Bland-Altman plots which were performed in MATLAB 2016b.

## III. RESULTS

The spread of the number of detected ocular events over a cycle across subjects was calculated as follows: blinks, Min = 0, Max = 22, M = 2, SD = 2, saccades, Min = 5, Max = 27, M = 12, SD = 3, and fixations, Min = 5, Max = 24, M = 11, SD = 3.

### A. THE YOUNG GROUP

Table 2 provides an overview of the measurements (i.e., the oculometrics, the performance metric, and the NASA-TLX scores) for the young group.

**TABLE 2.** Overview of the mean (standard deviation) values of variables on the two experimental days and at the three different task load levels in the young group. Statistically significantly changed metrics affected by load levels are indicated by underlines. (\* $p \leq 0.05$ )

	Day 1			Day 2		
	Low load	Medium load	High load	Low load	Medium load	High load
<b>Oculometrics</b>						
<b>Pupillary Response</b>						
<u>TPDL</u> (s)	0.70 (0.31)	0.81 (0.32)	0.83 (0.42)	0.74 (0.23)	0.87 (0.25)	0.91 (0.37)
<b>Saccades</b>						
<u>SF</u> (Hz)	1.35 (0.18)	1.29 (0.17)	1.19 (0.23)	1.25 (0.19)	1.19 (0.20)	1.23 (0.18)
<u>SA</u> (°)	6.4 (0.6)	6.0 (0.6)	5.8 (0.6)	5.9 (0.6)	5.9 (0.5)	5.7 (0.5)
<u>SCD</u> (ms)	64 (3)	64 (3)	63 (2)	64 (3)	64 (3)	63 (2)
<u>SPV</u> (°s <sup>-1</sup> )	159 (14)	143 (12)	138 (13)	154 (13)	141 (11)	134 (11)
SAAI (a.u.)	0.00 (0.02)	0.00 (0.02)	-0.02 (0.03)	0.00 (0.02)	-0.01 (0.02)	-0.01 (0.02)
SCR (°)	0.73 (0.10)	0.71 (0.10)	0.72 (0.11)	0.71 (0.10)	0.71 (0.10)	0.70 (0.11)
<u>SVA</u> (°s <sup>-1</sup> )	22.7 (1.0)	22.0 (0.9)	21.6 (1.3)	22.8 (1.4)	22.0 (0.7)	21.3 (0.9)
<b>Fixations</b>						
<u>FD</u> (s)	0.63 (0.11)	0.65 (0.10)	0.69 (0.12)	0.67 (0.09)	0.67 (0.10)	0.67 (0.08)
<u>FF</u> (Hz)	1.25 (0.18)	1.17 (0.18)	1.06 (0.22)	1.13 (0.19)	1.06 (0.19)	1.10 (0.17)
<b>Blinks</b>						
BF (Hz)	0.25 (0.15)	0.25 (0.10)	0.25 (0.11)	0.26 (0.12)	0.25 (0.13)	0.26 (0.12)
BD (sec)	0.30 (0.20)	0.33 (0.22)	0.34 (0.17)	0.36 (0.21)	0.45 (0.22)	0.40 (0.18)
<b>Nonlinear Metrics</b>						
<b>Pupillary Response</b>						
<u>PSaEn</u> (a.u.)	0.22 (0.06)	0.20 (0.06)	0.19 (0.06)	0.20 (0.60)	0.18 (0.05)	0.17 (0.05)
RRP (%)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
DTP (%)	99 (2)	99 (1)	99 (1)	99 (2)	99 (1)	99 (1)
<u>RMEP</u> (bit)	2.99 (0.45)	3.07 (0.42)	3.15 (0.42)	3.08 (0.55)	3.24 (0.43)	3.22 (0.50)
<b>Point of Gaze</b>						
<u>MSaEn</u> (a.u.)	1.05 (0.06)	1.03 (0.04)	1.03 (0.11)	1.06 (0.03)	1.03 (0.04)	1.05 (0.04)
<u>RRG</u> (%)	52 (2)	51 (2)	54 (4)	51 (1)	51 (1)	52 (1)
DTG (%)	100 (1)	100 (0)	100 (1)	100 (1)	100 (1)	100 (1)
<b>Performance Metric</b>						
<u>OP</u> (%)	51 (6)	43 (8)	38 (8)	52 (6)	46 (6)	44 (6)
<b>Perceived Load</b>						
<u>NASA-TLX</u> (%)	55 (13)	64 (12)	72 (12)	52 (19)	60 (16)	62 (17)

TPDL: Task-Evoked Pupillary Peak Dilation, SF: Saccade Frequency, SA: Saccade Amplitude, SCD: Saccade Duration, SPV: Saccade Peak Velocity, SAAI: Saccade Acceleration Asymmetry Index, SCR: Saccade Curvature, SVA: Saccade Velocity-Amplitude relationship (main sequence), FD: Fixation Duration, FF: Fixation Frequency, BF: Blink Frequency, BD: Blink Duration, PSaEn: Sample Entropy of Pupillary Response, RRP: Recurrence Rate of Pupillary Response, DTP: Determinism of Pupillary Response, RMEP: Recurrence Map Entropy of Pupillary Response, MSaEn: Multivariate Sample Entropy of Gaze Trajectory, RRG: Recurrence Rate of Gaze Trajectory, DTG: Determinism of Gaze Trajectory, OP: Overall Performance, NASA-TLX: Total Score of National Aeronautics and Space Administration Task Load Index.

## 1) OCULOMETRICS

The analysis of saccadic movements revealed that the SPV was significantly reduced with an increase in the load level of task,  $F(1.4, 23.5) = 121.0, p < 0.001, \eta_p^2 = 0.9$ . The SPV decreased significantly from low to medium,  $p < 0.001$ , low to high,  $p < 0.001$ , and medium to high load level,  $p < 0.001$ . The SVA also decreased significantly with an increase in the load level of task,  $F(1.5, 23.7) = 14.4, p < 0.001, \eta_p^2 = 0.5$ . The SPV was associated with SA (average  $r^2 = 0.86 \pm 0.02$ ). According to pairwise comparisons, significant reductions in SVA were observed from the low to medium,  $p = 0.015$ , low to high,  $p = 0.002$ , and medium to high load level,  $p = 0.023$ . The SF was affected significantly by the load levels,  $F(1.4, 23.6) = 4.7, p = 0.029, \eta_p^2 = 0.2$ . The SF declined from the low to medium,  $p = 0.032$ , and low to high load level,  $p = 0.015$ . An interaction between the

experimental days and the load levels was observed in the SF,  $F(1.4, 24.7) = 4.9, p = .025, \eta_p^2 = 0.2$ . The SA was reduced significantly with increased load level of the task,  $F(2, 34) = 27.4, p < 0.001, \eta_p^2 = 0.6$ . Further, an interaction was detected in the SA between the experimental days and load levels,  $F(2, 34) = 8.0, p = 0.016, \eta_p^2 = 0.2$ . The SA was also affected by the day of the experiment,  $F(1, 17) = 6.4, p = 0.021, \eta_p^2 = 0.3$ . The SCD was reduced significantly with increased load levels,  $F(2, 34) = 4.4, p = 0.020, \eta_p^2 = 0.2$ . The change in SCD between the low and high load levels was significant,  $p = 0.038$ .

The change in the load levels was reflected in fixations. The FD increased significantly in response to the increase in the load level,  $F(2, 34) = 3.3, p = .050, \eta_p^2 = 0.2$ . Further, a significant interaction in FD was detected between the experimental days and the load level,  $F(2, 34) = 4.1,$

$p = 0.025$ ,  $\eta_p^2 = 0.2$ . The FD increased significantly from the low to high load level,  $p = 0.049$ , on the first day of the experiment only. The FF was also changed significantly by the load levels,  $F(2, 34) = 5.6$ ,  $p = 0.008$ ,  $\eta_p^2 = 0.2$ . It decreased from the low to medium,  $p = 0.037$ , and low to high load level,  $p = 0.008$ .

The blinking oculometrics (i.e., BD and BF) were not significantly affected by the load levels, neither were they affected by the day of the experiment.

Regarding the pupillary responses, TPDL was significantly affected by the change in load levels,  $F(1.5, 29.0) = 10.0$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.3$ . TPDL increased significantly from the low to medium,  $p = 0.004$ , and low to high load level,  $p = 0.038$ .

The analysis of the pupillary response revealed that PSaEn was influenced significantly by the load levels,  $F(1.5, 28.7) = 21.3$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.5$ . PSaEn was reduced significantly from the low to medium,  $p = 0.003$ , and low to high load level,  $p < 0.001$ . RMEP was also affected significantly by change in the load levels,  $F(2, 38) = 13.3$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.4$ . It increased significantly from the low to medium,  $p = .006$ , and low to high load level,  $p = 0.001$ .

Likewise, the gaze coordinates dynamics revealed that MSaEn was reduced significantly with an increase in the load level,  $F(2, 36) = 5.934$ ,  $p = 0.006$ ,  $\eta_p^2 = 0.2$ . It was also significantly influenced by the experimental day,  $F(1, 18) = 6.9$ ,  $p = 0.017$ ,  $\eta_p^2 = 0.3$ . RRG changed significantly in response to the change in the load level,  $F(2, 38) = 15.9$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.4$ . RRG significantly increased from the low to high,  $p < 0.001$ , and medium to high load level,  $p = 0.001$ .

