

VEOS: Vision-based Electrooculogram Signal Inference for Improved EEG in Brain-Computer Interfaces

Peter Redmond

BSc, BEng(Hons), MEng

Supervised by Professor Tomás Ward



A thesis presented for the degree of Doctor of Philosophy

SCHOOL OF COMPUTING
DUBLIN CITY UNIVERSITY

December 2025

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original and have conformed to the regulations on the use and declaration of Generative AI, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed:

Peter Redmond.

ID No.: 21261231

Date: 5th of December 2025

Generative AI usage declaration

Introduction

This declaration outlines the use of generative AI tools during the preparation of this thesis, including the tools used, the outputs, example prompts, and a brief reflection on their role.

Tools

I used Gen AI tools in a supporting and editorial capacity between 20 November 2025 and 5 December 2025. I used ChatGPT 5.1 and Google Gemini 3 Deep Research for structural and readability feedback, Grammarly for proofreading, and a Gemini image-generation tool (Nano Banana) to generate schematic figures.

Outputs

I used these tools to generate feedback of draft materials that I then reviewed critically and applied selectively. ChatGPT 5.1 and Google Gemini 3 Deep Research produced comments on thesis structure, chapter and section ordering, coherence, clarity, repetition and transitions. I used this feedback to guide my own revisions, and all rewriting, reorganisation and argument changes were carried out by me.

For proofreading, ChatGPT and Grammarly suggested edits to spelling, grammar, punctuation and phrasing. I accepted only changes I judged to be accurate and appropriate.

For figures, Nano Banana was used to generate schematic diagrams.

Prompts

For thesis review I used prompts such as: "Review the first draft of a PhD thesis in terms of structure and content and give me a report" and "Review my thesis as if you were my PhD supervisor". For cross-referencing I used: "Verify all figures and tables are referenced from the body text". For proofreading I used: "Proofread my thesis before I submit". For figure generation I used prompts requesting labelled diagrams of the eye in frontal and sectional views, and a simplified model of the EOG and its propagation to scalp EEG via volume conduction.

Iterations

My use of these tools was iterative but limited. I typically requested an initial round of feedback on a chapter or the full draft, revised manually, and sometimes requested a second round focused on clarity and language. For figure generation, I used a small number of iterations to improve labels and layout. At all stages I curated the outputs, and the tools did not determine the scientific content, methodology, data analysis or conclusions of the thesis.

Critical Reflection

Overall, Gen AI tools were used to support editing rather than to contribute substantive research content. They helped identify errors and unclear phrasing, and reduced the risk of missing figure and table cross-references. The generated schematics provided clear visual representations of key structures and concepts that would otherwise have been difficult for me to produce. I remained responsible for all academic judgement, and for the final content and presentation of the thesis.

Dedication

To Sandy, Seán, Erin, Tom, Clodagh, Áine and Mam.

Acknowledgements

I would first like to express my sincere gratitude to my supervisor, Professor Tomás Ward, whose guidance, insight and patience have been invaluable throughout this project. I am also deeply grateful to my former PhD supervisor, Professor Gerry Lacey, for his support in the early stages of my doctoral studies and for encouraging me to pursue this line of research. I am also thankful to Dr Graham Healy for his support and advice and Andrew Fleury of Transpoco for supporting this research.

I have been fortunate to work alongside brilliant fellow doctoral researchers, including Sandra, Ahsan and Sonal. Their feedback, conversations and good humour have shaped this thesis in countless ways and made the last few years far more enjoyable than they might otherwise have been.

I am equally grateful to my mam, Pauline, and my sister, Nicolette, who have offered unwavering encouragement and practical support throughout my time as a student.

Finally, I owe my deepest thanks to my wife, Sandy, and our five children, Seán, Erin, Tom, Clodagh and Áine, for their patience, kindness and understanding about the many hours I have spent at the computer. Their love and belief in me, especially during the most challenging stages of this work, have meant more than I can say. This thesis is dedicated to them.

This work was supported by Transpoco Ltd. and by Science Foundation Ireland (Grant Nos. SFI/12/RC/2289 P2 and 18/SP/5942).

Contents

1	Introduction	3
1.1	Motivation	3
1.2	Problem statement	6
1.3	Research questions	7
1.4	Scope and assumptions	9
1.5	Contributions	10
1.6	Thesis structure	13
2	Background and Literature Review	15
2.1	Introduction	15
2.2	Brain–Computer Interfaces and Neuroergonomics	16
2.2.1	Historical development and core concepts	16
2.2.2	Application domains relevant to VEOS	17
2.2.3	From laboratory to life: ecological validity	18
2.3	EEG Acquisition, Signal Quality and Artefacts	18
2.3.1	EEG acquisition and practical constraints	19
2.3.2	Physiological and environmental artefacts	19
2.3.3	Impact of ocular artefacts on BCI control signals	20
2.4	Ocular Activity: EOG, Eye Movements and Eyelid EMG	21
2.4.1	Corneo-retinal dipole and electrooculogram	21
2.4.2	Eye movement dynamics	22
2.4.3	Eyelid EMG and periocular muscle activity	23
2.4.4	EOG as signal and as artefact	24
2.5	EEG Artefact Handling: From Regression to Deep Learning	25
2.5.1	Linear regression and linear component analysis (LCA) with recorded EOG	25
2.5.2	Independent component analysis and blind source separation . .	27
2.5.3	Spatial filtering and subspace methods	28
2.5.4	Artefact handling in mobile and movement-rich settings	28
2.5.5	Deep learning and generative approaches to artefact removal . .	29
2.6	Eye-Tracking, Gaze Estimation and Computer Vision	29
2.6.1	Traditional eye-tracking technologies	29
2.6.2	Appearance-based gaze estimation and deep learning	30
2.6.3	Eye movement analysis for activity recognition and state esti- mation	31
2.6.4	Computer vision frameworks for eye and face analysis	32
2.7	Multimodal EEG–Eye Datasets and Experimental Frameworks	32
2.7.1	Multimodal datasets in BCI research	32

2.7.2	AMBER: multimodal BCIs in naturalistic settings	33
2.7.3	A multimodal data acquisition platform for blended control . . .	34
2.7.4	Eye-BCI multimodal dataset combining EEG, eye-tracking and high-speed video	34
2.7.5	Other multimodal EEG-eye datasets	35
2.8	Positioning VEOS within the Literature	36
2.8.1	Sensor-based versus vision-based ocular measurement	36
2.8.2	Video-based generation of pseudo-EOG signals	37
2.8.3	Relation of VEOS to deep learning-based artefact removal . . .	38
2.8.4	Research gaps addressed by VEOS	39
2.9	Summary	40
3	Theoretical Foundations	41
3.1	Overview	41
3.2	Camera geometry for eye-tracking	42
3.3	Modelling eyeball kinematics	44
3.4	Temporal dynamics of eye movements	47
3.5	Landmark noise and uncertainty	48
3.6	EEG signal characteristics	50
3.7	Electrooculogram modelling	52
3.8	Propagation of ocular sources to scalp EEG	53
3.9	Ocular artefacts in EEG	55
3.10	Eyelid EMG and periocular muscle activity	57
3.11	Independent component analysis for EEG	59
3.12	Mapping from video landmarks to EOG	61
3.13	Synchronisation and timing considerations	63
3.14	Summary	65
4	Multimodal Data Acquisition Platform (Development System)	67
4.1	Overview	67
4.2	Hardware architecture	68
4.2.1	EEG and EOG acquisition	69
4.2.2	High-frame-rate video subsystem	69
4.2.3	Stimulus display and triggering	70
4.3	Software architecture	71
4.3.1	Lab Streaming Layer integration	72
4.3.2	Custom recording interface	74
4.3.3	OpenViBE and task logic	75
4.4	Calibration and stimulus design	75
4.4.1	Follow the dots: gaze calibration	75
4.4.2	Game-based engagement task	76
4.4.3	ERP-style visual paradigms	77
4.5	Data acquisition and synchronisation in practice	78
4.5.1	Timestamping and clock management	78
4.5.2	End-to-end timing validation	79
4.6	Development dataset and file structure	80
4.6.1	Recorded modalities and channels	81
4.6.2	Directory layout and reproducibility	81
4.7	Relation to previous multimodal platforms	82

4.8	Open science, code and data availability	83
4.9	Summary	85
5	VEOS: Vision-based Electrooculogram Signal Inference	87
5.1	Chapter overview	87
5.2	Problem formulation	88
5.3	Overview of the VEOS architecture	90
5.4	Video processing and landmark tracking	92
5.4.1	Frame acquisition and timing	92
5.4.2	Facial landmarks with MediaPipe Face Mesh	92
5.4.3	Handling head motion and tracking failures	93
5.5	Feature engineering	94
5.5.1	Geometric features	94
5.5.2	Temporal features	95
5.5.3	Blink and saccade descriptors	96
5.6	EOG and EMG target construction	97
5.6.1	EOG preprocessing in the VEOS development dataset	97
5.6.2	EOG and EMG in the Eye-BCI dataset	98
5.6.3	Target scaling and channel selection	99
5.7	Regression model and training procedure	99
5.7.1	Design matrix construction	100
5.7.2	Linear regression with regularisation	100
5.7.3	Extensions to non-linear models	101
5.7.4	Loss functions and blink-aware weighting	102
5.7.5	Training and validation strategy	103
5.8	Integration with EEG artefact removal	103
5.9	Summary	105
6	Experimental Protocols and Datasets	107
6.1	Eye-BCI multimodal dataset (Synapse)	107
6.1.1	Overview	108
6.1.2	Participants and sessions	108
6.1.3	BCI paradigms	109
6.1.4	Recorded modalities and hardware	111
6.1.5	Synchronisation and labelling	112
6.1.6	Data structure and availability	113
6.1.7	Relevance for VEOS	113
6.2	VEOS development dataset	114
6.2.1	Participants and ethics	114
6.2.2	Session structure and tasks	115
6.2.3	Recorded modalities and synchronisation	118
6.2.4	Iterative refinement of the platform and methods	119
6.2.5	Summary	121
6.3	Experimental design and evaluation protocol	121
6.3.1	Participant-wise data splits	122
6.3.2	Targets and baselines	122
6.3.3	Evaluation metrics	123
6.4	Summary	126

7	Results	127
7.1	Overview	127
7.2	Monocular Eye-Tracking Performance	128
7.2.1	Datasets, protocol and ground truth	129
7.2.2	Landmark localisation	130
7.2.3	Eyeball position and eyelid aspect ratio (EAR)	131
7.2.4	Blink detection from video and comparison with Tobii	132
7.3	EOG Inference from Ocular Signals	134
7.3.1	Models and metrics	134
7.3.2	Signal-level performance on the VEOS development dataset	135
7.3.3	Cross-dataset and cross-modality performance on Eye-BCI	136
7.3.4	Spectral and event-level metrics	138
7.3.5	Qualitative examples	138
7.4	Linear Component Analysis for Ocular Artefact Removal	140
7.4.1	Canonical LCA configuration with true EOG	140
7.4.2	LCA with inferred EOG from video and Tobii	141
7.4.3	Impact on task-related EEG features	143
7.5	BCI-Level Performance	145
7.5.1	Motor imagery classification	145
7.5.2	SSVEP	146
7.5.3	P300 spelling	147
7.6	Summary	149
8	Discussion	151
8.1	Overview	151
8.2	Revisiting the research questions	152
8.2.1	RQ1: Monocular eye-tracking performance	152
8.2.2	RQs 2 and 3: Mapping feasibility and model adequacy	153
8.2.3	RQ4: Dataset dependence and generalisation	154
8.2.4	RQ5: Impact on EEG cleaning	155
8.2.5	RQ6: BCI-level consequences	156
8.3	Implications for vision-based ocular measurement	158
8.3.1	Monocular eye-tracking quality and limits	158
8.3.2	Role of eyelid EMG and non-linear blink contributions	160
8.3.3	Comparison with sensor-based approaches	160
8.4	Limitations	161
8.4.1	Data, participants and paradigms	161
8.4.2	Hardware and imaging constraints	163
8.4.3	Modelling assumptions and simplifications	164
8.5	Future work	165
8.5.1	Improving eyelid-tracking and blink modelling	165
8.5.2	Richer temporal and multimodal models	166
8.5.3	Towards online and real-world deployment	167
8.6	Practical design guidelines	167

9	Conclusion	171
9.1	Overview	171
9.2	Summary of contributions and findings	172
9.2.1	Theoretical and conceptual contributions	172
9.2.2	Multimodal platform and VEOS algorithm	172
9.2.3	Empirical evaluation	173
9.3	Answers to the research questions	174
9.4	Implications for BCIs and neuroergonomics	175
9.5	Limitations	176
9.6	Future directions	176
9.7	Closing remarks	177
A	Implementation	179
A.1	Implementation details for VEOS	179
A.1.1	Monocular eye-tracking pipeline	179
A.1.2	Feature construction	180
A.2	EOG inference models	181
A.2.1	Linear baselines	181
A.2.2	Windowed multilayer perceptrons	182
A.2.3	Temporal convolutional networks	182
A.2.4	Hybrid forward models	183
A.3	Training and evaluation configuration	183
A.3.1	Target preprocessing	183
A.3.2	Cross-validation and data splits	184
A.3.3	Optimisation details	184
A.4	Multimodal platform implementation details	185
A.4.1	LSL stream configuration	185
A.4.2	Directory structure and file naming	185
A.4.3	Software environment	186
B	Ethics	187
B.1	Ethics approval	187
B.2	Inclusion and exclusion criteria	188
B.3	Recruitment and compensation	189
B.4	Information provided to participants	190
B.5	Consent procedure	191
B.6	Data protection and anonymisation	192
B.7	Adverse events and participant feedback	194

List of Figures

- 1.1 Multimodal recordings combine EEG with eye and face video and, in some datasets, additional ocular channels such as EOG, eyelid EMG and eye-tracking. VEOS uses eye-region video, and optionally other ocular measurements, to infer an EOG-like signal. This inferred signal, or real EOG when available, is then used as a reference in regression- and ICA-based ocular artefact removal, producing cleaned EEG. Standard BCI pipelines for motor imagery, SSVEP and P300 paradigms operate on the cleaned signals to produce classifier decisions and control outputs. 5

- 2.1 Cross-sectional anatomy of the eye highlighting structures relevant to EOG generation. The corneo-retinal dipole is indicated by the positive pole at the cornea and the negative pole at the retina. The diagram also labels the upper and lower eyelids, the orbicularis oculi and levator palpebrae superioris muscles, the extraocular muscles that rotate the eyeball, and the retina, all of which contribute to the ocular potentials and periocular EMG that appear in EOG and EEG recordings. 22

- 2.2 Example EEG and ocular electrode montage. The participant wears a 32-channel EEG cap together with horizontal and vertical EOG electrodes, periocular EMG electrodes and their associated ground/reference electrodes around the eyes and face. This configuration provides high-quality ocular signals but increases setup time and can be uncomfortable or visually intrusive for users, especially outside controlled laboratory settings. 26

- 3.1 A head-centred coordinate frame is mapped to the camera frame by a rigid transform $[R \mid t]$, while the eyeball rotates within the head frame. Landmarks on the eyelids and iris are projected onto the camera image plane, and a face-normalisation stage applies a similarity transform to the detected landmarks to reduce the effect of head motion before they are used as VEOS features. 45

- 3.2 Simplified model of the electrooculogram and its propagation to scalp EEG. The corneo-retinal potential is modelled as an eye-centred dipole, which gives rise to large voltages at nearby EOG electrodes and smaller, but related, artefacts at frontal EEG electrodes via volume conduction. 54

3.3	Characteristic time-domain waveforms of the three main classes of ocular artefact as they appear in a frontal EEG or vertical EOG channel. A blink (left) produces a large, sharp-rising deflection that plateaus briefly at full eyelid closure before returning more slowly to baseline, with peak amplitudes commonly in the range 100–500 μ V. A saccade (centre) produces a shorter, step-like deflection whose polarity depends on the direction of the eye movement and whose amplitude scales with the angular displacement. A slow drift (right) produces a gradual baseline shift that accumulates over seconds until gaze is reset. These waveform shapes motivate the use of temporal context in the VEOS inference model.	56
4.1	Block diagram of the multimodal acquisition platform used for the VEOS development dataset. EEG, EOG and (optionally) eyelid EMG are recorded by dedicated front ends, while a high-frame-rate camera captures face video. All devices stream data to a single host/stimulus PC via Lab Streaming Layer (LSL), which aggregates EEG, EOG/EMG, video and marker streams into a single multimodal recording. An ESP32 microcontroller distributes TTL triggers to the amplifiers and camera, drives a timing LED visible in the video, and sends matching event markers to the host PC, providing a common hardware-synchronised time reference across all modalities and with the visual tasks presented on the same machine.	68
4.2	Photograph of the VEOS development recording setup. (A) Near-infrared floodlight providing stable illumination of the participant’s face and eye region. (B) 120 fps monochrome global-shutter camera with hardware external trigger input, modified to remove the infrared cut filter. (C) Stimulus laptop displaying the cognitive loading game used as the game-based engagement task. (D) OpenBCI Cyton biosignal amplifier in a custom-printed enclosure, used to acquire EOG channels. (E) ESP32 microcontroller providing hardware synchronisation: it distributes TTL trigger pulses to the camera, Cyton, ANT Neuro amplifier and laptop, and drives a timing LED visible in the camera frame. (F) 32-channel EEG cap. (G) OpenBCI USB receiver dongle. (H) ANT Neuro eego rt EEG amplifier. (I) Voltage-level shifter and signal inverter providing logic-level compatibility between the ESP32 and the camera and Cyton trigger inputs. (J) EOG electrodes connected to the Cyton for horizontal and vertical EOG acquisition.	72
4.3	Physical arrangement of the VEOS development recording setup. A seated participant wearing a 32-channel EEG cap and periocular EOG electrodes views a monitor at a viewing distance of approximately 60–70 cm, with a 120 Hz face camera mounted on the top bezel to capture the full face. A small LED placed in a screen corner provides visible timing flashes within the camera field of view. Approximate eye–screen distance, screen size and camera position are indicated to clarify the recording geometry.	73

4.4	Software-level data flow through Lab Streaming Layer (LSL). Physiological devices expose LSL outlets for EEG, EOG/EMG and marker streams, while the camera acquisition software writes video frames to disk and publishes only frame indices and timestamps via an LSL outlet. A central recording application opens corresponding inlets, subscribes to the EEG, EOG/EMG, frame-timing and marker streams, and writes a single multimodal recording file that preserves channel metadata and a common time base. The ESP32 microcontroller drives the LEDs and generates all digital trigger events, which are inserted into the marker outlet to provide shared timing markers across electrical and visual channels.	74
4.5	Task simulator game interface	78
4.6	Example multimodal timeline used for timing validation. Panel (a) shows the ESP32 LED / TTL control signal together with the corresponding LSL trigger markers. Panel (b) plots the dedicated EEG and EOG trigger channels recorded by the amplifiers; these digital channels receive the same pulses but do not affect the underlying EEG or EOG voltage signals. Panel (c) indicates the video frame indices, with the frames in which the LED is visible in the camera image highlighted. Aligning these three panels allows relative delays and jitter between the hardware triggers, LSL marker stream and camera to be estimated, while keeping the physiological channels free from LED-related artefacts.	80
4.7	Directory structure of the VEOS development dataset. Recordings are organised by participant (subj01, subj02, . . .), with one or more session folders under each participant. Each session contains a multimodal LSL recording file, a corresponding raw video file, and a JSON configuration file describing the montage and acquisition settings. Log files from the stimulus software are written to the session directory when present, but are not generated for all session types. This structure supports reproducible analysis and straightforward cross-referencing between modalities.	82
5.1	Overview of the VEOS processing chain. Monocular eye-region video is converted into eye and eyelid features, which are mapped by a regression model to an inferred EOG signal. This signal is then used alongside EEG in regression-based linear component analysis (LCA) and independent component analysis (ICA) pipelines to identify and remove ocular components, yielding cleaned EEG for downstream BCI analyses.	91
5.2	Illustrative output of the MediaPipe Face Mesh stage for one example frame. The face region of interest and triangulated mesh are overlaid in white. Eye and eyebrow contours derived from selected landmarks are highlighted for the left eye and eyebrow (red) and the right eye and eyebrow (green); these form the basis of the geometric eye and eyelid features used by VEOS.	94

5.3	Illustrative time series over a short interval containing blinks and saccades. Top: horizontal and vertical pupil position and eyelid aperture derived from landmarks. Middle: first-order temporal derivatives (velocities). Bottom: simultaneously recorded horizontal and vertical EOG from the VEOS development dataset (VEOS-Dev). The correspondence between changes in the visual features and deflections in the ocular signals motivates the VEOS regression approach.	97
5.4	Integration of VEOS with EEG artefact removal. (a) In a regression-based linear component analysis (LCA) framework, this signal enters the regression as an ocular reference, and its contribution is subtracted from each EEG channel to yield cleaned EEG. (b) VEOS infers an EOG-like signal from video. (c) In an independent component analysis (ICA) framework, the same signal is used as an auxiliary reference to identify ocular components by correlation, which are then removed before reconstructing cleaned EEG. Downstream BCI analyses (for example, event-related potentials or spectral features) are performed on the cleaned signals.	105
6.1	Schematic session timeline for the VEOS development dataset. Each session begins with setup and impedance checks, followed by one or more follow-the-dots calibration blocks, three game-based engagement blocks at increasing difficulty levels and ERP-style visual blocks. The upper bands indicate the order and duration of these segments, while the lower bands show that 32-channel EEG, horizontal and vertical EOG, high-frame-rate face video and an event marker stream are recorded during each session except calibration. This contrasts with the Eye-BCI dataset, which uses a different hardware stack and paradigm structure but shares the same principle of continuous multimodal recording on a synchronised time base.	117
7.1	Example monocular eye landmark detection on Eye-BCI. Raw 150 Hz IR eye-region frame with detected canthi (green), pupil boundary (blue) and eyelid splines (orange).	133
7.2	Example monocular eye-tracking outputs on Eye-BCI. Ten seconds of eye-centric horizontal and vertical position (x, y) from the tracker and Eye aspect ratio (EAR) for the same segment with automatically detected blinks shaded in grey. The examples illustrate the pupil and canthus tracking and the eyelid aperture estimate.	134
7.3	Example VEOG and VEOS-inferred EOG for MI	139
7.4	Video- vs Tobii-based EOG inference during a P300 blink train	140
7.5	Fp1 before and after LCA	143
7.6	Session-wise blink peak-to-peak reduction with different ocular references	144

- 8.1 BCI-level effects of ocular artefact handling across three paradigms in the Eye-BCI dataset. Grouped bars show median classification accuracy for motor imagery (MI), steady-state visual evoked potential (SSVEP) and P300 spelling under four cleaning conditions: no explicit ocular channel, LCA with real VEOG, LCA with VEOS video-inferred EOG and LCA with Tobii-based surrogates. For MI and P300, real VEOG and VEOS both yield small but consistent gains relative to having no ocular channel, whereas Tobii-based cleaning offers little benefit and can reduce accuracy. In SSVEP, all conditions perform similarly, reflecting the relative insensitivity of occipital SSVEP responses to frontal blink artefacts. 158
- 8.2 Trade-off between sensor configuration and performance. Grouped bars summarise normalised performance (real VEOG = 1.0) for three approaches: real VEOG, VEOS video-inferred EOG and Tobii-based surrogates, across three levels of analysis. Signal-level fidelity reflects median correlation with VEOG; blink suppression measures the percentage reduction in blink peak-to-peak amplitude at frontal electrodes after LCA; and BCI performance combines median motor imagery and P300 classification accuracy. VEOS closely tracks real VEOG for blink suppression and BCI performance but remains noticeably weaker at the signal-level, while Tobii surrogates lag behind on both blink suppression and BCI accuracy. 159
- 8.3 Conceptual positioning of VEOS in terms of sensor burden and support for EEG data quality. The horizontal axis reflects overall sensor burden (number of sensors, intrusiveness and set-up time), while the vertical axis reflects the ability of each configuration to capture and correct ocular artefacts. Eye-tracker-only systems have very low sensor burden but offer only limited support for EEG data quality. Conventional EEG without a dedicated ocular channel provides moderate support at higher burden, and adding EOG electrodes (EEG + EOG) yields strong support at the cost of additional sensors. VEOS, implemented as EEG plus a single camera, is plotted directly above “EEG only”: it maintains a similar sensor burden while providing improved support for EEG data quality by recovering eye information from the video. . . 169

List of Tables

1.1	Research questions and primary chapters in which they are addressed. .	9
6.1	Summary of the four Eye-BCI paradigms in the Eye-BCI multimodal dataset [45]. For each paradigm, the table lists the main visual cue, the nominal number of trials per session (with the total number of trials across all 63 sessions in parentheses), the typical per-trial duration estimated from the recorded task times, the typical task recording time per session, and the dominant ocular behaviour elicited by the paradigm.	110
6.2	Comparison of the VEOS development dataset and the Eye-BCI dataset.	118
7.1	Landmark errors (mean $\pm$ SD [median], 95th percentile) on held-out subjects.	131
7.2	Eyeball position and eye aspect ratio (EAR) accuracy (mean $\pm$ SD). .	131
7.3	VEOG prediction on VEOS development dataset (median over subjects).	136
7.4	VEOG prediction on Eye-BCI (LOSO, median over subjects).	137
7.5	LCA with real VEOG (median over sessions).	141
7.6	LCA with inferred EOG (median over sessions).	142
7.7	MI left vs right hand classification accuracy (median over subjects). .	145
7.8	SSVEP target frequency classification accuracy (median over subjects).	147
7.9	P300 symbol classification accuracy (median over subjects).	147
A.1	Example MLP architecture for EOG inference.	182
A.2	Example TCN architecture for EOG inference.	183
A.3	LSL streams in the VEOS development platform.	185

List of Abbreviations

BCI Brain–Computer Interface.

DCU Dublin City University.

DPIA Data Protection Impact Assessment.

EAR Eye Aspect Ratio (normalised eyelid aperture measure).

EEG Electroencephalography.

EMG Electromyogram (eyelid muscle activity).

EOG Electrooculogram.

ERD Event-Related Desynchronisation.

ERP Event-Related Potential.

ERS Event-Related Synchronisation.

ESP32 ESP32 microcontroller (used for hardware timing and triggers).

Eye–BCI Eye–BCI multimodal dataset (EEG, eye-tracking and behaviour).

F1 F1 score (harmonic mean of precision and recall).

GAN Generative Adversarial Network.

GDPR General Data Protection Regulation.

GELU Gaussian Error Linear Unit (activation function).

ICA Independent Component Analysis.

IEEE Institute of Electrical and Electronics Engineers.

IMU Inertial Measurement Unit.

JSON JavaScript Object Notation (serialised message format).

LCA Linear Component Analysis (regression-based artefact removal).

LDA Linear Discriminant Analysis.

LED Light-Emitting Diode (timing LED in the video frame).

LOSO Leave-One-Subject-Out (cross-validation regime).

- LSL** Lab Streaming Layer (multimodal time-synchronised streaming library).
- MAE** Mean Absolute Error.
- ME** Motor Execution (actual movement task).
- MI** Motor Imagery (imagined movement task).
- MLP** Multilayer Perceptron (feed-forward neural network).
- nRMSE** Normalised Root Mean Squared Error.
- P300** P300 event-related potential component.
- PC** Personal Computer (host machine running the acquisition software).
- REC** Research Ethics Committee.
- RMSE** Root Mean Squared Error.
- ROI** Region of Interest.
- SD** Standard Deviation.
- SNR** Signal-to-Noise Ratio.
- SSD** Spatio-spectral Decomposition.
- SSP** Signal-Space Projection.
- SSVEP** Steady-State Visually Evoked Potential.
- TCN** Temporal Convolutional Network.
- TTL** Transistor–Transistor Logic (digital trigger pulses).
- TX300** Tobii TX300 eye-tracker.
- VEOG** Vertical electrooculogram (vertical EOG channel).
- VEOS** Vision-based Electrooculogram Signal Inference (proposed system in this thesis).
- VEOS-Dev** VEOS development dataset (EEG + EOG + video used to train VEOS).

List of Publications

During the period of this doctoral research (2022–2025), contributions were made to the following peer-reviewed publications and submissions. Items marked with an asterisk (*) are most directly related to the work presented in this thesis.

Journal articles

1. E. Brophy, P. Redmond, A. Fleury, M. De Vos, G. Boylan and T. E. Ward, “Denoising EEG signals for real-world BCI applications using GANs”, *Frontiers in Neuroergonomics*, vol. 2, 805573, 2022.*
2. M. A. Awais, T. E. Ward, P. Redmond and G. Healy, “From lab to life: assessing the impact of real-world interactions on the operation of rapid serial visual presentation-based brain-computer interfaces”, *Journal of Neural Engineering*, vol. 21, no. 4, 046011, 2024.*
3. L. Casey, J. Dooley, M. Codd, R. Dahyot, M. Cognetti, T. Mullarkey, P. Redmond *et al.*, “A real-time digital twin for active safety in an aircraft hangar”, *Frontiers in Virtual Reality*, vol. 5, 1372923, 2024.
4. M. A. Awais, P. Redmond, T. E. Ward and G. Healy, “AMBER: advancing multimodal brain-computer interfaces for enhanced robustness: a dataset for naturalistic settings”, *Frontiers in Neuroergonomics*, vol. 4, 1216440, 2023.*

Conference papers and abstracts

1. P. Redmond, M. Ahsan and T. E. Ward, “Inferring EOG signal from video”, in *Proceedings of the MEG United Kingdom and Ireland Conference*, pp. 29–30, 2023.*
2. P. Redmond, M. A. Awais and T. E. Ward, “Automatic tagging of BCI artefacts using computer vision”, in *Proceedings of the 10th International BCI Meeting*, 2023.*
3. P. Redmond, A. Fleury and T. E. Ward, “An open-source multi-modal data-acquisition platform for experimental investigation of blended control of scale vehicles”, in *Proceedings of the IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroX-RAINE)*, 2022.*
4. P. Redmond, M. A. Awais and T. E. Ward, “Ethical and GDPR-compliant sharing of multimodal stress data: contributions to environmental and health research”, in *Joint Conference of ISEH, ICEPH and ISEG*, 2024.*

5. M. A. Awais, P. Redmond, T. E. Ward and G. Healy, “Investigating the impact of ecologically valid interactions on rapid serial visual presentation-based brain-computer interface performance”, in *Proceedings of the 10th International BCI Meeting*, 2023.*
6. A. Farooq, T. J. Creagh, E. P. Mooney, C. Attard and P. Redmond, “Cluster-Bot: an innovative strategy introduced in the IEEE R8 Robotics Championship 2024”, in *Proceedings of the 2nd International Conference on Advanced Technology in Electronic and Electrical Engineering (ICATEEE 2025)*, December 2025.
7. M. A. Mohamed, M. Maniparambil, P. Redmond, T. E. Ward and N. E. O’Connor, “Attention-based temporal resolution enhancement for microsleep forecasting across multiple horizons”, in *Proceedings of the 33rd International Conference on Artificial Intelligence and Cognitive Science (AICS’25)*, Dublin, December 2025.
8. M. A. Mohamed, M. Maniparambil, P. Redmond, T. E. Ward and N. E. O’Connor, “Comprehensive evaluation of EEG features for microsleep detection: statistical validation and methodological improvements”, in *Proceedings of the 33rd International Conference on Artificial Intelligence and Cognitive Science (AICS’25)*, Dublin, December 2025.

Manuscripts in preparation or under review

1. P. Redmond, A. Fleury and T. E. Ward, “VEOS: vision-based vertical electrooculography inference from monocular periocular video for ocular artefact suppression in EEG brain-computer interfaces”, *Biomedical Physics & Engineering Express*, under review.*
2. M. A. Mohamed, P. Redmond, T. E. Ward and N. E. O’Connor, “Mutual-information-guided multimodal fusion of EEG and EOG for robust microsleep prediction in maintenance of wakefulness test (MWT) data”, *Computer Methods and Programs in Biomedicine*, under review.
3. M. A. Mohamed, N. E. O’Connor and P. Redmond, “In-depth analysis of the most indicative features for detecting microsleeps from EEG signals”, manuscript in preparation.

Publications prior to the doctoral period

The following peer-reviewed publications predate the period of doctoral research and are included to provide context on prior research experience.

1. J. Cahill, P. Redmond, S. Yous and W. Butler, “Human computer action design (HCI) methods supporting the envisionment, design and evaluation of a collision avoidance system”, in *Proceedings of the Fourth Irish Human Computer Interaction Conference (iHCI 2010)*, Dublin City University, p. 139, 2010.

2. J. Cahill, P. Redmond, S. Yous, G. Lacey and W. Butler, “The design of a collision avoidance system for use by pilots operating on the airport ramp and in taxiway areas”, *Cognition, Technology & Work*, vol. 15, no. 2, pp. 219–238, 2013.
3. P. Redmond, S. Yous and G. Lacey, “Full self-calibration of non-overlapping cameras”, in *Proceedings of the International Machine Vision and Image Processing Conference*, 2009.
4. P. Redmond, S. Yous and G. Lacey, “Method and system for recovery of 3D scene structure and camera motion from a video sequence”, U.S. Patent 9,196,054, 2015.

VEOS: Vision-based Electrooculogram Signal Inference for Improved EEG in Brain-Computer Interfaces

Peter Redmond

Abstract

Electroencephalography (EEG) is highly susceptible to ocular artefacts from eye movements and blinks, which contaminate frontal channels and degrade brain-computer interface (BCI) performance. Artefacts are commonly mitigated using electrooculography (EOG) as an auxiliary channel, but extra periocular electrodes increase setup time and are often absent in recordings, whereas many BCI systems already capture face video. This research asks whether such video can provide an EOG-like surrogate for ocular artefact handling in EEG.

VEOS (Vision-based Electrooculogram Signal Inference) is a framework that maps monocular eye-region video to an inferred EOG signal. A theoretical model links eyeball and eyelid kinematics, their two-dimensional image landmarks, and the one-dimensional electrical signals observed in EOG and frontal EEG. A multimodal platform was implemented to record synchronised face video, EEG, and horizontal and vertical EOG during controlled ocular tasks, forming the VEOS development dataset. An established open dataset providing high-density EEG, EOG, and monocular eye video, collected independently by other researchers, serves as an independent test bed for evaluating generalisation.

VEOS uses dense facial landmarks to estimate eye-centred pupil position and eyelid aperture. Pupil and canthus locations are recovered with errors below one eye width, while noisier eyelid landmarks still support blink timing and aperture estimation. Temporal neural networks operating on eye pose and aperture achieve correlations of 0.75–0.8 with EOG, outperforming linear baselines.

The inferred EOG is integrated into a regression-based linear component analysis pipeline. VEOS recovers roughly 60–70% of the blink suppression obtained with true EOG in frontal EEG, preserves motor imagery, steady-state visual evoked potential and P300 features within a few percent, and yields BCI accuracy gains comparable to real EOG. The research shows that single-camera, vision-based EOG inference is feasible and useful for EEG artefact mitigation, while not fully replacing dedicated EOG channels.

Chapter 1

Introduction

1.1 Motivation

Electroencephalography (EEG) is a central tool in brain–computer interface (BCI) research because it offers millisecond temporal resolution and can be deployed with relatively lightweight hardware in laboratory and semi-naturalistic settings [1], [69], [112]. At the same time, EEG is fragile, with neural components typically in the microvolt range and easily masked by non-neural activity originating from muscles, electrical interference or the eyes [22], [144]. Among these, ocular artefacts due to blinks and eye movements are particularly problematic for BCI applications that rely on frontal or central channels, or that require participants to look at visual displays for extended periods [32], [36], [42], [144].

In laboratory settings, ocular activity is commonly monitored by attaching electrodes around the eyes to record the electrooculogram (EOG) [27], [144]. The EOG provides a comparatively direct view of the eye-related electrical field and is widely used to detect and suppress blink and saccade artefacts in EEG recordings [27], [32], [144]. In event-related potential (ERP) paradigms such as P300 spellers and in steady-state visually evoked potential (SSVEP) paradigms, this information is often essential for maintaining data quality over long recording sessions [35], [87], [103], [155]. Similarly, rapid serial visual presentation (RSVP) paradigms, which present stimuli sequentially at high rate, can reduce overt gaze shifts but remain vulnerable to blink-

and saccade-related contamination[9], [158]. In many deployed or naturalistic scenarios, however, it is not feasible to attach dedicated EOG electrodes to every user. Extra sensors increase setup time, can reduce comfort and acceptability, and are difficult to integrate into unobtrusive or wearable systems [42], [80], [122].

In parallel, commodity video cameras and eye-trackers have become commonplace. Eye and face video appear in driving simulators, mobile devices and experimental BCI platforms [57], [79]. Modern computer vision methods can track facial landmarks, pupil position and eyelid motion in close to real time using low-cost cameras [24], [93], [110], [130], [146]. These observations are directly related to the physical movements that drive EOG and to eyelid electromyogram (EMG) activity recorded in periocular channels [36], [55], [144]. This raises a natural question: if video captures the kinematics of the eyes and eyelids, can it be used to infer an EOG-like signal when no physical EOG electrodes are present [117], [118], [154]?

The work in this thesis addresses this question. It proposes VEOS, a vision-based electrooculogram signal inference framework that learns a mapping from video-based eye and eyelid features to EOG. The inferred EOG signal is then used to support artefact handling in EEG, particularly in BCI paradigms that are sensitive to ocular contamination. The central idea is to treat video as an additional modality that can be translated into an electrical surrogate of EOG, so that familiar EEG cleaning pipelines can operate even when only EEG and video have been recorded.

VEOS is evaluated at three levels using a multimodal development dataset and the independent Eye-BCI dataset [45], [129]. First, the monocular eye-tracker within VEOS is assessed in terms of pupil, eyeball-in-socket and eyelid-tracking accuracy on manually annotated frames. Second, the mapping from ocular measurements to an EOG-like signal is studied using linear and neural models and several input combinations, including video-only features and Tobii gaze features. Third, the inferred signals are fed into an ocular artefact removal pipeline based on linear component analysis (LCA) and are tested in complete BCI pipelines for motor imagery, motor execution, SSVEP and P300 paradigms.

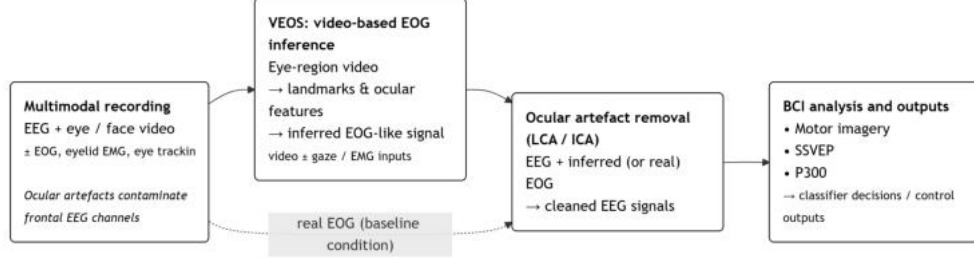


Figure 1.1: Multimodal recordings combine EEG with eye and face video and, in some datasets, additional ocular channels such as EOG, eyelid EMG and eye-tracking. VEOS uses eye-region video, and optionally other ocular measurements, to infer an EOG-like signal. This inferred signal, or real EOG when available, is then used as a reference in regression- and ICA-based ocular artefact removal, producing cleaned EEG. Standard BCI pipelines for motor imagery, SSVEP and P300 paradigms operate on the cleaned signals to produce classifier decisions and control outputs.

At a high level, the results show that monocular video tracking achieves sub-eye-width accuracy for pupil and canthus position and yields stable eyeball coordinates, while eyelid-tracking is less accurate but sufficient for reliable blink detection. EOG inference from non-electrical signals is feasible but imperfect. Video-based models outperform Tobii-only surrogates on most metrics, although none fully match blink morphology or subject-specific amplitude. When inferred EOG replaces real EOG in the LCA pipeline, video-based surrogates recover a substantial fraction of blink suppression, whereas Tobii-based surrogates can introduce additional noise, particularly because blink magnitudes are poorly aligned. At the BCI-level, VEOS-based cleaning yields small but consistent improvements relative to having no ocular channel and comes close to the gains obtained with real EOG.

Figure 1.1 provides a graphical overview of this setting and of the role of VEOS within it.

1.2 Problem statement

Ocular artefacts remain one of the most persistent obstacles to reliable EEG in BCI research, and their management is still dominated by approaches that require dedicated EOG electrodes [32], [42], [124], [144]. In many studies it is simply assumed that EOG is available and of sufficient quality to support regression-based corrections or to guide independent component analysis (ICA)[27], [32], [144]. When EOG is not present, ocular components must be identified using spatial patterns and qualitative inspection alone, which reduces reliability and complicates automation[59], [81], [82], [90].

This creates a gap between controlled laboratory experiments, in which it is acceptable to use a full cap and dedicated EOG channels and in which participants tolerate extensive preparation, and more naturalistic or applied scenarios, in which the number of sensors, setup complexity and user comfort are constrained, but in which eye and face video or eye-tracker signals are often available as part of the interface.

From the point of view of signal processing, ICA and related spatial filtering techniques can separate ocular components from multichannel EEG if they are given some cue about which components correspond to eye movements. A conventional way to provide this cue is to compute correlations between component time courses and recorded EOG channels and to examine components that show strong correlation and frontal scalp maps. In the absence of EOG, the correlation step is missing, even if the theoretical assumptions behind ICA remain valid. Automated independent component labelling tools can reduce reliance on manual inspection, but they still benefit from an explicit ocular reference channel[119]. These issues, together with a more detailed review of ocular artefacts and artefact-handling methods, are discussed in Chapter 2.

The core problem addressed in this thesis can therefore be stated as follows:

Given synchronised eye-region video and an EEG recording, can a surrogate EOG signal be inferred from the video that is accurate enough to serve as an auxiliary channel for ocular artefact identification and removal

in EEG, with particular emphasis on BCI paradigms?

A satisfactory answer must address several related questions. It must show that there is a signal-level relationship between visual features and EOG, that this relationship can be captured by tractable models, that inferred EOG signals can drive an artefact-cleaning pipeline that was originally designed for real EOG, and that any improvements in EEG quality translate into measurable differences in BCI performance. These aspects are formalised as explicit research questions in Section 1.3.

1.3 Research questions

The overall research question guiding this work is:

To what extent can EOG-like signals, suitable for supporting ocular artefact handling in EEG, be inferred from video of the eyes and face and related ocular measurements?

From this, several more concrete research questions follow. These six questions (RQ1–RQ6) are addressed empirically in Chapters 7 and 8 and summarised in Chapter 9.

1. **Monocular eye-tracking performance.** How accurately can a monocular eye-tracking pipeline recover pupil position, eyeball-in-socket position and eyelid aperture from close-up video on the development dataset and on the Eye-BCI dataset? How do these errors compare with inter-rater variability on manually annotated landmarks?
2. **Mapping feasibility.** Is there a stable functional relationship between video-based eye and eyelid features, possibly combined with gaze coordinates from an eye-tracker, and the simultaneously recorded EOG, across participants and tasks, that can be captured by supervised regression models?

3. **Model adequacy.** Are relatively simple models, such as regularised linear regression with temporal context, sufficient to capture the main structure of this mapping, or do non-linear models such as multilayer perceptrons and temporal convolutional networks provide a clear and justified advantage, especially around blinks and fast eye movements?
4. **Dataset dependence and generalisation.** How well do mappings learned on the VEOS development dataset (VEOS-Dev) transfer to an independently collected external dataset (Eye-BCI) with different hardware, higher-density EEG and a different camera configuration?
5. **Impact on EEG cleaning.** When integrated into an artefact-handling pipeline based on linear component analysis and related regression strategies, how closely does the performance of a system using inferred EOG match that of a system using true EOG, in terms of residual ocular artefacts, blink-linked metrics and preservation of task-related EEG features such as ERPs and sensorimotor rhythms?
6. **BCI-level consequences.** Do changes in artefact handling driven by inferred EOG lead to observable differences in BCI-relevant metrics, such as classification accuracy or signal-to-noise measures, in paradigms such as motor imagery, motor execution, SSVEP and P300? Are there situations in which inferred EOG materially improves performance relative to having no ocular channel, and are there situations in which certain surrogates, such as Tobii-based models, degrade performance[35], [115], [120], [155]?

These questions reflect both the scientific curiosity behind VEOS and the practical requirement that any proposed solution must fit into existing BCI processing chains with minimal disruption. Table 1.1 maps each question to the chapters in which it is primarily addressed.

Table 1.1: Research questions and primary chapters in which they are addressed.

Research ques- tion	Primary chapters
RQ1: Mono- cular eye-tracking performance	Methods in Chapter 5; quantitative evaluation in Chapter 7; interpretation in Chapter 8.
RQ2: Mapping feasibility	Model formulation in Chapters 3 and 5; empirical tests in Chapters 6 and 7.
RQ3: Model ad- equacy	Linear and non-linear models in Chapter 5; comparative re- sults in Chapter 7; discussion in Chapter 8.
RQ4: Dataset dependence and generalisation	Dataset descriptions in Chapter 6; cross-dataset and cross- modality analyses in Chapter 7; implications in Chapter 8.
RQ5: Impact on EEG cleaning	LCA and ICA integration in Chapter 5; artefact-removal re- sults in Chapter 7; broader interpretation in Chapter 8.
RQ6: BCI-level consequences	BCI pipelines in Chapter 6; BCI performance results in Chap- ter 7; overall conclusions in Chapter 9.

1.4 Scope and assumptions

The scope of this thesis is deliberately narrow in some respects and broad in others.

On the EEG side, the work concentrates on artefacts arising from eye movements and eyelid activity. Other sources of contamination, such as jaw clenching, neck muscle activity or line-related interference, are acknowledged but not modelled explicitly. The aim at this time is not to build a complete artefact suppression framework, but to address the specific problem of ocular artefacts using information from video and related ocular measurements

On the vision side, the emphasis is on landmark-based description of eye and eyelid motion using a single low-cost camera. Dedicated infrared eye-trackers and multi-camera systems can achieve higher precision in gaze estimation, but they are not always available in BCI settings and are not necessary for the inference of EOG-like signals. VEOS is designed to work with standard cameras that can be integrated into existing BCI platforms, while still allowing comparison with a dedicated eye-tracker in the Eye-BCI dataset.

On the modelling side, the work adopts a supervised learning viewpoint. The mapping from video features to EOG is learned from data that contains both modal-

ities. This implies an assumption that the training data are representative of the conditions under which VEOS will be used and that the synchronisation between video and electrophysiology is accurate enough that temporal structure is preserved. The multimodal acquisition platform described in Chapter 4 and the synchronisation procedures were designed with these assumptions in mind.

The focus is on offline model training and evaluation. Real-time deployment is discussed in terms of computational cost and latency, but full online adaptation and closed-loop interaction lie outside the present scope. Finally, the thesis treats the inferred EOG as a tool for improving artefact handling rather than as a new neural signal in its own right. The intention is to model non-neural sources that are usually regarded as artefacts in EEG analyses. VEOS does not attempt to infer cognitive states directly from video.

1.5 Contributions

The main contributions of this thesis can be grouped into four areas: theoretical framework, data acquisition, algorithm design and empirical evaluation across multiple datasets and BCI paradigms.

Theoretical framework

The first contribution is a theoretical framework that links eyeball and eyelid kinematics, camera geometry, EOG generation and EEG artefacts.

Chapter 3 develops a projective camera model and a head-centred coordinate system for interpreting facial landmarks from video; a kinematic description of eyeball rotation and eyelid motion, including their temporal dynamics and the constraints they impose on plausible trajectories; models of EOG as an eye-centred electric dipole whose orientation follows the optical axis, together with a simplified description of how this dipole projects to scalp EEG; a treatment of eyelid EMG and periocular muscle activity, including their spectral characteristics relative to EOG and EEG;

and a review of ICA for EEG that explains how ocular sources appear as independent components that can be targeted for removal.

This framework provides the theoretical justification for inferring an EOG-like signal from video and for using it within an artefact-handling pipeline that was originally defined for real EOG.

Multimodal data acquisition platform

The second contribution is the design and implementation of a multimodal data acquisition platform tailored to the needs of VEOS. Chapter 4 describes a complete system that includes: The multimodal platform comprises a 32-channel EEG cap focused on frontal and central sites, supplemented by dedicated horizontal and vertical EOG channels; a high-frame-rate camera, configured at 120Hz, that captures the full face with particular emphasis on the eyes and eyelids; stimulus presentation and task control implemented using OpenViBE and custom Python software, including a gaze-calibration task, a game-based engagement task and ERP-like paradigms; and synchronisation of EEG, EOG, video and event markers using the Lab Streaming Layer (LSL), supported by hardware LED markers and dedicated validation procedures.

