POLYTECHNIC OF TURIN

Master's Degree in Biomedical Engineering



Politecnico di Torino

Master's Degree Thesis

A light artefact-free method for real-world cognitive load monitoring through automated pupillometry: preliminary application in people with vestibular deficits and older adults

Supervisors

Prof. Andrea CEREATTI

Dr. Diletta BALTA

Dr. Marco CARUSO

Dr. Rachele ROSSANIGO

Ing. Paolo TASCA

Candidate

MATTEO ZICCHIERI

September 2026

Dedicato ai miei genitori

AI Use Disclosure

In the writing process of this thesis the following tools have been used to improve the readability and language:

- Claude (Anthropic) and NotebookLM (Google) for text generation and paraphrasing;

Turin, 28/09/2026

Abstract

The indirect estimation of cognitive workload offers valuable insight into a person's mental state across many tasks. It can be assessed with subjective scales, which lack temporal resolution, or neurophysiological techniques such as EEG and fNIRS, requiring complex setups and sensitive to motion artifacts. Pupillometry through smart glasses offers an objective, more practical alternative that also works during movement, exploiting task-evoked pupil dilations as a biomarker of cognitive load. This may help uncover the hidden attentional cost of movement when motor automaticity is reduced. However, isolating the cognitive response remains challenging, as ambient light variations elicit a larger pupil response than cognitive load. Consequently, most studies rely on strictly controlled lighting, reducing ecological validity.

This thesis investigated whether the cognitive component of pupil diameter can be isolated under uncontrolled lighting by modelling and removing the light-driven component, and if the resulting estimate is sensitive to the attentional demand of increasingly complex tasks in older adults and vestibular patients.

The study involved 12 participants: 7 older adults (66–82 y.o.) and 5 patients with unilateral or bilateral vestibular deficits (40–75 y.o.), equipped with eye-tracking glasses (Tobii Pro Glasses 3) custom-integrated with a synchronized ambient light sensor. Under natural laboratory lighting, they performed four tasks: a 4-m U-turn walk repeated twice, with and without a concurrent visual ball-tracking task; a simulated daily activity (SDA), requiring memorizing and executing a sequence of mobility and manual tasks; and the cognitive MoCA test. The cognitive workload associated with each task was self-assessed using the NASA-Task Load Index questionnaires.

Gaze-point luminance was estimated by combining scene-camera pixel intensity with light sensor measurements, following Pignoni et al. (2021). A pupillary-luminance reflex model then predicted the light-induced response, subtracted from the measured pupil diameter to obtain the residual dilation attributed to cognitive load.

Pearson correlation between estimated luminance and both the overall pupil diameter and its isolated cognitive component was computed. The cognitive component was then compared across tasks and groups, and against the NASA-TLX mental demand scores.

The overall pupil diameter correlated negatively with estimated luminance in both groups ($r = -0.60$, $p < 0.001$ in older adults; $r = -0.46$, $p < 0.01$ in vestibular patients), consistent with the expected pupillary constriction under stronger light. This correlation decreased by 54–88% for the cognitive component, as expected if

it reflects cognitive rather than light-driven fluctuations.

In both groups the cognitive pupillary component increased monotonically across tasks, from walking alone through walking with visual ball tracking and the SDA up to the MoCA, mirroring the NASA-TLX Mental Demand ratings and showing that pupillometric cognitive load aligns with self-reports. The MoCA elicited the largest cognitive component in both cohorts; however, while in the older adults it clearly stood out as the most demanding task, in the vestibular patients the SDA rose to nearly match it, suggesting that cognitive-motor interference is comparatively more demanding under vestibular impairment. Despite the small sample, the estimated cognitive dilation tracked the increasing task complexity and matched the perceived difficulty across varying ambient light, preliminarily supporting the potential of corrected pupillometry as an objective and ecologically valid measure of cognitive workload.

Acknowledgements

I would like to thank Professor Cereatti for giving me the opportunity to work on this project and to join this research group.

I would also like to thank Daniele, Diletta, Rachele, Marco, and Paolo for guiding me and placing their trust in me throughout this project. Your constant support, availability, and valuable advice during our endless meetings helped me bring this journey to completion.

A special thanks goes to Dr. Boldreghini for supporting me throughout the data acquisition campaign, for making the equipment and facilities of his laboratory available, and for the continuous and always constructive discussions that helped shape the direction of this work.

Last but not least, I would like to thank my amazing friends Mattia, Thea, Milena, Alessio, Alessandra A., and Alessandra B. Every break and data acquisition session we shared helped make this experience lighter, both in moments of joy and in times of difficulty. I'm grateful for your friendship and I wish you all the best.

Table of Contents

List of Tables x

List of Figures xii

1 Introduction 1
- 1.1 Cognitive workload: definition and relevance . . . 1
- 1.2 Assessing cognitive workload . . . 1
- 1.3 Pupillometry as a window on cognitive load . . . 2
- 1.4 The challenge of movement: cognitive-motor interference (CMI) . . . 3
- 1.5 Open challenges and thesis objectives . . . 4
- 1.6 Thesis outline . . . 4

2 Background 6
- 2.1 Pupil physiology . . . 6
  - 2.1.1 Pupillary light reflex (PLR) . . . 7
  - 2.1.2 Task-evoked pupillary response (TEPR) . . . 8
- 2.2 Pupil diameter measurement with infrared eye trackers . . . 9
- 2.3 Pupillometry in dynamic tasks . . . 10
- 2.4 PLR mitigation strategies . . . 11
  - 2.4.1 Baseline-related methods . . . 12
  - 2.4.2 Frequency domain methods . . . 12
  - 2.4.3 Brightness compensation methods . . . 13

3 Pupil diameter measurement validation 15
- 3.1 Experimental setup . . . 16
  - 3.1.1 Tobii Pro Glasses 3 . . . 16
  - 3.1.2 Videonystagmograph Natus ICS Dizcovery (VNG) . . . 16
  - 3.1.3 Session setup . . . 18
- 3.2 Acquisition protocol . . . 19
- 3.3 Signal processing . . . 19
  - 3.3.1 Manual segmentation . . . 19

3.3.2 Pixel-to-millimetre conversion . . . 20
3.3.3 Data synchronisation . . . 22
3.4 Metrics . . . 22

4 Ambient light sensor integration 24
4.1 Market survey . . . 24
4.2 ALS integration with INDIP system . . . 25
4.3 Sensor characterisation . . . 27
4.3.1 Acquisition protocol . . . 27
4.3.2 Linearity assessment . . . 28
4.3.3 Conversion factors and non-linearity correction . . . 31
4.3.4 Validation against the reference luxmeter . . . 33
4.4 Auto-ranging logic . . . 33
4.5 Case and positioning . . . 37
4.6 Data synchronisation with the eye tracker . . . 38

5 Cognitive workload study 41
5.1 Experimental design . . . 41
5.1.1 Study participants . . . 41
5.1.2 Experimental setup . . . 42
5.1.3 Experimental protocol . . . 44
5.2 Estimation method . . . 50
5.2.1 Overview of the estimation method . . . 50
5.2.2 Pupil diameter signal pre-processing . . . 51
5.2.3 ALS signal pre-processing . . . 52
5.2.4 Pignoni's method (2021) . . . 54
5.2.5 Baseline calibration and cognitive load estimation . . . 57
5.3 Data analysis . . . 59
5.3.1 Sub-phase segmentation (SDA and MoCA) . . . 59
5.3.2 Statistical analysis . . . 60

6 Results 63
6.1 Pupil diameter measurement validation . . . 63
6.1.1 Absolute diameter agreement . . . 63
6.1.2 Temporal tracking of pupillary fluctuations . . . 64
6.2 Assessment of pupil components separation . . . 67
6.2.1 Model accuracy over the ambient-light range . . . 67
6.2.2 Correlation of the pupillary components with luminance . . . 70
6.3 Cognitive workload analysis . . . 72
6.3.1 Comparison across tasks and cohorts . . . 72
6.3.2 Association with subjective ratings (NASA-TLX) . . . 74

6.3.3 Subphase analysis 78

7 Discussion 82
7.1 Reliability of Tobii Pro Glasses 3 pupil measurements 82
7.1.1 Agreement and temporal tracking results 82
7.1.2 Limitations of the validation protocol 83
7.2 Isolating the cognitive component under uncontrolled lighting . . . 85
7.2.1 Model accuracy and robustness across the luminance range . 85
7.2.2 Correlation analysis results 86
7.3 Sensitivity of the cognitive component to task demand and impairment 87
7.3.1 Pupil cognitive component across tasks 87
7.3.2 Association with subjective workload ratings 88
7.3.3 Subphase analysis 89
7.4 Limitations 91
7.5 Future work 92

8 Conclusions 93

Bibliography 94

List of Tables

3.1 The two systems compared. 18

4.1 Comparison of the ambient light sensors considered in the market survey. 25
4.2 Data fields added to the INDIP sample record to accommodate the ALS. 26
4.3 Nominal resolution (conversion factor, lx/count) of the VEML7700 as a function of gain and integration time, as declared by the manufacturer [86]. The row corresponding to the integration time adopted in this work is highlighted. 27
4.4 Summary of the calibration measurements. For each gain setting and each version of the setup, the number of paired acquisitions and the corresponding illuminance range measured by the reference luxmeter are reported. 28
4.5 Conversion factors obtained from the calibration at 25 ms integration time, compared with the nominal values declared by the manufacturer [86]. 32
4.6 Coefficients of the fourth-order correction of Equation (4.3), obtained by fitting the experimental data at 25 ms integration time, together with the centring and scaling parameters of the predictor. 32
4.7 Trade-off between resolution and maximum measurable illuminance for the three gain settings used in this work, at an integration time of 25 ms [86]. 33

5.1 Number of participants recruited and included in the final analyses. 42
5.2 Symbols used in the description of the estimation method. 52

6.1 Per-recording metrics (participant × eye). L = left; R = right; PD = pupil diameter; RMSE = root mean square error; r = Pearson correlation. 66

6.2 Aggregate metrics across participants (per-recording medians), by eye; LoA = limits of agreement; SD = standard deviation. 66
6.3 Accuracy of the calibrated Watson–Yellott prediction on the personalisation test, by ambient-light band. N = number of steady plateaus pooled across subjects; MAE = mean absolute residual; rel. error = mean absolute relative error. Each band is mapped to a typical everyday lighting condition, for context. 69
6.4 Pearson correlation with luminance (mean ± SD across subjects) for the three pupillary components, pooled over the four U-turn walking recordings and reported separately for the SDA ($n = 7$ healthy, $n = 5$ vestibular). 71
6.5 Pupil diameter and cognitive component per task and cohort (median [IQR] of the per-subject values; N varies for the vestibular group owing to missing recordings in two tasks). 74
6.6 Cognitive component ΔD_C (mm) in the seven SDA sub-phases: group median [IQR] across subjects ($n = 7$ healthy, $n = 5$ vestibular). 80

List of Figures

2.1 Constricted and dilated pupil captured by the eye camera 9
2.2 Examples of remote and wearable eye trackers 10
2.3 Wearable acquisition setup 14

3.1 The TG3: the head unit, shaped as a spectacle frame, and the cabled recording unit worn at the belt. 17
3.2 The Natus ICS Dizcovery videonystagmography goggle mask, used as the reference system in the validation. 17
3.3 Example of a raw eye frame recorded by the videonystagmograph, from which the pupil was segmented manually. 20
3.4 Manual segmentation of the pupil, illustrated on a single frame. (a) Raw eye frame. (b) Binary mask traced by the operator. (c) Ellipse having the same normalised second central moments as the mask. (d) Ellipse and its major axis superimposed on the eye frame, with the segmented contour in blue and the centroid marked; the length of the major axis is taken as the pupil diameter, in pixels. 21

4.1 The VEML7700 sensor connected to an INDIP unit via a flat ribbon cable. The bare board sits inside its 3D-printed case, deliberately left open at the front (Section 4.5). 26
4.2 Instruments and calibration setup. (a) The Vishay VEML7700 on its Adafruit breakout board, reproduced from the manufacturer's product page [89]. (b) The Delta OHM HD2302.0 photo-radiometer with its family of interchangeable probes, reproduced from the manufacturer's datasheet [87]; the probe used in this work is the LP471PHOT. (c) The shielding box positioned under the light source, with the HD2302.0 unit visible on the left. (d) Inside of the box seen from above: the sheet of graph paper fixed to the base allows the sensor position to be reproduced across successive measurements. 29

4.3 Raw VEML7700 output (counts) against the illuminance measured by the reference luxmeter, for the three gain settings and the four setups, on logarithmic axes. The dashed line is the ideal proportional response. Gain 1 departs from linearity above approximately 100 lx; gains 1/4 and 1/8 remain proportional up to about 1,000 lx and compress beyond 10 klx. 30
4.4 Cumulative linearity analysis. For each gain, the coefficient of determination R^2 of a first-order fit is plotted against the upper limit of the illuminance range included in the fit. Colours denote the setup to which the last point added belongs. Gain 1 loses linearity approximately one decade earlier than gains 1/4 and 1/8. 30
4.5 Effect of the polynomial correction at gains 1/4 and 1/8. Each panel compares the reference reading with the estimate obtained from the factory conversion factors, the estimate obtained with the calibrated factor k alone, and the final estimate including the fourth-order correction. The dashed line is the identity. 32
4.6 Agreement between the calibrated VEML7700 and the reference luxmeter over the whole calibration dataset, with the gain selected according to the auto-ranging logic. Colours denote the four setups. (a) Difference between the two instruments against their mean, with bias and limits of agreement. (b) Relative error against the reference reading; the dotted lines mark the $\pm 10\,\%$ band and the solid line the mean absolute percentage error. 34
4.7 Mean static pupil diameter as a function of ambient illuminance, obtained by collecting the values reported in five independent studies [90, 91, 92, 93, 35]. Error bars are the standard deviations reported in the original works. The curve is steep at low illuminance and flattens towards photopic levels, so that measurement resolution on the illuminance estimate matters most in the low range. 35
4.8 Flowchart of the auto-ranging logic, with the conversion factors obtained from the calibration and the correction formula applied at gains 1/4 and 1/8. 36
4.9 Response of the complete measurement chain under the auto-ranging logic: calibrated illuminance against the reference reading, on logarithmic axes, with the ideal one-to-one relationship shown as a dashed line. 37

4.10 Enclosure and positioning of the ALS. (a) The two assembled units, seen from the front (top) and from the back (bottom), each in its 3D-printed enclosure and connected to an INDIP unit by a ribbon cable; only one unit was used during the recordings, the second being kept as a spare. The front face is fully open, leaving the sensing element unobstructed. (b) The sensor in place on a participant, mounted on the temple alongside the TG3, close to the scene camera and to the eye. 38
4.11 Head angular velocity before and after synchronisation 40

5.1 Complete instrumentation of a participant: (a) posterior view and (b) anterior view, showing the head unit (ALS and TG3), the lower-back unit and the foot sensors; (c) the left-foot instep sensor and (d) the right-foot sensor. 43
5.2 Locomotor task. Participants walked a figure-of-eight path around two cones placed 4 m apart, completing five loops around each cone, first clockwise (left) and then counter-clockwise (right). The arrows indicate the walking direction. 45
5.3 Locomotor task with tracking. The same figure-of-eight walk was combined with a tracking task: on each loop a ball was thrown at random to the left or to the right (top) and the participant had to follow it. The task was performed clockwise (left) and counter-clockwise (right). 46
5.4 Simulated daily-living sequence (SDA). The dashed line traces the route through the workspace – chair, table, three-cone slalom and obstacle – and the numbered markers indicate the order of the phases carried out by the participant, described in the text. 47
5.5 Cognitive task (MoCA). The test was administered by a clinician while the participant, seated, fixated a red target on a sheet placed in front of them. 48
5.6 The five background levels of the fixation video, from black to white through three intermediate greys (normalised RGB values 0, 0.25, 0.5, 0.75 and 1). The red dot at the centre is the fixation target, held for the whole recording. 49
5.7 A participant during the personalisation test: (a) in the initial dark-room phase and (b) after the room lights were switched on. . . 49

5.8 Example recording from one personalisation test: the measured pupil diameter – right eye in red, left eye in blue – (left axis) and the illuminance recorded by the ALS (right axis) over time. The dashed lines delimit the eight stimulus phases – five background levels, two lighting steps and the raising of the blinds, sketched at the top. As the illuminance rises, the pupil constricts, illustrating the light-driven response that the test is designed to isolate. 50
5.9 Overview of the cognitive-component estimation pipeline 51
5.10 Pupil-diameter pre-processing pipeline. Left: block diagram of the processing steps, from the raw pupil diameter to the pre-processed signal D_T. Right: the effect of each step on a representative segment of the personalisation-test recording of participant 4009 (raw signal, blink reconstruction, velocity-threshold filtering with the removed samples marked in red, and low-pass filtering). 53
5.11 Ambient-light-sensor processing pipeline. Left: block diagram of the processing steps, from the raw sensor counts to the ambient luminance *avgL*. Right: the successive steps applied to an example signal, the full personalisation-test recording of participant 4009. . . 54
5.12 Scene-camera luminance-extraction pipeline. Left: block diagram of the processing steps. Right: three representative frames from the personalisation-test recording of participant 4009, illustrating frame extraction, the gaze point located in each frame, and the Magic-Wand area of interest grown around it. 56
5.13 Example output of the estimation method on the locomotor task of participant 6011, a healthy older adult. Top: the measured pupil diameter D_T, the baseline-calibrated light-driven component $D_{L,\mathrm{corr}}$ and the resulting cognitive component $D_C = D_T - D_{L,\mathrm{corr}}$; the shaded interval is the baseline used for the calibration. Bottom: the gaze-point (AOI) luminance over the same interval. The cognitive component stays close to zero and, unlike the total diameter, does not follow the oscillations of the ambient luminance. 59
5.14 Manual sub-phase labelling in LosslessCut, on the TG3 scene-camera video: (a) during the MoCA, and (b) during the SDA, as the participant approaches the slalom (the obstacle is visible in the background). The bottom bar shows the video timeline with all the labelled intervals; the panel on the right lists the corresponding segments, later exported to an Excel file read by the MATLAB labelling code. . . . 60

6.1 Bland–Altman analysis at the participant level (per-recording medians), separately for the left and the right eye. Lines: bias and limits of agreement (±1.96 SD); the ordinate axis is common to the two eyes. 64

6.2 Time course of the pupil diameter for each participant and eye: TG3 (blue) and manual VNG converted with the pixel-to-millimetre conversion factor (orange). Line interruptions correspond to discarded frames (blinks); the time axis starts at the first segmented Natus frame. 65
6.3 Plateau-based validation on the personalisation test (subject 4009, Test 19). The measured pupil (blue) and the baseline-calibrated prediction $D_{L,\mathrm{corr}}$ (green) are shown against the staircase luminance (orange, right axis), spanning approximately $0\,\mathrm{cd\,m^{-2}}$ to $195\,\mathrm{cd\,m^{-2}}$. Shaded bands mark the stable luminance plateaus retained (red segments) for the accuracy analysis. 68
6.4 Model accuracy versus ambient light level (log scale), pooled across subjects (coloured markers: mean ± SD per bin; grey dots: individual subjects). (A) mean relative error (%); (B) mean residual $D_T - D_{L,\mathrm{corr}}$ (mm). The prediction is accurate above about 3.4 lx (mean absolute residual $\leq 0.14\,\mathrm{mm}$, a few per cent) and increasingly under-estimates the pupil below it, reaching 0.81 mm (24.8 %) in the darkest band. . 69
6.5 Pearson correlation (mean ± SD across subjects) between luminance and the three pupillary components – total diameter (D_T), light-driven component (D_L) and cognitive component (D_C) – for each task (the four U-turn walking recordings and the SDA), in healthy older adults (top) and vestibular patients (bottom). D_T and D_L are strongly anti-correlated with luminance and vary with the light content of the task, whereas D_C stays close to zero and stable across tasks. 71
6.6 Per-task comparison of the raw pupil diameter D_T (left) and the extracted cognitive component D_C (right), by cohort (older adults in blue, vestibular patients in red) across four tasks (UTW, UTWT, SDA and MoCA); for the two U-turn walking tasks, the clockwise and counter-clockwise recordings were averaged per subject before plotting. Boxes show the distribution of the per-subject medians and the dots the individual subjects; the numbers above each panel are the mean illuminance ± SD (lx) recorded during the test, for that task and cohort (healthy top, vestibular bottom). The large between-cohort gap in the raw diameter, driven by differences in light exposure, collapses once the light component is removed, leaving a consistent task gradient shared by both cohorts. 73

6.7 Cognitive component versus NASA-TLX subscales in the healthy older adults, for the three tasks rated with the questionnaire (the UTWT, the SDA and the MoCA). Each panel overlays the cognitive component D_C (left axis, blue) and one NASA-TLX dimension (right axis, black), both as median [IQR] across participants. The Mental Demand subscale (top left) follows the cognitive component, both peaking at the MoCA. 76
6.8 Cognitive component versus NASA-TLX subscales in the vestibular patients, for the UTWT, the SDA and the MoCA. Panels and axes are as in Figure 6.7. Here both the Mental Demand subscale and the cognitive component show the SDA and the MoCA at nearly the same level, unlike the clear MoCA peak seen in the healthy group. 77
6.9 Dynamic analysis of the cognitive component during the MoCA, for the healthy older adults (top, $n = 7$) and the vestibular patients (bottom, $n = 5$). Left: per-exercise cognitive component split into briefing (open markers) and execution (filled markers) and connected in chronological order (grey/white bands = briefing/execution); faded dots are individual subjects, the line is the group mean. Right: macro-phase summary (baseline, briefing, execution). Exercise labels denote the MoCA macro-area: VSE = visuospatial/executive (VSE1 trail-making, VSE2 cube copy, VSE3 clock drawing), DEN = naming, MEM = memory, ATT = attention (ATT1 forward and ATT2 backward digit span, ATT3 target detection, ATT4 serial subtraction), LIN = language (LIN1 sentence repetition, LIN2 verbal fluency), AS = abstraction, RD = delayed recall. 79
6.10 Cognitive component ΔD_C in the seven SDA sub-phases, by cohort (older adults in blue, $n = 7$; vestibular patients in red, $n = 5$). Boxes show the distribution of the per-subject medians, dots the individual subjects, and the solid line the group median. The cognitive component is lowest in the passive listening phase and higher in the phases involving locomotion and balance control. 81

7.1 Temporal tracking: high- and low-correlation examples 84

Chapter 1

Introduction

1.1 Cognitive workload: definition and relevance

A widely accepted definition of mental workload is that proposed by Young and Stanton [1], who define it as «the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and past experience».

A non-optimal workload can be divided into overload, when task demands exceed the individual's available resources, leading to inefficiency and performance deterioration, and underload, when resources are underutilized, causing reduced vigilance and distraction [2].

Mental effort investment has been described as a form of energy mobilisation, representing a compensatory strategy used to protect performance under increased task demands [3]. Investigating cognitive workload therefore opens a window onto how much such compensatory effort an individual is investing to sustain a given level of performance. The aim of this study is specifically to investigate the allocation of additional mental resources by at-risk populations, such as individuals with vestibular dysfunction [4], during the execution of tasks that are normally automatic in young, healthy adults, such as walking and maintaining balance. Identifying a greater allocation of cognitive resources in these populations could help predict fall risk and inform the design of tailored rehabilitation strategies [5, 6].

1.2 Assessing cognitive workload

Several methods can be used to assess cognitive workload; broadly, they can be classified into two categories: self-report questionnaires, in which individuals subjectively report their own perceived difficulty of the task performed, and physiology-based

measures, which instead serve as objective instruments for measuring mental effort.

- Self-report instruments, most notably the NASA Task Load Index (NASA-TLX) [7], ask participants to rate their perceived workload after completing a task; they are inexpensive and easy to administer, but retrospective and susceptible to memory and rating bias.
- Cardiovascular measures such as heart rate and heart rate variability (HRV) can be recorded continuously and unobtrusively, but respond slowly to changing task demands and are strongly confounded by physical exertion.
- Electroencephalography (EEG) offers millisecond-level temporal resolution and direct access to neural activity, but is highly susceptible to motion and muscular artifacts, which limits its use in movement-based protocols.
- Functional near-infrared spectroscopy (fNIRS) indirectly indexes cortical activity through hemodynamic changes and tolerates motion better than EEG, but has a slower response and coarser spatial resolution.
- Eye tracking, and pupillometry in particular, estimates cognitive workload from oculomotor behavior and pupil diameter, modulated by brainstem nuclei tied to arousal and attentional control [8]; given its central role in this thesis, it is discussed in depth in Section 1.3.

1.3 Pupillometry as a window on cognitive load

Among the physiological measures reviewed above, pupillometry applications have attracted growing interest, driven in part by technical advances in eye-tracking technology. Pioneering studies by Hess and Polt [9] and Kahneman and Beatty [10] first demonstrated that pupil diameter increases systematically with the difficulty of a mental task, establishing the pupil as one of the earliest recognized psychophysiological correlates of cognitive effort. Subsequent work extended this finding to a wide range of paradigms, from information processing and memory load [11, 12] to individual differences in cognitive capacity [13] and clinical populations [14].

From a practical standpoint, pupillometry combines several advantages that make it particularly well suited to the aims of this thesis. It offers high temporal resolution, allowing workload to be tracked continuously as it fluctuates within a task rather than only assessed retrospectively. Modern video-based eye trackers, including wearable devices, allow pupil diameter to be recorded unobtrusively and without physical contact with the participant [15], and pupillometric measures have shown good convergent validity with established subjective measures such as the NASA-TLX [16]. This is especially valuable in populations such as older adults or

vestibular patients, for whom additional instrumentation can be uncomfortable or can itself interfere with the motor task under investigation.

At the same time, pupillometry carries a well-known limitation: pupil diameter reflects not only cognitive and attentional processes but also the pupillary light reflex (PLR), a much larger-amplitude response driven by retinal illuminance. Any attempt to use pupil diameter as a marker of mental effort must therefore isolate the cognitive component of the signal from the light-driven one — a methodological challenge that is central to this thesis and that is addressed in detail in Chapter 2.

1.4 The challenge of movement: cognitive-motor interference (CMI)

Pupillometry has typically been validated in seated, stationary settings, where the only task the participant performs is the cognitive one under study. Many everyday activities, however, require the simultaneous performance of a cognitive task and a motor task — for example, planning a conversation while walking, or maintaining balance while attending to instructions. The mutual interference that arises when cognitive and motor processes compete for shared attentional resources is commonly termed cognitive-motor interference (CMI).

CMI is typically studied through dual-task paradigms, in which participants perform a motor task alone and then again while simultaneously performing a cognitive task. Differences in performance between the single- and dual-task conditions — on either task, or both — are taken as evidence of interference between the two domains.

CMI is of particular relevance to populations in which either the motor or the cognitive domain is already compromised. In older adults, age-related decline in attentional and executive resources reduces the capacity to allocate resources to a secondary task [17], so that dual-tasking often produces disproportionately large costs to gait and balance compared to younger adults [18] — a phenomenon with direct implications for fall risk [19].

In patients with vestibular dysfunction, the picture is further complicated by the fact that maintaining balance itself becomes more attention-demanding, since the compromised vestibular system requires greater reliance on visual and cognitive strategies to compensate for the missing or degraded vestibular input [20, 21]. Growing evidence indicates that vestibular disorders are also associated with impairments in domains such as spatial memory, attention, and executive function even independently of an ongoing motor task [22, 23], and dedicated dual-task protocols have been developed specifically to characterize cognitive-motor interference in this population [24].

Taken together, this evidence supports the view that, in both older adults and vestibular patients, the cognitive cost of movement is a central feature of how these groups navigate everyday motor demands [4].

Quantifying this cognitive cost, however, requires a measure of mental effort that can be recorded continuously and unobtrusively while the participant is moving — a requirement that few of the classical workload measures satisfy, but that pupillometry is well positioned to meet.

This is precisely the gap that the present thesis addresses: applying pupillometry, combined with an ambient-light compensation strategy, to quantify cognitive workload during motor tasks of increasing complexity in older adults and vestibular patients.

1.5 Open challenges and thesis objectives

Applications of pupillometry during the execution of motor tasks remain limited [25], especially in ecological conditions, due to the challenge of separating the cognitive component of pupil dilation from the confounding effect of ambient luminance [26]. One of the objectives of this thesis is therefore, first, to validate the pupil-dilation measurements obtained from a wearable eye tracker such as the Tobii Pro Glasses 3 (TG3), since the entire thesis rests on the reliability of this measure. Building on this, the main objective is to apply pupillometry under more ecological conditions – without strict control over ambient lighting or participant mobility – by implementing and validating the performance of a method proposed in the literature for separating the cognitive component of pupil dilation from the component driven by ambient luminance [27]. Finally, the thesis aims to explore whether pupil dilation, once the cognitive component has been isolated, can serve as a biomarker for assessing the mental effort involved in performing motor tasks of varying complexity, in a sample of healthy older adults and vestibular patients, under the hypothesis that vestibular impairment requires the allocation of greater cognitive resources for walking and maintaining balance, as already suggested by Nascimbeni et al. [28].

1.6 Thesis outline

The thesis is organized as follows. The present chapter introduces the topic of the thesis, the open challenges in the field, and the objectives it sets out to achieve. Chapter 2 provides background on pupillometry: the physiological principles it relies on, the instrumentation required, and how it has been applied so far to studies involving motor tasks, with a particular focus on the approaches proposed

in the literature for handling the luminance confound. Chapter 3 describes the methods used to validate the pupil-diameter measurement obtained from the TG3. Chapter 4 describes how the ambient light sensor was selected, characterized, and integrated for luminance compensation. Chapter 5 then describes the experimental protocol, the signal-processing methods, and how the data were analyzed. Chapter 6 presents the results obtained from the methods described in Chapters 3, 4, and 5, and Chapter 7 discusses these results, outlining the limitations of the work and directions for future development. Finally, Chapter 8 summarizes the thesis's findings.

Chapter 2

Background

2.1 Pupil physiology

Pupil size is controlled primarily by two branches of the autonomic nervous system: the sympathetic pathway, which drives pupil dilation, and the parasympathetic pathway, which drives pupil constriction. These pathways connect the central control systems to the effector muscles of the iris: the iris sphincter muscle, which produces constriction, and the iris dilator muscle, which produces dilation.

Because they are driven by different systems, the two muscles also respond to different underlying mechanisms: pupillary constriction is primarily a reflexive response to light (the pupillary light reflex), but also occurs during near fixation. In contrast, pupillary dilation originates from central pathways involving the hypothalamus and the locus coeruleus (LC), a brainstem nucleus whose activity is closely tied to the engagement of cognitive processes such as attention, arousal, and mental effort [29, 30, 31]. Indeed, LC activity modulates pupil diameter through a dual mechanism, activating the sympathetic system while simultaneously inhibiting parasympathetic outflow at the level of the Edinger-Westphal nucleus (EWN), with both effects concurring to dilate the pupil [29]. Crucially, under typical ambient lighting conditions, the initial or moderate pupillary dilation observed during cognitive tasks is predominantly driven by this central parasympathetic inhibition, whereas sympathetic activation is further recruited during states of high arousal, emotional stress, or in darkness [30].

Three main types of stimuli elicit a pupillary response: the pupillary light reflex (PLR), a constriction in response to increased luminance; the task-evoked pupillary response (TEPR), a dilation in response to increased arousal and mental effort; and the pupillary near response (PNR), a constriction during near fixation [30]. The following sections focus on the first two of these phenomena, as the pupillary light reflex represents the main confounding factor when attempting to isolate the

second, cognitively-driven response.