## 2) PERFORMANCE METRICS

The performance metric, OP, was also influenced significantly by the load levels:  $F(2, 34) = 29.5$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.6$ . The OP decreased significantly in response to the increase in the load levels. No significant difference was detected between the days in OP. Significant interactions were observed between the days and the load levels in OP,  $F(2, 34) = 3.6$ ,  $p = 0.040$ ,  $\eta_p^2 = 0.2$ . According to pairwise comparisons, OP was reduced significantly from the low to medium,  $p = 0.002$ , low to high,  $p < 0.001$ , and medium to high load level,  $p = 0.013$ , on the first day, and from the low to medium,  $p = 0.018$ , and low to high load level,  $p < 0.001$ , on the second day.

## 3) SUBJECTIVE ASSESSMENTS

A significant effect of change in the mental load was observed in the total NASA-TLX scores,  $F(2, 36) = 20.4$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.5$ , confirming that the perceived load changed in concordance with the manipulation of the task demands. The NASA-TLX scores increased significantly from the low to medium,  $p = 0.001$ , and low to high load level,  $p < 0.001$ .

## B. THE ELDERLY GROUP

An overview on the measurements (i.e., the oculometrics, the performance metric, and the NASA-TLX scores) for the elderly group is provided in Table 3.

### 1) OCULOMETRICS

Regarding the saccadic eye movements, the SPV was reduced significantly in response to the increase in the load level,  $F(2, 34) = 27.7$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.6$ . It significantly decreased from the low to medium,  $p < 0.001$ , low to high,  $p < 0.001$ , and medium to high load level,  $p = 0.049$ . The SVA was also significantly affected by the load levels,  $F(2, 34) = 12.1$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.4$ . It was significantly reduced from the low to high,  $p < 0.001$ , and medium to high load level,  $p = 0.021$ . The SPV was associated with SA (average  $r^2 = 0.86 \pm 0.02$ ). The SA did not change significantly with the load levels, but an interaction between the experimental days and the load levels was observed,  $F(2, 34) = 11.3$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.3$ . The SF was significantly affected by the load levels,  $F(1.5, 25.8) = 21.4$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.6$ . It was significantly reduced from the low to high level and medium to high level, both with  $p < 0.001$ . The SAAI was affected significantly by the load levels  $F(2, 34) = 3.6$ ,  $p = 0.037$ ,  $\eta_p^2 = 0.2$ . It was significantly reduced from the low to medium level,  $p = 0.029$ .

The fixations were also influenced by the load levels. The FD increased significantly in response to the increase in the load level,  $F(2, 34) = 8.8$ ,  $p = 0.001$ ,  $\eta_p^2 = 0.3$ . Pairwise comparisons revealed an increase from the low to high,  $p = 0.012$ , and medium to high load level,  $p = 0.001$ . The FF decreased significantly with the increase in the load level,  $F(2, 34) = 20.6$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.5$ . The reductions from the low to medium,  $p = 0.046$ , and low to high,  $p < 0.001$ , and medium to high load level,  $p = 0.001$ , were significant.

The blinking oculometrics (i.e., BD and BF) were not significantly affected by the change in load levels, neither were they affected by the day of the experiment.

The analysis of pupillary response revealed that the TPDL was influenced significantly by the load levels,  $F(1.4, 24.6) = 7.9$ ,  $p = 0.005$ ,  $\eta_p^2 = 0.3$ . It increased significantly from the low to high,  $p = 0.002$ , and medium to high load level,  $p = 0.003$ , according to the pairwise comparisons. The PSaEn was affected significantly by the change in load levels,  $F(1.4, 23.6) = 12.5$ ,  $p < .001$ ,  $\eta_p^2 = 0.4$ . It was reduced significantly from the low to medium,  $p = .003$ , and low to high load level,  $p = 0.009$ .

Regarding the nonlinear dynamics of the gaze coordinates, the MSaEn decreased significantly with increase in the load levels,  $F(2, 34) = 11.0$ ,  $p < .001$ ,  $\eta_p^2 = 0.4$ . Pairwise comparisons indicate that the reduction in MSaEn was significant from the low to medium,  $p = .003$ , and low to high load level,  $p = 0.002$ . The RRG increased significantly with the load levels,  $F(2, 32) = 23.0$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.6$ . The

**TABLE 3.** Overview of the mean (standard deviation) values of variables on the two experimental days and at the three different task load levels in the Elderly group. Statistically significantly changed metrics affected by load levels are indicated by underlines. (\* $p \leq 0.05$ ).

	Day 1			Day 2		
	Low load	Medium load	High load	Low load	Medium load	High load
<b>Oculometrics</b>						
<b>Pupillary Response</b>						
<u>TPDL</u> (s)	0.97 (0.29)	1.02 (0.31)	1.18 (0.51)	0.89 (0.25)	0.95 (0.38)	1.11 (0.48)
<b>Saccades</b>						
<u>SF</u> (Hz)	1.32 (0.18)	1.29 (0.13)	1.19 (0.16)	1.26 (0.24)	1.20 (0.18)	1.15 (0.20)
SA (°)	5.8 (0.5)	5.5 (0.6)	5.7 (0.6)	5.4 (0.5)	5.6 (0.6)	5.6 (0.7)
<u>SCD</u> (ms)	60 (4)	60 (5)	60 (5)	60 (4)	60 (4)	60 (4)
<u>SPV</u> (°s <sup>-1</sup> )	145 (13)	139 (14)	135 (16)	147 (13)	140 (14)	136 (12)
<u>SAAI</u> (a.u.)	-0.02 (.02)	-0.02 (0.02)	-0.02 (0.02)	-0.01 (0.02)	-0.02 (0.03)	-0.02 (0.03)
SCR (°)	0.68 (0.07)	0.70 (0.11)	0.69 (0.10)	0.68 (0.10)	0.69 (0.10)	0.66 (0.10)
<u>SVA</u> (°s <sup>-1</sup> )	23.0 (2.0)	22.5 (2.5)	21.7 (2.0)	23.3 (2.1)	22.7 (1.8)	22.0 (1.7)
<b>Fixations</b>						
<u>FD</u> (s)	0.68 (0.08)	0.68 (0.07)	0.72 (0.07)	0.71 (0.11)	0.72 (0.08)	0.74 (0.09)
<u>FF</u> (Hz)	1.21 (0.17)	1.17 (0.13)	1.08 (0.16)	1.14 (0.21)	1.07 (0.17)	1.04 (0.19)
<b>Blinks</b>						
BF (Hz)	0.22 (0.10)	0.26 (0.16)	0.26 (0.16)	0.22 (0.11)	0.25 (0.13)	0.22 (0.10)
BD (sec)	0.68 (0.10)	0.67 (0.10)	0.67 (0.11)	0.66 (0.11)	0.66 (0.11)	0.65 (0.12)
<b>Nonlinear Metrics</b>						
<b>Pupillary Response</b>						
<u>PSaEn</u> (a.u.)	0.21 (0.07)	0.19 (0.06)	0.18 (0.06)	0.20 (0.06)	0.18 (0.04)	0.18 (0.05)
RRP (%)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)
DTP (%)	98 (4)	99 (2)	98 (3)	99 (2)	99 (1)	99 (1)
RMEP (bit)	2.97 (0.62)	3.06 (0.57)	3.04 (0.65)	2.99 (0.54)	3.10 (0.47)	3.08 (0.53)
<b>Gaze Trajectory</b>						
<u>MSaEn</u> (a.u.)	1.05 (0.03)	1.02 (0.02)	1.02 (0.04)	1.05 (0.04)	1.03 (0.02)	1.02 (0.02)
<u>RRG</u> (%)	50 (1)	51 (1)	52 (2)	50 (1)	50 (2)	52 (1)
DTG (%)	99 (0)	100 (1)	100 (1)	100 (1)	100 (1)	100 (1)
<b>Performance Metric</b>						
<u>OP</u> (%)	34 (9)	25 (9)	21 (9)	38 (10)	33 (10)	29 (9)
<b>Perceived Load</b>						
<u>NASA-TLX</u> (%)	58 (13)	67 (15)	66 (13)	56 (15)	63 (13)	64 (14)

TPDL: Task-Evoked Pupillary Peak Dilation, SF: Saccade Frequency, SA: Saccade Amplitude, SCD: Saccade Duration, SPV: Saccade Peak Velocity, SAAI: Saccade Acceleration Asymmetry Index, SCR: Saccade Curvature, SVA: Saccade Velocity-Amplitude relationship (main sequence), FD: Fixation Duration, FF: Fixation Frequency, BF: Blink Frequency, BD: Blink Duration, PSaEn: Sample Entropy of Pupillary Response, RRP: Recurrence Rate of Pupillary Response, DTP: Determinism of Pupillary Response, RMEP: Recurrence Map Entropy of Pupillary Response, MSaEn: Multivariate Sample Entropy of Gaze Trajectory, RRG: Recurrence Rate of Gaze Trajectory, DTG: Determinism of Gaze Trajectory, OP: Overall Performance, NASA-TLX: Total Score of National Aeronautics and Space Administration Task Load Index

increases from the medium to high and low to high level were significant, both with  $p < 0.001$ .

## 2) PERFORMANCE METRIC

OP decreased significantly by the increase in load levels  $F(2, 34) = 53.8, p < 0.001, \eta_p^2 = 0.8$ . A significant increase was detected in OP,  $F(1, 17) = 26.9, p < 0.001, \eta_p = 0.6$ , from day 1 to day 2. The difference in OP between the days was  $M = 0.07, SD = 0.01$ . The significant reductions in OP in response to the levels of load were as follows: from the low to medium,  $p < 0.001$ , low to high,  $p < 0.001$ , and medium to high load level,  $p = 0.010$ .