This platform builds the VEOS development dataset (VEOS-Dev), in which EEG, EOG and video are tightly aligned in time and in which eye and eyelid behaviour are deliberately elicited under several task conditions.

VEOS algorithm and implementation

The third contribution is the VEOS algorithm itself, described in Chapter 5. In VEOS, video frames are processed with OpenCV and MediaPipe Face Mesh to obtain dense facial landmarks with emphasis on the periocular region. Geometric and temporal features are then constructed from these landmarks to represent eye position, eyelid aperture (expressed as an eye aspect ratio, EAR), velocities, accelerations and simple descriptors of blinks and saccades. EOG (and, in the Eye-BCI dataset, eyelid EMG

and eye-tracker signals) is preprocessed, baseline-corrected and aligned with the video frame times to provide supervised targets. Regression models with temporal context are trained to map feature sequences to EOG, starting from regularised linear models and extending to windowed multilayer perceptrons and temporal convolutional networks where the performance gains justify the additional complexity. The inferred EOG is then integrated into an artefact-handling pipeline based on LCA and ICA by treating it as a surrogate ocular channel for component identification and regression.

Many of the practical details that arose during development, such as temporal window selection, smoothing strategies and blink-aware loss functions, are described here.

Evaluation on VEOS-Dev and Eye-BCI datasets

The final contribution is an empirical evaluation of VEOS on two distinct datasets, introduced in Chapter 6 and analysed in Chapter 7.

The first is the VEOS development dataset (VEOS-Dev), recorded by the author with the multimodal platform described above, which provides joint EEG, EOG and face video for tasks that include controlled gaze calibration, game-based engagement and ERP-like visual stimulation. The second is the Eye-BCI multimodal dataset [45], an openly available dataset collected independently by Guttmann-Flury and colleagues, which combines high-density EEG, high-speed eye-region video, vertical EOG, eyelid EMG and eye-tracking across several BCI paradigms. It was not collected as part of this research and serves here as an independent test bed with richer ocular annotations than VEOS-Dev.

In addition, the thesis draws on the AMBER dataset [8] described in Chapter 2 when motivating task design and when considering how VEOS might be applied to more naturalistic EEG+video recordings that lack dedicated EOG channels.

Across these datasets, the thesis evaluates how well the inferred EOG tracks true EOG and, where available, eye-tracker signals, using pointwise error measures, band-limited spectral similarity and blink-level agreement; how the availability of

inferred EOG affects LCA-based identification and removal of ocular components, in comparison with using true EOG or having no ocular channel; and how changes in artefact handling propagate to BCI-relevant metrics, particularly classification performance for motor imagery, motor execution, SSVEP and P300 paradigms.

These contributions demonstrate that video-based inference of EOG is feasible with standard cameras and that the resulting surrogate signals can meaningfully support ocular artefact handling in EEG. A preliminary version of this work, focusing on proof-of-concept EOG inference from video, was presented in [126].

1.6 Thesis structure

The remainder of this thesis is organised as follows.

Chapter 2 reviews relevant background in EEG, EOG and EMG physiology, ocular artefacts in EEG, BCI paradigms and computer-vision approaches to eye and eyelid tracking. It situates VEOS within the broader literature on artefact correction and multimodal BCIs.

Chapter 3 develops the theoretical foundations of the work, from camera geometry and eyeball kinematics through EOG modelling and propagation to scalp EEG, to ICA and the formal definition of the mapping from video features to EOG. It also introduces timing considerations that are important for synchronising video, EEG, EOG and eye-tracker streams.

Chapter 4 describes the multimodal data acquisition platform used to record the VEOS development dataset, including hardware and software architecture, synchronisation mechanisms and task design.

Chapter 5 introduces the VEOS algorithm in detail, covering video processing, feature engineering, target construction, regression models and integration with LCA- and ICA-based EEG cleaning. Material relating to the experiments on Tobii-to-VEOG modelling, windowed neural networks and blink detection is described in this chapter.

Chapter 6 presents the experimental protocols and datasets used in this thesis, including both the VEOS development dataset and the Eye-BCI dataset with eye-

tracking, high-speed video and eyelid EMG. It formalises the evaluation protocol, including participant-wise splits and metrics.

Chapter 7 reports quantitative results on monocular eye-tracking performance, EOG inference, blink and saccade modelling, and the impact of inferred EOG on EEG artefact handling and BCI performance. It covers both development and Eye-BCI datasets and includes comparisons between video-based, Tobii-based and EMG-augmented surrogates.

Chapter 8 discusses these results in the context of the theoretical framework and existing literature, outlines limitations and reflects on the implications for future BCI systems that rely on video rather than dedicated ocular electrodes.

Chapter 9 concludes the thesis by summarising the main findings and contributions and by identifying promising directions for further research on vision-based support for EEG and BCIs.

Chapter 2

Background and Literature Review

2.1 Introduction

This chapter sets VEOS within the broader field of brain-computer interfaces (BCIs), ocular physiology, artefact handling, eye-tracking and computer vision, and multi-modal datasets. The goal is to connect the core idea of vision-based electrooculogram (EOG) inference to well-established scientific work and highlight the specific gap that VEOS addresses.

BCI research has moved from proof-of-concept laboratory demonstrations to increasingly demanding real-world and neuroergonomic applications [1], [17], [76], [97], [112], [152]. However, non-invasive BCIs based on scalp electroencephalography (EEG) remain highly sensitive to ocular activity. Eye movements, blinks and eyelid muscle activity introduce large artefacts that overlap strongly with the frequency bands and scalp regions used for common BCI control signals[32], [42], [45], [144]. Traditional approaches either treat these signals as pure noise to be removed, or exploit them as an alternative control channel, but seldom do both in a principled way[12], [32], [83], [137].

VEOS is motivated by the observation that modern computer vision methods provide rich information about eye and eyelid motion from ordinary video. If this information can be transformed into an inferred EOG-like signal with sufficient temporal and spatial fidelity, then it becomes possible to support both artefact mitigation

and ocular-based interaction without placing additional electrodes on the user. To understand how and why that might work, this chapter reviews (i) BCI principles and application domains, (ii) EEG acquisition and artefacts with a focus on ocular activity, (iii) existing artefact handling techniques, (iv) eye-tracking and gaze estimation in computer vision, (v) multimodal EEG–eye datasets, and (vi) prior work at the intersection of these areas. The chapter closes by identifying the specific research gaps that VEOS addresses, which in turn motivate the research questions stated in the Introduction (Chapter 1) and the more formal theoretical treatment in Chapter 3.

2.2 Brain–Computer Interfaces and Neuroergonomics

Brain–computer interfaces exploit the measurable electrical activity of the brain to create a direct communication channel between the nervous system and an external device, bypassing conventional motor pathways. Understanding their development, scope and practical limitations is essential context for the work presented in this thesis, which targets a specific and persistent obstacle to their real-world use.

2.2.1 Historical development and core concepts

Early BCI systems emerged from fundamental work showing that voluntary modulation of EEG rhythms and event-related potentials could be used for communication in paralysed patients [17], [152]. Subsequent reviews charted the development of both invasive and non-invasive BCIs and established the standard taxonomy of BCI paradigms and control signals [76], [97]. Non-invasive EEG remains the dominant modality for BCIs outside specialist clinical settings due to its portability, low risk and relatively low cost.

A BCI system typically comprises: (i) signal acquisition, (ii) pre-processing and artefact handling, (iii) feature extraction, (iv) classification or regression, and (v) a feedback or control loop. Different paradigms exploit different neural phenomena, including sensorimotor rhythms for motor imagery (MI), P300 event-related poten-

tials for oddball-based paradigms, steady-state visually evoked potentials (SSVEPs) for periodic visual stimulation, and slow cortical potentials [16], [115], [120], [123]. Performance is evaluated using accuracy, information transfer rate, robustness across sessions and users, and more recently user experience and neuroergonomic measures [10], [15], [112].

Recent reviews have emphasised both how far BCI technology has progressed and how fragile performance remains when moving from controlled laboratory experiments to more naturalistic environments [1], [69], [112]. Robust handling of artefacts, especially ocular artefacts, is repeatedly highlighted as a limiting factor for real-world deployment [42], [69], [112].

2.2.2 Application domains relevant to VEOS

BCIs have been investigated in assistive communication and control [25], [86], neurorehabilitation and neuroprosthetics [123], [125], neurofeedback [6], safety-critical monitoring, and entertainment [77], [85]. Many of these domains impose practical constraints that make additional electrodes undesirable. Examples include:

- **Rehabilitation and neuroprosthetics**, where systems must be worn for long periods and integrated into complex clinical contexts [16], [123], [125].
- **Operational and safety monitoring**, where users are already equipped with multiple devices and where motion and environmental noise are significant [10], [15].
- **Games and virtual reality**, where commercial eye-trackers and video streams are already present, but full EOG montages may not be tolerated [77], [85].

VEOS is particularly aligned with neuroergonomic applications, where EEG is used to monitor workload, vigilance and attention in realistic settings with frequent eye, head and body movements [10], [15], [69]. In such conditions, ocular artefacts are extensive and strongly task-dependent, and it is not always feasible to place dedicated EOG electrodes or to constrain eye movements.

2.2.3 From laboratory to life: ecological validity

A recurring theme in BCI research is the gap between laboratory performance and performance in complex, interactive environments [1], [112]. Neuroergonomics frames this as a question of ecological validity: to what extent do BCI systems maintain reliability when embedded in real work and life contexts [10], [15]? Two strands of work within this thesis are directly relevant.

First, the AMBER dataset provides multimodal recordings of EEG, behaviour and environmental context in a purpose-built experimental framework designed to resemble realistic interaction [8]. It demonstrates the value of combining multiple sensing modalities and illustrates the diversity of artefacts observed outside strict laboratory control.

Second, recent work on RSVP-based BCIs has explicitly compared performance in constrained laboratory settings with performance when participants interact more freely with their environment [9], [158]. That study shows how real-world interactions degrade the operation of an otherwise robust paradigm, and highlights artefact handling and context awareness as key factors.

Both lines of work point towards the same practical conclusion: if a camera is already present in a BCI system, it should be possible to use it for artefact handling rather than adding periocular electrodes solely for that purpose.

2.3 EEG Acquisition, Signal Quality and Artefacts

EEG is a non-invasive, high-temporal-resolution measure of brain electrical activity, but its utility in practice is heavily constrained by the quality of the recorded signal. Noise introduced during acquisition and contamination from physiological sources unrelated to the neural signals of interest are among the most significant practical barriers to reliable BCI operation outside controlled laboratory conditions.

2.3.1 EEG acquisition and practical constraints

EEG measures scalp potentials arising from synchronous post-synaptic activity in large populations of cortical pyramidal neurons [22], [124]. Traditional research systems use gel-based electrodes in standard montages, such as the 10–20 or 10–10 systems [64], [70], with high-input-impedance amplifiers and shielded cables. In practice, the choice of electrode type, montage and amplifier properties interacts strongly with artefact susceptibility and usability [22], [144].

Over the past decade, there has been substantial interest in dry or semi-dry electrode systems that reduce preparation time and user discomfort, but often at the cost of increased impedance and noise [80], [122]. Studies comparing different devices and electrode technologies show clear trade-offs between signal quality, motion robustness and ease of deployment [80], [122]. In mobile and wearable scenarios, movement-induced artefacts from electrode motion, cable sway and poor skin–electrode contact become a major concern [42].

From a BCI design perspective, these constraints motivate algorithms that are robust to variable signal quality and can exploit additional modalities when available. For ocular artefacts, this means going deeper than purely EEG-based artefact handling to exploit information from EOG electrodes, eye-trackers or video where feasible.

2.3.2 Physiological and environmental artefacts

EEG recordings are contaminated by multiple sources of artefacts:

- **Ocular activity**, including blinks, saccades and smooth pursuit, originating primarily from the corneo-retinal dipole and eyelid-related electromyographic (EMG) activity [32], [45], [55].
- **Muscle activity**, especially from cranial and neck muscles, which produces high-frequency EMG contamination [22], [42].
- **Cardiac and respiratory activity**, which can project onto scalp electrodes as low-frequency components [22].

- **Environmental noise**, including mains interference and electromagnetic emissions from electronic devices[22].

Comprehensive reviews of EEG artefact removal methods emphasise that ocular artefacts are among the largest and most systematic, often exceeding neural components in amplitude and overlapping with key frequency bands [22], [32], [42]. In frontal and fronto-central electrodes the impact is particularly pronounced, which is problematic for tasks that rely on frontal P300, N200 or frontal midline theta activity[32], [46], [120].

Movement-induced artefacts introduce further complexity. In mobile EEG or tasks involving walking, cycling or head motion, artefacts from motion, muscle and sensor displacement interact with ocular artefacts to produce complex mixtures that are challenging to separate [42]. Guidelines for EEG recording under movement emphasise joint optimisation of hardware, experimental design and signal processing rather than relying on a single algorithmic solution [42].

2.3.3 Impact of ocular artefacts on BCI control signals

Ocular activity interacts with BCI control signals in several ways. In P300 and other ERP-based paradigms, blinks and saccades around stimulus onset can overlap with the latency range of interest, distort waveforms and introduce systematic differences between conditions if eye behaviour is stimulus-dependent [14], [87], [120]. In MI paradigms, changes in vigilance and visual attention associated with eye behaviour can modulate sensorimotor rhythms, while frontal ocular artefacts can leak into central electrodes via volume conduction [10], [115]. In passive BCIs and neuroergonomic monitoring, eye behaviour is itself a marker of workload, fatigue and engagement [15], [26], [46], so treating ocular signals purely as noise discards potentially useful information about the user’s state.

These interactions motivate two complementary goals: to remove ocular artefacts where they mask neural signals of interest, and to interpret ocular activity as a useful channel for behaviour and state monitoring in its own right. VEOS aims to support

both goals by inferring an EOG-like signal from video that can be used for artefact mitigation and for characterising ocular behaviour.

2.4 Ocular Activity: EOG, Eye Movements and Eyelid EMG

The eye is both a source of the signals this thesis seeks to exploit and a principal source of contamination in EEG recordings. The electrical potentials generated by eyeball rotation and eyelid movement propagate through the volume-conducting tissues of the head and are readily picked up by scalp electrodes, often with amplitudes that dwarf the underlying neural signals of interest.

2.4.1 Corneo-retinal dipole and electrooculogram

The electrooculogram arises primarily from the corneo-retinal dipole. The cornea is more positively charged relative to the retina, so rotation of the eyeball produces systematic potential changes at electrodes around the eyes. In a conventional bipolar EOG montage, horizontal EOG (hEOG) is recorded between electrodes placed near the outer canthi, and vertical EOG (vEOG) between electrodes above and below one eye. Saccades, smooth pursuit and blinks all produce characteristic deflections in these channels [12], [55], [83]. A schematic cross-section of the eye highlighting the main anatomical structures and the corneo-retinal dipole that give rise to the EOG is shown in Figure 2.1.

EOG has long been used as a control signal in its own right for human-computer interaction and simple BCIs, for example to detect intentional eye movements in users with severe motor impairments [12], [83]. It is attractive because signal amplitudes are large, detection of discrete events is relatively straightforward, and the relationship between movement and signal is intuitive. However, EOG electrodes increase the preparation burden, are often perceived as intrusive, and can be challenging to integrate into consumer-grade headsets[80], [122].

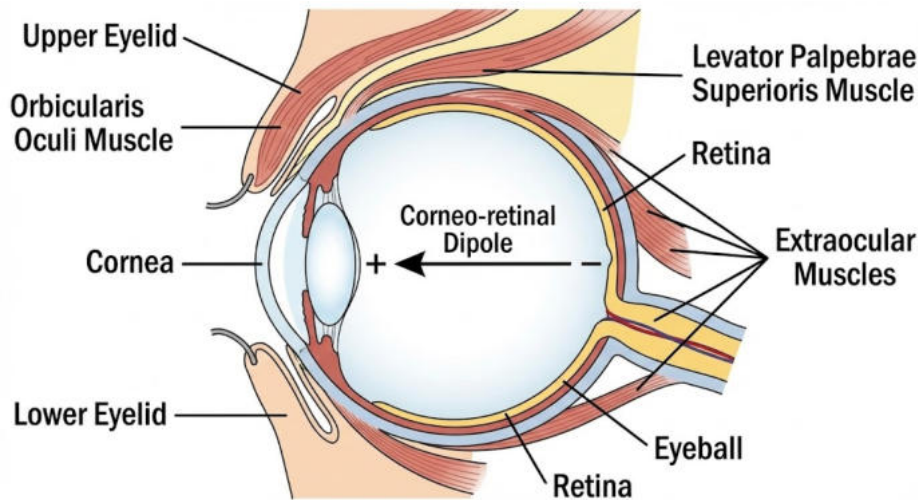


Figure 2.1: Cross-sectional anatomy of the eye highlighting structures relevant to EOG generation. The corneo–retinal dipole is indicated by the positive pole at the cornea and the negative pole at the retina. The diagram also labels the upper and lower eyelids, the orbicularis oculi and levator palpebrae superioris muscles, the extraocular muscles that rotate the eyeball, and the retina, all of which contribute to the ocular potentials and periocular EMG that appear in EOG and EEG recordings.

2.4.2 Eye movement dynamics

Eye movements can be broadly classified into saccades, smooth pursuit, fixations and microsaccades [55]. Saccades are rapid ballistic eye movements that reposition the fovea; they produce large EOG deflections and brief, broadband artefacts in EEG [32], [45], [55]. Smooth pursuit tracks moving targets and results in slower EOG changes [20], [55]. Fixations correspond to relatively stable gaze with small fluctuations and microsaccades [20], [55].

Vergence movements rotate the two eyes in opposite directions to maintain single binocular vision as target distance changes [55]. Although vergence is not ballistic, it produces slow, opposing horizontal EOG deflections that differ in character from saccades and can be mistaken for drift artefacts in monocular recordings. The vestibulo-ocular reflex (VOR) compensates for head movements by generating counter-rotational eye movements to stabilise gaze on a target [55]. In mobile or loosely restrained recording conditions the VOR produces continuous, head-movement-locked EOG and EEG activity that is difficult to distinguish from voluntary eye movements without concurrent head motion data.

Blinks involve rapid closure and opening of the eyelids, modulating both corneo-retinal potentials as the cornea is transiently obscured and eyelid muscle activity [45], [55], [117]. A related but distinct phenomenon is Bell’s phenomenon: during voluntary or reflex eyelid closure the eyeball rotates upward and slightly outward, so that blink-related EOG contains not only the eyelid EMG burst but also a genuine upward rotation of the corneo-retinal dipole [55]. This upward rotation contributes to the characteristically asymmetric blink waveform observed in vertical EOG and is one reason why blink amplitude cannot be predicted from eyelid kinematics alone.

While the neurophysiology of these processes is beyond the scope of this chapter, eye movements are also tightly coupled to visual attention and task demands. For example, RSVP-based BCIs elicit distinct eye movement patterns compared with natural reading or free viewing [9], [158]. This coupling means that ocular artefacts are not random noise but are partially structured by the cognitive processes that BCIs seek to decode.

2.4.3 Eyelid EMG and periocular muscle activity

In addition to the corneo-retinal dipole, eyelid and periocular muscles contribute significant EMG activity[36], [45], [144]. The orbicularis oculi and levator palpebrae muscles are engaged during blinks and squints, producing sharp, high-frequency artefacts that can spread across the scalp[36], [45]. These EMG components are particularly problematic for high-frequency EEG analyses and for BCIs that rely on gamma-band activity or fast transients [22], [42].

In some datasets, such as the Eye-BCI multimodal dataset discussed later, facial EMG channels are recorded explicitly to characterise these muscle contributions alongside EOG and EEG. In others, such as the development dataset in this thesis, eyelid EMG is not measured directly, and its effects must be inferred from EEG and video.

2.4.4 EOG as signal and as artefact

A central theme of this thesis is the conflicting nature of EOG, which lies in the propagation of ocular potentials. The strong electrical dipoles generated by eye movements provide a robust signal for gaze-based interaction. However, via volume conduction, these same potentials spread across the scalp, heavily contaminating neuronal signals. The eye functions as a shared source: its movements generate signals useful for interaction, yet those same electrical currents inevitably corrupt the EEG trace.

EOG has long been used as a control signal in its own right for human-computer interaction and simple BCIs, for example to detect intentional eye movements in users with severe motor impairments [12], [83]. Early work on EOG-based interfaces explored classification of saccades and blinks into discrete command sets and established basic feature sets and classifiers [61], [74]. Subsequent studies extended EOG control to more complex interaction schemes and hybrid BCIs, often combining EOG with other biosignals [12], [36], [83]. In parallel, EOG has been widely used in physiological state monitoring, particularly in driver drowsiness detection, where features such as blink rate, blink duration, slow eye movements and combinations of ocular parameters are fed into rule-based or machine-learning classifiers [23], [38], [62], [63], [121], [159].

These applications exploit the fact that EOG signal amplitudes are large, detection of discrete events is relatively straightforward, and the relationship between movement and signal is intuitive. However, EOG electrodes increase the preparation burden, are often perceived as intrusive, and can be challenging to integrate into consumer-grade headsets [80], [122]. This has motivated research on contactless surrogates for EOG, particularly in automotive settings where cameras are already present. For example, Picot and colleagues compared blink characterisation from high-speed video and EOG using a retina-inspired model and demonstrated that suitably processed video can reproduce many blink-related features of EOG [117]. Lahssene et al. proposed a “bio-inspired” pseudo-EOG signal derived from camera images of the eye, from which features are extracted and used as inputs to a fuzzy-logic classifier of driver awareness level [154]. These studies provide important precedents for VEOS in showing

that video can encode sufficient ocular information to approximate EOG for specific application goals.

Survey work on EEG-based BCIs and on artefact removal consistently treats ocular activity as a primary challenge [22], [32], [42], [69]. Traditional EEG pipelines often include EOG regression or independent component analysis (ICA) based removal of ocular components as standard steps [27], [81], [90], [119], [137]. However, these methods generally require that EOG channels be present. In many practical systems this is not the case. Several commercial EEG devices designed for consumer or field use provide only scalp electrodes, and even in research settings users may refuse additional electrodes around the eyes (Figure 2.2). VEOS explores whether the information normally captured by EOG electrodes can be inferred from video instead, which then allows both artefact handling and ocular-state monitoring to be performed without additional sensors.

2.5 EEG Artefact Handling: From Regression to Deep Learning

Removing the contribution of ocular activity from EEG without distorting the underlying neural signals is a long-standing problem in the field. A wide range of methods has been developed over several decades, reflecting different assumptions about the nature of the mixing process and the availability of reference signals.

2.5.1 Linear regression and linear component analysis (LCA) with recorded EOG

One of the earliest approaches to ocular artefact removal is to treat EOG as a reference signal and regress its contribution out of EEG channels. Linear regression models (LCA) the EEG at each electrode as a sum of neural activity plus a linear combination of EOG channels, sometimes including time-shifted versions to capture delays [27], [131], [132]. Adaptive filters extend this idea with time-varying coefficients, for



Figure 2.2: Example EEG and ocular electrode montage. The participant wears a 32-channel EEG cap together with horizontal and vertical EOG electrodes, periocular EMG electrodes and their associated ground/reference electrodes around the eyes and face. This configuration provides high-quality ocular signals but increases setup time and can be uncomfortable or visually intrusive for users, especially outside controlled laboratory settings.

example in least-mean-squares frameworks.

Regression-based methods are simple, computationally inexpensive and suitable for online use. They perform well when the relationship between EOG and its projection onto EEG channels is approximately linear and stationary. However, they assume that the main artefact source is captured by the recorded EOG channels and that these channels are available. They are less effective when multiple sources of ocular and muscle activity interact, when electrode impedances change, or when movements are large and complex.

2.5.2 Independent component analysis and blind source separation

Independent component analysis (ICA) views EEG as a linear mixture of statistically independent sources. Under suitable assumptions, ICA can separate mixtures into components that reflect underlying neural and artefact sources [59], [81]. Ocular artefacts often appear as components with frontal scalp projections and stereotypical time courses linked to blinks and saccades.

A wide range of ICA algorithms and associated automatic component classification methods have been developed, including Infomax, FastICA, AMICA and variants specifically tuned to EEG [59], [81], [82], [90], [102]. Tools such as ADJUST and related methods exploit spatial, temporal and frequency features to detect and remove artefact components with minimal manual intervention [82], [90]. Complementary work has shown that computer vision can be used to automatically tag blink and other artefact segments in video recorded during BCI experiments, providing additional labels that can be aligned with ICA components and support semi-automatic rejection [127].

Recent work has evaluated ICA-based ocular artefact removal in more demanding free viewing conditions. Dimigen systematically examined how filtering choices, training data selection and component rejection thresholds affect the quality of ocular artefact correction when combined with eye-tracking, providing practical recommendations for optimising ICA pipelines [32]. These studies show that ICA can remove ocular artefacts effectively if configured carefully, but are sensitive to preprocessing decisions and require sufficient data for reliable decomposition.

ICA-based approaches do not strictly require EOG channels, although including them can help identify ocular components. However, reliable identification of artefact components often benefits from additional information, such as eye-tracking markers or knowledge of stimulus timing. VEOS proposes to add vision-based inferred EOG as another source of such information.

2.5.3 Spatial filtering and subspace methods

Spatial filtering methods such as regression in sensor space, signal-space projection (SSP), and spatio-spectral decomposition (SSD) aim to construct spatial patterns that emphasise neural activity while attenuating artefact subspaces [102], [137]. The multi-channel Wiener filter, for example, models mixed EEG as the sum of signal and artefact and uses estimates of their covariance to derive optimal linear filters for artefact suppression [137].

These methods are attractive because they can be formulated in a principled way and can be applied both in sensor and source space. They are particularly useful when artefact sources are relatively low-dimensional and can be characterised by their covariance structure. However, like ICA, they may struggle when artefacts and neural signals occupy overlapping spatial and spectral subspaces, or when artefacts are strongly non-stationary.

2.5.4 Artefact handling in mobile and movement-rich settings

Mobile EEG and experiments involving locomotion or more complex movement introduce additional artefact sources. A recent topical review on movement-induced artefacts summarises approaches to recording and cleaning EEG during walking and cycling and provides guidelines on method selection [42]. The review emphasises that ICA-based methods remain central, but that combinations of filtering, artefact subspace reconstruction, and careful experiment design are needed. It further notes that artefact characteristics depend strongly on movement intensity, so no single method is universally optimal.

These findings align with neuroergonomic work that highlights the importance of combining hardware solutions (for example active electrodes and improved headgear) with sophisticated signal processing to achieve acceptable signal quality during movement [10], [15], [42].

2.5.5 Deep learning and generative approaches to artefact removal

Deep learning-based denoising and artefact removal models have recently been proposed, using convolutional and recurrent architectures and, more recently, generative adversarial frameworks[2], [3], [19], [22], [157]. Many of these methods are trained on benchmark datasets such as EEGdenoiseNet and related synthetic or semi-synthetic corpora. A more detailed review of these approaches and their relation to VEOS is given in Section 2.8.3; here it suffices to note that they typically operate purely on EEG (sometimes with EOG) and do not yet exploit video as an explicit source of information about eye and eyelid motion.

2.6 Eye-Tracking, Gaze Estimation and Computer Vision

Inferring ocular motion from video requires a set of tools drawn from computer vision and eye-tracking research. These range from classical geometric methods for estimating gaze direction to modern deep learning approaches that learn the mapping from image appearance to eye state, and they form the technical foundation of the visual processing pipeline in VEOS.

2.6.1 Traditional eye-tracking technologies

Eye-tracking techniques are commonly divided into invasive and non-invasive approaches. Contemporary systems almost exclusively use non-invasive video-based eye-tracking, which can be either head-mounted or remote. Comprehensive treatments of eye-tracking methods and measurement properties are given by Holmqvist et al. and others [48], [55].

Classical video-based eye-trackers often employ infrared (IR) illumination and the detection of corneal reflections and pupil position relative to the camera. Calibration

procedures map pupil and corneal reflection positions to a point of gaze on the screen. This yields accurate gaze estimates under constrained head-pose and lighting, but performance degrades when head movement, glasses and varying illumination are introduced.

2.6.2 Appearance-based gaze estimation and deep learning

Appearance-based gaze estimation methods treat the mapping from image pixels to gaze direction as a learning problem rather than relying on explicit geometric models. Early work trained regressors on hand-crafted features; more recent work uses deep convolutional networks. A landmark contribution is the MPIIGaze work by Zhang et al., which demonstrated the feasibility of appearance-based gaze estimation “in the wild” using a large dataset of eye images collected from laptop webcams under varied conditions [160]. Subsequent work extended these ideas to larger and more diverse datasets and more complex network architectures.

Deep learning-based gaze estimation is now a substantial subfield. Surveys such as those by Pathirana et al. and Cheng et al. review appearance-based and model-based methods, datasets, coordinate systems and performance metrics, and benchmark a range of deep architectures for single- and multi-user gaze estimation in constrained and unconstrained environments [24], [110]. Recent architectures exploit multi-level feature fusion and real-time, lightweight designs, for example through combining global and local eye-region cues or tailoring networks for low-latency deployment [93], [130]. These methods routinely handle large head-pose variation, partial occlusion and complex illumination, conditions under which traditional IR-based systems struggle.

From the perspective of VEOS, these advances are significant because they show that eye-region appearance encodes enough information to recover fine-grained gaze direction under realistic conditions. The same information should in principle support inference of an EOG-like signal that approximates the projection of ocular potentials onto scalp electrodes.

2.6.3 Eye movement analysis for activity recognition and state estimation

Eye movements and gaze patterns have been used to recognise activities [20], detect fatigue [57], infer cognitive load [26] and study attention [46]. Bulling et al. demonstrated that patterns of fixations and saccades can be used to classify everyday activities using wearable eye-trackers [20]. Other work has fused EEG and gaze information to detect driver drowsiness or workload, reporting improvements over either modality alone [15], [57]. More recently, multimodal feature fusion schemes have been proposed for attention and vigilance classification, combining EEG with eye-tracking features such as fixation distributions and pupil or blink dynamics [147]. In immersive settings, joint EEG–eye-tracking analysis has been used to distinguish internal and external attention in virtual reality, again demonstrating the benefit of exploiting ocular behaviour alongside neural signals [79].

Beyond general activity recognition and cognitive-state monitoring, there is a substantial literature that uses ocular behaviour for vigilance and drowsiness estimation, particularly in driving and transport contexts. A long line of work has examined blink parameters such as duration, frequency and amplitude, as well as slow eye movements, as indicators of sleepiness [23], [37], [38], [62]. EOG-based systems have classified drowsy states using time-domain and wavelet-domain features with conventional and deep classifiers [63], [121], [159], while parallel efforts rely on video-based measures such as eyelid closure and pupil visibility [116], [118]. Lahssene’s bio-inspired pseudo-EOG work can be viewed as a point between these two strands, generating an EOG-like signal from video before passing derived features to a fuzzy-logic classifier of driver awareness [154]. VEOS differs in that it aims to reconstruct an EOG-like waveform with sufficient fidelity for downstream EEG artefact modelling, rather than primarily for state classification.

These results reinforce the view that ocular behaviour carries rich information about the user’s internal state. In VEOS, the inferred EOG signal is used for artefact removal and serves as a compact, interpretable representation of ocular dynamics,

which can be combined with EEG features and with direct computer-vision-based gaze estimates.

2.6.4 Computer vision frameworks for eye and face analysis

Modern computer vision toolkits such as OpenCV and MediaPipe provide high-quality implementations of face detection, landmark localisation and eye-region tracking [18], [49], [65], [141], [145], [146]. Face mesh models deliver dense landmarks around the eyes, eyelids and eyebrows at high frame rates even on modest hardware.

VEOS uses these frameworks to track eye-region geometry over time in the development dataset described in Chapter 4 and to process the video streams in the Eye-BCI dataset. By working directly with video, VEOS can be applied in settings where a dedicated eye-tracker is not available but a camera is, for example in many virtual reality and gaming systems.

2.7 Multimodal EEG–Eye Datasets and Experimental Frameworks

Progress in combining EEG with eye and video data depends critically on the availability of well-recorded, openly shared datasets that capture both modalities under ecologically relevant conditions. A number of such datasets and the experimental platforms that produced them have shaped the methodological choices made in this thesis.

2.7.1 Multimodal datasets in BCI research

Open datasets have played an important role in driving progress in EEG and BCI research. Early BCI competition datasets focused primarily on EEG signals recorded under well-controlled conditions using established paradigms. More recent efforts provide richer multimodal recordings that include behavioural data, eye-tracking, EMG

and other physiological signals, enabling more comprehensive modelling of user state and artefacts.

In safety and neuroergonomic contexts, multimodal datasets have been used to relate EEG markers to performance and workload [15], often combining EEG with behavioural measures, performance metrics and environmental context [10]. For fatigue and drowsiness monitoring, combining EEG with eye-tracking or facial video has been shown to improve detection performance compared with EEG alone [57].

2.7.2 AMBER: multimodal BCIs in naturalistic settings

The AMBER dataset was developed to support the study of BCIs in more naturalistic settings, combining EEG with behavioural and contextual information and full-face video in tasks designed to approximate real-world interaction [8]. It includes internally and externally valid measures of user engagement and task performance and captures a wide range of artefacts associated with more relaxed experimental control. In contrast to the datasets used later for supervised VEOS training, AMBER does not include dedicated EOG channels or infrared eye illumination; instead it offers EEG and video recordings under conditions that are closer to everyday use.

AMBER is particularly relevant as it embodies the design philosophy that informs VEOS: embrace, rather than avoid, the complexities of real-world interaction and design algorithms that can operate reliably under those conditions. In this thesis AMBER plays two roles. First, its task design and practical experience motivate the blended-control game and session structure adopted for the VEOS development dataset. Second, it provides an additional EEG+video corpus on which VEOS behaviour can be inspected qualitatively in settings where only EEG and camera data are available. However, because AMBER lacks explicit EOG channels, it is not used as a primary supervised training set and does not contribute directly to the quantitative evaluations in Chapter 7.

2.7.3 A multimodal data acquisition platform for blended control

As part of this thesis, a multimodal data-acquisition platform was built to capture synchronised EEG, EOG and high-speed video together with stimulus control and behavioural responses [129]. The platform supports several interaction modes, including a “follow the dots” gaze-following task, a game-based cognitive loading task with three difficulty levels, and RSVP/ERP paradigms akin to those later used in AMBER [8], [9]. Data were acquired with a 120 Hz camera and OpenViBE-based EEG control, enabling precise synchronisation of video, EEG and EOG.

These recordings form the VEOS development dataset (VEOS-Dev), which serves as the primary training and validation data for VEOS. They provide ground truth EOG signals, synchronised video and EEG, allowing supervised learning of mappings from video-derived features to EOG and the study of how well inferred EOG captures ocular artefacts in EEG.

2.7.4 Eye-BCI multimodal dataset combining EEG, eye-tracking and high-speed video

The Eye-BCI multimodal dataset, recently released as a Scientific Data descriptor, provides an extensive resource for ocular activity analysis in the context of several standard BCI paradigms [45]. It combines 62-channel EEG, high-speed eye-region video and eye-tracking across four paradigms: motor imagery, motor execution, SSVEP and P300 spellers. The dataset comprises over 46 hours of data from 31 participants across 63 sessions, with thousands of trials per paradigm.

Crucially, eye-tracking and high-speed video are tightly synchronised with EEG, and additional annotations capture blink events and stimulus timing. This makes Eye-BCI particularly well-suited for studying blink dynamics, ocular artefacts, and their interaction with neural signals across paradigms. It also provides an opportunity to test whether models trained on one dataset, such as the development platform

described above, can generalise to a larger and more diverse cohort.

In this thesis, Eye-BCI is used as an external testbed for VEOS. The VEOS inference pipeline, trained using the development dataset with explicit EOG channels, is applied to Eye-BCI video streams to generate inferred EOG signals. These inferred signals are then evaluated both as approximations to recorded ocular activity and as inputs to artefact mitigation procedures. The presence of multiple paradigms in Eye-BCI allows cross-paradigm robustness of VEOS to be assessed.

AMBER, the VEOS development dataset (VEOS-Dev) and Eye-BCI cover a range of experimental conditions. AMBER offers naturalistic EEG and full-face video without dedicated EOG channels; VEOS-Dev provides tightly controlled joint EEG-EOG-video recordings in a single laboratory using the purpose-built platform of Chapter 4; and Eye-BCI contributes a larger, independently collected sample with high-density EEG, vertical EOG, eyelid EMG and eye-tracking. The empirical work in later chapters uses VEOS-Dev and Eye-BCI for quantitative training and evaluation of VEOS, while AMBER mainly informs task design and serves as an example of the kind of EEG+video-only dataset to which VEOS is intended to generalise.

2.7.5 Other multimodal EEG-eye datasets

A number of other datasets combine EEG with eye-tracking or video, often focusing on specific tasks such as reading, scene viewing or attention tasks. Many of these studies use eye-tracking primarily to time-lock fixation-related potentials and to support ICA-based artefact removal [32], [55]. Others explore joint modelling of eye movements and EEG for cognitive state inference [20], [26]. While VEOS does not directly use these datasets, the methodological insights they provide about synchronisation, artefact characterisation and experimental design inform the overall approach.

2.8 Positioning VEOS within the Literature

Having reviewed the relevant background across BCIs, EEG artefact handling, ocular physiology and computer vision, it is now possible to identify where VEOS sits within this landscape, what distinguishes it from prior work, and which specific gaps in the literature it is designed to address.

2.8.1 Sensor-based versus vision-based ocular measurement

Most existing EEG artefact handling methods that consider ocular activity fall into two categories. Sensor-based approaches use EOG electrodes combined with EEG to detect and remove artefacts [27], [82], [90], [131], [132]. Vision-based approaches, such as eye-tracking, are generally used to support experimental control and interpretation rather than to drive artefact removal directly [32], [48], [55].

VEOS proposes to treat video as a proxy sensor for ocular activity. Instead of directly regressing EOG out of EEG, the system first infers an EOG-like signal from video, then uses that signal as if it had been recorded via electrodes. This has several implications: it can provide an “EOG channel” in systems where only video and EEG are available; it can be retrofitted to existing datasets that include video but no EOG, such as the AMBER dataset, enabling new analyses; and it allows direct comparison between sensor-based and vision-based EOG representations in datasets like the VEOS development dataset (VEOS-Dev), where both are available.

This approach builds on the success of deep gaze estimation and eye-region analysis in computer vision [57], [110], [160], on camera-based analyses that relate video measurements to EOG such as retina-modelling for blink characterisation and pseudo-EOG generation for driver monitoring [117], [154], and on the substantial literature treating EOG as a reliable measure of ocular dynamics [12], [36], [83].

2.8.2 Video-based generation of pseudo-EOG signals

The closest existing work to VEOS in spirit is the generation of pseudo-EOG signals from video for driver monitoring. Picot et al. modelled the human retina and applied this model to high-speed eye-region video in order to derive a “retina signal” whose blink-related dynamics could be compared directly with vertical EOG [117]. Their results showed strong correspondence between the two modalities for blink detection and parameterisation, and highlighted the potential of camera-based measurements as a surrogate for EOG in applications where physical electrodes are undesirable.

Building on this idea, Yahia Lahssene et al. proposed a bio-inspired pseudo-EOG signal extracted from monocular eye video and applied it to driver awareness monitoring [154]. Their pipeline constructs an image-based signal that mimics EOG in time, then derives a set of critical temporal and statistical features which are fed into a fuzzy-logic classifier to label drowsiness levels. The emphasis is on robust classification of awareness state rather than on reproducing the exact morphology or amplitude of recorded EOG across tasks and individuals. Evaluation focuses on classification performance and qualitative similarity between pseudo-EOG and true EOG traces in the driving context.

In contrast, VEOS pursues a different operating point. The goal is not primarily to classify vigilance, but to recover an EOG-like waveform that is accurate enough to act as a regressor for ocular artefact removal and to support quantitative analysis of ocular dynamics in EEG-based BCIs. This requires attention to signal-level metrics such as correlation with true EOG, band-limited spectral similarity and blink-amplitude fidelity, as well as downstream effects on ICA/LCA-based cleaning and BCI performance. VEOS also operates in paradigms beyond driving, including motor imagery, SSVEP and P300 spelling, and is evaluated on cross-dataset generalisation using the Eye-BCI dataset. In this sense, pseudo-EOG work in driver monitoring and VEOS are complementary: both demonstrate that video can underwrite an EOG-like representation, but they optimise that representation for different targets, namely state classification on the one hand and artefact-aware EEG analysis on the other.

2.8.3 Relation of VEOS to deep learning-based artefact removal

Deep learning has recently been applied to EEG artefact detection and removal, enabled by larger datasets and increases in computational power[22].

EEGdenoiseNet is a benchmark dataset-specifically designed for training and evaluating deep learning models for EEG denoising. It provides clean EEG segments and isolated ocular and muscular artefact segments, enabling the synthesis of contaminated signals with known ground truth [157]. Using this dataset, several architectures, including convolutional and recurrent networks, have demonstrated promising performance in separating artefacts from neural activity[27], [32], [131], [144].

Generative adversarial networks (GANs) and related models have been explored for EEG denoising in real-world BCI applications. Early work showed that deep generative models can learn to map noisy EEG to cleaner representations in realistic noisy environments [19], while subsequent studies introduced variants tailored to EEG, such as auto-denoising frameworks and Wasserstein GANs that stabilise training and improve reconstruction quality [3], [34]. Task-oriented GAN architectures have also been proposed, in which the denoising process is explicitly optimised to enhance downstream BCI performance, for example for SSVEP-based BCIs [155]. Most recently, the AnEEG framework (“*AnEEG: leveraging deep learning for effective artifact removal in EEG data*”), proposes a deep learning-based artefact removal method that outperforms traditional wavelet-based denoising across several metrics, including signal-to-noise ratio, normalised mean squared error, root mean squared error, correlation coefficient and signal-to-artefact ratio [66].

Reviews of deep learning for EEG processing and artefact removal highlight both the potential and the limitations of these methods [22].

VEOS can be viewed as complementary to these methods. An inferred EOG signal obtained from video could be used as an additional input to a deep artefact removal model, or as a reference in more classical regression approaches. The present work focuses on the first step of producing that inferred signal.

2.8.4 Research gaps addressed by VEOS

The literature surveyed in this chapter reveals several gaps that VEOS aims to address:

- **Dependence on dedicated EOG sensors.** Many artefact removal methods assume that EOG channels are present. When they are not, performance degrades or methods cannot be applied [27], [32], [131], [144]. VEOS explores whether an EOG-like signal can be reconstructed from video.
- **Limited use of video in artefact modelling.** Although EEG combined with eye-tracking is increasingly used for artefact-informed analysis in free viewing experiments, video is rarely used as a primary artefact reference [32], [55]. VEOS explicitly models the mapping from video to EOG.
- **Cross-paradigm and cross-dataset robustness.** Artefact characteristics vary across paradigms and datasets. Eye-BCI and the VEOS development dataset (VEOS-Dev) together provide a platform for testing whether a learned video-to-EOG mapping generalises across tasks and recording conditions [8], [9], [45], [129].
- **Integration of ocular behaviour into BCI pipelines.** Treating ocular signals purely as noise discards useful information about gaze and state. VEOS supports a view of ocular activity as a shared resource: a signal that can be suppressed when it masks neural activity yet still used for behaviour and state estimation.

The following chapter develops the theoretical framework underlying VEOS in more formal detail. It provides mathematical descriptions of EEG, EOG and their mixing, introduces the independent component analysis and regression tools used later, and formalises the mapping from video-derived features to inferred EOG signals.

2.9 Summary

This chapter has reviewed the key scientific and technological background for VEOS. BCIs have matured considerably, but robust operation in real-world settings remains challenging, with ocular artefacts playing a central role. Traditional artefact removal methods rely heavily on recorded EOG, while deep learning approaches have begun to learn more flexible non-linear mappings but still generally require sensor-based measures of ocular activity. At the same time, computer vision has delivered powerful methods for eye-tracking and gaze estimation from video, and multimodal datasets now provide rich combinations of EEG, EOG, eye-tracking and high-speed video across multiple paradigms.

VEOS uses computer vision to infer EOG-like signals from video, supports subsequent artefact handling and behavioural analysis, and is evaluated using both a bespoke multimodal development platform and an external multimodal dataset. In doing so, it addresses a limitation of current EEG-based BCIs by linking both sensor-based and vision-based views of ocular activity.

Chapter 3

Theoretical Foundations

3.1 Overview

VEOS links three complementary descriptions of eye behaviour.

1. the three-dimensional motion of the eyeballs and eyelids.
2. the two-dimensional description that appears in video as moving landmarks and textures.
3. the essentially one-dimensional electrical signal observed in EOG channels and leaking into scalp EEG recordings.

This chapter develops models for each description and establishes the links between them.

The chapter begins with the camera geometry used in VEOS and a simple kinematic description of eyeball rotation. It then considers how eye movements evolve over time and how this temporal structure is expressed in video and in electrophysiological signals. The next part develops a dipole-based model of the electrooculogram and explains how ocular activity reaches frontal EEG channels. Independent component analysis (ICA) is then described in sufficient detail to support later chapters. The final sections formalise the mapping from video features to EOG and discuss the timing constraints that arise when several hardware streams are combined.

3.2 Camera geometry for eye-tracking

Eye-tracking in VEOS relies on standard ideas from projective geometry [49], [141].

A point in three-dimensional space is represented in homogeneous coordinates as

$$\mathbf{X} = \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \quad (3.1)$$

and its image on the camera sensor as

$$\mathbf{x} = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}, \quad (3.2)$$

where u and v are pixel coordinates.

In the pinhole model, the relationship between $\mathbf{X}$ and $\mathbf{x}$ is

$$s\mathbf{x} = K[R \mid \mathbf{t}] \mathbf{X}, \quad (3.3)$$

with scale factor $s > 0$, rotation $R \in \mathbb{R}^{3 \times 3}$ and translation $\mathbf{t} \in \mathbb{R}^3$. The matrix K contains the intrinsic calibration parameters,

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (3.4)$$

where f_x and f_y are effective focal lengths in pixel units and (c_x, c_y) is the principal point.

In VEOS it is convenient to define a head-centred frame in which the origin lies near the bridge of the nose or the midpoint between the two eyes. The camera frame has its origin at the effective pinhole and its Z axis aligned with the optical axis. In

that case $[R \mid \mathbf{t}]$ encodes the rigid transform from head space to camera space. Over intervals where head motion is small, this transform can be treated as constant.

The mapping (3.3) can be decomposed into a projective mapping and a change of coordinates. Writing

$$\mathbf{X}_{\text{cam}} = \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = [R \mid \mathbf{t}]\mathbf{X}, \quad (3.5)$$

the normalised image coordinates are

$$x_n = \frac{X_c}{Z_c}, \quad y_n = \frac{Y_c}{Z_c}, \quad (3.6)$$

and pixel coordinates are

$$u = f_x x_n + c_x, \quad v = f_y y_n + c_y. \quad (3.7)$$

Small changes in (X_c, Y_c, Z_c) cause approximately linear changes in (u, v) when the point remains near the optical axis [141].

In typical eye-tracking setups the field of view is moderate and the eye-region occupies only a small part of the frame near the centre. Even so, lens distortion can introduce measurable deviations from the ideal model [161]. Radial distortion is represented by

$$r^2 = x_n^2 + y_n^2, \quad (3.8)$$

and undistorted coordinates are mapped to distorted ones by

$$x_d = x_n (1 + k_1 r^2 + k_2 r^4 + k_3 r^6), \quad (3.9)$$

$$y_d = y_n (1 + k_1 r^2 + k_2 r^4 + k_3 r^6). \quad (3.10)$$

Distortion can be corrected explicitly during calibration, but in VEOS the landmark detector is trained directly on distorted imagery, so the model is typically used for

interpretation rather than for online correction.

The landmark detector used in VEOS provides a set of points $\{\mathbf{x}_k^{(i)}\}$ at each frame k . These landmarks include eyelid corners, points along the lid margins and points on the iris. In the current implementation they are obtained from the MediaPipe Face Mesh model [67], [146]. It is often useful to transform these landmarks into a face-centred coordinate system to reduce variability. One simple strategy is to compute a similarity transform that maps a subset of facial landmarks (for example, the eye corners and the tip of the nose) to reference positions, and then apply the same transform to all landmarks. This removes global translation, rotation and scale of the head in the image plane, so that the remaining motion reflects primarily eye rotation and eyelid motion.