2.1.1 Pupillary light reflex (PLR)

The pupillary light reflex (PLR) is a well-characterized phenomenon that strongly shapes the behaviour of pupil diameter. Indeed, pupil diameter fluctuations driven by ambient luminance span a range of approximately 1.5–9 mm [32], and numerous studies have quantified this effect in terms of magnitude [33, 34, 35] and latency [36, 37]. The pupillary response to changes in ambient luminance is not instantaneous, but unfolds over roughly 200–500 ms; this is mainly due to the inertia and activation dynamics of the iris smooth muscle [34]. Both the static and dynamic pupillary response are influenced by several factors:

- **Age**: several studies show this to be the most relevant factor, demonstrating that light-adapted pupil diameter decreases approximately linearly with age at all illuminance levels, although the rate of decline is more pronounced at low intensities [33, 35, 38].
- **Field of view**: the pupil does not respond to point luminance alone, but integrates light over the entire viewed area [39].
- **Monocular vs. binocular vision**: monocular adaptation produces a significantly larger pupil than binocular adaptation [39]. In one study, Telek et al. [38] report a pupil 4–6% smaller under bilateral illumination compared to unilateral illumination, with more pronounced differences at intermediate light levels.

These findings have, over time, led to the development of mathematical formulas for predicting pupil diameter in response to a target luminance value (cd/m^2), usable to compute the expected pupil diameter for a given light stimulus. Holladay [40] was the first to propose an empirical, exponential-type formula, based on three healthy observers. Moon and Spencer [41] introduced the use of the hyperbolic tangent function to establish upper and lower biological asymptotes. Stanley and Davies [42] were the first to integrate the visual field area (deg^2), and Winn et al. [33] subsequently provided the first robust quantification of the age-based regression slope. It is then Watson and Yellott [39] who unify all of the preceding parameters into a single predictive model. They define the Effective Corneal Flux Density ($F = L \cdot a \cdot M(e)$), where a is the adapting field area (in deg^2) and $M(e)$ is the monocular/binocular factor; they use the Stanley & Davies formula as the reference diameter at a reference age ($y_0 = 28.58$ years), and the variation due to age (y) is computed as a linear function of the age slope, which in turn depends on the

residual dynamic amplitude of the pupil – the smaller the reference pupil, the narrower the margin for further constriction with age:

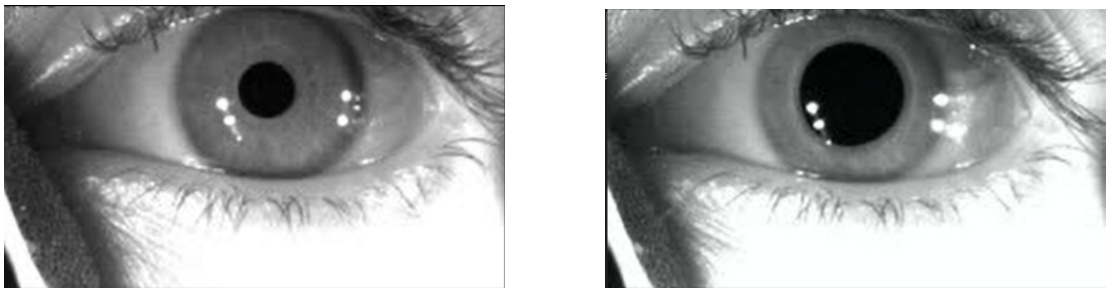
$$D_U(L, a, y, y_0, e) = D_{SD}(F, 1) + (y - y_0)\,[0.02132 - 0.009562\,D_{SD}(F, 1)] \quad (2.1)$$

2.1.2 Task-evoked pupillary response (TEPR)

It is well established in the literature that variations in pupil diameter are closely linked to mental effort and arousal. Graded changes in pupil diameter have been experimentally observed in tasks where workload is parametrically manipulated, such as during mental arithmetic tasks [9, 13] or short-term memory tasks [10]. At the same time, when task demand exceeds the subject's cognitive resources, a sudden decrease or plateau in pupil diameter is observed – a phenomenon regarded in the literature as an index of cognitive overload [11, 12, 14].

Some studies further suggest that not only the mean amplitude, but also high-frequency oscillations in pupil diameter, may be associated with cognitive activity and workload [8, 43, 44]. These findings sparked considerable interest in pupil dilation as an index of cognitive functioning, leading to the widespread adoption of pupillometry across many fields – spanning cognitive science, with research on perception, language processing, memory and decision-making, emotion, and cognitive development [32], as well as human factors [8, 45] and clinical and medical assessment [5, 46].

However, when compared with the large changes induced by the pupillary light reflex (PLR) and the pupillary near response (PNR), pupil diameter variations due exclusively to cognitive load or arousal are of modest magnitude, rarely exceeding a threshold of 0.5 mm [32, 30]. This intrinsic subtlety of the signal makes it essential, in pupillometry studies, to carefully control for and isolate possible environmental and visual confounding factors – first and foremost the PLR – as will be discussed in Section 2.4.



(a) Constricted pupil **(b)** Dilated pupil

Figure 2.1: Constricted (a) and dilated (b) pupil of the same participant, captured by the videonystagmograph (Natus ICS Dizcovery) used as the reference instrument during the TG3 validation protocol. The two bright points on the iris are the corneal reflections of the instrument's infrared illuminators (own data).

2.2 Pupil diameter measurement with infrared eye trackers

While handheld pupillometers are already available and commonly used in clinical practice for monitoring patients' PLR as a tool for measuring pupil diameter, most other use cases rely on video-based eye-tracking systems, which provide pupil size in addition to gaze direction.

Video-based eye trackers illuminate the eye with near-infrared light, invisible to the participant and therefore unable to elicit a pupillary response, and recover both gaze direction and pupil size from the resulting images [47]. Two features are extracted: the pupil, which appears as a dark, high-contrast region when the illuminators sit off the optical axis of the camera (*dark pupil* technique), and the first Purkinje image, that is, the specular reflection of the illuminators on the anterior surface of the cornea (*corneal reflection*), visible in Figure 2.1 as the two bright points on the iris. The vector joining the two is largely invariant to small translations of the camera relative to the head, which is what makes the technique usable in a wearable device. Data acquisition rate generally varies from 60 to 1000 Hz [48].

Two broad families of instruments can be distinguished:

- **Remote systems**: these typically require the subject to be stabilized with a chinrest. This eliminates head movements and allows the cameras to capture the eye on-axis, ensuring excellent sub-millimetre accuracy (<0.3 mm) [49], but at the cost of reduced naturalness of the subject's behaviour. Figure 2.2a shows the Tobii Pro Fusion, a representative remote system.

- **Wearable systems**: integrated into lightweight eyeglass frames, these leave the subject free to move and interact, albeit at the cost of lower accuracy compared to remote systems. This has enabled a number of studies that would not otherwise have been possible using screen-based eye-trackers alone [50]. Figure 2.2b shows the TG3, a representative wearable system.

These systems – such as the Tobii Pro Glasses or Pupil Labs used in eye-tracking research – combine miniaturized cameras and infrared illuminators mounted on the glasses frame worn by the participant. Precisely because of their light weight and portability, they are well suited to applications in ecological, dynamic contexts, allowing eye movements to be recorded while the wearer is free to move and interact with the surrounding environment. At the same time, this comes at the cost of an increased risk of artefacts introduced by these less controlled recording conditions. One study documented the effects of direct sunlight on the tracking performance of smart glasses [51], while other work has examined artefacts related to head movement – in particular, glasses slippage [50] – or to eye movements themselves, which affect the accuracy of the eye tracker's pupil size measurements [52].



(a) Tobii Pro Fusion (remote) **(b)** TG3 (wearable)

Figure 2.2: Examples of a remote (a) and a wearable (b) eye tracker. Images: Tobii AB.

2.3 Pupillometry in dynamic tasks

As discussed in the previous section, head-worn eye trackers have extended the use of pupillometry – the study of pupil-diameter variations as a function of cognitive processes – to more ecological settings, including dynamic tasks. Indeed, understanding the relationship between the execution of single or dual motor tasks and the cognitive load they impose has important implications for performance on tasks required for daily living [26]. This is especially relevant when motor automaticity is reduced, as in pathological conditions such as Parkinson's disease, vestibular

disorders, or older age, where individuals must recruit additional attentional resources to perform a given task. Detecting this additional load through a tool such as pupillometry could prove valuable for identifying individuals at risk of falling, or for developing novel rehabilitation strategies [5].

This has motivated a number of studies that have used pupil-diameter variations to estimate cognitive load during motor tasks. Kahya et al. [5] compared the pupillary response across postural tasks of increasing demand, induced through vision occlusion and dual tasking, between individuals with Parkinson's disease and healthy controls, while Tang et al. [53] sought to quantify the cognitive effort required during balance tasks in patients fitted with a cochleo-vestibular implant. Francavilla et al. [46] combined pupillometry with static posturography to gain insight into the cognitive demands associated with postural control in individuals with Ménière's disease. Other studies have focused more specifically on gait: Saeedpour-Parizi et al. [26] examined pupil and eye-movement behaviour in relation to cognitive effort during goal-directed walking in university students, while Manz et al. [25] used fixation and pupil-dilation metrics in an attempt to estimate cognitive load across a broad set of locomotor tasks in lower-limb prosthesis users.

Building on this line of work, the present thesis extends the study of the cognitive-motor relationship, as captured through pupillometry, to older adults and vestibular patients performing tasks of progressively increasing complexity, under ecologically valid, uncontrolled lighting conditions.

2.4 PLR mitigation approaches

However, all pupillometry studies face a common obstacle – the main confounding factor when conducting a pupillometry study – namely the PLR. As discussed in Section 2.1.2, ambient light elicits a much larger response than cognitive load, making it difficult to detect the weak dilations associated with cognitive effort. This problem is typically addressed by tightly controlling ambient luminance – often in windowless rooms with uniformly distributed lighting, so that no variation is present to elicit unwanted pupillary constrictions [5] – and participants are sometimes asked to fixate a reference point throughout the task. In other cases, recordings are carried out in complete darkness to sidestep the problem of light-driven constriction altogether [53]. While these solutions help mitigate the effect of the PLR on the recordings, they inevitably compromise the ecological validity of the study, limiting its applicability outside the laboratory and under real-world conditions. The luminance confound can, however, also be tackled algorithmically; overall, three broad categories of approaches can be identified by which pupillometry studies attempt to isolate the cognitive component of pupil dilation.

2.4.1 Baseline-related methods

Most studies rely on acquiring a mean baseline value of pupil diameter immediately before the task begins, or during its first milliseconds [54], under the same lighting conditions in which the task will subsequently be performed. Once the task begins, this baseline value is subtracted, and the resulting residual is no longer dependent on the PLR, since the task is carried out under the same lighting conditions in which the baseline was recorded. The fundamental limitation of this approach is that both the environment and the visual stimulus must be kept at a constant luminance level throughout the experiment, which makes it unusable in more complex, dynamic, and realistic scenarios [55].

2.4.2 Frequency domain methods

Other studies have attempted to separate the cognitive and light-driven components of pupil-diameter variation in the frequency domain. Peysakhovich et al. [43] showed that the ratio between the low- (0–1.6 Hz) and high-frequency (1.6–4 Hz) bands of the power spectral density of the pupillary signal (LF/HF ratio) is sensitive to cognitive load but not to luminance, under controlled experimental conditions. Other studies have instead focused on wavelet analysis, based on the assumption that an increase in the occurrence of abrupt discontinuities in the pupillometric signal reflects an increase in cognitive load. Building on this idea, Marshall [8] developed the Index of Cognitive Activity (ICA), in which the pupil signal is processed through wavelet decomposition to capture sudden changes at a chosen decomposition level; cognitive load is then quantified as the frequency of occurrence of these events. However, the ICA is a proprietary algorithm, and its implementation details are not disclosed. The only attempt to replicate the ICA in an open form is that of Duchowski et al. [44], who developed the Index of Pupillary Activity (IPA), sharing the same underlying principles as the ICA; this was later refined into the Low/High Index of Pupillary Activity (LHIPA) [56], which instead exploits a low/high-frequency ratio of pupil oscillation, in a similar spirit to the LF/HF approach of Peysakhovich et al. [43]. However, while the theoretical formulation of these indices claims that frequency-domain analysis can fully exclude the pupillary light reflex, subsequent experimental studies and a meta-analysis have disproved this neutrality: a meta-analytic review found the ICA to be significantly influenced by lighting conditions [57], echoing Rerhaye et al. [58], who report inconsistent ICA values across tasks performed under different lighting conditions; a direct comparison further showed that neither the ICA nor the IPA reliably differentiate levels of cognitive strain, with both remaining sensitive to screen luminance [59]. Together, these findings cast serious doubt on the applicability of these methods under field conditions, where lighting is highly variable.

2.4.3 Brightness compensation methods

Other studies instead rely on estimating and subtracting the light-driven component through pupil-diameter prediction models, in order to obtain a "residual" signal that preserves the cognitive or emotional effects:

$$D_C = D_T - D_L \tag{2.2}$$

To determine D_T, some works rely on static models describing the light-adapted pupil diameter, such as the unified formula of Watson and Yellott [39], while others rely on dynamic models, which aim to track the rapid transients of pupillary constriction and dilation: Pamplona et al. [60] developed a dynamic model based on a delay differential equation (DDE), while Fan and Yao [61] developed a nonlinear, second-order dynamic model, grounded in a viscoelastic description of the iris muscles, specifically optimised to estimate the transient PLR elicited by a brief (100 ms) light flash. These models are then used to predict, instant by instant, a baseline pupil-diameter value attributable to ambient luminance alone, which is subtracted from the eye tracker's measurement to isolate the cognitive component.

Regardless of the model chosen, accurately computing the PLR requires knowing how much light is actually stimulating the user's eye, and researchers adopt different strategies depending on the available hardware. In screen-based applications, perceived luminance is often extracted directly from the display's RGB values, which are linearised and converted into perceived-luminance values [55]; a similar approach was applied by John et al. [62], who fit personalised pupillary light-response models to known grayscale display intensities for both 2D screens and VR head-mounted displays. In a fully immersive VR environment, Eckert et al. [63] instead extract the display's sRGB values directly from the Unity game engine, again linearising them to derive relative luminance. In other cases, an external light sensor is used to obtain a direct, calibrated measurement of ambient illuminance, rather than relying on screen or camera pixel values. Pignoni et al. [27] propose one such combined solution, integrating an ambient light sensor with the relative luminance extracted from the eye tracker's scene camera.

Among these approaches, baseline-related methods are unsuitable for the dynamic, ecologically valid scenarios targeted by this work, as they require constant lighting throughout the task. Frequency-domain methods, despite their appeal as a light-invariant measure, have been shown to remain sensitive to luminance and therefore cannot be relied upon under uncontrolled illumination. Brightness-compensation models, by contrast, explicitly account for the light stimulus reaching the eye and can in principle be applied under any lighting condition, provided that stimulus luminance can be estimated with sufficient accuracy. For this reason, this thesis adopts the approach proposed by Pignoni et al. [27], combining a synchronized ambient light sensor with the eye tracker's scene camera to estimate

gaze-point luminance – a solution particularly well suited to a wearable, ecological setup such as the one used in this work, where lighting cannot be controlled and participants are free to move. The resulting luminance estimate, together with a pupillary light reflex model, is then used to isolate the cognitive component of pupil dilation, as described in the following chapters.



Figure 2.3: The wearable acquisition setup used in this thesis: a TG3 eye tracker fitted with a custom ambient light sensor, mounted on a headband above the frame. The design is inspired by the setup of Pignoni et al. [27], who instead employed a Pupil Labs eye tracker together with a different ambient light sensor (own data).

Chapter 3

Pupil diameter measurement validation

The methods proposed in this thesis rest on a single measured quantity: the pupil diameter reported by the eye-tracking glasses. Every subsequent step – estimating the light-driven component, subtracting it, and interpreting the residual as a correlate of cognitive workload – is therefore tied to how well that signal is acquired.

The magnitude of the effect of interest makes this dependence delicate. In fact, as task-evoked pupillary responses are of the order of 0.1–0.5 mm [30], establishing how faithfully the device reproduces pupil diameter, and above all its variations over time, is an essential prerequisite for the analysis that follows.

The validity of pupil-size measurements is very rarely established in the previous works, and to our knowledge there is no study in the literature comparing pupil-diameter measurement by eye trackers against a proper gold standard pupillometer. Studies on remote eye trackers have quantified gaze-dependent errors by exploiting conditions in which no pupillary variation was expected, typically constant-luminance tasks [48]. Others have compared a device against a clinical screening procedure [49], or against a research-grade stationary tracker taken as reference, finding good agreement at group level but consistent subject-specific biases [64]. For wearable eye trackers the picture is thinner still. The TG3 specifically, the smart glasses used in this study, have been evaluated repeatedly, but always in terms of *gaze* data quality: angular accuracy and precision in static and dynamic conditions [65], robustness to head and body movements [66], sensitivity to headset slippage [50], and standardised data-quality reporting [67]. Pupil size appears in this literature as a factor *affecting* gaze estimation quality [68], not as an output to be validated in its own right. To the best of the author's knowledge its pupil diameter has never been compared against an independent reference.

Since neither the algorithm nor any independent assessment is publicly available,

this chapter describes how the absolute pupil-size measurement of the TG3 was verified against a reference obtained by manual, frame-by-frame segmentation of the eye videos recorded by a clinical videonystagmograph (VNG) worn at the same time. Because that reference is itself a two-dimensional image measurement and subject to the same optics described in Section 2.2, the validation used only recordings acquired with the gaze held fixed straight ahead: keeping the gaze angle constant removes the pupil foreshortening error from the reference, so that the comparison between the two systems is not confounded by it.

3.1 Experimental setup

3.1.1 Tobii Pro Glasses 3

The system under validation is the TG3 (Tobii AB, Danderyd, Sweden), a wearable binocular eye tracker shaped as a spectacle frame. The head unit weighs 76.5 g including the cable and carries, per eye, eight infrared illuminators and two eye-tracking cameras integrated into the lens edge, for a total of sixteen illuminators and four sensors in a fixed geometry [69]. Gaze is estimated with a dark-pupil, corneal-reflection technique exploiting the stereo geometry of the two cameras per eye, and frame slippage on the face is compensated through a three-dimensional eye model. Nominal gaze accuracy is 0.6°, and calibration requires a single point or can be performed automatically. Pupil diameter is provided as an absolute measure in millimetres for each eye.

A scene camera mounted on the nose bridge records the participant's field of view at 1920 × 1080 pixels and 25 frames per second, with a field of view of 95° horizontally and 63° vertically; its use for gaze-point luminance estimation is described in Section 5.2.4. The head unit also embeds an ST LSM9DS1 magneto-inertial platform, whose gyroscope and accelerometer are sampled at 100 Hz and whose magnetometer is sampled at 10 Hz; these signals are what make temporal alignment with the ambient light sensor possible (Section 4.6). The head unit is cabled to a recording unit worn at the belt, which stores all streams on an SD card and is controlled from the TG3 controller application; data are then exported through Tobii Pro Lab.

Depending on the package, the device samples the eyes at 50 Hz or 100 Hz; the 100 Hz configuration was used throughout this work.

3.1.2 Videonystagmograph Natus ICS Dizcovery (VNG)

The reference system is the Natus ICS Dizcovery (Natus Medical Inc., Middleton, WI, USA), a videonystagmography device used clinically to assess vestibular



Figure 3.1: The TG3: the head unit, shaped as a spectacle frame, and the cabled recording unit worn at the belt.

function. It consists of a light-tight goggle mask weighing 395 g, operated through the Otosuite Vestibular software. Each eye is imaged by an infrared camera at 1280×1024 pixels under 940 nm infrared illumination; the tracking range spans $\pm 30°$ horizontally and $\pm 25°$ vertically. Eye position signals are sampled at up to 200 frames per second, whereas the eye videos are stored at a lower, separately configurable rate [70]. In the recordings used here, the exported signals are sampled at 200 Hz and the videos are stored at 30 frames per second.



Figure 3.2: The Natus ICS Dizcovery videonystagmography goggle mask, used as the reference system in the validation.

The device was adopted for its ability to store the raw eye videos alongside the derived signals, which the TG3 do not provide. This allowed the pupil to be

measured directly on the images, but it also means that the reference is not a gold standard in the strict sense, and it is worth being explicit about why: its output is a measurement in pixels on the image plane, and a metric calibration is required before it can be compared with the millimetres reported by the eye tracker. How this was obtained, and with what uncertainty, is the subject of Section 3.3.2. Table 3.1 summarises the two systems side by side.

Table 3.1: The two systems compared.

	TG3	Natus ICS Dizcovery
Role	System under validation	Reference
Form factor	Wearable spectacle frame	Head-mounted goggle mask
Weight	76.5 g (head unit, cable included)	395 g
Eye cameras	2 per eye	1 per eye
Eye illumination	8 infrared illuminators per eye	Infrared LEDs, 940 nm
Eye image resolution	Not accessible to the user	1280×1024 px
Eye signal sampling rate	50 or 100 Hz (100 Hz here)	Up to 200 Hz (200 Hz here)
Raw eye videos stored	No	Yes (30 fps here)
Pupil diameter output	Absolute, in mm	Measured offline on the images, in pixels
Gaze tracking range	Not specified	$\pm 30°$ horizontal, $\pm 25°$ vertical
Nominal gaze accuracy	0.6°	Not specified

3.1.3 Session setup

Data were drawn from an existing dataset acquired at the Audiology Department of Molinette Hospital (Turin) as part of an independent technical validation of the same eye tracker. The original cohort consisted of twenty university students (8 female, 12 male; age range 22–30 years), six of whom (3 female, 3 male) were included in the present analysis.

The two systems were worn concurrently throughout each acquisition session, with the TG3 positioned directly on the face and the Natus goggles mounted over the frame.

3.2 Acquisition protocol

Of the trials included in the full protocol, only those performed under the static condition were analysed in the present work. In this condition participants were seated on a rotational chair and instructed to maintain a fixed forward gaze for approximately 10 s.

Each session was preceded by task instruction and by an independent calibration of both devices. The TG3 were calibrated with the printed one-point target, held still at 0.5 m to 1 m in front of the participant, who was asked to fixate its centre; this procedure adapts the three-dimensional eye model of the device to the individual eye geometry [71]. For the Natus system, a region of interest was first placed around each pupil and the detection threshold adjusted until the pupil was tracked reliably; the participant, seated 1 m to 2 m from a blank wall with the head held still, was then asked to alternate the gaze between two targets projected on the wall by the integrated lasers [72].

Acquisition was subsequently initiated on the TG3 and then on the Natus, an order dictated by the synchronisation procedure described in Section 3.3.3. Each trial was performed once, unless repetition was required by acquisition failures. Five left–right head rotations were executed immediately after the onset and immediately before the offset of every recording, and were subsequently exploited for the temporal alignment of the two data streams.

3.3 Signal processing

3.3.1 Manual segmentation

As anticipated in Sections 3.1.1 and 3.1.2, the TG3 output an absolute measure of pupil diameter sampled at 100 Hz – the quantity to be validated – whereas the videonystagmograph provides no direct estimate of pupil size. It does, however, record and store the raw eye videos (Figure 3.3), from which pupil diameter could be obtained by manual segmentation and adopted as the reference measure.

Segmentation was carried out in MATLAB by two trainees of Politecnico di Torino, instructed on the task through a shared set of worked examples. Participants were divided equally between the two operators, and no recording was segmented by both. For each eye video, the first and last frames to be segmented were selected manually: the former as a frame with stable gaze following the initial head rotations, the latter as a frame with stable gaze preceding the final ones, so as to isolate the fixed-forward-gaze phase of the recording. Frames in which the pupil was even partially occluded, whether by a blink or by the eyelid, were left unsegmented and excluded from the subsequent analysis. The valid interval and the discarded frames



Figure 3.3: Example of a raw eye frame recorded by the videonystagmograph, from which the pupil was segmented manually.

were logged for every recording.

Within the selected interval, the pupil contour was traced frame by frame using the MATLAB `roipoly` function, which returns a binary mask of the region enclosed by the operator. Each mask was stored and processed with `regionprops`, which computes the ellipse having the same normalised second central moments as the region; the pupil diameter was taken as the length of its major axis. A value in pixels was thus obtained for every segmented frame of both eyes (Figure 3.4).

3.3.2 Pixel-to-millimetre conversion

The manual segmentation returns the pupil diameter in pixels on the image plane, whereas the eye tracker reports it in millimetres; a metric calibration of the reference is therefore required before the two can be compared. As the VNG eye-cameras frame one eye each, for this calibration it was chosen to use the horizontal palpebral fissure length (HPFL), the distance between the nasal (medial) and temporal (lateral) canthi, as a reference, which is clearly identifiable in the eye images and documented by normative studies [73, 74].

The horizontal palpebral fissure length was therefore measured directly for each participant, rather than estimated indirectly. Each participant measured their own HPFL using a ruler, with a resolution of 0.5 mm. The same distance was subsequently identified by the operator on the central frame of the corresponding video and measured in pixels. The conversion factor was then computed, for each recording, as the ratio between the two:

$$k = \frac{\mathrm{HPFL}_{\mathrm{mm}}}{\mathrm{HPFL}_{\mathrm{px}}}. \tag{3.1}$$

Two sources of uncertainty affect this factor beyond the resolution of the pixel measurement itself: the 0.5 mm resolution of the ruler used for the direct

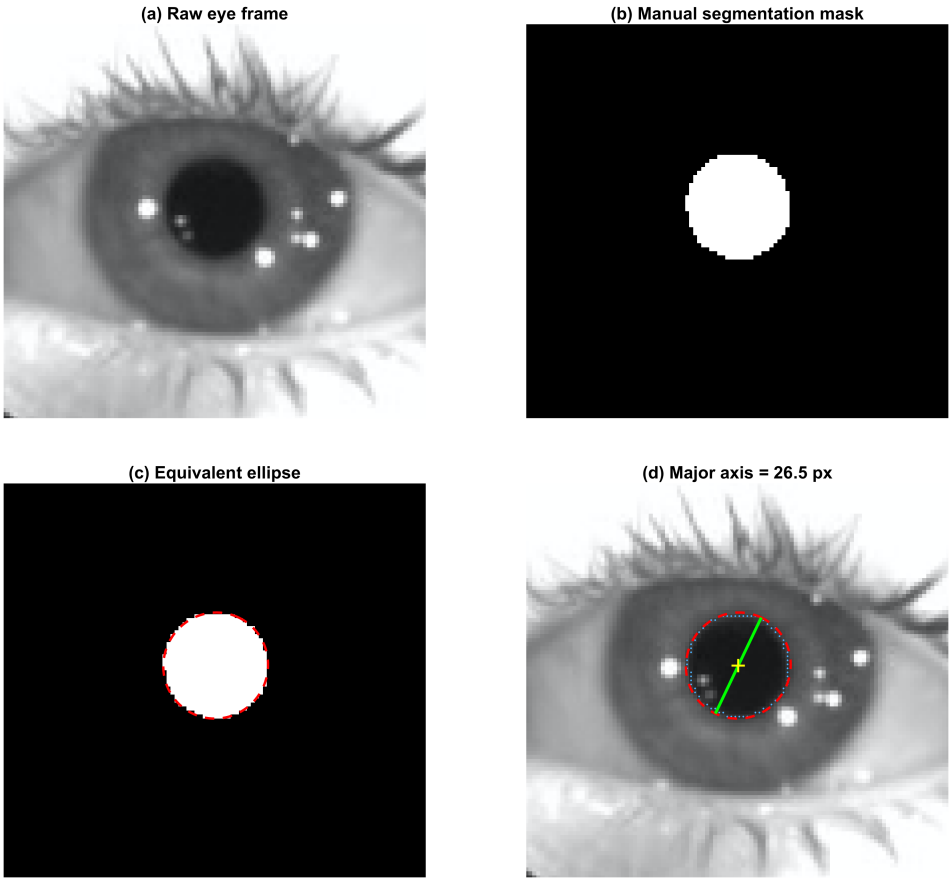


Figure 3.4: Manual segmentation of the pupil, illustrated on a single frame. (a) Raw eye frame. (b) Binary mask traced by the operator. (c) Ellipse having the same normalised second central moments as the mask. (d) Ellipse and its major axis superimposed on the eye frame, with the segmented contour in blue and the centroid marked; the length of the major axis is taken as the pupil diameter, in pixels.

measurement, and the possible mismatch between the two canthal points identified by the participant and those subsequently located by the operator on the video frame, since the two measurements were performed independently and at different times.

3.3.3 Data synchronisation

The two systems recorded independently, with their own clocks and sampling rates, so the streams had to be aligned in time before any sample-by-sample comparison. Alignment exploited the five left–right rotations performed at the beginning and end of each recording (Section 3.2), which produce a large, distinctive horizontal eye excursion visible to both devices: the horizontal eye position provided by the VNG and the horizontal gaze angle of the eye tracker, obtained as the arctangent of the horizontal over the depth component of its three-dimensional gaze point. Both signals were resampled onto a common 100 Hz grid and mean-subtracted, and the lag between them was estimated by cross-correlation, evaluated on the absolute-valued signals to remove the effect of the opposite sign conventions of the two systems. The estimated lag was then applied to the pupil-diameter series, which were resampled onto the same 100 Hz grid and truncated to their common interval, yielding, for each eye, a pair of time-aligned pupil-diameter signals.

3.4 Metrics

For each recording the comparison reduced to a pair of time-aligned pupil-diameter signals, the eye tracker in millimetres and the manual reference converted to millimetres. Let x_i and y_i $(i = 1, \ldots, N)$ denote the N samples at which both measurements were simultaneously defined, for the manual reference and for the eye tracker respectively, and $d_i = y_i - x_i$ their difference. Two complementary aspects were quantified.

Absolute agreement was summarised, per recording, by the median of each signal and by two paired-difference statistics. The mean signed difference, or bias (eye tracker minus reference),

$$\text{bias} = \frac{1}{N} \sum_{i=1}^{N} d_i, \tag{3.2}$$

isolates the systematic component of the discrepancy, whereas the root-mean-square error,

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} d_i^2}, \tag{3.3}$$

also incorporates its random dispersion.

At the group level the two systems were compared through a Bland–Altman analysis. For each recording j the between-method difference was taken as the difference of the two median diameters, $\Delta_j = \tilde{y}_j - \tilde{x}_j$, and the 95 % limits of agreement were defined as

$$\mathrm{LoA} = \bar{\Delta} \pm 1.96\, s_\Delta, \qquad \bar{\Delta} = \frac{1}{R}\sum_{j=1}^{R} \Delta_j, \tag{3.4}$$

with s_Δ the sample standard deviation of the differences over the R recordings, the interval being that expected to contain 95 % of the between-method differences.

Temporal agreement – the fidelity with which the device reproduces the variations of pupil diameter over time, which is the quantity of primary interest here – was quantified by the Pearson correlation coefficient between the two aligned signals,

$$r = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\,\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}, \tag{3.5}$$

with $\bar{x}$ and $\bar{y}$ the sample means. Being invariant to affine transformations, r is unaffected by the conversion factor and therefore by the calibration described above.

Chapter 4

Ambient light sensor integration

As discussed in Chapter 2, the pupillary luminance reflex model used in this work (Section 2.1.1) requires, as input, the luminance actually reaching the participant's gaze. The scene camera of the eye-tracking glasses cannot provide this value reliably, since it is limited by the exposure settings and the image post-processing of the camera [27]. A dedicated ambient light sensor (ALS) was therefore integrated into the acquisition system to provide an independent, calibrated estimate of the illuminance incident on the participant's eyes, to be combined with the scene-camera estimate to produce a gaze-point luminance estimate (Section 5.2.4). This chapter describes the choice of the sensor, its calibration, its integration into the existing wearable acquisition system, the auto-ranging logic implemented to cover the wide dynamic range of natural lighting, its physical placement, and its synchronisation with the eye tracker.