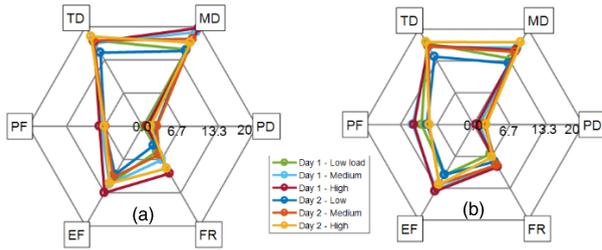
## 3) SUBJECTIVE ASSESSMENTS

The total score of NASA-TLX increased with the load levels,  $F(2, 34) = 23.0, p < 0.001, \eta_p^2 = 0.6$ . The increase was

observed from the low to medium,  $p < 0.001$ , and low to high load level,  $p < 0.001$ . Fig 2 shows how the participants assessed the task load in terms of NASA-TLX subscales. It seems that mental and temporal demand predominated the different aspects of the task load compared with the other subscales in both groups. It is also commonly observable that the physical demand scored low ( $M: 1.43, SD: 1.40$ ) in both groups.

## C. RELIABILITY ANALYSIS

The reliability of the affected oculometrics by the levels of load without significant difference across days was analyzed. No sign of heteroscedasticity was found by visually inspecting the plots of the difference and the mean values from the first and second experimental sessions (days). The results for the relative and absolute reliability in the young and elderly groups are summarized in Table 4.



**FIGURE 2.** Mean scores of NASA-TLX subscales for each load level (i.e., low, medium and high) and each experimental day (i.e., 1 and 2) in (a) the young group and (b) the elderly group. NASA-TLX subscales are as follows: MD: Mental Demand, TD: Temporal Demand, PF: Performance, EF: Effort, FR: Frustration, PD: Physical Demand.

The SVA showed good relative reliability in both groups with  $ICC > 0.44$  in the young and  $ICC > 0.61$  in the elderly group for all load levels. The absolute reliability of SVA was acceptable in both groups with a maximum LOA% of 11 in the young and 18 in the elderly group across the load levels. The relative reliability in SPV was also evaluated as good with the lowest ICC values of 0.50 and 0.54 in the young and elderly group, respectively. The LOA also revealed that the absolute reliability in SPV was within the acceptable range, with the highest LOA% of 16 and 19 for the young and elderly group, respectively. The SCD exhibited a good relative reliability with the lowest ICC values of 0.48 and 0.87 in the young and elderly groups, respectively. The LOA% supports the acceptable absolute reliability in SCD with the highest values equal to 8 in both groups. The SF did not provide a good relative reliability with  $ICC = 0.39$  in the high load level in the young group and  $ICC = 0.35$  in the medium load level in the elderly group. In neither groups, the LOA suggested an acceptable absolute reliability for the SF ( $> 30$ ). Relative reliability in the SAAI was not within a good range in terms of ICC in the young group ( $< 0.40$  in low and high load levels). Furthermore, the LOA% revealed an unacceptable absolute reliability of SAAI in both groups ( $< -30$ ).

Regarding the pupillary responses, the ICC values for the TPDL indicated good relative reliability in both groups with the lowest ICC values of 0.66 and 0.70 in the young and elderly group, respectively. However, the absolute reliability in the TPDL was not within the acceptable range. The PSaEn also exhibited a reliability state similar to the TPDL, in which the ICC values represented good reliability but the LOA% did not. The assessment of the relative reliability in the RMEP revealed a good reliability with the lowest ICC values of 0.51 and 0.78 in the young and elderly group, respectively. The LOA% was within the acceptable range of absolute reliability for the RMEP in both groups.

Concerning the reliability of the fixational oculometrics, the FD exhibited a good relative reliability with the lowest ICC values of 0.62 and 0.41 in the young and elderly group, respectively. The LOA% was also within the acceptable range for FD in both groups with the highest LOA% of 26 and 27 for the young and elderly group, respectively. The FF did not provide a good relative reliability with  $ICC < 0.40$  in the low and medium load levels in the elderly group as well as in

the high load level in the young group. However, the LOA% indicated an acceptable absolute reliability of FF in both groups.

The RRG did not provide good relative reliability with the lowest ICC values of 0.32 and 0.36 in the young and elderly group, respectively. However, LOA% for RRG was within the acceptable reliability range in both groups with the highest LOA% values of 5 and 6 in the young and elderly group, respectively.

The Bland-Altman plots (Supplementary Materials, Fig. 2, 3, and 4) represent agreement between the two experimental days where the difference between the two days was plotted against the average of the two days for saccadic, fixational, and pupillary oculometrics in each load level for both young and elderly groups. According to the Bland-Altman plots, no systematic difference between the young and elderly groups was found for any of the oculometrics. Commonly for all oculometrics, it can be observed that the mean LOA values are close to zero in all the levels of load suggesting no significant bias between days supported by a uniform scattering of points close to midline compared with the upper and lower LOAs.

#### IV. DISCUSSION

In quest for a sensitivity and reliability analysis of psychophysiological metrics in response to mental load variation in an occupationally relevant task, this study focused on important oculometrics advantages such as unobtrusive acquisition, quickness of the reaction, and specificity to mental load variation compared with other potential physiological metrics [110]–[114]. The present study took advantage of video-based eye-tracking to provide an ecologically valid model to monitor changes in mental load on oculometrics. With this aim, we conducted the study and found that the saccadic, fixational, and pupillary metrics can reflect mental load variation in both young and elderly adults. Further, as reflected in the SVA, SPV, FD, and SCD, the responses remained consistent across the experimental days. The results of the sensitivity analysis of the oculometrics support our first hypothesis that the mental load variation can be reflected in oculometrics. In addition, we have found the affected oculometrics to be reliable which supports our second hypothesis. This suggests that oculometrics can be adopted to monitor the variability of human mental load in a functional computer task with concurrent mental demands.

We did not find considerable differences in the level of reliability of oculometrics in the young and elderly groups. Furthermore, the level of load did not change the reliability indices noticeably. This evidence implies that the eye-tracking technology could serve as a robust interaction modality to detect different levels of mental load in both young and old adults.

##### A. SENSITIVITY TO MENTAL LOAD

The SPV and SVA were influenced significantly by the load levels in both groups. This negative correlation of SPV and

**TABLE 4.** The test-retest reliability of the significantly affected oculometrics by the task load variations in the young and elderly groups.