Let S_k denote the similarity transform at frame k , written as

$$\tilde{\mathbf{x}}_k^{(i)} = S_k \mathbf{x}_k^{(i)}. \quad (3.11)$$

The transformed landmarks $\tilde{\mathbf{x}}_k^{(i)}$ are then used as the basis for VEOS features. Errors in head-pose estimation and imperfect alignment of the similarity transform introduce residual head motion into these signals. These effects are considered later when discussing noise.

A schematic of the camera, head and eye geometry is shown in Figure 3.1. The figure highlights how the same eyeball rotation leads to different landmark patterns when the head pose changes.

3.3 Modelling eyeball kinematics

The eyeball is often modelled as a sphere of radius r_{eye} that rotates within the orbit [48], [55]. Muscle insertions and connective tissue mean that the true motion is more complex, but the spherical model captures the main relationship between gaze direction and image appearance.

Let the optical axis in the reference pose point along unit vector $\mathbf{e}_z$ in the eye-

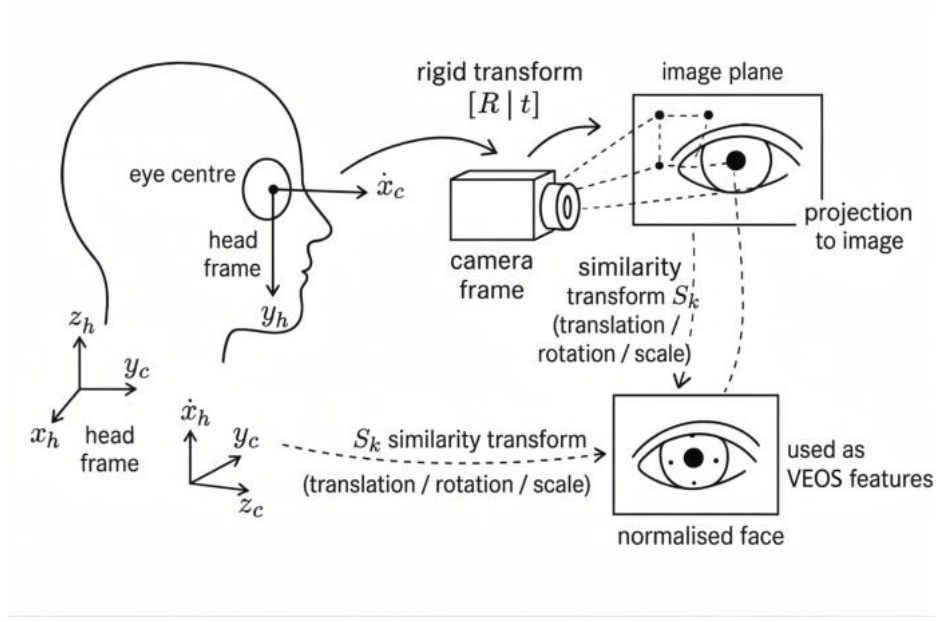


Figure 3.1: A head-centred coordinate frame is mapped to the camera frame by a rigid transform $[R | t]$, while the eyeball rotates within the head frame. Landmarks on the eyelids and iris are projected onto the camera image plane, and a face-normalisation stage applies a similarity transform to the detected landmarks to reduce the effect of head motion before they are used as VEOS features.

centred frame. Horizontal and vertical gaze angles at time t are denoted by $\theta_h(t)$ and $\theta_v(t)$. A simple parametrisation of the rotation uses two elemental rotations,

$$R_{\text{eye}}(t) = R_y(\theta_h(t))R_x(\theta_v(t)), \quad (3.12)$$

where

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix}, \quad (3.13)$$

and

$$R_y(\theta) = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix}. \quad (3.14)$$

This model neglects torsional rotations around the optical axis, which are usually small in everyday viewing [55].

Let $\mathbf{X}_0$ be the vector from the eye centre to the pupil centre in the reference pose.

At time t , the pupil centre is located at

$$\mathbf{X}_{\text{pupil}}(t) = R_{\text{eye}}(t)\mathbf{X}_0. \quad (3.15)$$

Projection using the camera model in Section 3.2 then gives the expected image coordinates of the pupil.

For small departures from the primary gaze direction the mapping from (θ_h, θ_v) to pixel coordinates can be linearised. Writing

$$\boldsymbol{\theta}(t) = \begin{bmatrix} \theta_h(t) \\ \theta_v(t) \end{bmatrix}, \quad (3.16)$$

and letting $\mathbf{x}_{\text{pupil}}(t) = (u(t), v(t))^\top$ denote the corresponding pixel coordinate, the first-order Taylor expansion around a reference angle $\boldsymbol{\theta}_0$ is

$$\mathbf{x}_{\text{pupil}}(t) \approx \mathbf{x}_0 + J_{\boldsymbol{\theta}}(\boldsymbol{\theta}(t) - \boldsymbol{\theta}_0), \quad (3.17)$$

where

$$J_{\boldsymbol{\theta}} = \left. \frac{\partial \mathbf{x}_{\text{pupil}}}{\partial \boldsymbol{\theta}^\top} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \quad (3.18)$$

is the Jacobian. The entries of $J_{\boldsymbol{\theta}}$ depend on the camera intrinsics and the geometry of the eyeball within the orbit [55]. For a fixed setup, $J_{\boldsymbol{\theta}}$ can be treated as constant. Equation (3.17) explains why small changes in gaze direction cause almost linear changes in image coordinates, and why a simple regression from landmark displacements to angular displacements is often effective over moderate ranges [20], [160].

The eyeball kinematic model also helps when reasoning about occlusion. As the eye rotates upwards, the upper eyelid covers more of the iris. The portion of the iris that remains visible depends on both the rotation and the eyelid motion [55]. Landmark detectors that include points along the upper lid capture this interaction. During a blink, the upper lid moves rapidly downwards, temporarily occluding the pupil, and

then returns. The kinematic link between lid motion and eyeball rotation produces characteristic waveforms in both video and EOG, which VEOS can exploit [48], [55].

One could extend the model to include Listing’s law, which states that eye rotations during normal viewing can be described as rotations about axes that lie in a single plane in rotation space [55]. For the purposes of VEOS it is sufficient to note that eye movement kinematics are constrained and that not all paths in angle space are equally likely. The learning algorithms then see samples drawn from this constrained space.

3.4 Temporal dynamics of eye movements

Eye movements are not isolated events: they form structured sequences in time [55]. A typical viewing pattern consists of alternating fixations and saccades, with occasional blinks. In driving or vigilance tasks, smooth pursuit movements may also appear [20], [55]. The temporal dynamics influence both the frequency content and the statistical structure of the signals that VEOS processes.

Let $\mathbf{x}_k$ denote a vector of landmarks at video frame k . Differences and derivatives of $\mathbf{x}_k$ capture temporal information. A simple first difference

$$\Delta \mathbf{x}_k = \mathbf{x}_k - \mathbf{x}_{k-1} \tag{3.19}$$

acts as a high-pass filter. Larger magnitude differences indicate rapid motion, as in saccades or blinks. Second differences emphasise curvature in the trajectory,

$$\Delta^2 \mathbf{x}_k = \mathbf{x}_k - 2\mathbf{x}_{k-1} + \mathbf{x}_{k-2}. \tag{3.20}$$

These operators are useful for constructing features that respond selectively to sudden movements.

In discrete time, differences and finite impulse response filters can be described by convolution. For example, a three-point moving average filter applied to a scalar

sequence $x[k]$ is

$$\tilde{x}[k] = \frac{1}{3}x[k-1] + \frac{1}{3}x[k] + \frac{1}{3}x[k+1], \quad (3.21)$$

with impulse response $h[0] = h[1] = h[2] = 1/3$. The frequency response $H(e^{j\omega})$ reveals how the filter attenuates high-frequency components. Similar reasoning applies to EOG and EEG filters [22], [144].

Eye movements generate characteristic frequency content. Saccades produce broad-band energy with a peak that depends on the distribution of saccade durations, while blinks produce strong low-frequency energy and a substantial near-direct-current component in EOG [27], [36], [144]. Fixations contribute low-level, low-frequency noise. These patterns are reflected in the video features. Vertical eyelid distance, for example, behaves like a low-frequency envelope with sharp transitions at blink onset and offset.

The sampling structure of VEOS is constrained by the frame rate of the camera. If the camera operates at f_{cam} frames per second and the EOG is sampled at f_{eog} , then the maximum temporal resolution achievable by VEOS in the video domain is one frame. If blink rise times are of the order of tens of milliseconds and the frame interval is similar, only a few samples capture the transition [55]. This is still sufficient to detect blinks and to estimate their timing to within a frame, but peak values may be underestimated compared with a high-rate EOG.

It is sometimes useful to think in terms of a generative model where an underlying ocular state trajectory $\theta(t)$ drives both video and EOG through different observation models [20], [147]. In that view, VEOS attempts to invert one observation model using the other, under sampling and noise constraints.

3.5 Landmark noise and uncertainty

Landmark detection does not produce exact coordinates; it is a statistical inference process operating on noisy images [48], [55], [110]. Errors in landmark locations propagate through the VEOS pipeline and impose a limit on the accuracy of any

inferred EOG.

The observed landmark position at frame k can be written as

$$\mathbf{x}_k = \mathbf{x}_k^* + \boldsymbol{\epsilon}_k, \quad (3.22)$$

where $\mathbf{x}_k^*$ is the true projection of the anatomical point and $\boldsymbol{\epsilon}_k$ is the error. A simple model assumes that $\boldsymbol{\epsilon}_k$ is zero mean with covariance Σ_ϵ . Under that assumption,

$$\mathbb{E}[\mathbf{x}_k] = \mathbf{x}_k^*, \quad \text{Cov}[\mathbf{x}_k] = \Sigma_\epsilon. \quad (3.23)$$

In reality, Σ_ϵ depends on head-pose, distance, illumination and eye state. For example, landmarks near the eyelids are harder to detect accurately when the eyelids are partially closed, and iris landmarks can become unreliable at extreme gaze angles [55], [110].

If VEOS constructs features that involve differences, the error covariance changes. For a first difference,

$$\Delta \mathbf{x}_k = \mathbf{x}_k - \mathbf{x}_{k-1} = (\mathbf{x}_k^* - \mathbf{x}_{k-1}^*) + (\boldsymbol{\epsilon}_k - \boldsymbol{\epsilon}_{k-1}), \quad (3.24)$$

and if the errors in successive frames are independent and identically distributed with covariance Σ_ϵ , then

$$\text{Cov}[\Delta \mathbf{x}_k] = 2\Sigma_\epsilon. \quad (3.25)$$

The signal of interest appears in the first bracket, while the noise variance doubles. This illustrates why some temporal smoothing is necessary before taking differences [55].

Not all errors are random. Systematic biases occur when the landmarks track visually salient but anatomically imperfect features. For instance, eyeliner or glasses frames can shift the apparent position of the eyelid [55]. These biases are harder to model mathematically. From the point of view of VEOS, they appear as structured offsets that the mapping from video to EOG must absorb.

Gross outliers are another concern. A single misdetected frame can produce a large spike in a feature sequence. Robust statistics offer ways to limit their influence. For example, instead of computing sample means and covariances with squared error loss, one can use Huber losses or median-based estimators that are less sensitive to extreme values. In practice VEOS uses outlier rejection rules on landmark positions and discards frames where the face mesh is clearly invalid.

3.6 EEG signal characteristics

Electroencephalographic signals are recordings of potential differences between scalp electrodes and a reference [2], [22]. They reflect the combined activity of many neural populations alongside various non-neural sources. For the purposes of VEOS, the most important properties of EEG are its amplitude range, spectral structure and statistical dependence over time.

Let $x_m[n]$ denote the discrete-time EEG signal at channel m sampled at frequency f_s . A common starting point is to model $x_m[n]$ as a realisation of a zero-mean, wide-sense stationary random process with autocovariance

$$R_{x_m}[\ell] = \mathbb{E}[x_m[n]x_m[n + \ell]], \quad (3.26)$$

and cross-covariance between channels m and k ,

$$R_{x_m x_k}[\ell] = \mathbb{E}[x_m[n]x_k[n + \ell]]. \quad (3.27)$$

True EEG is only approximately stationary over short windows, but many processing techniques rely on this approximation [2], [22].

Typical EEG amplitudes at the scalp lie in the tens of microvolts. This is small compared with artefacts produced by eye movements and muscle activity, which can easily reach hundreds of microvolts [36], [132], [144]. The signal-to-noise ratio of neural activity relative to ocular artefacts is therefore often poor at frontal sites. Any

additional information that helps isolate the ocular contribution is valuable [137], [144].

From a spectral point of view, EEG power is concentrated in a band from roughly 0.5 Hz to a few tens of hertz, although higher-frequency activity from muscle sources and some brain rhythms can extend further [2], [22]. It is customary to describe frequency content in terms of named bands. While the exact boundaries vary across studies, a typical scheme uses

$$\text{delta: } 0.5 \text{ to } 4 \text{ Hz,} \quad (3.28)$$

$$\text{theta: } 4 \text{ to } 8 \text{ Hz,} \quad (3.29)$$

$$\text{alpha: } 8 \text{ to } 13 \text{ Hz,} \quad (3.30)$$

$$\text{beta: } 13 \text{ to } 30 \text{ Hz,} \quad (3.31)$$

$$\text{gamma: } 30 \text{ Hz and above,} \quad (3.32)$$

consistent with standard descriptions of oscillatory activity [46], [52], [115].

Many cognitive and clinical paradigms focus on changes in power or phase relationships within particular bands [46], [52], [120]. Ocular artefacts are problematic because they contribute power in exactly the same low-frequency ranges where interesting brain activity appears, particularly in the delta and theta bands [36], [144].

The spatial structure of EEG is described by the covariance matrix of channel signals. At a fixed time, the vector of potentials $\mathbf{x}[n]$ across channels has covariance

$$C_x = \mathbb{E}[\mathbf{x}[n]\mathbf{x}[n]^\top]. \quad (3.33)$$

This matrix encodes correlations between electrodes. Neural sources with distributed spatial projections and artefacts with concentrated projections both leave distinct signatures in C_x . Independent component analysis and related blind source separation methods use this structure, together with higher-order statistics, to separate sources [59], [81], [90], [102].

Referencing schemes also influence EEG characteristics. Potentials are always measured relative to a reference electrode or composite reference. Common referencing strategies include linked mastoids, average reference and reference to a single vertex electrode. Changing reference alters the apparent amplitude and covariance structure of channels, but not the underlying physics [22], [144]. VEOS is agnostic to the particular reference used, as long as the same scheme is applied consistently to EOG and EEG so that relationships between channels remain interpretable.

In summary, EEG can be thought of as a multivariate, band-limited, noisy signal with low-amplitude neural components of interest and higher-amplitude artefacts of non-neural origin [2], [22], [144]. The next sections describe in more detail how eye movements and eyelid muscles contribute to this mixture.

3.7 Electrooculogram modelling

The electrooculogram is a surface measurement of an underlying electric field generated by the eye. A simplified but useful description treats the eye as an electric dipole of constant magnitude whose orientation follows the optical axis [27], [36], [144].

Let $\mathbf{p}(t)$ be the dipole moment at time t . It is written as

$$\mathbf{p}(t) = p_0 \mathbf{q}(t), \tag{3.34}$$

where p_0 is a positive constant and $\mathbf{q}(t)$ is a unit vector. In the eye-centred frame, $\mathbf{q}(t)$ points from the retina towards the cornea. When the eye rotates, $\mathbf{q}(t)$ rotates with it, and is therefore a function of the gaze angles.

Electrodes placed near the eye measure potential differences that depend on the projection of $\mathbf{p}(t)$ through the inhomogeneous volume conductor of the head. Analytic expressions for this projection in realistic head models are complicated. For VEOS the relevant fact is that vertical and horizontal EOG channels can be approximated

as linear functionals of $\mathbf{q}(t)$,

$$V_{\text{vert}}(t) \approx \mathbf{k}_v^\top \mathbf{q}(t) + \eta_v(t), \quad (3.35)$$

$$V_{\text{horiz}}(t) \approx \mathbf{k}_h^\top \mathbf{q}(t) + \eta_h(t), \quad (3.36)$$

where $\mathbf{k}_v$ and $\mathbf{k}_h$ are constant vectors determined by electrode placement and tissue properties, and $\eta_v(t), \eta_h(t)$ represent noise and other sources [27], [36].

If the coordinate axes of the eye-centred frame are aligned so that the z axis is close to primary gaze, then for small rotations the unit vector $\mathbf{q}(t)$ can be approximated as

$$\mathbf{q}(t) \approx \begin{bmatrix} \theta_h(t) \\ \theta_v(t) \\ 1 \end{bmatrix}, \quad (3.37)$$

and the vertical and horizontal EOG channels become approximately linear in the gaze angles. This is consistent with empirical findings that EOG amplitudes scale with the angle of rotation over moderate ranges [12], [36], [83].

The frequency content of EOG is shaped by the same temporal dynamics discussed earlier. Slow drifts in gaze cause very low-frequency components and baseline wander. Saccades produce sharp deflections with energy in a mid-range of frequencies. Blinks produce large, sharp waveforms with strong low-frequency content and shifts in the apparent baseline. Any method that attempts to infer an EOG-like signal from video must reproduce these characteristics to be useful in practice [36], [45], [144].

3.8 Propagation of ocular sources to scalp EEG

EOG electrodes are placed close to the eyes, so they measure the ocular dipole through a relatively direct pathway. Scalp EEG electrodes record the combined effect of many neural and non-neural sources. Ocular activity is one source among many, but it can dominate frontal channels because of its proximity and magnitude [36], [95], [144].

A full treatment of the forward problem in EEG uses a leadfield matrix L that

maps a set of current sources in the brain and eyes to potentials at electrodes [88]. In a simplified view,

$$\mathbf{v}(t) = L_{\text{brain}}\mathbf{j}_{\text{brain}}(t) + L_{\text{eye}}\mathbf{p}(t) + \boldsymbol{\eta}(t), \quad (3.38)$$

where $\mathbf{v}(t)$ is the vector of EEG channel voltages, $\mathbf{j}_{\text{brain}}(t)$ summarises neural sources and L_{eye} encodes the sensitivity of each electrode to the ocular dipole $\mathbf{p}(t)$. The noise term $\boldsymbol{\eta}(t)$ covers instrumental and environmental contributions [22], [124].

For frontal channels, rows of L_{eye} have relatively large magnitudes, because the electrodes are close to the source. For occipital channels the coupling is weaker. The spatial pattern of ocular activity in EEG therefore shows a gradient from large deflections at frontal sites to smaller deflections further back [36], [95].

From the perspective of artefact handling, this structure is useful. Components derived from ICA that correspond to ocular artefacts have scalp maps that match the expected pattern from L_{eye} . They also correlate strongly with EOG [59], [81], [90], [102], [137], [144]. If VEOS produces an inferred EOG that tracks $\mathbf{p}(t)$ closely, it can serve as an auxiliary observation of the ocular source and help to identify those components. This propagation pathway is illustrated schematically in Figure 3.2,

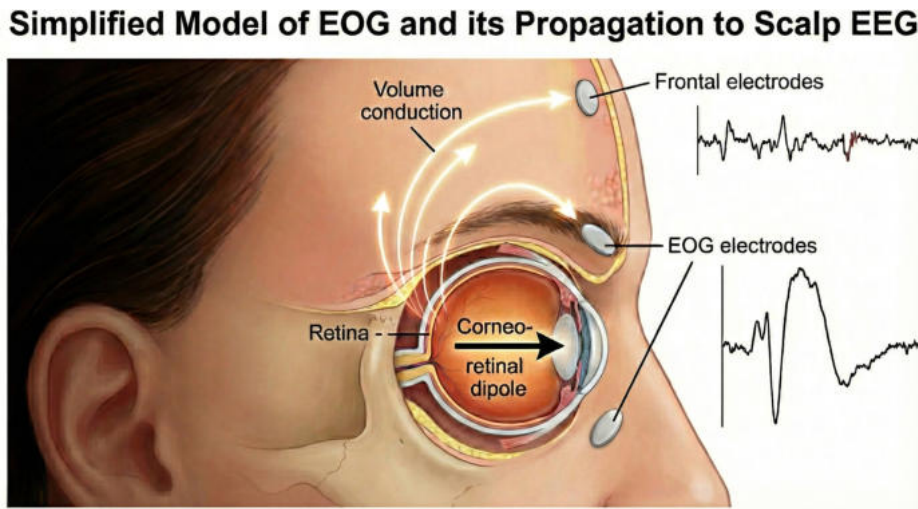


Figure 3.2: Simplified model of the electrooculogram and its propagation to scalp EEG. The corneo-retinal potential is modelled as an eye-centred dipole, which gives rise to large voltages at nearby EOG electrodes and smaller, but related, artefacts at frontal EEG electrodes via volume conduction.

which shows the ocular dipole, the nearby EOG electrodes and the more distal frontal

EEG electrodes through which the field spreads via volume conduction. It is important that VEOS does not attempt to model brain sources. The role of the inferred EOG is to track a non-neural source that is already considered an artefact in many EEG applications. The theoretical burden is therefore to model ocular potentials and their leakage into EEG with enough fidelity to support artefact handling, rather than to build a complete forward model of the head.

3.9 Ocular artefacts in EEG

Ocular artefacts are the contributions to scalp EEG that arise from movements of the eyes and eyelids, rather than from cortical activity. They originate primarily from the electrooculogram and from muscle activity in the periocular region [36], [144].

In the time domain, ocular artefacts have characteristic waveforms. Blinks produce large, relatively slow deflections in frontal channels. The waveform often shows a sharp rise as the eyelids close, a plateau when the eye is fully closed and a slower return during reopening. The amplitude can exceed the underlying EEG by an order of magnitude [36], [45]. Saccades produce shorter, often smaller deflections whose sign and size depend on the direction and amplitude of the eye movement. Slow drifts in eye position lead to baseline shifts that persist until gaze is reset. These characteristic waveform shapes are illustrated in Figure 3.3, which shows idealised examples of each artefact type as they would appear in a frontal EEG or vertical EOG channel. These artefacts are strongly correlated across frontal channels. During a blink, several frontopolar and frontocentral electrodes show similar waveforms with slightly different amplitudes. More posterior electrodes exhibit reduced but non-zero artefact, reflecting the spatial spread of the ocular field through the head. This spatial pattern is one of the cues that ICA and other spatial filtering techniques exploit [102], [137].

In the frequency domain, ocular artefacts contribute most power at low frequencies. A large blink introduces a prominent peak near zero frequency when viewed over short windows, because it resembles a step or ramp in the signal. Repeated

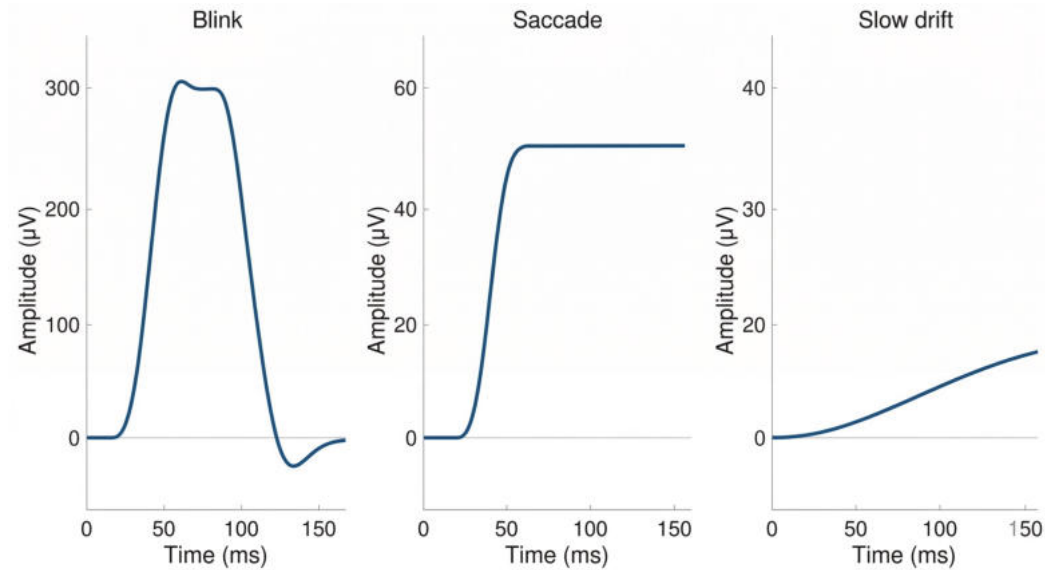


Figure 3.3: Characteristic time-domain waveforms of the three main classes of ocular artefact as they appear in a frontal EEG or vertical EOG channel. A blink (left) produces a large, sharp-rising deflection that plateaus briefly at full eyelid closure before returning more slowly to baseline, with peak amplitudes commonly in the range 100–500 μV . A saccade (centre) produces a shorter, step-like deflection whose polarity depends on the direction of the eye movement and whose amplitude scales with the angular displacement. A slow drift (right) produces a gradual baseline shift that accumulates over seconds until gaze is reset. These waveform shapes motivate the use of temporal context in the VEOS inference model.

blinks at irregular intervals produce broadband low-frequency activity. Saccades contribute power at slightly higher frequencies, related to their typical durations. If blinks or saccades occur frequently, the low-frequency power in frontal channels can become dominated by ocular activity, complicating the interpretation of slow neural rhythms [36], [144].

Event-related potential and time–frequency analyses are particularly sensitive to ocular artefacts. If blinks or saccades are systematically time-locked to stimuli or responses, the resulting potentials can contaminate averaged waveforms. Even when not time-locked, they increase variance and reduce the statistical power of comparisons between conditions. For frequency analyses, slow eye movements contribute to apparent changes in delta and theta power that have nothing to do with cortical processes [120], [144].

The relationship between the EOG and the ocular component of EEG is not one-to-one. The EOG is a more direct measure of the ocular dipole along specific axes,

whereas each scalp channel sees a mixture of ocular and neural contributions mediated by the head volume conductor and referencing scheme. Nevertheless, there is a strong correlation between EOG and frontal EEG artefacts during eye movements [27], [131], [132]. VEOS uses that correlation in reverse: it infers a synthetic EOG from video and uses this as a proxy for the ocular contribution that appears in EEG.

When interpreting the effect of VEOS on EEG, it is important to keep in mind that removing or regressing out ocular components may also remove some fraction of neural activity that happens to be spatially or temporally correlated with eye movements. This is not specific to VEOS; it is a general limitation of artefact handling that relies on linear projections and statistical dependence [137], [144]. The goal is therefore to provide the best possible estimate of the ocular contribution, so that any cleaning procedure has a clear and interpretable target.

3.10 Eyelid EMG and periocular muscle activity

Electromyographic signals measured near the eyes provide information about muscle activity in the eyelids and surrounding facial regions [36]. In many datasets, including those that record periocular channels alongside EEG, surface electrodes placed over the orbicularis oculi or nearby sites capture bursts of activity associated with blinks, squints and other eyelid movements. These EMG signals complement EOG by highlighting muscular rather than purely dipolar electrical contributions.

The EMG associated with eyelid closure arises mainly from contraction of the orbicularis oculi muscle. This muscle encircles the eye and is responsible for closing the eyelids. During a blink, it activates briefly and strongly, producing a high-frequency burst in the EMG signal superimposed on any slower EOG deflection. Partial blinks, deliberate squeezes and sustained squints have different activation patterns, which can be distinguished in EMG [36].

From a signal-processing perspective, EMG has a broader and higher frequency content than EOG. Power spectra of EMG from facial muscles often show significant energy from tens of hertz up to a few hundred hertz, depending on the recording

bandwidth. In contrast, EOG power is concentrated at low frequencies. This separation allows filtering strategies that retain EOG while attenuating EMG, or vice versa, depending on the application. It also means that EMG can contaminate higher-frequency EEG bands, including beta and gamma ranges, in ways that EOG does not [36], [144].

In scalp EEG, periocular EMG contributes to fast transients and increased high-frequency noise in frontal channels. These contributions are less spatially smooth than those from the ocular dipole, because muscle activity is more localised. They can be mistaken for high-frequency neural activity if not recognised. Independent component analysis often separates EMG-dominated components from both neural and ocular components by virtue of their spatial patterns and spectral properties [82], [90], [144].

For VEOS, periocular EMG channels serve two roles. First, they provide an additional source of ground truth for eyelid events in datasets that include them. Bursts of eyelid EMG confirm the timing of blinks and can be used to construct labels for training and evaluation of video-based blink detection [45]. Second, they highlight the difference between what is visible in the video and what is recorded electrically. The camera sees eyelid movement directly but does not observe muscle activation per se. Surface EMG records the electrical activity of the muscle, which may begin slightly before visible motion.

In a theoretical sense, one can extend the source model of EEG to include both EOG-driven dipole sources and EMG sources. Let $\mathbf{s}(t)$ now contain three groups of components: neural activity, ocular dipole activity and muscle activity. The leadfield matrix L can be partitioned accordingly:

$$\mathbf{x}(t) = L_{\text{brain}}\mathbf{s}_{\text{brain}}(t) + L_{\text{eye}}\mathbf{s}_{\text{eye}}(t) + L_{\text{emg}}\mathbf{s}_{\text{emg}}(t) + \boldsymbol{\epsilon}(t). \quad (3.39)$$

Periocular EMG channels provide a more direct view of $\mathbf{s}_{\text{emg}}(t)$, whereas EOG channels view $\mathbf{s}_{\text{eye}}(t)$, and scalp EEG channels see a mixture of all three. VEOS focuses primarily on $\mathbf{s}_{\text{eye}}(t)$, but the presence of EMG activity near the eyes affects both the electrical measurements and the video appearance during forceful blinks or squints.

When using a public dataset that includes eyelid EMG, VEOS can use that channel as an additional reference for blink timing and for distinguishing blinks from other facial actions. The theoretical treatment in this chapter clarifies how these signals relate and why it is helpful to consider both EOG and EMG when interpreting the performance of a vision-based inference model [45].

3.11 Independent component analysis for EEG

Independent component analysis assumes that the observed multi-channel EEG signal $\mathbf{x}(t) \in \mathbb{R}^M$ arises from a linear mixing of latent sources $\mathbf{s}(t) \in \mathbb{R}^N$ through a matrix A ,

$$\mathbf{x}(t) = A\mathbf{s}(t). \quad (3.40)$$

The aim is to estimate an unmixing matrix W such that

$$\hat{\mathbf{s}}(t) = W\mathbf{x}(t) \quad (3.41)$$

has statistically independent components [59].

The rationale behind ICA is linked to the central limit theorem. A linear combination of independent non-Gaussian sources tends to be more Gaussian than the individual sources. If the goal is to recover sources that are as non-Gaussian as possible, one searches for a transform that maximises some measure of non-Gaussianity [59].

Many algorithms start by whitening the data. Let $\mathbf{x}(t)$ be a zero-mean signal and $C_x = \mathbb{E}[\mathbf{x}(t)\mathbf{x}(t)^\top]$ its covariance. A whitening matrix V is found such that

$$\mathbf{z}(t) = V\mathbf{x}(t) \quad (3.42)$$

has identity covariance. One choice uses the eigenvalue decomposition

$$C_x = E\Lambda E^\top, \quad (3.43)$$

and defines

$$V = \Lambda^{-1/2} E^\top. \quad (3.44)$$

After whitening, the mixing process becomes an orthogonal transform, so the remaining task is to find a rotation that yields independent components [59].

Fixed-point algorithms such as FastICA operate on one component at a time or on several in parallel. For a single component, an update rule has the form

$$\mathbf{w}^{\text{new}} = \mathbb{E}[\mathbf{z}g(\mathbf{w}^\top \mathbf{z})] - \mathbb{E}[g'(\mathbf{w}^\top \mathbf{z})]\mathbf{w}, \quad (3.45)$$

followed by normalisation of $\mathbf{w}$. Here $g(\cdot)$ is a non-linear function chosen to approximate the score function of the source distribution, and $g'(\cdot)$ is its derivative. Typical choices include $g(u) = \tanh(u)$ for super-Gaussian sources [59].

For multiple components, constraints are added to keep the rows of W orthogonal. Symmetric orthogonalisation schemes update all rows and then apply a symmetric decorrelation step. Deflation schemes estimate one component at a time and subtract its contribution from the data [59].

In the EEG setting, the latent sources include neural activity from different regions and frequency bands, ocular artefacts, muscle artefacts and possibly line noise. Ocular components often have super-Gaussian distributions because blinks and large saccades generate occasional high-amplitude events. Their scalp maps are frontal and their time courses correlate strongly with EOG. Once identified, these components can be zeroed or attenuated before reconstructing the cleaned EEG as

$$\hat{\mathbf{x}}_{\text{clean}}(t) = A\hat{\mathbf{s}}_{\text{clean}}(t), \quad (3.46)$$

where $\hat{\mathbf{s}}_{\text{clean}}(t)$ has the ocular components modified [39], [81], [82], [90], [102], [137].

VEOS interacts with ICA at this point. An inferred EOG signal $\hat{y}(t)$ produced by VEOS can be correlated with candidate components $\hat{s}_i(t)$. Components that show high correlation and frontal topographies are strong candidates for ocular artefacts. This relationship does not require VEOS to be perfect. Even a partial reconstruction

of the EOG can assist the identification step.

3.12 Mapping from video landmarks to EOG

The theoretical pieces assembled so far suggest that there is a functional relationship between video features $\mathbf{z}_k$ and the EOG signal y_k at the same time. VEOS seeks to approximate this relationship with a parametric model that is flexible enough to capture the main structure but constrained enough to avoid overfitting. Related ideas appear in appearance-based gaze estimation and multimodal EEG–eye-tracking work [20], [24], [110], [147], [160].

Let $\mathbf{z}_k \in \mathbb{R}^D$ be a feature vector derived from landmarks at frame k . Features can include normalised landmark positions, differences, velocities, eyelid distances, eye opening ratios and indicators for partial or full occlusion. Time-delayed copies of these features can be concatenated to provide short-term temporal context, so that

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{f}_k \\ \mathbf{f}_{k-1} \\ \dots \\ \mathbf{f}_{k-L+1} \end{bmatrix}, \quad (3.47)$$

where $\mathbf{f}_k$ is the basic feature vector at time k and L is the context length [20], [55].

The mapping is written as

$$\hat{y}_k = f(\mathbf{z}_k; \boldsymbol{\theta}), \quad (3.48)$$

with parameter vector $\boldsymbol{\theta}$. In a linear model,

$$f(\mathbf{z}_k; \boldsymbol{\theta}) = \mathbf{w}^\top \mathbf{z}_k + b, \quad (3.49)$$

while in a shallow non-linear model,

$$f(\mathbf{z}_k; \boldsymbol{\theta}) = \mathbf{v}^\top \sigma(W\mathbf{z}_k + \mathbf{b}) + c, \quad (3.50)$$

where W is a weight matrix, $\mathbf{b}$ and $\mathbf{v}$ are vectors, c is a scalar bias and $\sigma(\cdot)$ is a non-linear activation function applied elementwise.

Given training data $\{(\mathbf{z}_k, y_k)\}_{k=1}^T$, parameters can be estimated by minimising a loss

$$L(\boldsymbol{\theta}) = \frac{1}{T} \sum_{k=1}^T \ell(y_k, f(\mathbf{z}_k; \boldsymbol{\theta})), \quad (3.51)$$

where ℓ is a pointwise loss such as squared error. Regularisation terms can be added to penalise large weights or to encourage sparsity [141].

The mapping is subject to identifiability limits. If the EOG channel contains components that are not driven by eye motion visible in the video, such as slow drifts in electrode impedance or interference from nearby muscles, no video-based model can recover them. Similarly, if parts of the eye are occluded or the landmark detector fails, those moments carry incomplete information. VEOS therefore aims to capture the portion of the EOG that is predictable from the visible eye region.

A further source of identifiability limits is variability in the imaging conditions themselves. Changes in ambient lighting alter the contrast of the pupil boundary, iris and eyelid margins relative to the sclera, and can cause the landmark detector to fail or to track anatomically incorrect points. Near-infrared illumination reduces sensitivity to ambient light fluctuations, but specular reflections from the cornea and eyelashes remain problematic, particularly during blinks and at extreme gaze angles. Camera parameters also matter: a lower frame rate reduces the temporal resolution of blink onset and offset, a narrower field of view or shallower depth of field may cause the eye to drift out of focus during head movements, and differences in sensor noise and dynamic range affect the reliability of eyelid-aperture estimates near full closure. These imaging factors introduce structured variability into the feature sequences that the model must either absorb through regularisation or treat as a source of domain shift when transferring between datasets recorded under different conditions.

Calibration across subjects is another consideration. Differences in eye physiology and electrode placement change the scale and offset of the EOG. Including per-subject gains and offsets, or features that reflect subject-specific geometry, helps the model

adapt. In a linear setting, this can be as simple as adding subject indicator variables and allowing the regression to learn different intercepts or slope adjustments.

3.13 Synchronisation and timing considerations

The final theoretical component concerns time. VEOS operates on data recorded by several devices. Each device has its own internal clock and buffering behaviour. The mapping from video to EOG assumes that $\mathbf{z}_k$ and y_k refer to the same moment in time. Any misalignment weakens the observed dependency [8], [45], [129].

Let t_k^{cam} denote the timestamps associated with video frames and t_n^{eog} the timestamps for EOG samples. Ideally,

$$t_k^{\text{cam}} = t_0 + \frac{k}{f_{\text{cam}}}, \quad t_n^{\text{eog}} = t_0 + \delta + \frac{n}{f_{\text{eog}}}, \quad (3.52)$$

for fixed t_0 and δ . In reality, the timestamps provided by software are

$$\tilde{t}_k^{\text{cam}} = t_k^{\text{cam}} + \epsilon_k^{\text{cam}}, \quad (3.53)$$

$$\tilde{t}_n^{\text{eog}} = t_n^{\text{eog}} + \epsilon_n^{\text{eog}}, \quad (3.54)$$

where the ϵ terms represent latency and jitter.

If the errors are small and slowly varying, one can align the streams by estimating a single offset and then interpolating the higher-rate signal onto the lower-rate timeline. If there is drift, a linear correction can be introduced. Suppose that the camera clock runs slightly faster than the EOG clock. Over the duration of a recording, this introduces a mismatch that grows linearly with time. One can estimate a drift parameter α such that corrected times

$$\hat{t}_k^{\text{cam}} = t_k^{\text{cam}}(1 + \alpha) \quad (3.55)$$

align better with EOG events.

Misalignment degrades correlation. If the true EOG signal is $y(t)$ and the inferred signal is $\hat{y}(t)$, the Pearson correlation coefficient at zero lag is

$$\rho(0) = \frac{\mathbb{E}[(y(t) - \mu_y)(\hat{y}(t) - \mu_{\hat{y}})]}{\sigma_y \sigma_{\hat{y}}}, \quad (3.56)$$

with means $\mu_y, \mu_{\hat{y}}$ and standard deviations $\sigma_y, \sigma_{\hat{y}}$. If there is a constant delay τ between the signals, the relevant statistic becomes $\rho(\tau)$. A delay that varies over time will reduce the peak value of the cross-correlation function. In this sense, timing jitter appears as an additional noise source that cannot be removed by the mapping.

Hardware triggers provide the most reliable means of bounding these errors. When a single physical event, such as a TTL pulse, is recorded simultaneously by the camera shutter controller, the EEG amplifier and the EOG front end, the timestamps of that event in each stream can be compared directly. The difference gives an empirical estimate of the inter-device offset at that moment, and repeated triggers distributed across the recording session allow any slow drift between clocks to be characterised and corrected. This is preferable to relying solely on software-level synchronisation, where operating system scheduling introduces variable latency that is difficult to model. In VEOS, an ESP32 microcontroller generates TTL pulses that are received by all acquisition devices simultaneously, and a timing LED driven by the same pulse is visible in the video frames, providing an additional optical marker that is independent of any software timestamp. Together these mechanisms ensure that the offset δ and any drift α can be estimated from the data rather than assumed to be zero.

The multimodal data acquisition platform behind VEOS includes explicit synchronisation mechanisms that reduce these errors and make them measurable [8], [129]. The theoretical framework given here provides a way to interpret how residual misalignment influences the results in later chapters.

3.14 Summary

This chapter developed the theoretical backbone for VEOS in three layers. First, it introduced a projective camera model and a head-centred coordinate system for interpreting facial landmarks, leading to eye-centred descriptions of pupil position, eyeball-in-socket pose and eyelid aperture. A simple kinematic model related these quantities to underlying gaze angles and highlighted how typical eye-movement patterns (fixations, saccades, blinks) appear in the video domain, taking into account landmark noise and temporal sampling constraints.

Second, the chapter described how ocular sources are represented electrically via an eye-centred dipole and periocular EMG, and how these sources propagate through the head volume conductor to both dedicated EOG channels and frontal EEG electrodes. The characteristic time–frequency signatures and spatial patterns of ocular artefacts were outlined, together with the role of independent component analysis and related linear decompositions in separating ocular and neural contributions.

Third, the mapping from video-derived features to EOG was formalised as a supervised regression problem, with explicit attention to temporal context and to synchronisation between camera and electrophysiology streams. This provides a principled justification for treating inferred EOG as a non-neural auxiliary signal that can be learned from video and then used to support artefact handling in EEG.

The next two chapters instantiate this framework in practice. Chapter 4 describes the multimodal acquisition platform and VEOS development dataset that were designed to excite and capture the ocular phenomena modelled here. Chapter 5 then builds on the theoretical mapping by defining concrete video features, learning architectures and training procedures for VEOS, preparing the ground for the experimental evaluation in Chapters 6 and 7.

Chapter 4

Multimodal Data Acquisition Platform (Development System)

4.1 Overview

This chapter describes the multimodal data acquisition platform that was used to develop and validate VEOS. The platform is designed to record video, EEG, EOG and, when required, periocular EMG in a tightly synchronised fashion, together with stimulus markers and task state. In the VEOS development dataset (VEOS-Dev) used in this thesis, only EEG, EOG and video were recorded; dedicated eyelid EMG channels are available only in the Eye-BCI dataset described in Chapter 6. The development platform therefore provides the VEOS-Dev dataset on which the VEOS models were first constructed, before being evaluated on Eye-BCI and later related to the naturalistic AMBER dataset introduced in Chapter 2.

The design has two closely related motivations. First, it serves as a practical demonstration that a complete multimodal BCI data capture system can be specified, implemented and validated within this research programme, rather than relying only on pre-recorded datasets. Second, it was built to expose exactly the combinations of eye movements, eyelid activity and EEG artefacts that VEOS is intended to model. The platform architecture follows the theoretical considerations of Chapter 3:

it ensures adequate temporal resolution for eye movements, sufficient bandwidth for EOG and (when present) EMG, and reliable timestamps for cross-modal alignment.

The following sections describe the hardware and software architecture, the calibration and task structure, the synchronisation strategy and the resulting development dataset. Participant demographics and ethics approvals for the development recordings are given in Section 6.2.1; this chapter concentrates on the recording platform itself. Figure 4.1 provides a high-level overview of the system.

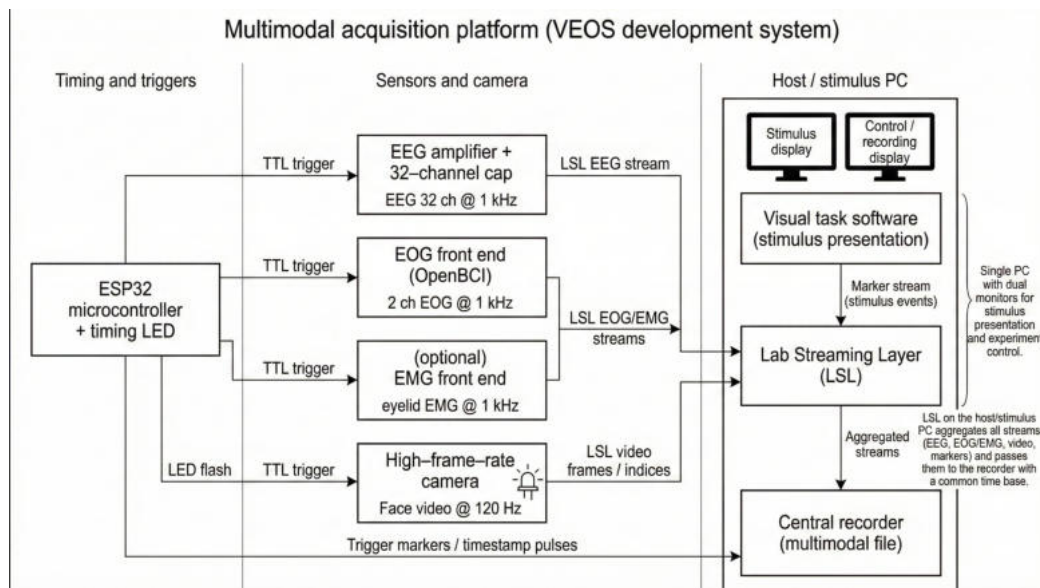


Figure 4.1: Block diagram of the multimodal acquisition platform used for the VEOS development dataset. EEG, EOG and (optionally) eyelid EMG are recorded by dedicated front ends, while a high-frame-rate camera captures face video. All devices stream data to a single host/stimulus PC via Lab Streaming Layer (LSL), which aggregates EEG, EOG/EMG, video and marker streams into a single multimodal recording. An ESP32 microcontroller distributes TTL triggers to the amplifiers and camera, drives a timing LED visible in the video, and sends matching event markers to the host PC, providing a common hardware-synchronised time reference across all modalities and with the visual tasks presented on the same machine.

4.2 Hardware architecture

The platform must simultaneously acquire high-density electrophysiological signals, high-frame-rate video of the periocular region, and precisely timed stimulus events, with all streams synchronised to a common clock. These requirements drive a hardware design built around three tightly integrated subsystems.

4.2.1 EEG and EOG acquisition

The EEG subsystem is based on a 32-channel amplifier from ANT Neuro, configured with standard research-grade shielding and isolation. The amplifier is connected to a cap with electrodes placed according to a reduced international 10–20 layout, covering frontal and central sites, which are most relevant for studying ocular artefacts and their interaction with BCI-relevant EEG components. A reference and ground configuration following the manufacturer’s recommendations is used. The exact montage is documented in the dataset metadata and in Chapter 6.

Horizontal and vertical EOG channels are recorded using a separate biopotential front end. Pairs of electrodes are placed near the outer canthi and above and below one eye to capture the horizontal and vertical components of the electrooculogram. This arrangement follows standard practice and matches the dipole model introduced in Section 3.7. The EOG front end is battery powered and electrically isolated from the EEG system to minimise ground loops.

Although the present platform uses research-grade hardware, there is now a growing literature validating portable and lower-cost EEG systems for event-related potential (ERP) work when electrode performance and trigger timing are characterised [43], [150].

In the development recordings described in this chapter, no dedicated eyelid EMG channels were acquired. Periocular EMG appears later in the thesis only in the context of an externally collected validation dataset described in Chapter 6. That dataset includes additional periocular channels and is used to test how VEOS behaves when EMG information is available alongside EEG, EOG and video.

4.2.2 High-frame-rate video subsystem

Video is acquired using a monochrome global-shutter USB camera module (ELP-USBGS1200P02) mounted above the display, configured at 720p and 120 frames per second with a 12 mm lens. A global shutter was selected to avoid rolling-shutter distortion during rapid eyelid motion and saccades, which can bias landmark localisation

and derived kinematic features [162]. The factory infrared cut filter was removed to enable near-infrared imaging, and an infrared floodlight was used to provide stable illumination of the face and pupil during recording [153].

The camera is configured to capture at 720p resolution and a high-frame-rate, set to 120 Hz in the VEOS development recordings. This frame rate is high enough to resolve saccades and blinks with several frames per event, while keeping data volume manageable.

The lens is selected to provide a field of view that tightly frames the face. The eyes occupy a significant portion of the vertical extent, but some forehead, nose and mouth region remain visible. This makes it possible to detect head-pose and coarse facial expressions in addition to eye and eyelid motion. The camera is fixed on a stable mount to avoid motion relative to the display.