4.1 Market survey

The choice of the ALS was first constrained by the range of illuminance encountered in real-world recording conditions. Field studies report large variability across everyday locations. Indeed, while illuminance in indoor environments is typically found below about 500 lx [75], outdoor illuminance is typically found above 1,000 lx under daylight and varies considerably with environmental conditions, reaching up to about 90 klx [76] and beyond under direct sunlight.

A further requirement concerned temporal resolution. The pupillary light reflex is not instantaneous: pupillary response latency is defined as the time interval between the onset of a light stimulus and the effective onset of pupillary constriction. In healthy subjects, this latency typically falls between 200 and 500 ms from stimulus onset [30], decreasing as stimulus intensity increases; the same holds for the velocity of both constriction and dilation, which likewise increase with stimulus intensity

[34]. Resolving these dynamics therefore requires the ALS to have a temporal resolution well below these timescales.

Since the overarching aim of this work was to monitor cognitive workload under uncontrolled-lighting conditions, the objective was to identify an ALS able to cover an illuminance range consistent with real-world conditions and to sample at a rate adequate to resolve the dynamics of the pupillary light reflex, while providing a spectral sensitivity matching the photopic response of the human eye. Table 4.1 compares the main candidates considered against these criteria.

Table 4.1: Comparison of the ambient light sensors considered in the market survey.

Sensor	Manufacturer	Range (lux)	Output	Configurable gain / IT
VEML7700	Vishay	0 – 120 000	Calibrated lux	Yes (4 gains, 25–800 ms)
BH1750FVI	ROHM	1 – 65 535	Calibrated lux	No (2 fixed modes)
TSL2591	ams	up to 88 000 (600M:1 range)	Raw counts (dual photodiode)	Yes (4 gains, 100–600 ms)
VEML6030	Vishay	0 – 120 000	Raw counts	Yes
NOA1305	onsemi	0.165 – >100 000	Raw counts	Yes (8 IT steps, 6.25–800 ms)
LTR-329ALS-01	Lite-On	0.01 – 64 000	Raw counts	Yes (6 gains)
TSL2585	ams	4096× range (via gain)	Raw counts, plus UV/flicker channels	Yes

Datasheets: VEML7700 [77]; BH1750FVI [78]; TSL2591 [79]; VEML6030 [80]; NOA1305 [81]; LTR-329ALS-01 [82]; TSL2585 [83].

A market survey was carried out, and the Adafruit VEML7700 (Vishay) was ultimately selected as the best balance across these criteria: it provides an output already expressed in calibrated lux, an integration time configurable down to 25 ms, a photopic spectral response and straightforward I^2C integration [77].

4.2 ALS integration with INDIP system

The ALS was integrated into the INDIP system, a wearable multi-sensor platform developed at the University of Sassari for the instrumented analysis of gait in both laboratory and real-world settings [84]. Each INDIP unit is built around

a magneto-inertial measurement unit (MIMU) combining an LSM6DSO and an LIS2MDL device (STMicroelectronics, Switzerland), which provide a three-axial accelerometer with a full scale of ±16 g, a three-axial gyroscope with a full scale of ±2000 dps and a three-axial magnetometer with a full scale of ±50 gauss [85]. All channels are sampled synchronously at 100 Hz.

Besides its internal sensors, the INDIP unit exposes an I^2C bus that allows additional sensors to be integrated into the same acquisition chain. The VEML7700 evaluation board (Figure 4.2a) was therefore connected to this bus, so that the ambient light reading would be acquired by the same microcontroller that samples the inertial signals, stored in the same log file and, above all, referred to the same timebase: no post-hoc alignment between light and inertial data is required.

Integration required a modification of the INDIP firmware, which in its original form had no provision for an ambient light channel. Each sample record was extended by three bytes, encoding two additional fields (Table 4.2): two bytes for the raw ALS reading, expressed in counts, and one byte for the gain at which the sensor was operating when that reading was taken. Since the conversion from counts to lux depends on the gain (Section 4.3), a reading is not interpretable unless the gain in force at that instant is known. Storing both quantities also means the conversion to lux can be deferred to post-processing.

Table 4.2: Data fields added to the INDIP sample record to accommodate the ALS.

Field	Content	Size	Unit / encoding
`Lux`	Raw ALS output of the VEML7700	2 bytes	counts (0 – 65 535)
`Lux_Gain`	Gain in force at that sample	1 byte	0 = ×1, 1 = ×2, 2 = ×1/8, 3 = ×1/4

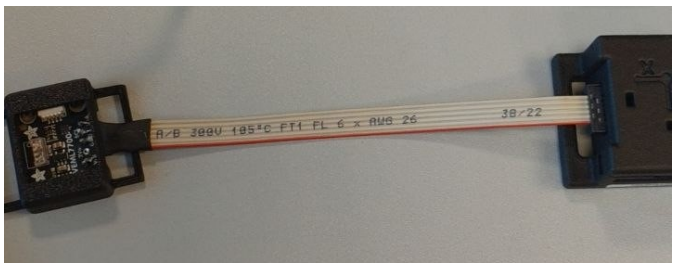


Figure 4.1: The VEML7700 sensor connected to an INDIP unit via a flat ribbon cable. The bare board sits inside its 3D-printed case, deliberately left open at the front (Section 4.5).

4.3 Sensor characterisation

The VEML7700 does not output an illuminance value directly: the ambient light reading is made available over the I^2C bus as a 16-bit raw digital value (*counts*, from 0 to 65 535) [77]. This raw output is related to illuminance by a purely multiplicative conversion factor, whose nominal value depends on the two configurable acquisition parameters, gain and integration time (IT): halving either of them doubles the conversion factor, and therefore the maximum measurable illuminance (Table 4.3). These nominal factors are, however, subject to a device-to-device output tolerance of up to 10 % [77], and to a further ±30 % tolerance at the shortest integration time of 25 ms [86], which is the setting adopted in this work.

An empirical calibration against a traceable reference instrument was therefore carried out to determine the actual conversion factor of the specific device used.

Table 4.3: Nominal resolution (conversion factor, lx/count) of the VEML7700 as a function of gain and integration time, as declared by the manufacturer [86]. The row corresponding to the integration time adopted in this work is highlighted.

IT (ms)	Gain 1	Gain 1/4	Gain 1/8
800	0.0084	0.0336	0.0672
400	0.0168	0.0672	0.1344
200	0.0336	0.1344	0.2688
100	0.0672	0.2688	0.5376
50	0.1344	0.5376	1.0752
25	**0.2688**	**1.0752**	**2.1504**

Since the auto-ranging logic implemented in the acquisition firmware (Section 4.4) switches between the gain settings 1, 1/4 and 1/8 according to the incident illuminance, the calibration was repeated independently for each of these three gains.

The same procedure also served to identify, for each gain, the range over which the response is effectively linear: the manufacturer reports that gains 1 and 2 become non-linear above approximately 100 lx, whereas gains 1/4 and 1/8 remain linear up to approximately 1,000 lx and, beyond that value, require a fourth-order polynomial correction to remain usable up to about 100 klx [86].

4.3.1 Acquisition protocol

The reference instrument was a Delta OHM HD2302.0 portable photo-radiometer equipped with an LP471PHOT photometric probe [87], whose spectral response

matches the standard photopic curve $V(\lambda)$ (Class B according to CIE No. 69), with a measuring range from 0.1 lx to 2×10^5 lx and a calibration uncertainty below 4 %.

The calibration setup, inspired by the approach of Coramik [88], is shown in Figure 4.2. A cardboard box, open on one side to allow the sensors to be inserted, shielded the sensing element from every light source other than the controlled one under test, and a sheet of graph paper fixed to its base allowed the sensor position to be reproduced exactly across successive measurements. Since the tip of the reference probe sits 38 mm above its mounting surface, the VEML7700 was raised on a support of matching height, so that both sensors would sample the light field at the same effective distance from the source. The desired illuminance level was set by changing the light source and its distance from the box.

Because the source-to-sensor distance was relatively short – particularly at the highest illuminance levels – the two sensors could not be placed inside the box simultaneously without one shadowing the other. Each condition was therefore acquired twice in immediate succession: once with the VEML7700 inside the box, and once with the reference probe in exactly the same position. The procedure was carried out across four successive versions of the setup (Setup 1 to Setup 4), progressively refined to improve light shielding and to extend the illuminance range that could be reached, for a total of 187 paired measurements (Table 4.4).

Table 4.4: Summary of the calibration measurements. For each gain setting and each version of the setup, the number of paired acquisitions and the corresponding illuminance range measured by the reference luxmeter are reported.

	Gain 1		Gain 1/4		Gain 1/8	
Setup	N	Range (lux)	N	Range (lux)	N	Range (lux)
1	21	0.61 – 184	21	3 – 196	21	0.64 – 182
2	17	171 – 1 321	17	211 – 1 329	15	247 – 1 424
3	15	1 450 – 14 732	15	1 492 – 16 105	15	1 533 – 15 477
4	10	18 330 – 61 987	10	19 395 – 58 070	10	19 790 – 59 410
Total	**63**		**63**		**61**	

4.3.2 Linearity assessment

Each recording was reduced to a single value by averaging the raw counts over a two-second window, and the resulting pairs (reference illuminance, raw counts) were examined for departures from linearity (Figure 4.3).

A cumulative linearity analysis was performed: for each gain, a first-order



(a) VEML7700 **(b)** HD2302.0 **(c)** Setup **(d)** Box interior

Figure 4.2: Instruments and calibration setup. (a) The Vishay VEML7700 on its Adafruit breakout board, reproduced from the manufacturer's product page [89]. (b) The Delta OHM HD2302.0 photo-radiometer with its family of interchangeable probes, reproduced from the manufacturer's datasheet [87]; the probe used in this work is the LP471PHOT. (c) The shielding box positioned under the light source, with the HD2302.0 unit visible on the left. (d) Inside of the box seen from above: the sheet of graph paper fixed to the base allows the sensor position to be reproduced across successive measurements.

polynomial was fitted to the measurements sorted by increasing illuminance, progressively extending the upper limit of the fitting range one point at a time, and the coefficient of determination R^2 was recorded as a function of that upper limit (Figure 4.4). At gain 1, R^2 remains essentially unitary up to about 150 lx and then degrades monotonically, falling below 0.90 above 10 klx. At gains 1/4 and 1/8, R^2 stays above 0.99 up to roughly 10 klx before declining, even if non-linearities already appear around 1,000 lx. This behaviour reproduces the thresholds declared by the manufacturer [86].

Part of the local variability visible in all three panels coincides with the transition between Setup 1 and Setup 2, that is, with a change of light source: such a dependence on the spectral content of the source is expected, since the manufacturer specifies an agreement within about 10 % with a reference luxmeter across LED, fluorescent and daylight sources [86].

On this basis, the fitting range used to estimate the conversion factor was restricted to 150 lx for gain 1 and to 1,000 lx for gains 1/4 and 1/8.

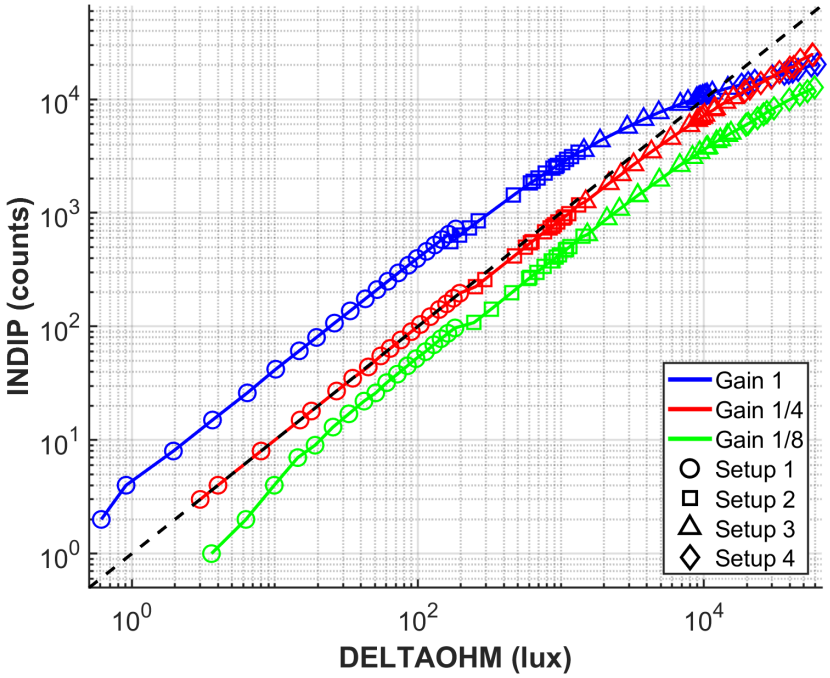


Figure 4.3: Raw VEML7700 output (counts) against the illuminance measured by the reference luxmeter, for the three gain settings and the four setups, on logarithmic axes. The dashed line is the ideal proportional response. Gain 1 departs from linearity above approximately 100 lx; gains 1/4 and 1/8 remain proportional up to about 1,000 lx and compress beyond 10 klx.

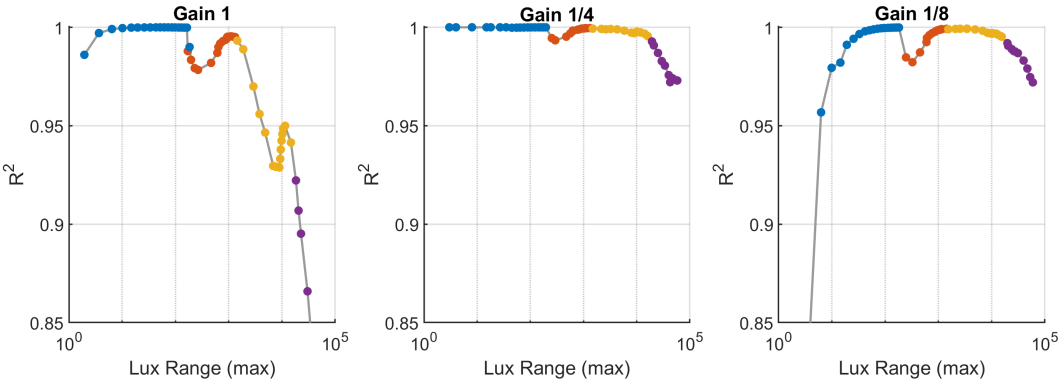


Figure 4.4: Cumulative linearity analysis. For each gain, the coefficient of determination R^2 of a first-order fit is plotted against the upper limit of the illuminance range included in the fit. Colours denote the setup to which the last point added belongs. Gain 1 loses linearity approximately one decade earlier than gains 1/4 and 1/8.

4.3.3 Conversion factors and non-linearity correction

Inside the linear range, the raw count N is converted into illuminance by a single proportionality constant:

$$E_{\mathrm{lin}} = k\,N. \tag{4.1}$$

The constant was estimated by least squares, forcing the fit through the origin:

$$k = \frac{\sum_i N_i E_i}{\sum_i N_i^2}. \tag{4.2}$$

The fit is forced through the origin for a physical reason: a zero photocurrent must correspond to zero illuminance. An intercept would distort the readings exactly where the counts are few, and where the resolution of the sensor is already coarsest relative to the measured value.

Table 4.5 compares the resulting factors with the nominal ones. All three deviate by less than 10 %, that is, within the tolerance declared by Vishay [77]. The calibration therefore removed a systematic error of a few per cent, and at the same time confirmed that the device behaves as specified.

Above the linear range the response compresses (Figure 4.3). The manufacturer recommends compensating this by applying a fourth-order polynomial to the linearly converted value [86]:

$$E = a\,z^4 + b\,z^3 + c\,z^2 + d\,z + e, \qquad z = \frac{E_{\mathrm{lin}} - \mu_1}{\mu_2}, \tag{4.3}$$

The predictor is centred on the mean μ_1 and scaled by the standard deviation μ_2 of E_{lin} over the fitting set, which improves the numerical conditioning of the fit. The polynomial was re-fitted on the data of this work rather than taken from the application note; its coefficients are reported in Table 4.6. The correction is applied above 100 lx, and only at gains 1/4 and 1/8.

Figure 4.5 shows what the correction achieves. Up to a few thousand lux the linear model alone is already accurate. Above approximately 5 klx it progressively under-reads, whereas the corrected estimate stays on the identity line up to the top of the tested range.

Table 4.5: Conversion factors obtained from the calibration at 25 ms integration time, compared with the nominal values declared by the manufacturer [86].

Gain	Calibrated k (lx/count)	Nominal (lx/count)	Deviation
1	0.2501	0.2688	−7.0 %
1/4	1.1326	1.0752	+5.3 %
1/8	2.2681	2.1504	+5.5 %

Table 4.6: Coefficients of the fourth-order correction of Equation (4.3), obtained by fitting the experimental data at 25 ms integration time, together with the centring and scaling parameters of the predictor.

Coefficient	Gain 1/4	Gain 1/8
a	−214.85	77.64
b	700.07	−254.98
c	2 376.0	2 301.0
d	9 983.4	10 116
e	5 823.4	5 660.6
μ_1 (lux)	5 157.8	5 059.0
μ_2 (lux)	7 755.7	7 591.8

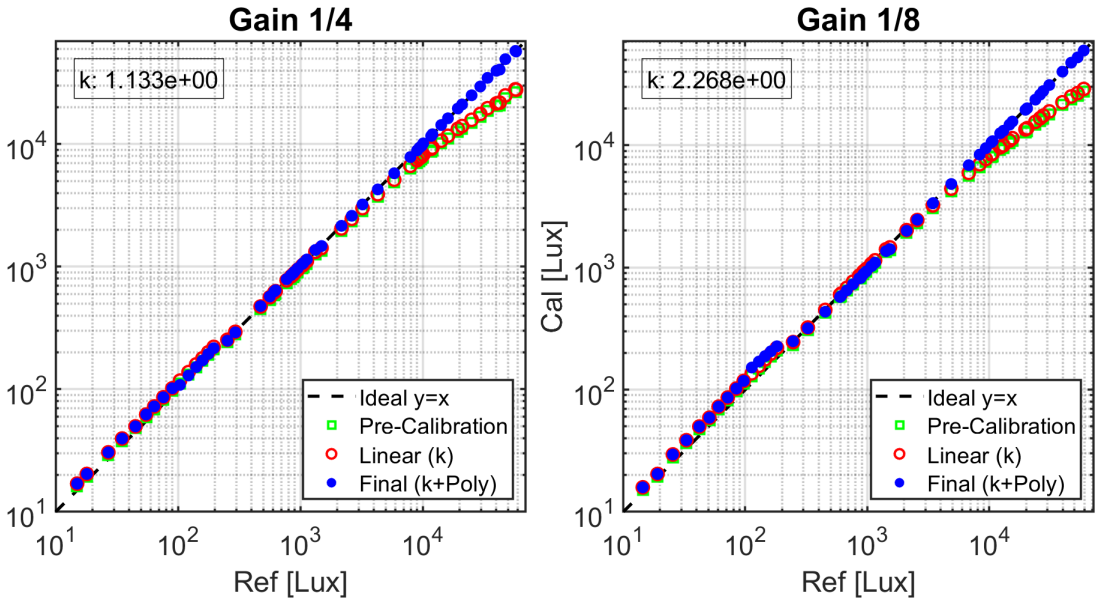


Figure 4.5: Effect of the polynomial correction at gains 1/4 and 1/8. Each panel compares the reference reading with the estimate obtained from the factory conversion factors, the estimate obtained with the calibrated factor k alone, and the final estimate including the fourth-order correction. The dashed line is the identity.

4.3.4 Validation against the reference luxmeter

The calibrated model was evaluated over the whole set of 187 paired measurements, taking for each point the gain that the auto-ranging logic of Section 4.4 would have selected at that illuminance.

The Bland–Altman analysis (Figure 4.6(a)) gives a mean difference of 0.15 lx with limits of agreement of +275.4 and −275.1 lx (SD = 140.5 lx). The absolute differences remain negligible over most of the range and grow only above 10 klx, where a given count step corresponds to several lux and where the polynomial correction operates. Expressed in relative terms (Figure 4.6(b)), the estimate carries a bias of 1.31 %, a mean absolute percentage error of 2.42 % and a maximum error of 18.01 %. The maximum is reached at the very bottom of the range, below 1 lx, where the reading amounts to a couple of counts and quantisation dominates; over the remainder of the range the error stays within the ±10 % band that corresponds to the sensor's own declared tolerance.

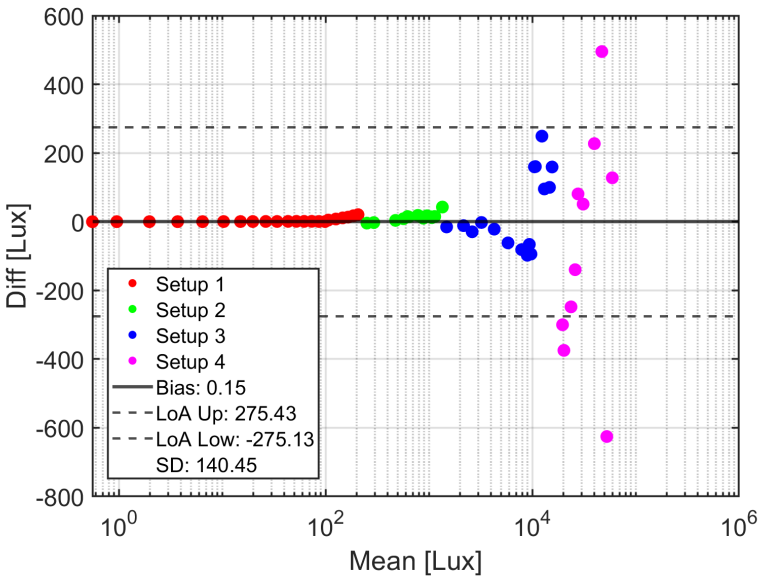
4.4 Auto-ranging logic

Covering a light detection range of more than seven decades, from 0.007 lx to 120 klx, requires the acquisition settings of the sensor to be adapted to the incident light level. A fixed gain would either saturate outdoors or waste resolution indoors. Moreover, the response of the pupil itself is not linear with respect to ambient illuminance: pupil diameter varies steeply with illuminance at low light levels, in the scotopic regime, and progressively flattens as photopic levels are approached, so that at high illuminance a large change in lux produces only a small change in diameter (Figure 4.7).

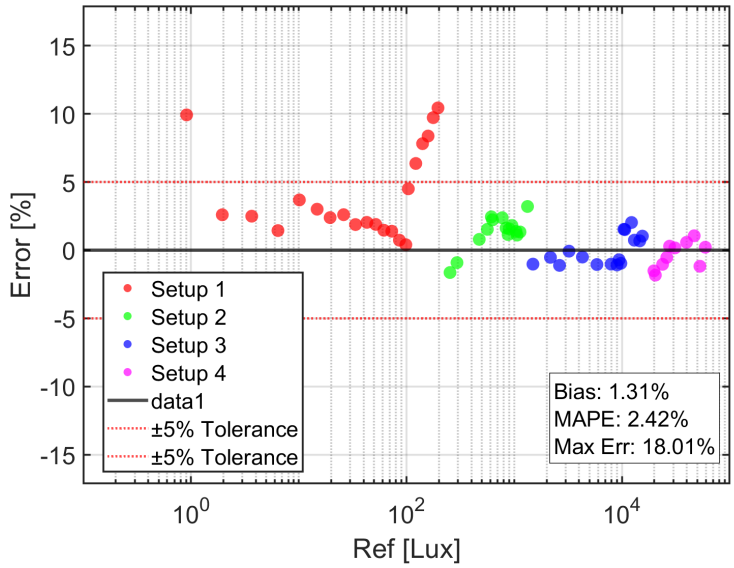
No single combination of gain and integration time provides both the resolution and the range required by this work (Table 4.7). A software logic was therefore implemented, in which the gain adapts to the incident illuminance so as to provide fine resolution at the lowest illuminance values, and to prioritise the avoidance of saturation where fine resolution is no longer needed.

Table 4.7: Trade-off between resolution and maximum measurable illuminance for the three gain settings used in this work, at an integration time of 25 ms [86].

Gain	Resolution (lx/count)	Maximum illuminance (lux)	Linear up to (lux)
1	0.2688	17 616	∼100
1/4	1.0752	70 463	∼1 000
1/8	2.1504	140 926	∼1 000



(a) Bland–Altman analysis.



(b) Relative error.

Figure 4.6: Agreement between the calibrated VEML7700 and the reference luxmeter over the whole calibration dataset, with the gain selected according to the auto-ranging logic. Colours denote the four setups. (a) Difference between the two instruments against their mean, with bias and limits of agreement. (b) Relative error against the reference reading; the dotted lines mark the ±10 % band and the solid line the mean absolute percentage error.

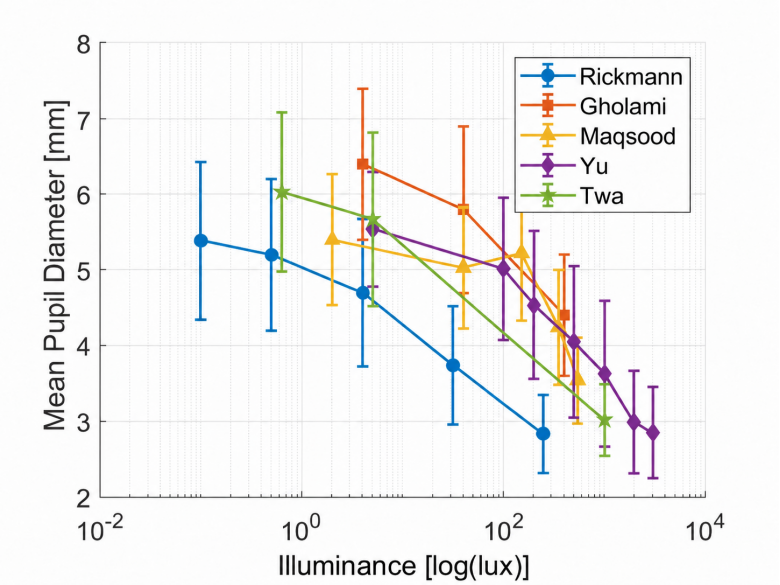


Figure 4.7: Mean static pupil diameter as a function of ambient illuminance, obtained by collecting the values reported in five independent studies [90, 91, 92, 93, 35]. Error bars are the standard deviations reported in the original works. The curve is steep at low illuminance and flattens towards photopic levels, so that measurement resolution on the illuminance estimate matters most in the low range.

The auto-ranging logic switches between the three gains at two illuminance boundaries. The lower boundary was placed at 100 lx, the limit beyond which gain 1 ceases to be linear. The upper boundary was placed at 10 klx, leaving a margin with respect to the compression observed at gain 1/4 beyond that level. Both boundaries are enforced on the raw counts rather than on converted illuminance values, so that no conversion is required on board in order to decide whether to switch.

The complete logic is summarised in Figure 4.8. On power-up the sensor is configured with an integration time of 25 ms and gain 1/4, the intermediate setting, so that the first reading is already usable over most indoor and outdoor conditions without risking immediate saturation and while still ensuring a fine resolution. Each read-out is then compared against the thresholds reported in Figure 4.8: if the count falls below the lower threshold the gain is raised to 1, to ensure a better resolution where the pupil is most sensitive; if it exceeds the upper threshold the gain is lowered to 1/8, to avoid saturation. From gain 1 and from gain 1/8 the logic returns to gain 1/4 as soon as the respective return threshold is crossed.

The illuminance is finally obtained from the raw count according to the gain in force: Equation (4.1) alone at gain 1, which by construction operates only

within its linear range, and Equation (4.1) followed by the polynomial correction of Equation (4.3) at gains 1/4 and 1/8.

Figure 4.9 shows the response of the complete chain, from the raw count through the gain selection to the corrected illuminance, against the reference luxmeter over five decades.

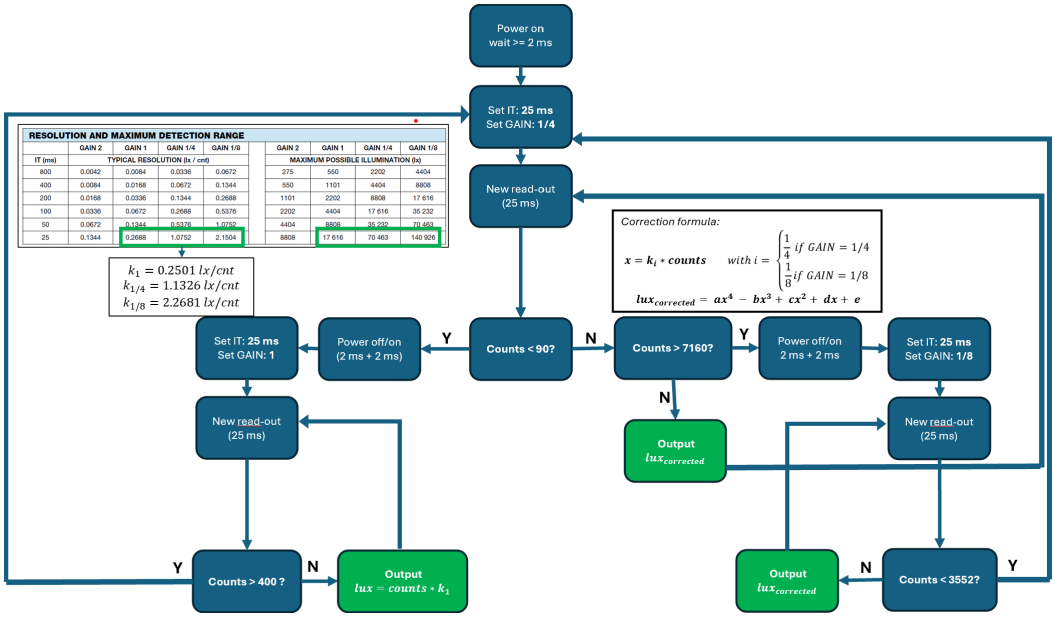


Figure 4.8: Flowchart of the auto-ranging logic, with the conversion factors obtained from the calibration and the correction formula applied at gains 1/4 and 1/8.

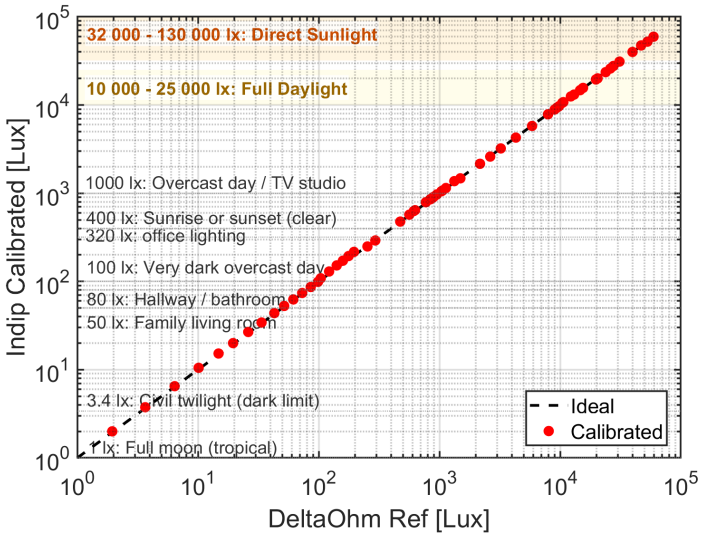
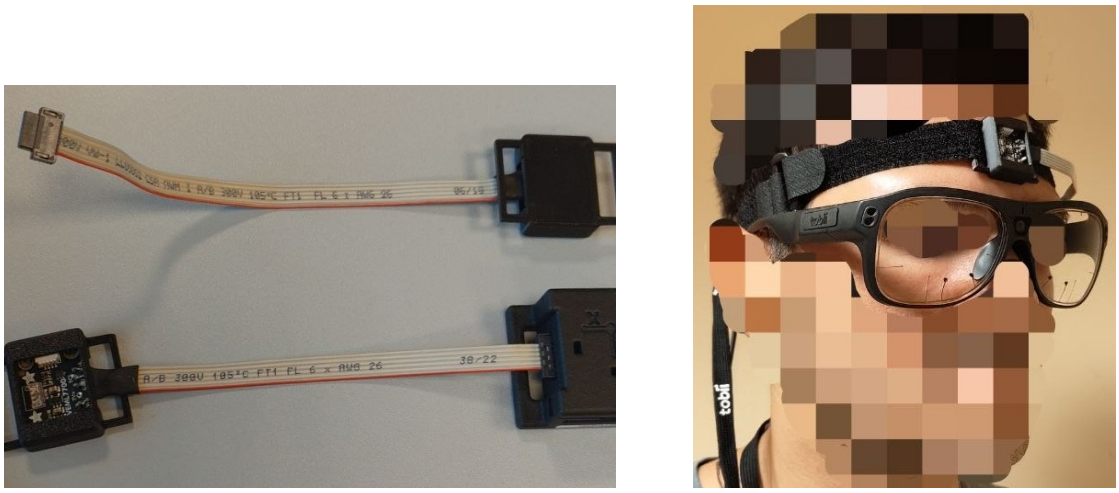


Figure 4.9: Response of the complete measurement chain under the auto-ranging logic: calibrated illuminance against the reference reading, on logarithmic axes, with the ideal one-to-one relationship shown as a dashed line.