	Young Group			Elderly Group				
	ICC (95% CI)	p-value	LOA (95% CI)	LOA%	ICC (95% CI)	p-value	LOA (95% CI)	LOA%
<b>Saccades</b>								
<b>SVA (<math>^{\circ}\text{s}^{-1}</math>)</b>								
Low load	0.47 [0.31 0.75]	0.020	-0.2 [-2.9 2.4]	11	0.79 [0.51 0.92]	<0.001	-0.2 [-2.9 2.4]	11
Medium load	0.54 [0.11 0.79]	0.009	0.1 [-1.6 1.7]	7	0.61 [0.21 0.83]	0.003	-0.2 [-4.4 3.4]	18
High load	0.44 [0.04 0.73]	0.026	0.3 [-2.0 2.6]	11	0.81 [0.57 0.92]	<0.001	-0.4 [-3.5 2.6]	13
<b>SPV (<math>^{\circ}\text{s}^{-1}</math>)</b>								
Low load	0.54 [0.15 0.79]	0.005	4.2 [-20.6 29.1]	16	0.72 [0.40 0.88]	<0.001	-1.8 [-21.1 17.4]	13
Medium load	0.50 [0.10 0.77]	0.010	2.7 [-19.0 24.6]	15	0.54 [0.11 0.80]	0.010	-1.5 [-28.2 25.2]	19
High load	0.69 [0.37 0.87]	<0.001	3.3 [-14.1 20.8]	13	0.65 [0.27 0.85]	0.002	-1.4 [-24.7 21.9]	17
<b>SAAI (a.u.)</b>								
Low load	0.10 [-0.38 0.53]	0.342	0.00 [-0.06 0.06]	-2383	0.61 [0.23 0.83]	0.002	-0.01 [-0.04 0.03]	-269
Medium load	0.46 [0.05 0.74]	0.015	0.00 [-0.03 0.05]	-796	0.51 [0.05 0.78]	0.016	0.00 [-0.03 0.05]	-201
High load	0.38 [-0.08 0.71]	0.051	0.00 [-0.05 0.05]	-401	0.55 [0.13 0.80]	0.008	0.00 [-0.05 0.06]	-262
<b>SCD (s)</b>								
Low load	0.67 [0.31 0.86]	0.001	0.0 [-0.005 0.005]	8	0.89 [0.73 0.96]	<0.001	0.0 [-0.004 0.004]	7
Medium load	0.61 [0.24 0.83]	0.002	0.0 [-0.005 0.005]	7	0.88 [0.71 0.95]	<0.001	0.0 [-0.004 0.005]	8
High load	0.48 [0.04 0.76]	0.018	0.0 [-0.006 0.005]	8	0.87 [0.71 0.95]	<0.001	0.0 [-0.004 0.005]	8
<b>SF (Hz)</b>								
Low load	0.40 [-0.01 0.71]	0.030	0.09 [-0.33 0.51]	32	0.41 [-0.04 0.73]	0.039	0.06 [-0.39 0.50]	34
Medium load	0.41 [0.01 0.71]	0.022	0.09 [-0.32 0.50]	32	0.35 [-0.06 0.69]	0.044	0.09 [-0.25 0.43]	53
High load	0.39 [-0.07 0.71]	0.048	-0.04 [-0.48 0.40]	36	0.43 [-0.04 0.74]	0.036	0.03 [-0.35 0.41]	32
<b>Fixations</b>								
<b>FD (s)</b>								
Low load	0.62 [0.26 0.83]	0.001	-0.03 [-0.20 0.14]	26	0.51 [0.09 0.78]	0.011	-0.03 [-0.22 0.15]	27
Medium load	0.66 [0.33 0.85]	0.001	-0.02 [-0.18 0.14]	24	0.55 [0.07 0.81]	0.002	-0.05 [-0.18 0.08]	18
High load	0.67 [0.33 0.86]	0.001	0.01 [-0.14 0.17]	23	0.41 [-0.06 0.73]	0.043	-0.02 [-0.19 0.16]	24
<b>FF (Hz)</b>								
Low load	0.42 [0.01 0.72]	0.020	0.1 [-0.3 0.5]	34	0.36 [-0.08 0.69]	0.057	0.1 [-0.3 0.1]	36
Medium load	0.44 [0.04 0.72]	0.016	0.1 [-0.3 0.5]	36	0.31 [-0.10 0.65]	0.068	0.1 [-0.3 0.4]	30
High load	0.31 [-0.17 0.67]	0.099	0.0 [-0.5 0.4]	41	0.42 [-0.04 0.73]	0.039	0.0 [-0.3 0.4]	35
<b>Pupillary Response</b>								
<b>TPDL (s)</b>								
Low load	0.80 [0.56 0.91]	<0.001	0.0 [-0.3 0.3]	48	0.73 [0.41 0.90]	<0.001	0.1 [-0.3 0.5]	40
Medium load	0.66 [0.33 0.85]	0.001	0.0 [-0.5 0.4]	55	0.81 [0.58 0.92]	<0.001	0.1 [-0.3 0.5]	41
High load	0.69 [0.38 0.86]	<0.001	-0.1 [-0.7 0.5]	70	0.68 [0.33 0.87]	0.001	0.1 [-0.7 0.8]	69
<b>PSaEn (a.u.)</b>								
Low load	0.45 [0.05 0.73]	0.013	0.03 [-0.09 0.15]	59	0.62 [0.22 0.84]	0.003	0.00 [-0.11 0.11]	55
Medium load	0.54 [0.15 0.78]	0.003	0.02 [-0.07 0.12]	51	0.72 [0.40 0.89]	<0.001	0.01 [-0.07 0.08]	42
High load	0.40 [0.08 0.71]	0.032	0.01 [-0.11 0.13]	67	0.78 [0.51 0.91]	<0.001	0.00 [-0.07 0.08]	42
<b>RMEP (bit)</b>								
Low load	0.51 [0.10 0.77]	0.009	-0.1 [-1.1 0.9]	32	0.80 [0.55 0.92]	<0.001	-0.02 [-0.7 0.7]	24
Medium load	0.60 [0.23 0.82]	0.001	-0.1 [-0.9 0.5]	23	0.78 [0.50 0.91]	<0.001	-0.03 [-0.7 0.5]	23
High load	0.56 [0.16 0.80]	0.005	-0.1 [-0.9 0.8]	27	0.84 [0.63 0.94]	<0.001	-0.04 [-0.7 0.8]	22
<b>Gaze Trajectory</b>								
<b>RRG (%)</b>								
Low load	0.56 [0.16 0.81]	0.005	0.5 [-1.4 2.4]	4	0.54 [0.14 0.78]	0.003	0.4 [-1.8 2.5]	4
Medium load	0.47 [0.40 0.76]	0.019	-0.4 [-3.0 2.2]	5	0.36 [0.06 0.68]	0.048	0.5 [-2.4 3.4]	6
High load	0.32 [0.18 0.68]	0.099	0.0 [-2.1 2.1]	4	0.40 [0.05 0.71]	0.040	0.0 [-2.3 2.3]	4

TPDL: Task-Evoked Pupillary Peak Dilation, SF: Saccade Frequency, SA: Saccade Amplitude, SD: Saccade Duration, SPV: Saccade Peak Velocity, SAAI: Saccade Acceleration Asymmetry Index, SVA: Saccade Velocity-Amplitude relationship, FD: Fixation Duration, FF: Fixation Frequency, PSaEn: Sample Entropy of Pupillary Response, RMEP: Recurrence Map Entropy of Pupillary Response, RRG: Recurrence Rate of Gaze Trajectory

mental load is in agreement with previous studies, e.g., [115]. As indicated in previous studies [43], [91], SVA reflected mental load variations identified from the distribution of saccades. Some studies have linked the association between saccadic dynamics and task demands to the activation of the sympathetic nervous system to modulate arousal levels [42]. The total number of spikes of the excitatory burst neurons in the paramedian pontine reticular formation, which encodes SA, and their instantaneous firing rate, which encodes SPV, are closely related [20], explaining the correlation between saccadic amplitudes and peak velocities [55]. The magnitude of the excitatory burst neuron discharge determines the

saccadic velocity profile [20] which could be reflected in SVA. As a result of the higher number of pattern points and geometrical complexity of the patterns in higher load levels compared with the lower load levels, the increasing uncertainty levels could also be a reason for the decreasing saccade velocity as the uncertainty level and neural activity in superior colliculus have been found to be negatively correlated [116].

Although an alteration in the morphology of the saccadic profile was expected [117], we did not find any significant change in the SCR. Still, the SAAI was found to reflect varying levels of mental load in the elderly group only. The

SAAI has been reported to quantify accelerations and decelerations of eye movements during saccades and to reflect the neural circuitry involved in the production of the oculomotor control [93]. The acceleration profile of saccadic movements during horizontal saccades from a fixation point in the center to predefined targets was studied in [118] where the participants were informed about the positions of the targets to avoid spatial uncertainty. The saccades followed the same model in young and old adults in [118]. However, as this was not examined with variable levels of mental load, it is conceivable that the variation in mental load may have led to alteration in the oculomotor control strategy and resulted in the difference in the SAAI between the young and elderly group in the current study.

In addition to the aforementioned saccadic metrics, the change in the SF was significant in both groups with higher effects in the elderly group (with a fold change of 9%) compared with the young group (with a fold change of 6%) from low to high load level. On the other hand, the SCD was only affected by the levels of load in the young group even though the functional relevance of such difference remains questionable.

The FF and FD were influenced significantly by the levels of load in both groups. The changes in the FF and FD were in agreement with the previous findings where FF decreased and FD increased with the increase in mental load [87]. The increase in uncertainty levels, from the lower to the higher levels of task, may also contribute to the longer FD [119]; thus decreasing FF. The effects on FF and FD were more pronounced in the elderly group (with a fold change of 9% and 6.5%, respectively) compared with the young group (with a fold change of 7% and 4.5%, respectively) from low to high load level. These differences between the young and elderly group may arise from the cognitive declines due to aging resulting in elongated fixations during visual search to recognize each point of the patterns to be reproduced [120]. Taken together with the SF in relation to mental load, the influence of aging on visual attention [121] and working memory [122] was highlighted in the current study.

The BD and BF did not change significantly with the load variations in any of the groups. Although some studies have report the association of blinking with mental load [82], [85], [123], our results are in line with those studies which have found that the blinking is independent of mental load variation [124], [125]. These discrepancies among studies may reflect the differences in the visual demands of the tasks [28] and not necessarily the mental demands of the tasks.

The analysis of the pupillary response revealed that in both groups of participants, the TPD, PSaEn, and RMEP were influenced significantly by the level of load, whereas the RRP and DTP were not. As expected, in both groups the TPD increased with the level of load in association with the increase in the required efforts to accomplish the tasks [126]. Although the RMEP changed significantly with the load levels in the young group only, a similar tendency

was found in the elderly group. RMEP was positively correlated with the load levels, mostly reflecting a growth of regularity in the pupillary response. The links between the working memory and the pupillary response, especially in visual search, may explain this effect [127], [128]. The PSaEn decreased in both groups in response to the higher load levels implying less complex patterns in pupillary responses during increasing levels of mental load [34]. In many studies pupillary response is suggested to serve as a psychophysiological index quantifying mental load in both young and old adults which was also confirmed in this study [110], [129].

The increase in mental load during the proposed task was corroborated by a significant reduction in the performance metric and an increase in the subjective metric of the load perception. The reduction in the performance metric with the load levels was in agreement with a previous study [91]. The observed changes in performance metric across days in the elderly group were mostly due to a learning effect despite our efforts to acclimatize the participants to the task, suggesting that the training duration should be extended in future studies. Despite the increased performance over days, it did not have any impact on most of the oculometrics.

## B. RELIABILITY OF THE OCULOMETRICS

The SVA, SPV, SCD, and FD were found to be reliable oculometrics according to Table 4. This finding was in line with a previous study suggesting that SVA is highly reliable [49]. The SPV has also been reported to be a reliable metric [48], [49]. The saccade and fixation have been found to provide good to excellent reliability in the associated temporal characterizing oculometrics [46]–[48], [130].

A part of the variations observed in saccadic eye movement may arise from the underlying variability inherent to motor control. For example, purely horizontal and vertical saccades have been shown to be less variable compared with saccades in oblique directions [131]. This might also be the reason why saccades exhibited lower relative reliability in the current study compared with the standardized pro-saccade and anti-saccade tasks in [49] involving only horizontal saccades.

The RMEP showed good to excellent relative reliability to differentiate mental load variations across the levels. In [132] the underlying dynamics of pupillary responses were better captured by employing the RQA compared with linear metrics, e.g., pupil size. This suggests the potential applicability of nonlinear techniques on pupillary responses in empirical studies.