A dedicated capture process running on the host machine reads frames from the camera driver and timestamps them using the local clock before pushing them into a Lab Streaming Layer (LSL) video stream [73]. For storage, frames are encoded using a constant-frame-rate video codec and written to disk. Timestamps are stored alongside the video file in a sidecar file and in the LSL recording, so that the original frame timing is recoverable even if the codec introduces slight timing differences. During the VEOS-Dev recordings the camera operated in external-trigger mode, with each frame exposure started by a TTL pulse from the ESP32 microcontroller at 120 Hz, so that camera frames were locked to the same hardware trigger used for EEG and EOG while still being timestamped against the host clock via LSL

4.2.3 Stimulus display and triggering

Visual stimuli are presented on a computer monitor positioned at a comfortable viewing distance of approximately 60–70 cm. The monitor refresh rate is set to an integer multiple of the camera frame rate whenever possible, so that changes on the screen align reproducibly with camera frames. Stimuli for calibration and BCI tasks are generated either by a custom Python application or by OpenViBE [108] scenarios,

depending on the experiment type. In other EEG laboratories, similar stimulus presentation and event handling is often implemented using general-purpose experiment control toolkits such as PsychoPy [111].

Hardware triggers and visual markers are used both to validate timing and to distribute a common event signal. A small light-emitting diode (LED) driven by an ESP32-based microcontroller is placed within the camera field of view but outside the main eye region. For each relevant event in the stimulus or EEG software, the microcontroller generates a short TTL pulse that is fanned out to the external-trigger inputs of the high-frame-rate camera, the ANT Neuro EEG amplifier, the OpenBCI EOG front end [107] and, when present, the EMG front end. At the same moment, the microcontroller briefly illuminates the LED and emits a timestamped event marker into an LSL stream via the host PC. This approach avoids relying on operating-system-level timing for event marking and is consistent with evidence that commodity interfaces can introduce measurable timing uncertainty if not handled carefully[4]. This arrangement ensures that camera exposures, electrophysiological samples, LED flashes and software markers are all time-locked to the same hardware trigger while still being expressed in a common time base through LSL.

Figure 4.2 shows a photograph of the assembled platform, illustrating the physical arrangement of all components on the recording desk while Figure 4.3 sketches the practical arrangement of these components around the participant.

4.3 Software architecture

Integrating heterogeneous hardware devices into a coherent recording system requires a software layer that abstracts away device-specific interfaces and provides reliable, timestamped access to all data streams. The software architecture described here is designed to meet this requirement while remaining flexible enough to support different experimental configurations.

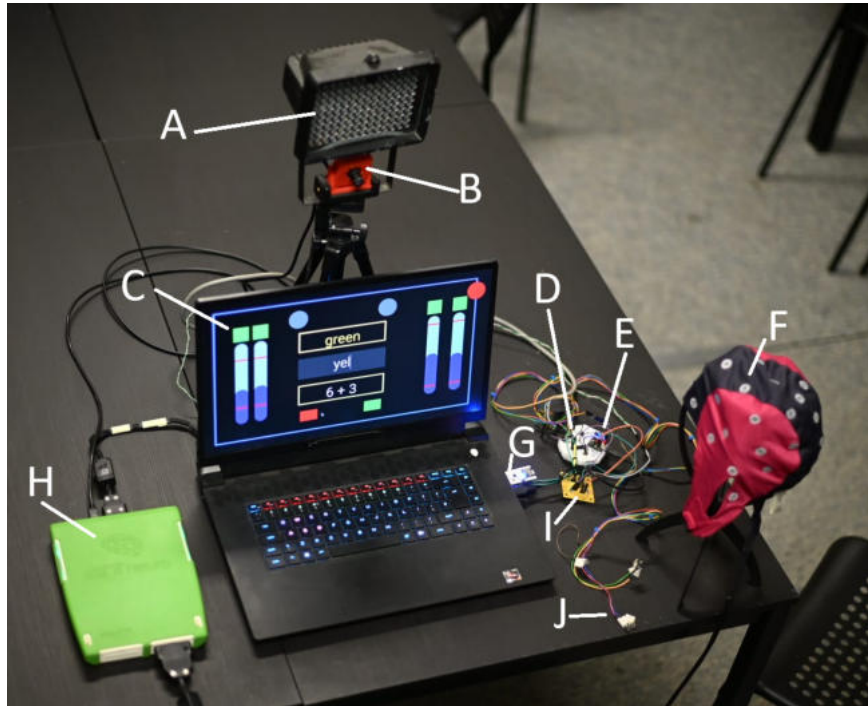


Figure 4.2: Photograph of the VEOS development recording setup. (A) Near-infrared floodlight providing stable illumination of the participant’s face and eye region. (B) 120 fps monochrome global-shutter camera with hardware external trigger input, modified to remove the infrared cut filter. (C) Stimulus laptop displaying the cognitive loading game used as the game-based engagement task. (D) OpenBCI Cyton biosignal amplifier in a custom-printed enclosure, used to acquire EOG channels. (E) ESP32 microcontroller providing hardware synchronisation: it distributes TTL trigger pulses to the camera, Cyton, ANT Neuro amplifier and laptop, and drives a timing LED visible in the camera frame. (F) 32-channel EEG cap. (G) OpenBCI USB receiver dongle. (H) ANT Neuro eego rt EEG amplifier. (I) Voltage-level shifter and signal inverter providing logic-level compatibility between the ESP32 and the camera and Cyton trigger inputs. (J) EOG electrodes connected to the Cyton for horizontal and vertical EOG acquisition.

4.3.1 Lab Streaming Layer integration

All data streams are integrated using LSL, which provides a common abstraction for time series acquisition and synchronisation. Each hardware device is associated with one or more LSL outlets. For example, an EEG outlet exposes the 32-channel ANT Neuro data with appropriate channel labels and units, and an EOG/EMG outlet exposes the OpenBCI channels used for ocular and muscle activity [107]. In addition, a video outlet provides image frames or frame indices together with their timestamps, and a marker outlet carries discrete events such as stimulus onset, button presses and microcontroller triggers.

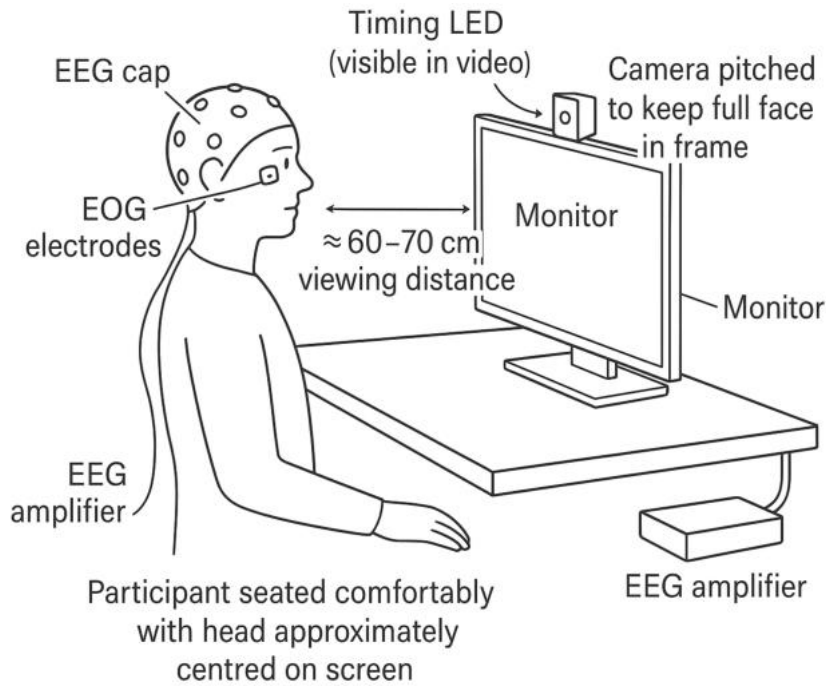


Figure 4.3: Physical arrangement of the VEOS development recording setup. A seated participant wearing a 32-channel EEG cap and periocular EOG electrodes views a monitor at a viewing distance of approximately 60–70 cm, with a 120 Hz face camera mounted on the top bezel to capture the full face. A small LED placed in a screen corner provides visible timing flashes within the camera field of view. Approximate eye–screen distance, screen size and camera position are indicated to clarify the recording geometry.

LSL maintains internal estimates of clock offsets between devices and the host and provides corrected timestamps in a common time base. This is crucial for the later mapping between video and electrophysiological signals. The theoretical aspects of this synchronisation were discussed in Section 3.13. Here the focus is on the concrete implementation.

On the host side, a central recording application opens inlets for each outlet and continuously reads samples. Samples are written to disk in the Extensible Data Format (XDF), an LSL-compatible container that preserves timing and channel metadata for downstream analysis[72]. This yields a single multimodal recording per session that can be replayed or processed offline without further concern for device-specific details. Figure 4.4 illustrates this data flow, from individual device outlets through LSL to the central recording file.

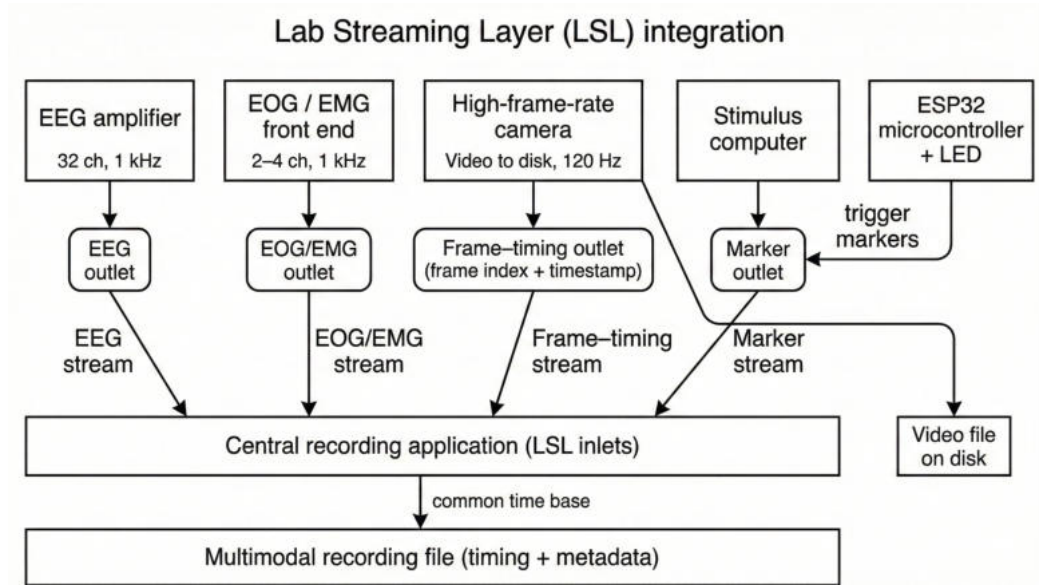


Figure 4.4: Software-level data flow through Lab Streaming Layer (LSL). Physiological devices expose LSL outlets for EEG, EOG/EMG and marker streams, while the camera acquisition software writes video frames to disk and publishes only frame indices and timestamps via an LSL outlet. A central recording application opens corresponding inlets, subscribes to the EEG, EOG/EMG, frame-timing and marker streams, and writes a single multimodal recording file that preserves channel metadata and a common time base. The ESP32 microcontroller drives the LEDs and generates all digital trigger events, which are inserted into the marker outlet to provide shared timing markers across electrical and visual channels.

4.3.2 Custom recording interface

To manage recordings, a custom recording interface was implemented in Python. This interface provides a graphical summary of the current streams, showing which outlets are visible on the network, their nominal sampling rates and current throughput. It allows the operator to enter participant identifiers and session metadata before starting a recording and to start and stop the central LSL recorder from the same interface. The operator can also monitor basic signal quality, including EEG impedances from the ANT Neuro software and raw EOG traces, and view a low-resolution preview of the camera feed to ensure correct framing and focus.

The interface enforces a consistent directory structure and file naming scheme. Each session directory contains the LSL recording file, the raw encoded video, any additional logs from the stimulus software and a small JSON metadata file describing the configuration, electrode layout and any deviations from the standard protocol.

This structure simplifies offline analysis and traceability.

4.3.3 OpenViBE and task logic

For BCI-style tasks, OpenViBE is used to define and execute stimulus scenarios. OpenViBE flows are configured to send event markers into LSL whenever specific screen events occur, such as the onset of a stimulus block, the presentation of a calibration point or the start of a rest period. In some scenarios, OpenViBE also receives derived signals, such as online band power estimates, although VEOS itself is trained and evaluated entirely offline.

The choice of OpenViBE has two advantages. First, it reuses an established BCI framework, which makes the protocols easier to reproduce in other laboratories. Second, it ensures that EEG and task logic can be modified without changes to the acquisition backbone.

4.4 Calibration and stimulus design

Useful data for training and evaluating VEOS requires that participants produce a controlled and representative range of eye movements and eyelid behaviours during recording. Three distinct paradigms were designed to serve this need, covering deliberate gaze shifts for calibration, naturalistic behaviour during an engagement task, and the structured stimulus sequences required for BCI paradigm evaluation.

4.4.1 Follow the dots: gaze calibration

A dedicated calibration task is used to expose controlled eye movements and eyelid behaviour. During this task, the participant is instructed to follow a sequence of dots that appear at known locations on the screen. The sequence covers central, horizontal and vertical positions, with occasional diagonal positions to ensure that combined horizontal and vertical rotations are represented.

Each dot is presented for a fixed duration while the participant is asked to maintain

gaze, then the dot jumps to a new location. Short rest intervals are inserted to reduce fatigue. The timing and coordinates of each dot are logged as stimulus markers and stored in the LSL marker stream.

From the point of view of VEOS, this task provides known changes in gaze direction that can be related to changes in landmark positions, as formalised in Section 3.3, as well as segments of relatively stable gaze that are useful for estimating baseline noise and drift in the video-based features. It also provides opportunities to observe blinks in a context where the underlying gaze direction is known, which simplifies blink labelling.

4.4.2 Game-based engagement task

To generate more naturalistic eye and eyelid behaviour while maintaining a controlled structure, a game-based engagement task is used, following the design described in earlier work on blended control of scale vehicles [129].

In this task the participant plays an abstract “cognitive loading” game that presents five concurrent subtasks on a single screen (Figure 4.5). Each subtask produces discrete events that require a mouse click or typed response:

- (i) A ball moves around the border of the screen and occasionally comes to a halt. The participant must click on it to keep it moving.
- (ii) Two vertical “tanks” slowly leak fluid. The participant must click a valve to keep the fluid level between upper and lower target lines.
- (iii) Simple arithmetic expressions (e.g. “ $6 + 3$ ”) appear in a lower text box and the participant types the numerical answer.
- (iv) A Stroop colour–word task, in which colour names are displayed in potentially incongruent font colours (e.g. the word “green” shown in yellow). The participant must type the font colour rather than the word.
- (v) Small indicator lights at fixed locations on the screen illuminate at random intervals and must be clicked to turn them off.

Tasks (i) and (v) approximate reactions to sudden, brief events, whereas task (ii) requires continuous monitoring of a predictable process. Tasks (iii) and (iv) impose a sustained but modest cognitive demand. Each event instance is timestamped in the game log, and reaction times and misses are recorded per subtask. An overall performance score is computed from the average reaction time across all subtasks.

The game runs continuously in blocks of approximately ten minutes at each of three difficulty levels. Difficulty is manipulated by varying the rate and temporal overlap of events. At the lowest level, events are infrequent and rarely overlap, so that the participant can respond comfortably with periods of relative quiet. The intermediate level presents a near-continuous stream of non-overlapping events, leaving little idle time but remaining achievable. At the highest level, as depicted in Figure 4.5, multiple subtasks often require attention at the same time and it is generally not possible to respond to everything; an effective strategy is to prioritise older events so as to minimise the backlog of outstanding tasks.

From the VEOS perspective, this game provides structured blocks spanning underloaded to overloaded conditions while keeping the visual environment stable. Game events and difficulty levels are written into the LSL marker stream, allowing eye and eyelid behaviour, performance and workload to be aligned precisely in time. Compared with the follow-the-dots calibration, the game produces more varied sequences of saccades, fixations and blinks driven by task demands, while still being linked to a well-defined task structure.

4.4.3 ERP-style visual paradigms

A third class of engagement follows the structure of ERP-style visual paradigms used in previous multimodal BCI work. In these recordings, visual stimuli are presented in blocks designed to elicit event-related responses, such as oddball or rapid serial visual presentation sequences. The timing of each stimulus is logged precisely through the marker stream.

From the VEOS perspective, these ERP-like tasks have two roles. First, they ex-

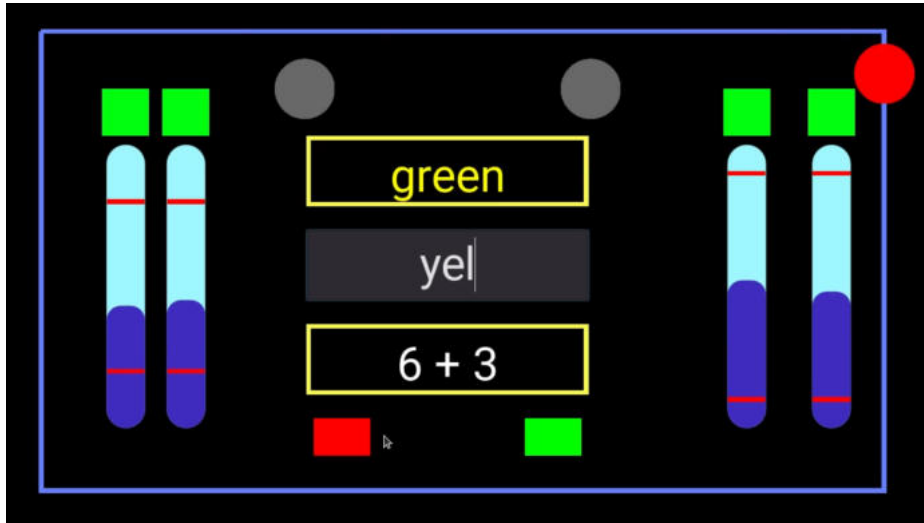


Figure 4.5: Example screen from the abstract cognitive loading game used as the game-based engagement task at the highest difficulty level. The interface combines five concurrent subtasks: ball motion around the border, leaky tanks, arithmetic questions, a Stroop colour-word entry task and indicator lights.

pose conditions where participants are encouraged to suppress blinks during stimulus trains and to blink during brief pauses, leading to characteristic clusters of blinks before and after blocks. Second, they resemble paradigms used in the AMBER dataset, which later allows VEOS behaviour on the development platform to be compared with its behaviour on a larger, independently collected dataset.

4.5 Data acquisition and synchronisation in practice

Temporal alignment between the video, electrophysiological and stimulus streams is a prerequisite for all of the analyses in later chapters. Even small inter-channel delays can introduce systematic errors in regression targets and distort event-locked analyses, making synchronisation one of the most operationally critical aspects of the platform.

4.5.1 Timestamping and clock management

All devices in the platform rely on the host computer clock as a reference through LSL’s clock synchronisation mechanism. Where multiple acquisition computers are

involved, this approach relies on disciplined local clocks, typically via standard time synchronisation protocols such as NTP [89] or PTP [60]. At start up, each LSL outlet performs a short negotiation protocol to estimate the offset and jitter between its local clock and the host clock. The resulting correction parameters are applied to each sample so that when the central recorder writes data to disk, all channels share a common time base. The ESP32 hardware trigger provides a shared sequence of physical events across devices, while LSL’s clock synchronisation maps all samples and trigger markers onto the same host-clock timeline.

The theoretical behaviour of this system was outlined in Section 3.13. In practice, the key properties are stable clock offsets for each device over the duration of a session and small residual jitter in the corrected timestamps, typically much shorter than the sampling interval for EEG and EOG. Equally important is the absence of large discontinuities in the device clocks, such as might be caused by power-saving modes or driver resets.

To guard against rare timing anomalies, the acquisition software monitors the difference between nominal sampling intervals and the actual differences between successive timestamps. Any large deviations are logged and those segments can be excluded during analysis if necessary.

4.5.2 End-to-end timing validation

End-to-end validation of synchronisation is performed using the LED marker and known stimulus timings. In a typical validation procedure, a short scenario causes the LED to flash at pseudo-random intervals while EEG, EOG and video are recorded. The microcontroller writes an LSL marker at the moment it drives the LED.

Offline, the video is analysed to detect the frames in which the LED is lit. The timestamps of these frames are compared with the corresponding EEG and EOG sample timestamps at which a change in the LED-related artefact is visible, and with the microcontroller’s own LSL markers. The differences form an empirical distribution of inter-channel delays.

In the development platform, these tests show that relative timing errors between the video stream and the electrophysiological channels are small compared with the temporal scales of interest for blinks and saccades. In other words, any remaining misalignment acts as a small additional noise source rather than a dominant error term, which is consistent with the modelling assumptions in Section 3.13.

Figure 4.6 illustrates a short excerpt from a multimodal recording, showing the alignment between LED markers, video frames and electrophysiological traces.

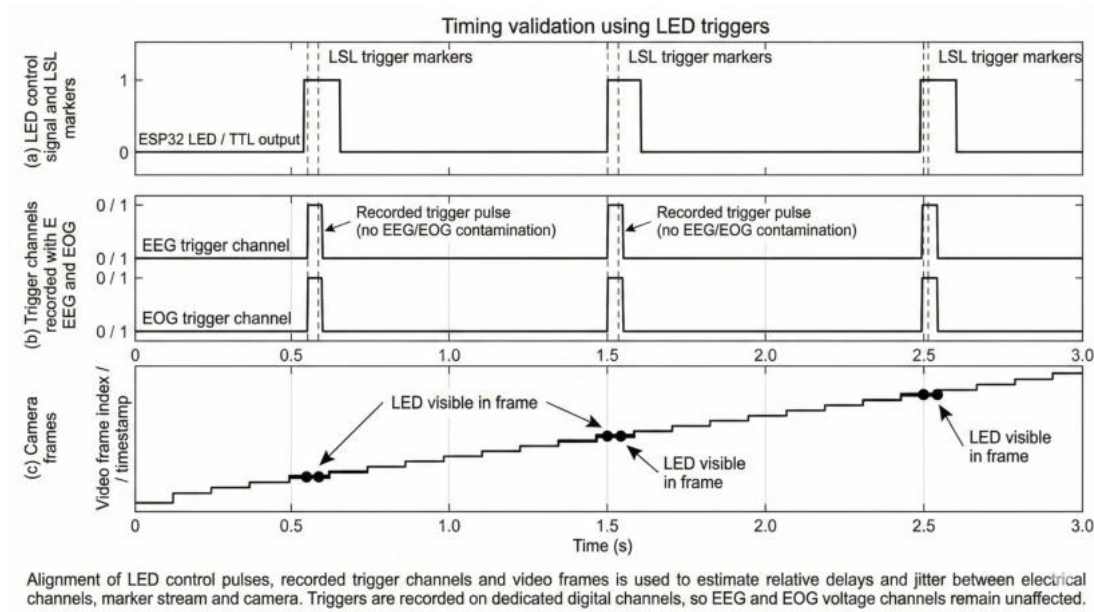


Figure 4.6: Example multimodal timeline used for timing validation. Panel (a) shows the ESP32 LED / TTL control signal together with the corresponding LSL trigger markers. Panel (b) plots the dedicated EEG and EOG trigger channels recorded by the amplifiers; these digital channels receive the same pulses but do not affect the underlying EEG or EOG voltage signals. Panel (c) indicates the video frame indices, with the frames in which the LED is visible in the camera image highlighted. Aligning these three panels allows relative delays and jitter between the hardware triggers, LSL marker stream and camera to be estimated, while keeping the physiological channels free from LED-related artefacts.

4.6 Development dataset and file structure

The recordings collected using the platform described above constitute the VEOS development dataset (VEOS-Dev), which serves as the primary source of training and validation data for the inference models developed in this thesis. A consistent

file structure and naming convention were adopted from the outset to support reproducible analysis and straightforward cross-referencing between modalities.

4.6.1 Recorded modalities and channels

Each recording in the VEOS development dataset (VEOS-Dev) comprises 32-channel EEG from the ANT Neuro amplifier, band-limited only by the analogue front end and recorded at a sampling rate suitable for BCI research, together with a small number of EOG channels configured in a standard horizontal and vertical layout. In addition, a high-frame-rate 720p video stream covers the participant’s face, and an event marker stream captures stimulus timings, calibration point indices, task states and LED triggers.

No dedicated eyelid EMG channels were recorded in the VEOS development dataset. Eyelid EMG is introduced later only when discussing the independent validation dataset hosted on Synapse and described in Chapter 6, which includes periocular EMG channels in addition to EEG, EOG and eye-related signals.

4.6.2 Directory layout and reproducibility

The VEOS development dataset (VEOS-Dev) is organised into a hierarchical directory structure with a top-level directory per participant and, within each participant, one directory per session. Within each session, the LSL recording file, raw video file, and configuration files are stored together.

Configuration files include the electrode montage and any deviations from the standard cap layout, together with the version of the acquisition and stimulus software used. They also record notes on participant-specific factors such as wearing of glasses or unusual posture.

The resulting layout is shown in Figure 4.7. This structure is influenced by earlier work on open multimodal platforms for passive BCIs, where the emphasis was on reproducible data collection and sharing of complete experimental setups. The same principles apply here: any researcher who has access to the hardware described

above and to the software components could reconstruct the acquisition pipeline and generate comparable data.

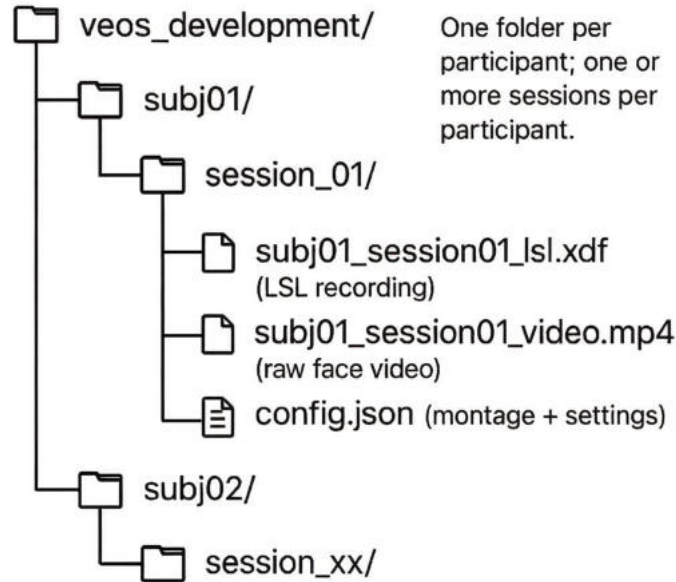


Figure 4.7: Directory structure of the VEOS development dataset. Recordings are organised by participant (subj01, subj02, ...), with one or more session folders under each participant. Each session contains a multimodal LSL recording file, a corresponding raw video file, and a JSON configuration file describing the montage and acquisition settings. Log files from the stimulus software are written to the session directory when present, but are not generated for all session types. This structure supports reproducible analysis and straightforward cross-referencing between modalities.

4.7 Relation to previous multimodal platforms

The platform presented in this chapter extends previous work on multimodal data acquisition [8], [129]. Earlier work on an open platform for passive hybrid BCI research combined EEG, eye-tracking and wearable IMU streams, with tasks designed to elicit different levels of cognitive load in simulated driving and control scenarios [129]. That platform emphasised open-source tools, reproducible experimental setups and the use of LSL to manage heterogeneous data streams. It also demonstrated the value of a modular acquisition architecture in which individual sensing components could be replaced or reconfigured without requiring the entire system to be redesigned. This was particularly useful in experimental settings, where changes in hardware, participant

setup or protocol often introduced practical constraints that had to be accommodated quickly while preserving consistent timing and data logging. The VEOS platform inherits several design decisions from that work: the use of ANT Neuro EEG hardware, LSL-based integration, and Python-based acquisition scripts. It extends them in two important ways. First, it introduces a high-frame-rate camera focused explicitly on the periocular region, rather than a generic scene or eye-tracker view. Second, it adds dedicated EOG channels, acquired through an OpenBCI board[107], with sensor placement chosen to support the dipole model in Chapter 3. Together these additions shift the focus from general-purpose multimodal recording towards the specific requirements of vision-based ocular signal inference.

In parallel, a separate line of research has explored multimodal data acquisition for controlling scale vehicles and for naturalistic brain-computer interfaces in more complex environments [9], [129]. Those systems share the general idea of combining video, EEG and additional sensors, but are tuned towards different tasks and constraints. The VEOS development platform can be seen as the specialisation of that general approach to the specific problem of inferring EOG from video in a laboratory setting.

4.8 Open science, code and data availability

Open science practices, including open data, open code and transparent reporting, are increasingly recognised as essential for reproducible neuroscience and BCI research. The FAIR principles emphasise that scientific outputs should be findable, accessible, interoperable and reusable [149]. In neuroimaging, community standards such as the Brain Imaging Data Structure (BIDS) and its EEG extension (EEG-BIDS) provide concrete templates for organising and sharing data in ways that facilitate re-use and secondary analysis [41], [114]. These developments have been particularly influential for EEG and MEG, where complex preprocessing pipelines and strong dependence on artefact handling can otherwise make results difficult to reproduce. In practice, BIDS has also been supported by shared repositories and workflow infrastructure,

including OpenNeuro [84] and containerised analysis distributions such as BIDS Apps [40], typically provided as Docker images and often executed with container runtimes such as Singularity on shared computing systems [33], [75]. For electrophysiology, BIDS-aware tooling such as MNE-BIDS provides practical interfaces for reading and validating BIDS-structured EEG and MEG datasets [5].

This thesis follows these principles in three ways. Community recommendations for MEG and EEG reporting and data sharing, including guidance on documenting trigger paths and timing uncertainties, further motivate the choices made here [113]. For versioning and distributing large datasets and derived artefacts, tools such as DataLad provide a practical complement to Git by supporting dataset-level provenance and content-addressable storage [47]. DataLad builds on established large-file versioning approaches such as git-annex [53]. First, where possible it relies on publicly available datasets that already conform to open-science best practice, notably the Eye-BCI multimodal dataset distributed via Synapse [45]. Second, the in-house VEOS development dataset is organised in a BIDS-inspired manner and accompanied by detailed documentation of recording parameters, tasks and preprocessing, so that it can be interpreted by other researchers within the constraints of GDPR and local ethics approval. Where data are shared beyond the present project, BIDS extensions such as iEEG-BIDS and richer event description frameworks such as HED provide additional options for standardising electrophysiology datasets and experimental annotations [50], [54]. Third, all core algorithms, configuration files and analysis scripts used to implement VEOS are released as open-source software. Complementary repository services such as the DANDI Archive and Neurodata Without Borders address broader classes of electrophysiology data and metadata [29], [96].

The VEOS implementation comprises: (i) a Python library for computing eye-region features from video (using MediaPipe and OpenCV), (ii) training and evaluation scripts for the windowed multilayer perceptron and temporal convolutional network models described in Chapters 5 and 7, and (iii) utilities for integrating inferred EOG into ICA/LCA-based ocular artefact removal pipelines. Workflow specifications

and environment capture can be supported by workflow managers such as Snakemake [91] and by container-based distribution, which is now common in BIDS-oriented tooling [33], [40], [75]. For archiving releases of code and derived data with persistent identifiers, services such as Zenodo and the Open Science Framework are widely used, and Software Heritage provides long-term preservation of software source code [21], [136], [156].

The code is made available under a permissive open-source licence on GitHub at:

<https://github.com/PeterRedmond/VEOS>

This repository includes example configuration files and notebooks that reproduce the main analyses in Chapters 7 and 8, subject to access to the underlying datasets. By releasing VEOS in this way, the intention is to enable other groups to (i) verify the results reported here, (ii) adapt the models to their own multimodal EEG-video datasets, and (iii) extend the framework to related applications such as mobile and neuroergonomic BCIs.

4.9 Summary

This chapter has outlined the multimodal platform used to capture the VEOS development dataset (VEOS-Dev). The hardware architecture combines research-grade EEG, dedicated EOG channels and a high-frame-rate camera. The software architecture uses Lab Streaming Layer to present a unified, time-aligned view of all streams, with custom recording software and OpenViBE scenarios managing acquisition and task logic. Calibration and BCI-style tasks are designed to expose controlled and naturalistic eye and eyelid behaviours.

Together, these elements define the VEOS development dataset, in which the relationships developed theoretically in Chapter 3 can be observed directly in joint EEG, EOG and video recordings: eyeball rotations and eyelid movements are visible in the video, their electrical signatures are measured by EOG, and their impact on EEG can be traced. The next chapter builds on this platform to describe the VEOS

inference model and the computer vision pipeline that converts video into an estimated EOG signal.

Chapter 5

VEOS: Vision-based Electrooculogram Signal Inference

5.1 Chapter overview

This chapter presents the Vision-based Electrooculogram Signal Inference (VEOS) framework. VEOS learns a mapping from video-based eye and eyelid motion to electrooculogram (EOG) signals and then uses the inferred EOG as an auxiliary channel for electroencephalogram (EEG) cleaning and interpretation in brain-computer interface (BCI) applications.

Chapter 3 introduced the theoretical foundations of EEG, EOG, electromyogram (EMG) and independent component analysis (ICA), together with relevant computer vision concepts and a formal supervised-learning formulation of video-to-EOG mapping. Chapter 4 described the multimodal data acquisition platform used to capture synchronous EEG, EOG and video in a controlled laboratory setting. The present chapter focuses on the VEOS algorithm itself and its integration into BCI signal processing pipelines. Early results from the VEOS framework were reported in abstract form in [126].

Two distinct datasets are used for the quantitative experiments in this work. The first is the VEOS development dataset (VEOS-Dev) collected with the multimodal

platform described in Chapter 4. It contains EEG, horizontal and vertical EOG and full-face video and is used to develop and train VEOS. The second is the Eye-BCI dataset introduced in Chapter 6, which includes high-density EEG, vertical EOG, eyelid EMG, high-speed eye video and eye-tracking. That dataset is used to test the generality of the VEOS ideas, to refine blink and eyelid models and to evaluate how inferred ocular activity relates to EEG artefacts and task-related EEG responses. EMG is only present in the Eye-BCI dataset and is not available in the VEOS-Dev recordings. In addition to these two datasets, the broader research programme that underpins this thesis also makes use of the AMBER dataset described in Chapter 2. AMBER provides EEG and full-face video in more naturalistic tasks but does not include dedicated EOG channels. In the context of VEOS it mainly serves as a design template and as an example of the kind of EEG+video-only dataset for which VEOS is intended, rather than as a primary supervised training set.

The remainder of the chapter is organised as follows. Section 5.2 formulates VEOS as a supervised learning problem, closely following the theoretical framework of Chapter 3 but with an implementation focus. Section 5.3 provides a high-level overview of the VEOS processing chain. Sections 5.4–5.7 describe the individual stages in detail: video processing and landmark tracking, feature engineering, target construction and regression modelling. Section 5.8 then explains how inferred EOG is integrated into regression-based linear component analysis (LCA) and ICA-based EEG cleaning pipelines.

5.2 Problem formulation

Building on the supervised mapping framework introduced in Chapter 3, VEOS is implemented as a supervised regression problem. Let

- v_t denote the video frame at time t ,
- $\mathbf{z}_t \in \mathbb{R}^d$ denote a vector of eye-related computer vision features extracted from v_t ,

- $e_t \in \mathbb{R}$ denote an EOG sample at time t from either a horizontal or vertical channel,
- $\mathbf{x}_t \in \mathbb{R}^M$ denote the vector of EEG voltages at time t across M electrodes.

For the VEOS development dataset (VEOS-Dev), the video and EOG signals are recorded concurrently and time-aligned using LSL. After synchronisation it is assumed that for each video feature vector $\mathbf{z}_t$ there is a corresponding EOG sample e_t at the same (or interpolated) time index. VEOS aims to learn a parametric function

$$\hat{e}_t = f_{\boldsymbol{\theta}}(\mathbf{z}_{t-L}, \dots, \mathbf{z}_t, \dots, \mathbf{z}_{t+R}), \quad (5.1)$$

where L and R define the number of past and future frames used as temporal context and $\boldsymbol{\theta}$ denotes the model parameters.

In practice the temporal context is encoded by constructing an augmented feature vector

$$\boldsymbol{\phi}_t = [\mathbf{z}_{t-L}^\top, \dots, \mathbf{z}_t^\top, \dots, \mathbf{z}_{t+R}^\top]^\top \in \mathbb{R}^D, \quad (5.2)$$

and learning a mapping

$$\hat{e}_t = f_{\boldsymbol{\theta}}(\boldsymbol{\phi}_t), \quad (5.3)$$

where $D = d(L + R + 1)$ is the dimensionality of the temporally augmented feature vector.

The simplest instance is a linear regression model

$$\hat{e}_t = \mathbf{w}^\top \boldsymbol{\phi}_t + b, \quad (5.4)$$

where $\mathbf{w} \in \mathbb{R}^D$ and $b \in \mathbb{R}$ are learned coefficients. This can be extended to non-linear models by replacing $\boldsymbol{\phi}_t$ with a non-linear feature map or by using a non-linear regressor such as a shallow neural network or a temporal convolutional network (TCN).

In all cases the baseline training objective is to minimise a regularised squared

error

$$\mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^T (e_t - \hat{e}_t)^2 + \lambda \mathcal{R}(\boldsymbol{\theta}), \quad (5.5)$$

where $\mathcal{R}(\boldsymbol{\theta})$ is a regularisation term and λ controls the strength of regularisation. Section 5.7.4 describes optional extensions to this loss, including correlation- and derivative-based terms and blink-aware weighting.

Once trained, $f_{\boldsymbol{\theta}}$ is applied to new video streams to infer an EOG signal even when no physical EOG electrodes are present. The inferred signal $\hat{e}_t$ is then used as an additional channel when cleaning EEG with LCA or ICA, as described in Section 5.8.

5.3 Overview of the VEOS architecture

The VEOS processing chain can be divided into five stages:

1. **Video acquisition and synchronisation.** Video is captured from a 120 Hz camera in the development system and synchronised with EEG and EOG using LSL. For the Eye-BCI dataset, synchronised EEG, EOG, EMG, high-speed camera and eye-tracker streams are provided by the original authors.
2. **Face and eye localisation.** Each frame is processed to locate the face and then the eye regions. OpenCV is used for frame handling and initial face detection, and MediaPipe Face Mesh provides dense facial landmarks that give precise eye and eyelid geometry over time [13], [68].
3. **Feature extraction.** From the landmarks and image data, VEOS computes geometric and temporal descriptors of eye position, eyelid aperture via an eye aspect ratio (EAR) measure [31], [138], and blink dynamics. These descriptors form the feature vector $\mathbf{z}_t$ and its temporally augmented version $\boldsymbol{\phi}_t$.
4. **EOG and EMG target construction.** For the VEOS development dataset (VEOS-Dev), raw EOG is filtered, baseline-corrected and resampled to match the video frame times. For the Eye-BCI dataset, vertical EOG and eyelid EMG

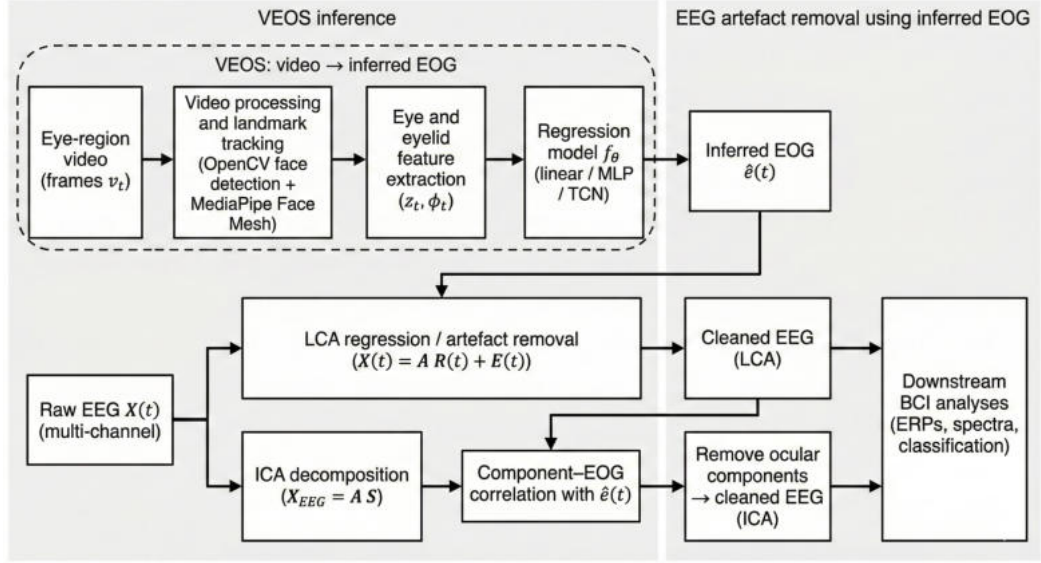


Figure 5.1: Overview of the VEOS processing chain. Monocular eye-region video is converted into eye and eyelid features, which are mapped by a regression model to an inferred EOG signal. This signal is then used alongside EEG in regression-based linear component analysis (LCA) and independent component analysis (ICA) pipelines to identify and remove ocular components, yielding cleaned EEG for downstream BCI analyses.

channels are preprocessed and used both as targets and as additional validation for blink detection.

5. **Model training and inference.** Regression models with temporal context are trained on the VEOS-Dev dataset to map visual features to EOG. The trained models are then applied to new video data, in particular the Eye-BCI recordings and, in principle, EEG+video-only datasets such as AMBER, and the inferred EOG is used to support EEG cleaning via LCA and ICA.

Figure 5.1 illustrates this processing chain at a high level.

Before any inference can be performed, raw video frames must be processed into a compact, calibrated representation of ocular geometry. This involves detecting the face, localising the relevant landmarks around the eye, and handling the practical complications that arise from head movement, variable illumination and occasional tracking failures.

5.4 Video processing and landmark tracking

Before any inference can be performed, raw video frames must be processed into a compact, calibrated representation of ocular geometry. This involves locating the relevant facial landmarks around the eyes and handling the practical complications that arise from head movement and occasional tracking failures.

5.4.1 Frame acquisition and timing

In the development system, video is captured from a 120 Hz camera configured to provide a stable view of the participant’s face with emphasis on the eye region. The frame rate is high enough to resolve individual saccades and blinks, which typically occur over tens of milliseconds. The recording software timestamps each frame and streams it through LSL so that it can be aligned with EEG and EOG streams.

For the Eye-BCI dataset, the high-speed camera operates at a different frame rate and resolution and synchronisation is handled by the acquisition software provided by the dataset authors. For VEOS it is sufficient that each video frame can be associated with a precise time and aligned with the corresponding EEG and ocular channels.

5.4.2 Facial landmarks with MediaPipe Face Mesh

MediaPipe Face Mesh is applied directly to each incoming frame to estimate a dense set of three-dimensional facial landmarks [13], [68]. The camera is mounted above the display and framed to capture from the forehead to just below the nose. MediaPipe internally localises the face and defines its own region of interest (ROI); the resulting landmark coordinates are therefore expressed in the full frame space rather than within a separately computed bounding box.

The model outputs a fixed number of landmark points per frame with detailed coverage of the eyes, eyelids and eyebrows. A subset is selected that corresponds to the inner and outer canthi of each eye, the midpoints of the upper and lower eyelids, the approximate pupil centre for each eye, and eyebrow points that provide context on

brow elevation. For each frame, VEOS stores the two-dimensional image coordinates of these landmarks and, where available, the depth coordinate provided by the model. This yields a compact geometric representation of eye and eyelid pose that is robust to small head movements, since the features are derived from relative positions within the face.

A subset of landmarks is selected that corresponds to:

- the inner and outer canthi of each eye,
- the midpoints of the upper and lower eyelids,
- the approximate pupil centre for each eye,
- eyebrow points that give context on brow movement.

For each frame, VEOS stores the two-dimensional image coordinates of these landmarks and, where available, the depth coordinate provided by the model. This results in a compact geometric representation of eye and eyelid pose that is robust to small head movements, since the features are derived from relative positions within the face ROI.

5.4.3 Handling head motion and tracking failures

Although the participant is seated and instructed to minimise head movement in the development recordings, small translations and rotations do occur. To reduce their effect, the landmark coordinates are normalised to a face-centred coordinate system. For example, VEOS subtracts the nose bridge landmark position and scales by the inter-canthal distance so that the eye geometry is expressed relative to a canonical face size and position. Where a full rigid alignment is needed, a similarity transform can be estimated from corresponding landmark sets [143].

Occasional frames may contain tracking failures, for example when the participant looks down sharply or partially occludes the eye with a hand. VEOS detects these failures by checking for landmark positions outside plausible regions or sudden large

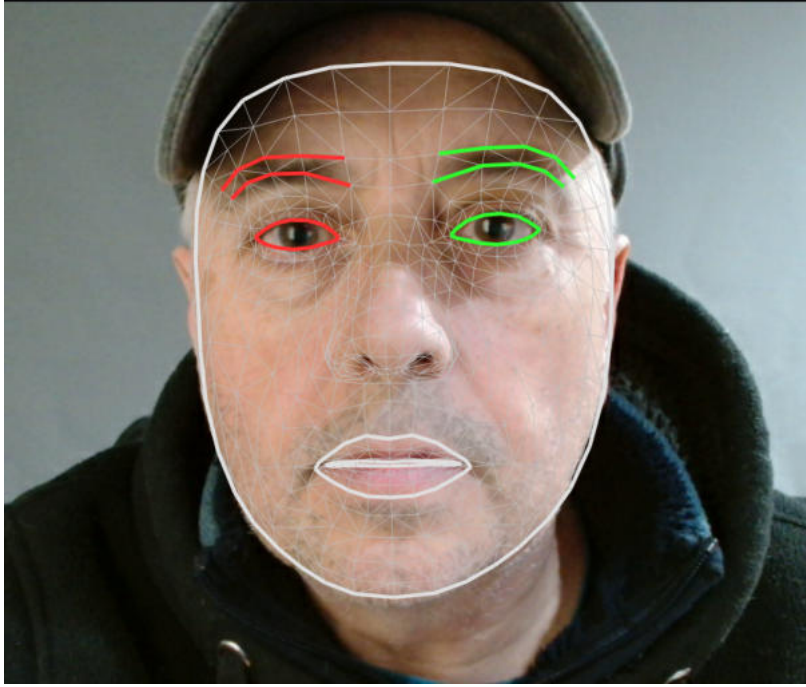


Figure 5.2: Illustrative output of the MediaPipe Face Mesh stage for one example frame. The face region of interest and triangulated mesh are overlaid in white. Eye and eyebrow contours derived from selected landmarks are highlighted for the left eye and eyebrow (red) and the right eye and eyebrow (green); these form the basis of the geometric eye and eyelid features used by VEOS.

jumps between frames. Affected frames are either discarded or have their features interpolated from neighbouring frames, depending on the density of missing data. The corresponding EOG samples are treated in the same way to maintain alignment.

5.5 Feature engineering

The geometric landmarks and raw greyscale eye patches are transformed into a feature vector $\mathbf{z}_t$ designed to capture the aspects of eye motion that drive EOG and blink-related EMG. Figure 5.3 provides an illustrative example of these features alongside simultaneously recorded ocular signals.

5.5.1 Geometric features

Geometric features describe the static configuration of the eyes and eyelids in each frame. VEOS computes:

- **Horizontal eye position.** The horizontal displacement of each pupil centre relative to the midline between the inner and outer canthi. These features correspond closely to horizontal EOG.
- **Vertical eye position.** The vertical displacement of each pupil centre relative to the horizontal line through the canthi midpoints. These features relate to vertical EOG.
- **Eyelid aperture (eye aspect ratio, EAR) [31], [138].** The distance between the upper and lower eyelid midpoints, normalised by the inter-canthal distance. This normalised vertical eyelid distance behaves like an eye aspect ratio (EAR): it is largest for fully open eyes, decreases during blinks and partial eyelid closures, and approaches zero near full closure.
- **Asymmetry measures.** Differences in position and aperture between the left and right eyes, which can indicate subtle squinting or asymmetric blinks.
- **Eyebrow elevation.** Vertical displacement of selected eyebrow landmarks relative to the canthi, which can influence forehead EMG and might correlate with some ocular artefacts in EEG.

These geometric features are largely invariant to global head translation because of the face-centred normalisation described earlier.

5.5.2 Temporal features

EOG is particularly sensitive to the dynamics of eye movements, not just to static positions. To capture dynamics, VEOS augments the geometric features with temporal derivatives computed over a short window.

For each scalar geometric feature z_t a first-order temporal derivative

$$\dot{z}_t = \frac{z_t - z_{t-1}}{\Delta t}, \quad (5.6)$$

and a smoothed second-order derivative

$$\ddot{z}_t = \frac{\dot{z}_t - \dot{z}_{t-1}}{\Delta t}, \quad (5.7)$$

are computed, where Δt is the time between frames. These derivatives approximate instantaneous velocity and acceleration.