4.5 Case and positioning

A VEML7700 unit was assembled in a 3D-printed plastic enclosure and connected to an INDIP unit through a flat ribbon cable (Figure 4.10(a)). The enclosure was designed to leave the sensing element as unobstructed as possible: the device collects light over a wide acceptance cone, its relative sensitivity remaining above half of the on-axis value out to an angular displacement of approximately $\pm 55°$ [77], and any aperture narrower than the housing would cut into a region in which the sensor is still substantially sensitive.

The positioning followed the same rationale adopted by Pignoni et al. [27]. Two requirements acted in the same direction. The first was proximity to the scene camera of the eye tracker, so that the two devices sample as nearly as possible the same portion of the light field. The second was that the quantity of interest is the illuminance actually reaching the eye, which is best approximated by placing the sensor as close to it as the wearability of the system allows. The enclosure was therefore fixed to the temple module worn alongside the eye tracker, with the sensing face oriented forward along the direction of gaze, a few centimetres from the eye (Figure 4.10(b)).



(a) Enclosure **(b)** Positioning

Figure 4.10: Enclosure and positioning of the ALS. (a) The two assembled units, seen from the front (top) and from the back (bottom), each in its 3D-printed enclosure and connected to an INDIP unit by a ribbon cable; only one unit was used during the recordings, the second being kept as a spare. The front face is fully open, leaving the sensing element unobstructed. (b) The sensor in place on a participant, mounted on the temple alongside the TG3, close to the scene camera and to the eye.

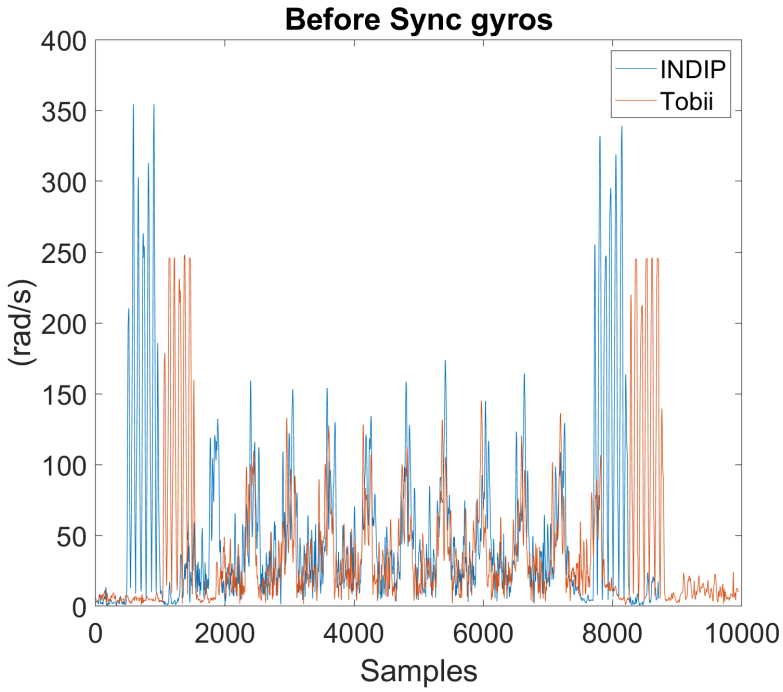
4.6 Data synchronisation with the eye tracker

As the eye tracker is an independent device with respect to the INDIP system, with its own clock and its own recording session, the two systems had to be aligned a posteriori. Both systems carry a magneto-inertial unit, and this redundancy was exploited for the alignment: the angular velocity of the head is measured simultaneously by the TG3 head unit and by the INDIP unit worn at the ear, and provides a common reference signal.

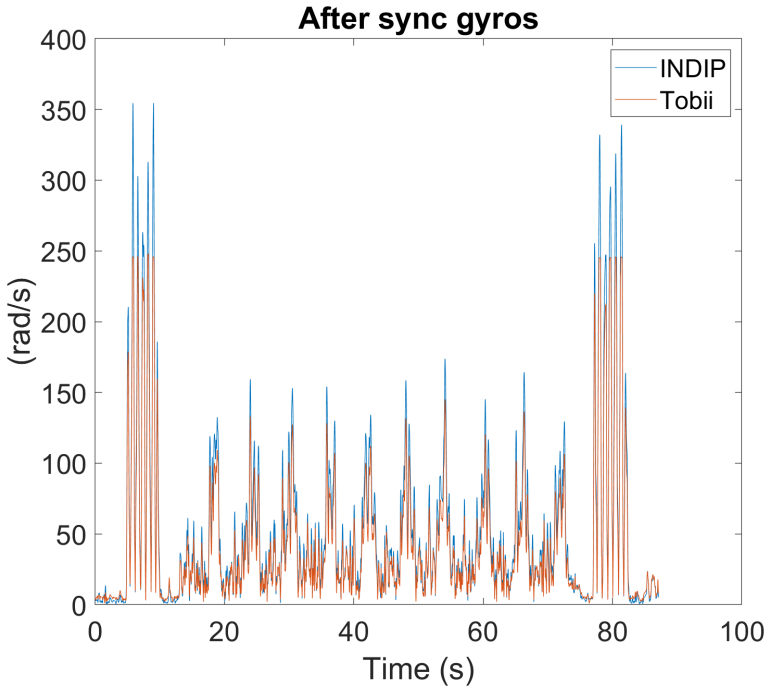
Each recording was organised so that the TG3 recording was always started before the INDIP recording and stopped after it, leaving an unused tail at both ends. Immediately after the INDIP was started, the participant was asked to perform three consecutive head rotations, followed by approximately ten seconds of steady forward gaze; the same sequence was repeated immediately before the INDIP was stopped. The rotations produce a distinctive, high-amplitude feature in the angular velocity of both devices, and the still interval that follows separates it cleanly from the rest of the recording.

So, in post-processing, the inertial signals of the eye tracker were first rotated

into the same axis convention as the INDIP unit, then the alignment was performed computing the lag maximising the cross-correlation between the two norms (Figure 4.11), taken as the delay of the INDIP start with respect to the TG3 start. The eye-tracker signals were shifted by that lag and truncated to the length of the INDIP recording, which removes the leading and trailing tails. Finally, all signals from both systems were interpolated onto a single time axis at 100 Hz. The output of this step is a single record in which pupil diameter, gaze, head angular velocity and the raw output of the ambient light sensor, with the gain in force at each sample, are all expressed on the same time base and are directly comparable.



(a) Before synchronisation



(b) After synchronisation

Figure 4.11: Magnitude (norm) of the head angular velocity measured by the INDIP head unit and by the TG3, (a) before and (b) after alignment. The head rotations performed at the beginning and at the end of each acquisition provide the distinctive feature on which the cross-correlation is computed.

Chapter 5

Cognitive workload study

This chapter presents the central study of the thesis, in which the light-compensated pupillometry approach is applied to the estimation of cognitive workload during cognitive–motor tasks in healthy older adults and vestibular patients. Building on the validation of the pupil-diameter measurement (Chapter 3) and on the integration of the ALS (Chapter 4), it brings the two together into a single acquisition and processing pipeline. The chapter first sets out the experimental design – participants, setup and protocol (Section 5.1); it then describes the estimation method that turns the raw device outputs into a light-corrected cognitive component (Section 5.2); and it finally details how the resulting signal was analysed, both to assess the quality of the light–cognitive separation and to test the sensitivity of the cognitive component to task demand and clinical status (Section 5.3).

5.1 Experimental design

The experimental design is described in three parts: the two cohorts of participants (Section 5.1.1), the wearable acquisition setup (Section 5.1.2), and the protocol of tasks (Section 5.1.3), the latter chosen to span a graded range of cognitive–motor demand under uncontrolled, ecologically realistic lighting.

5.1.1 Study participants

The study involved two cohorts. The first consisted of healthy older adults, recruited with an age of at least 65 years and reporting no neurological, vestibular or musculoskeletal disorder that could affect postural control or locomotion. The second consisted of patients with a diagnosed vestibular disorder. The two cohorts were chosen in order to compare a population with intact vestibular function against one in which the deficit is expected to increase the cognitive cost of locomotion, an

aspect discussed in Section 1.4 on the influence of vestibular pathology on spatial working memory. Participants requiring vision correction wore the corrective lenses supplied with the TG3, fitted into the eye-tracker frame.

Participation was voluntary, and all subjects gave written informed consent before the experimental activities; the data, including the scene-camera videos, were anonymised and processed in accordance with the current regulations on the protection of personal data (EU Regulation 2016/679).

Because of problems in the data acquisition, not all the collected recordings were usable, so the analyses reported here were conducted on a subset of the participants. In all, 11 healthy older adults and 7 vestibular patients were recruited, of whom 7 and 5 respectively were retained (Table 5.1), for a total of 12 subjects: the healthy group comprised three men and four women (age range 66–82 years) and the vestibular group three women and two men (age range 40–75 years), four with a unilateral and one with a bilateral deficit.

Table 5.1: Number of participants recruited and included in the final analyses.

Cohort	Recruited	Included in the analysis
Healthy older adults	11	7
Vestibular patients	7	5
Total	18	12

5.1.2 Experimental setup

The recordings were carried out at two sites: the vestibular patients and two of the healthy participants were tested at the Audiology Department of Molinette Hospital (Turin), and the remaining healthy participants at the Movement Analysis Laboratory of Politecnico di Torino.

Each participant was equipped with four INDIP units: two units at the feet coupled to pressure insoles, two distance sensors on the right foot, one unit at the waist and one at the left ear. The present work relies only on the two head-referred signals: the ALS carried by the ear-worn unit (Section 4.2) and the pupil diameter and scene video provided by the TG3 (Section 3.1.1). The remaining channels belong to the gait analysis for which the wider protocol was designed and are not used here. Prior to each session the TG3 were calibrated following the device procedure, and synchronisation between the INDIP system and the eye tracker was performed in post-processing as described in Section 4.6.

Lighting was deliberately left uncontrolled. Sessions took place under ordinary laboratory lighting, combining artificial light and daylight, with no attempt to keep

it constant. This was a central aim of the study: to test the pupillary light reflex compensation under conditions as close as possible to everyday ones, and thereby to support the use of pupillometry beyond the tightly controlled settings it usually demands. These remain laboratory conditions, and cannot yet be regarded as a real-world setting, but they represent a clear step in that direction. The illuminance measured over the sessions ranged from 2.00 lx to 1,557.68 lx at Politecnico and from 0.25 lx to 978.58 lx at Molinette.

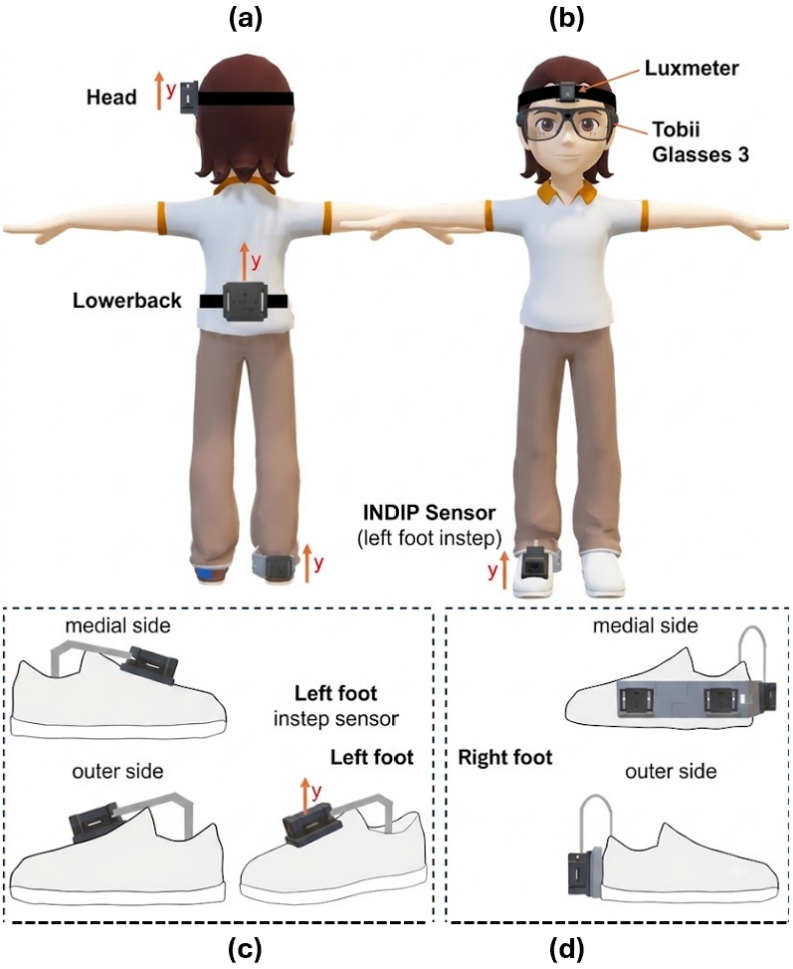


Figure 5.1: Complete instrumentation of a participant: (a) posterior view and (b) anterior view, showing the head unit (ALS and TG3), the lower-back unit and the foot sensors; (c) the left-foot instep sensor and (d) the right-foot sensor.

5.1.3 Experimental protocol

The full protocol comprised several activities, of which six were retained for the estimation of cognitive workload. They were selected to span a range of cognitive–motor interference, from an almost purely motor task to a purely cognitive one, so that the light-corrected pupil signal could be examined across increasing cognitive demand. After the walking task with tracking, the simulated daily-living sequence and the MoCA, participants completed a NASA-TLX questionnaire [7] rating the perceived workload; for this purpose the two repetitions of the walking-with-tracking task (clockwise and counter-clockwise) were treated as a single activity. The tasks were administered in the order MoCA, daily-living sequence, walking, and walking with tracking, an order dictated only by practical convenience and without consequence for the analysis. They are described below from the least to the most cognitively demanding.

Locomotor task. Participants walked continuously around two cones placed 4 m apart, completing five loops around each cone along a figure-of-eight path. The task was performed twice, first clockwise and then counter-clockwise (hereafter UTWcw and UTWccw, with UTW standing for the U-turn walking task). It carries little explicit cognitive load and serves as the motor baseline.

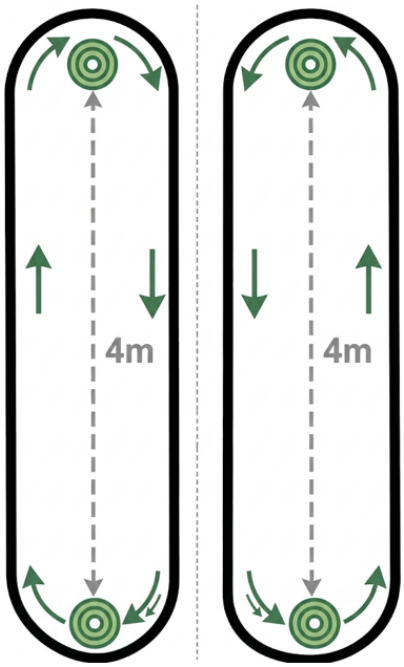


Figure 5.2: Locomotor task. Participants walked a figure-of-eight path around two cones placed 4 m apart, completing five loops around each cone, first clockwise (left) and then counter-clockwise (right). The arrows indicate the walking direction.

Locomotor task with tracking. The same figure-of-eight walk was combined with a tracking task: before the participant reached one of the two turning points, a ball was thrown at random to the left or to the right, and the participant had to follow it. One throw was made on each of the five loops, and the task was again repeated clockwise and counter-clockwise (UTWTcw and UTWTccw, with UTWT standing for U-turn walking with tracking). The reaction to the ball adds an attentional demand on top of the locomotor one.

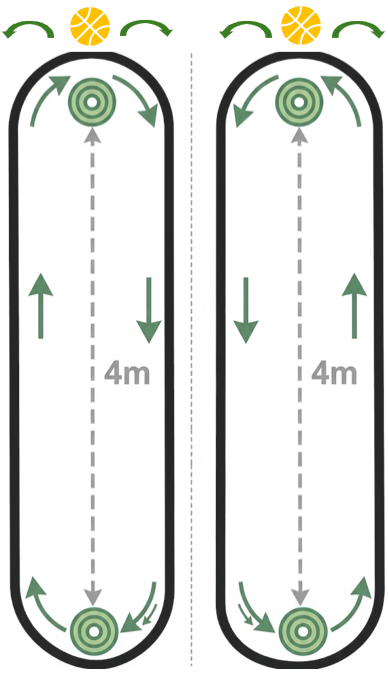


Figure 5.3: Locomotor task with tracking. The same figure-of-eight walk was combined with a tracking task: on each loop a ball was thrown at random to the left or to the right (top) and the participant had to follow it. The task was performed clockwise (left) and counter-clockwise (right).

Simulated daily-living sequence. Participants were required to memorise and then execute an ordered sequence of everyday actions, numbered here as in Figure 5.4: (1) standing up from a chair; (2) reaching the table and setting it; (3) moving a chair in front of the table, sitting down, miming eating and drinking for about 30 s, standing up, putting the chair back and clearing the table; (4) walking a slalom between three cones; (5) stepping over an obstacle; (6) listening to a person speaking for about one minute; and (7) walking the slalom again while collecting the cones, before sitting down. By combining a realistic motor circuit with the encoding and retrieval of an ordered sequence, this task places a high demand on the motor and cognitive systems simultaneously.

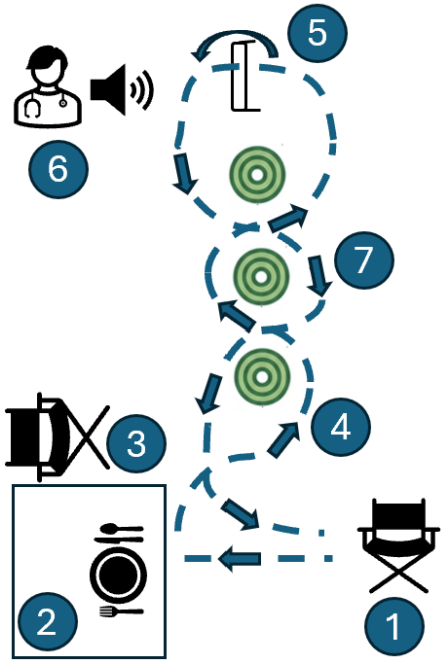


Figure 5.4: Simulated daily-living sequence (SDA). The dashed line traces the route through the workspace – chair, table, three-cone slalom and obstacle – and the numbered markers indicate the order of the phases carried out by the participant, described in the text.

Cognitive task. A purely cognitive condition was provided by the Montreal Cognitive Assessment (MoCA), a brief paper-based screening of global cognition administered while seated [94, 95]. Performed at rest, it represents the purely cognitive end of the interference range.

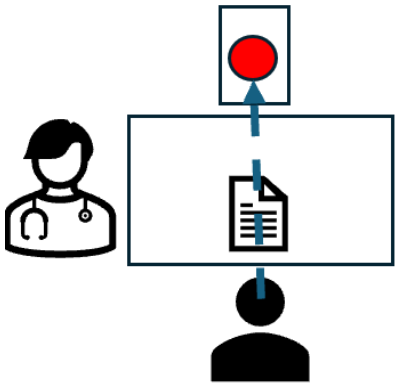


Figure 5.5: Cognitive task (MoCA). The test was administered by a clinician while the participant, seated, fixated a red target on a sheet placed in front of them.

Personalisation test. Alongside the six tasks above, all twelve participants performed an additional acquisition, the personalisation test, designed to characterise the subject-specific response of pupil size to light alone, in the absence of any cognitive demand. Its purpose was to probe the light-driven component of the pupillary response in isolation, so that the prediction of the Watson–Yellott model (Section 2.1.1) could be compared with the pupil diameter measured by the TG3 across a wide range of luminance, where, with no cognitive load, the two are expected to coincide. This condition therefore provides the reference against which the accuracy and the limits of the light model – saturation and residual error – are assessed.

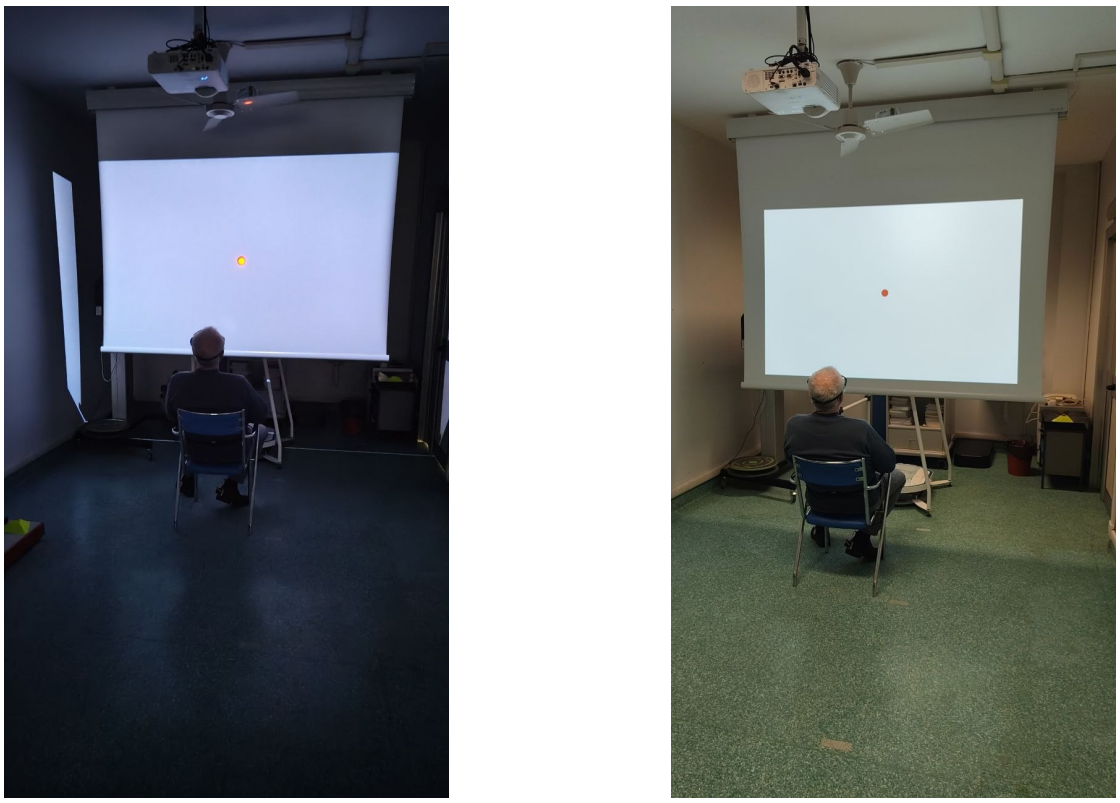
Performed last in the session, the test required the participant to sit in front of a display in a room initially in complete darkness. After 1.5 min to 2 min of dark adaptation, once the pupil diameter had stabilised, a fixation video was started (the same video for every participant). A red dot at the centre of the display served as a fixation target, which the participant was instructed to fixate for the entire duration of the recording. The background changed every 20 s, stepping from black to white through three intermediate shades of grey (normalised RGB levels 0, 0.25, 0.5, 0.75 and 1); after the brightest level the room lights were switched on in two steps, through two separate switches operated one at a time, to raise the illuminance further with artificial lighting, and finally the blinds were raised to add natural light on top of the artificial one. The test thus comprised eight phases – five background levels, two lighting steps and the raising of the blinds – each held for about 20 s, for a total duration of about 160 s. The display differed between sites: at Molinette the video was projected onto a dedicated screen and viewed from 1.8 m, whereas at Politecnico, where no projector was available, it was shown on a monitor viewed from 0.6 m. As in every other acquisition, the luminance

reaching the participant was measured throughout by the ALS together with the scene camera (Section 4.2). In this way the stimulus drove the pupil across a broad luminance range, from scotopic to photopic vision, under light stimulation alone.



(a) RGB 0 **(b)** RGB 0.25 **(c)** RGB 0.5 **(d)** RGB 0.75 **(e)** RGB 1

Figure 5.6: The five background levels of the fixation video, from black to white through three intermediate greys (normalised RGB values 0, 0.25, 0.5, 0.75 and 1). The red dot at the centre is the fixation target, held for the whole recording.



(a) Dark room **(b)** Room lights on

Figure 5.7: A participant during the personalisation test: (a) in the initial dark-room phase and (b) after the room lights were switched on.

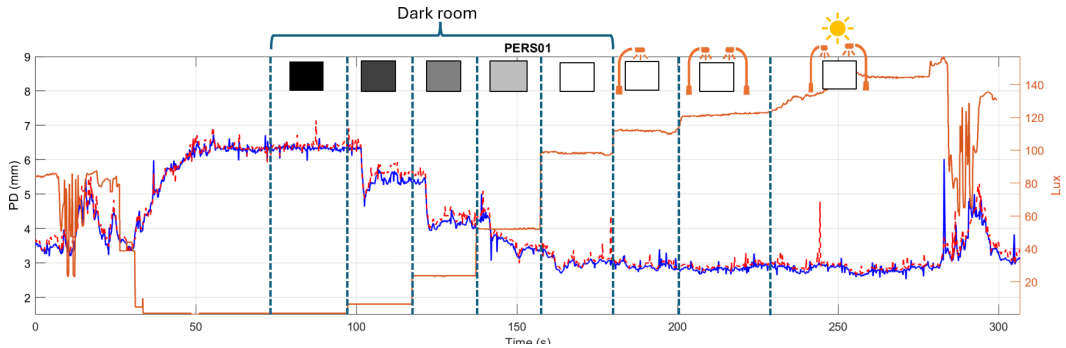


Figure 5.8: Example recording from one personalisation test: the measured pupil diameter – right eye in red, left eye in blue – (left axis) and the illuminance recorded by the ALS (right axis) over time. The dashed lines delimit the eight stimulus phases – five background levels, two lighting steps and the raising of the blinds, sketched at the top. As the illuminance rises, the pupil constricts, illustrating the light-driven response that the test is designed to isolate.

5.2 Estimation method

5.2.1 Overview of the estimation method

As we have already discussed before, pupil size is driven by two main superimposed mechanisms: the pupillary light reflex, which adapts the pupil to the light entering the eye, and the task-evoked response, which reflects cognitive and autonomic load (Chapter 2). Under the uncontrolled lighting of the present recordings the light-driven variation is by far the larger of the two – of the order of millimetres against the tenths of a millimetre of the cognitive one – so that the cognitive signal is buried in the response to light and cannot be read directly.

The strategy adopted is to estimate the light-driven contribution from an independent measurement of the light actually reaching the eye and to remove it from the total, leaving the cognitive component as a residual. Assuming that the two contributions add,

$$D_T(t) = D_L(t) + D_C(t), \tag{5.1}$$

the cognitive component is obtained as $D_C(t) = D_T(t) - D_{L,\mathrm{corr}}(t)$, where $D_{L,\mathrm{corr}}$ is the light-driven diameter predicted by a model and anchored to the individual subject (Section 5.2.5).

Building the two terms of this subtraction requires two distinct lines of processing, described in the subsections that follow. The first (Sections 5.2.2 and 5.2.3) is a signal-preprocessing pipeline assembled for this work from established methods and the specific needs of the sensors, which turns the raw outputs of the two wearable

devices – the pupil diameter from the eye tracker and the illuminance from the ALS – into the clean total diameter D_T and the absolute ambient luminance $avgL$. The second (Sections 5.2.4–5.2.5) is the separation algorithm proper, adapted from the hybrid camera–sensor method of Pignoni et al. [27]: it reconstructs the absolute luminance at the point the subject is looking at by combining the scene camera with the ALS, feeds it to the Watson–Yellott light-adapted pupil model to predict D_L, and calibrates that prediction on a per-subject baseline before the final subtraction. Figure 5.9 summarises the whole chain, and the symbols used throughout are collected in Table 5.2.

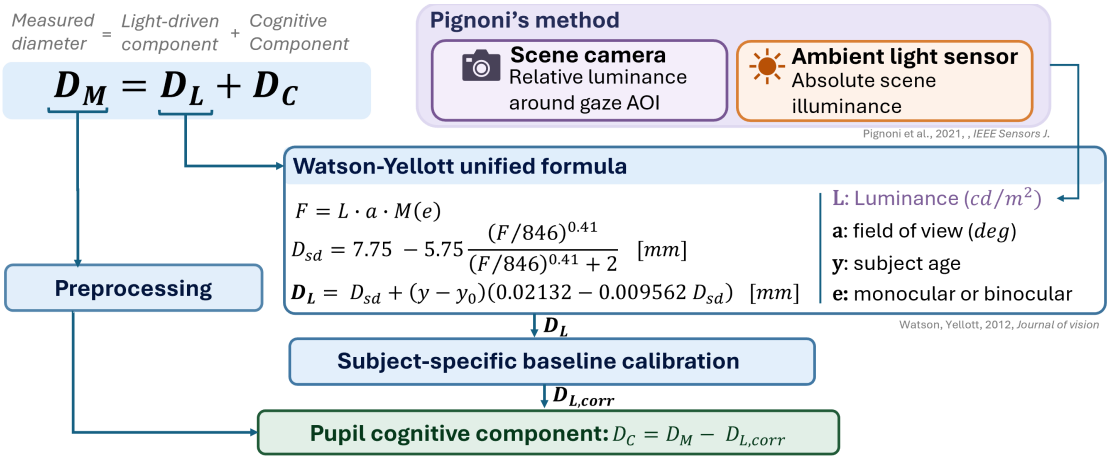


Figure 5.9: Overall block diagram of the estimation method, from the four system outputs – the pupil diameter and the scene camera from the eye tracker, and the raw counts and gain from the ALS – to the cognitive component D_C. The upper branch pre-processes the pupil diameter into the total signal D_T; the lower branch reconstructs the absolute gaze-point luminance by combining the ALS with the scene camera, feeds it to the Watson–Yellott model to predict the light-driven diameter D_L, and calibrates it on a per-subject baseline before the final subtraction $D_C = D_T - D_{L,\mathrm{corr}}$.

5.2.2 Pupil diameter signal pre-processing

The eye tracker outputs the pupil diameter of each eye in millimetres, sampled at 100 Hz and already brought onto the common time base of Section 4.6. Each eye was processed independently through the same sequence, illustrated in Figure 5.10.

The raw signal is first cleaned of blink artefacts. During a blink the eye tracker returns NaN, as the pupil is not visible; immediately before and after it, the measured diameter collapses and then recovers while the eyelid crosses the pupil, so

Table 5.2: Symbols used in the description of the estimation method.