## C. LIMITATIONS, CHALLENGES, AND SUGGESTIONS

In addition to the application of oculometrics in quantifying mental load supported by the results, other psychophysiological measures, e.g., electroencephalography or heart rate variability could be employed for further studies. Sympathetic and parasympathetic fibers dictate the pupil diameter to contract or dilate by balanced activation of the dilator and sphincter pupillae muscles in changing the light exposures towards the eyes [110]. Given this fact, the use of oculometrics based

on pupillary response, e.g., the TPDL, would be limited to paradigms with controlled light settings. Furthermore, the use of task-evoked oculometrics, e.g., TPDL, may impose the limitation of task-specificity. However, the idea of measuring latencies and defining the onset of a mental processing event may be customizable in future applications.

Response time would quantify processing speed as a performance measure, e.g. [120]. In experimental settings involving various cognitive processes such as current study, however, some qualitative aspects of cognitive abilities could also be captured using combined high-level performance measures including response time as well as mouse cursor and gaze trace patterns. This may especially highlight the different strategies recruited by young and elderly individuals.

Since nonlinear oculometrics exhibited promising results in this study reflecting the sensitivity of eye movement dynamics to changes in mental load, further studies in this field are warranted. For example, the categorical RQA measure for gaze trajectory may provide a straightforward interpretation of spatiotemporal exploration strategies in the oculomotor system [38], [133]–[135].

This study highlighted some important issues regarding the study of oculometrics under mentally demanding tasks. Firstly, the differences between individuals should be acknowledged in experimental studies, including social class and education, which this study considered when recruiting the participants. Participants had to be familiar with computer work and understand English since the task involved verbal comprehension of the shape names of the pattern points. Another important notion is how to precisely classify an individual into a young or an elderly group. For this study, we selected the working age definition in developed countries which partially corresponds to the structural changes in the visual system due to aging in healthy individuals, e.g., [58], [136]. Other important factors to be noted in oculometrics-based human-centered design are the consumption of alcoholic or caffeinated beverages, drugs or medications, smoking, any mental disorder, and sleep deprivation; which could all affect the dynamics of the eye movements. Finally, yet importantly, since financial incentives have been shown to motivate participant to engage more in tasks involving mental demands [137], this was also taken into account in this study by asking the participants to perform their best to get a financial reward (100 Danish Kroner) if they achieved the highest performance among each group of participants (social comparison).

## V. CONCLUSION

In conclusion, we obtained a good to excellent test-retest reliability and an acceptable absolute reliability between the two experimental days in the SPV, SVA, SCD, and FD quantifying the variation in mental load during a functional task. These oculometrics were significantly changed in both the young and elderly group in response to mental load variations. The use of the video-based eye tracker allowed an ecologically valid study of the oculomotor system during the computer

task. The results of this study may provide a basis for the design of a cognition-aware human computer interface which may increase work productivity and reduce potential health risks due to sustained mental loading.

## APPENDIX

Firstly, an initial saccadic peak velocity threshold ( $PT_1$ ) of  $100^\circ/s$  was chosen for the entire task segment, then the optimal peak velocity threshold was iteratively found to satisfy the condition  $|PT_n - PT_{n-1}| < 1$  where  $PT_n = \mu_{n-1} + 6\sigma_{n-1}$ ,  $\mu_{n-1}$ : average and  $\sigma_{n-1}$ : standard deviation of the angular gaze velocity samples ( $\dot{\theta}$ ) of the entire task segment with values lower than  $PT_n$ ,  $n$ : iteration number. Once saccade peak threshold,  $PT_m$ , was obtained at  $m$ -th iteration, potential saccadic segments were identified where  $\dot{\theta} > PT_m$ . For each potential saccade, the onset sample was determined as the first velocity sample ( $i$ ) going below the local saccade onset threshold of  $\dot{\theta}_{ST}^{onset} = \mu_{m-1} + 3\sigma_{m-1}$  and  $(\dot{\theta}_i - \dot{\theta}_{i+1}) \geq 0$  in backward search in time. The offset threshold was determined in forward search in time as the first sample going below saccade offset threshold  $\dot{\theta}_{ST}^{offset} = 0.7\dot{\theta}_{ST}^{onset} + 0.3\dot{\theta}_i$  and  $(\dot{\theta}_i - \dot{\theta}_{i+1}) \leq 0$ , where  $\dot{\theta}_i = \mu_i + 3\sigma_i$  was a locally adaptive noise factor calculated within a window of 40 ms preceding the onset of the saccade. The detected saccadic samples between the onset and offset had to satisfy the saccade criteria (Maximum saccade velocity of  $500^\circ/s$ , Maximum saccade acceleration of  $50000^\circ/s$ , and Minimum saccade duration of 10 ms). Additionally, a saccade trajectory had not to deviate more than  $60^\circ$  from its main direction which was calculated as the average of sample-to-sample ( $k$  to  $k+1$ ) direction, i.e.,  $Arctan(v(k+1) - v(k)/h(k+1) - h(k))$  of three consecutive horizontal ( $h$ ) and vertical ( $v$ ) gaze coordinate samples centred around the sample with the highest velocity within the saccade segment [76].

## REFERENCES

- [1] C. D. Wickens, "Processing resources in attention, dual task performance, and workload assessment," in *Varieties of Attention*. New York, NY, USA: Academic, 1984, pp. 63–102.
- [2] W. Karwowski, *International Encyclopedia of Ergonomics and Human Factors*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2006.
- [3] S. Chen and J. Epps, "Using task-induced pupil diameter and blink rate to infer cognitive load," *Human-Comput. Interaction*, vol. 29, no. 4, pp. 390–413, 2014.
- [4] D. van der Linden, M. Frese, and S. Sonnentag, "The impact of mental fatigue on exploration in a complex computer task: Rigidity and loss of systematic strategies," *Hum. Factors*, vol. 45, no. 3, pp. 483–494, 2003.
- [5] S. Stansfeld and B. Candy, "Psychosocial work environment and mental health—a meta-analytic review," *Scand. J. Work. Environ. Health*, pp. 443–462, 2006.
- [6] T. A. Salthouse, *Theoretical Perspectives on Cognitive Aging*. New York, NY, USA: Psychology Press, 2016.
- [7] T. A. Salthouse, "What and when of cognitive aging," *Current Directions Psychol. Sci.*, vol. 13, no. 4, pp. 140–144, 2004.
- [8] J. M. Jansma, N. F. Ramsey, R. Coppola, and R. S. Kahn, "Specific versus nonspecific brain activity in a parametric N-back task," *NeuroImage*, vol. 12, no. 6, pp. 688–697, 2000.
- [9] J. Beatty, "Task-evoked pupillary responses, processing load, and the structure of processing resources," *Psychol. Bull.*, vol. 91, no. 2, p. 276, 1982.