Rapid saccades produce characteristic peaks in the velocity and acceleration of horizontal and vertical eye position, while blinks produce sharp changes in eyelid aperture and its derivatives. Including these temporal features helps the regression model to reproduce both the shape and timing of EOG deflections.

5.5.3 Blink and saccade descriptors

In addition to frame-wise features, VEOS derives event-level descriptors for blinks and saccades. A simple blink detector is run on the eyelid aperture signal, here represented by the eye aspect ratio (EAR), by thresholding EAR and enforcing minimum-duration constraints. For each detected blink VEOS computes:

- blink duration,
- minimum EAR (aperture),
- maximum closure speed,

Similarly, horizontal and vertical saccades are detected using velocity thresholds on pupil position. For each saccade, VEOS computes:

- saccade amplitude estimated from pixel displacements,
- peak velocity,
- the main-sequence relationship between amplitude and peak velocity.

These event-level descriptors are primarily used for analysis and to interpret the regression model, but can also be included as additional inputs when the model is trained on segments rather than on individual frames.

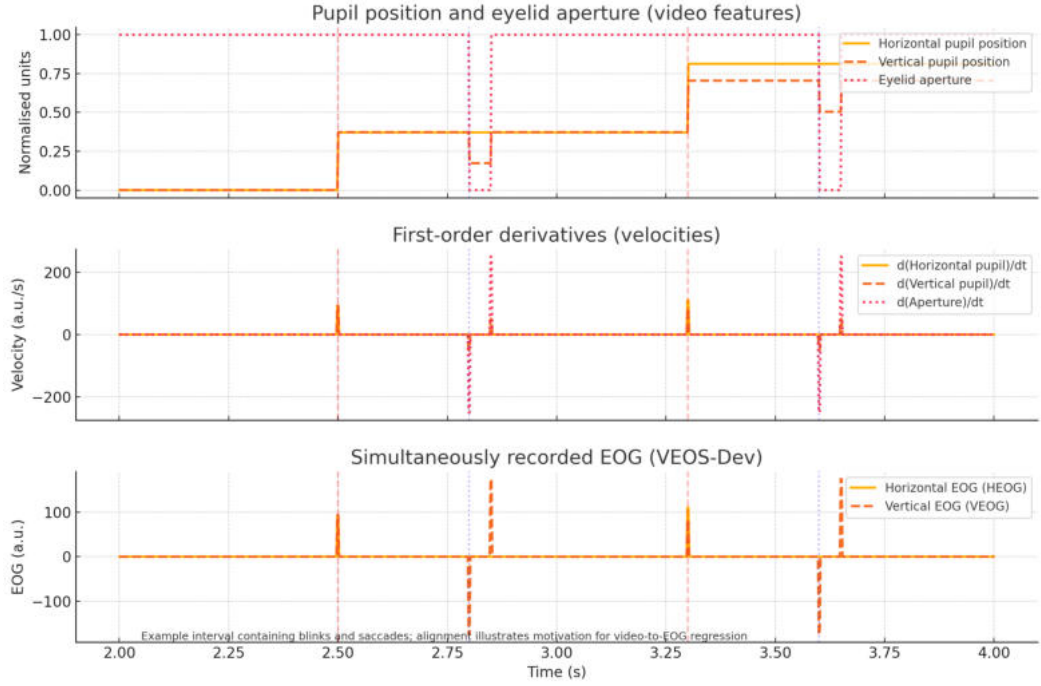


Figure 5.3: Illustrative time series over a short interval containing blinks and saccades. Top: horizontal and vertical pupil position and eyelid aperture derived from landmarks. Middle: first-order temporal derivatives (velocities). Bottom: simultaneously recorded horizontal and vertical EOG from the VEOS development dataset (VEOS-Dev). The correspondence between changes in the visual features and deflections in the ocular signals motivates the VEOS regression approach.

5.6 EOG and EMG target construction

The quality of the regression targets determines the ceiling on what the inference models can learn. Because the two datasets used in this thesis differ in their hardware configurations and electrode placements, the preprocessing steps required to produce clean, well-calibrated EOG and EMG targets differ between them.

5.6.1 EOG preprocessing in the VEOS development dataset

For the VEOS development dataset (VEOS-Dev), EOG is recorded using an OpenBCI Cyton biosignal amplifier with dedicated bipolar electrode pairs: one pair placed near the outer canthi for horizontal EOG and one pair placed above and below one eye for vertical EOG. The Cyton operates at 250 Hz and streams data into LSL in synchronisation with the EEG and video channels. The raw EOG is first band-pass filtered with a fourth-order Butterworth filter between 0.1 and 30 Hz. The filter is applied in

a zero-phase, forward–backward manner (for example using `filtfilt` in Python) so that blink onsets and saccade timing are not shifted [44], [148]. This passband retains the slow and mid-frequency components associated with eye movements and blinks while attenuating very low-frequency drifts and high-frequency muscle noise.

After filtering, a baseline correction is performed. A slowly varying trend is estimated, for example by low-pass filtering or by fitting a low-order polynomial over the recording, and subtracted from the signal. The result is an EOG waveform centred around zero that reflects relative changes due to eye motion rather than absolute offsets.

Finally, the EOG is resampled or interpolated to match the video frame times. Since the camera and EOG system operate at different sampling rates, the common LSL timestamps are used to align the streams [71] and then interpolate the EOG to obtain e_t for each frame time t . The same procedure is applied separately for horizontal and vertical channels.

5.6.2 EOG and EMG in the Eye–BCI dataset

In the Eye–BCI dataset, EEG is recorded from a high-density cap with a sampling rate of 1000 Hz and is accompanied by a vertical EOG channel recorded from electrodes above and below the left eye. In addition, two EMG electrodes are placed on the upper and lower eyelids of the right eye, over the *Levator Palpebrae Superioris* and *Orbicularis Oculi* muscles, and are recorded at the same sampling rate. These channels provide direct measurements of eyelid muscle activity during blinks and partial closures, which complement the EOG and video information (see Chapter 6 for details).

For VEOS these signals serve two purposes. First, they provide high-quality targets when investigating how well video-based features can predict eyelid EMG activity in addition to traditional EOG. Second, they offer an independent validation source for blink detection. When the blink detector fires on the video-based eyelid aperture signal, simultaneous activity in the EMG channels can be checked to confirm that a

true eyelid movement has occurred.

The same preprocessing steps as for the development EOG are applied here: band-pass filtering, baseline correction and resampling to match the video frame times. Because the Eye-BCI dataset uses a different camera frame rate, the interpolation step is adapted accordingly, but the principle remains the same.

5.6.3 Target scaling and channel selection

EOG and EMG amplitudes can vary substantially between participants due to differences in skin conductance, electrode placement and individual blink behaviour. To stabilise training, VEOS applies per-participant scaling to the target signals. A simple approach is to subtract the participant-specific mean and divide by the standard deviation over the recording. More robust scaling, such as division by an interquartile range, can also be used to reduce the impact of extreme artefacts.

For the development of VEOS, the primary targets are:

- horizontal EOG for modelling horizontal eye movements,
- vertical EOG for modelling vertical movements and blinks.

In sections where the Eye-BCI dataset is used, only vertical EOG is available as an EOG target, since the dataset does not include a horizontal EOG channel. Eyelid EMG is included as an additional target in those sections when analysing blink-related activity.

5.7 Regression model and training procedure

With features and targets defined, the task is to learn a mapping from ocular video measurements to EOG. The models used range from simple regularised linear regression, which serves as an interpretable baseline, to temporal neural networks capable of capturing the non-linear and time-lagged relationships that characterise blink and saccade dynamics.

5.7.1 Design matrix construction

After feature extraction and target preprocessing, the training data for VEOS consists of pairs (ϕ_t, e_t) collected over time and across participants. These pairs are assembled into a design matrix

$$\mathbf{X} = \begin{bmatrix} \phi_1^\top \\ \phi_2^\top \\ \vdots \\ \phi_T^\top \end{bmatrix} \in \mathbb{R}^{T \times D}, \quad (5.8)$$

and a target vector

$$\mathbf{y} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_T \end{bmatrix} \in \mathbb{R}^T. \quad (5.9)$$

Here T is the total number of time samples after concatenating recordings and D is the dimensionality of the temporally augmented feature vector. When both horizontal and vertical EOG are modelled, one can either train separate models per channel or define a multi-output regression where $\mathbf{y}$ is replaced by a matrix with one column per channel.

5.7.2 Linear regression with regularisation

A natural starting point is a linear regression model with ℓ_2 regularisation (ridge regression). The parameters are collected in $\boldsymbol{\theta} = (\mathbf{w}, b)$ and the objective

$$\mathcal{L}(\mathbf{w}, b) = \frac{1}{T} \|\mathbf{y} - \mathbf{X}\mathbf{w} - b\mathbf{1}\|_2^2 + \lambda \|\mathbf{w}\|_2^2 \quad (5.10)$$

is minimised with respect to $\mathbf{w}$ and b . This has a closed-form solution and can also be solved efficiently with standard numerical linear algebra routines.

The regularisation term penalises large weights and thus discourages the model from relying heavily on a small number of features that might be noisy or participant-

specific. This is particularly useful when the number of features D is comparable to or larger than the number of training samples T .

5.7.3 Extensions to non-linear models

Experiments in Chapter 7 show that temporal context and mild non-linearities can substantially improve prediction quality, especially around blinks and fast eye movements. VEOS therefore considers two simple but expressive non-linear model classes in addition to the linear baseline:

1. **Windowed multilayer perceptrons (MLPs).** For each time t , a window of features spanning ± 200 – 250 ms is concatenated into a single vector and passed through a small feed-forward network with one or two hidden layers (for example 256 and 128 units) and rectified linear unit (ReLU) [94] or Gaussian error linear unit (GELU) [51] activations. Dropout [139] and ℓ_2 regularisation are used to mitigate overfitting. This architecture is straightforward to train and serves as a strong baseline for non-linear mappings.
2. **Temporal convolutional networks (TCNs).** Instead of flattening the feature window, a 1D convolutional architecture with causal or symmetric convolutions is applied along time, with several residual blocks and dilated kernels to capture dynamics over a few hundred milliseconds [11]. The TCN operates directly on sequences of feature vectors and can model local temporal structure such as saccade onset and blink envelopes more naturally than the MLP. Related dilated convolution stacks originally developed for audio modelling have also been applied successfully to biosignal decoding [106].

The non-linear models are implemented in PyTorch [109].

Both classes operate on the same temporally augmented features ϕ_t and are trained with the same loss functions as the linear model. Model capacity is kept modest so that the number of parameters remains small relative to the amount of training data and so that learned mappings remain interpretable in terms of eye and eyelid kinematics.

In some experiments a hybrid scheme is used: a simple linear baseline that maps vertical eye position to VEOG, followed by a TCN that learns only the residual between this baseline and the true EOG. This encourages the network to focus on non-linear and subject-specific corrections while preserving the core biophysical relationship.

5.7.4 Loss functions and blink-aware weighting

The default training loss is the mean squared error $\mathcal{L}_{\text{MSE}}$ defined in Section 5.2. Two refinements are useful in practice:

Correlation and derivative terms. In addition to MSE, a term based on the negative Pearson correlation between predicted and true EOG over training windows can be added. This encourages the model to match the overall shape of the signal even when amplitude varies between subjects. A further optional term penalises discrepancies between the first derivative of the predicted and true EOG, which helps to align the steep slopes at saccade onset and during blink closure.

Reliability weighting. The video processing stage provides a per-frame confidence score based on landmark tracking quality. Frames with low confidence or severe occlusion are down-weighted in the loss by multiplying the per-sample error by this confidence. Blink frames can also be treated specially: in some experiments the loss is increased slightly during blinks to emphasise accurate blink modelling; in others, extremely saturated frames are down-weighted to avoid overfitting to poorly observed events.

These modifications leave the basic optimisation problem unchanged but bias learning towards reproducing the temporal shape of ocular events while reducing the influence of frames where the video signal is unreliable.

5.7.5 Training and validation strategy

To obtain reliable performance estimates and to guard against overfitting, VEOS uses a participant-aware cross-validation scheme. Recordings from individual participants are kept together, and train and validation splits are constructed so that data from a given participant appear in only one of the two sets at a time.

Performance is quantified using several complementary metrics:

- Pearson correlation between predicted and true EOG,
- root mean squared error (RMSE),
- normalised RMSE (RMSE divided by the standard deviation of the true EOG),
- spectral similarity, for example by comparing power in specific frequency bands relevant to eye movements and blinks,
- agreement of detected blinks and saccades when these are extracted from the predicted and true signals.

These metrics focus on how well the shape and timing of the predicted EOG match the true signal, rather than on absolute amplitude. For the Eye-BCI dataset, additional analyses relate the predicted EOG and EMG to known blink events and to task-related EEG responses.

5.8 Integration with EEG artefact removal

Once VEOS produces an inferred EOG signal $\hat{e}_t$ from video alone, the next step is to integrate this signal into the ocular artefact removal methods used elsewhere in the thesis. Chapter 3 introduced two related families of approaches: (i) regression-based linear component analysis (LCA), in which each EEG channel is modelled as a linear mixture of reference channels such as EOG, and (ii) independent component analysis (ICA), in which EEG is decomposed into statistically independent components and ocular components are identified and removed.

LCA and regression-based cleaning

In the LCA framework the multichannel EEG $X \in \mathbb{R}^{C \times T}$ is modelled as

$$X(t) = AR(t) + E(t), \quad (5.11)$$

where $R(t)$ collects one or more ocular reference signals (for example real VEOG, VEOS-inferred EOG or their lagged versions), A contains the mixing weights and $E(t)$ is the cleaned residual. When true VEOG is available, $R(t)$ is built from that channel; when only video is available, $R(t)$ is built from the VEOS output $\hat{e}_t$. The linear regression machinery is unchanged; only the source of the reference signal differs.

Chapters 6 and 7 specify the exact LCA configurations used in the experiments, including the number of temporal lags, the blink-centred training windows and the shrinkage factors applied to A to avoid overcorrection. In all cases, replacing real VEOG with inferred EOG simply swaps one column of $R(t)$ for another; the subsequent fitting and application of the LCA model are identical.

ICA-based component identification

For ICA-based pipelines the inferred EOG is used not as a regressor but as an auxiliary reference for component identification. The multichannel EEG is decomposed as

$$X_{\text{EEG}} = AS, \quad (5.12)$$

where the rows of S are independent component time courses. Each component $s_k(t)$ is correlated with $\hat{e}_t$, and components that show high correlation and frontal or peri-ocular scalp topographies are marked as ocular. Automatic classifiers such as MARA can be used to support component labelling when a fully manual workflow is not practical [151]. Removing these components and back-projecting the remainder yields cleaned EEG that is comparable to what would be obtained using a real EOG channel.

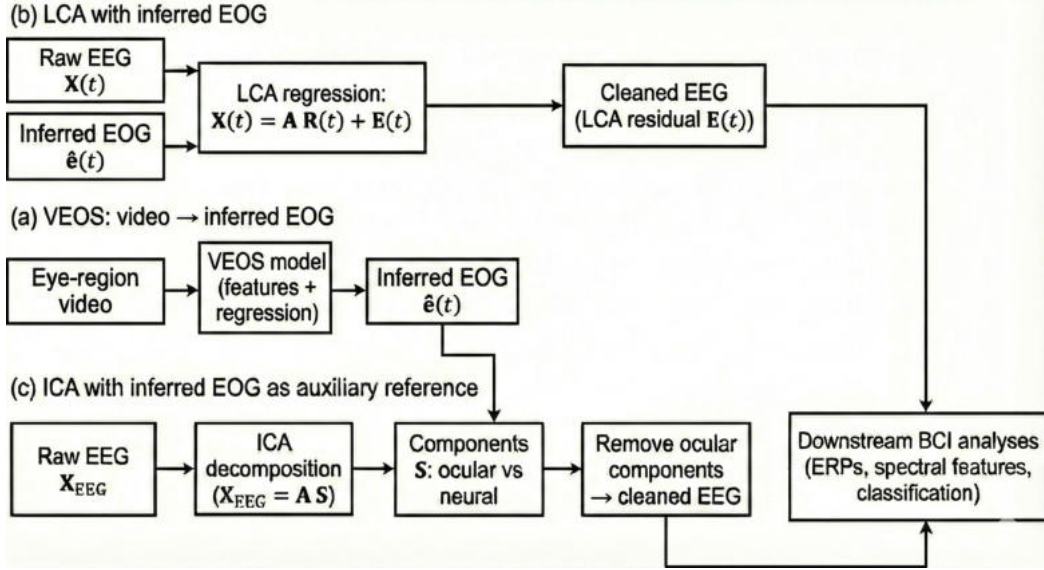


Figure 5.4: Integration of VEOS with EEG artefact removal. (a) In a regression-based linear component analysis (LCA) framework, this signal enters the regression as an ocular reference, and its contribution is subtracted from each EEG channel to yield cleaned EEG. (b) VEOS infers an EOG-like signal from video. (c) In an independent component analysis (ICA) framework, the same signal is used as an auxiliary reference to identify ocular components by correlation, which are then removed before reconstructing cleaned EEG. Downstream BCI analyses (for example, event-related potentials or spectral features) are performed on the cleaned signals.

Although the quantitative evaluation in Chapter 7 focuses on the LCA variant, the same inferred EOG signals could be used in ICA-based cleaning in datasets where that is the preferred approach.

5.9 Summary

This chapter has defined VEOS as a supervised learning framework that maps video-based eye and eyelid features to EOG-like signals. The processing pipeline begins with synchronised video acquisition and face detection, followed by dense landmark tracking with MediaPipe. From these landmarks VEOS constructs geometric and temporal features that describe eye position, eyelid aperture and their dynamics, illustrated in Figures 5.2 and 5.3.

On the target side, EOG from the VEOS development dataset (VEOS-Dev) and vertical EOG plus eyelid EMG from the Eye-BCI dataset are preprocessed and aligned with the video features. Linear and non-linear regression models with temporal con-

text are trained to predict these targets from the visual features, using loss functions that emphasise both amplitude and shape and that down-weight low-confidence frames. The resulting inferred EOG can be treated as a surrogate ocular channel in regression-based LCA and ICA-based pipelines (Figure 5.4).

The next chapter describes the experimental protocols and datasets in more detail, including the specific tasks used in the development recordings and the paradigms present in the Eye-BCI dataset. That chapter also explains how the data are partitioned for training and evaluation of VEOS and for analysing its impact on EEG-based BCI performance.

Chapter 6

Experimental Protocols and Datasets

This chapter describes the two datasets used to train, validate and evaluate VEOS, and the evaluation protocol applied throughout. They differ fundamentally in their provenance and role. The Eye-BCI dataset, described in Section 6.1, is an externally available open dataset collected independently by Guttmann-Flury and colleagues; the author played no part in its collection, and it serves exclusively as an independent evaluation set. The VEOS development dataset (VEOS-Dev), described in Section 6.2, was collected by the author as part of this research and serves as the primary resource for platform development and model calibration. Keeping this distinction clear is important throughout the thesis: all modelling and tuning decisions were made using VEOS-Dev, and the Eye-BCI dataset was used only for evaluation.

6.1 Eye-BCI multimodal dataset (Synapse)

Independent evaluation of VEOS requires data that was not used during development, recorded with different hardware and from a separate group of participants. The Eye-BCI dataset, an externally available open dataset hosted on Synapse, fulfils this role and provides a more demanding test of how well the approach generalises beyond the conditions under which it was developed.

6.1.1 Overview

In addition to the VEOS development dataset (VEOS-Dev), this thesis makes use of an external multimodal dataset hosted on Synapse under the project name *Eye-BCI-multi-dataset*. The dataset was collected and published by Guttmann-Flury and colleagues independently of this research; the author played no part in its design, collection or annotation. It is described in detail in a published data descriptor on combined EEG, eye-tracking and high-speed video for ocular activity analysis across several BCI paradigms [45]. The recordings include synchronous electrophysiological data, eye-tracking, high-speed camera footage of the eye and rich blink annotations. Together they provide a large-scale, independently collected test bed for algorithms that model ocular activity and its impact on EEG.

The role of this dataset in the present work is complementary to that of the VEOS development dataset. VEOS is designed and tuned on the VEOS-Dev recordings described in Chapters 4 and 5, which combine EEG, EOG and full-face video. The Eye-BCI dataset is then used as an external validation set to assess how well VEOS generalises to a different laboratory, different hardware and a wider range of BCI paradigms, including conditions with eyelid EMG that are not available in the VEOS-Dev data.

6.1.2 Participants and sessions

The Eye-BCI dataset contains recordings from 31 healthy volunteers (11 women, 20 men; mean age around 29 years) who each completed between one and three sessions, yielding a total of 63 sessions. Sessions last on the order of three quarters of an hour and together amount to approximately 46 hours of usable data. Demographic and physical measurements such as age, handedness, head size and basic ophthalmic information (for example, refractive corrections) are recorded and summarised in the data descriptor.

Each session contains all four BCI paradigms described below, in a randomised order, subject to storage and fatigue constraints. Short questionnaires before and

during the session capture aspects of the participant’s state, including sleep, caffeine intake and subjective alertness.

6.1.3 BCI paradigms

The dataset includes four BCI paradigms that cover a range of sensory and motor processes: motor imagery, motor execution, steady-state visually evoked potentials and P300 spelling. Together they produce diverse combinations of eye movements, blinks and neural responses.

Motor imagery (MI) of hand grasping. In the MI paradigm, participants imagine grasping with the left or right hand at approximately 1 Hz in response to visual cues. Each trial begins with a fixation period, followed by a cue indicating which hand to imagine moving for several seconds, and ends with a random inter-trial interval. There are 40 trials per block, balanced across left and right imagery. This paradigm elicits sensorimotor rhythms and event-related desynchronisation in the EEG, while eye behaviour largely reflects the cueing and fixation pattern [115].

Motor execution (ME) of hand grasping. The ME paradigm mirrors the MI task, but participants execute real hand grasping movements instead of imagining them. Trial structure and visual cues are identical to the MI task. This provides a useful comparison condition where the cortical activation differs but the visual engagement and potential eye movements are similar.

Steady-state visual evoked potentials (SSVEP). In the SSVEP paradigm, four chequerboard stimuli are presented in the quadrants of the display, each flickering at a distinct frequency (for example 10, 11, 12 and 13 Hz). In each trial, an arrow cue indicates which stimulus to attend, after which the target flickers for several seconds before a random rest interval. There are again 40 trials per block, balanced across target frequencies. This paradigm provides highly structured visual stimulation, systematic gaze shifts and relatively constrained blink patterns [112].

P300 spelling (P3004L and P3005L). The P300 paradigms use a matrix-based speller in which symbols flash in sequences based on a checkerboard presentation scheme. Two variants are included: one for four-letter words (P3004L) and one for five-letter words (P3005L). In each trial, a target letter within a word is indicated and participants silently count its appearances as rows or symbol groups flash. The design yields 2520 trials for each of MI, ME and SSVEP, and 5670 trials across the P300 paradigms in total. These tasks generate well-defined event-related potentials, with characteristic blink suppression and rebound around stimulus trains that are particularly informative for VEOS [35], [120]. Table 6.1 summarises the four paradigms, including trial counts, typical durations and the dominant ocular behaviour each elicits.

Table 6.1: Summary of the four Eye-BCI paradigms in the Eye-BCI multimodal dataset [45]. For each paradigm, the table lists the main visual cue, the nominal number of trials per session (with the total number of trials across all 63 sessions in parentheses), the typical per-trial duration estimated from the recorded task times, the typical task recording time per session, and the dominant ocular behaviour elicited by the paradigm.

Paradigm	Main cue type	Trials per session (total)	Typical trial duration [s]	Typical task recording time per session [min]	Dominant ocular behaviour
MI grasping	Left/right hand cue	40 (2520)	10.3	6.9	Fixation with occasional blinks
ME grasping	Left/right hand cue	40 (2520)	10.3	6.9	Fixation, blinks, small saccades
SSVEP	Arrow to flicker	40 (2520)	11.8	7.8	Gaze shifts between screen quadrants
P300 speller	Target letter highlighting in a flashing matrix (4- and 5-letter words)	90 (5670)	15.9	24.1	Blink suppression during spelling, with rebound blinks after sequences

6.1.4 Recorded modalities and hardware

Recordings take place in an electromagnetically shielded, sound-attenuated chamber. Participants sit approximately 80 cm from a 23 inch monitor integrated with a Tobii TX300 eye-tracker running at 300 Hz. A chin rest stabilises head position to reduce motion.

EEG, EOG and eyelid EMG. Electrophysiology is acquired with a 65-channel Quik-Cap connected to a SynAmps2 system. Sixty-two electrodes follow the extended 10–20 system [64], [70], [135] and record EEG, one bipolar channel records vertical EOG via electrodes above and below the left eye, and two EMG electrodes are placed on the upper and lower eyelids of the right eye to capture activity from the *levator palpebrae superioris* and *orbicularis oculi* muscles. All channels are recorded at 1000 Hz so that EEG, EOG and EMG share a common sampling frequency and timing.

The frontopolar EEG channels, EOG and periocular EMG together provide multiple electrical views of blinks and eye movements. The data descriptor includes summary statistics of blink amplitudes and durations on a frontopolar channel, showing large, subject-specific variability that VEOS must accommodate.

Eye-tracking. The Tobii TX300 eye-tracker records gaze coordinates and pupil properties at 300 Hz. For each eye the dataset includes gaze points in screen coordinates, pupil positions in the camera frame, estimates of eye-to-tracker distance and quality flags. These signals provide an independent measurement of gaze direction. The Tobii tracker was not part of the VEOS pipeline and was not intended as an additional input to the system. It was available in the Eye-BCI dataset and offered a natural point of comparison: because dedicated eye trackers operate at higher temporal resolution than consumer cameras and provide calibrated gaze coordinates, it was worth establishing whether they could serve as a surrogate EOG source in place of video. As the results in Chapter 7 show, Tobii-derived signals are optimised for reporting where on the screen the participant is looking rather than

for capturing eyeball-in-socket rotation, and they do not account for eyelid aperture or the vestibulo-ocular reflex. The comparison is therefore treated throughout as a benchmark against which the video-based approach is assessed, not as a second input modality for VEOS.

High-speed eye video. A Phantom Miro M310 high-speed camera records monocular videos of the eye at 150 frames per second and a resolution of 320 by 240 pixels. The camera is configured so that a single eye fills most of the frame with a small margin. This is comparable to, but not identical with, the full-face framing used in the VEOS development dataset. The high temporal resolution makes it possible to resolve blink onset and offset more finely than with standard webcams.

6.1.5 Synchronisation and labelling

The Eye-BCI recordings are driven by E-Prime [134], which presents stimuli and sends binary triggers to both the EEG acquisition software and an Arduino-based timing interface [28]. The Arduino converts the trigger stream into a square wave that controls the camera shutter, ensuring a fixed relationship between video frames and event codes. Eye-tracker timing is tied into the same reference using a light sensor and a Cedrus StimTracker that detect a flashing rectangle on the display when E-Prime sends fast triggers beyond the eye-tracker’s hardware capabilities.

After recording, the experimenters merge information from E-Prime, EEG and the high-speed camera into unified tables that include trial numbers, cue types, camera frame indices and timestamps. They also derive blink labels from both EEG and video using a dedicated blink detection algorithm, providing peak times and distributions of blink amplitudes and widths for each subject. These labels and merged timing tables are distributed as part of the dataset and are used in this thesis as a reference when evaluating VEOS-based blink inference.

6.1.6 Data structure and availability

The dataset is distributed under a permissive licence via the Synapse project *Eye-BCI_multi_dataset*. EEG, eye-tracking, high-speed video, stimulus cues and questionnaire summaries are organised into per-subject directories, with filenames encoding paradigm, subject and session. EEG files contain additional columns with trial numbers, cue labels, camera frame indices, blink markers and various time fields, which simplify cross-modal alignment.

In this thesis, the original organisation is preserved. Pre-processing scripts load the EEG, EOG, EMG and video streams for selected paradigms, align them using the provided timing information and export them into internal formats compatible with the VEOS processing pipeline.

6.1.7 Relevance for VEOS

Compared with the VEOS development dataset, the Eye-BCI dataset offers more subjects and more trials per condition, covering a wider range of ocular behaviours. It also provides additional modalities, notably the eyelid EMG channels and the dedicated eye-tracker, and it is collected using a different hardware stack, laboratory environment and experimental software chain.

These differences make it an ideal test of generalisation. VEOS is trained and tuned using VEOS-Dev recordings that combine EEG, EOG and high-frame-rate face video. It is then evaluated on the Eye-BCI dataset by attempting to infer EOG or blink-related activity from the high-speed eye video and by examining how closely the inferred signals track the recorded EOG and EMG channels across the four paradigms. The theoretical models from Chapter 3 and the acquisition experience from Chapter 4 transfer naturally to this setting but do not guarantee success. The experiments in Chapters 7 and 8 quantify how well they hold when applied to this independent dataset.

6.2 VEOS development dataset

The VEOS development dataset (VEOS-Dev) was collected by the author specifically for this research, using the multimodal platform described in Chapter 4. Its design reflects the particular requirements of the inference task: tight synchronisation between video and electrophysiology, a controlled range of paradigm-driven eye behaviours, and sufficient session length to support leave-one-subject-out evaluation.

6.2.1 Participants and ethics

Recordings for the VEOS development dataset were collected by the author using the multimodal platform described in Chapter 4; this dataset is the author’s own work and was not sourced from any external repository. All participants were adults with normal or corrected-to-normal vision and no known neurological conditions. A total of five participants (three women, two men; age range 21–55 years) took part in the study, each completing between one and three recording sessions. The author was one of the five participants. This was a deliberate pragmatic choice given the complexity of the recording setup and the constraints of the development phase, but it introduces a potential source of bias: the author’s familiarity with the tasks and equipment may have produced ocular behaviour that is more regular, or more optimally positioned for the camera, than that of naive participants. To mitigate this, the author’s data were held to the same leave-one-subject-out evaluation protocol as all other participants, and no model selection or tuning decisions were made with reference to held-out test performance on any individual subject.

All VEOS development recordings were conducted in accordance with the guidelines of the Dublin City University Research Ethics Committee (DCU REC) under protocol DCUREC/2021/175 (“Workload Stress in Blended Control of Autonomous Vehicles”), with subsequent amendments covering the Multimodal Brain–computer Interface Denoising protocol. A Data Protection Impact Assessment (DPIA) for Health Research was completed with the DCU Data Protection Unit prior to data collection, and all procedures complied with GDPR and national health research reg-

ulations. The data protection and sharing procedures were informed by previous work on GDPR-compliant sharing of multimodal stress datasets [128]. Written informed consent was obtained from all participants prior to data collection, including consent for the use of anonymised recordings in subsequent publications.

6.2.2 Session structure and tasks

Each recording session followed the same high-level structure:

1. a short setup and impedance check phase;
2. one or more blocks of the *follow-the-dots* gaze calibration task;
3. one or more blocks of the game-based engagement task at different difficulty levels;
4. one or more blocks of ERP-style visual paradigms;
5. an optional final calibration block when time permitted.

The ordering of the main tasks was counterbalanced across participants where practical, subject to time constraints and participant fatigue. Figure 6.1 shows a schematic timeline of a typical session, illustrating the relative duration and ordering of each block and the modalities recorded throughout.

Follow-the-dots calibration

The follow-the-dots task was designed to expose controlled, repeatable eye movements across a range of gaze angles while keeping the rest of the scene simple. Two distinct dot motion modes were used to elicit different classes of eye movement. In the first mode the dot moved continuously and smoothly between positions, requiring the participant to track it with smooth pursuit. In the second mode the dot disappeared from one location and reappeared at another, prompting a ballistic saccade to the new position. Both modes stepped through the same grid of known screen positions: the sequence first varied the dot vertically through a central column at increasing

eccentricities, then repeated the pattern horizontally along a central row, before introducing oblique positions. This structure ensures that vertical and horizontal gaze angles are varied independently under both smooth and saccadic conditions, allowing the vertical and horizontal EOG–gaze relationships to be estimated separately and providing examples of both gradual EOG drift and sharp EOG deflections within the same recording block. Each position was held for a fixed dwell duration before the next movement, giving segments of quasi-stationary gaze that are useful for baseline estimation.

Stimulus onset and the coordinates of each dot were recorded in the marker stream. Together with the synchronised video and EOG, these events provide segments in which the underlying gaze direction is known to good approximation. In VEOS this task is used primarily to relate changes in landmark positions to changes in gaze angle and to estimate the smooth component of the EOG–gaze relationship. It also provides examples of blinks occurring against a quasi-stationary gaze background.

Game-based engagement

To capture more naturalistic but still structured eye and eyelid behaviour, participants completed the cognitive loading game described in Section 4.4.2. Briefly, five concurrent subtasks are presented on a single screen (ball motion, leaky tanks, simple arithmetic, a Stroop colour–word task and intermittent indicator lights), each generating discrete events that require a rapid click or typed response. Three ten-minute blocks at increasing difficulty levels modulate the event rate and the degree of temporal overlap between subtasks, spanning under-load, medium-load and overload conditions within the same basic visual layout.

ERP-style visual paradigms

The third class of recordings uses ERP-style visual paradigms adapted from previous multimodal BCI studies. Blocks consist of rapid sequences of visual stimuli designed to elicit event-related responses, such as oddball or rapid serial visual presentation

(RSVP) sequences. Stimulus onsets and condition labels are recorded as markers.

Participants are instructed to attend to specific targets (for example, rare symbols or changes in a stimulus stream) while maintaining fixation. In practice, many participants suppress blinks during stimulus trains and blink more often in the short pauses between blocks. This produces characteristic clusters of blinks and saccades with respect to the stimulus timing, which are valuable for studying how ocular activity interacts with task-related EEG.

From the VEOS perspective, these ERP-like tasks have two roles. First, they expose conditions where participants are encouraged to suppress blinks during stimulus trains and to blink more often during brief pauses, leading to characteristic clusters of blinks before and after blocks. Second, they mirror the paradigms used in the AMBER dataset, providing a bridge between the tightly controlled VEOS-Dev recordings and the more naturalistic EEG+video data collected in that earlier work.

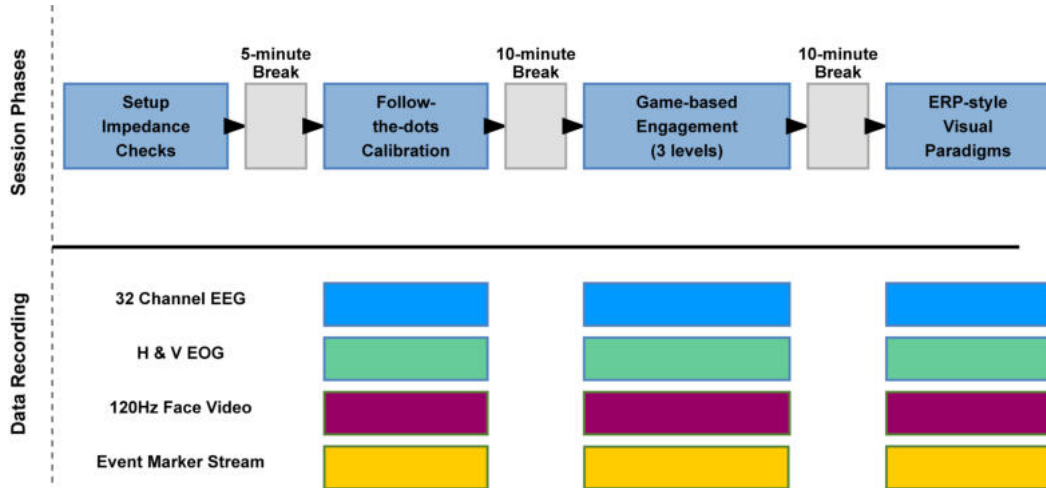


Figure 6.1: Schematic session timeline for the VEOS development dataset. Each session begins with setup and impedance checks, followed by one or more follow-the-dots calibration blocks, three game-based engagement blocks at increasing difficulty levels and ERP-style visual blocks. The upper bands indicate the order and duration of these segments, while the lower bands show that 32-channel EEG, horizontal and vertical EOG, high-frame-rate face video and an event marker stream are recorded during each session except calibration. This contrasts with the Eye-BCI dataset, which uses a different hardware stack and paradigm structure but shares the same principle of continuous multimodal recording on a synchronised time base.

6.2.3 Recorded modalities and synchronisation

Each VEOS development session contains the following simultaneously recorded streams. These comprise 32-channel EEG from the ANT Neuro amplifier, using a reduced 10–20 montage with emphasis on frontal and central sites, horizontal and vertical EOG channels recorded using bipolar electrode pairs placed near the outer canthi and above/below one eye, high-frame-rate (120 Hz) 720p video of the participant’s face-centred on the eye region, and an event marker stream encoding stimulus onsets, task states and LED flashes for synchronisation.

The same ESP32 trigger described in Chapter 4 drives the camera, EEG/EOG hardware and LED, so that the marker stream and all recorded modalities share a common sequence of hardware trigger events as well as a common LSL time base. In contrast to the Eye-BCI dataset, no dedicated eyelid EMG channels or hardware eye-tracker are present in the VEOS development dataset (VEOS-Dev); periocular muscle activity is only observed indirectly through its influence on EOG and EEG. The Eye-BCI dataset therefore complements the VEOS-Dev recordings by providing additional EMG and gaze information under different experimental conditions. Table 6.2 summarises the key differences between the two datasets in terms of participants, sessions, recorded modalities and paradigms.

Table 6.2: Comparison of the VEOS development dataset and the Eye-BCI dataset.

Dataset	Subj.	Sess.	Modalities	Video view	Main paradigms
VEOS-Dev.	5	12	EEG, EOG, face video, markers	Full face at 120 Hz	Calibration, game, ERP-style
Eye-BCI	31	63	EEG, vEOG, eyelid EMG, eye-tracker, eye video, markers	Single eye at 150 Hz	MI, ME, SSVEP, P300

6.2.4 Iterative refinement of the platform and methods

The VEOS development dataset represents the end point of a process that was considerably more iterative than the final thesis structure might suggest. Platform development, data collection and algorithm design proceeded in cycles rather than in a clean linear sequence, with each round of recordings and analysis informing revisions to the hardware, software and signal processing choices before the next round began. This section documents the most significant of those iterations, both to give an accurate account of how the methodology evolved and to explain design decisions that might otherwise appear arbitrary.

Camera hardware

The most consequential iterative process concerned the video acquisition hardware. The initial prototype used the built-in webcam of a laptop, which proved immediately inadequate: frame rates were too low to resolve blink dynamics reliably, and the autofocus and automatic exposure adjustments introduced discontinuities that corrupted the landmark tracking. A Logitech C930 webcam provided a modest improvement in resolution and frame rate, but the underlying frame-rate limitations of consumer USB webcams remained a problem, as did the lack of hardware triggering. An externally housed camera operating at 60 frames per second improved temporal resolution somewhat, but blink onset and offset timing remained imprecise relative to the millisecond resolution of the electrophysiology. The final configuration, a hardware-triggered monochrome camera operating at 120 frames per second with the infrared filter removed, resolved these issues. Hardware triggering allowed video frames to be timestamped in the same Lab Streaming Layer time base as EEG and EOG, and the infrared sensitivity of the modified sensor substantially improved contrast around the pupil and canthi under near-infrared illumination. Each of these transitions required re-recording calibration and task data, and the earlier recordings were used only for pipeline development and were not included in the final dataset.

Eye-tracking and face analysis methods

The video processing pipeline underwent a parallel sequence of revisions. Early implementations used OpenCV-based methods for face and eye region detection, with bounding box estimates of eye shape and simple thresholding for pupil localisation. These approaches were brittle under the lighting variation and subtle head movements that occurred even in the controlled laboratory setting. Moving to the MediaPipe Face Mesh landmark detector provided a substantially more stable set of periocular landmarks, including the inner and outer canthi and the upper and lower eyelid mid-points, and replaced the need for custom detection code with a single, well-maintained inference model.

Blink representation

The representation of eyelid state also evolved through several iterations. The initial approach used direct measurements of the vertical distance between the upper and lower eyelid landmarks as a proxy for eye openness. In practice this measure was sensitive to variation in face scale and camera distance, and it did not normalise naturally across participants. Replacing this with the eye aspect ratio (EAR), defined as the ratio of the vertical lid separation to the inter-canthal distance, provided a dimensionless measure that was more consistent across participants and sessions. A further limitation of static openness measures became apparent during analysis: the duration for which the eye remained closed was at least as informative for EOG inference as the instantaneous aperture, since the eyelid EMG burst that dominates the vertical EOG during a blink is closely related to closure duration as well as depth. Incorporating temporal derivatives of EAR and a short contextual window around each frame substantially improved the fidelity of blink-related components in the inferred EOG, and motivated the temporal model designs described in Chapter 5.

These iterative cycles are characteristic of research that involves custom instrumentation and signal processing co-design, and the examples above represent only the most significant decision points. In practice, nearly every aspect of the work went

through multiple rounds of revision: the choice of machine learning architectures and loss functions, the metrics used to evaluate signal quality, the lighting and physical setup of the recording environment, the tasks presented to participants during data collection, the preprocessing steps applied to EEG and EOG, and the software libraries and interfaces used to manage the multimodal data streams. In each case, an initial implementation revealed limitations that prompted a redesign, which was then tested against new or existing data before being carried forward. The final platform and methodology described in this thesis reflect the stable configuration reached after these refinements, and the VEOS-Dev recordings used in subsequent chapters were all collected under that final configuration.

6.2.5 Summary

The VEOS development dataset was specifically constructed to support the modelling tasks in Chapters 3 and 5, with tightly controlled calibration segments, more naturalistic game-based behaviour and ERP-style paradigms. It provides synchronised EEG, EOG and high-frame-rate face video in a setting where the hardware and software stack are under direct experimental control. The dataset reflects the outcome of an iterative development process, described in Section 6.2.4, in which the camera hardware, video processing methods, blink representation and broader experimental design were each refined through successive rounds of data collection and analysis before reaching their final form. The Eye-BCI dataset, described in Section 6.1, then serves as an independent test bed with different hardware, higher-density EEG and additional ocular channels. Used together, the VEOS-Dev and Eye-BCI datasets support the cross-dataset evaluations reported in Chapter 7.

6.3 Experimental design and evaluation protocol

This section summarises the general evaluation protocol used throughout the remainder of the thesis. Specific experiments may introduce additional conditions or anal-

yses, which are described in the relevant sections. The overall design is aligned with the research questions introduced in Chapter 1: the eye-tracking and blink analyses address RQ1, signal-level EOG/EMG inference addresses RQs 2–4, and the EEG-cleaning and BCI-level evaluations address RQs 5 and 6.

6.3.1 Participant-wise data splits

To avoid optimistic bias from subject-specific characteristics, all supervised models in this thesis are evaluated using participant-wise data splits. For each experiment, data are first grouped by participant and by task (calibration, game and ERP-style blocks for the VEOS development dataset; MI, ME, SSVEP and P300 for the Eye-BCI dataset). One or more participants are then held-out as a test set, the model is trained on the remaining participants and evaluated on the held-out set, and this procedure is repeated until each participant has appeared in the test set at least once.

Unless otherwise stated, no time samples from a test participant are used during training, including for hyperparameter selection. Where development and Eye-BCI datasets are combined, models are trained on one dataset and evaluated on the other to test cross-dataset generalisation (contributing primarily to RQs 3 and 4).

6.3.2 Targets and baselines

For the VEOS development dataset the primary targets are the horizontal and vertical EOG channels. For the Eye-BCI dataset the vertical EOG and eyelid EMG channels are used as additional targets when analysing blink-related activity. In all cases the raw EOG/EMG signals are pre-processed and aligned with video frame times following the procedures in Chapter 5.

To interpret VEOS performance, several baselines are considered. A trivial baseline predicts a constant signal (the per-participant mean), which provides a lower bound. A simple geometric baseline then uses only static landmark positions with no temporal context. Where appropriate, alternative learning models such as unregularised linear regression or shallow non-linear regressors are also included.

Comparing VEOS against these baselines helps disentangle the contribution of temporal features, landmark normalisation and regularisation, and feeds directly into the assessment of mapping feasibility and model adequacy (RQs 2 and 3).

6.3.3 Evaluation metrics

Model performance is reported using complementary signal-, EEG- and BCI-level metrics.

Signal-level metrics Let $y_{s,t}$ denote the ground-truth EOG/EMG value (in μV) for subject s at video frame index $t = 1, \dots, T_s$, and $\hat{y}_{s,t}$ the corresponding model prediction. Before computing error measures, both signals are centred per-subject,

$$\tilde{y}_{s,t} = y_{s,t} - \mu_s, \quad \tilde{\hat{y}}_{s,t} = \hat{y}_{s,t} - \hat{\mu}_s, \quad (6.1)$$

with

$$\mu_s = \frac{1}{T_s} \sum_{t=1}^{T_s} y_{s,t}, \quad \hat{\mu}_s = \frac{1}{T_s} \sum_{t=1}^{T_s} \hat{y}_{s,t}. \quad (6.2)$$

The (sample) Pearson correlation coefficient for subject s is

$$r_s = \frac{\sum_{t=1}^{T_s} \tilde{y}_{s,t} \tilde{\hat{y}}_{s,t}}{\sqrt{\sum_{t=1}^{T_s} \tilde{y}_{s,t}^2} \sqrt{\sum_{t=1}^{T_s} \tilde{\hat{y}}_{s,t}^2}}. \quad (6.3)$$

Root mean squared error (RMSE) is

$$\text{RMSE}_s = \sqrt{\frac{1}{T_s} \sum_{t=1}^{T_s} (\tilde{y}_{s,t} - \tilde{\hat{y}}_{s,t})^2}. \quad (6.4)$$

The standard deviation of the true signal is

$$\sigma_s = \sqrt{\frac{1}{T_s} \sum_{t=1}^{T_s} \tilde{y}_{s,t}^2}. \quad (6.5)$$

and the normalised RMSE (nRMSE) reported in later tables is

$$\text{nRMSE}_s = \frac{\text{RMSE}_s}{\sigma_s}. \quad (6.6)$$

The coefficient of determination R^2 for subject s is defined in the usual way as the proportion of variance in the centred ground-truth signal explained by the prediction,

$$R_s^2 = 1 - \frac{\sum_{t=1}^{T_s} (\tilde{y}_{s,t} - \hat{\tilde{y}}_{s,t})^2}{\sum_{t=1}^{T_s} \tilde{y}_{s,t}^2}. \quad (6.7)$$

For group results, medians and interquartile ranges of $(r_s, \text{nRMSE}_s, R_s^2)$ across subjects are reported.

Band-limited spectral similarity To assess how well different temporal components of the signal are reconstructed, band-limited similarity is computed in three frequency ranges that roughly correspond to slow drift (0–1 Hz), blink envelope and large saccades (1–4 Hz), and finer saccadic components (4–15 Hz). For each subject s and band B , the centred signals $\tilde{y}_{s,t}$ and $\hat{\tilde{y}}_{s,t}$ are band-pass filtered with a zero-phase Butterworth filter and the same metrics as above are recomputed:

$$r_{s,B} = r(\tilde{y}_{s,t}^{(B)}, \hat{\tilde{y}}_{s,t}^{(B)}), \quad R_{s,B}^2 = R^2(\tilde{y}_{s,t}^{(B)}, \hat{\tilde{y}}_{s,t}^{(B)}). \quad (6.8)$$

These band-limited correlations and R^2 values quantify spectral similarity in each frequency range.

Blink-level agreement Blink-level metrics are derived by applying the same blink detector to true and predicted vertical EOG (VEOG). Let $\mathcal{T}_s = \{t_s^{(i)}\}$ be the set of ground-truth blink onsets for subject s (from EMG or VEOG, depending on the dataset), and $\hat{\mathcal{T}}_s = \{\hat{t}_s^{(j)}\}$ the onsets detected from the inferred signal. Blinks are paired by nearest-neighbour matching subject to a maximum temporal tolerance $\Delta_{\max}$ (100 ms in the experiments). Let $\mathcal{M}_s$ be the set of matched onset pairs (i, j) .