Symbol	Meaning
D_T	Total (measured) pupil diameter
D_L	Light-driven pupil diameter predicted by the Watson–Yellott model
$D_{L,\mathrm{corr}}$	Light-driven diameter after the per-subject baseline calibration
D_C	Cognitive component of the pupil diameter
L	Absolute luminance ($\mathrm{cd\,m^{-2}}$)
$avgL$	Average absolute ambient luminance from the ALS
$avgRL$	Average relative luminance of the scene-camera frame
$aoiRL$	Average relative luminance of the gaze area of interest (AOI)
$aoiL$	Absolute luminance of the gaze AOI
k	Per-subject baseline calibration factor
$a,\ y,\ y_0,\ e$	Watson–Yellott parameters (field area, subject age, reference age, viewing condition)

these surrounding samples must be discarded as well. Both were handled with the procedure of Marshall [8], which identifies the blink and its neighbourhood from the signal itself and interpolates linearly across them. A second step addresses the residual transients that survive this reconstruction, together with other spurious artefacts: following John et al. [62], any sample whose velocity exceeds $5.65\,\mathrm{mm\,s^{-1}}$ was rejected as physiologically unrealistic. Finally, the signal was low-pass filtered with a Butterworth filter at 4 Hz, which removes measurement noise while preserving the pupillary dynamics of interest, known to lie below this cut-off frequency.

Both eyes were carried through this pipeline, but the analysis used a single eye per recording: the left by default, and the right only when the left was affected by excessive data loss. As no systematic difference is expected between the two, this choice does not affect the result. The output of this stage is the total pupil-diameter signal $D_T(t)$.

5.2.3 ALS signal pre-processing

The ALS delivers, at each 100 Hz sample, a raw count together with the gain at which it was acquired (Section 4.2); turning this pair into a usable luminance follows the chain of Figure 5.11.

The raw counts are first repaired of sporadic drop-outs – isolated samples that fall to zero regardless of the actual light level – which are detected and replaced by interpolation. The corrected counts are then converted to illuminance in lux by applying, for the gain in force at each sample, the calibration established in Section 4.3. Because the auto-ranging logic changes gain during a recording, the

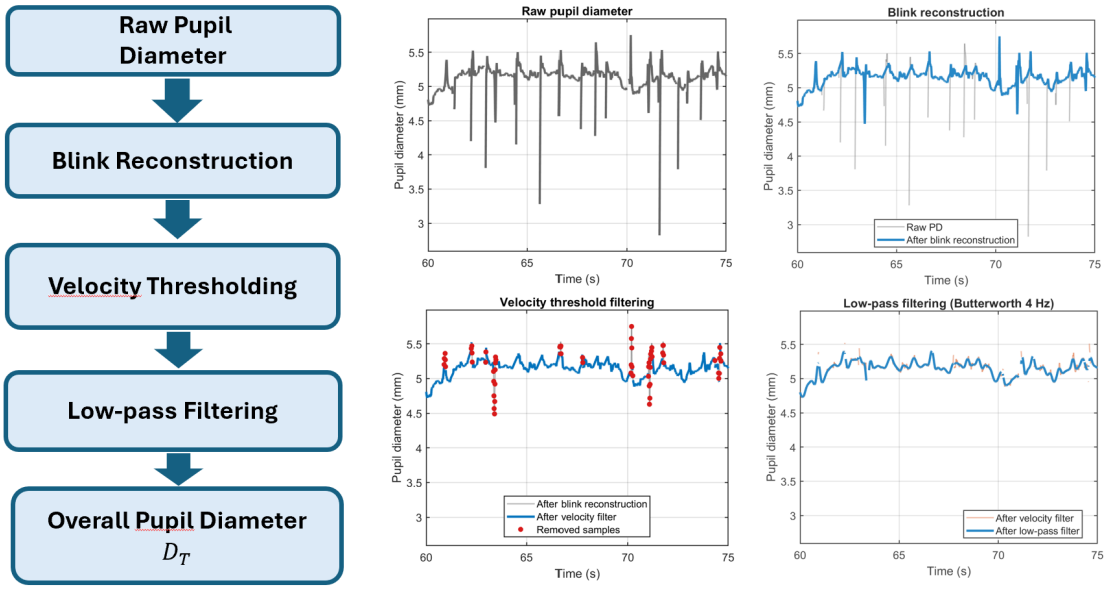


Figure 5.10: Pupil-diameter pre-processing pipeline. Left: block diagram of the processing steps, from the raw pupil diameter to the pre-processed signal D_T. Right: the effect of each step on a representative segment of the personalisation-test recording of participant 4009 (raw signal, blink reconstruction, velocity-threshold filtering with the removed samples marked in red, and low-pass filtering).

transitions between gains leave short transients in the converted signal, which were corrected next so that the illuminance is continuous across a switch. Finally, the illuminance E was converted into an average ambient luminance $avgL$ (in cd m^{-2}) through $L = E/2.2$.

This quantity is expressed in a physical unit that can be fed directly to the Watson–Yellott model (Section 2.1.1), and it is an objective measurement, independent of the auto-exposure of any camera. Nevertheless, the sensor integrates the light over its whole acceptance cone – whose relative sensitivity stays above half the on-axis value out to about ±55° off-axis – and therefore returns a single average value for the entire field of view ahead of the head. The luminance that actually drives the pupil, by contrast, is the one around the point the subject is looking at, which can differ substantially from that field average. Recovering this gaze-specific luminance is the role of the scene camera, and is the subject of the next section (Section 5.2.4).

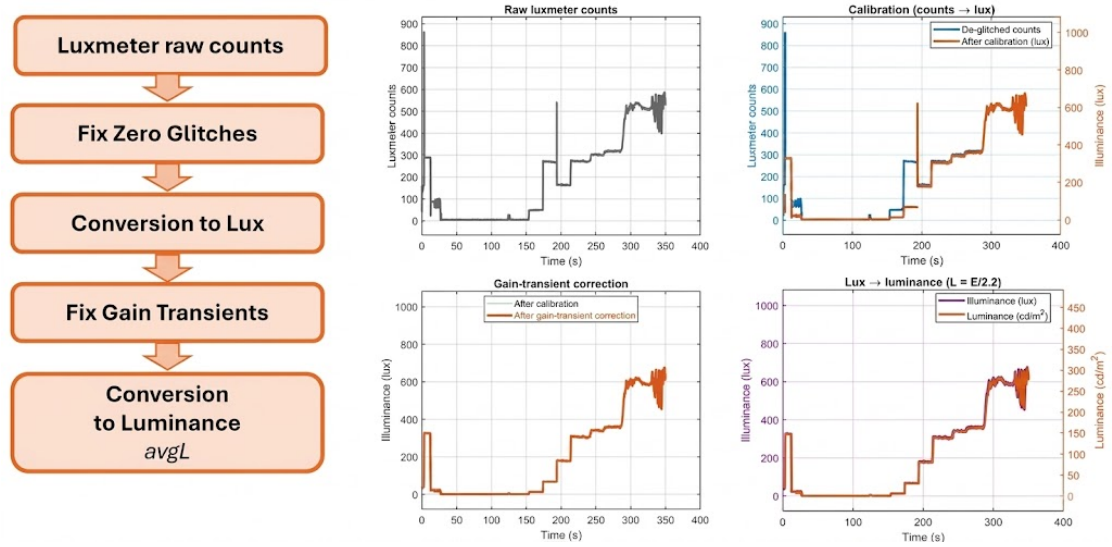


Figure 5.11: Ambient-light-sensor processing pipeline. Left: block diagram of the processing steps, from the raw sensor counts to the ambient luminance *avgL*. Right: the successive steps applied to an example signal, the full personalisation-test recording of participant 4009.

5.2.4 Pignoni's method (2021)

Scene-camera luminance extraction. The scene camera records the participant's field of view at 1920×1080 pixels and 25 frames per second (Section 3.1.1). Combined with the gaze stream, it serves to estimate the luminance in the region the subject is actually looking at, following the procedure of Pignoni et al. [27] and summarised in Figure 5.12.

For every frame the corresponding gaze samples are retrieved and mapped to pixel coordinates, $x = g_x\, W$ and $y = (1 - g_y)\, H$, with W and H the frame width and height. Because gaze is sampled at 100 Hz while the scene camera runs at 25 frames per second, each frame is associated with about four gaze samples, which brings the camera-derived luminance onto the 100 Hz rate of the other signals.

Around each gaze point an area of interest (AOI) is delineated by a Magic-Wand region growing (adapted from Reynolds' OpenCV implementation of the Photoshop tool): taking the gaze pixel as seed, all connected pixels (8-connectivity) whose colour lies within a fixed tolerance of the seed value in each channel are aggregated into the AOI. The tolerance is set to 1.5 times the average of the three per-channel standard deviations of the frame, so that it adapts to the contrast of each image. The AOI thus isolates the homogeneous surface the subject is fixating.

The whole frame and the AOI are each reduced to a single relative-luminance

value. The RGB channels are averaged over the pixels of the region,

$$\bar{I}_c^A = \frac{1}{|A|} \sum_{p \in A} I_c(p), \qquad A \in \{\text{frame}, \text{AOI}\},\ c \in \{R, G, B\}, \tag{5.2}$$

normalised to the unit range ($V_c = \bar{I}_c/255$) and linearised from the gamma-compressed sRGB encoding,

$$c_{\text{lin}} = \begin{cases} V_c/12.92 & V_c \leq 0.03928, \\ \left(\dfrac{V_c + 0.055}{1.055}\right)^{2.4} & \text{otherwise}, \end{cases} \tag{5.3}$$

before applying the photopic luminosity function [96]

$$Y^A = 0.2126\, R_{\text{lin}} + 0.7152\, G_{\text{lin}} + 0.0722\, B_{\text{lin}} \in [0,1]. \tag{5.4}$$

Evaluated on the whole frame this gives the frame relative luminance *avgRL*, and on the AOI the gaze relative luminance *aoiRL*, both at 100 Hz.

Combining the camera with the gaze provides exactly what the ALS cannot: a luminance referred to the region the subject is actually looking at, which is the light that drives the pupil. On its own, however, it is insufficient for two reasons. First, the value produced is a relative luminance in [0,1] derived from the RGB image, not a measurement in a physical unit, and so cannot be fed to the Watson–Yellott model as is. Second, the camera has no manual exposure control and applies its own auto-exposure and image processing, so its pixel values are not a stable, calibrated photometric quantity: the same scene at the same real luminance can yield different RGB depending on the camera's response. The relative, gaze-specific luminance of the camera and the absolute, field-averaged luminance of the sensor are therefore complementary, and combining them is the object of the next section.

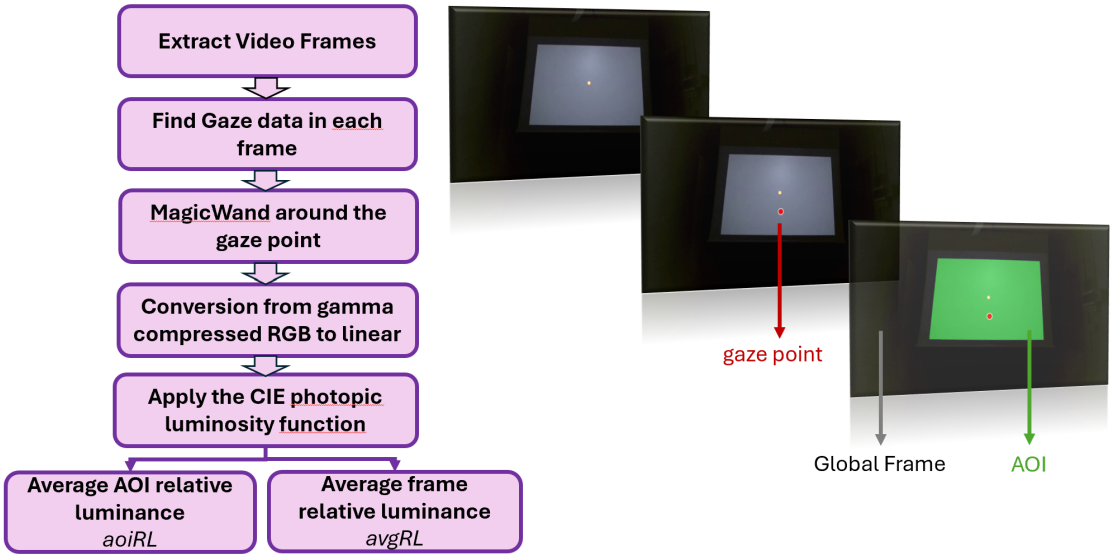


Figure 5.12: Scene-camera luminance-extraction pipeline. Left: block diagram of the processing steps. Right: three representative frames from the personalisation-test recording of participant 4009, illustrating frame extraction, the gaze point located in each frame, and the Magic-Wand area of interest grown around it.

Scene camera – ALS integration. The two luminance descriptions obtained so far are complementary: the sensor provides an absolute value ($\mathrm{cd\,m^{-2}}$) averaged over the field of view, $avgL$, while the camera provides relative values in [0,1], both frame-wide ($avgRL$) and gaze-specific ($aoiRL$). The integration combines them into a single absolute luminance at the gaze point, $aoiL$ ($\mathrm{cd\,m^{-2}}$), which is the input required by the Watson–Yellott model: the camera supplies the spatial detail – where the subject looks, relative to the frame – and the sensor supplies the absolute scale.

Before being combined the signals are conditioned. The camera-derived series, native at 25 frames per second, are aligned onto the 100 Hz sensor time axis; the sensor luminance is smoothed with a Savitzky–Golay filter (111 samples $\approx 1.1\,\mathrm{s}$, order 6) and the gaze relative luminance with a heavier Savitzky–Golay filter (301 samples $\approx 3\,\mathrm{s}$, order 1), which suppresses the abrupt jumps produced by rapid gaze shifts and changes of AOI.

For each sample the absolute gaze-point luminance is then computed as

$$lum_{\min} = \frac{avgL}{10\,avgRL + 1}, \qquad lum_{\max} = 11\,lum_{\min}, \tag{5.5}$$

$$aoiL = \frac{lum_{\max}\,aoiRL + lum_{\min}\,(1 - aoiRL)}{2}, \tag{5.6}$$

which simplifies to

$$aoiL = \frac{avgL\,(1 + 10\,aoiRL)}{2\,(1 + 10\,avgRL)}. \tag{5.7}$$

This $aoiL(t)$ is the absolute luminance at the gaze point that feeds the Watson–Yellott model in the next section.

Application of the Watson–Yellott model. The light-driven component of the pupil diameter is predicted with the Watson–Yellott model of light-adapted pupil size, introduced in Section 2.1.1 and applied here sample by sample to yield $D_L(t)$. The model computes the effective corneal flux density, the Stanley–Davies steady-state diameter and an age correction,

$$F = L\,a\,M(e), \tag{5.8}$$

$$D_{sd} = 7.75 - 5.75\,\frac{(F/846)^{0.41}}{(F/846)^{0.41} + 2}, \tag{5.9}$$

$$D_L = D_{sd} + (y - y_0)\,(0.02132 - 0.009562\,D_{sd}), \tag{5.10}$$

where L is the field luminance, a the field area (in deg^2), $M(e)$ the ocular attenuation ($M(1) = 0.1$ monocular, $M(2) = 1$ binocular), y the subject age, y_0 the reference age, D_{sd} the Stanley–Davies steady-state diameter and D_L the predicted light-adapted diameter, in millimetres.

The inputs were set as follows. The luminance is the gaze-point luminance $L = aoiL(t)$ from Section 5.2.4. The field area was set to $a = 167$ deg^2: in free-viewing, ecological conditions, with no fixed fixation target, the widest plausible field is assumed. The subject age y was taken from each participant's record, individualising the age term, while the reference age was kept at its nominal value $y_0 = 28.58$ years; the recordings being binocular, $e = 2$.

Being a generalist model, its prediction is not expected to match every individual in absolute terms: how $D_L(t)$ is anchored to each subject and then subtracted from the measured diameter to isolate the cognitive component is the subject of the next section (Section 5.2.5).

5.2.5 Baseline calibration and cognitive load estimation

The predicted diameter D_L reflects the population-average response to light and, on its own, cannot match the pupil of a given subject: at equal luminance, individuals differ. To anchor the prediction to each participant, a calibration factor is estimated on a baseline phase – the roughly 10 s of quiet frontal fixation that precedes each task, during which the cognitive load is assumed negligible. Over this window the

factor

$$k = \frac{\overline{D_T}\Big|_{\text{baseline}}}{\overline{D_L}\Big|_{\text{baseline}}} \tag{5.11}$$

(computed on a smoothed trend of the measured diameter, for robustness to noise) rescales the prediction to the subject's own pupil, giving the corrected light-driven component $D_{L,\text{corr}}(t) = k\, D_L(t)$.

The cognitive component is then the residual of the additive model of Equation 5.1,

$$D_C(t) = D_T(t) - D_{L,\text{corr}}(t), \tag{5.12}$$

computed sample by sample. By construction $D_C \approx 0$ over the baseline, so the signal is read relative to that reference: positive values indicate a cognitive load above the baseline, negative values below it. Figure 5.13 shows the resulting decomposition on an example recording.

This is where the method departs from Pignoni's. There the eye tracker returns the pupil in pixels, which must first be converted to millimetres; the conversion factor is the ratio between the mean predicted and the mean measured diameter over the test, and the resulting D_C is interpreted relative to the test average. Here the TG3 already provide the diameter in millimetres, so no such conversion is needed: the factor k instead calibrates the prediction curve to the subject's baseline, shifting the interpretation from "relative to the test average" to "relative to a cognitively unloaded baseline".

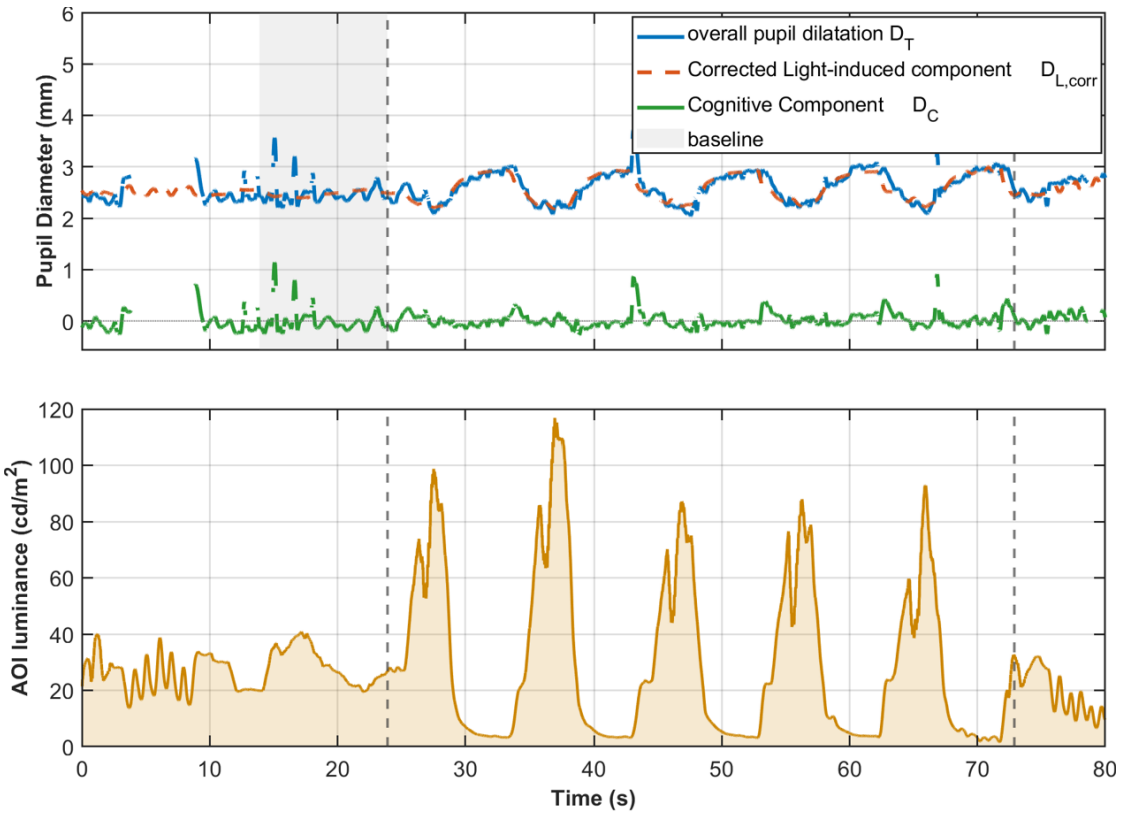


Figure 5.13: Example output of the estimation method on the locomotor task of participant 6011, a healthy older adult. Top: the measured pupil diameter D_T, the baseline-calibrated light-driven component $D_{L,\text{corr}}$ and the resulting cognitive component $D_C = D_T - D_{L,\text{corr}}$; the shaded interval is the baseline used for the calibration. Bottom: the gaze-point (AOI) luminance over the same interval. The cognitive component stays close to zero and, unlike the total diameter, does not follow the oscillations of the ambient luminance.

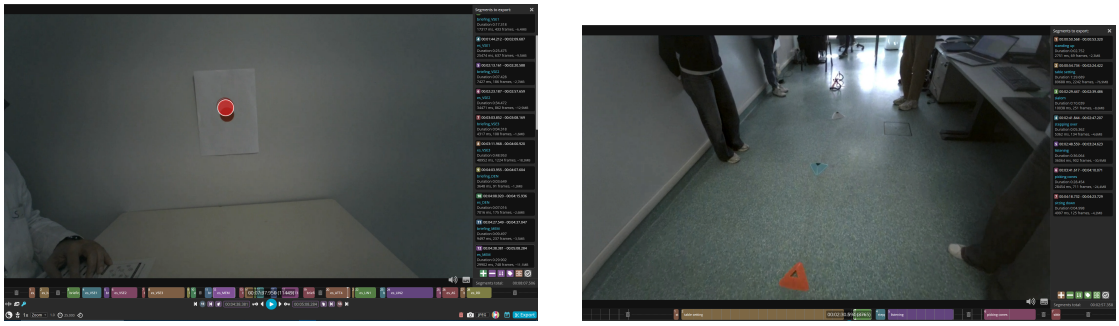
5.3 Data analysis

5.3.1 Sub-phase segmentation (SDA and MoCA)

Two of the tasks, the MoCA and the simulated daily-living sequence, are not homogeneous: each is a succession of distinct actions or exercises. To examine how the cognitive component varies within a task, and not only its overall level, both were divided into sub-phases. In each case the sub-phases were delimited manually with the LosslessCut software, by visual inspection of the TG3 scene-camera video of the recording; the resulting time boundaries were then applied to the synchronised signals.

The MoCA was preceded by the usual baseline, about ten seconds of frontal fixation, and the test itself was segmented into alternating briefing and execution phases: during a briefing the examiner explained the exercise to be performed, and during the following execution the participant carried it out, the two alternating from one exercise to the next. The cognitive load was expected to be lowest during the baseline, intermediate during the briefings and highest during the executions.

The simulated daily-living sequence was divided into the sub-phases corresponding to the parts of the circuit, namely: standing up, table setting, slalom, stepping over the obstacle, listening, picking up the cones (the return slalom, in which the cones are collected) and sitting down. The goal was to reveal any difference in cognitive effort between the phases of the task and among populations.



(a) MoCA **(b)** SDA

Figure 5.14: Manual sub-phase labelling in LosslessCut, on the TG3 scene-camera video: (a) during the MoCA, and (b) during the SDA, as the participant approaches the slalom (the obstacle is visible in the background). The bottom bar shows the video timeline with all the labelled intervals; the panel on the right lists the corresponding segments, later exported to an Excel file read by the MATLAB labelling code.

5.3.2 Statistical analysis

The analysis addressed two questions in turn: how well the method separates the light-driven and cognitive components, and what the estimated cognitive component reveals about the tasks and the two populations.

Assessment of pupil components separation. The performance of the light-driven pupil diameter estimation pipeline was first characterised through the personalisation test. Since this test was designed so that luminance would be the only source of variation in pupil diameter, the prediction of the implemented model should approximately coincide with the diameter measured by the TG3, as the

acquisition conditions allow us to assume that cognitive load and other possible causes of pupil dilation or constriction were, at the very least, minimised.

First, the signal recorded by the ALS during acquisition was used to identify intervals of near-constant illuminance within the test, isolating phases in which pupil diameter is stable with respect to the surrounding ambient luminance. For each of the selected intervals, the median pupil diameter measured by the TG3 and the pupil diameter predicted by the Pignoni method were extracted. As already discussed in Section 5.2.5, the light-driven component of pupil diameter requires a subject-specific correction factor before it can be compared with the instrument measurement; this conversion factor is determined over a baseline period. Since extracting the conversion factor from a portion of the personalisation test itself would introduce a circularity issue, undermining the objectivity of the agreement assessment between the two measures, the average correction factor obtained from the other tests was reused instead. These aspects were evaluated by computing the residual error between $D_{L,\mathrm{corr}}$ and D_T, and the percentage residual error, for each of the identified stable-illuminance intervals, in order to characterise the method's behaviour across different luminance ranges.

$$e_i = D_{T,i} - D_{L,\mathrm{corr},i}, \tag{5.13}$$

$$\mathrm{MAPE} = \frac{100}{N} \sum_{i=1}^{N} \frac{|D_{T,i} - D_{L,\mathrm{corr},i}|}{D_{L,\mathrm{corr},i}}, \tag{5.14}$$

The separation was then assessed on the tasks through the correlation between the ambient luminance and the pupil components. For every recording the Pearson correlation was computed between the ambient luminance from the ALS and each component of the diameter (the total D_T, the predicted light-driven D_L and the cognitive residual D_C) over the activity phase; the MoCA, performed at rest and under nearly constant light, was excluded. A strong negative correlation was expected for D_T and D_L, reflecting the dominance of the light reflex, and a weak or negligible one for D_C, which would indicate that the light-driven part had been effectively removed.

Comparison across tasks, populations and sub-phases. To compare the cognitive load itself, each recording was reduced to the median of D_C over its activity phase, and the median of these per-subject values was taken for every task and each population. Comparing these values across tasks and between the healthy and vestibular groups addressed whether more demanding tasks elicit a larger cognitive component and whether the vestibular deficit alters it; no inferential test was applied to these comparisons, which are reported descriptively. The per-task median of these values was then placed alongside the subjective NASA-TLX ratings given by the participants for the corresponding activities, in particular the Mental

Demand subscale, to look for agreement between the estimated index and the perceived workload. Finally, using the segmentation of Section 5.3.1, the median cognitive component was computed within each sub-phase of the MoCA and of the simulated daily-living sequence and compared across sub-phases and populations, testing in particular the expected ordering baseline $<$ briefing $<$ execution for the MoCA.

Chapter 6

Results

6.1 Pupil diameter measurement validation

The following subsections report the comparison between the pupil diameter measured by the TG3 and the frame-by-frame manual segmentation of the eye videos recorded by the Natus VNG, taken as the reference. The final dataset comprises five participants (P15, P16, P17, P19 and P20), whose left- and right-eye recordings were treated separately; participant P13, whose right-eye recording had already been excluded owing to the almost complete absence of data recorded by the TG3, was left out of this analysis entirely, leaving 10 recordings for the analysis (5 left, 5 right).

6.1.1 Absolute diameter agreement

Using the pixel-to-millimetre conversion factor described in the methods (Section 3.3.2), the within-recording RMSE averaged 1.30 mm (SD 0.96 mm) for the left eyes and 1.53 mm (SD 1.04 mm) for the right eyes, with an overall range across subjects from 0.13 mm (smallest) to 2.60 mm (largest) (Table 6.1). The largest errors were observed for participants P15, P16 and P17, whose directly-measured HPFL (37.5 mm to 41 mm) was considerably larger than that of P19 and P20 (27.5 mm to 35 mm), and this difference accounts for most of the between-subject variability in the results.

The Bland–Altman analysis of the per-recording median diameters (Figure 6.1) showed a systematic offset between the two methods, with the eye tracker underestimating the manually segmented reference on average by 1.24 mm for the left eyes (LoA $[-3.29, +0.81]$ mm) and by 1.53 mm for the right eyes (LoA $[-3.66, +0.61]$ mm) (Table 6.2).

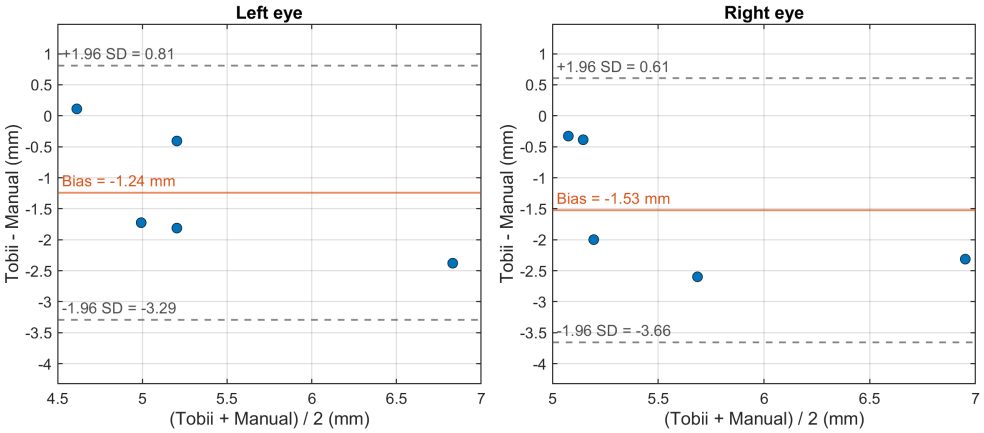


Figure 6.1: Bland–Altman analysis at the participant level (per-recording medians), separately for the left and the right eye. Lines: bias and limits of agreement (±1.96 SD); the ordinate axis is common to the two eyes.

6.1.2 Temporal tracking of pupillary fluctuations

The consistency of the temporal dynamics of the two measurements was also assessed, through the Pearson correlation between the two time series (Figure 6.2). This scale-invariant analysis revealed a moderate-to-strong correlation across all subjects, with marked between-participant variability: Pearson's r averaged 0.69 (SD 0.26) for the left eye and 0.73 (SD 0.19) for the right eye. The correlation was excellent for P17 ($r = 0.94/0.92$, left/right), high for P15 (0.87/0.88) and P19 (0.71/0.81), moderate for P16 (0.61/0.57) and weak for P20 (0.29/0.49) — the same participant that had shown the largest RMSE. The left and the right eye produced mutually consistent estimates for all the metrics analysed.