- [10] D. Srinivasan, S. E. Mathiassen, D. M. Hallman, A. Samani, P. Madeleine, and E. Lyskov, "Effects of concurrent physical and cognitive demands on muscle activity and heart rate variability in a repetitive upper-extremity precision task," *Eur. J. Appl. Physiol.*, vol. 116, no. 1, pp. 227–239, 2016.
- [11] G. Matthews, L. E. Reinerman-Jones, D. J. Barber, and J. Abich, "The psychometrics of mental workload: Multiple measures are sensitive but divergent," *Human Factors*, vol. 57, no. 1, pp. 125–143, 2014.
- [12] M. Grassmann, E. Vlemingx, A. von Leupoldt, and O. Van den Bergh, "The role of respiratory measures to assess mental load in pilot selection," *Ergonomics*, vol. 59, no. 6, p. 745, 2016.
- [13] P. Richter, T. Wagner, R. Heger, and G. Weise, "[Curvature] psychophysiological analysis of mental load during driving on rural roads—A quasi-experimental field study," *Ergonomics*, vol. 41, no. 5, pp. 593–609, 1998.
- [14] M. Hitchen, D. A. Brodier, and J. B. Harnesst, "Cardiac responses to demanding mental load," *Ergonomics*, vol. 23, no. 4, pp. 379–385, 1980.
- [15] P. Averty, S. Athenes, C. Collet, and A. Dittmar, "Evaluating a new index of mental workload in real ATC situation using psychophysiological measures," in *Proc. 21st Digit. Avionics Syst. Conf.*, vol. 2, 2002, pp. 7A4-1–7A4-13.
- [16] A. Yuce, H. Gao, G. L. Cuendet, and J. P. Thiran, "Action units and their cross-correlations for prediction of cognitive load during driving," *IEEE Trans. Affect. Comput.*, vol. 8, no. 2, pp. 161–175, Feb. 2017.
- [17] N. Nourbakhsh, F. Chen, Y. Wang, and R. A. Calvo, "Detecting users' cognitive load by galvanic skin response with affective interference," *ACM Trans. Interaction Intell. Syst.*, vol. 7, no. 3, pp. 1–20, 2017.
- [18] D. P. Munoz, J. R. Broughton, J. E. Goldring, and I. T. Armstrong, "Age-related performance of human subjects on saccadic eye movement tasks," *Exp. Brain Res.*, vol. 121, no. 4, pp. 391–400, 1998.
- [19] N. J. Gandhi and H. A. Katmani, "Motor functions of the superior colliculus," *Annu. Rev. Neurosci.*, vol. 34, pp. 205–231, Mar. 2011.
- [20] R. J. Leigh and D. S. Zee, *The Neurology of Eye Movements*. Oxford, U.K.: Oxford Univ. Press, 2015.
- [21] J. L. McFarland and A. F. Fuchs, "Discharge patterns in nucleus prepositus hypoglossi and adjacent medial vestibular nucleus during horizontal eye movement in behaving macaques," *J. Neurophysiol.*, vol. 68, no. 1, pp. 319–332, 1992.
- [22] H. W. Magoun and S. W. Ranson, "The central path of the light reflex: A study of the effect of lesions," *Arch. Ophthalmol.*, vol. 13, no. 5, pp. 791–811, May 1935.
- [23] R. Z. Marandi and S. H. Sabzpoushan, "Qualitative modeling of the decision-making process using electrooculography," *Behav. Res. Methods*, vol. 47, no. 4, pp. 1404–1412, 2014.
- [24] R. Zargari Marandi and S. Sabzpoushan, "Using eye movement analysis to study auditory effects on visual memory recall," *Basic Clin. Neurosci.*, vol. 5, no. 1, pp. 55–65, 2014.
- [25] T. Halverson and A. J. Hornof, "A computational model of 'active vision' for visual search in human-computer interaction," *Human-Comput. Interaction*, vol. 26, no. 4, pp. 285–314, 2012.
- [26] A. J. Hornof, "Cognitive strategies for the visual search of hierarchical computer displays," *Human-Comput. Interaction*, vol. 19, no. 3, pp. 183–223, 2004.
- [27] C. M. Masciocchi and J. D. Still, "Alternatives to eye tracking for predicting stimulus-driven attentional selection within interfaces," *Human-Comput. Interaction*, vol. 28, no. 5, pp. 417–441, 2013.
- [28] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford, U.K.: Oxford Univ. Press, 2011.
- [29] L. L. Di Stasi *et al.*, "Saccadic peak velocity sensitivity to variations in mental workload," *Aviat. Sp. Environ. Med.*, vol. 81, no. 4, pp. 413–417, 2010.
- [30] J. Cegarra and A. Chevalier, "The use of Tholos software for combining measures of mental workload: Toward theoretical and methodological improvements," *Behav. Res. Methods*, vol. 40, no. 4, pp. 988–1000, 2008.
- [31] S. Martin, "Measuring cognitive load and cognition: Metrics for technology-enhanced learning," *Edu. Res. Eval.*, vol. 20, nos. 7–8, pp. 592–621, 2014.
- [32] X. Liu, T. Chen, G. Xie, and G. Liu, "Contact-free cognitive load recognition based on eye movement," *J. Elect. Comput. Eng.*, vol. 2016, Oct. 2016, Art. no. 1601879.
- [33] A. Samani, D. Srinivasan, S. E. Mathiassen, and P. Madeleine, "Non-linear metrics assessing motor variability in a standardized pipetting task: Between-and within-subject variance components," *J. Electromyogr. Kinesiol.*, vol. 25, no. 3, pp. 557–564, 2015.
- [34] P. Zarjam, J. Epps, and N. H. Lovell, "Characterizing mental load in an arithmetic task using entropy-based features," in *Proc. 11th Int. Conf. Inf. Sci., Signal Process. Appl. (ISSPA)*, 2012, pp. 199–204.
- [35] R. A. McKinley, L. K. McIntire, R. Schmidt, D. W. Repperger, and J. A. Caldwell, "Evaluation of eye metrics as a detector of fatigue," *Human Factors*, vol. 53, no. 4, pp. 403–414, 2011.
- [36] S. S. Monfort, C. M. Sibley, and J. T. Coyne, "Using machine learning and real-time workload assessment in a high-fidelity UAV simulation environment," *Proc. SPIE*, vol. 9851, p. 98510B, May 2016.
- [37] N. C. Anderson, W. F. Bischof, K. E. W. Laidlaw, E. F. Risko, and A. Kingstone, "Recurrence quantification analysis of eye movements," *Behav. Res. Methods*, vol. 45, no. 3, pp. 842–856, 2013.
- [38] D. W. Wu, N. C. Anderson, W. F. Bischof, and A. Kingstone, "Temporal dynamics of eye movements are related to differences in scene complexity and clutter," *J. Vis.*, vol. 14, no. 2014, pp. 1–14, 2014.
- [39] J. F. Ruiz-Rabelo *et al.*, "Validation of the NASA-TLX score in ongoing assessment of mental workload during a laparoscopic learning curve in bariatric surgery," *Obes. Surg.*, vol. 25, no. 12, pp. 2451–2456, 2015.
- [40] L. L. Di Stasi *et al.*, "Gaze entropy reflects surgical task load," *Surg. Endosc.*, vol. 30, no. 11, pp. 1–10, 2016.
- [41] L. L. Di Stasi *et al.*, "Saccadic eye movement metrics reflect surgical residents' fatigue," *Ann. Surg.*, vol. 259, no. 4, pp. 824–829, 2014.
- [42] L. L. Di Stasi, A. Catena, J. J. Cañas, S. L. Macknik, and S. Martinez-Conde, "Saccadic velocity as an arousal index in naturalistic tasks," *Neurosci. Biobehav. Rev.*, vol. 37, no. 5, pp. 968–975, 2013.
- [43] L. L. Di Stasi, M. Marchitto, A. Antolí, T. Baccino, and J. J. Cañas, "Approximation of on-line mental workload index in ATC simulated multitasks," *J. Air Transp. Manag.*, vol. 16, no. 6, pp. 330–333, 2010.
- [44] G. Marquart, C. Cabrall, and J. de Winter, "Review of eye-related measures of drivers' mental workload," *Proc. Manuf.*, vol. 3, pp. 2854–2861, Mar. 2015.
- [45] L. P. Prieto, K. Sharma, L. Kidzinski, and P. Dillenbourg, "Orchestration load indicators and patterns: In-the-wild studies using mobile eye-tracking," *IEEE Trans. Learn. Technol.*, to be published.
- [46] C. Klein and B. Fischer, "Instrumental and test-retest reliability of saccadic measures," *Biol. Psychol.*, vol. 68, no. 3, pp. 201–213, 2005.
- [47] M. Versino *et al.*, "Quantitative evaluation of saccadic and smooth pursuit eye movements. Is it reliable?" *Invest. Ophthalmol. Vis. Sci.*, vol. 34, pp. 1702–1709, 1993.
- [48] U. Ettinger, V. Kumari, T. J. Crawford, R. E. Davis, T. Sharma, and P. J. Corr, "Reliability of smooth pursuit, fixation, and saccadic eye movements," *Psychophysiology*, vol. 40, no. 4, pp. 620–628, 2003.
- [49] G. Bargary, J. M. Bosten, P. T. Goodbourn, A. J. Lawrence-Owen, R. E. Hogg, and J. D. Mollon, "Individual differences in human eye movements: An oculomotor signature?" *Vis. Res.*, vol. 141, pp. 157–169, Dec. 2017.
- [50] G. Horgen and T. Langaas, "Saccade latency and fixation stability: Repeatability and reliability," *J. Eye Movement Res.*, vol. 9, no. 2, pp. 1–13, 2016.
- [51] P. Roy-Byrne, A. Radant, D. Wingerson, and D. S. Cowley, "Human oculomotor function: Reliability and diurnal variation," *Biol. Psychiatry*, vol. 38, no. 2, pp. 92–97, 1995.
- [52] I. Meyhöfer, K. Bertsch, M. Esser, and U. Ettinger, "Variance in saccadic eye movements reflects stable traits," *Psychophysiology*, vol. 53, no. 4, pp. 566–578, 2016.
- [53] T. H. Turner, J. B. Renfroe, A. Dupstadt-Delambo, and V. K. Hinson, "Validation of a behavioral approach for measuring saccades in Parkinson's disease," *J. Motion Behav.*, vol. 49, no. 6, pp. 657–667, 2017.
- [54] M. R. MacAskill and T. J. Anderson, "Eye movements in neurodegenerative diseases," *Current Opin. Neurol.*, vol. 29, no. 1, pp. 61–68, 2016.
- [55] E. Bollen *et al.*, "Variability of the main sequence," *Invest. Ophthalmol. Vis. Sci.*, vol. 34, no. 13, pp. 3700–3704, 1993.
- [56] S. Stuart, L. Alcock, A. Godfrey, S. Lord, L. Rochester, and B. Galna, "Accuracy and re-test reliability of mobile eye-tracking in Parkinson's disease and older adults," *Med. Eng. Phys.*, vol. 38, no. 3, pp. 308–315, 2016.
- [57] S. Dowiasch, S. Marx, W. Einhäuser, and F. Bremmer, "Effects of aging on eye movements in the real world," *Frontiers Human Neurosci.*, vol. 9, p. 46, Feb. 2015.
- [58] G. Nelles, A. de Greiff, A. Pscherer, and J. Esser, "Age-related differences of saccade induced cortical activation," *Neurosci. Lett.*, vol. 458, no. 1, pp. 15–18, 2009.