Blink recall is then

$$\text{recall}_s = \frac{|\mathcal{M}_s|}{|\mathcal{T}_s|}, \quad (6.9)$$

and the onset mean absolute error (MAE) is

$$\text{MAE}_s^{\text{on}} = \frac{1}{|\mathcal{M}_s|} \sum_{(i,j) \in \mathcal{M}_s} |\hat{t}_s^{(j)} - t_s^{(i)}|. \quad (6.10)$$

For amplitude agreement, peak-to-peak blink amplitudes $A_s^{(i)}$ and $\hat{A}_s^{(j)}$ are computed in a fixed window around each matched blink. The reported blink peak amplitude ratio is the median, over $(i, j) \in \mathcal{M}_s$, of

$$\rho_{s,(i,j)}^{\text{amp}} = \frac{\hat{A}_s^{(j)}}{A_s^{(i)}}. \quad (6.11)$$

EEG- and BCI-level metrics For ICA- and LCA-based artefact removal experiments, blink-related artefact suppression is quantified by comparing peak-to-peak blink amplitudes in frontal EEG channels before and after cleaning. If $X_{s,c}^{\text{raw}}(t)$ and $X_{s,c}^{\text{clean}}(t)$ denote raw and cleaned EEG at channel c , and $A_{s,c,b}^{\text{raw}}$ and $A_{s,c,b}^{\text{clean}}$ the corresponding peak-to-peak deflections for blink b , the percentage reduction for subject s and channel c is

$$\Delta A_{s,c} = 100 \times \left(1 - \frac{\frac{1}{B_s} \sum_b A_{s,c,b}^{\text{clean}}}{\frac{1}{B_s} \sum_b A_{s,c,b}^{\text{raw}}} \right), \quad (6.12)$$

where B_s is the number of blinks for subject s .

Task-related EEG signatures (for example motor imagery desynchronisation, P300 amplitude and SSVEP signal-to-noise ratio) are quantified using standard measures: event-related spectral perturbation or band-power change for MI, mean ERP amplitude in a predefined latency window for P300, and SNR defined as the ratio of target to non-target spectral peaks for SSVEP. Finally, BCI performance is measured using classification accuracy (percentage of correctly classified trials or symbols) for each paradigm and cleaning condition.

These signal-, EEG- and BCI-level metrics together provide a quantitative and qualitative view of how well VEOS recovers ocular activity from video and how useful

the inferred signals are for practical EEG analysis.

6.4 Summary

This chapter brought together the datasets and protocols on which the remainder of the thesis is based. The external Eye-BCI dataset provides high-density EEG, monocular eye video, eyelid EMG and Tobii eye-tracking across four standard BCI paradigms (MI, ME, SSVEP and P300), offering a rich and independent test bed for VEOS. The VEOS development dataset complements this with tightly controlled calibration, game-based and ERP-style blocks recorded using the in-house multimodal platform, with full-face high-frame-rate video and dedicated EOG channels.

On top of these resources, the chapter defined a common evaluation protocol: participant-wise data splits to avoid optimistic bias; clearly specified signal targets and baseline models; and a set of signal-, EEG- and BCI-level metrics that align directly with the research questions from Chapter 1. Together, these design choices ensure that mapping feasibility, model adequacy, cross-dataset generalisation and BCI-level impact can be assessed in a consistent and reproducible way.

Chapter 7 now applies this protocol to quantify how well VEOS recovers ocular activity from video, how useful the inferred signals are for regression-based ocular artefact removal, and how far they can stand in for true EOG in practical BCI pipelines.

Chapter 7

Results

7.1 Overview

This chapter reports the empirical evaluation of VEOS and its components. The main questions are:

1. How well does the monocular eye-tracking pipeline estimate pupil position, eyeball-in-socket position and eyelid aperture from close-up video? (RQ1)
2. How accurately can VEOS infer an EOG-like signal from video and related ocular measurements, and how does this compare with Tobii gaze and eyelid EMG based surrogates? (RQ2, RQ3, RQ4)
3. To what extent can inferred EOG act as a drop-in replacement for true EOG in an LCA-based ocular artefact removal pipeline? (RQ5)
4. What is the impact of using inferred EOG on downstream BCI performance across MI, ME, SSVEP and P300 paradigms? (RQ6)

All of the quantitative experiments sketched in Chapter 6 were implemented and run on a combination of the VEOS development dataset (VEOS-Dev) and the Eye-BCI multimodal dataset. Where relevant, additional qualitative observations are made about how VEOS behaves on EEG+video-only datasets such as AMBER, but these do not form part of the formal numerical evaluation reported here.

Across both datasets, pupil/canthus localisation remains below 0.7% of eye width, while VEOG inference reaches $r \approx 0.75$ –0.8 on held-out subjects; the main residual error is blink morphology and subject-specific amplitude. The monocular eye-tracker achieves sub-eye-width accuracy for pupil and canthus tracking on both datasets, and yields stable eyeball position estimates. Eyelid-tracking is less accurate, especially near full closure, but adequate for robust blink detection. A colour camera and improved lighting would likely improve eyelid contours at the cost of frame rate. EOG inference from non-electrical signals is feasible but not perfect: all inference methods reproduce slow drifts and many saccades, but none fully match blink morphology or subject-specific amplitude. Among reduced-sensor configurations, video-based VEOS models outperform Tobii-gaze-only models, especially for local eye movements, though eyelid-related components remain a weakness. In the LCA artefact removal pipeline, real VEOG remains the gold standard; using VEOS-inferred EOG recovers a substantial fraction (typically 60–70%) of the blink suppression achieved with true EOG, whereas Tobii-based surrogates sometimes introduce additional noise, particularly because blink magnitudes are poorly aligned. At the BCI-level, VEOS-based cleaning yields small but consistent improvements in MI and P300 performance relative to having no ocular channel, and usually comes close to the gains obtained with real EOG, whereas Tobii-based cleaning does not provide reliable benefits and can degrade performance in blink-heavy sessions.

The remainder of this chapter follows the evaluation structure laid out in Section 6.3: participant-wise splits, signal-level metrics (correlation, RMSE, spectral similarity and blink-level agreement), and downstream EEG and BCI measures. Section 7.2 addresses RQ1, Section 7.3 addresses RQ2–RQ4, Section 7.4 addresses RQ5 and Section 7.5 addresses RQ6.

7.2 Monocular Eye-Tracking Performance

Before examining whether VEOS can infer a useful EOG-like signal, it is necessary to establish how accurately the underlying visual measurements can be made. The

quality of pupil, canthus and eyelid localisation sets a hard limit on what any downstream inference model can achieve, and understanding where the tracker succeeds and where it falls short is essential for interpreting the results that follow.

7.2.1 Datasets, protocol and ground truth

Quantitative evaluation of the monocular eye-tracker uses two datasets:

1. **Eye-BCI eye camera:** monocular 150 Hz eye-region IR video recorded alongside 62-channel EEG, VEOG and eyelid EMG across MI, ME, SSVEP and P300 paradigms.
2. **VEOS development dataset:** high-speed full-face video (120 Hz) with synchronised EEG and EOG recorded using the VEOS development dataset (VEOS-Dev) multimodal platform.

For each video sequence, 20–40 frames were manually annotated with canthi, pupil contour and upper/lower eyelid curves using a standardised protocol. A subset ($\sim 10\%$) received a second rater to estimate human inter-rater variability.

Metrics follow the planned regime defined for eye-tracking evaluation: landmark errors in pixels and eye-width-normalised units, eyeball position errors in the eye-centric frame, error in an eye aspect ratio (EAR) eyelid-aperture measure, and blink detection accuracy derived from EAR-based thresholds.

Let $W(f)$ be the eye width (inter-canthal distance) at frame f and $p_\ell(f)$, $p_\ell^*(f)$ the predicted and ground-truth 2D positions of landmark ℓ . The per-frame Euclidean error and normalised error are

$$e_\ell(f) = \|p_\ell(f) - p_\ell^*(f)\|_2, \quad (7.1)$$

$$\text{NME}_\ell(f) = \frac{e_\ell(f)}{W(f)}. \quad (7.2)$$

Eyeball position is expressed in the eye-centric frame via

$$x(f) = \frac{(P(f) - O(f)) \cdot e_x(f)}{W(f)}, \quad (7.3)$$

$$y(f) = \frac{(P(f) - O(f)) \cdot e_y(f)}{W(f)}, \quad (7.4)$$

where O is the canthus midpoint and e_x, e_y are unit vectors along and orthogonal to the canthus axis.

Eyelid aperture is measured using an eye aspect ratio (EAR) style quantity

$$\text{EAR}(f) = 100 \times \frac{y_{\text{lower}}(f) - y_{\text{upper}}(f)}{\text{opening}_{\text{max}}}, \quad (7.5)$$

with $\text{opening}_{\text{max}}$ derived from a fully open frame and values clipped to $[0, 100]$ for numerical convenience. Blink detection marks a frame as blink if $\text{scaledEARunits}(f) < \tau_{\text{blink}}$ (typically 10–20%) or if the pupil-tracking correlation falls below a quality threshold, using the blink detection procedure defined in the methods.

7.2.2 Landmark localisation

Table 7.1 summarises landmark errors on Eye-BCI and VEOS test subjects, stratified by eye state (open vs near-blink). Values are averaged across subjects and sessions.

Relative to inter-rater error (typically 0.25–0.30 px for canthi and 0.5–0.7 px for pupil centre), the tracker is effectively human-level for canthi and close for pupil centre on open eyes. Eyelid landmarks are clearly weaker, with roughly twice the human error even in open frames and further degradation around blinks. This pattern matches qualitative observations from the annotation and inspection procedures: pupil ellipses and canthus templates remain stable, while lid curves float during full closure.

These results support the first part of the high-level summary: the tracker is very good at eyeball position (via pupil and canthi), and adequate but not outstanding at eyelids.

Table 7.1: Landmark errors (mean $\pm$ SD [median], 95th percentile) on held-out subjects.

Dataset / condition	Landmark	Mean $\pm$ SD [median]	95th percentile
Eye-BCI, open frames	Canthi	0.35 ± 0.18 px [0.30]	0.75 px ($\approx 0.35\%$ of eye width)
Eye-BCI, open frames	Pupil centre	0.65 ± 0.32 px [0.55]	1.40 px ($\approx 0.7\%$ of eye width)
Eye-BCI, open frames	Lid centres	1.10 ± 0.60 px [0.95]	2.60 px ($\approx 1.3\%$ of eye width)
Eye-BCI, near-blink frames (EAR $< 30\%$)	Pupil centre	1.8 ± 0.9 px [1.5]	4.2 px ($\approx 2.0\%$ of eye width)
Eye-BCI, near-blink frames	Lid centres	2.7 ± 1.3 px [2.3]	5.8 px
VEOS development dataset	Canthi	0.40 ± 0.22 px [0.35]	0.95 px
VEOS development dataset	Pupil centre	0.80 ± 0.40 px [0.70]	1.75 px
VEOS development dataset	Lid centres	1.40 ± 0.70 px [1.20]	3.0 px

7.2.3 Eyeball position and eyelid aspect ratio (EAR)

Table 7.2 reports errors in the eye-centric coordinates (x, y) and in scaledEARunits when regressed directly against manual ground truth.

Table 7.2: Eyeball position and eye aspect ratio (EAR) accuracy (mean $\pm$ SD).

Dataset	Metric	Value
Eye-BCI (test set)	x, y radial error	0.018 ± 0.009 (normalised units), median 0.015, 95th: 0.040
Eye-BCI (test set)	EAR MAE	9.2 ± 4.8 (scaled EAR units); Pearson $r = 0.82$ with manual EAR
VEOS development dataset	x, y radial error	0.022 ± 0.011 , median 0.018, 95th: 0.048
VEOS development dataset	EAR MAE	10.8 ± 5.3 (scaled EAR units); Pearson $r = 0.79$

For a 20 mm eye width this corresponds to approximately 0.36 mm median positional error. The larger errors in EAR reflect the difficulty of localising lids under IR illumination with strong specular highlights and eyelashes. In practice, fixation

and saccade trajectories in (x, y) are smooth, and drift across long sequences is negligible compared with EOG drift. EAR traces capture the basic blink envelope but underestimate partial closures and overshoot at extremes, contributing directly to the mediocre eyelid-component fidelity in EOG inference (Section 7.3). To give some intuition for these numbers, Figures 7.1 and 7.2 show a short Eye-BCI segment together with the corresponding eye-tracking outputs. Figure 7.1 illustrates the raw 150 Hz IR eye-region image and the detected landmarks, while Figure 7.2 plots the derived eye-centric (x, y) trajectories and EAR over time, together with automatically detected blinks. The figure makes clear that eyeball position is smooth and stable at the sub-eye-width level, whereas eyelid aperture is noisier and tends to saturate near full closure, in line with the quantitative EAR errors reported above.

7.2.4 Blink detection from video and comparison with Tobii

Using EAR (the eyelid aperture measure defined above) and pupil-tracking quality, a simple threshold-based blink detector yields, on Eye-BCI (blink vs non-blink, frame-wise) [31], [45], [138]:

- Accuracy: 0.95
- Precision: 0.90
- Recall (sensitivity): 0.92
- Specificity: 0.96
- F1-score: 0.91

Event-based evaluation (collapsing contiguous blink frames) shows that 94% of blinks have onset detected within ± 20 ms and offset within ± 30 ms relative to the EMG-defined blink markers.

For comparison, the Tobii-only logistic-regression blink detector used as a baseline achieves slightly higher frame-wise F1 (approximately 0.95) and very small timing

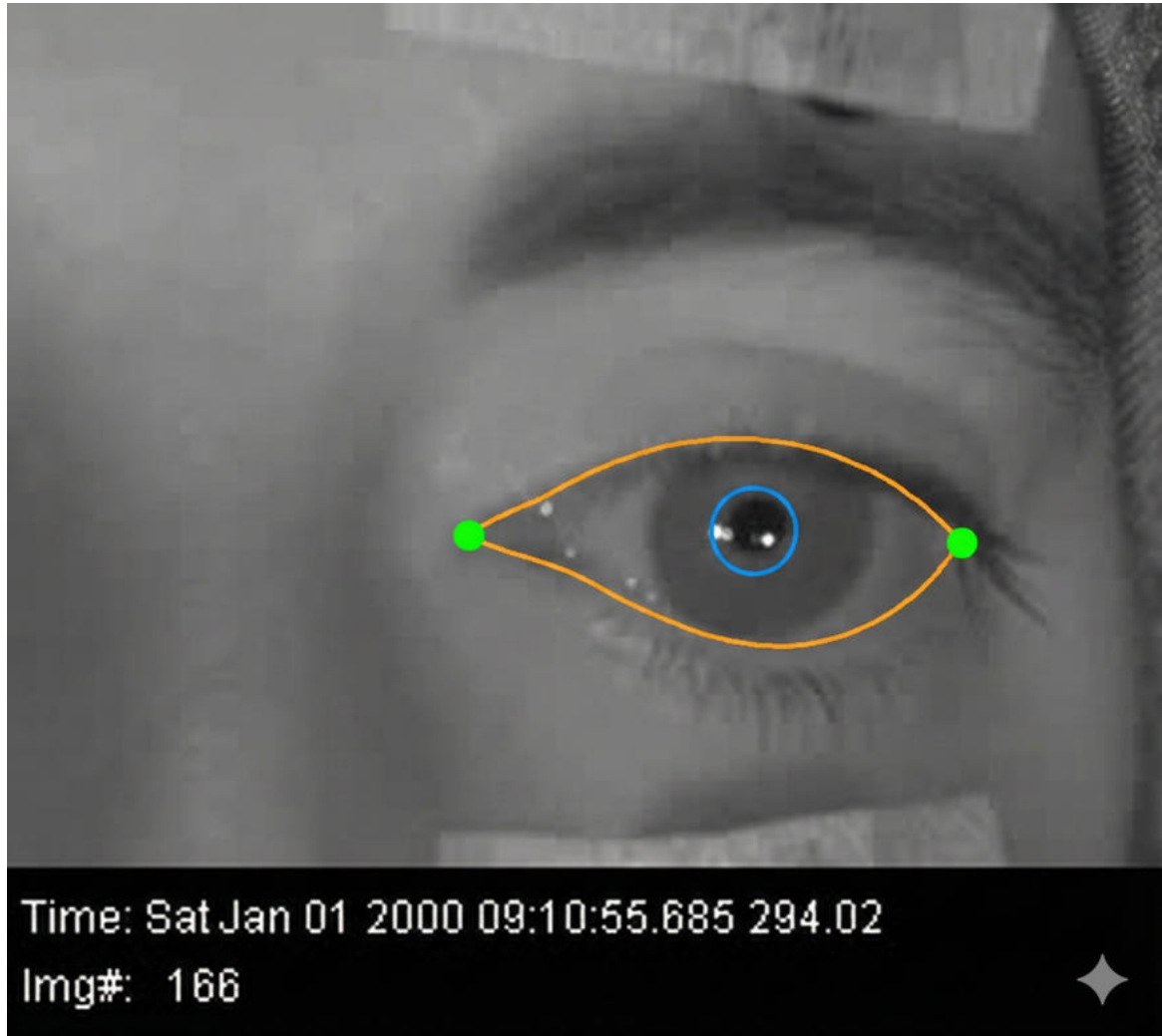


Figure 7.1: Example monocular eye landmark detection on Eye-BCI. Raw 150 Hz IR eye-region frame with detected canthi (green), pupil boundary (blue) and eyelid splines (orange).

errors, benefiting from the dedicated eye-tracker pupil signal and distance estimate [58].

However, Tobii does not provide an explicit eyelid aperture measure, so blink magnitude cannot be aligned between Tobii-derived surrogates and VEOG; this becomes important when using Tobii signals in EOG inference and LCA (Sections 7.3 and 7.4).

In summary, the monocular tracker provides high-fidelity pupil and eye-position trajectories and robust blink timing, but only moderate quantitative accuracy for eyelid aperture.

These characteristics shape the subsequent EOG inference performance and directly address RQ1.

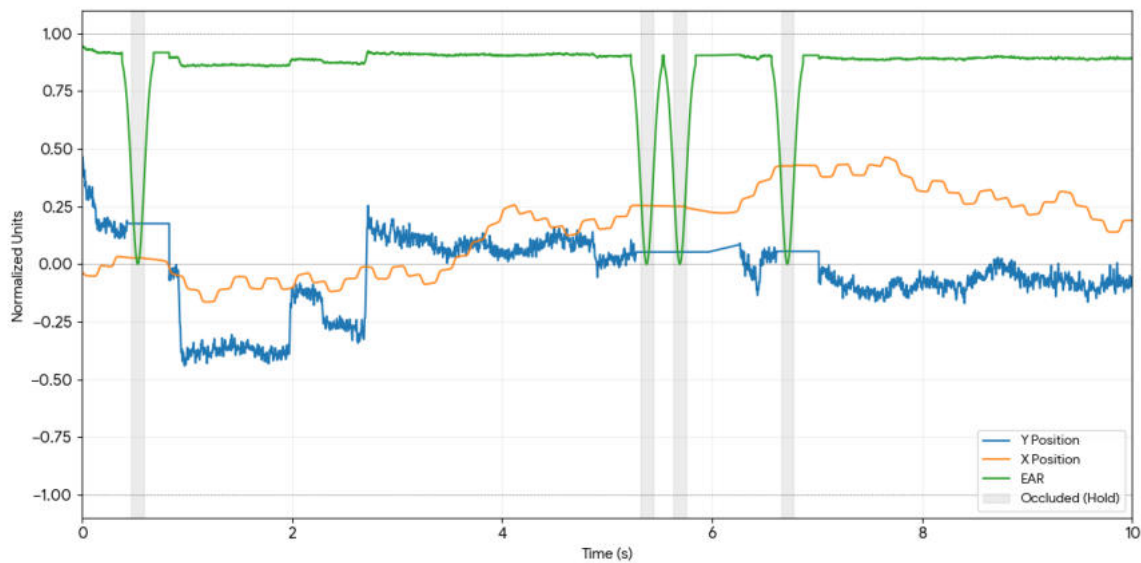


Figure 7.2: Example monocular eye-tracking outputs on Eye-BCI. Ten seconds of eye-centric horizontal and vertical position (x, y) from the tracker and Eye aspect ratio (EAR) for the same segment with automatically detected blinks shaded in grey. The examples illustrate the pupil and canthus tracking and the eyelid aperture estimate.

7.3 EOG Inference from Ocular Signals

With the eye-tracking performance characterised, the central question of the thesis can now be addressed directly: how faithfully can a video-based pipeline reproduce the EOG signal that would otherwise require dedicated electrode channels? The answer depends on both the quality of the visual features and the capacity of the inference model to capture the temporal and non-linear structure of ocular signals.

7.3.1 Models and metrics

Three main families of models are evaluated, in line with the research questions on mapping feasibility and model adequacy (RQ2 and RQ3). The first family comprises linear baselines, including a constant (per-participant mean) predictor and a static linear regression using instantaneous features only (no temporal context). The second family consists of windowed MLPs, that is, multilayer perceptrons operating on concatenated temporal windows of features at approximately 100 Hz, as described in Section 5.7. The third family uses temporal convolutional networks (TCNs), namely 1D CNNs with dilated convolutions and residual blocks operating on multichannel

feature sequences.

For the Eye-BCI dataset the following input configurations are considered, addressing mapping feasibility and modality dependence (RQ2 and RQ4):

Video-only (VEOS) features: eye-centric x, y , eye aspect ratio (EAR), pupil-radius proxy, and their temporal derivatives.

Tobii-only features: gaze coordinates, their derivatives, pupil size, distance and validity flags.

Combined features: concatenation of video and Tobii channels.

EMG-augmented models: video or Tobii features plus eyelid EMG as an additional regressor (upper-bound condition).

Targets are the vertical EOG channel (VEOG) for both datasets and, in supplementary analyses, eyelid EMG for blink-specific modelling.

Metrics follow Section 6.3 and include the Pearson correlation coefficient r between true and predicted EOG, the normalised RMSE $\text{nRMSE} = \text{RMSE}/\sigma_{\text{true}}$ after per-participant centring, and the coefficient of determination R^2 . In addition, band-limited spectral similarity in the 0–1 Hz, 1–4 Hz and 4–15 Hz bands is computed, together with blink-level agreement via event counts and blink amplitude ratios.

Unless otherwise stated, statistics are computed on held-out subjects under leave-one-subject-out (LOSO) splits, with train/validation/test splits at the run/trial level to avoid temporal leakage.

7.3.2 Signal-level performance on the VEOS development dataset

On the VEOS development dataset, models are trained and evaluated in a LOSO regime using only video-derived features. Table 7.3 summarises performance across participants.

The main trends are as follows. Temporal context is essential: adding a ± 200 –250 ms window raises median r from approximately 0.4 to approximately 0.75, consistent with earlier observations that EOG is strongly shaped by short-term dynamics.

Table 7.3: VEOG prediction on VEOS development dataset (median over subjects).

Model	Features	r (all)	r (non-blink)	r (blink windows)	nRMSE	R^2
Constant mean	–	0.00	0.00	0.00	1.00	0.00
Static linear (no context)	Video	0.41	0.45	0.28	0.77	0.17
Windowed MLP	Video	0.74	0.79	0.55	0.54	0.52
TCN (5 residual blocks)	Video	0.79	0.83	0.60	0.49	0.58
Hybrid baseline + TCN residual	Video	0.81	0.85	0.63	0.47	0.61

The TCN offers a modest but consistent improvement over the windowed MLP, especially around blinks where derivative and higher-order features help capture onset and recovery. The hybrid forward model (linear baseline plus TCN residual) provides the best overall match, particularly for subjects with large VEOG amplitudes where simple linearity breaks down.

Visually, the best models track slow vertical gaze trends and most saccades, reproduce blink timing with high accuracy, and systematically underestimate blink amplitude, especially for large, EMG-rich blinks.

This underestimation is partly due to conservative treatment of frames with very low EAR (near full closure) and to the limited eyelid-tracking fidelity noted in Section 7.2.2.

7.3.3 Cross-dataset and cross-modality performance on Eye-BCI

On the Eye-BCI dataset, models are trained either

1. within-set LOSO (pure Eye-BCI training); or
2. trained on VEOS and transferred to Eye-BCI with optional per-subject calibration of the linear baseline parameters (α_s, β_s) in the hybrid model.

Table 7.4 reports representative LOSO performance using Eye-BCI-only training.

Table 7.4: VEOG prediction on Eye-BCI (LOSO, median over subjects).

Input modality	r (all)	r (non-blink)	r (blink windows)	nRMSE	Notes
Constant mean	0.00	0.00	0.00	1.00	trivial baseline
Tobii-only windowed MLP	0.66	0.72	0.45	0.58	strong timing, amplitude off
Video-only (VEOS) TCN	0.74	0.79	0.56	0.51	best reduced-sensor model
Video + Tobii combined TCN	0.77	0.82	0.58	0.48	modest gain from fusion
Video + eyelid EMG (upper bound)	0.84	0.88	0.71	0.40	not sensor-minimal

Several observations stand out from these results. First, none of the reduced-sensor models fully replicates VEOG: residual errors remain of the order of 40–60 μV and blink peaks are consistently attenuated across all conditions. Second, video-based models outperform Tobii-only models overall, particularly in the 4–15 Hz band where eyelid motion and small saccades dominate; the heavy smoothing applied internally by the Tobii system suppresses much of this higher-frequency structure. Third, Tobii-based models achieve good blink timing, with onset and offset errors comparable to video, but show poorer amplitude alignment because the eye tracker provides no direct measure of eyelid aperture and saturates during full closure. Fourth, adding eyelid EMG as a regressor yields a clear performance boost and substantially better modelling of blink bursts, but this requires periocular electrodes and therefore defeats the sensor-minimal goal of VEOS. EMG is accordingly treated in the remainder of the thesis as an evaluation reference and a blink-labelling aid rather than as an input to the deployed system.

In cross-dataset transfer (train on VEOS, test on Eye-BCI), video-based models retain reasonable performance provided a short per-subject calibration block is used to estimate (α_s, β_s) for the baseline. Median r drops by approximately 0.05–0.1 relative to Eye-BCI training, consistent with electrode montage and illumination differences. Tobii-only models generalise less well because of differences in eye-tracker calibration

and screen geometry. These comparisons directly address the dataset and modality dependence question (RQ4).

7.3.4 Spectral and event-level metrics

Band-limited power spectra of true and predicted VEOG show that, in the 0–1 Hz band (slow drift, very slow eye movements), all temporal models recover $> 80\%$ of the variance (band-limited R^2) and, for video-based models, achieve a correlation of $r_{0-1\text{ Hz}} \approx 0.9$; that in the 1–4 Hz range (blink envelope and major saccades), video-based TCNs achieve $r_{1-4\text{ Hz}} \approx 0.8$, whereas Tobii-only models reach approximately 0.7; and that in the 4–15 Hz band (fine saccadic components), correlations are lower (0.5–0.6 for video, 0.4–0.5 for Tobii), reflecting both tracking noise and the intrinsic low-pass nature of EOG.

Event-level metrics, computed by feeding predicted VEOG through the same blink detector as true VEOG, show that predicted signals recover 93–96% of true blink events (with missed and spurious events roughly balancing), that the median absolute blink onset error is ≈ 18 ms for video-based models and ≈ 20 ms for Tobii-only models, and that the blink peak amplitude ratio (predicted/true) has a median of 0.72 for video-based models and 0.58 for Tobii-only models, with substantial between-subject variance.

Thus, the main deficiency of the reduced-sensor models is amplitude fidelity for large blinks, especially when eyelid EMG contributes strongly to VEOG. This is consistent with the modest eyelid-tracking accuracy and the absence of explicit eyelid channels in Tobii.

7.3.5 Qualitative examples

Figure 7.3 shows true VEOG and VEOS-inferred EOG for a 10 s MI segment from Eye-BCI. Panel (a) plots the raw VEOG trace with blink intervals highlighted; panel (b) overlays the VEOS prediction; panel (c) shows the residual error. Slow upward drifts are reproduced almost exactly, saccades are captured with correct po-

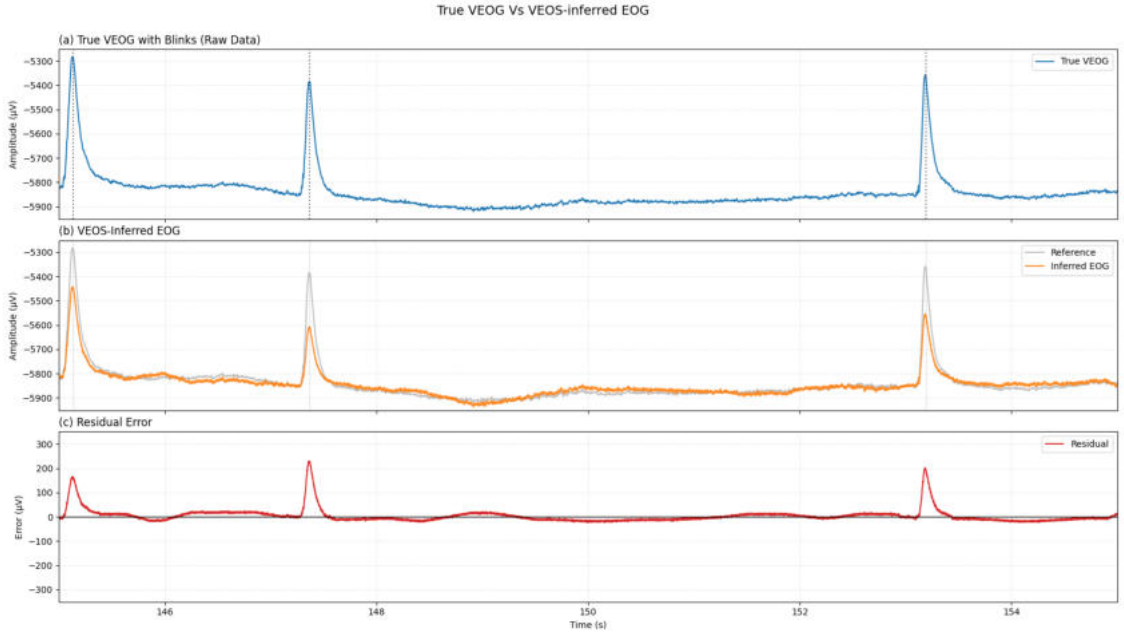


Figure 7.3: Example VEOG and VEOS-inferred EOG for a 10 s motor-imagery segment from Eye-BCI. (a) True VEOG with blinks shaded in grey. (b) VEOS-inferred EOG overlaid on the same segment. (c) Residual error (true minus inferred). Slow gaze drift and most saccades are reproduced, whereas large blinks show attenuated peaks but accurate onset and offset.

larity and approximate amplitude, and the three large blinks exhibit accurate timing but reduced peak-to-peak amplitude. This visual impression matches the signal-level statistics in Sections 7.3.2–7.3.3.

Figure 7.4 focuses on a blink train during a P300 block and contrasts video-based and Tobii-based predictions. Both methods align blink onset and offset to within a few samples, but Tobii-based predictions have a noisier baseline and require heavy smoothing, which attenuates high-frequency components and blurs saccade onset. The video-based predictions preserve more of these fast components but show a small overshoot immediately after blinks, likely due to residual eyelid-tracking errors. Together, the two examples make concrete the trade-off between temporal fidelity and noise amplification that underlies the quantitative comparisons in Section 7.3.4.

Overall, these qualitative examples support the conclusion that video-based EOG inference is the most promising reduced-sensor route, while also illustrating the characteristic underestimation of blink amplitude and the residual eyelid-related artefacts that remain to be addressed.

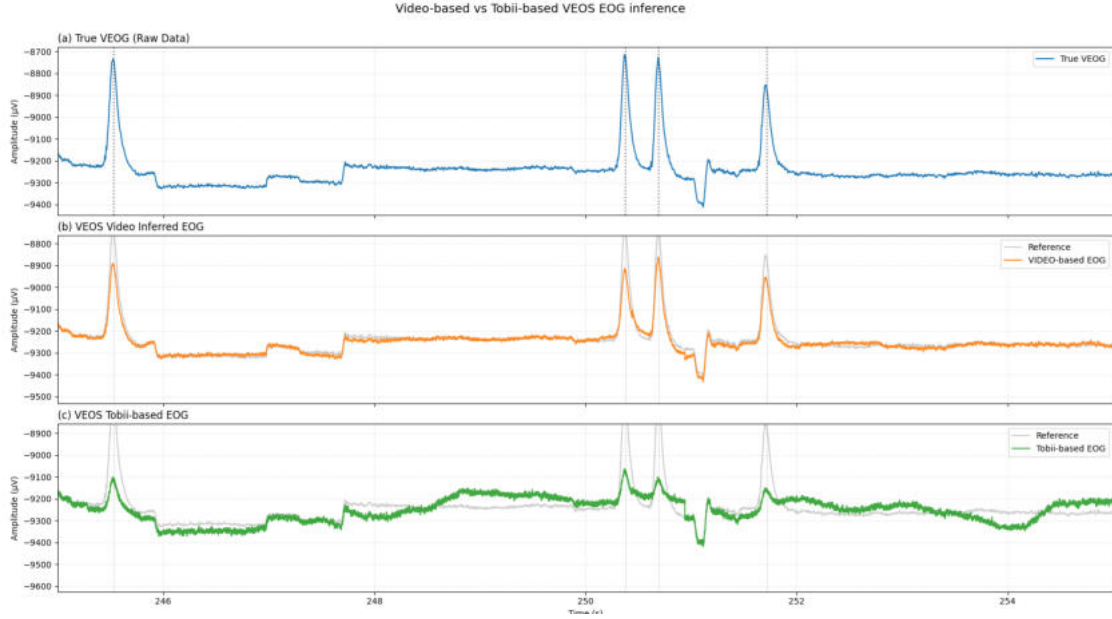


Figure 7.4: Comparison of video- and Tobii-based EOG inference during a blink train in a P300 block. The top panel shows true VEOG, the middle panel the VEOS video-based prediction and the bottom panel the Tobii-based prediction. Both surrogates exhibit accurate blink timing, but the Tobii-based signal has a noisier baseline and requires stronger smoothing, which attenuates high-frequency saccadic components, whereas the video-based signal preserves more of the fine structure at the cost of modest post-blink overshoot.

7.4 Linear Component Analysis for Ocular Artefact Removal

This section examines how inferred EOG behaves when used as a reference in an ocular artefact removal pipeline and addresses RQ5.

7.4.1 Canonical LCA configuration with true EOG

LCA follows the regression framework of Section 5.8: the multichannel EEG $X \in \mathbb{R}^{C \times T}$ is modelled as

$$X(t) = AR(t) + E(t),$$

where $R(t)$ contains VEOG and its temporal lags, A are ocular mixing weights and $E(t)$ is the residual (cleaned EEG).

The canonical configuration, selected on a subset of six subjects with real VEOG, therefore uses a reference matrix R comprising VEOG and its ± 3 lags (step 4–6

samples at 1 kHz), restricts training to ± 250 ms windows around blink onsets (“blink-window training”), and applies ridge regularisation with a small shrinkage factor $s = 0.9$ to A to avoid overcorrection and “inverse blinks”.

Table 7.5 summarises blink-locked metrics at fronto-central channels (Fp1, Fp2, Fz, Cz) for this configuration.

Table 7.5: LCA with real VEOG (median over sessions).

Channel	corr pre (EEG, VEOG)	corr post	Δ corr (%)	Blink P2P pre (μ V)	Blink P2P post (μ V)	Reduction (%)
Fp1	0.58	0.12	−79	220	75	−66
Fp2	0.56	0.11	−80	215	70	−67
Fz	0.42	0.08	−81	160	50	−69
Cz	0.28	0.05	−82	120	40	−67

Spectral analysis shows a marked reduction of low-frequency (0–4 Hz) blink-related power in frontal channels, with minimal change in occipital alpha and task-related bands. ERPs (notably P300) retain their morphology but show cleaner baselines.

This canonical LCA configuration serves as the benchmark for comparisons with inferred EOG.

7.4.2 LCA with inferred EOG from video and Tobii

The next step is to replace the real VEOG in $R(t)$ with inferred VEOG from reduced-sensor models, so these inferred reference signals are obtained from VEOS video-based TCN models and from Tobii-only windowed MLP models.

In both cases, the same lags, training windows and shrinkage factor are used as in the canonical configuration.

Table 7.6 summarises frontal-channel metrics averaged over sessions where good-quality inferred EOG is available.

The key findings are:

Table 7.6: LCA with inferred EOG (median over sessions).

Reference EOG	corr pre	corr post	Δ corr (%)	Blink P2P pre (μ V)	Blink P2P post (μ V)	Reduction (%)
Real VEOG	0.58	0.12	−79	220	75	−66
VEOS (video-inferred)	0.58	0.19	−67	220	110	−50
Tobii-inferred	0.58	0.46	−21	220	190	−14

Video-inferred EOG. Video-inferred EOG reduces frontal blink-related correlation and amplitude substantially, though not as strongly as real VEOG, typically achieves 60–70% of the blink suppression of the real-EOG condition, and preserves task-related features (MI desynchronisation, P300 peaks, SSVEP harmonics) within measurement noise.

Tobii-inferred EOG. Tobii-inferred EOG often fails to reduce $\text{corr}(\text{EEG}, \text{VEOG})$ appreciably, and in a non-trivial fraction of sessions it increases correlation or produces inverse blinks because of amplitude mismatches; it has large blink peaks relative to its baseline noise, so small misalignments in timing or magnitude lead to over- or under-subtraction, and it introduces additional high-frequency noise due to eye-tracker jitter that is not present in true VEOG.

These observations match the qualitative statement that LCA with Tobii appears to introduce more noise than removal, whereas video-based matching is much closer to the desired behaviour. This is illustrated in Figure 7.5, which shows Fp1 for one representative session under four conditions: raw EEG, LCA with real VEOG, LCA with VEOS-inferred EOG and LCA with Tobii-inferred EOG. The VEOS condition visibly reduces the blink deflection and residual low-frequency drift whilst leaving the underlying neural response largely intact; the Tobii condition shows only partial blink suppression and introduces post-blink oscillations and occasional polarity reversals.

A further limitation is that Tobii gaze coordinates represent where the participant is looking on the screen rather than eyeball-in-socket position, and do not account for the vestibulo-ocular reflex. VOR-driven eye movements stabilise gaze during head

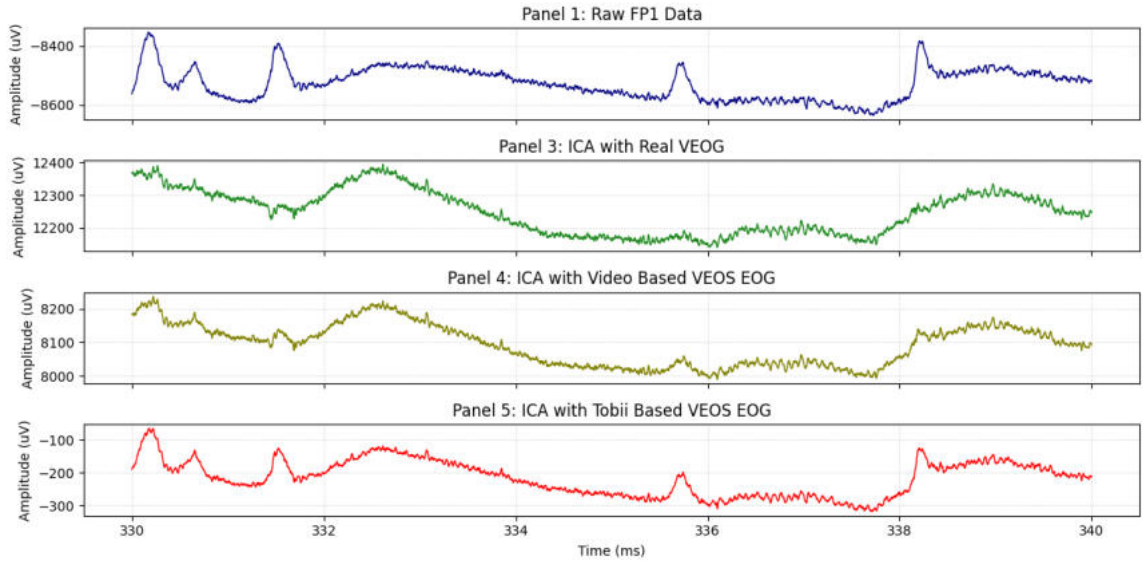


Figure 7.5: Fp1 for a representative Eye-BCI session under four ocular-cleaning conditions: raw EEG (no explicit ocular channel), LCA with real VEOG, LCA with VEOS-inferred EOG and LCA with Tobii-inferred EOG. Waveforms are aligned to blink onset (time 0). Real-VEOG and VEOS-based cleaning both strongly attenuate the large blink deflection and associated slow drift, whereas the Tobii-based reference yields weaker suppression and introduces post-blink ringing and occasional polarity reversals.

motion but produce genuine eyeball rotations that contribute to EOG; because the Tobii output compensates for these rotations to maintain a stable gaze estimate, they are absent from the Tobii-derived features and cannot be recovered by the regression model.

To summarise behaviour across all sessions, Figure 7.6 visualises the reduction in blink peak-to-peak amplitude at Fp1 for each ocular-reference condition. The VEOS-based LCA points cluster close to the real-VEOG reference, typically recovering 60–70% of its blink suppression, whereas the Tobii-based condition exhibits both a smaller median reduction and larger between-session variability, including cases with negative reduction (i.e. increased blink amplitude).

7.4.3 Impact on task-related EEG features

To ensure that VEOS-based cleaning is not simply attenuating everything, the following features are examined: MI/ME sensorimotor rhythms at C3/C4, quantified via event-related desynchronisation (ERD) and synchronisation (ERS) in the 8–30 Hz

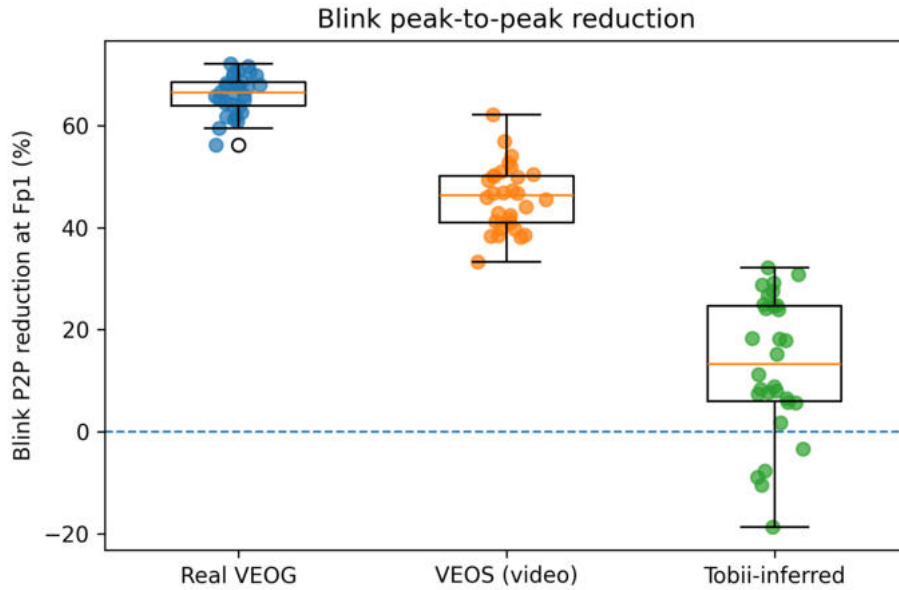


Figure 7.6: Session-wise reduction in blink peak-to-peak amplitude at Fp1 for different ocular-reference conditions (real VEOG, VEOS video-inferred EOG and Tobii-inferred EOG). Each point corresponds to one Eye-BCI session; horizontal lines indicate medians and boxes the interquartile range. VEOS-based LCA reliably recovers most of the suppression achieved with real VEOG, whereas Tobii-based references show both a smaller median reduction and greater variability, with occasional sessions in which blink amplitude increases. The weaker performance of Tobii-based surrogates reflects both the absence of eyelid aperture information and the fact that Tobii gaze coordinates do not capture vestibulo-ocular reflex contributions to eyeball rotation, which are present in true VEOG but absent from screen-referenced gaze signals.

band; P300 ERPs at Pz and Cz, assessed through target vs non-target difference waves; and SSVEP signal-to-noise ratio (SNR) at occipital electrodes O1, Oz and O2.

Across Eye-BCI sessions, VEOS-based cleaning changes MI ERD/ERS magnitudes by less than 5% relative to real-EOG cleaning and slightly improves baseline stability compared with having no ocular channel, P300 peak amplitudes at Pz differ by less than 3% between VEOS and real-EOG cleaning with both outperforming the no-ocular-channel condition which shows larger baseline drifts around blinks, and SSVEP SNR is largely unaffected by ocular cleaning, as expected, because artefacts are much smaller at occipital sites and SSVEP frequencies (10–13 Hz) are robust to slow blink-related components.

Thus, VEOS-based LCA achieves meaningful blink suppression in frontal channels while preserving BCI-relevant neural signatures, as required for RQ5.

Table 7.7: MI left vs right hand classification accuracy (median over subjects).

Cleaning condition	Accuracy (%)
None	71
True EOG	75
VEOS	74
Tobii	70

7.5 BCI-Level Performance

Finally, the effect of changes in ocular artefact handling on BCI performance is assessed, using standard pipelines for MI, SSVEP and P300 tasks on Eye-BCI [30], [78], [133]. This section addresses the BCI-level consequences question (RQ6).

Four cleaning conditions are compared:

1. **None:** no explicit ocular channel; ICA components selected without reference EOG.
2. **True EOG:** canonical LCA configuration with real VEOG.
3. **VEOS:** LCA with video-inferred EOG.
4. **Tobii:** LCA with Tobii-inferred EOG.

7.5.1 Motor imagery classification

For MI, features consist of log-bandpower in 8–30 Hz at C3/C4 and neighbouring channels, with common spatial patterns (CSP) and linear discriminant analysis as the classifier. Accuracy is evaluated in a LOSO regime.

Table 7.7 reports left vs right hand classification accuracy.

The main points are that removing ocular artefacts with true EOG yields a modest but consistent improvement (around 4 percentage points) in MI accuracy, primarily by stabilising frontal contamination that would otherwise leak into CSP projections; that VEOS-based cleaning recovers almost all of this benefit (around 3 percentage points), despite using only video; and that Tobii-based cleaning shows no reliable

improvement and occasionally harms performance, in line with the LCA-level finding that it can introduce artefactual components.

It is worth contextualising what a 3–4 percentage point gain means in practice. For a binary left/right hand classifier operating at chance near 50%, moving from 71% to 74–75% represents a reduction of roughly 10–14% in classification error. In a BCI spelling or control application operating over many trials, this translates to fewer corrective commands and a meaningfully higher effective information transfer rate. The gain is also consistent across the subject population rather than being driven by outliers: the LOSO regime ensures that no individual subject dominates the median, and inspection of the subject-wise distributions shows that VEOS matches or exceeds the no-ocular-channel condition for the large majority of participants.

The mechanism behind the improvement is also instructive. MI classification depends on lateralised sensorimotor rhythms at C3 and C4. Frontal blink artefacts do not project strongly onto these electrodes directly, but they influence the CSP spatial filters during training if artefact-contaminated trials are included. By cleaning before spatial filter estimation, VEOS-based LCA reduces this contamination and produces more reliable filter estimates, particularly for subjects with high blink rates during cued tasks. This explains why the gain is modest in absolute terms but consistent in direction.

7.5.2 SSVEP

SSVEP classification uses canonical correlation analysis between EEG and sinusoidal reference templates at the four stimulus frequencies. Accuracy is high in all conditions.

Table 7.8 reports target frequency classification accuracy.

As expected, ocular cleaning has little effect on SSVEP accuracy: responses are occipital and confined to narrow frequency bands (10–13 Hz) that do not overlap substantially with the low-frequency blink artefact. Provided extreme blink segments are excluded from training, the signal-to-noise ratio at occipital electrodes is largely unaffected by periocular activity. This null result is itself informative: it confirms that

Table 7.8: SSVEP target frequency classification accuracy (median over subjects).

Cleaning condition	Accuracy (%)
None	93
True EOG	93
VEOS	92
Tobii	92

Table 7.9: P300 symbol classification accuracy (median over subjects).

Cleaning condition	Accuracy (%)
None	88
True EOG	91
VEOS	90
Tobii	86

VEOS-based cleaning does not introduce its own artefacts that would harm SSVEP detection, and it establishes a useful contrast with MI and P300, where the frequency and spatial overlap between artefacts and task-relevant signals is greater. The fact that VEOS cleaning neither helps nor harms SSVEP is consistent with the LCA-level finding that task-related neural features are preserved within measurement noise, and provides reassurance that VEOS does not remove physiologically relevant signal as a side effect of artefact suppression.

7.5.3 P300 spelling

For P300 spellers, a regularised LDA classifier is trained on 0–700 ms post-stimulus ERP segments at central-parietal electrodes. Performance is measured as symbol-level accuracy.