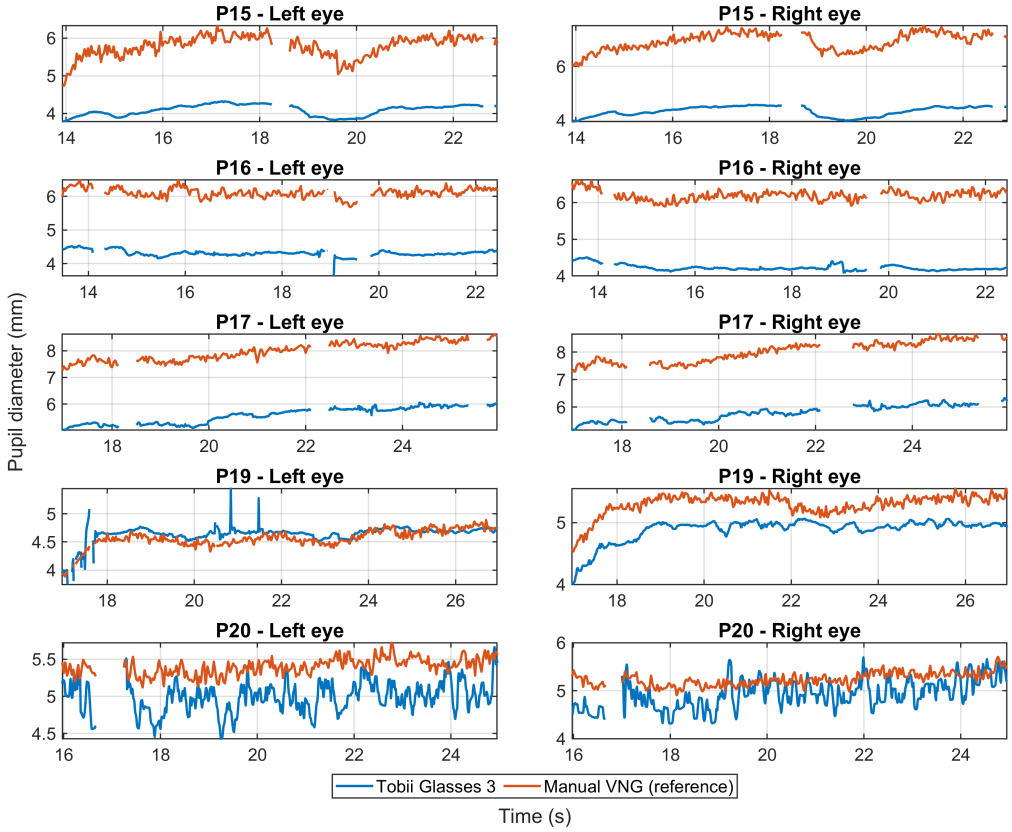


Figure 6.2: Time course of the pupil diameter for each participant and eye: TG3 (blue) and manual VNG converted with the pixel-to-millimetre conversion factor (orange). Line interruptions correspond to discarded frames (blinks); the time axis starts at the first segmented Natus frame.

Table 6.1: Per-recording metrics (participant × eye). L = left; R = right; PD = pupil diameter; RMSE = root mean square error; r = Pearson correlation.

ID	Eye	HPFL	Median PD manual (mm)	Median PD TG3 (mm)	Bias (mm)	RMSE (mm)	r
P15	L	40.0	5.86	4.13	−1.697	1.706	0.87
P16	L	41.0	6.11	4.30	−1.802	1.806	0.61
P17	L	37.5	8.02	5.64	−2.404	2.407	0.94
P19	L	27.5	4.56	4.67	+0.076	0.131	0.71
P20	L	35.0	5.41	5.00	−0.409	0.456	0.29
P15	R	40.0	6.99	4.39	−2.593	2.600	0.88
P16	R	41.0	6.19	4.20	−1.974	1.976	0.57
P17	R	37.5	8.11	5.80	−2.223	2.228	0.92
P19	R	27.5	5.34	4.95	−0.409	0.423	0.81
P20	R	35.0	5.24	4.91	−0.316	0.409	0.49

Table 6.2: Aggregate metrics across participants (per-recording medians), by eye; LoA = limits of agreement; SD = standard deviation.

Eye	HPFL (SD)	Bias (mm)	LoA (mm)	r (SD)
Left	36.20 (5.39)	−1.24	$[-3.29, +0.81]$	0.69 (0.26)
Right	36.20 (5.39)	−1.53	$[-3.66, +0.61]$	0.73 (0.19)

6.2 Assessment of pupil components separation

The method assumes that the measured pupil diameter (D_T) can be decomposed into a light-driven component ($D_{L,\text{corr}}$), reconstructed by the baseline-calibrated Watson–Yellott model (Section 5.2.5), and a residual cognitive component ($D_C = D_T - D_{L,\text{corr}}$). The effectiveness of this separation was evaluated in two complementary ways: first, by validating the pipeline, together with the Watson–Yellott predictor, on the personalisation test (Section 5.1.3) – a light-only protocol that, being nominally free of cognitive demand and covering a wide range of luminances, made it possible to characterise the accuracy of the model as a function of the light level; and second, by correlating each pupillary component with the ambient luminance obtained from the ALS, across the tasks and both groups.

6.2.1 Model accuracy over the ambient-light range

Across all participants, the personalisation test covered an overall ambient-light range from about 0 lx (near-complete darkness) to 740 lx. The values could differ between acquisitions because, with natural daylight forming the highest step of the ramp, the illuminance actually reached depended on the conditions of the specific day. An example for a representative subject is shown in Figure 6.3.

The comparison between the pupil diameter predicted by the pipeline (luminance estimation followed by the Watson–Yellott model) and the diameter recorded by the TG3 showed a consistent behaviour across participants: the magnitude of the residual error is strongly influenced by the ambient light level and grows almost monotonically as the illuminance decreases (Figure 6.4, Table 6.3, where each light band is also mapped to a typical everyday condition, for context). In the darkest conditions, below 1 lx ($n = 3$), the mean absolute residual is 0.81 mm, corresponding to a mean relative error of 24.8 %; in the 1 lx to 3.4 lx band ($n = 7$) the error remains high, 0.60 mm (18.0 %). In both bands the residual is consistently positive, that is, the model systematically under-estimates the true diameter: in keeping with scotopic vision, in the dark the pupil dilates more than the model — calibrated on the photopic regime — can predict.

Above 3.4 lx the error drops sharply and quickly settles at small values. In the 3.4 lx to 50 lx band ($n = 16$) the residual is already down to 0.13 mm (4.8 %), and it stays around this level for the rest of the tested range: 0.14 mm (5.6 %) between 50 lx to 80 lx ($n = 6$), 0.11 mm (5.2 %) between 100 lx to 320 lx ($n = 28$), and down to 0.06 mm (3.3 %) in the brightest bands, between 320 lx to 740 lx ($n = 10$ and $n = 14$). The only departure from this trend is the 80 lx to 100 lx band, where the error rises to 0.25 mm (9.1 %) and the residual reverses sign relative to the neighbouring bands (a slight over-estimation).

Overall, therefore, above about 3.4 lx the model error stays at only a few per

cent — mean absolute residual no larger than 0.14 mm, and mean relative error at most about 5.6 % (the 80 lx to 100 lx band aside) — over an interval that coincides with the entire range of light levels actually encountered during the protocol tasks.

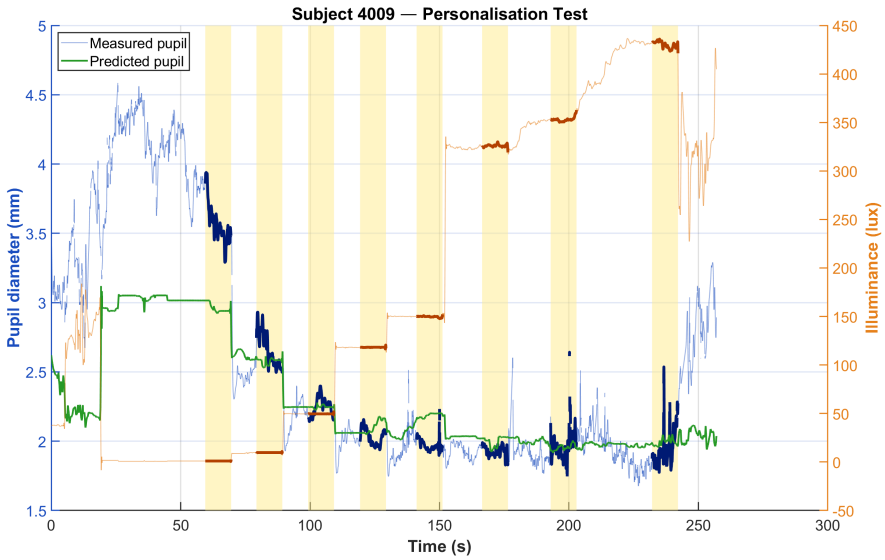


Figure 6.3: Plateau-based validation on the personalisation test (subject 4009, Test 19). The measured pupil (blue) and the baseline-calibrated prediction $D_{L,\mathrm{corr}}$ (green) are shown against the staircase luminance (orange, right axis), spanning approximately 0 cd m^{-2} to 195 cd m^{-2}. Shaded bands mark the stable luminance plateaus retained (red segments) for the accuracy analysis.

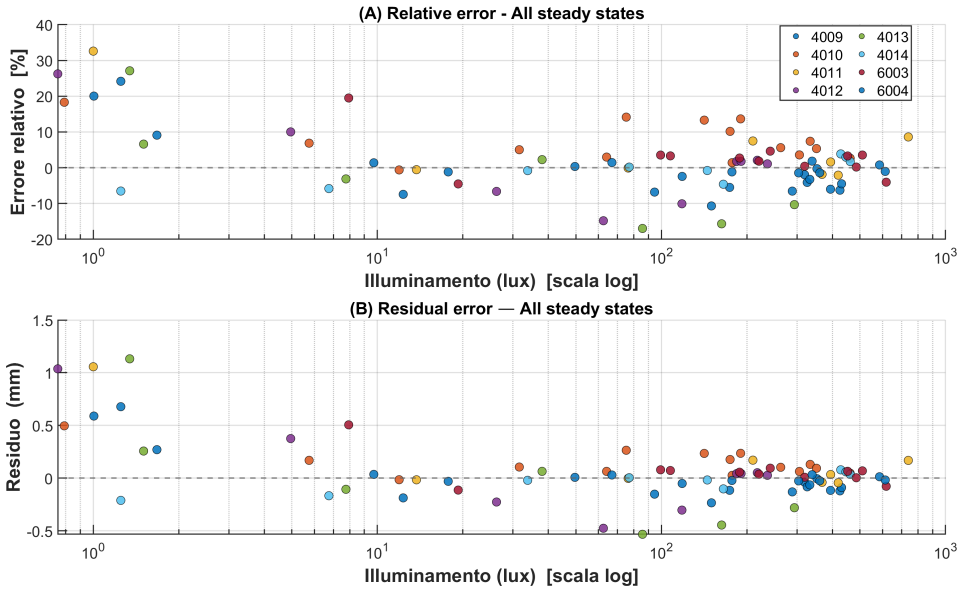


Figure 6.4: Model accuracy versus ambient light level (log scale), pooled across subjects (coloured markers: mean ± SD per bin; grey dots: individual subjects). (A) mean relative error (%); (B) mean residual $D_T - D_{L,\mathrm{corr}}$ (mm). The prediction is accurate above about 3.4 lx (mean absolute residual ≤ 0.14 mm, a few per cent) and increasingly under-estimates the pupil below it, reaching 0.81 mm (24.8 %) in the darkest band.

Table 6.3: Accuracy of the calibrated Watson–Yellott prediction on the personalisation test, by ambient-light band. N = number of steady plateaus pooled across subjects; MAE = mean absolute residual; rel. error = mean absolute relative error. Each band is mapped to a typical everyday lighting condition, for context.

Light level (lux)	Everyday condition	N	MAE (mm)	Rel. error (%)
0.0–1.0	< Full moon	3	0.81	24.8
1.0–3.4	Full moon → Civil twilight	7	0.60	18.0
3.4–50	Civil twilight → Living room	16	0.13	4.8
50–80	Family living room → Hallway	6	0.14	5.6
80–100	Hallway/bathroom → Very dark overcast day	3	0.25	9.1
100–320	Very dark overcast day → Office lighting	28	0.11	5.2
320–400	Office lighting → Sunrise	10	0.06	3.3
400–1000	Sunrise → overcast day	14	0.06	3.3

6.2.2 Correlation of the pupillary components with luminance

Having established the accuracy and the robustness of the light model, its removal was then assessed on the tasks through the correlation between the ambient luminance and the pupillary components (Section 5.3.2). The MoCA was excluded from this analysis: performed with the participant seated at the table, it entails very little variation of the ambient luminance, which makes the Pearson coefficient unreliable – dominated by noise rather than by a genuine dependence of the pupil on luminance. The analysis therefore covers the four U-turn walking recordings (UTWcw, UTWccw, UTWTcw and UTWTccw; Section 5.1.3) and the SDA.

Figure 6.5 reports, for each task and separately for the two groups, the Pearson correlation of the three pupillary components with luminance (mean ± SD across subjects; $n = 7$ healthy, $n = 5$ vestibular; see also Table 6.4). As expected, the overall diameter and the light-driven component were negatively correlated with luminance in every task. The light-driven component captured almost all of the light-related variance, with correlations that were both high and consistent across tasks (healthy: -0.88 ± 0.09 on the U-turn walking tests and -0.81 ± 0.08 on the SDA; vestibular: -0.87 ± 0.06 and -0.88 ± 0.05, respectively). The correlation of the overall diameter was instead more variable and task-dependent: it was strongest during the U-turn walking tests, where the subject is in continuous motion and the gaze-contingent light input dominates the pupillary signal (healthy -0.58 ± 0.26; vestibular -0.54 ± 0.21), and weaker during the SDA (healthy -0.40 ± 0.07; vestibular -0.32 ± 0.15), which also comprises quasi-static phases (e.g. table setting, listening) with little light variation and, therefore, a larger relative weight of the non-light components. Most importantly, the cognitive component was close to zero in almost every task, indicating that the separation behaves consistently across very different conditions. In the healthy group the task-wise correlations ranged from -0.18 to -0.02 (U-turn walking pooled -0.09 ± 0.36; SDA -0.13 ± 0.14). In the vestibular group the four U-turn walking tests showed a small but consistent positive residual ($+0.09$ to $+0.13$; pooled $+0.11 \pm 0.30$), while the SDA was essentially null (-0.07 ± 0.19).

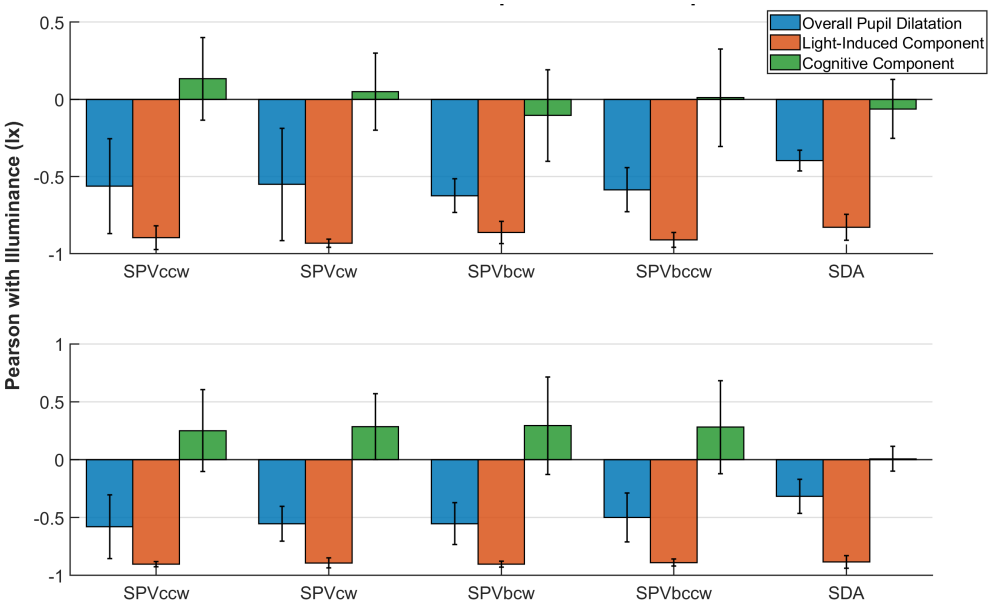


Figure 6.5: Pearson correlation (mean ± SD across subjects) between luminance and the three pupillary components – total diameter (D_T), light-driven component (D_L) and cognitive component (D_C) – for each task (the four U-turn walking recordings and the SDA), in healthy older adults (top) and vestibular patients (bottom). D_T and D_L are strongly anti-correlated with luminance and vary with the light content of the task, whereas D_C stays close to zero and stable across tasks.

Table 6.4: Pearson correlation with luminance (mean ± SD across subjects) for the three pupillary components, pooled over the four U-turn walking recordings and reported separately for the SDA ($n = 7$ healthy, $n = 5$ vestibular).

Component	Group	U-turn walk	SDA
Overall diameter	Healthy	-0.58 ± 0.26	-0.40 ± 0.07
Overall diameter	Vestibular	-0.54 ± 0.21	-0.32 ± 0.15
Light-driven comp.	Healthy	-0.88 ± 0.09	-0.81 ± 0.08
Light-driven comp.	Vestibular	-0.87 ± 0.06	-0.88 ± 0.05
Cognitive comp.	Healthy	-0.09 ± 0.36	-0.13 ± 0.14
Cognitive comp.	Vestibular	$+0.11 \pm 0.30$	-0.07 ± 0.19

6.3 Cognitive workload analysis

6.3.1 Comparison across tasks and cohorts

To assess whether the cognitive component is a more valid index of cognitive engagement than the raw pupil diameter, the two measures were compared across four tasks (UTW, UTWT, SDA and MoCA; for the two U-turn walking tasks, the clockwise and counter-clockwise recordings were averaged per subject) and the two cohorts (Figure 6.6, Table 6.5). The raw pupil diameter (Figure 6.6, left) showed large between-subject variability, especially in the vestibular cohort (IQR width 0.69 mm to 1.80 mm, versus 0.41 mm to 0.58 mm in the healthy group), and the vestibular medians were systematically higher than the healthy ones in every task. This apparent group difference, however, is largely an artefact of light exposure rather than a genuine difference in cognitive engagement. The cohort gap was largest at UTWT, reaching 0.70 mm (median 3.40 mm versus 2.70 mm). This is also the task with the largest between-cohort difference in illuminance, the healthy participants being exposed to about 2.7 times more light on average (265 lx, SD 112 lx, versus 100 lx, SD 44 lx), so their smaller pupils reflect a stronger light reflex, not a lower cognitive load.

Isolating the cognitive component removed most of this spurious variance (Figure 6.6, right; Table 6.5). Between-subject variability (IQR width) decreased markedly in both cohorts, most clearly in the vestibular one, where it fell from about 0.69 mm to 1.80 mm for the raw diameter to about 0.06 mm to 0.15 mm for the cognitive component; in the healthy cohort it fell from about 0.41 mm to 0.58 mm to about 0.20 mm to 0.25 mm. A consistent task gradient then emerged in both cohorts alike, one not visible in the raw signal: the cognitive component increased monotonically from the pure U-turn walking task, which involves the least cognitive demand, through the ball-tracking and daily-living tasks, up to the MoCA. In the healthy group the median rose from −0.15 mm (UTW) to 0.07 mm (UTWT), 0.23 mm (SDA) and 0.37 mm (MoCA); in the vestibular group the same ordering held, from −0.03 mm (UTW) to 0.18 mm (UTWT), 0.27 mm (SDA) and 0.30 mm (MoCA) (full IQRs in Table 6.5). This ordering was recovered only after removing the light component: in the raw diameter it was masked, and in the healthy group the tracking task even appeared lower than the pure walk because of its higher illuminance.

The two cohorts differed, however, in how distinctly the purely cognitive task separated from the rest. In the healthy group the MoCA stood out clearly as the task eliciting the largest cognitive component, 0.14 mm above the SDA, consistent with it being the most cognitively demanding task of the protocol. In the vestibular group this distinction was far less pronounced: the SDA and the MoCA settled on very similar median values (0.27 mm and 0.30 mm, respectively, a gap of only

0.03 mm, with largely overlapping interquartile ranges), so that, unlike in the healthy group, the purely cognitive task did not clearly emerge as more demanding than the multi-step motor-cognitive one.

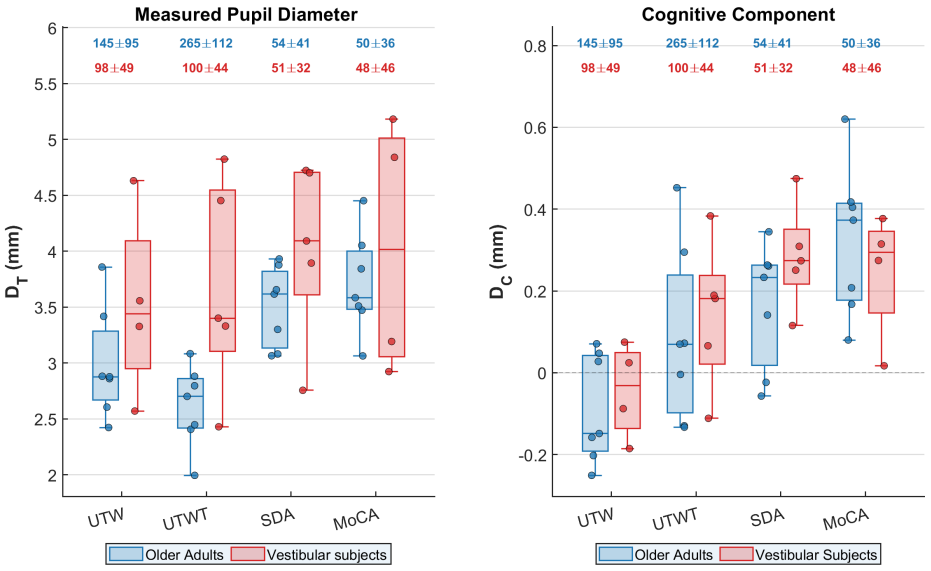


Figure 6.6: Per-task comparison of the raw pupil diameter D_T (left) and the extracted cognitive component D_C (right), by cohort (older adults in blue, vestibular patients in red) across four tasks (UTW, UTWT, SDA and MoCA); for the two U-turn walking tasks, the clockwise and counter-clockwise recordings were averaged per subject before plotting. Boxes show the distribution of the per-subject medians and the dots the individual subjects; the numbers above each panel are the mean illuminance ± SD (lx) recorded during the test, for that task and cohort (healthy top, vestibular bottom). The large between-cohort gap in the raw diameter, driven by differences in light exposure, collapses once the light component is removed, leaving a consistent task gradient shared by both cohorts.

Table 6.5: Pupil diameter and cognitive component per task and cohort (median [IQR] of the per-subject values; N varies for the vestibular group owing to missing recordings in two tasks).

Task	Cohort	N	Pupil diameter (mm)	Cognitive comp. (mm)
UTW	Healthy	7	2.88 [2.73, 3.15]	−0.15 [−0.18, +0.04]
	Vestibular	4	3.44 [3.14, 3.83]	−0.03 [−0.11, +0.04]
UTWT	Healthy	7	2.70 [2.43, 2.84]	+0.07 [−0.07, +0.18]
	Vestibular	5	3.40 [3.33, 4.45]	+0.18 [+0.07, +0.19]
SDA	Healthy	7	3.62 [3.19, 3.77]	+0.23 [+0.06, +0.26]
	Vestibular	5	4.09 [3.89, 4.70]	+0.27 [+0.25, +0.31]
MoCA	Healthy	7	3.58 [3.49, 3.95]	+0.37 [+0.19, +0.41]
	Vestibular	4	4.02 [3.13, 4.93]	+0.30 [+0.21, +0.33]

6.3.2 Association with subjective ratings (NASA-TLX)

The isolated cognitive component was compared with the participants' own ratings of task difficulty, collected at the end of each activity, so as to benchmark the pupil-based index of cognitive load against one of the most widely used tools for its subjective assessment, the NASA-TLX questionnaire [7]. Since the questionnaire was administered only after the UTWT, SDA and MoCA tasks, the comparison is restricted to these three tasks and is purely descriptive, given the small number of activities and participants.

Of the six subscales of the NASA-TLX, the most relevant one for this study was Mental Demand, since it was expected to reflect the pupillary cognitive component. The two measures agree on how the SDA and the MoCA rank against each other, but this ranking is different in the two cohorts. In the healthy group, the MoCA was rated as clearly more mentally demanding than the SDA (median 12.0 [IQR 10.25, 15.0] versus 5.0 [IQR 3.25, 8.75], on the 0 to 20 scale of the questionnaire), and the cognitive component shows the same pattern (median 0.37 mm [IQR +0.18, +0.41] versus 0.23 mm [IQR +0.02, +0.26]) (Figure 6.7). In the vestibular group, instead, the two tasks were rated as almost equally demanding (MoCA median 10.5 [IQR 5.5, 17.5] versus SDA 10.0 [IQR 5.75, 15.25]), and the cognitive component again shows this same near-equality (0.29 mm [IQR +0.15, +0.35] versus 0.27 mm [IQR +0.22, +0.35]) (Figure 6.8). The UTWT task, however, shows a mismatch between the two measures, in both cohorts. Its perceived mental demand was rated as similar to the SDA (healthy: median 5.0 [IQR 2.75, 9.5]; vestibular: median 10.0 [IQR 4.25, 10.0], the same as the corresponding SDA rating), but its cognitive component was clearly lower than the SDA's in both groups (healthy: median

0.04 mm [IQR −0.06, +0.38]; vestibular: median 0.09 mm [IQR +0.01, +0.27]). In the healthy group, this value was also much more variable across subjects than for the other two tasks.

The subjective ratings therefore corroborate what the cognitive-component analysis had already shown about the relative cognitive load of the protocol tasks, including the population-dependent difference that appears on both indices: the healthy adults reached their clear maximum at the MoCA, whereas in the vestibular patients the SDA and the MoCA were nearly indistinguishable on both measures. What the questionnaires cannot reveal, unlike pupillometry, is how the cognitive effort evolves from moment to moment within a single task, being compiled only retrospectively. It is precisely here that the value of the pupil as a cognitive-load estimator emerges, in its ability to pinpoint the critical, most demanding moments within a task, as well as the least demanding ones. This aspect is developed in the next section (Section 6.3.3), which examines the two most cognitively demanding tasks, the SDA and the MoCA, at the level of their sub-phases.

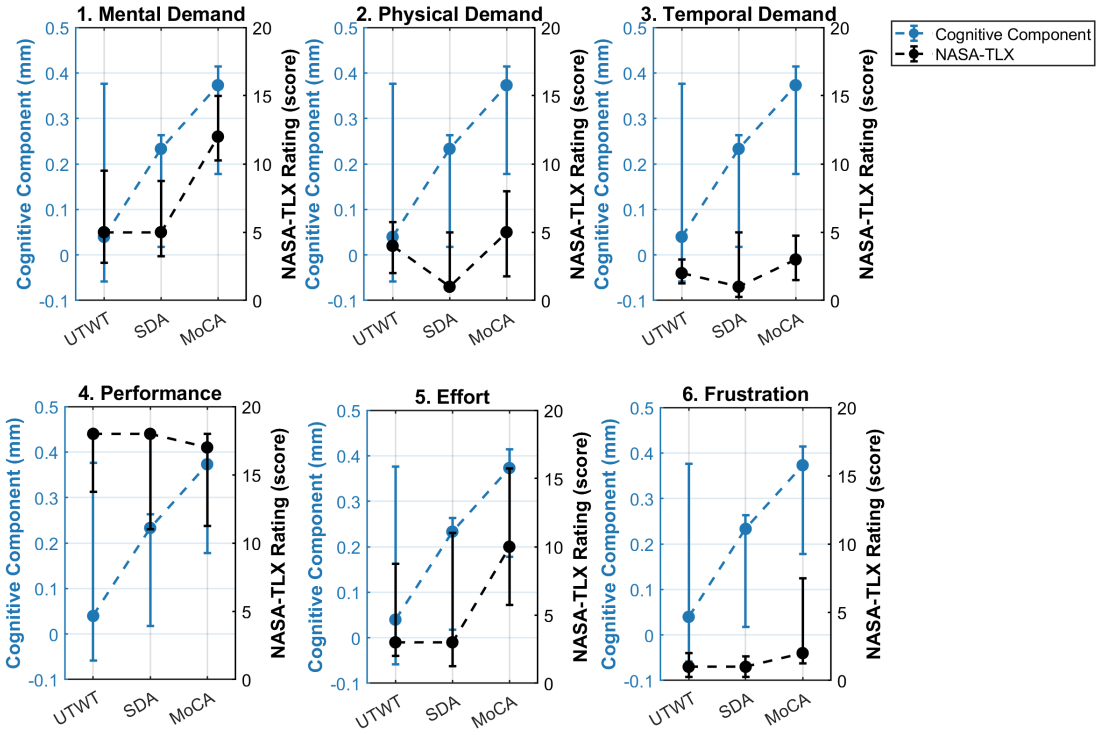


Figure 6.7: Cognitive component versus NASA-TLX subscales in the healthy older adults, for the three tasks rated with the questionnaire (the UTWT, the SDA and the MoCA). Each panel overlays the cognitive component D_C (left axis, blue) and one NASA-TLX dimension (right axis, black), both as median [IQR] across participants. The Mental Demand subscale (top left) follows the cognitive component, both peaking at the MoCA.

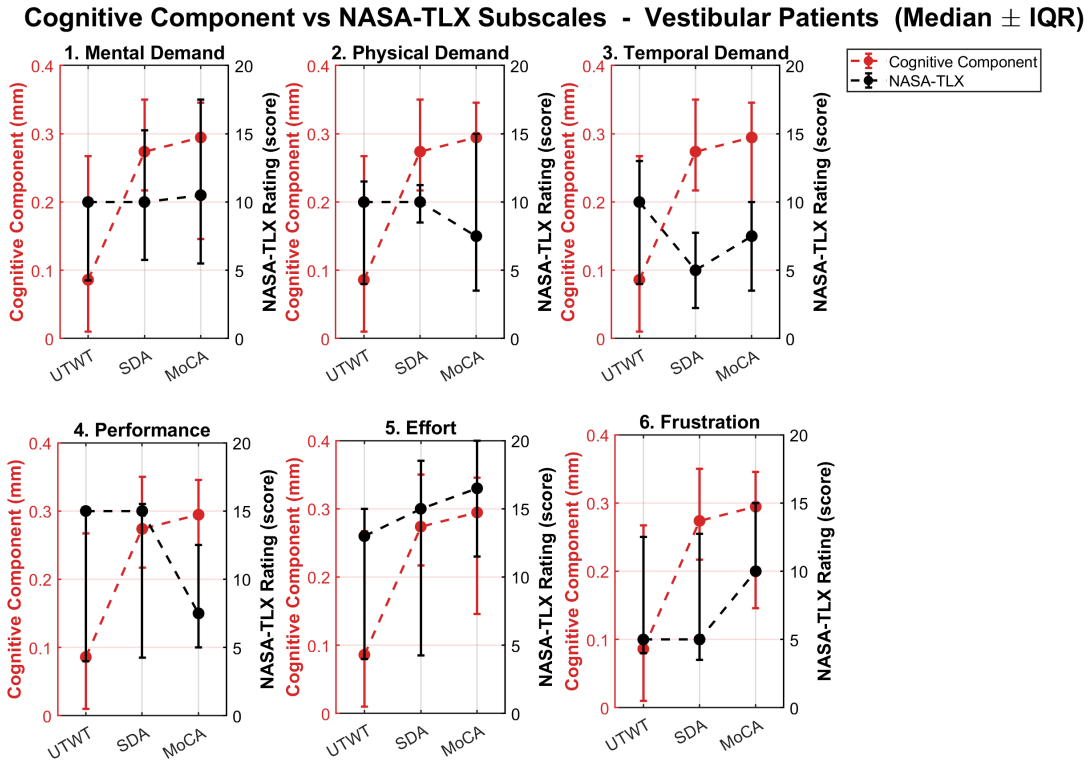


Figure 6.8: Cognitive component versus NASA-TLX subscales in the vestibular patients, for the UTWT, the SDA and the MoCA. Panels and axes are as in Figure 6.7. Here both the Mental Demand subscale and the cognitive component show the SDA and the MoCA at nearly the same level, unlike the clear MoCA peak seen in the healthy group.

6.3.3 Subphase analysis

MoCA. The sub-phase analysis of the MoCA showed that the cognitive component is not held constantly high throughout the test, but is modulated in line with the nature of its phases. The baseline intervals at the start of the test are clearly those with the lowest cognitive component; from then on, exercise after exercise, the cognitive component rises on average during the briefing — when the participant listens to and memorises the task to be performed — and rises further during the execution of the exercise, the phases that emerge as the most demanding. Because the test is a repeated sequence of briefing and execution phases, averaging across subjects produces a clear alternating pattern — very regular in the healthy group, more irregular in the vestibular one — in which the cognitive component rises and falls between briefing and execution (Figure 6.9). The same ordering is recovered at the macro-phase level in both groups, the group-mean cognitive component increasing from baseline (healthy 0.02 ± 0.02, vestibular 0.00 ± 0.04 mm) through briefing (0.22 ± 0.11 and 0.18 ± 0.03) to execution (0.39 ± 0.16 and 0.31 ± 0.03).