- [59] C. Klein, B. Fischer, K. Hartnegg, W. H. Heiss, and M. Roth, "Optomotor and neuropsychological performance in old age," *Exp. Brain Res.*, vol. 135, no. 2, pp. 141–154, 2000.
- [60] R. G. Ross et al., "The effects of age on a smooth pursuit tracking task in adults with schizophrenia and normal subjects," *Biol. Psychiatry*, vol. 46, no. 3, pp. 383–391, 1999.
- [61] L. A. Abel, B. T. Troost, and L. F. Dell'Osso, "The effects of age on normal saccadic characteristics and their variability," *Vis. Res.*, vol. 23, no. 1, pp. 33–37, 1983.
- [62] J. C. R. Bergstrom, E. L. Olmsted-Hawala, and M. E. Jans, "Age-related differences in eye tracking and usability performance: Website usability for older adults," *Int. J. Human-Comput. Interaction*, vol. 29, pp. 541–548, 2013.
- [63] S. D. Walter, M. Eliasziw, and A. Donner, "Sample size and optimal designs for reliability studies," *Statist. Med.*, vol. 17, no. 1, pp. 101–110, 1998.
- [64] R. C. Oldfield, "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, 1971.
- [65] P. Frati et al., "Smart drugs and synthetic androgens for cognitive and physical enhancement: Revolving doors of cosmetic neurology," *Current Neuropharmacol.*, vol. 13, no. 1, pp. 5–11, 2015.
- [66] J. Kunimatsu and M. Tanaka, "Striatal dopamine modulates timing of self-initiated saccades," *Neuroscience*, vol. 337, pp. 131–142, Nov. 2016.
- [67] L. Birch, L. Arendt-Nielsen, T. Graven-Nielsen, and H. Christensen, "An investigation of how acute muscle pain modulates performance during computer work with digitizer and puck," *Appl. Ergon.*, vol. 32, no. 3, pp. 281–286, 2001.
- [68] A. Samani, A. Holtermann, K. Sogaard, and P. Madeleine, "Active pauses induce more variable electromyographic pattern of the trapezius muscle activity during computer work," *J. Electromyogr. Kinesiol.*, vol. 19, no. 6, pp. e430–e437, 2009.
- [69] R. Z. Marandi, A. Samani, and P. Madeleine, "The level of mental load during a functional task is reflected in oculometrics," in *Proc. IFMBE*, 2018, pp. 57–60.
- [70] K. H. E. Kroemer, H. B. Kroemer, and K. E. Kroemer-Elbert, *Ergonomics: How to Design for Ease and Efficiency*. London, U.K.: Pearson College Division, 2001.
- [71] Y. Chen and H. Sundaram, "Estimating complexity of 2D shapes," in *Proc. IEEE 7th Workshop Multimedia Signal Process.*, vol. 3, Nov. 2005, pp. 2–5.
- [72] C. Sun, Y. Wang, and J. Zheng, "Dissecting pattern unlock: The effect of pattern strength meter on pattern selection," *J. Inf. Secur. Appl.*, vol. 19, nos. 4–5, pp. 308–320, 2014.
- [73] D. Chaffin, G. Andersson, and B. Martin, *Occupational Biomechanics*. New York, NY, USA: Wiley, 1999.
- [74] M. Nyström and K. Holmqvist, "An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data," *Behav. Res. Methods*, vol. 42, no. 1, pp. 188–204, 2010.
- [75] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.
- [76] L. Larsson, M. Nystrom, and M. Stridh, "Detection of saccades and postsaccadic oscillations in the presence of smooth pursuit," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 9, pp. 2484–2493, Sep. 2013.
- [77] P. P. Caffier, U. Erdmann, and P. Ullsperger, "Experimental evaluation of eye-blink parameters as a drowsiness measure," *Eur. J. Appl. Physiol.*, vol. 89, nos. 3–4, pp. 319–325, 2003.
- [78] R. S. Hessels, D. C. Niehorster, C. Kemner, and I. T. C. Hooge, "Noise-robust fixation detection in eye movement data: Identification by two-means clustering (I2MC)," *Behav. Res. Methods*, vol. 49, no. 5, pp. 1802–1823, 2016.
- [79] H. Pashler et al., "Using task-induced pupil diameter and blink rate to infer cognitive load," *A Handbook Process Tracing Methods Decision*, vol. 88, no. 4, pp. 69–72, 2014.
- [80] J. F. Hopstaken, D. van der Linden, A. B. Bakker, and M. A. Kompier, "The window of my eyes: Task disengagement and mental fatigue covary with pupil dynamics," *Biol. Psychol.*, vol. 110, pp. 100–106, Sep. 2015.
- [81] T. Knapen, J. W. de Gee, J. Brascamp, S. Nuiten, S. Hoppenbrouwers, and J. Theeuwes, "Cognitive and ocular factors jointly determine pupil responses under equiluminance," *PLoS One*, vol. 11, no. 5, p. e0155574, 2016.
- [82] F. Chen, *Robust Multimodal Cognitive Load Measurement*. Cham, Switzerland: Springer, 2016.
- [83] J. Klingner, B. Tversky, and P. Hanrahan, "Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks," *Psychophysiology*, vol. 48, no. 3, pp. 323–332, 2011.
- [84] G. J. Siegle, N. Ichikawa, and S. Steinhauer, "Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses," *Psychophysiology*, vol. 45, no. 5, pp. 679–687, 2008.
- [85] J. G. May, R. S. Kennedy, M. C. Williams, W. P. Dunlap, and J. R. Brannan, "Eye movement indices of mental workload," *Acta Psychol.*, vol. 75, no. 1, pp. 75–89, 1990.
- [86] A. C. Smit and J. A. van Gisbergen, "A short-latency transition in saccade dynamics during square-wave tracking and its significance for the differentiation of visually-guided and predictive saccades," *Exp. Brain Res.*, vol. 76, no. 1, pp. 64–74, 1989.
- [87] M. Nakayama, K. Takahashi, and Y. Shimizu, "The act of task difficulty and eye-movement frequency for the 'Oculo-motor indices,'" in *Proc. ETRA*, 2002, pp. 37–42.
- [88] E. App and G. Debus, "Saccadic velocity and activation: Development of a diagnostic tool for assessing energy regulation," *Ergonomics*, vol. 41, no. 5, pp. 689–697, 1998.
- [89] I. P. Bodala, J. Li, N. V. Thakor, and H. Al-Nashash, "EEG and eye tracking demonstrate vigilance enhancement with challenge integration," *Frontiers Human Neurosci.*, vol. 10, p. 273, Jun. 2016.
- [90] A. T. Bahill, M. R. Clark, and L. Stark, "The main sequence, a tool for studying human eye movements," *Math. Biosci.*, vol. 24, nos. 3–4, pp. 191–204, 1975.
- [91] L. L. Di Stasi, A. Antolí, and J. J. Cañas, "Main sequence: An index for detecting mental workload variation in complex tasks," *Appl. Ergon.*, vol. 42, no. 6, pp. 807–813, 2011.
- [92] O. I. Jóhannesson and A. Kristjánsson, "Violating the main sequence: Asymmetries in saccadic peak velocities for saccades into the temporal versus nasal hemifields," *Exp. Brain Res.*, vol. 227, no. 1, pp. 101–110, 2013.
- [93] H. Cui, X.-H. Liu, K.-Y. Wang, C.-Y. Zhu, C. Wang, and X.-H. Xie, "Association of saccade duration and saccade acceleration/deceleration asymmetry during visually guided saccade in schizophrenia patients," *PLoS ONE*, vol. 9, no. 5, p. e97308, 2014.
- [94] M. Doyle and R. Walker, "Curved saccade trajectories: Voluntary and reflexive saccades curve away from irrelevant distractors," *Exp. Brain Res.*, vol. 139, no. 3, pp. 333–344, 2001.
- [95] J. A. Veltman and A. W. K. Gaillard, "Physiological workload reactions to increasing levels of task difficulty," *Ergonomics*, vol. 41, no. 5, pp. 656–669, 1998.
- [96] X. Li, W. Fang, and Y. Zhou, "Mental workload prediction model based on information entropy," *Comput. Assist. Surg.*, vol. 21, no. 1, pp. 117–124, 2016.
- [97] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Amer. J. Physiol. Heart Circuits Physiol.*, vol. 278, no. 6, pp. H2039–H2049, 2000.
- [98] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," *Phys. D, Nonlinear Phenomena*, vol. 65, pp. 117–134, May 1993.
- [99] S. Wallot, "Recurrence quantification analysis of processes and products of discourse: A tutorial in R," *Discourse Process.*, vol. 54, nos. 5–6, pp. 382–405, Jul. 2017.
- [100] C. L. Webber, Jr., and J. P. Zbilut, "Recurrence quantification analysis of nonlinear dynamical systems," *Tuts. Contemp. Nonlinear methods Behav. Sci.*, vol. 2, pp. 26–94, Mar. 2005.
- [101] M. U. Ahmed and D. P. Mandic, "Multivariate multiscale entropy analysis," *IEEE Signal Process. Lett.*, vol. 19, no. 2, pp. 91–94, Feb. 2012.
- [102] S. Wallot, A. Roepstorff, and D. Mønster, "Multidimensional recurrence quantification analysis (MdrQA) for the analysis of multidimensional time-series: A software implementation in MATLAB and its application to group-level data in joint action," *Frontiers Psychol.*, vol. 7, p. 1835, Nov. 2016.
- [103] C. Green and D. Bavelier, "Action video game modifies visual selective attention," *Nature*, vol. 423, no. 6939, pp. 534–537, 2003.
- [104] J. Moran and R. Desimone, "Selective attention gates visual processing in the extrastriate cortex," *Science*, vol. 229, no. 4715, pp. 782–784, 1985.
- [105] D. Sharek, "A useable, online NASA-TLX tool," *Proc. Human Factors Ergon. Soc. Annu. Meeting*, vol. 55, no. 1, pp. 1375–1379, 2011.
- [106] G. Atkinson and A. M. Nevill, "Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine," *Sports Med.*, vol. 26, no. 4, pp. 217–238, 1998.