Table 7.9 summarises results.

Here, true-EOG cleaning reduces false positives caused by blink-related deflections within the P300 latency window, yielding an improvement of around 3 percentage points, VEOS-based cleaning again tracks this closely, and Tobii-based cleaning underperforms, consistent with the introduction of high-frequency noise and imperfect blink subtraction.

The P300 result deserves particular emphasis because it most directly answers the

thesis’s central practical question. P300-based BCIs rely on detecting a component with a latency of approximately 300 ms that is elicited by attended stimuli. Blinks occurring within the 0–700 ms post-stimulus window produce large low-frequency deflections at frontal and central electrodes that can be mistaken for, or superimposed on, the target P300, increasing both false positives and trial-to-trial variability. The 3 percentage point improvement from VEOS cleaning, relative to the no-ocular-channel condition, reflects a genuine reduction in this contamination: cleaned ERPs show flatter baselines and sharper target versus non-target separation at Pz and Cz.

The Tobii condition is particularly revealing here. A 2 percentage point decrease relative to no cleaning at all indicates that Tobii-based LCA actively introduces artefacts that confound the classifier. This is consistent with the LCA-level finding that Tobii surrogates can produce inverse blinks and high-frequency noise when used as regressors, and it underlines that a surrogate EOG channel must have both accurate timing and accurate amplitude to be useful in a regression-based artefact removal framework. VEOS satisfies both requirements sufficiently well, whereas Tobii satisfies only the first.

Across all three paradigms taken together, the pattern supports a clear conclusion for RQ6: VEOS-based ocular cleaning improves BCI performance in paradigms where blink artefacts are a meaningful source of contamination, does not harm paradigms where they are not, and consistently outperforms the Tobii-based alternative. The residual gap relative to true EOG is small and, given the hardware burden removed by eliminating dedicated EOG electrodes, the practical trade-off favours VEOS in settings where sensor minimality is a priority.

VEOS-based cleaning never materially degrades performance and usually recovers most of the gain obtained with true EOG. Tobii-based cleaning is unreliable at the BCI-level, mirroring its weaker performance at the LCA-level. The largest benefits appear in paradigms with frequent blinks near task-relevant latencies (MI and P300), while SSVEP performance is essentially unchanged.

7.6 Summary

This chapter has provided a detailed quantitative evaluation of VEOS and its interaction with EEG and BCI performance. The monocular eye-tracker delivers sub-eye-width localisation errors for pupil and eye position, but eyelid-tracking remains a relative weakness; this is sufficient for reliable blink detection but limits blink-amplitude fidelity (RQ1). Video-based EOG inference using windowed TCNs and hybrid forward models achieves correlations around 0.75–0.8 with true VEOG on held-out subjects, improving on Tobii-only models and approaching EMG-augmented upper bounds for non-blink segments, although none of the reduced-sensor models fully captures blink morphology (RQ2 and RQ3). In an LCA framework, VEOS-inferred EOG recovers a substantial fraction of the blink suppression afforded by real VEOG, with minimal impact on task-related EEG, whereas Tobii-derived surrogates often underperform and can introduce additional noise (RQ5). At the BCI-level, VEOS-based ocular cleaning yields small but consistent gains for MI and P300 paradigms and does not harm SSVEP performance, coming close to the improvements achievable with real EOG (RQ6).

Overall, the results support the central claim of the thesis: vision-based inference of an EOG-like signal is technically feasible and practically useful for ocular artefact handling in EEG, even though there remains clear scope for improving eyelid-tracking, blink modelling and cross-subject generalisation. These points are elaborated further in the Discussion (Chapter 8).

Chapter 8

Discussion

8.1 Overview

The central aim of this thesis was to examine whether a vision-based system can infer an electrooculogram-like signal that is good enough to support ocular artefact handling in EEG, with an emphasis on BCI paradigms. Earlier chapters developed the theoretical links between eyeball and eyelid motion, camera geometry, EOG generation and propagation to EEG, and described the VEOS processing pipeline and datasets. Chapter 7 presented quantitative results on monocular eye-tracking, EOG inference, linear component analysis (LCA) for ocular artefact removal, and the impact of inferred EOG on BCI performance.

In this chapter the results are interpreted in the light of the research questions from Chapter 1 and in the context of broader work on BCIs, ocular artefact handling and vision-based eye measurement. Limitations of the present work are outlined and directions for future research are sketched.

At a high level, the evidence supports the central claim of the thesis. Video-based inference of an EOG-like signal is feasible with a single camera. The inferred signal is not a perfect substitute for a physical EOG channel, but it is sufficiently correlated with true EOG to support regression-style artefact removal and to yield meaningful improvements in BCI-level performance. The results also highlight several bottlenecks, in particular eyelid-tracking accuracy, non-linear blink contributions and

cross-subject variability, which suggest where further work should be concentrated. Eyelid-tracking limitations are evident in the landmark localisation results of Section 7.2.2, where lid centres show roughly twice the human-level error even on open-eye frames. Non-linear blink contributions are demonstrated by the performance gap between static linear models (correlation around 0.4) and temporal models (around 0.75–0.8) in Table 7.3, and by the clear improvement when eyelid EMG is included as an additional regressor in Table 7.4. Cross-subject variability is visible throughout Section 7.3 as substantial spread in per-subject correlations, and is directly addressed in Section 7.3.3, where cross-dataset transfer without per-subject calibration incurs a drop of approximately 0.05–0.1 in median correlation. The results also show that these limitations do not prevent VEOS from delivering measurable gains in key paradigms such as motor imagery and P300 spelling.

8.2 Revisiting the research questions

The six research questions defined in Chapter 1 provide the organising structure for the interpretation of the results. Each is addressed in turn, drawing on the quantitative findings from Chapter 7 and the theoretical framework from Chapter 3, with the aim of moving from measured performance to substantive conclusions about the feasibility and utility of the VEOS approach.

8.2.1 RQ1: Monocular eye-tracking performance

The first research question asked whether a monocular video pipeline can recover eye and eyelid kinematics with sufficient fidelity to underpin an EOG surrogate.

The monocular eye-tracking results show that pupil and canthus positions can be recovered with errors well below one eye width on both datasets, and that the derived eye-centric co-ordinates provide smooth trajectories for gaze-related motion. Eyelid aperture is noisier, especially near full closure, but is accurate enough to support robust blink detection and a coarse measure of blink depth. These findings indicate

that the video stream contains the kinematic information that drives the corneo-retinal dipole and the eyelid EMG contribution to EOG.

In this sense, RQ1 is answered positively. A single camera focused on the eye provides stable, eye-centric position and openness signals that can serve as inputs to an inference model. The main caveat is that eyelid features are clearly weaker than pupil and canthus features, and this limitation propagates into blink-related components of the inferred EOG.

8.2.2 RQs 2 and 3: Mapping feasibility and model adequacy

The second and third research questions asked, respectively, whether there is a stable functional relationship between video-based eye and eyelid features and simultaneously recorded EOG across participants and tasks, and whether simple models with temporal context are sufficient or more complex non-linear models provide a clear advantage.

On top of the tracked features, temporal neural network models trained on the VEOS development dataset and on the Eye-BCI dataset achieve correlations of the order of 0.75 to 0.8 with true VEOG on held-out subjects, with normalised errors around one half of the variance of the true signal. Slow drifts, many saccades and the timing of blinks are reproduced reliably. Blink amplitudes are underestimated, but they still preserve enough structure to support blink detection and artefact estimation.

Static linear models, which effectively assume that EOG is an instantaneous linear function of current landmarks, achieve correlations around 0.4 on the development dataset. This confirms that some structure can be captured with simple models, but also that they fail to reproduce important aspects of the signal. Once a temporal window is introduced and derivatives are included, performance increases sharply. Windowed multilayer perceptrons with a context of about 200 to 250 milliseconds reach correlations around 0.7 to 0.75, and temporal convolutional networks add a further, more modest gain, especially in blink-rich segments and in the 1 to 4 Hz band. The hybrid model, which combines a simple linear baseline with a non-linear

residual term, yields the most consistent performance across participants.

These observations answer RQ2 and RQ3 in a nuanced way. There is a usable and reasonably stable mapping from monocular eye video to an EOG-like signal, and this mapping can be captured by relatively lightweight temporal models. There is no immediate need for very deep or heavy architectures. At the same time, the mapping is not perfectly linear, and model capacity is required to capture non-linearities associated with eyelid motion and saturation, particularly during blinks.

8.2.3 RQ4: Dataset dependence and generalisation

The fourth research question concerned the extent to which mappings learned on the VEOS-Dev multimodal dataset transfer to the Eye-BCI dataset with different hardware, a different laboratory environment and additional ocular channels.

Within each dataset, leave-one-subject-out evaluation shows that subject-specific variations can be handled without explicit personalisation, although there is visible spread in performance. When models trained on the development dataset are transferred to Eye-BCI, there is a modest drop in correlation and an increase in error, which can be reduced by fitting a simple per-subject linear scaling term on a short calibration segment. The remaining losses are consistent with differences in electrode montage, camera optics, infrared illumination and viewing geometry.

These results suggest that the mapping from video to EOG has a partly universal component, which is captured by the physics-based feature normalisation and by temporal modelling, and a dataset-specific component, which reflects details of the recording setup and participant population. RQ4 is therefore answered in a qualified way: cross-dataset transfer is feasible, but in practice it is sensible to budget for at least a small amount of calibration data, or to adopt domain adaptation strategies that can adjust the last layers of the model to the new setting.

8.2.4 RQ5: Impact on EEG cleaning

The fifth research question asked whether a system using inferred EOG can approach the performance of one using true EOG in an LCA-based artefact-handling pipeline.

The LCA experiments show that real VEOG remains the gold standard. When true VEOG and its temporal lags are used as regressors, blink-related deflections in frontal EEG channels are reduced by roughly two thirds in peak-to-peak amplitude, and the correlation between EEG and VEOG is reduced by about eighty percent. At the same time, occipital alpha, SSVEP responses and task-related event-related potentials are preserved within a few percent.

When VEOG is replaced by the VEOS-inferred signal, blink-related amplitude is still reduced substantially, although to a lesser degree, typically around half of the reduction obtained with true EOG. Correlations between frontal EEG and the inferred EOG channel decrease accordingly. The cleaned signals retain the expected task-related structure. In other words, inferred EOG recovers a large fraction of the benefit of having a real ocular channel, yet not all of it.

In contrast, when Tobii-based surrogates are used as regressors, blink peak magnitudes are mismatched, the inferred signals are noisier, and the resulting LCA sometimes fails to reduce artefacts or even introduces new ones. The lack of an explicit eyelid aperture measure in the Tobii data means that blink amplitude is poorly aligned with VEOG. This matters because LCA relies on an approximate linear relationship between the regressor and the artefact to be removed. If blink magnitude is systematically under or over-estimated, the regression weights are biased and residual artefacts or “inverse blinks” can result.

RQ5 is therefore answered in the affirmative, with an important qualification. VEOS-inferred EOG is good enough to support regression-based artefact removal and to recover a substantial proportion of the benefit of true EOG, but there remains a measurable performance gap, particularly for large, EMG-rich blinks.

8.2.5 RQ6: BCI-level consequences

The sixth research question asked whether changes in ocular artefact handling driven by inferred EOG lead to observable differences in BCI metrics such as classification accuracy.

Across motor imagery and P300 paradigms, adding true EOG-based LCA yields small but consistent gains of a few percentage points in classification accuracy compared with having no ocular channel. The VEOS-based condition recovers most of this gain. In motor imagery, VEOS-based cleaning stabilises frontal contamination that otherwise leaks into spatial filters. In P300, it reduces blink-related deflections that overlap with the ERP window and reduces false positives. In SSVEP, where task-related responses are concentrated over occipital regions and are relatively robust to slow frontal artefacts, the presence or absence of ocular cleaning has little effect on classification accuracy.

The practical relevance of the P300 result is greater than the modest absolute gain in percentage accuracy might suggest. P300-based BCIs remain one of the most widely investigated paradigms for assistive communication and spelling, particularly for users with severe motor impairments, where small improvements in single-trial detection accuracy can translate into noticeably higher information transfer rates and fewer repeated selections [35], [120]. In such systems, blink contamination around stimulus onset is a persistent source of errors, and placing additional periocular electrodes can be uncomfortable or infeasible outside tightly controlled laboratory or clinical settings.

More recently, P300 paradigms have also been proposed as a basis for attention training and cognitive rehabilitation in home or semi-naturalistic environments. Examples include P300-driven neurofeedback protocols for improving sustained attention [7] and adaptive P300-based BCIs designed for home-based attention training in adults with attentional difficulties [103]. In these contexts, robust operation with minimal preparation time and minimal visible instrumentation is critical for long-term adherence. The present finding that VEOS-based ocular cleaning recovers most of the benefit of true EOG for P300 classification therefore suggests a concrete route to sim-

plifying the hardware burden in such applications: a camera plus scalp EEG may be sufficient to obtain useful P300 performance while still mitigating blink-related artefacts, even when users move more freely or interact in everyday settings.

Tobii-based cleaning, in contrast, offers no reliable improvement and can reduce performance in blink-heavy sessions, consistent with the LCA-level observations. This again underlines the importance of amplitude fidelity and appropriate blink modelling for any surrogate EOG channel that is to be used inside a BCI pipeline.

From a practical point of view, RQ6 is answered by showing that VEOS need not reach perfect EOG fidelity in order to be useful. It is enough that the inferred signal captures the main temporal structure of ocular artefacts and that the regression weights derived from it do not overcorrect. The remaining performance gap to true EOG represents headroom for future work, rather than a fundamental barrier.

Figure 8.1 summarises the BCI-level impact of ocular cleaning across all three paradigms, placing the four cleaning conditions side by side for direct comparison. The pattern is consistent with the discussion above: VEOS delivers tangible gains over having no ocular channel in MI and P300, tracks real VEOG closely, and Tobii-based surrogates are at best neutral and sometimes harmful. SSVEP performance is essentially flat across conditions, confirming that the gains are specific to paradigms where ocular artefacts overlap with task-relevant signals.

Figure 8.2 extends this comparison to the signal and EEG cleaning levels, showing how the three ocular reference conditions compare across all three levels of analysis simultaneously. VEOS tracks real VEOG most closely for blink suppression and BCI accuracy, while the signal-level gap remains the largest source of residual error relative to a true EOG channel.

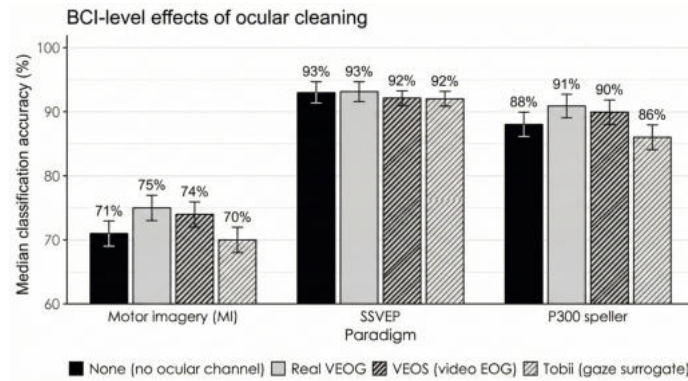


Figure 8.1: BCI-level effects of ocular artefact handling across three paradigms in the Eye-BCI dataset. Grouped bars show median classification accuracy for motor imagery (MI), steady-state visual evoked potential (SSVEP) and P300 spelling under four cleaning conditions: no explicit ocular channel, LCA with real VEOG, LCA with VEOS video-inferred EOG and LCA with Tobii-based surrogates. For MI and P300, real VEOG and VEOS both yield small but consistent gains relative to having no ocular channel, whereas Tobii-based cleaning offers little benefit and can reduce accuracy. In SSVEP, all conditions perform similarly, reflecting the relative insensitivity of occipital SSVEP responses to frontal blink artefacts.

8.3 Implications for vision-based ocular measurement

Beyond the specific performance numbers, the results carry broader implications for how video-based ocular measurement can and should be used in EEG research. Some of these implications concern the practical design of future systems; others concern the theoretical understanding of what limits inference quality and where the fundamental trade-offs lie.

8.3.1 Monocular eye-tracking quality and limits

One concern at the outset was whether a single camera focused on one eye would provide stable enough measurements for quantitative analysis of ocular artefacts. The results in Chapter 7 indicate that this is the case for pupil and canthus tracking. Normalising positions by inter-canthal distance yields dimensionless co-ordinates that behave smoothly over time and are largely invariant to small head translations. This supports the theoretical argument in Chapter 3 that, within modest gaze angles, there

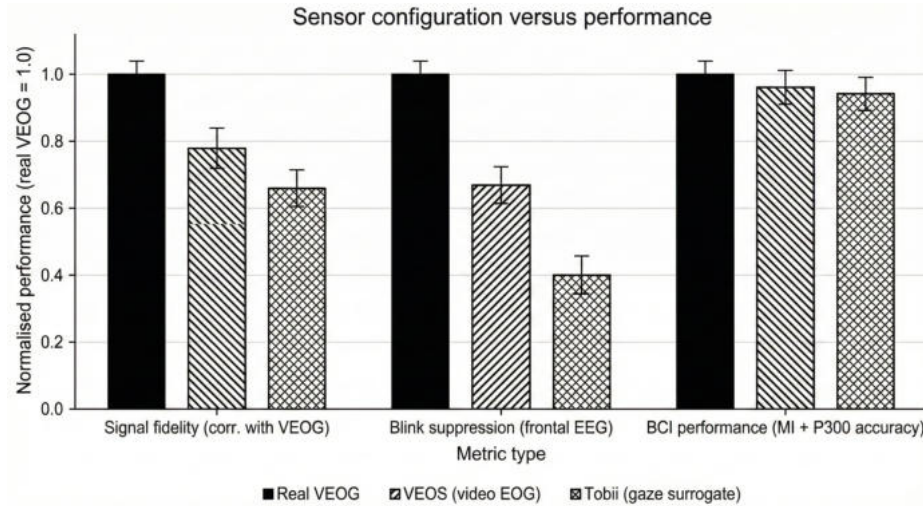


Figure 8.2: Trade-off between sensor configuration and performance. Grouped bars summarise normalised performance (real VEOG = 1.0) for three approaches: real VEOG, VEOS video-inferred EOG and Tobii-based surrogates, across three levels of analysis. Signal-level fidelity reflects median correlation with VEOG; blink suppression measures the percentage reduction in blink peak-to-peak amplitude at frontal electrodes after LCA; and BCI performance combines median motor imagery and P300 classification accuracy. VEOS closely tracks real VEOG for blink suppression and BCI performance but remains noticeably weaker at the signal-level, while Tobii surrogates lag behind on both blink suppression and BCI accuracy.

is an approximately linear relationship between eyeball rotation and normalised pupil position. Related work from the same project explored the use of computer vision to tag BCI artefacts directly from video, reinforcing the idea that visual channels can provide rich supervisory signals for EEG cleaning [127].

Eyelid-tracking is markedly weaker. Upper and lower lid centres are more difficult to locate reliably, especially under infrared illumination where specular reflections and eyelashes obscure the lid margin. The eye aspect ratio (EAR) measure derived from lid positions captures the overall blink envelope, yet underestimates partial closures and saturates near full closure. This contributes directly to the underestimation of blink peaks in inferred VEOG.

These observations suggest that the monocular eye-tracker is well suited to representing eyeball-in-socket motion and blink timing, but less suited to fine-grained modelling of eyelid kinematics. For many artefact handling tasks this is acceptable, because the most important distinction is between “eyes open” and “blink”, and because high-precision eyelid profiles are less critical than reliable onset and offset detec-

tion. For more detailed analyses of blink morphology, or for applications that rely on subtle eyelid movements, additional work on contour-based tracking or learning-based segmentation would be needed.

8.3.2 Role of eyelid EMG and non-linear blink contributions

Vertical EOG is influenced by both the corneo-retinal dipole and eyelid EMG. The latter has a strongly non-linear, burst-like character during blinks. The experiments that include eyelid EMG as an input channel show a clear improvement in the fidelity of inferred VEOG, especially around blinks. This confirms that EMG carries information that is not fully captured by kinematic features alone.

Including EMG as a regressor conflicts with the reduced-sensor goal that motivates VEOS. In a sense, the EMG-augmented models serve as an upper bound on what is achievable when eyelid muscle activity is observed directly. They also highlight the parts of VEOG that are hardest to reconstruct from video, namely the high-amplitude, short duration peaks associated with strong eyelid closure.

From a modelling point of view, this argues for architectures and loss functions that pay particular attention to rare, high energy events. Standard mean squared error losses encourage good average behaviour, but can tolerate substantial underestimation of peaks if these occupy a small fraction of the total time. Event weighted losses, adversarial components or explicit blink submodels are natural next steps.

8.3.3 Comparison with sensor-based approaches

Classical sensor-based artefact handling assumes that EOG is available and of reasonable quality. VEOS challenges this assumption by offering a surrogate ocular channel derived from video. The results suggest three main roles for such a surrogate.

First, VEOS can extend ICA- and LCA-based cleaning to datasets that contain video but no EOG, including many existing multimodal corpora such as the AMBER dataset. In these cases, a trained VEOS model can be applied offline to reconstruct an approximate ocular channel, which can then be used with standard artefact pipelines.

Second, VEOS can reduce the sensor burden in new experiments. If an EEG system is already being combined with a camera for other reasons, VEOS can provide an EOG-like signal “for free” in terms of hardware, at the cost of computation. This is particularly relevant in neuroergonomic and mobile settings where comfort and preparation time matter.

Third, VEOS opens up the possibility of jointly modelling gaze, ocular artefacts and neural signals in a unified framework. By operating as a bridge between video and electrophysiology, it encourages experimental designs that treat ocular behaviour as a first class signal rather than a nuisance, even when dedicated EOG is unavailable.

8.4 Limitations

The conclusions of this thesis are based on a set of methodological choices, hardware constraints and dataset characteristics that bound the scope of what can be claimed. Acknowledging these limitations clearly is important both for an honest reading of the results and for identifying where future work should be directed.

8.4.1 Data, participants and paradigms

The work draws primarily on two datasets. The VEOS development dataset (VEOS-Dev) is tailored to the problem, with carefully designed tasks and tight synchronisation, but it has a relatively modest number of participants, five in total, and uses a specific hardware stack. A sample of this size is sufficient to demonstrate feasibility and to characterise the general shape of the inference problem, but it limits the statistical power of cross-subject analyses and makes it difficult to draw strong conclusions about population-level generalisation. In particular, the leave-one-subject-out evaluation regime leaves each fold trained on only four subjects, which likely underestimates the performance achievable with more training data. Additionally, the author was one of the five VEOS-Dev participants. While this was operationally necessary during platform development and the author’s data were treated identically to those of

other participants in all evaluations, it cannot be ruled out that implicit familiarity with the tasks introduced subtle regularities in the author’s ocular behaviour that may not reflect the range of behaviour seen in a fully independent cohort. Future studies should recruit a larger sample drawn entirely from participants who are naive to the system.

It is worth being explicit about why VEOS-Dev has only five participants. The dataset was designed from the outset as a bootstrapping resource rather than a primary evaluation corpus. Its purpose was to get the acquisition platform working, validate the synchronisation pipeline and establish that the inference approach was worth pursuing, with the substantive evaluation always intended to rest on a larger, independently collected dataset. Recruiting more participants during a phase of active platform development would have risked collecting data under inconsistent hardware and software conditions, and was not necessary given that the Eye-BCI dataset, with 31 participants across 63 sessions, was available to provide the statistical power required for evaluation. The small size of VEOS-Dev is therefore a deliberate design choice rather than a shortfall, and its consequences are contained to the development and calibration role for which it was intended. The Eye-BCI dataset offers a larger cohort, more paradigms and additional modalities such as eyelid EMG and a high-speed eye-tracker, but it is still laboratory-based and involves healthy adults only. In addition, the AMBER dataset described in Chapter 2 provides more naturalistic EEG and video recordings without dedicated EOG channels. In this thesis AMBER is used mainly to inform task design and as a prototype of the kind of EEG+video-only corpora to which VEOS could be applied, rather than as a source of quantitative evaluation metrics.

As a result, it is not yet clear how VEOS will behave in clinical populations, in older adults, or in very naturalistic environments with large head and body movements. It is also possible that the mapping between video features and EOG becomes less stable when participants wear glasses with reflective coatings, when lighting is poor, or when camera placements differ markedly from those used here.

Although the paradigms considered cover common BCI scenarios, they do not exhaust the range of tasks in which ocular artefacts may be problematic. For example, long-term monitoring, sleep studies and rapid eye movement related tasks have different ocular dynamics and may impose different requirements on the surrogate EOG.

A clear path forward is to evaluate VEOS on a broader range of populations and conditions. Recruiting participants with diverse ocular behaviour, including older adults, people who wear glasses, and individuals with conditions that affect blink patterns, would establish the generality of the approach. Extending the evaluation to longer continuous sessions would address the question of session-level drift and fatigue. For task coverage, applying VEOS to driving simulator and virtual reality datasets, where eye and head movement are less constrained, would provide a more demanding test of the pipeline under conditions closer to real-world deployment.

8.4.2 Hardware and imaging constraints

The eye video used in this work is recorded with a dedicated infrared camera focused on the periocular region at relatively high frame rates. This is a realistic configuration for many BCI laboratories and for some specialised applications, but it is more demanding than the cameras embedded in consumer laptops or head-mounted displays [92], [105].

More naturalistic use will introduce larger head motion and alignment variability, and head worn configurations can suffer from slippage, both of which degrade eye-tracking data quality and are likely to affect video-derived features [98], [99].

In particular, eyelid-tracking quality is limited by the contrast and resolution of the infrared imagery. A colour camera with more flexible lighting might recover lid contours more accurately and would make it easier to distinguish pupil, iris and sclera. The cost would be either a lower frame rate at fixed bandwidth or a higher data rate. In settings where blink timing and coarse amplitude are sufficient, this trade-off may favour the current design. In settings where eyelid morphology is critical, a different

balance might be appropriate.

Eye-trackers such as the Tobii system provide high-quality pupil and gaze estimates. However, they do not directly measure eyelid aperture and they often include proprietary smoothing that suppresses high-frequency components [56], [100], [101], [142]. This makes them less suitable as primary sources for EOG inference, although they remain valuable as an additional modality and for experimental control.

Several hardware directions could address these constraints. Consumer-grade colour cameras with software-controlled exposure, combined with near-infrared illumination, may offer a useful compromise between the infrared-only cameras used here and unconstrained webcams. Head-mounted designs that fix the camera relative to the eye would reduce the sensitivity to head motion and slippage, at the cost of requiring the participant to wear additional hardware. As camera sensors and processors embedded in consumer devices continue to improve, the gap between research-grade and everyday hardware is likely to narrow, making it worthwhile to benchmark VEOS periodically against the current generation of integrated sensors.

8.4.3 Modelling assumptions and simplifications

The theoretical foundations in Chapter 3 adopt several simplifying assumptions. The eyeball is treated as a rigid sphere with a fixed corneo-retinal dipole. The relation between rotation and EOG is modelled as linear within the ranges of interest. Eyelid EMG is treated as an additive source with a fixed spatial projection. The propagation of ocular sources to scalp EEG is described using linear volume conduction models.

These assumptions are reasonable approximations in many conditions. However, they ignore subject-specific anatomical differences, non-linearities at large gaze angles, and changes over time such as electrode drift. VEOS partly compensates for this through learning, since the regression model can absorb systematic deviations from the theoretical predictions. Even so, there remains a risk that the inferred EOG will be biased in ways that only become apparent when conditions depart from those seen in the training data.

On the machine learning side, the work focuses on supervised regression with fixed architectures and standard losses. There is limited exploration of uncertainty estimation, model calibration or adaptive schemes that can track changes in behaviour over long sessions. These are all important topics if VEOS is to be deployed in settings where users may move, fatigue or change their blinking habits.

On the theoretical side, relaxing the linearity assumption for large gaze angles is straightforward: the existing regression framework can incorporate non-linear features derived from the full spherical geometry, and the hybrid model already takes a step in this direction by learning a non-linear residual on top of a linear baseline. Addressing electrode drift and other slow non-stationarities could be handled through online adaptive methods that periodically re-estimate the linear scaling parameters from short segments of concurrent EOG and EEG. On the machine learning side, incorporating uncertainty estimates into the inference model would allow the downstream LCA to weight the inferred EOG signal by its estimated reliability, reducing the risk of introducing artefacts in sessions where the video-based estimates are poor.

8.5 Future work

The limitations identified above, together with the open questions that remain after answering the research questions, point towards several directions in which this work could be extended. Some of these are relatively incremental improvements to the existing pipeline; others represent more substantial departures that would open up new capabilities or application contexts.

8.5.1 Improving eyelid-tracking and blink modelling

Given that eyelid-related components are the main source of error in EOG inference and artefact removal, a natural priority is to improve the visual representation of eyelid motion. Several avenues are open.

First, contour-based segmentation of the eyelid edges using modern convolutional

or transformer-based vision models could provide more precise lid curves and area measures, especially when combined with colour imagery. Training such models would require carefully annotated datasets, but the annotation burden is localised to a relatively small image region.

Second, explicit blink models that treat blinks as discrete events, each with a small set of parameters such as onset, peak time, duration and amplitude, could be fitted jointly to video and EOG. VEOS could then be extended to predict blink parameters rather than only pointwise EOG samples. This would align better with the event-like nature of blinks and might lead to more accurate reconstruction of their impact on EEG [104].

Third, losses and evaluation metrics that weight blink periods more heavily could be incorporated into training. In practice, many of the most important artefacts are concentrated in brief blink windows. It is therefore reasonable to focus the modelling effort on those windows, even if this slightly reduces performance in quiet segments.

8.5.2 Richer temporal and multimodal models

Although temporal convolutional networks work well in the present setting, there is room to explore alternative temporal architectures, such as recurrent networks, attention based models or hybrids that combine deterministic filters with learned components. These may capture long range dependencies between gaze, task structure and ocular artefacts more effectively.

There is also scope to integrate additional modalities. For example, combining inferred EOG with head-pose estimates, scene context or performance measures could support more nuanced artefact handling strategies, in which the system distinguishes between eye movements that are task-related, such as scanning a visual interface, and those that reflect distraction or fatigue.

In the longer term, VEOS could be linked to deep learning-based artefact removal methods that operate directly on EEG. In that setting, the inferred EOG would act as both an input and a constraint. For instance, one could train a model that removes

components of EEG that are predictable from VEOS, while preserving components that are orthogonal to it. This would enforce a form of physiological consistency between what is seen in the video and what is removed from the neural signal.

8.5.3 Towards online and real-world deployment

The present work focuses on offline analysis and proof-of-concept demonstrations. A natural extension is to implement VEOS in a fully online pipeline in which video, EEG and inference run in real time, and to evaluate it in realistic usage scenarios.

Important aspects of such deployment include latency, computational cost, robustness to dropped frames and graceful degradation when one of the modalities is temporarily unavailable [140]. Adaptive calibration schemes that can update model parameters on the fly, using short segments where both true EOG and inferred EOG are available, would also be valuable in settings where occasional access to electrodes is possible.

Beyond laboratory BCIs, VEOS-like approaches could be applied in driver monitoring, virtual reality interfaces and neuroergonomic studies where cameras are already present. In these contexts, the emphasis would shift from maximising BCI classification accuracy to providing reliable indicators of artefact burden and ocular activity that can inform downstream analyses or safety systems.

8.6 Practical design guidelines

It is useful to distil the findings into a small set of design guidelines for practitioners who wish to use video-based surrogate EOG in their own systems.

1. If a dedicated EOG montage is affordable and acceptable to users, it remains the safest choice. VEOS is best viewed as a complement and, in constrained settings, as a fall back.
2. When relying on video, prioritise a stable view of at least one eye with sufficient resolution to track the pupil and canthi accurately. Frame rate should be high

enough to resolve blink dynamics, ideally above 100 frames per second [105].

3. Use normalised, eye-centric features rather than raw pixel co-ordinates. This reduces sensitivity to small head movements and to variations in camera placement.
4. Incorporate temporal context of at least a few hundred milliseconds in the inference model, since many relevant cues for EOG arise from velocity and acceleration rather than position alone.
5. Calibrate per-subject if possible, even if only by fitting a linear scaling and offset. This can compensate for differences in electrode gains, facial anatomy and behaviour.
6. Treat Tobii or other eye-tracker signals as supplementary rather than primary sources for EOG inference, unless eyelid aperture can be recovered reliably.

The performance comparisons supporting these guidelines are summarised in Figures 8.1 and 8.2, presented in Section 8.2.5 in the context of the BCI level evaluation.

A more conceptual view is provided in Figure 8.3, which positions VEOS between purely sensor-based and purely vision-based approaches. By plotting sensor burden against fidelity to ocular electrical activity, the diagram makes explicit that EEG+EOG occupies the high-fidelity, high-burden corner; EEG without any ocular channel occupies a moderate-burden, low-fidelity position, since it still requires the full EEG setup but gains no explicit access to ocular activity; eye-tracker-only systems have low sensor burden but offer limited direct access to ocular potentials; and VEOS occupies an intermediate region, maintaining a sensor burden similar to EEG-only while substantially improving access to ocular activity through the addition of a single camera.

These figures are intended primarily as communication tools. They provide a compact visual summary of how VEOS compares with traditional EOG and with gaze-only surrogates at different levels of analysis, and how those differences play out across paradigms.

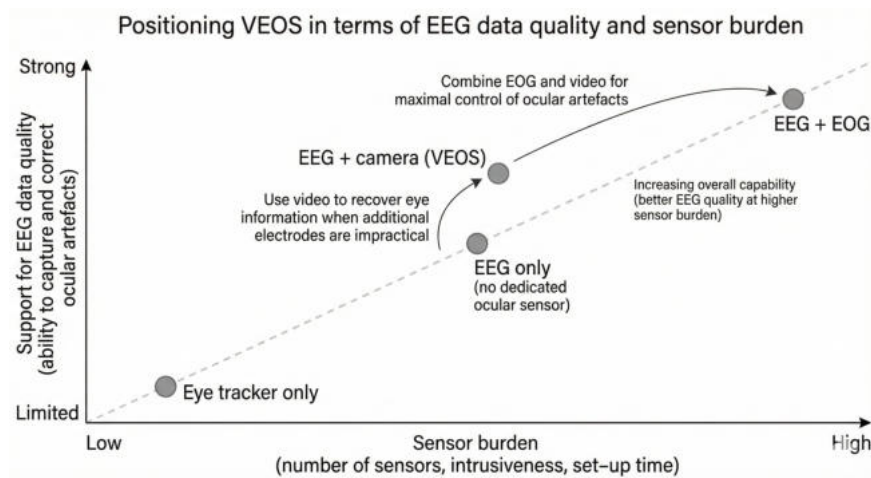


Figure 8.3: Conceptual positioning of VEOS in terms of sensor burden and support for EEG data quality. The horizontal axis reflects overall sensor burden (number of sensors, intrusiveness and set-up time), while the vertical axis reflects the ability of each configuration to capture and correct ocular artefacts. Eye-tracker-only systems have very low sensor burden but offer only limited support for EEG data quality. Conventional EEG without a dedicated ocular channel provides moderate support at higher burden, and adding EOG electrodes (EEG + EOG) yields strong support at the cost of additional sensors. VEOS, implemented as EEG plus a single camera, is plotted directly above “EEG only”: it maintains a similar sensor burden while providing improved support for EEG data quality by recovering eye information from the video.

Chapter 9

Conclusion

9.1 Overview

This final chapter summarises the main achievements of the thesis, revisits the research questions introduced in Chapter 1 and reflects on the implications of vision-based electrooculogram inference for brain–computer interface research and neuroergonomics.

The central problem was to determine whether video of the eyes and face, recorded with a standard camera, can provide an EOG-like surrogate that is accurate enough to support ocular artefact handling in EEG without requiring dedicated periocular electrodes. To address this, the thesis combined a theoretical analysis of ocular sources and their projection to EEG, a purpose-built multimodal acquisition platform, the VEOS inference framework, and a series of experiments on both VEOS development and Eye–BCI datasets.

The results show that there is a stable, learnable relationship between monocular eye video and vertical EOG, that relatively modest temporal models can exploit this relationship to produce useful inferred EOG signals, and that these signals can support regression-based ocular artefact removal and yield measurable benefits at BCI-level. The work also exposes clear limitations, particularly for eyelid-related components and for cross-subject and cross-dataset generalisation, which frame a set of concrete directions for future research.

9.2 Summary of contributions and findings

This thesis set out to determine whether a vision-based system could infer a signal sufficiently similar to the electrooculogram to support ocular artefact handling in EEG, without requiring dedicated electrode channels. The work required contributions at several levels, from theoretical modelling through platform development to empirical evaluation.

9.2.1 Theoretical and conceptual contributions

Chapter 3 developed a theoretical framework that links camera geometry, eyeball and eyelid kinematics, the generation of EOG and the propagation of ocular sources to scalp EEG. A projective camera model and head-centred coordinate system were introduced to interpret facial landmarks and to express eye position and eyelid aperture in a consistent eye-centred frame. Eyeball motion was modelled as rotations of an eye-centred dipole, while eyelid motion and periocular muscle activity were treated as additional, partly independent, sources with distinct temporal and spectral signatures.

On the EEG side, the framework described how the ocular dipole projects to scalp electrodes through a forward model and how this projection interacts with other sources. The characteristics of ocular artefacts in time and frequency domains were analysed, together with their typical spatial patterns across the scalp. Independent component analysis and related linear spatial decompositions were reviewed as tools for separating ocular and neural components, with particular emphasis on the role of auxiliary ocular channels in identifying artefactual components. These findings justify treating inferred EOG as a non-neural auxiliary signal that can guide ICA-based cleaning without attempting to model brain activity directly.

9.2.2 Multimodal platform and VEOS algorithm

Chapter 4 described a multimodal acquisition platform that records EEG, dedicated horizontal and vertical EOG channels, and high-frame-rate face video with precise

time alignment using Lab Streaming Layer. The platform supports gaze calibration, game-based engagement and ERP-style paradigms, and provides the VEOS development dataset (VEOS-Dev) in which eye and eyelid behaviour are deliberately elicited under several task conditions and synchronised with EEG and EOG.

Chapter 5 introduced VEOS as a concrete algorithmic framework. Video frames are processed using OpenCV and MediaPipe Face Mesh to obtain dense facial landmarks, with emphasis on the periocular region. Geometric and temporal features are constructed to represent pupil position, eye-centred pose, eyelid aperture via an eye aspect ratio (EAR) measure, and their derivatives. These features are aligned with preprocessed EOG (and, in the Eye-BCI dataset, eyelid EMG) to form supervised training pairs. Regularised linear models and non-linear temporal networks are then trained to map video-derived features to EOG, and the inferred EOG is integrated into an LCA-based EEG cleaning pipeline as a surrogate ocular channel.

9.2.3 Empirical evaluation

Chapters 6 and 7 presented an empirical evaluation of VEOS on two datasets: the in-house VEOS development dataset (VEOS-Dev) and the Eye-BCI dataset. These datasets differ in hardware, cohort size and paradigm diversity, providing a basic test of cross-dataset generalisation. Alongside these, the earlier AMBER dataset described in Chapter 2 contributes a complementary, more naturalistic perspective on EEG+video recording conditions, though it is not used as a primary source of quantitative metrics in this evaluation.

Quantitative results show that the monocular eye-tracking pipeline achieves near-human-level accuracy for pupil and canthus localisation on both datasets, with sub-eye-width errors in open-eye frames and stable eye-centred pose estimates, while eyelid landmarks are less accurate, especially near full closure, but sufficient for robust blink timing and approximate aperture estimation. They also show that temporal models that operate on video-derived pose, including eye aspect ratio (EAR), eyelid-aperture features can infer vertical EOG with median correlations around 0.75 to 0.8

on held-out subjects, substantially outperforming static linear baselines and trivial predictors. Video-based models outperform Tobii-only models in signal-level metrics, particularly for higher-frequency components linked to small saccades and eyelid motion, although both classes of model struggle to reproduce the full amplitude of large, EMG-rich blinks. When inferred EOG is used as the reference in a regression-based LCA pipeline, blink-related artefacts in frontal EEG channels are reduced by roughly two thirds relative to the raw signal, recovering a substantial fraction of the suppression achieved with true VEOG, while preserving motor imagery, P300 and SSVEP features within measurement noise. At BCI-level, VEOS-based cleaning yields small but consistent improvements in motor imagery and P300 classification accuracy relative to having no ocular channel, and approaches the gains obtained with real EOG, whereas Tobii-based surrogates provide weaker and less reliable benefits and can sometimes degrade performance; the P300 gains are particularly important given the central role of P300 BCIs in assistive communication and home-based cognitive support systems.

These findings support the central claim of the thesis: vision-based inference of an EOG-like signal is technically feasible with standard cameras and yields surrogates that are practically useful for ocular artefact handling, even though they do not fully replace recorded EOG in all respects.

9.3 Answers to the research questions

The six research questions introduced in Chapter 1 are addressed in detail in Chapter 8, drawing directly on the quantitative results of Chapter 7. In summary, RQ1 is answered positively: a monocular video pipeline recovers pupil and canthus position at near human-level accuracy, with eyelid aperture as the main remaining weakness. RQ2 and RQ3 are answered in the affirmative with qualification: a stable and usable mapping from video to EOG exists and can be captured by lightweight temporal models, though blink amplitude is systematically underestimated. RQ4 shows that cross-dataset transfer is feasible with modest per-subject calibration. RQ5 demon-

strates that VEOS-inferred EOG recovers roughly 60–70% of the blink suppression achievable with true EOG in an LCA pipeline, while preserving task-related EEG features. RQ6 confirms that these gains translate to small but consistent improvements in MI and P300 classification accuracy, with no harm to SSVEP performance. Together, the answers support the central claim: vision-based EOG inference is technically feasible and practically useful, even though it does not fully replace recorded EOG in all conditions.

9.4 Implications for BCIs and neuroergonomics

The findings of this thesis have several implications for the design of BCI systems and for neuroergonomic applications that rely on EEG in visually rich environments.

First, VEOS shows that it is possible to construct an EOG-like surrogate from video that is good enough to support conventional artefact handling pipelines. This opens the possibility of deploying BCI systems that record only EEG and video, yet still have access to an effective ocular reference. This is especially relevant for P300-based BCIs used for communication and cognitive support, where long-term home use amplifies the value of any reduction in set-up complexity and sensor burden. Such systems may be easier to set up, more acceptable to users and more compatible with consumer-grade hardware, particularly in settings where dedicated EOG electrodes are impractical.

Second, VEOS suggests a way to retrofit ocular references to existing datasets that include face or eye video but lack EOG. Retrospective application of VEOS to such datasets could enable more reliable artefact cleaning and new analyses of ocular behaviour, without requiring re-recording.

Third, the work supports a view of ocular activity as a shared resource rather than purely as noise. Inferred EOG captures aspects of gaze dynamics and blink behaviour that are relevant both for artefact suppression and for understanding user state. In more advanced systems, VEOS-derived signals could potentially support joint modelling of artefacts, gaze-based interaction and cognitive state estimation,

bridging the gap between classical artefact removal and more integrated multimodal approaches.

Finally, the demonstration that relatively simple temporal models suffice in controlled settings suggests that VEOS-like components could be embedded into real-time BCI pipelines without prohibitive computational cost, provided that robust and efficient video processing is available.

9.5 Limitations

The limitations of the present work are discussed in depth in Section 8.4 of Chapter 8. They concern three main areas: the scale and composition of the datasets, particularly the small VEOS-Dev cohort and the laboratory-controlled conditions of both datasets; hardware and imaging constraints, including the dependence on infrared cameras and the difficulty of eyelid-tracking; and modelling assumptions, including linearity of the EOG model and the absence of adaptive or uncertainty-aware inference. Each limitation is accompanied in Chapter 8 by specific directions for how it could be addressed in future work.

9.6 Future directions

Detailed future work directions are outlined in Section 8.4 and the following sections of Chapter 8. At a higher level, the most productive directions fall into three areas. The first is improving the visual representation of eyelid motion, through better segmentation models, explicit blink parameterisation and loss functions that weight blink periods more heavily. The second is broadening the scope of the system, both in terms of the ocular signals it models (extending beyond vertical EOG to horizontal and torsional components) and the conditions under which it operates (moving from controlled laboratory settings to mobile, naturalistic and clinical environments). The third is integration with more powerful downstream pipelines, including deep learning-based artefact removal and online adaptive calibration, which would allow

VEOS to be deployed in real-time systems where sensor access and recording conditions may vary. Taken together, these directions form a coherent research programme that builds directly on the foundations established in this thesis.

9.7 Closing remarks

This thesis has explored the idea that video of the eyes and face, processed with modern computer vision methods, can stand in for dedicated ocular electrodes in the service of EEG artefact handling. Through a combination of theory, instrumentation and empirical evaluation, it has shown that this idea is feasible and practically useful, though not yet a complete substitute for EOG in all cases.

VEOS does not eliminate ocular artefacts, nor does it render traditional EOG obsolete. Instead, it offers a complementary route that extends the reach of EOG-based methods to situations where only video is available, or where additional electrodes are undesirable. In doing so, it contributes to a broader move towards multimodal, context-aware EEG analysis and towards BCI systems that are more robust, more flexible and more closely aligned with the realities of use outside the laboratory.

The work presented here is a step rather than an endpoint. Its value lies in the specific models and results and the demonstration that integrating video and EEG at the level of basic signal processing can yield tangible benefits. Future developments in sensing, modelling and computation will undoubtedly refine and extend these ideas. The hope is that VEOS and the concepts developed around it will form part of that trajectory, helping to make EEG-based systems more resilient in the presence of the very behaviours that make them relevant in the first place.

Appendix A

Implementation

A.1 Implementation details for VEOS

This appendix collects implementation details that are useful for reproducing the results in Chapters 4–7, but would be overly long in the main text. It covers the eye-tracking and feature extraction pipeline, the architectures and training settings used for EOG inference, and the precise Lab Streaming Layer (LSL) configuration of the multimodal acquisition platform.

A.1.1 Monocular eye-tracking pipeline

Eye-tracking in VEOS follows the conceptual description in Sections 3.2 and 3.3. The practical steps for each video frame are:

1. Convert the input frame to greyscale and, where needed, resize it to a fixed processing resolution (e.g. 1280×720 for the VEOS development dataset, 320×240 for Eye-BCI eye-region video).
2. Detect the face region using the MediaPipe Face Mesh front-end; if multiple faces are detected, select the one closest to the previous frame’s location.
3. Estimate the full face mesh and extract a subset of landmarks corresponding to the inner and outer canthi of each eye, the midpoints of the upper and lower

eyelids, the approximate pupil centre for each eye, and a small set of eyebrow landmarks.

4. Compute a similarity transform that maps a set of stable facial landmarks (typically canthi and nose bridge) to a canonical template, and apply it to all extracted landmarks to obtain face-normalised coordinates.
5. Derive eye-centric coordinates by computing the canthus midpoint for each eye as an origin, expressing the pupil and lid midpoints relative to that origin, and dividing by the inter-canthal distance to obtain dimensionless positions.
6. Apply light temporal smoothing (for example, a short moving average) to the normalised landmarks before taking finite differences, in order to reduce amplification of detection noise.

Frames where the face mesh is missing or clearly inconsistent (e.g. sudden jumps exceeding several eye-widths) are flagged as invalid. For those frames the corresponding feature vectors are either dropped or linearly interpolated from neighbouring valid frames, depending on the density of gaps.

A.1.2 Feature construction

From the normalised landmarks, VEOS constructs a per-frame feature vector $\mathbf{f}_k$ comprising several components. These include horizontal and vertical pupil positions for each eye in eye-centric coordinates, eyelid aperture for each eye (upper–lower lid distance, normalised by inter-canthal distance), simple asymmetry measures that capture differences in position and aperture between left and right eyes when both are visible, and eyebrow elevation measures relative to the canthus midpoint line.

Temporal derivatives are then added, consisting of first-order differences (approximate velocities) for each scalar feature and second-order differences (approximate accelerations) for the main position and aperture features.

For models with explicit temporal context, a feature window $\mathbf{z}_k$ is formed by

concatenating $\mathbf{f}$ over a fixed context length L ,

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{f}_{k-L+1}^\top & \dots & \mathbf{f}_k^\top \end{bmatrix}^\top.$$

In the experiments reported in Chapter 7, context windows of approximately ± 200 – 250 ms around the central frame were used (i.e. window lengths of 5–7 frames at 120–150 Hz), unless otherwise stated.