Across the individual exercises, labelled by the MoCA macro-area to which they belong (see caption), the highest execution-phase values were elicited, in the healthy group, by the attention item ATT4 (0.50 ± 0.28 mm) and the visuospatial/executive items VSE1–VSE3 (0.48 ± 0.27, 0.44 ± 0.34 and 0.43 ± 0.16), together with ATT2 (0.43 ± 0.38); in the vestibular group, by the attention items ATT2 and ATT4 (0.47 ± 0.12 and 0.45 ± 0.14) and by the memory-encoding item MEM (0.46 ± 0.16). The two attention items ATT2 and ATT4 were thus among the most demanding in both populations.

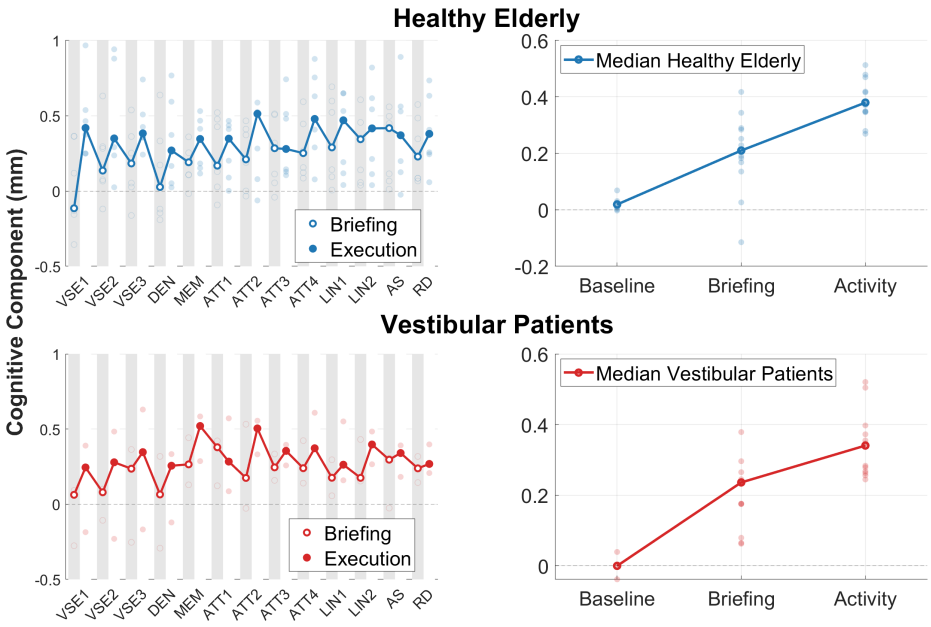


Figure 6.9: Dynamic analysis of the cognitive component during the MoCA, for the healthy older adults (top, $n = 7$) and the vestibular patients (bottom, $n = 5$). Left: per-exercise cognitive component split into briefing (open markers) and execution (filled markers) and connected in chronological order (grey/white bands = briefing/execution); faded dots are individual subjects, the line is the group mean. Right: macro-phase summary (baseline, briefing, execution). Exercise labels denote the MoCA macro-area: VSE = visuospatial/executive (VSE1 trail-making, VSE2 cube copy, VSE3 clock drawing), DEN = naming, MEM = memory, ATT = attention (ATT1 forward and ATT2 backward digit span, ATT3 target detection, ATT4 serial subtraction), LIN = language (LIN1 sentence repetition, LIN2 verbal fluency), AS = abstraction, RD = delayed recall.

Simulated daily-living sequence. The cognitive component (D_C), estimated as the residual of the pupil diameter after removing the light-induced response, was quantified in each of the seven SDA sub-phases by summarising every phase with the per-subject median and comparing the populations through the group median (Table 6.6, Figure 6.10). In both populations the cognitive load is not uniform along the activity but follows a structured profile: the minimum falls systematically in the passive listening phase (0.14 [0.20] in the healthy group, 0.12 [0.16] in the vestibular; median [IQR] throughout), while higher values occur in the phases involving motor coordination and balance control. The between-population comparison shows a consistent pattern: the vestibular patients present a higher cognitive component than the healthy adults precisely in the phases that engage balance and locomotion — standing up (0.35 [0.17] vs 0.16 [0.16]), slalom (0.35 [0.11] vs 0.15 [0.25]) and stepping over (0.36 [0.20] vs 0.21 [0.16]) — whereas the two groups converge in the final phase (sitting down, 0.34 [0.20] vs 0.33 [0.25]) and in listening. The picking-cones phase is the exception, with the healthy group showing the higher value (0.27 [0.18] vs 0.16 [0.22]).

Table 6.6: Cognitive component ΔD_C (mm) in the seven SDA sub-phases: group median [IQR] across subjects ($n = 7$ healthy, $n = 5$ vestibular).

Sub-phase	Healthy ($n = 7$)	Vestibular ($n = 5$)
Standing up	0.16 [0.16]	0.35 [0.17]
Table setting	0.30 [0.12]	0.37 [0.16]
Slalom	0.15 [0.25]	0.35 [0.11]
Stepping over	0.21 [0.16]	0.36 [0.20]
Listening	0.14 [0.20]	0.12 [0.16]
Picking cones	0.27 [0.18]	0.16 [0.22]
Sitting down	0.33 [0.25]	0.34 [0.20]

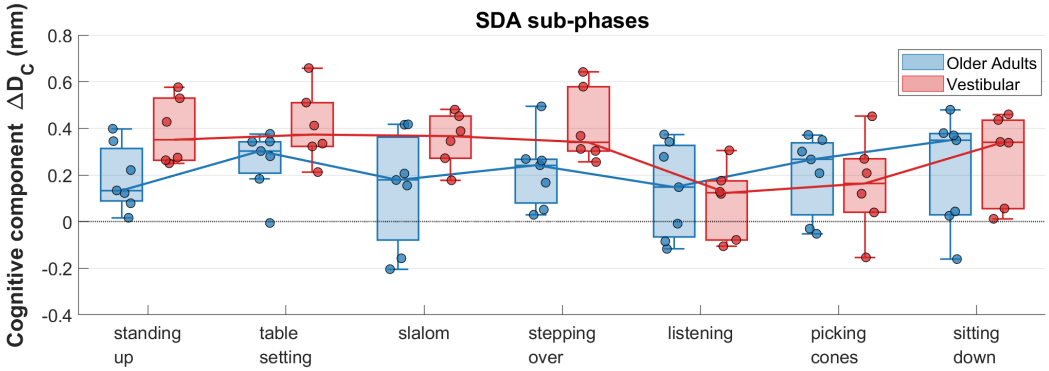


Figure 6.10: Cognitive component ΔD_C in the seven SDA sub-phases, by cohort (older adults in blue, $n = 7$; vestibular patients in red, $n = 5$). Boxes show the distribution of the per-subject medians, dots the individual subjects, and the solid line the group median. The cognitive component is lowest in the passive listening phase and higher in the phases involving locomotion and balance control.

Chapter 7

Discussion

7.1 Reliability of Tobii Pro Glasses 3 pupil measurements

7.1.1 Agreement and temporal tracking results

This study compared the pupil-diameter measurements of a wearable eye tracker (TG3) with the frame-by-frame manual segmentation of Natus VNG eye videos, under a static protocol. Two main findings emerge.

The pixel-to-millimetre conversion of the manual reference relies on an independent measurement, the horizontal palpebral-fissure length (HPFL) each participant measured directly with a ruler (Section 3.3.2). With this scale, absolute agreement was modest (RMSE, mean ± SD across subjects: right eye 1.53 mm (SD 1.04 mm), left eye 1.30 mm (SD 0.96 mm)), with the eye tracker under-estimating the manually segmented reference by 1.24 to 1.53 mm on average. This discrepancy is not uniform across participants, however, and traces largely back to the reference scale itself: for three of the five participants (P15, P16, P17) the directly-measured HPFL (37.5 mm to 41 mm) lies far outside the normative adult range reported in the literature (27 mm, SD $\approx$ 1.6 mm, pooling male and female values from Knezi et al. [74]; a comparable range is reported by Read et al. [73]), and it is precisely these three recordings that show the largest bias and RMSE. The remaining two participants, whose HPFL falls within or close to this normative range (P19: 27.5 mm; P20: 35 mm), show a markedly smaller discrepancy, reinforcing the interpretation that inaccuracies in the self-administered ruler measurement, rather than a genuine disagreement between the two instruments, are the principal driver of the observed bias.

A second, independent factor may also contribute, affecting the manual segmentation itself rather than the conversion factor. The TG3 reportedly derive

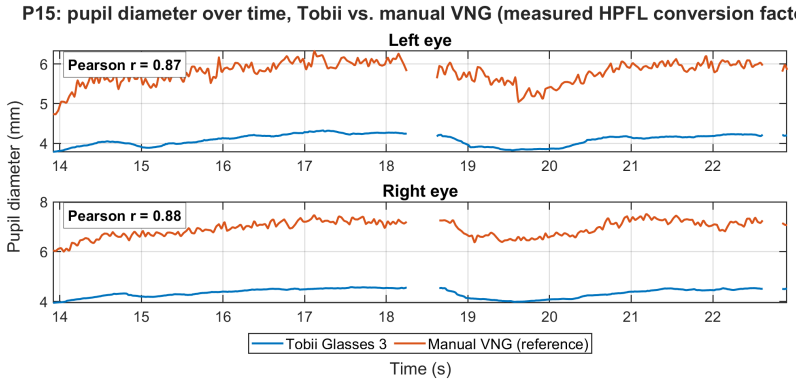
pupil diameter from a three-dimensional reconstruction of the eye that explicitly compensates for the optical effects of the cornea, whereas the manual segmentation used here measures only the apparent, two-dimensional pupil visible in the raw VNG image, with no such correction. Petersch and Dierkes [52] show that this kind of uncorrected measurement is subject to a systematic bias from corneal refraction and from the foreshortening of the pupil image with gaze angle.

The choice of conversion factor, on the other hand, does not affect the temporal-tracking metrics, which are scale-independent by construction. The device tracked the temporal fluctuations of the pupil with an on-average moderate-to-strong correlation (mean $\pm$ SD Pearson coefficient across subjects: right eye 0.73 ± 0.19, left eye 0.69 ± 0.26), albeit with marked inter-individual variability. Since pupillometry is primarily concerned with detecting the pupil's variations associated with cognitive load – recognising a dilation or a constriction driven by a change in cognitive effort or in ambient light, rather than the pupil's absolute size – these temporal-tracking results are the more relevant of the two.

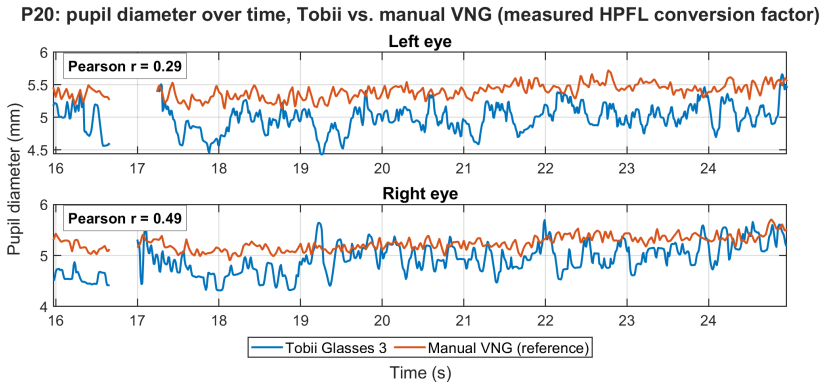
The between-subject spread in temporal tracking is attributable to the low signal-to-noise ratio inherent in the static protocol: with the pupil nominally constant, the informative content reduces to small physiological fluctuations, so that any measurement or segmentation noise weighs heavily on the correlation. This is consistent with the anomalous behaviour of P20, whose TG3 trace was visibly noisier, and with the intermediate performance of P16; conversely, the subjects with larger, cleaner fluctuations (P17 and P15) exceeded correlations of 0.85 (Figure 7.1).

7.1.2 Limitations of the validation protocol

These results must be read against several limitations. First, the sample is small, comprising only five subjects. Second, the reference scale itself is a limitation: HPFL was self-measured by each participant with a ruler of 0.5 mm resolution, and the implausible values obtained for three of the five participants suggest this procedure is prone to a sizeable, uncontrolled error well beyond the nominal ruler resolution; a measurement performed or verified by the operator, rather than relying on self-report, would be needed to rule this out. Third, the manual reference is imperfect in a further respect, since pupil diameter is taken as the major axis of the segmented ellipse with no correction for corneal refraction or for the fixed foreshortening imposed by the VNG camera geometry, unlike the three-dimensional, refraction-aware reconstruction reportedly used by the TG3 [52]. Fourth, only a single static condition was examined, which limits the dynamics available for assessing the tracking.



(a) P15



(b) P20

Figure 7.1: Pupil-diameter time courses measured by the TG3 and by the Natus VNG reference for two representative participants: (a) P15, whose clean fluctuations yield a high correlation between the two traces, and (b) P20, whose visibly noisier TG3 trace yields a low correlation. The corresponding Pearson coefficients are reported in Table 6.1 (P15: $r = 0.87/0.88$; P20: $r = 0.29/0.49$, left/right eye).

7.2 Isolating the cognitive component under uncontrolled lighting

The protocol adopted for this study set a demanding test for the removal of the light-driven component of the pupil diameter, since recordings were carried out under lighting that was not controlled a priori and with no constraint on the subject's movement. The aim was to probe the separation pipeline proposed by Pignoni et al. [27] to assess whether pupillometry can support the estimation of cognitive load in settings closer to real-world use – and thereby open the way to its application in dynamic contexts such as the assessment of cognitive-motor interference.

7.2.1 Model accuracy and robustness across the luminance range

A second aspect the study set out to address was the behaviour of the Watson–Yellott prediction across the range of ambient illuminance it must operate under, and in particular its accuracy at the low-light end, where the model is most exposed to error. The prediction proved accurate from about 3.4 lx upward, with the mean absolute residual never exceeding 0.14 mm (at most about 5.6 % relative error, the 80 lx to 100 lx band aside) across the entire range of light levels actually encountered during the six tasks. Below this threshold, however, the error grew sharply and consistently in one direction: the pupil was increasingly under-estimated, reaching a mean absolute residual of 0.60 mm (18.0 %) between 1 lx to 3.4 lx, and 0.81 mm (24.8 %) below 1 lx, in near-total darkness.

This systematic under-estimation is best read as a limitation of the light model itself rather than of the pipeline that feeds it. This is consistent with the fact that the core shape of the Watson–Yellott formula and its age-adjustment term were fitted on photopic-range data (9 cd m^{-2} to 4,400 cd m^{-2} for the age effect), with the authors themselves cautioning against extrapolating beyond that range [39]. As ambient light approaches zero, the true pupil continues to dilate toward the physiological limits of scotopic vision, while the model's prediction saturates well before reaching that limit, unable to reproduce the full extent of the eye's dark-adapted response. The residual therefore grows precisely because the model, calibrated mainly on the photopic regime, cannot follow the pupil as it approaches its maximum achievable diameter – an asymptotic mismatch rather than a random source of error, and consistent with its systematically one-signed (under-estimating) character in both low-light bands.

Beyond accuracy at any single light level, robustness to changes in luminance is a further requirement the model must satisfy, one that matters especially in a

dynamic, real-world context such as the one this pipeline was applied to in the present study: unlike a controlled laboratory setting, a subject moving through an ecological protocol is continuously exposed to changing illumination, so that a model accurate only in the neighbourhood of its calibration point would be of limited practical use. That the residual error stays bounded within a few per cent across the entire span of light levels actually encountered during the test indicates that the model's accuracy is not confined to a narrow operating point, but holds broadly enough to support its use throughout an uncontrolled, ecologically realistic protocol.

7.2.2 Correlation analysis results

The correlation analysis first confirmed the essential role of the ALS in tracking the pupillary light reflex, evidenced by the moderate-to-high negative correlations between luminance and the total pupil diameter across tasks (healthy: -0.58 ± 0.26 on the UTW tasks and -0.40 ± 0.07 on the SDA; vestibular: -0.54 ± 0.21 and -0.32 ± 0.15, respectively). At the same time, the marked inter-subject variability of these coefficients, together with their dependence on the task being performed, shows that ambient luminance accounts for only part of the fluctuations of the pupil diameter, the remainder reflecting other cofactors – cognitive load, arousal, hippus, stress, among others. The task dependence is itself informative: the correlation is strongest during the UTW tasks, where the subject's locomotion produces large, continuous changes in the luminance reaching the eye and the light reflex dominates the pupillary signal, and comparatively weaker during the SDA, whose quasi-static phases (table setting, listening) contribute little light variation and therefore let non-light components weigh relatively more. The MoCA was deliberately excluded from this analysis for a related reason: performed seated and at rest, it involves almost no variation in ambient luminance, which would make the Pearson coefficient a measure of noise rather than of a genuine light dependence.

The sign reversal of this correlation observed in the vestibular group deserves comment, as it may point to an intrinsic limitation of the Watson–Yellott prediction model. Being a static model, it predicts the pupil diameter that would correspond, at each instant, to the ambient luminance at that same instant, without accounting for the latency of the pupillary response [36]. This can produce an overshoot in the model's prediction: the model anticipates a variation in pupil diameter larger than the one that actually occurs, since the true response is attenuated by this latency. Such a mismatch, particularly in the presence of rapid luminance changes, can generate apparent variations in the estimated cognitive load that are in fact an artefact of the model rather than a genuine change in cognitive effort. This limitation was also reported by Pignoni and Komandur [45] in their work on pupillometry during maritime navigation – an uncontrolled, real-world setting

closely comparable to the one examined here.

A purely physiological interpretation of the phenomenon is also conceivable: the greater cognitive load required by vestibular patients to perform the task could account for a smaller-than-predicted pupillary constriction. This explanation, however, appears to fail on closer inspection, since it would equally imply an amplified dilation relative to the model's prediction whenever luminance decreases – a pattern that is not observed in the data.

Taken together, and notwithstanding these limitations, the analysis demonstrates that the pipeline is able to isolate the cognitive component of the pupil diameter even under uncontrolled, real-world-like lighting: it substantially attenuates the impact of the pupillary light reflex on the diameter's fluctuations, allowing the residual to be interpreted as an expression of the subject's cognitive load during the task.

7.3 Sensitivity of the cognitive component to task demand and impairment

Beyond its correlation with luminance, the cognitive component was examined across the six tasks to assess whether it tracks differences in cognitive-motor demand. Given the small sample sizes ($n = 7$ healthy, $n = 5$ vestibular) and the visible presence of outliers in several tasks – most notably in the UTWT conditions – the median and interquartile range (IQR) are reported here as the primary summary statistics, being considerably less sensitive to extreme values than the mean and standard deviation; the corresponding mean ± SD values remain available in Table 6.5 for reference, but are not used to draw conclusions about central tendency.

7.3.1 Pupil cognitive component across tasks

Across both cohorts, the cognitive component shows a consistent ordering by task: UTW was the least demanding, followed by UTWT, and finally SDA emerged as the most demanding among motor tasks (healthy: UTW = −0.15 mm [IQR 0.23 mm], UTWT = 0.07 mm [IQR 0.34 mm], SDA = 0.23 mm [IQR 0.24 mm]; vestibular: UTW = −0.03 mm [IQR 0.19 mm], UTWT = 0.18 mm [IQR 0.22 mm], SDA = 0.27 mm [IQR 0.13 mm]). This gradient is coherent with the nominal demand hierarchy of the protocol: the plain U-turn walking task (UTW) requires only locomotor control, the addition of a concurrent cognitive task (UTWT) introduces a dual-task load, and the SDA – a multi-step, memory-dependent activity of daily living performed while standing – combines sustained attentional and working-memory demand with postural control over a longer time window. The fact that

this ordering holds, at least at the level of the median, for both the healthy and the vestibular group suggests that the cognitive component is sensitive to task demand independently of clinical status, and supports its use as a general marker of cognitive load rather than one specific to a single population or task type.

This task-demand gradient is consistent with findings reported in the broader literature on pupillometry during motor tasks. Working within a considerably more controlled paradigm, Saeedpour-Parizi et al. [26] showed that the pupil dilates more during the preplanning phase of a walking task that requires working-memory-based route selection than during a simple, unplanned walk, directly linking pupil-derived cognitive load to the working-memory component of a motor task – a parallel to the elevated cognitive component observed here during the SDA, whose multi-step, memory-dependent structure plausibly recruits similar resources. Along a related axis, Kahya et al. [5] used the Index of Cognitive Activity to show that pupillary responses in Parkinson's disease patients scale with postural demand across a hierarchy of standing conditions of increasing difficulty, offering an independent confirmation, obtained with a coarser light-correction approach and a different clinical population, that pupil-derived measures track graded increases in task demand.

7.3.2 Association with subjective workload ratings

A further way to assess the validity of the cognitive component as a marker of task demand is to compare it against an independent, subjective measure of workload. As detailed in Section 6.3.2, the NASA-TLX questionnaire was administered only after the UTWT, SDA and MoCA, so the comparison is restricted to these three tasks and read qualitatively. On the SDA and the MoCA the two measures agree: in the healthy group the MoCA was rated as markedly more mentally demanding than the SDA, and the cognitive component followed the same ranking; in the vestibular group the two tasks were instead rated as almost equally demanding, and the cognitive component again tracked this near-equality. This agreement is itself informative, since the MoCA is a purely cognitive task performed seated and at rest while the SDA is a multi-step motor task performed while standing and moving: that the pupil-derived cognitive component tracks perceived mental demand consistently across both settings is a preliminary but encouraging indication that it can index cognitive load not only in the seated conditions for which pupillometry is traditionally used, but within a motor task as well. A broadly similar agreement between pupil dilation and perceived difficulty was reported by Manz et al. [25], using a single-item, 1–9 subjective rating of cognitive load collected after each task, following the approach of Paas [97].

The UTWT task, however, exposes a clear divergence between the two measures, present in both cohorts. Its perceived mental demand was rated as comparable

to that of the SDA (healthy: median 5.0 for both; vestibular: median 10.0 for both), yet its cognitive component was markedly lower than the SDA's in both groups (healthy: 0.04 mm versus 0.23 mm; vestibular: 0.09 mm versus 0.27 mm). A plausible explanation lies in the physical structure of the two tasks: unlike the UTWT, whose continuous, fairly uniform walking-and-tracking pattern adds only the tracking demand on top of level walking, the SDA combines several distinct postural transitions – standing up, a multi-cone slalom, stepping over an obstacle, sitting down – each requiring the coordinated processing of vestibular and visual input to maintain balance. Li et al. [98] recently showed that whole-body motion itself, independently of nominal task difficulty, significantly increases pupil dilation, attributing this to the added demand of simultaneously processing vestibular and visual information while maintaining balance.

Taken together, these comparisons should be read with the same caution applied throughout this chapter: the sample is small and the comparison purely descriptive, with no formal test applied. Still, the pattern is coherent – agreement where the two measures track the same construct, and a divergence that a plausible, literature-supported mechanism can account for – and it reinforces the case for pupillometry as a complement to subjective workload ratings, particularly for capturing the physically demanding, within-task components of cognitive-motor load that a retrospective questionnaire may under-represent.

7.3.3 Subphase analysis

Although the SDA and the MoCA emerged as the most demanding tasks overall (Section 7.3.1), the cognitive component did not remain constant within either of them: it fluctuated in line with the cognitive demand of each specific phase of the task.

MoCA. Within the MoCA, the phases eliciting the highest cognitive load were those in which the participant actually performed the item, reflecting the need to recruit executive and attentional resources, as opposed to the phases in which the examiner was still instructing the participant on the task to be carried out. This is coherent with the pattern already reported for the pipeline validation protocol applied by Pignoni et al. [27] in their study.

Beyond this instruction/execution distinction, not all MoCA items elicited the same degree of effort. In the healthy group, the items associated with the greatest cognitive component were VSE1, VSE2 and VSE3 – respectively, connecting numbers and letters in ascending order with a continuous line (a trail-making-type task), copying a cube, and drawing a clock face with the hands set to eleven-ten. In the vestibular group, the item associated with the greatest effort was instead the

memory task, which required encoding and repeating a list of five words across two trials. Two further items – backward digit span (three digits) and serial subtraction of 7 from 100 – were among the most demanding items for both groups. The convergence of backward digit span and serial subtraction across both cohorts, without a clear population-specific effect, is coherent with their nature as general working-memory and executive tasks that do not specifically engage the postural or vestibular system, and is in line with Camp et al. [99]'s demonstration, in a healthy population, that increasing the cognitive complexity of a task systematically increases its cost, independently of any vestibular-specific mechanism. This pattern also finds support in the broader pupillometry literature on tasks of a similar nature [100, 56]. Since the MoCA is not expected to engage the vestibular system in any direct way, no population-specific difference attributable to it was expected here, either in the observed cognitive load or in performance, which in this task depend instead on the participants' cognitive abilities and level of education.

SDA. The fact that the cognitive component varies across the activity and reaches its minimum during the passive-listening phase – the least attentionally demanding – is broadly consistent with what would be expected of the index: it tends to increase when the demand on executive and attentional resources grows (postural transitions, obstacle navigation, fine manipulation) and to decrease when the subject is stationary and receptive. This suggests, without amounting to a strict validation, that the index is able to track, within a single ecological task, moments that are more or less cognitively demanding – although, as discussed below, this pattern does not hold uniformly across every phase for the vestibular group. A further caveat concerns the listening phase itself: for the vestibular patients, one might also have expected a higher cognitive effort here, since some of them had a vestibular disorder that also affected the auditory system, to the point of wearing a hearing aid. This is worth keeping in mind when interpreting the results, since a study has shown that cochleo-vestibular implant use can reduce the cognitive effort required during balance-maintenance tasks [53]; its use during testing may therefore have played some role in the effort required to complete this phase.

The difference between healthy and vestibular participants is not generalised, but concentrated in the phases that most challenge balance and gait control. This is fully consistent with the pathophysiology of vestibular deficit: in the absence of reliable vestibular information, maintaining posture and locomotion cease to be largely automatic processes and instead require a greater recruitment of cognitive and attentional resources – the classic cost of the postural-cognitive dual task. The phases in which the vestibular patients show the highest load – standing up, stepping over the obstacle, moving through the slalom – are exactly those in which sensorimotor integration for balance is most heavily taxed. The transition from

sitting to standing is a complex motor event that abruptly shifts the centre of gravity and requires a rapid readjustment of the limits of stability [101]. Pupillometric research in patients with vestibular hypofunction shows that peak pupil dilation occurs right at the onset of the balance task, in the first seconds of the postural transition, reflecting the attentional "shock" and the recalibration effort required of the system to regain stability [53]. Table setting can be regarded as a dual task in its own right, since it requires both executing a manual activity and keeping track of the remaining steps of the sequence: because the patient's attentional resources are limited and already substantially taken up by simply maintaining an upright stance, adding the manual task demands a further, consciously mediated allocation of attention, consistent with the higher cognitive component recorded here relative to the healthy group [28]. The slalom and the obstacle crossing both require walking while performing active head movements, which the literature identifies as one of the most demanding rehabilitative and assessment tasks for vestibular patients [102], owing to the deficit in the vestibulo-ocular reflex (VOR).

These results, although strictly preliminary given the small sample, put forward the SDA as an ecological test capable of discriminating between healthy and vestibular subjects, and pupil dilation as a useful index for revealing where motor automaticity breaks down and conscious, effortful control becomes necessary.

7.4 Limitations

The most consequential limitation of this study is the small size of the sample examined: as a result, the analysis was carried out at a purely descriptive level, since formal statistical inference is not possible, and this places an important constraint on the interpretation of the results, particularly given the considerable inter-subject variability that emerges from them.

A further limitation is related to the use of eye tracking in dynamic contexts: the frequent eye and head movements involved likely had a negative effect on signal quality, causing noise and data loss [48, 50].

Regarding the implemented pipeline, limitations may be both hardware- and software-related. From a hardware standpoint, a first limitation concerns the positioning of the ALS, which cannot guarantee a field of view identical to that of the scene camera, whose settings in turn limit the accuracy of the luminance estimate (distortion, limited dynamic range and vignetting) [27]. A further, more practical constraint follows from this: integrating the sensor into the current setup required a magneto-inertial unit and the sensor's placement on the forehead, increasing the bulk of the device and affecting its aesthetics, which makes it less suited to prolonged use in real-world conditions. It is also worth noting that, although lighting was not controlled a priori, all acquisitions took place in indoor laboratories,

where the predominant light source was artificial: the lighting conditions tested, while uncontrolled, therefore remained confined to an indoor setting and do not capture the full range of illumination encountered in genuinely ecological contexts. Finally, the fact that the acquisitions for the two populations were carried out under conditions that were not fully homogeneous may represent a further source of confounding, given the presence of a luminance residual on pupil dilation.

Some limitations, lastly, relate to the model used. The assumption of an additive model, whereby variations in pupil diameter depend only on luminance and cognitive load, is a simplification that neglects other factors known to influence the pupil, such as stress, fatigue, and the near pupillary reflex. Moreover, the Watson–Yellott model was developed to estimate the pupillary light response under static conditions, whereas it is applied here under variable-light conditions.

7.5 Future work

Future developments could include the use of dynamic prediction models, better suited to contexts in which luminance varies continuously and able to account for other factors that affect the pupillary light response, such as latency and dependence on stimulus intensity. The experimental setup could also be further improved: integrating the ALS directly into the glasses frame, with a direct connection to the smart glasses, would represent a step toward a lighter device, better suited to use in real-world conditions.

Validation on larger and more homogeneous samples, including a genuine control group, will also be necessary to assess the ability of the cognitive component to discriminate between pathological populations and healthy controls. To this end, a broader comparison with other cognitive-workload assessment metrics could prove useful, within a set of activities characterised by a clear, well-defined increase in the required cognitive component.

Finally, it would be worthwhile to test the algorithm's performance under more ecological conditions, including movement across different environments and outdoor settings, so as to extend its application to situations closer to everyday life (for example, stairs, walking through crowded places, or more or less uneven terrain). Such an extension, however, would place the use of eye tracking under even more challenging conditions, in particular owing to the interference of the infrared component of sunlight with data quality – the TG3 manual indeed advises against outdoor use – and to its effect on the pupillary response, due to the frequent transitions between light and shadow and to the narrowing of the palpebral aperture under direct light, which makes pupil identification more difficult [51].

Chapter 8

Conclusions

This work set out to assess the feasibility of using pupillometry to quantify cognitive workload during cognitive-motor tasks, under uncontrolled ambient lighting conditions. The first relevant finding concerns precisely the separation of the components affecting pupil dilation under these testing conditions: the results obtained are encouraging regarding the possibility of isolating the cognitive component even in uncontrolled settings, using an open-source method, thereby overcoming one of the long-standing limitations of pupillometry, which has largely confined its application to highly controlled experimental conditions. The second finding concerns instead the sensitivity of the resulting index to the different demand levels of the tasks performed, which proved consistent with what was expected on the basis of the experimental design.

Overall, this preliminary application provides evidence that motivates further research into the use of pupillometry in ecological contexts, with possible implications as a clinical biomarker or as a tool for monitoring at-risk populations – a prospect made all the more concrete by the growing availability on the market of smart glasses equipped with eye-tracking capabilities.

Bibliography

[1] Mark S. Young and Neville A. Stanton. «Mental Workload». In: *Handbook of Human Factors and Ergonomics Methods*. Ed. by Neville A. Stanton, Alan Hedge, Karel Brookhuis, Eduardo Salas, and Hal W. Hendrick. Boca Raton, FL: CRC Press, 2005. Chap. 39, pp. 39-1–39-9 (cit. on p. 1).