- [107] J. R. Landis and G. G. Koch, "Agreement measures for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
- [108] N. Becser, T. Sand, and J. A. Zwart, "Reliability of cephalic thermal thresholds in healthy subjects," *Cephalalgia*, vol. 18, no. 8, pp. 574–582, 1998.
- [109] L. A. H. Critchley and J. A. J. H. Critchley, "A meta-analysis of studies using bias and precision statistics to compare cardiac output measurement techniques," *J. Clin. Monitor. Comput.*, vol. 15, no. 2, pp. 85–91, 1999.
- [110] M. K. Eckstein, B. Guerra-Carrillo, A. T. M. Singley, and S. A. Bunge, "Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development?" *Develop. Cognit. Neurosci.*, vol. 25, pp. 69–91, Jun. 2017.
- [111] K. F. Van Orden, W. Limbert, S. Makeig, and T. P. Jung, "Eye activity correlates of workload during a visuospatial memory task," *Human Factors*, vol. 43, no. 1, pp. 111–121, 2001.
- [112] J. Klingner, "Measuring cognitive load during visual tasks by combining pupillometry and eye tracking," Ph.D. dissertation, Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, 2010.
- [113] M. L. Mele and S. Federici, "Gaze and eye-tracking solutions for psychological research," *Cognit. Process.*, vol. 13, no. 1, pp. S261–S265, 2012.
- [114] M. Borys *et al.*, "An analysis of eye-tracking and electroencephalography data for cognitive load measurement during arithmetic tasks," in *Proc. 10th Int. Symp. Adv. Topics Elect. Eng. (ATEE)*, 2017, pp. 287–292.
- [115] L. L. Di Stasi *et al.*, "Saccadic peak velocity sensitivity to variations in mental workload," *Aviat. Space Environ. Med.*, vol. 81, no. 4, pp. 413–417, 2010.
- [116] M. A. Basso and R. H. Wurtz, "Modulation of neuronal activity by target uncertainty," *Nature*, vol. 389, no. 6646, pp. 66–69, 1997.
- [117] U. J. Ilg, Y. Jin, S. Schumann, and U. Schwarz, "Preparation and execution of saccades: The problem of limited capacity of computational resources," *Exp. Brain Res.*, vol. 171, no. 1, pp. 7–15, 2006.
- [118] R. A. Abrams, J. Pratt, and A. L. Chasteen, "Aging and movement: Variability of force pulses for saccadic eye movements," *Psychol. Aging*, vol. 13, no. 3, pp. 339–387, 1998.
- [119] M. Causse, B. Baracat, J. Pastor, and F. Dehaiss, "Reward and uncertainty favor risky decision-making in pilots: Evidence from cardiovascular and oculometric measurements," *Appl. Psychophysiol. Biofeedback*, vol. 36, no. 4, pp. 231–242, 2011.
- [120] L. C. McPhee, C. T. Scialfa, W. M. Dennis, G. Ho, and J. K. Caird, "Age differences in visual search for traffic signs during a simulated conversation," *Human Factors*, vol. 46, no. 4, pp. 674–685, 2004.
- [121] D. J. Madden, "Aging and visual attention," *Current Direction Psychol. Sci.*, vol. 16, no. 2, pp. 70–74, 2007.
- [122] R. Cabeza, S. M. Daselaar, F. Dolcos, S. E. Prince, M. Budde, and L. Nyberg, "Task-independent and task-specific age effects on brain activity during working memory, visual attention and episodic retrieval," *Cereb. Cortex*, vol. 14, no. 4, pp. 364–375, 2004.
- [123] A. Poole and L. J. Ball, "Eye tracking in human-computer interaction and usability Research: Current status and future. Prospects," in *Encyclopedia of Human-Computer Interaction*, C. Ghaoui, Ed. Doylestown, PA, USA: Idea Group, Inc, 2005.
- [124] J. A. Veltman and A. W. K. Gaillard, "Physiological indices of workload in a simulated flight task," *Biol. Psychol.*, vol. 42, no. 3, pp. 323–342, 1996.
- [125] E. Siegenthaler *et al.*, "Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes," *Eur. J. Neurosci.*, vol. 39, no. 2, pp. 287–294, 2014.
- [126] A. A. Zekveld, S. E. Kramer, and J. M. Festen, "Pupil response as an indication of effortful listening: The influence of sentence intelligibility," *Earth Heart*, vol. 31, no. 4, pp. 480–490, 2010.
- [127] N. Attar, M. H. Schneps, and M. Pomplun, "Working memory load predicts visual search efficiency: Evidence from a novel pupillary response paradigm," *Mem. Cognit.*, vol. 44, no. 7, pp. 1038–1049, 2016.
- [128] G. Porter, T. Troscianko, and I. D. Gilchrist, "Effort during visual search and counting: Insights from pupillometry," *Quart. J. Exp. Psychol.*, vol. 60, no. 2, pp. 211–229, 2007.
- [129] M. Ranchet, J. C. Morgan, A. E. Akinwuntan, and H. Devos, "Cognitive workload across the spectrum of cognitive impairments: A systematic review of physiological measures," *Neurosci. Biobehav. Rev.*, vol. 80, pp. 516–537, Sep. 2017.
- [130] N. M. Wöstmann, D. S. Aichert, A. Costa, K. Rubia, H. J. Müller, and U. Ettinger, "Reliability and plasticity of response inhibition and interference control," *Brain Cognit.*, vol. 81, no. 1, pp. 82–94, 2013.
- [131] R. J. van Beers, "The sources of variability in saccadic eye movements," *J. Neurosci.*, vol. 27, no. 33, pp. 8757–8770, 2007.
- [132] L. Mesin, A. Monaco, and R. Cattaneo, "Investigation of nonlinear pupil dynamics by recurrence quantification analysis," *Biomed. Res. Int.*, vol. 2013, Aug. 2013, Art. no. 420509.
- [133] P. Vaidyanathan, J. Pelz, C. Alm, P. Shi, and A. Haake, "Recurrence quantification analysis reveals eye-movement behavior differences between experts and novices," in *Proc. Symp. Eye Tracking Res. Appl., Safety Harbor*, Mar. 2014, pp. 303–306.
- [134] N. C. Anderson, F. Anderson, A. Kingstone, and W. F. Bischof, "A comparison of scanpath comparison methods," *Behav. Res. Methods*, vol. 47, no. 4, pp. 1377–1392, 2015.
- [135] G. J. F. Solman and A. Kingstone, "Endogenous strategy in exploration," *J. Exp. Psychol. Human Perception Perform.*, vol. 41, no. 6, pp. 1634–1649, 2015.
- [136] M. Raemaekers, M. Vink, M. P. van den Heuvel, R. S. Kahn, and N. F. Ramsey, "Effects of aging on BOLD fMRI during prosaccades and antisaccades," *J. Cognit. Neurosci.*, vol. 18, no. 4, pp. 594–603, 2006.
- [137] M. A. S. Boksem, T. F. Meijman, and M. M. Lorist, "Mental fatigue, motivation and action monitoring," *Biol. Psychol.*, vol. 72, no. 2, pp. 123–132, 2006.



**RAMTIN ZARGARI MARANDI** received the M.Sc. degree in biomedical engineering from Iran University of Science and Technology, Iran, in 2013, and the B.Sc. degree in electrical engineering from Shahid Rajaei University, Iran, in 2008. He is currently pursuing the Ph.D. degree with the Department of Health Science and Technology, Aalborg University (AAU), Denmark, jointly with the Université Grenoble Alpes, France. Since then, he has been a member of

Ergonomics and Work-related Disorders Laboratory, AAU. His current research interests include machine learning, eye movements, and cognitive science.



**PASCAL MADELEINE** was born in Toulouse, France, in 1969. He received the Ph.D. and Dr.Sc. degrees from Aalborg University, Denmark. He is currently a Professor in sports and ergonomics with the Department of Health Science and Technology, Aalborg University, where he is also the Head of Sport Sciences Group. He has published over 190 peer-reviewed scientific journal publications and book chapters. His current research interests include the development and application of novel methods, and technologies in ergonomics and sport sciences.



**ØYVIND OMLAND** received the M.D. and Ph.D. degrees from the Aarhus University, Denmark, in 1980 and 2001, respectively. Since 1991, he has been a Specialist in occupational medicine, from 1999 at the Occupational Clinic, Aalborg University Hospital, as a Consultant, and since 2013 as a Professor. He has authored or co-authored several articles mainly in the field of occupational medicine. His current research interests include environmental and occupational lung diseases including cancer and aging of workforce and ability to maintain physical hard work.



**NICOLAS VUILLERME** received the Ph.D. degree in sciences and technologies in physical activities and sports with speciality in neurosciences and the Habilitation degree in medicine from the Joseph Fourier University of Grenoble, France, in 2002 and 2010, respectively.

He is currently an Assistant Professor in human movement sciences with the Université Grenoble Alpes, France, and an Adjunct Professor in ergonomics and sport with Aalborg University,

Denmark. He is an Honorary Member of the Institut Universitaire de France (promotion IUF 2012).

His research and teaching concern e-Health, human factors, ergonomics, neurosciences, and sport sciences. His ongoing works also include activities of transfer of technology and know-how, and industrial valorization of research and development results, and creation of innovative technology businesses.



**AFSHIN SAMANI** received the Ph.D. degree in biomedical engineering and science from Aalborg University, Denmark, in 2010. He is currently an Associate Professor in sports science and ergonomics with the Department of Health Science and Technology, Aalborg University. He is the Director of the Laboratory for Ergonomics and Work-Related Disorders. His current research interests include the methods of quantification of work exposures and risk factors for the develop-

ment of musculoskeletal disorders in computer users.

...