[2] Da Tao, Haibo Tan, Hailiang Wang, Xu Zhang, Xingda Qu, and Tingru Zhang. «A Systematic Review of Physiological Measures of Mental Workload». In: *International Journal of Environmental Research and Public Health* 16.15 (2019), p. 2716. DOI: 10.3390/ijerph16152716 (cit. on p. 1).

[3] Stephen H. Fairclough and Kim Houston. «A Metabolic Measure of Mental Effort». In: *Biological Psychology* 66 (2004), pp. 177–190. DOI: 10.1016/j.biopsycho.2003.10.001 (cit. on p. 1).

[4] Giorgio Guidetti. «The Role of Cognitive Processes in Vestibular Disorders». In: *Hearing, Balance and Communication* 11.1 (2013), pp. 3–35. DOI: 10.3109/21695717.2013.765085 (cit. on pp. 1, 4).

[5] Melike Kahya, Kelly E. Lyons, Rajesh Pahwa, Abiodun E. Akinwuntan, Jianghua He, and Hannes Devos. «Pupillary Response to Postural Demand in Parkinson's Disease». In: *Frontiers in Bioengineering and Biotechnology* 9 (2021), p. 617028. DOI: 10.3389/fbioe.2021.617028 (cit. on pp. 1, 8, 11, 88).

[6] Melike Kahya, Kelly E. Lyons, Rajesh Pahwa, Abiodun E. Akinwuntan, Jianghua He, and Hannes Devos. «Reliability and Validity of Pupillary Response during Dual-Task Balance in Parkinson's Disease». In: *Archives of Physical Medicine and Rehabilitation* 102.3 (2021), pp. 448–455. DOI: 10.1016/j.apmr.2020.08.008 (cit. on p. 1).

[7] Sandra G. Hart and Lowell E. Staveland. «Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research». In: *Human Mental Workload*. Ed. by Peter A. Hancock and Najmedin Meshkati. Vol. 52. Advances in Psychology. North-Holland, 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9 (cit. on pp. 2, 44, 74).

[8] Sandra P. Marshall. «The Index of Cognitive Activity: measuring cognitive workload». In: *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*. 2002, pp. 7-5–7-9. DOI: 10.1109/HFPP.2002.1042860 (cit. on pp. 2, 8, 12, 52).

[9] Eckhard H. Hess and James M. Polt. «Pupil size in relation to mental activity during simple problem-solving». In: *Science* 143.3611 (1964), pp. 1190–1192. DOI: 10.1126/science.143.3611.1190 (cit. on pp. 2, 8).

[10] Daniel Kahneman and Jackson Beatty. «Pupil diameter and load on memory». In: *Science* 154.3756 (1966), pp. 1583–1585. DOI: 10.1126/science.154.3756.1583 (cit. on pp. 2, 8).

[11] W. S. Peavler. «Pupil size, information overload, and performance differences». In: *Psychophysiology* 11.5 (1974), pp. 559–566. DOI: 10.1111/j.1469-8986.1974.tb01114.x (cit. on pp. 2, 8).

[12] Gary K. Poock. «Information processing vs pupil diameter». In: *Perceptual and Motor Skills* 37.3 (1973), pp. 1000–1002. DOI: 10.2466/pms.1973.37.3.1000 (cit. on pp. 2, 8).

[13] Susan Ahern and Jackson Beatty. «Pupillary responses during information processing vary with Scholastic Aptitude Test scores». In: *Science* 205.4412 (1979), pp. 1289–1292. DOI: 10.1126/science.472746 (cit. on pp. 2, 8).

[14] Eric Granholm, Sarah K. Morris, Andrew J. Sarkin, Robert F. Asarnow, and Dilip V. Jeste. «Pupillary responses index overload of working memory resources in schizophrenia». In: *Journal of Abnormal Psychology* 106.3 (1997), pp. 458–467. DOI: 10.1037/0021-843x.106.3.458 (cit. on pp. 2, 8).

[15] Bhanuka Mahanama, Yasith Jayawardana, Sundararaman Rengarajan, Gavindya Jayawardena, Leanne Chukoskie, Joseph Snider, and Sampath Jayarathna. «Eye Movement and Pupil Measures: A Review». In: *Frontiers in Computer Science* 3 (2022), p. 733531. DOI: 10.3389/fcomp.2021.733531 (cit. on p. 2).

[16] Hannes Devos, Kathleen Gustafson, Pedram Ahmadnezhad, Ke Liao, Jonathan D. Mahnken, William M. Brooks, and Jeffrey M. Burns. «Psychometric Properties of NASA-TLX and Index of Cognitive Activity as Measures of Cognitive Workload in Older Adults». In: *Brain Sciences* 10.12 (2020), p. 994. DOI: 10.3390/brainsci10120994 (cit. on p. 2).

[17] Karen Z. H. Li, Louis Bherer, Anat Mirelman, Inbal Maidan, and Jeffrey M. Hausdorff. «Cognitive Involvement in Balance, Gait and Dual-Tasking in Aging: A Focused Review From a Neuroscience of Aging Perspective». In: *Frontiers in Neurology* 9 (2018), p. 913. DOI: 10.3389/fneur.2018.00913 (cit. on p. 3).

[18] Hui-Ting Goh, Miranda Pearce, and Asha Vas. «Task Matters: An Investigation on the Effect of Different Secondary Tasks on Dual-Task Gait in Older Adults». In: *BMC Geriatrics* 21 (2021), p. 510. DOI: 10.1186/s12877-021-02464-8 (cit. on p. 3).

[19] Fuzhong Li and Peter Harmer. «Prevalence of Falls, Physical Performance, and Dual-Task Cost While Walking in Older Adults at High Risk of Falling with and Without Cognitive Impairment». In: *Clinical Interventions in Aging* 15 (2020), pp. 945–952. DOI: 10.2147/CIA.S254764 (cit. on p. 3).

[20] G. Guidetti, D. Monzani, M. Trebbi, and V. Rovatti. «Impaired Navigation Skills in Patients with Psychological Distress and Chronic Peripheral Vestibular Hypofunction Without Vertigo». In: *Acta Otorhinolaryngologica Italica* 28.1 (2008), pp. 21–25 (cit. on p. 3).

[21] G. Guidetti, R. Guidetti, M. Manfredi, and M. Manfredi. «Vestibular Pathology and Spatial Working Memory». In: *Acta Otorhinolaryngologica Italica* 40 (2020), pp. 72–78. DOI: 10.14639/0392-100X-2189 (cit. on p. 3).

[22] L. J. Smith, D. Wilkinson, M. Bodani, and S. S. Surenthiran. «Cognition in Vestibular Disorders: State of the Field, Challenges, and Priorities for the Future». In: *Frontiers in Neurology* 15 (2024), p. 1159174. DOI: 10.3389/fneur.2024.1159174 (cit. on p. 3).

[23] D. A. Chari, A. Madhani, J. D. Sharon, and R. F. Lewis. «Evidence for Cognitive Impairment in Patients with Vestibular Disorders». In: *Journal of Neurology* 269 (2022), pp. 5831–5842. DOI: 10.1007/s00415-022-11289-3 (cit. on p. 3).

[24] M. Danneels, R. Van Hecke, L. Leyssens, S. Degeest, D. Cambier, R. van de Berg, V. Van Rompaey, and L. Maes. «2BALANCE: A Cognitive-Motor Dual-Task Protocol for Individuals with Vestibular Dysfunction». In: *BMJ Open* 10 (2020), e037138. DOI: 10.1136/bmjopen-2020-037138 (cit. on p. 3).

[25] Sabina Manz, Thomas Schmalz, Michael Ernst, Thomas Maximilian Köhler, Jose Gonzalez-Vargas, and Strahinja Dosen. «Using mobile eye tracking to measure cognitive load through gaze behavior during walking in lower limb prosthesis users: A preliminary assessment». In: *Clinical Biomechanics*

115 (2024), p. 106250. DOI: 10.1016/j.clinbiomech.2024.106250 (cit. on pp. 4, 11, 88).

[26] Mahdiyeh Rahmani Saeedpour-Parizi, Charles H. Shea, Sunee Bovonsunthonchai, and Hiroki Ihara. «Pupil diameter as a biomarker of effort in goal-directed gait». In: *Experimental Brain Research* 238 (2020), pp. 2615–2623. DOI: 10.1007/s00221-020-05915-7 (cit. on pp. 4, 10, 11, 88).

[27] Giovanni Pignoni, Sashidharan Komandur, and Frode Volden. «Accounting for Effects of Variation in Luminance in Pupillometry for Field Measurements of Cognitive Workload». In: *IEEE Sensors Journal* 21.5 (2021), pp. 6393–6400. DOI: 10.1109/JSEN.2020.3038291 (cit. on pp. 4, 13, 14, 24, 37, 51, 54, 85, 89, 91).

[28] Alberto Nascimbeni, Andrea Gaffuri, Arminio Penno, and Mara Tavoni. «Dual Task Interference during Gait in Patients with Unilateral Vestibular Disorders». In: *Journal of NeuroEngineering and Rehabilitation* 7 (2010), p. 47. DOI: 10.1186/1743-0003-7-47 (cit. on pp. 4, 91).

[29] Sebastiaan Mathôt. «Pupillometry: Psychology, Physiology, and Function». In: *Journal of Cognition* 1.1 (2018), pp. 1–23. DOI: 10.5334/joc.18 (cit. on p. 6).

[30] Stuart R. Steinhauer, Margaret M. Bradley, Greg J. Siegle, Kathryn A. Roecklein, and Annika Dix. «Publication guidelines and recommendations for pupillary measurement in psychophysiological studies». In: *Psychophysiology* 59.4 (2022), e14035. DOI: 10.1111/psyp.14035 (cit. on pp. 6, 8, 15, 24).

[31] Siddhartha Joshi, Yin Li, Rishi M. Kalwani, and Joshua I. Gold. «Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus, Colliculi, and Cingulate Cortex». In: *Neuron* 89.1 (2016), pp. 221–234. DOI: 10.1016/j.neuron.2015.11.028 (cit. on p. 6).

[32] Sylvain Sirois and Julie Brisson. «Pupillometry». In: *Wiley Interdisciplinary Reviews: Cognitive Science* 5.6 (2014), pp. 679–692. DOI: 10.1002/wcs.1323 (cit. on pp. 7, 8).

[33] Barry Winn, David Whitaker, David B. Elliott, and Nicholas J. Phillips. «Factors affecting light-adapted pupil size in normal human subjects». In: *Investigative Ophthalmology & Visual Science* 35.3 (1994), pp. 1132–1137 (cit. on p. 7).

[34] C. J. K. Ellis. «The pupillary light reflex in normal subjects». In: *British Journal of Ophthalmology* 65 (1981), pp. 754–759. DOI: 10.1136/bjo.65.11.754 (cit. on pp. 7, 25).

[35] Annekatrin Rickmann, Maria Waizel, Sara Kazerounian, Peter Szurman, Helmut Wilhelm, and Karl T. Boden. «Digital Pupillometry in Normal Subjects». In: *Neuro-Ophthalmology* 41.1 (2017), pp. 12–18. DOI: 10.1080/01658107.2016.1226345 (cit. on pp. 7, 35).

[36] Richard Feinberg and Edward Podolak. «Latency of Pupillary Reflex to Light Stimulation and Its Relationship to Aging». In: *Behavior, Aging and the Nervous System*. Ed. by A. T. Welford and James E. Birren. Springfield, IL: Charles C Thomas, 1965 (cit. on pp. 7, 86).

[37] Ken Asakawa, Akari Ito, Hinako Kobayashi, Aya Iwai, Chihiro Ito, and Hitoshi Ishikawa. «Adaptation time, electroretinography, and pupillography in healthy subjects». In: *Documenta Ophthalmologica* 139 (2019), pp. 33–44. DOI: 10.1007/s10633-019-09693-8 (cit. on p. 7).

[38] Hande Husniye Telek, Hidayet Erdol, and Adem Turk. «The Effects of Age on Pupil Diameter at Different Light Amplitudes». In: *Beyoglu Eye Journal* 3.2 (2018), pp. 80–85. DOI: 10.14744/bej.2018.43534 (cit. on p. 7).

[39] Andrew B. Watson and John I. Yellott. «A unified formula for light-adapted pupil size». In: *Journal of Vision* 12.10 (2012), pp. 1–16. DOI: 10.1167/12.10.12 (cit. on pp. 7, 13, 85).

[40] Lloyd L. Holladay. «The Fundamentals of Glare and Visibility». In: *Journal of the Optical Society of America* 12.4 (1926), pp. 271–319. DOI: 10.1364/JOSA.12.000271 (cit. on p. 7).

[41] Parry Moon and Domina Eberle Spencer. «On the Stiles-Crawford Effect». In: *Journal of the Optical Society of America* 34.6 (1944), pp. 319–329. DOI: 10.1364/JOSA.34.000319 (cit. on p. 7).

[42] P. A. Stanley and A. K. Davies. «The effect of field of view size on steady-state pupil diameter». In: *Ophthalmic and Physiological Optics* 15.6 (1995), pp. 601–603. DOI: 10.1046/j.1475-1313.1995.9400019v.x (cit. on p. 7).

[43] Vsevolod Peysakhovich, Mickaël Causse, Sébastien Scannella, and Frédéric Dehais. «Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort». In: *International Journal of Psychophysiology* 97 (2015), pp. 30–37. DOI: 10.1016/j.ijpsycho.2015.04.019 (cit. on pp. 8, 12).

[44] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, et al. «The Index of Pupillary Activity: Measuring Cognitive Load vis-à-vis Task Difficulty with Pupil Oscillation». In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, 2018. DOI: 10.1145/3173574.3173856 (cit. on pp. 8, 12).

[45] Giovanni Pignoni and Sashidharan Komandur. «Development of a Quantitative Evaluation Tool of Cognitive Workload in Field Studies Through Eye Tracking». In: *Engineering Psychology and Cognitive Ergonomics (HCII 2019)*. Ed. by Don Harris. Vol. 11571. Lecture Notes in Computer Science. Springer, 2019, pp. 106–122. DOI: 10.1007/978-3-030-22507-0_9 (cit. on pp. 8, 86).

[46] Beatrice Francavilla, Gianluca Velletrani, Carlo Chiaramonte, Stefano Di Girolamo, and Pier Giorgio Giacomini. «Assessing Cognitive Effort in Ménière's Disease: Pupillometry as a Novel Tool for Postural Control». In: *The Journal of the International Advanced Otology* 20.1 (2024), pp. 69–75. DOI: 10.5152/iao.2024.231150 (cit. on pp. 8, 11).

[47] Enkelejda Kasneci, Gjergji Kasneci, Tobias Appel, et al. *Introduction to Eye Tracking: A Hands-On Tutorial for Students and Practitioners*. 2024. arXiv: 2404.15435 (cit. on p. 9).

[48] Julie Brisson, Marc Mainville, Dominique Mailloux, Christelle Beaulieu, Josée Serres, and Sylvain Sirois. «Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers». In: *Behavior Research Methods* 45.4 (2013), pp. 1322–1331. DOI: 10.3758/s13428-013-0327-0 (cit. on pp. 9, 15, 91).

[49] Nicholas P. Murray, Melissa Hunfalvay, and Takumi Bolte. «The Reliability, Validity, and Normative Data of Interpupillary Distance and Pupil Diameter Using Eye-Tracking Technology». In: *Translational Vision Science & Technology* 6.4 (2017), p. 2. DOI: 10.1167/tvst.6.4.2 (cit. on pp. 9, 15).

[50] Diederick C. Niehorster, Thiago Santini, Roy S. Hessels, Ignace T. C. Hooge, Enkelejda Kasneci, and Marcus Nyström. «The impact of slippage on the data quality of head-worn eye trackers». In: *Behavior Research Methods* 52 (2020), pp. 1140–1160. DOI: 10.3758/s13428-019-01307-0 (cit. on pp. 10, 15, 91).

[51] Kamran Binaee, Christian Sinnott, Kaylie Jacleen Capurro, Paul MacNeilage, and Mark D. Lescroart. «Pupil Tracking Under Direct Sunlight». In: *ACM Symposium on Eye Tracking Research and Applications (ETRA '21 Adjunct)*. ACM, 2021. DOI: 10.1145/3450341.3458490 (cit. on pp. 10, 92).

[52] Bernhard Petersch and Kai Dierkes. «Gaze-angle dependency of pupil-size measurements in head-mounted eye tracking». In: *Behavior Research Methods* 54 (2022), pp. 763–779. DOI: 10.3758/s13428-021-01657-8 (cit. on pp. 10, 83).

[53] Joyce Tang, Ángel Ramos de Miguel, Juan Carlos Falcón González, Silvia Borkoski Barreiro, Isaura Rodriguez Montesdeoca, and Ángel Ramos Macías. «Using Pupillometry to Evaluate Balance in Patients Implanted with a Cochleo-Vestibular Implant». In: *Journal of Clinical Medicine* 13.13 (2024), p. 3797. DOI: 10.3390/jcm13133797 (cit. on pp. 11, 90, 91).

[54] Sebastiaan Mathôt and Ana Vilotijević. «Methods in Cognitive Pupillometry: Design, Preprocessing, and Statistical Analysis». In: *Behavior Research Methods* 55 (2023), pp. 3055–3077. DOI: 10.3758/s13428-022-01957-7 (cit. on p. 12).

[55] Marios Fanourakis and Guillaume Chanel. «Attenuation of the Dynamic Pupil Light Response During Screen Viewing for Arousal Assessment». In: *Frontiers in Virtual Reality* 3 (2022), p. 971613. DOI: 10.3389/frvir.2022.971613 (cit. on pp. 12, 13).

[56] Andrew T. Duchowski, Krzysztof Krejtz, Nina A. Gehrer, Tanya Bafna, and Per Bækgaard. «The Low/High Index of Pupillary Activity». In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, 2020. DOI: 10.1145/3313831.3376394 (cit. on pp. 12, 90).

[57] Julia N. Czerniak, Nikolas Schierhorst, Christopher Brandl, Alexander Mertens, Maximilian Schwalm, and Verena Nitsch. «A Meta-Analytic Review of the Reliability of the Index of Cognitive Activity Concerning Task-Evoked Cognitive Workload and Light Influences». In: *Acta Psychologica* 220 (2021), p. 103402. DOI: 10.1016/j.actpsy.2021.103402 (cit. on p. 12).

[58] Lisa Rerhaye, Talke Blaser, and Thomas Alexander. «Evaluation of the Index of Cognitive Activity (ICA) as an Instrument to Measure Cognitive Workload Under Differing Light Conditions». In: *Advances in Human Factors in Simulation and Modeling (IEA 2018)*. Vol. 827. Advances in Intelligent Systems and Computing. Springer, 2019, pp. 350–359. DOI: 10.1007/978-3-319-96059-3_38 (cit. on p. 12).

[59] Nikolas J. Schierhorst, Vera B. Rick, Julia N. Czerniak, Christopher Brandl, Alexander Mertens, and Verena Nitsch. «Comparison of Two Pupillometric Measures as Indicators of Cognitive Strain and Under the Influence of Screen Luminance». In: *Applied Ergonomics* 117 (2024), p. 104242. DOI: 10.1016/j.apergo.2024.104242 (cit. on p. 12).

[60] Vitor F. Pamplona, Manuel M. Oliveira, and Gladimir V. G. Baranoski. «Photorealistic Models for Pupil Light Reflex and Iridal Pattern Deformation». In: *ACM Transactions on Graphics* 28.4 (2009). DOI: 10.1145/1559755.1559763 (cit. on p. 13).

[61] Xiaofei Fan and Gang Yao. «Modeling transient pupillary light reflex induced by a short light flash». In: *IEEE Transactions on Biomedical Engineering* 58.1 (2010), pp. 36–42. DOI: 10.1109/TBME.2010.2080678 (cit. on p. 13).

[62] Brendan John, Pallavi Raiturkar, Arunava Banerjee, and Eakta Jain. «An evaluation of pupillary light response models for 2D screens and VR HMDs». In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18)*. ACM, 2018. ISBN: 9781450360869. DOI: 10.1145/3281505.3281538 (cit. on pp. 13, 52).

[63] Marie Eckert, Emanuël A. P. Habets, and Olli S. Rummukainen. «Cognitive Load Estimation Based on Pupillometry in Virtual Reality with Uncontrolled Scene Lighting». In: *2021 Thirteenth International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2021. DOI: 10.1109/QoMEX51781.2021.9465417 (cit. on p. 13).

[64] Benedikt V Ehinger, Katharina Groß, Inga Ibs, and Peter König. «A new comprehensive eye-tracking test battery concurrently evaluating the Pupil Labs glasses and the EyeLink 1000». In: *PeerJ* 7 (2019), e7086 (cit. on p. 15).

[65] Vishal Onkhar, Dimitra Dodou, and Joost CF de Winter. «Evaluating the Tobii Pro Glasses 2 and 3 in static and dynamic conditions». In: *Behavior Research Methods* 56.5 (2024), pp. 4221–4238 (cit. on p. 15).

[66] Ignace TC Hooge, Diederick C Niehorster, Roy S Hessels, Jeroen S Benjamins, and Marcus Nyström. «How robust are wearable eye trackers to slow and fast head and body movements?» In: *Behavior Research Methods* 55.8 (2023), pp. 4128–4142 (cit. on p. 15).

[67] Diederick C Niehorster, Roy S Hessels, Jeroen S Benjamins, Marcus Nyström, and Ignace TC Hooge. «GlassesValidator: A data quality tool for eye tracking glasses». In: *Behavior research methods* 56.3 (2024), pp. 1476–1484. DOI: 10.3758/s13428-023-02105-5 (cit. on p. 15).

[68] Mohammadhossein Salari, Diederick C Niehorster, Marcus Nyström, and Roman Bednarik. «The effect of pupil size on data quality in head-mounted eye trackers». In: *Behavior Research Methods* 58.1 (2025), p. 17. DOI: 10.3758/s13428-025-02880-3 (cit. on p. 15).

[69] Tobii AB. *Tobii Pro Glasses 3 Product Description*. v 1.14, June 2024. Tobii AB. Danderyd, Sweden, 2024 (cit. on p. 16).

[70] Natus Medical Incorporated. *ICS Dizcovery – Sfida le aspettative. Cambia i risultati.* Product brochure, 047337-IT RevA. Natus Medical Incorporated. Middleton, WI, USA, 2024 (cit. on p. 17).

[71] Tobii AB. *Tobii Pro Glasses 3 User Manual*. v 1.20. Tobii AB. Danderyd, Sweden, 2024 (cit. on p. 19).

[72] Natus Medical Incorporated. *ICS Dizcovery User Guide*. Rev. 02. Natus Medical Incorporated. Middleton, WI, USA, 2024 (cit. on p. 19).

[73] Scott A Read, Michael J Collins, Leo G Carney, and D Robert Iskander. «The morphology of the palpebral fissure in different directions of vertical gaze». In: *Optometry and vision science* 83.10 (2006), pp. 715–722. DOI: 10.1097/01.opx.0000236811.78177.97 (cit. on pp. 20, 82).

[74] Nikola Knezi, Vladimir Isaretovic, Ivan Adjic, Nikola Babic, Dusica Maric, and Ljubica Stojsic Dzunja. «Morphometric Analysis of the Palpebral Fissure and Canthal Distance in Serbian Young Adults». In: *International Journal of Morphology* 38.5 (2020), pp. 1381–1385. DOI: 10.4067/S0717-95022020000501381 (cit. on pp. 20, 82).

[75] Shashank Kishore Bhandary, Rohit Dhakal, Vishwa Sanghavi, and Pavan Kumar Verkicharla. «Ambient light level varies with different locations and environmental conditions: Potential to impact myopia». In: *PLOS ONE* 16.7 (2021), e0254027. DOI: 10.1371/journal.pone.0254027 (cit. on p. 24).

[76] Ritesh K Chaurasiya, Samir Sutar, Akansha Gupta, Rishabh Chaudhary, Rishiraj Saini, Pradeep Agarwal, Ashi Khurana, and Lokesh Chauhan. «Variation of ambient illumination across different locations: an impact on Antimyopia strategy». In: *Middle East African Journal of Ophthalmology* 30.4 (2024), p. 250 (cit. on p. 24).

[77] Vishay Semiconductors. *VEML7700 - High Accuracy Ambient Light Sensor with I2C Interface*. Tech. rep. 84286. Rev. 1.0, 19-Nov-2015. Vishay Intertechnology, Inc., 2015 (cit. on pp. 25, 27, 31, 37).

[78] ROHM Semiconductor. *BH1750FVI - Digital 16bit Serial Output Type Ambient Light Sensor IC*. Tech. rep. 11046EDT01. Rev. D. ROHM Co., Ltd., 2011 (cit. on p. 25).

[79] ams AG. *TSL2591 - Light-to-Digital Converter*. Tech. rep. v2-04, 2018-Jun-05. ams AG, 2018 (cit. on p. 25).

[80] Vishay Semiconductors. *VEML6030 - High Accuracy Ambient Light Sensor With I2C Interface*. Tech. rep. 84366. Rev. 1.3, 12-May-2017. Vishay Intertechnology, Inc., 2017 (cit. on p. 25).

[81] ON Semiconductor. *NOA1305 - Ambient Light Sensor with I2C Interface and Dark Current Compensation*. Tech. rep. NOA1305/D. Rev. 0, October 2012. Semiconductor Components Industries, LLC, 2012 (cit. on p. 25).

[82] LITE-ON Optoelectronics. *Optical Sensor Product Data Sheet: LTR-329ALS-01*. Tech. rep. DS86-2014-0006. Revision C, effective 03/14/2020. LITE-ON Technology Corp., 2020 (cit. on p. 25).

[83] ams OSRAM Group. *TSL2585 - Miniature Ambient Light Sensor with UV and Light Flicker Detection*. Tech. rep. v4-00, 2022-Dec-22. ams-OSRAM AG, 2022 (cit. on p. 25).

[84] Francesca Salis, Stefano Bertuletti, Kirsty Scott, Marco Caruso, Tecla Bonci, Ellen Buckley, Ugo Della Croce, Claudia Mazzà, and Andrea Cereatti. «A wearable multi-sensor system for real world gait analysis». In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico*. 2021, pp. 7020–7023. DOI: 10.1109/EMBC46164.2021.9630392 (cit. on p. 25).

[85] Stefano Bertuletti, Andrea Cereatti, Daniele Comotti, Michele Caldara, and Ugo Della Croce. «Static and Dynamic Accuracy of an Innovative Miniaturized Wearable Platform for Short Range Distance Measurements for Human Movement Applications». In: *Sensors* 17.7 (2017), p. 1492. DOI: 10.3390/s17071492 (cit. on p. 26).

[86] Reinhard Schaar. *Designing the VEML7700 Into an Application*. Application Note 84323. Rev. 06-Mar-2025. Vishay Intertechnology, Inc., 2025 (cit. on pp. 27, 29, 31–33).

[87] Delta OHM. *HD2302.0 – Fotoradiometro*. Tech. rep. Technical datasheet, Rev. 1.1, 05/2021. Delta OHM S.r.l., 2021 (cit. on pp. 27, 29).

[88] Mustafa Coramik. «Calibration of an LDR-Lux meter using a smartphone». In: *Physics Education* 56.3 (2021), p. 035009. DOI: 10.1088/1361-6552/abe1d5 (cit. on p. 28).

[89] Adafruit Industries. *Adafruit VEML7700*. Learn guide, "Overview" page. 2021. URL: https://learn.adafruit.com/adafruit-veml7700/overview (cit. on p. 29).

[90] Shiva Gholami, Nicolaas J. Reus, and Thomas J. T. P. van den Berg. «The significance of changes in pupil size during straylight measurement and with varying environmental illuminance». In: *Journal of Optometry* 11 (2018), pp. 167–173. DOI: 10.1016/j.optom.2017.08.004 (cit. on p. 35).

[91] Farah Maqsood. «Effects of varying light conditions and refractive error on pupil size». In: *Cogent Medicine* 4.1 (2017), p. 1338824. DOI: 10.1080/2331205X.2017.1338824 (cit. on p. 35).

[92] Hanyang Yu, Wentao Li, Ziping Chen, Mengzhen Chen, Junwen Zeng, Xijiang Lin, and Feng Zhao. «Is Ocular Accommodation Influenced by Dynamic Ambient Illumination and Pupil Size?» In: *International Journal of Environmental Research and Public Health* 19.17 (2022), p. 10490. DOI: 10.3390/ijerph191710490 (cit. on p. 35).

[93] Michael D. Twa, Melissa D. Bailey, John Hayes, and Mark Bullimore. «Estimation of pupil size by digital photography». In: *Journal of Cataract and Refractive Surgery* 30 (2004), pp. 381–389. DOI: 10.1016/S0886-3350(03)00619-9 (cit. on p. 35).

[94] Ziad S. Nasreddine, Natalie A. Phillips, Valérie Bédirian, Simon Charbonneau, Victor Whitehead, Isabelle Collin, Jeffrey L. Cummings, and Howard Chertkow. «The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool for Mild Cognitive Impairment». In: *Journal of the American Geriatrics Society* 53.4 (2005), pp. 695–699. DOI: 10.1111/j.1532-5415.2005.53221.x (cit. on p. 47).

[95] Gabriella Santangelo et al. «Normative data for the Montreal Cognitive Assessment in an Italian population sample». In: *Neurological Sciences* 36.4 (2015), pp. 585–591. DOI: 10.1007/s10072-014-1995-y (cit. on p. 47).

[96] W3C. *Web Content Accessibility Guidelines (WCAG) 2.1*. W3C Recommendation; relative luminance definition. World Wide Web Consortium. 2018. URL: https://www.w3.org/TR/WCAG21/ (cit. on p. 55).

[97] Fred G. W. C. Paas. «Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach». In: *Journal of Educational Psychology* 84.4 (1992), pp. 429–434. DOI: 10.1037/0022-0663.84.4.429 (cit. on p. 88).

[98] Yuzhang Li, Bin Zheng, Peter Grant, and Xinming Li. «Assessing Cognitive Load through Eye Metrics in-Motion vs. Stationary Environments». In: *Applied Ergonomics* 129 (2025), p. 104598. DOI: 10.1016/j.apergo.2025.104598 (cit. on p. 89).

[99] Nicola Camp, Roberto Vagnetti, Maria Bisele, Paul Felton, Kirsty Hunter, and Daniele Magistro. «The Effect of Cognitive Task Complexity on Healthy Gait in the Walking Corsi Test». In: *Brain Sciences* 13.7 (2023), p. 1019. DOI: 10.3390/brainsci13071019 (cit. on p. 90).

[100] Tepring Piquado, Derek Isaacowitz, and Arthur Wingfield. «Pupillometry as a Measure of Cognitive Effort in Younger and Older Adults». In: *Psychophysiology* 47.3 (2010), pp. 560–569. DOI: 10.1111/j.1469-8986.2009.00947.x (cit. on p. 90).

[101] Brooke N. Klatt, Wendy J. Carender, Cathy C. Lin, Saud F. Alsubaie, Catherine R. Kinnaird, Kathleen H. Sienko, and Susan L. Whitney. «A Conceptual Framework for the Progression of Balance Exercises in Persons with Balance and Vestibular Disorders». In: *Physical Medicine and Rehabilitation International* 2.4 (2015), p. 1044 (cit. on p. 91).

[102] Gregory F. Marchetti, Susan L. Whitney, Philip J. Blatt, Laura O. Morris, and Joan M. Vance. «Temporal and Spatial Characteristics of Gait during Performance of the Dynamic Gait Index in People with and People without Balance or Vestibular Disorders». In: *Physical Therapy* 88.5 (2008), pp. 640–651. DOI: 10.2522/ptj.20070130 (cit. on p. 91).