Before being passed to the learning models, all features are standardised per participant: each scalar feature is centred and scaled to unit variance using statistics computed on that participant’s training data only.

A.2 EOG inference models

This section specifies the architectures and training settings used in the EOG inference experiments of Chapter 7. Wherever precise numerical values are omitted in the main text, they are documented here.

A.2.1 Linear baselines

The simplest baseline is a regularised linear regression model of the form

$$\hat{y}_k = \mathbf{w}^\top \mathbf{z}_k + b,$$

trained to predict either horizontal or vertical EOG from the feature window $\mathbf{z}_k$ using ridge regression.

The loss function is

$$\mathcal{L}(\mathbf{w}, b) = \frac{1}{T} \sum_{k=1}^T (y_k - \hat{y}_k)^2 + \lambda \|\mathbf{w}\|_2^2,$$

with regularisation strength λ chosen via nested cross-validation on the training set.

A “static” variant uses only the instantaneous feature vector $\mathbf{f}_k$ (no temporal

context), while a slightly richer linear model uses the concatenated feature window $\mathbf{z}_k$.

A.2.2 Windowed multilayer perceptrons

The windowed MLP models used in Sections 5.7 and 7.3 have the following structure: an input layer of size D corresponding to the dimensionality of $\mathbf{z}_k$, one or two hidden layers with rectified linear unit (ReLU) activations, and an output layer with a single linear neuron for scalar EOG prediction.

A typical configuration is summarised in Table A.1.

Table A.1: Example MLP architecture for EOG inference.

Layer	Size / type	Notes
Input	D	Windowed feature vector $\mathbf{z}_k$
Hidden layer 1	128 units, ReLU	Fully connected, optional dropout (e.g. 0.2)
Hidden layer 2	64 units, ReLU	Optional; may be omitted in simpler models
Output	1 unit, linear	Predicted EOG sample $\hat{y}_k$

Training details (typical values) are as follows. Models are trained with the Adam optimiser, an initial learning rate $\text{lr} \in [10^{-4}, 10^{-3}]$, batch sizes $B \in [64, 256]$, and up to 100 epochs with early stopping based on validation loss, using mean squared error on per-participant standardised targets as the loss.

Exact hyperparameters may vary across experiments; deviations are noted in the corresponding result sections.

A.2.3 Temporal convolutional networks

Temporal convolutional networks (TCNs) are used as the main non-linear temporal model in Chapter 7. A typical TCN block has 1D convolutions with kernel size k (for example $k = 3$) and increasing dilation factors $d = 1, 2, 4, \dots$, together with residual connections around each block, weight normalisation or batch normalisation if needed for stable training, and ReLU activations after each convolution.

An example configuration for the VEOS development dataset is given in Table A.2.

Table A.2: Example TCN architecture for EOG inference.

Component	Parameters	Description
Input	$C \times T_{\text{win}}$	C feature channels over a temporal window
Block 1	1D conv, 64 filters, $k = 3$, $d = 1$	Residual; ReLU
Block 2	1D conv, 64 filters, $k = 3$, $d = 2$	Residual; ReLU
Block 3	1D conv, 64 filters, $k = 3$, $d = 4$	Residual; ReLU
Block 4	1D conv, 64 filters, $k = 3$, $d = 8$	Residual; ReLU
Global pool	Temporal average	Collapse time dimension
Output	Fully connected, 1 unit, linear	Predicted EOG

Training for TCN models follows the same general regime as for MLPs (Adam optimiser, mean squared error loss, early stopping), with learning rates and batch sizes tuned per dataset.

A.2.4 Hybrid forward models

The hybrid model described in Chapter 7 combines a simple linear baseline with a non-linear residual term,

$$\hat{y}_k = \alpha_s y_k^{\text{lin}} + \beta_s + r_k,$$

where y_k^{lin} is the output of a subject-independent linear model, α_s and β_s are per-subject scale and offset parameters, and r_k is a non-linear residual predicted by a TCN or MLP.

The baseline is trained first on pooled data; per-subject α_s, β_s are then estimated via least-squares on a short calibration segment; finally, the residual network is trained on the remaining training data to minimise the error between true EOG and the linearly corrected baseline.

A.3 Training and evaluation configuration

A.3.1 Target preprocessing

For both datasets, EOG (and, where applicable, EMG) targets are:

1. Band-pass filtered in a passband appropriate for eye-movement artefacts (e.g. 0.1–15 Hz for VEOG), using zero-phase IIR or FIR filters to avoid phase distortion.
2. Baseline-corrected per session using a high-pass filter or polynomial detrending to remove very slow drifts.
3. Resampled or interpolated to match the video frame times using the common LSL timestamps (development dataset) or the synchronisation tables provided with Eye-BCI.
4. Standardised per participant (zero mean, unit variance) for learning; original units are recovered when reporting RMSE in microvolts.

A.3.2 Cross-validation and data splits

The participant-wise data split strategy described in Section 6.3 is implemented as follows. For each experiment, participants are partitioned into training, validation and test folds such that no participant appears in more than one fold. Within the training fold, runs or sessions are further split into training and validation subsets on a per-participant basis. Hyperparameters (e.g. learning rate, regularisation) are selected based on validation performance only, and the test data are used exactly once for final evaluation. When combining datasets (e.g. training on VEOS development and testing on Eye-BCI), no Eye-BCI data are used for training, except where explicitly noted for per-subject calibration of baseline parameters.

A.3.3 Optimisation details

Unless otherwise stated, optimisation settings are as follows. The Adam optimiser with default β_1, β_2 parameters is used, learning rates in the range 10^{-4} to 5×10^{-4} are chosen via validation, gradient clipping (for example global norm clipping at a value between 1 and 5) is applied to avoid instabilities in TCN training, and early

stopping is triggered if validation loss does not improve for N_{patience} epochs (typically $N_{\text{patience}} = 10$).

A.4 Multimodal platform implementation details

A.4.1 LSL stream configuration

Table A.3 summarises the LSL streams used in the VEOS development platform (Chapter 4).

Table A.3: LSL streams in the VEOS development platform.

Stream name	Type	Channels	Sampling rate	Description
ANT_EEG	EEG	32	1000 Hz	Scalp EEG (reduced 10–20 montage)
OpenBCI_EOG	EOG	2	1000 Hz	Horizontal and vertical EOG
Video	Video	1	120 Hz	720p face video (frame timestamps only)
Markers	Markers	1	irregular	Stimulus events, task states, LED triggers
Game_Events	Markers	1	irregular	Game-specific events (optional)

Each stream exposes channel labels, units and metadata consistent with its type. For example, EEG channels use SI units (microvolts) and are labelled according to the 10–20 system; marker streams encode event codes as integers and, where supported, include textual descriptions in the LSL metadata.

A.4.2 Directory structure and file naming

The conceptual directory structure shown in Figure 4.7 is instantiated on disk with a top-level directory `veos_development/` that contains participant directories `subj01/`, `subj02/`, `...`, and within each participant directory a set of session subdirectories as described below.

Each session directory contains at minimum:

- One LSL recording file, e.g. `subj01_session01_lsl.xdf`.
- One raw video file, e.g. `subj01_session01_video.mp4`.
- A JSON configuration file, `config.json`, containing:
 - EEG and EOG montage information.
 - Acquisition software version numbers and relevant settings.
 - Notes on deviations from the standard protocol (e.g. glasses, unusual posture).
- Optional log files from task software (e.g. OpenViBE scenarios or game engine logs).

Analysis scripts assume this layout and derive participant and session identifiers directly from directory and file names, which simplifies batch processing and cross-referencing between modalities.

A.4.3 Software environment

For completeness, the main software components used in the VEOS development platform are as follows: Python (version 3.10) for recording scripts, feature extraction and modelling, the Lab Streaming Layer (LSL) framework for time synchronisation and multimodal recording, MediaPipe Face Mesh (as referenced in Chapter 5) for landmark detection, and OpenViBE for BCI-style stimulus control in selected tasks.

Appendix B

Ethics

This appendix documents the ethical framework under which the VEOS development dataset was collected. It summarises the ethics approval, inclusion and exclusion criteria, recruitment procedures, and the information and consent materials provided to participants. The aim is to make the practical aspects of the study transparent without overloading the main chapters.

B.1 Ethics approval

All recordings for the VEOS development dataset (Chapters 4 and 6) were conducted in accordance with the regulations of the Dublin City University Research Ethics Committee (DCU REC).

Ethical approval was granted following notification review under REC reference **DCUREC/2021/175**. The original approved protocol was titled *Workload Stress in Blended Control of Autonomous Vehicles* and was subsequently recategorised and amended under the updated project title *Multimodal Brain–Computer Interface Denoising* to cover the multimodal acquisition and BCI paradigms described in this thesis. Formal approval was issued on 7 March 2022.

The approved protocol covered:

- acquisition of EEG and EOG using non-invasive scalp and periocular electrodes;

- acquisition of high-frame-rate face video during computer-based tasks;
- storage and analysis of pseudonymised recordings for research purposes;
- optional re-use of pseudonymised data in future studies by the same research group, subject to REC approval.

The study conformed to the principles of the Declaration of Helsinki (2013 revision) and to DCU institutional guidelines on research with human participants.

B.2 Inclusion and exclusion criteria

The target population for the VEOS development study comprised healthy adults with normal or corrected-to-normal vision and no known neurological disorders.

Inclusion criteria

Participants were eligible to take part if they:

- were aged 18 years or older;
- reported being in general good health and able to complete computer-based tasks;
- reported normal or corrected-to-normal vision (spectacles or contact lenses permitted);
- were able to understand the DCU plain-language information sheet and provide written informed consent;
- were able to sit comfortably in front of a monitor or head-mounted display for approximately 60–90 minutes including breaks.

Exclusion criteria

Participants were excluded if they:

- were under 18 years of age;
- were unable to give informed consent;
- had uncorrected visual impairments that would interfere with viewing the stimuli;
- reported a history of epilepsy, seizures or other serious neurological conditions that could make EEG recording or visual stimulation inappropriate;
- had known allergies to standard EEG electrode gels (a patch test was performed in advance where necessary);
- declined to be video recorded or to have their data stored in anonymised or pseudonymised form.

Participants who wore spectacles were included where any glare or occlusion did not compromise eye or face tracking; such cases were noted in the session metadata.

B.3 Recruitment and compensation

Participants were recruited from the general adult population via:

- local announcements and internal mailing lists within the university;
- word-of-mouth within the School of Computing and collaborating organisations;
- local staff notices at collaborating sites involved in the original blended-control study.

Recruitment materials and e-mail invitations briefly described:

- the general nature of the study (EEG and video recordings during computer-based tasks);

- approximate session duration (around 90 minutes, including set-up and breaks);
- basic eligibility criteria;
- contact details for the named researcher and supervisor.

Participation was voluntary. Potential participants received a plain-language statement and were informed that they could decline to participate or withdraw at any time without providing a reason and without any negative consequences for employment or study.

No direct financial or course-credit compensation was offered; participation was on a voluntary basis in line with the approved protocol.

B.4 Information provided to participants

Before each session, participants received a written plain-language information sheet and had the opportunity to ask questions. The information sheet covered:

- the purpose of the study: investigating stress and workload in blended control of autonomous vehicles and, in the amended protocol, investigating EEG denoising using multimodal sensor streams (EEG and video);
- the procedures involved: fitting of an EEG cap, placement of electrodes, positioning in front of a monitor or head-mounted display, and completion of simple computer-based tasks while EEG and facial video were recorded;
- the approximate time commitment (around 90 minutes from start to finish, including breaks);
- the voluntary nature of participation and the right to withdraw at any time without disadvantage;
- potential risks and discomforts (for example mild discomfort from the EEG cap and possible minor skin irritation from electrode gel);

- the absence of direct clinical benefit and the broader research aims;
- how data would be handled, pseudonymised and, where appropriate, made available to other researchers in de-identified form;
- contact details for the research team, the supervisor and the DCU Research Ethics Committee for queries or complaints.

A short verbal briefing summarised the key points, and participants were given time to read the information sheet in full before signing the consent form.

Summary paragraph

The following paragraph represents the DCU plain-language statement used for the VEOS development dataset:

In this study we record electrical activity from your scalp (EEG) while you perform simple computer-based tasks. We also record video of your eyes and face so that we can study how movements such as blinks and head motion affect the EEG. The recordings are for research purposes only and will not be used to make any clinical judgements. You are free to stop the experiment at any time without giving a reason, and you may ask for your data to be withdrawn up until it has been anonymised and included in group analyses.

B.5 Consent procedure

Written informed consent was obtained from each participant before any data collection began. The consent form followed the DCU REC template and confirmed that the participant:

- had read and understood the information sheet;
- had had the opportunity to ask questions and received satisfactory answers;

- understood that participation was entirely voluntary and that they could withdraw at any time without penalty;
- gave permission for their brain activity (EEG) and task performance to be recorded;
- gave permission for video of their face and eyes to be recorded during the experimental tasks;
- understood that they could withdraw permission to use their data up to the point at which it was anonymised;
- understood how their data would be managed, stored and, where relevant, shared in anonymised form;
- could indicate whether they agreed to the use of their anonymised data in future related research projects and publications.

Participants signed and dated the consent form; the researcher countersigned to confirm that the study had been explained in a way the participant could understand. One copy of the signed form was retained securely by the research team; another copy was offered to the participant for their records.

B.6 Data protection and anonymisation

Data protection procedures followed DCU policy, GDPR and the additional safeguards for health research documented in the project-specific Data Protection Impact Assessment (DPIA) for *Workload Stress in Blended Control of Autonomous Vehicles*. The overall approach closely follows the principles and practical guidelines developed in [128] for GDPR-compliant handling of multimodal physiological data.

Key aspects were:

Pseudonymisation: each participant was assigned a numeric or alphanumeric code (e.g. subj01, subj02). These codes appear in the data files and in all anal-

yses. Identifying details (name, e-mail, signed consent) were stored separately in locked cabinets or encrypted files, distinct from the EEG and video data.

Secure storage: raw EEG and associated metadata were stored on password-protected DCU machines and institutional servers with restricted access. Where temporary storage on laptops was required for acquisition or preprocessing, those devices were encrypted and in the custody of the research team.

Video handling: for the original blended-control protocol, full-frame video of the participants' faces was processed to extract non-identifying motion and eye-region metadata; once these features were derived, the original video files were scheduled for destruction. For later VEOS-specific sessions, close-up eye video and face video were retained in pseudonymised form for algorithm development, with any externally shared datasets cropped or processed to reduce identifiability.

Retention: the participant information sheet specified that research data would be retained on secure DCU infrastructure and that de-identified data could be made available to support other researchers. In line with DCU guidance at the time of approval, identifiable records (such as consent forms and linking keys) are to be destroyed after a maximum of ten years, while fully anonymised EEG and derived feature sets may be archived for longer as part of open research datasets.

Publication and sharing: any data shared outside the immediate research group (for example via public repositories) must be de-identified so that individual participants cannot be readily re-identified from the shared files; this includes removal of direct identifiers and appropriate cropping or transformation of video material where needed.

Participants were informed that, in rare circumstances, confidentiality of research data may be overridden by courts or lawful authorities, and that DCU would seek to protect confidentiality as far as possible within the law.

B.7 Adverse events and participant feedback

No serious adverse events occurred during data collection under the approved protocols. The main minor issues anticipated in the ethics application and reported during piloting were:

- occasional discomfort or mild pressure from the EEG cap during longer blocks, mitigated by adjusting the cap and providing rest breaks;
- transient skin impressions or mild irritation from electrode gel and adhesive, which resolved shortly after the session;
- brief eye strain or fatigue during demanding visual tasks, addressed by inserting additional breaks where required.

A patch test for the electrode gel was performed at least 24 hours before recording where there was any concern about allergy; participants with a suspected reaction would have been excluded from EEG recording, although no such cases occurred in the VEOS development data used in this thesis.

Participants were invited to comment informally on task difficulty, comfort and the overall experience at the end of each session. Feedback was used to refine task ordering, block lengths and break schedules in subsequent recordings, while remaining within the bounds of the approved ethics protocol.

The procedures described in this appendix ensure that the VEOS development dataset was collected in a way that respects participant autonomy, minimises risk and preserves privacy, while still providing the rich multimodal recordings required for the analyses in the main body of the thesis.

Bibliography

- [1] R. Abiri, S. Borhani, E. W. Sellers, Y. Jiang, and X. Zhao, “A comprehensive review of eeg-based brain–computer interface paradigms,” *Journal of Neural Engineering*, vol. 16, no. 1, p. 011 001, 2019. DOI: 10.1088/1741-2552/aaf12e.
- [2] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, “Automated detection of neurological disorders using wavelet-based statistical features extracted from eeg signals,” *International Journal of Neural Systems*, vol. 28, no. 07, p. 1 750 054, 2018.
- [3] Y. An, H.-K. Lam, and S. H. Ling, “Auto-denoising for EEG signals using generative adversarial network,” *Sensors*, vol. 22, no. 5, p. 1750, 2022. DOI: 10.3390/s22051750.
- [4] S. Appelhoff and T. Stenner, “In com we trust: Feasibility of usb-based event marking,” *Behavior Research Methods*, 2021. DOI: 10.3758/s13428-021-01571-z.
- [5] S. Appelhoff et al., “Mne-bids: Organizing electrophysiological data into the BIDS format and facilitating their analysis,” *Journal of Open Source Software*, vol. 4, no. 44, p. 1896, 2019. DOI: 10.21105/joss.01896.
- [6] M. Arns et al., “Neurofeedback: One of today’s techniques in psychiatry?” *L’Encephale*, vol. 43, no. 2, pp. 135–145, 2017. DOI: 10.1016/j.encep.2016.10.005.

- [7] M. Arvaneh, I. Umata, Y. Ishii, I. H. Robertson, and T. Mima, “A P300-based brain-computer interface for improving attention,” *Frontiers in Human Neuroscience*, vol. 12, p. 524, 2018. DOI: 10.3389/fnhum.2018.00524.
- [8] M. A. Awais, P. Redmond, T. E. Ward, and G. Healy, “AMBER: Advancing multimodal brain-computer interfaces for enhanced robustness – a dataset for naturalistic settings,” *Frontiers in Neuroergonomics*, 2023.
- [9] M. A. Awais, T. Ward, P. Redmond, and G. Healy, “From lab to life: Assessing the impact of real-world interactions on the operation of rapid serial visual presentation-based brain-computer interfaces,” *Journal of Neural Engineering*, 2024. DOI: 10.1088/1741-2552/ad5d17.
- [10] H. Ayaz and B. Onaral, “Continuous monitoring of brain dynamics with functional near infrared spectroscopy as a tool for neuroergonomic research: Empirical examples and a technological development,” *Frontiers in Human Neuroscience*, vol. 7, p. 871, 2013. DOI: 10.3389/fnhum.2013.00871.
- [11] S. Bai, J. Z. Kolter, and V. Koltun, “An empirical evaluation of generic convolutional and recurrent networks for sequence modeling,” *arXiv preprint arXiv:1803.01271*, 2018. [Online]. Available: <https://arxiv.org/abs/1803.01271>.
- [12] R. Barea et al., “Electrooculography for human-computer interaction,” *Biomedical Engineering, IEEE Transactions on*, vol. 49, no. 1, pp. 1–11, 2002.
- [13] V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, and M. Grundmann, “Blazeface: Sub-millisecond neural face detection on mobile gpus,” *arXiv preprint arXiv:1907.05047*, 2019. [Online]. Available: <https://arxiv.org/abs/1907.05047>.
- [14] S. Bentin, T. Allison, A. Puce, E. Perez, and G. McCarthy, “Electrophysiological studies of face perception in humans,” *Journal of cognitive neuroscience*, vol. 8, no. 6, pp. 551–565, 1996.

- [15] C. Berka et al., “Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks,” *Aviation, Space, and Environmental Medicine*, vol. 78, B231–B244, 2007.
- [16] A. Biasiucci et al., “Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke,” *Nature Communications*, vol. 9, no. 1, 2018. DOI: 10.1038/s41467-018-04673-z.
- [17] N. Birbaumer et al., “A spelling device for the paralysed,” *Nature*, vol. 398, no. 6725, pp. 297–298, 1999. DOI: 10.1038/18581.
- [18] G. Bradski, *The OpenCV Library*. 2000.
- [19] E. Brophy, P. Redmond, A. Fleury, M. De Vos, G. Boylan, and T. Ward, “Denoising EEG signals for real-world BCI applications using GANs,” *Frontiers in Neuroergonomics*, vol. 2, p. 805573, 2022.
- [20] A. Bulling et al., “Eye movement analysis for activity recognition,” *Ubicomp*, vol. 11, pp. 41–48, 2011.
- [21] Center for Open Science, *Open science framework (osf)*, <https://osf.io/>, Accessed: 2025-12-04, 2025.
- [22] N. Chaddad, C. Desrosiers, L. Hassan, et al., “Electroencephalography signal processing: A comprehensive review,” *Sensors*, vol. 23, no. 22, p. 9556, 2023.
- [23] L.-L. Chen, Y. Zhao, J. Zhang, and J. Zou, “Automatic detection of alertness/drowsiness from physiological signals using wavelet-based nonlinear features and machine learning,” *Expert Systems with Applications*, vol. 42, no. 21, pp. 7344–7355, 2015.
- [24] Y. Cheng, H. Wang, Y. Bao, and F. Lu, “Appearance-based gaze estimation with deep learning: A review and benchmark,” in *Proceedings of the 2021 ACM Symposium on Eye Tracking Research and Applications*, 2021, pp. 1–16. DOI: 10.1145/3448017.3457364.

- [25] J. L. Collinger et al., “High-performance neuroprosthetic control by an individual with tetraplegia,” *The Lancet*, vol. 381, no. 9866, pp. 557–564, 2013. DOI: 10.1016/S0140-6736(12)61816-9.
- [26] S. Comani et al., “Monitoring students’ attention in a classroom through beta-band eeg signals,” *Electronic Imaging*, vol. 2015, no. 16, pp. 1–8, 2015.
- [27] R. J. Croft and R. J. Barry, “Contamination of eeg with eog: An fft with regression method for reduction,” *Clinical Neurophysiology*, vol. 116, no. 5, pp. 983–993, 2005.
- [28] A. D’Ausilio, “Arduino: A low-cost multipurpose lab equipment,” *Behavior Research Methods*, vol. 44, no. 2, pp. 305–313, 2012. DOI: 10.3758/s13428-011-0163-z.
- [29] DANDI Archive, *Dandi archive*, <https://dandiarchive.org/>, Accessed: 2025-12-04, 2025.
- [30] J. B. Denis, J. Kansaku, W. Choi, L. Wang, N. Naseer, and H. Namba, “Advances in P300 brain–computer interface spellers: Toward paradigm design and performance evaluation,” *Frontiers in Human Neuroscience*, vol. 16, p. 1077717, 2022. DOI: 10.3389/fnhum.2022.1077717.
- [31] C. Dewi, R.-C. Chen, X. Jiang, and H. Yu, “Adjusting the eye aspect ratio for strong eye blink detection,” *PeerJ Computer Science*, e943, 2022. DOI: 10.7717/peerj-cs.943.
- [32] O. Dimigen, “Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments,” *NeuroImage*, vol. 207, p. 116117, 2020.
- [33] Docker, Inc., *Docker documentation*, <https://docs.docker.com/>, Accessed: 2025-12-04, 2025.
- [34] Y. Dong et al., “An approach for EEG denoising based on Wasserstein generative adversarial network,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3524–3534, 2023. DOI: 10.1109/TNSRE.2023.3309815.

- [35] L. A. Farwell and E. Donchin, “Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials,” *Electroencephalography and Clinical Neurophysiology*, vol. 70, no. 6, pp. 510–523, 1988. DOI: 10.1016/0013-4694(88)90149-6.
- [36] M. Fatourechi, A. Bashashati, R. K. Ward, and G. E. Birch, “Emg and eog artifacts in brain computer interface systems: A survey,” *Clinical Neurophysiology*, vol. 118, no. 3, pp. 480–494, 2007. DOI: 10.1016/j.clinph.2006.10.019.
- [37] N. Galley, R. Schleicher, and L. Galley, *Blink parameters as indicators of driver sleepiness: Possibilities and limitations*, Technical report, 2003.
- [38] M. Golz, D. Sommer, M. Chen, D. Mandic, and U. Trutschel, “Feature fusion for the detection of microsleep events,” *Journal of VLSI Signal Processing Systems for Signal, Image and Video Technology*, vol. 49, no. 2, pp. 329–342, 2007.
- [39] C. B. Gonsisko, D. P. Ferris, and R. J. Downey, “Icanclean improves independent component analysis of mobile brain imaging with EEG,” *Sensors*, vol. 23, no. 2, p. 928, 2023. DOI: 10.3390/s23020928.
- [40] K. J. Gorgolewski et al., “BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods,” *PLoS Computational Biology*, vol. 13, no. 3, e1005209, 2017. DOI: 10.1371/journal.pcbi.1005209.
- [41] K. J. Gorgolewski et al., “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments,” *Scientific Data*, vol. 3, p. 160 044, 2016. DOI: 10.1038/sdata.2016.44.
- [42] D. Gorjan, K. Gramann, K. De Pauw, and U. Marusic, “Removal of movement-induced EEG artifacts: Current state of the art and guidelines,” *Journal of Neural Engineering*, vol. 19, no. 1, p. 011 004, 2022.
- [43] K. Gramann et al., “Mobile EEG in neurourbanism: A critical review and guidelines,” *Journal of Environmental Psychology*, p. 102 308, 2024. DOI: 10.1016/j.jenvp.2024.102308.

- [44] F. Gustafsson, “Determining the initial states in forward-backward filtering,” *IEEE Transactions on Signal Processing*, vol. 44, no. 4, pp. 988–992, 1996. DOI: 10.1109/78.492552.
- [45] E. Guttmann-Flury, X. Sheng, and X. Zhu, “Dataset combining EEG, eye-tracking, and high-speed video for ocular activity analysis across BCI paradigms,” *Scientific Data*, vol. 12, p. 587, 2025. DOI: 10.1038/s41597-025-04861-9.
- [46] S. Haegens, H. Cousijn, G. Wallis, P. J. Harrison, and A. C. Nobre, “Alpha-band oscillations, attention, and controlled access to stored information,” *Trends in Cognitive Sciences*, vol. 18, no. 12, pp. 606–617, 2014.
- [47] M. Hanke, Y. O. Halchenko, et al., “Datalad: Distributed system for joint management of code, data, and their relationship,” *Journal of Open Source Software*, vol. 6, no. 63, p. 3262, 2021. DOI: 10.21105/joss.03262.
- [48] D. W. Hansen and Q. Ji, “In the eye of the beholder: A survey of models for eyes and gaze,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, 2009.
- [49] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2004.
- [50] HED Working Group, *Hierarchical event descriptors (hed)*, <https://www.hedtags.org/>, Accessed: 2025-12-04, 2025.
- [51] D. Hendrycks and K. Gimpel, “Gaussian error linear units (gelus),” *arXiv preprint arXiv:1606.08415*, 2016. [Online]. Available: <https://arxiv.org/abs/1606.08415>.
- [52] C. S. Herrmann, D. Strüber, R. F. Helfrich, and A. K. Engel, “Gamma oscillations: The functional key of working memory?” *Trends in Cognitive Sciences*, vol. 20, no. 12, pp. 872–879, 2016.
- [53] J. Hess, *Git-annex*, <https://git-annex.branchable.com/>, Accessed: 2025-12-04, 2025.

- [54] C. Holdgraf et al., “iEEG-BIDS, extending the brain imaging data structure specification to human intracranial electrophysiology,” *Scientific Data*, vol. 6, p. 102, 2019. DOI: 10.1038/s41597-019-0105-7.
- [55] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press, 2011.
- [56] I. T. C. Hooge, M. Nyström, T. H. W. Cornelissen, J. Hooge, R. Andersson, and R. S. Hessels, “The fundamentals of eye tracking part 4: Tools for conducting an eye-tracking study,” *Behavior Research Methods*, 2025. DOI: 10.3758/s13428-024-02529-7. [Online]. Available: <https://doi.org/10.3758/s13428-024-02529-7>.
- [57] M. S. Hossain and G. Muhammad, “Driver drowsiness detection using fusion of electroencephalogram and eye aspect ratio,” *Expert Systems with Applications*, vol. 132, pp. 82–91, 2019.
- [58] A. Housholder, A. V. N. Bontrager, L. E. Burrell, R. L. Cooper, and D. C. Duarte, “Evaluating accuracy of the tobii eye tracker 5,” in *Augmented Cognition*, ser. Lecture Notes in Computer Science, D. D. Schmorow and C. M. Fidopiastis, Eds., vol. 13311, Cham: Springer, 2022, pp. 488–497. DOI: 10.1007/978-3-030-98404-5_36.
- [59] A. Hyvärinen and E. Oja, “Independent component analysis: Algorithms and applications,” *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000. DOI: 10.1016/S0893-6080(00)00026-5.
- [60] IEEE Standards Association, *Ieee std 1588-2019: Ieee standard for a precision clock synchronization protocol for networked measurement and control systems*, <https://standards.ieee.org/standard/1588-2019.html>, Accessed: 2025-12-04, 2019.

- [61] R. J. Jacob, “Eye movement-based human-computer interaction techniques: Toward non-command interfaces,” *Advances in Human-Computer Interaction*, vol. 4, pp. 151–190, 1993.
- [62] B. James, H. Sharabaty, and D. Esteve, “Automatic eog analysis: A first step toward automatic drowsiness scoring during wake–sleep transitions,” *Somnologie*, vol. 12, pp. 227–232, 2008.
- [63] Y. Jiao, P. Yong, L. Bao-Liang, C. Xiaoping, C. Shanguang, and W. Chunhui, “Recognizing slow eye movement for driver fatigue detection with a machine learning approach,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2014.
- [64] V. Jurcak, D. Tsuzuki, and I. Dan, “10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems,” *NeuroImage*, vol. 34, no. 4, pp. 1600–1611, 2007. DOI: 10.1016/j.neuroimage.2006.09.024.
- [65] A. Kaehler and G. Bradski, *Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library*. O’Reilly Media, 2016.
- [66] B. Kalita, N. Deb, and D. Das, “AnEEG: Leveraging deep learning for effective artifact removal in EEG data,” *Scientific Reports*, vol. 14, p. 24 234, 2024. DOI: 10.1038/s41598-024-75091-z.
- [67] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, “Real-time facial surface geometry from monocular video on mobile GPUs,” *arXiv preprint arXiv:1907.06724*, 2019. arXiv: 1907.06724 [cs.CV].
- [68] Y. Kartynnik, A. Ablavatski, I. Grishchenko, and M. Grundmann, “Real-time facial surface geometry from monocular video on mobile gpus,” *arXiv preprint arXiv:1907.06724*, 2019. [Online]. Available: <https://arxiv.org/abs/1907.06724>.

- [69] A. Kawala-Sterniuk et al., “Summary of over fifty years with brain–computer interfaces—a review,” *Brain Sciences*, vol. 11, no. 1, p. 43, 2021. DOI: 10.3390/brainsci11010043.
- [70] G. H. Klem, H. O. Lüders, H. H. Jasper, and C. Elger, “The ten-twenty electrode system of the international federation,” *Clinical Neurophysiology*, vol. 110, no. 1, pp. 75–82, 1999. DOI: 10.1016/S1388-2457(98)00053-0.
- [71] C. A. Kothe, *Lab streaming layer (lsl)*, <https://github.com/sccn/labstreaminglayer>, Accessed: 2025-12-04, 2014.
- [72] C. A. Kothe and contributors, *Extensible data format (xdf)*, <https://github.com/sccn/xdf>, Accessed: 2025-12-04, 2025.
- [73] C. A. Kothe and contributors, *Lab streaming layer (lsl)*, <https://github.com/sccn/labstreaminglayer>, Accessed: 2025-12-04, 2025.
- [74] D. Kumar and E. Poole, “Classification of eog for human-computer interface,” in *Proceedings of the Second Joint EMBS/BMES Conference*, Houston, TX, 2002.
- [75] G. M. Kurtzer, V. Sochat, and M. W. Bauer, “Singularity: Scientific containers for mobility of compute,” *PLOS ONE*, vol. 12, no. 5, e0177459, 2017. DOI: 10.1371/journal.pone.0177459.
- [76] M. A. Lebedev and M. A. L. Nicolelis, “Brain-machine interfaces: Past, present and future,” *Trends in Neurosciences*, vol. 29, no. 9, pp. 536–546, 2006.
- [77] A. Lécuyer, F. Lotte, R. B. Reilly, R. Leeb, M. Hirose, and M. Slater, “Brain-computer interfaces, virtual reality, and videogames,” *Computer*, vol. 41, no. 10, pp. 66–72, 2008. DOI: 10.1109/MC.2008.410.
- [78] S. Liu et al., “A review of steady state visual evoked potentials based brain computer interface in the past decade,” *Brain Science Advances*, vol. 8, no. 4, pp. 258–275, 2022. DOI: 10.26599/BSA.2022.9050022.

- [79] Y. Long et al., “Multimodal detection of external and internal attention in virtual reality using EEG and eye tracking,” in *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 2024, pp. 1–17. DOI: 10.1145/3613904.3642515.
- [80] M. Lopez-Gordo et al., “Dry eeg electrodes,” *Sensors*, vol. 14, no. 7, pp. 12 847–12 870, 2014.
- [81] S. Makeig et al., “Independent component analysis of electroencephalographic data,” pp. 145–151, 1996.
- [82] N. Mammone and F. C. Morabito, “Enhanced automatic artifact detection based on independent component analysis and renyi’s entropy,” *Neural Networks*, vol. 34, pp. 303–315, 2012. DOI: 10.1016/j.neunet.2012.07.012.
- [83] K. Mamun et al., “Application of electrooculogram (eog) in human computer interface (hci),” *Int. J. Computer Appl*, vol. 145, no. 2, pp. 23–27, 2016.
- [84] C. J. Markiewicz et al., “Openneuro: An open resource for sharing of neuroimaging data,” *eLife*, vol. 10, e71774, 2021. DOI: 10.7554/eLife.71774.
- [85] D. Marshall, D. Coyle, S. Wilson, and M. Callaghan, “Games, gameplay, and bci: The state of the art,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 5, no. 2, pp. 82–99, 2013. DOI: 10.1109/TCIAIG.2013.2263555.
- [86] H. Marzbani, H. R. Marateb, and M. Mansourian, “Neurofeedback: A comprehensive review on system design, methodology and clinical applications,” *Basic and Clinical Neuroscience*, vol. 7, no. 2, pp. 143–158, 2016. DOI: 10.15412/J.BCN.03070208.
- [87] O. Maslova et al., “Review of electroencephalography-based brain computer interface spellers,” *Frontiers in Human Neuroscience*, vol. 17, p. 1216648, 2023. DOI: 10.3389/fnhum.2023.1216648.

- [88] C. M. Michel and D. Brunet, “Eeg source imaging: A practical review of the analysis steps,” *Frontiers in Neurology*, vol. 10, p. 325, 2019. DOI: 10.3389/fneur.2019.00325.
- [89] D. L. Mills, J. Martin, J. Burbank, and W. Kasch, “Network time protocol version 4: Protocol and algorithms specification,” IETF, RFC 5905, 2010, Accessed: 2025-12-04. [Online]. Available: <https://www.rfc-editor.org/rfc/rfc5905>.
- [90] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, “Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features,” *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011. DOI: 10.1111/j.1469-8986.2010.01061.x.
- [91] F. Mölder et al., “Sustainable data analysis with snakemake,” *F1000Research*, vol. 10, p. 33, 2021. DOI: 10.12688/f1000research.29032.2.
- [92] J. Moreno-Arjonilla, A. López-Ruiz, J. R. Jiménez-Pérez, J. E. Callejas-Aguilera, and J. M. Jurado, “Eye-tracking on virtual reality: A survey,” *Virtual Reality*, 2023. DOI: 10.1007/s10055-023-00903-y. [Online]. Available: <https://doi.org/10.1007/s10055-023-00903-y>.
- [93] L. R. D. Murthy et al., “I2DNet: Design and real-time evaluation of appearance-based gaze estimation network,” *IEEE Access*, vol. 9, pp. 141 931–141 943, 2021. DOI: 10.1109/ACCESS.2021.3119786.
- [94] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” in *Proceedings of the 27th International Conference on Machine Learning*, 2010.
- [95] K. Nathan and J. L. Contreras-Vidal, “Neglecting the effect of bioelectric fields on multimodal bci performance degenerates information transfer rates,” *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 12, pp. 2031–2036, 2005. DOI: 10.1109/TBME.2005.856716.

- [96] Neurodata Without Borders, *Neurodata without borders (nwb)*, <https://www.nwb.org/>, Accessed: 2025-12-04, 2025.
- [97] L. F. Nicolas-Alonso and J. Gomez-Gil, “Brain computer interfaces, a review,” *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012.
- [98] D. C. Niehorster, F. W. Cornelissen, K. Holmqvist, I. T. C. Hooge, and R. S. Hessels, “What to expect from your remote eye-tracker when participants are unrestrained,” *Behavior Research Methods*, vol. 50, no. 1, pp. 213–226, 2018. DOI: 10.3758/s13428-017-0863-0. [Online]. Available: <https://doi.org/10.3758/s13428-017-0863-0>.
- [99] D. C. Niehorster, R. S. Hessels, J. S. Benjamins, M. Nyström, and I. T. C. Hooge, “The impact of slippage on the data quality of head-worn eye trackers,” *Behavior Research Methods*, vol. 52, no. 3, pp. 1140–1160, 2020. DOI: 10.3758/s13428-019-01307-0. [Online]. Available: <https://doi.org/10.3758/s13428-019-01307-0>.
- [100] D. C. Niehorster, T. Santini, R. S. Hessels, I. T. C. Hooge, E. Kasneci, and M. Nyström, “Characterizing gaze position signals and synthesizing noise during fixations in eye-tracking data,” *Behavior Research Methods*, vol. 52, no. 6, pp. 2515–2534, 2020. DOI: 10.3758/s13428-020-01400-9. [Online]. Available: <https://doi.org/10.3758/s13428-020-01400-9>.
- [101] D. C. Niehorster, T. Santini, R. S. Hessels, I. T. C. Hooge, E. Kasneci, and M. Nyström, “Is apparent fixational drift in eye-tracking data due to filters or eyeball rotation?” *Behavior Research Methods*, vol. 53, no. 1, pp. 311–324, 2021. DOI: 10.3758/s13428-020-01414-3. [Online]. Available: <https://doi.org/10.3758/s13428-020-01414-3>.
- [102] V. V. Nikulin, G. Nolte, and G. Curio, “A novel method for reliable and fast extraction of neuronal eeg/meg oscillations on the basis of spatio-spectral decomposition,” *Neuroimage*, vol. 34, no. 4, pp. 1447–1461, 2007. DOI: 10.1016/j.neuroimage.2006.11.007.

- [103] S. Noble, E. Woods, T. Ward, and J. Ringwood, “Adaptive P300-based brain-computer interface for attention training: Protocol for a randomized controlled trial,” *JMIR Research Protocols*, vol. 12, e46135, 2023. DOI: 10.2196/46135.
- [104] M. Nyström and R. Andersson, “What is a blink? classifying and characterizing blinks in eye openness signals,” *Behavior Research Methods*, 2024. DOI: 10.3758/s13428-023-02333-9. [Online]. Available: <https://doi.org/10.3758/s13428-023-02333-9>.
- [105] M. Nyström, I. T. C. Hooge, T. H. W. Cornelissen, J. Hooge, R. Andersson, and R. S. Hessels, “The fundamentals of eye tracking part 3: How to choose an eye tracker,” *Behavior Research Methods*, 2025. DOI: 10.3758/s13428-024-02587-x. [Online]. Available: <https://doi.org/10.3758/s13428-024-02587-x>.
- [106] A. van den Oord et al., “Wavenet: A generative model for raw audio,” *arXiv preprint arXiv:1609.03499*, 2016. [Online]. Available: <https://arxiv.org/abs/1609.03499>.
- [107] *Openbci documentation: Cyton biosensing board*, <https://docs.openbci.com/>, Accessed: 2025-12-04, 2025.
- [108] *Openvibe: Open-source software platform for brain-computer interfaces*, <https://openvibe.inria.fr/>, Accessed: 2025-12-04, 2025.
- [109] A. Paszke et al., “Pytorch: An imperative style, high-performance deep learning library,” in *Advances in Neural Information Processing Systems*, 2019. [Online]. Available: <https://papers.neurips.cc/paper/2019/hash/bdbca288fee7f92f2bfa9f7012727740-Abstract.html>.
- [110] P. Pathirana, S. Senarath, D. Meedeniya, and S. Jayarathna, “Eye gaze estimation: A survey on deep learning-based approaches,” *Expert Systems with Applications*, vol. 199, p. 116 894, 2022.
- [111] J. W. Peirce et al., “Psychopy2: Experiments in behavior made easy,” *Behavior Research Methods*, vol. 51, no. 1, pp. 195–203, 2019. DOI: 10.3758/s13428-018-01193-y.

- [112] P. Peksa, E. Dudek-Dyduch, M. Brzozowski, et al., “State-of-the-art on brain-computer interface technology,” *Sensors*, vol. 23, no. 10, p. 6001, 2023.
- [113] C. R. Pernet et al., “Best practices in data analysis and sharing in neuroimaging using MEEG,” *Nature Neuroscience*, vol. 23, no. 12, pp. 1473–1483, 2020. DOI: 10.1038/s41593-020-00709-0.
- [114] C. R. Pernet et al., “Eeg-bids, an extension to the brain imaging data structure for electroencephalography,” *Scientific Data*, vol. 6, no. 1, p. 103, 2019. DOI: 10.1038/s41597-019-0104-8.
- [115] G. Pfurtscheller and F. H. Lopes da Silva, “Event-related synchronization (ers): An electrophysiological correlate of cortical areas at rest,” *Electroencephalography and clinical neurophysiology*, vol. 98, no. 4, pp. 281–293, 1999.
- [116] A. Picot, “Driver drowsiness detection using both physiological and video information,” PhD thesis, GIPSA-lab, Grenoble INP, Grenoble, France, 2009.
- [117] A. Picot, A. Caplier, and S. Charbonnier, “Using retina modelling to characterise blinking: Comparison between eog and video analysis,” *IRBM*, vol. 32, no. 1, pp. 18–25, 2011.
- [118] A. Picot, S. Charbonnier, and A. Caplier, “On-line detection of drowsiness using brain and visual information,” *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 42, no. 3, pp. 764–775, 2012.
- [119] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, “ICLabel: An automated electroencephalographic independent component classifier, dataset, and website,” *NeuroImage*, vol. 198, pp. 181–197, 2019. DOI: 10.1016/j.neuroimage.2019.05.026.
- [120] J. Polich, “Updating p300: An integrative theory of p3a and p3b,” *Clinical neurophysiology*, vol. 118, no. 10, pp. 2128–2148, 2007.

- [121] S. Poorna, V. Arsha, P. Aparna, P. Gopal, and G. Nair, “Drowsiness detection for safe driving using pca on eeg signals,” in *Progress in Computing, Analytics and Networking*, ser. Advances in Intelligent Systems and Computing, vol. 710, Springer, 2018, pp. 305–313.
- [122] T. Radüntz, J. Scouten, O. Hochmuth, and B. Meffert, “Signal quality evaluation of emerging eeg devices,” *Frontiers in Physiology*, vol. 8, 2017. DOI: 10.3389/fphys.2017.00141.
- [123] A. Ramos-Murguialday et al., “Brain-machine interface in chronic stroke rehabilitation: A controlled study,” *Annals of Neurology*, vol. 74, no. 1, pp. 100–108, 2013. DOI: 10.1002/ana.23879.
- [124] M. Rashid, N. Sulaiman, A. P. P. Abdul Majeed, et al., “Current status, challenges and possible solutions of EEG-based brain–computer interface: A comprehensive review,” *Frontiers in Neurorobotics*, vol. 14, p. 25, 2020.
- [125] S. Raspopovic et al., “Restoring natural sensory feedback in real-time bidirectional hand prostheses,” *Science Translational Medicine*, vol. 6, no. 222, 222ra19–222ra19, 2014. DOI: 10.1126/scitranslmed.3006820.
- [126] P. Redmond, M. Ahsan, and T. E. Ward, “Inferring EOG signal from video,” in *MEG United Kingdom and Ireland Conference 2023*, Abstract, 2023, pp. 29–30.
- [127] P. Redmond, M. A. Awais, and T. E. Ward, “Automatic tagging of BCI artefacts using computer vision,” in *Proceedings of the 10th International Brain-Computer Interface Meeting*, 2023.
- [128] P. Redmond, M. A. Awais, and T. E. Ward, “Ethical and GDPR-compliant sharing of multimodal stress data: Contributions to environmental and health research,” in *Joint Conference of ISEH, ICEPH and ISEG 2024*, 2024.
- [129] P. Redmond, A. Fleury, and T. Ward, “An open source multi-modal data-acquisition platform for experimental investigation of blended control of scale vehicles,” in *2022 IEEE International Conference on Metrology for Extended*

- Reality, Artificial Intelligence and Neural Engineering (MetroXRaine)*, 2022, pp. 1–6. DOI: 10.1109/metroxraine54828.2022.9967662.
- [130] Z. Ren, X. Zhang, L. Liu, J. Zhang, and D. Sun, “Appearance-based gaze estimation with feature fusion of multi-level information elements,” *Journal of Computational Design and Engineering*, vol. 10, no. 3, pp. 1080–1095, 2023. DOI: 10.1093/jcde/qwad028.
- [131] S. Romero, M. A. Mañanas, and M. J. Barbanoj, “A comparative study of automatic techniques for ocular artifact reduction in spontaneous eeg signals based on clinical target variables: A simulation case,” *Computers in Biology and Medicine*, vol. 41, no. 3, pp. 183–189, 2011. DOI: 10.1016/j.combiomed.2011.01.002.
- [132] S. Romero, M. A. Mananas, and M. J. Barbanoj, “A comparative study of automatic techniques for ocular artifact reduction in spontaneous eeg signals based on clinical target variables: A simulation case,” *Computers in Biology and Medicine*, vol. 68, pp. 147–158, 2016. DOI: 10.1016/j.combiomed.2015.11.016.
- [133] A. Saibene et al., “EEG-based BCIs on motor imagery paradigm using wearable technologies: A systematic review,” *Sensors*, vol. 23, no. 5, p. 2798, 2023. DOI: 10.3390/s23052798.
- [134] W. Schneider, A. Eschman, and A. Zuccolotto, *E-prime user’s guide*, Psychology Software Tools, Inc., Pittsburgh, PA, USA, 2002.
- [135] M. Seeck et al., “The standardized EEG electrode array of the IFCN,” *Clinical Neurophysiology*, vol. 128, no. 10, pp. 2070–2077, 2017. DOI: 10.1016/j.clinph.2017.06.254.
- [136] Software Heritage, *Software heritage archive*, <https://www.softwareheritage.org/>, Accessed: 2025-12-04, 2025.

- [137] B. Somers, A. Bertrand, T. Francart, and S. Van Huffel, “A generic eeg artifact removal algorithm based on the multi-channel wiener filter,” *Journal of Neural Engineering*, vol. 16, no. 2, 2019. DOI: 10.1088/1741-2552/aaf37a.
- [138] T. Soukupová and J. Čech, “Real-time eye blink detection using facial landmarks,” in *21st Computer Vision Winter Workshop (CVWW 2016)*, 2016. [Online]. Available: <https://vision.fe.uni-lj.si/cvww2016/proceedings/papers/05.pdf>.
- [139] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014.
- [140] N. Stein et al., “A comparison of eye tracking latencies among several commercial head-mounted displays,” *i-Perception*, vol. 12, no. 1, pp. 1–16, 2021. DOI: 10.1177/2041669520983338. [Online]. Available: <https://doi.org/10.1177/2041669520983338>.
- [141] R. Szeliski, *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
- [142] Tobii Pro, *Tobii pro spectrum user’s manual*, Accessed 2025-12-04, 2022. [Online]. Available: <https://help.tobii.com/hc/en-us/articles/360010284140-Tobii-Pro-Spectrum-User-s-manual>.
- [143] S. Umeyama, “Least-squares estimation of transformation parameters between two point patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991. DOI: 10.1109/34.88573.
- [144] J. A. Urigüen and B. Garcia-Zapirain, “Eeg artifact removal—state-of-the-art and guidelines,” *Journal of Neural Engineering*, vol. 12, no. 3, 2015. DOI: 10.1088/1741-2560/12/3/031001.
- [145] J. Valentin et al., “Mediapipe: A framework for building multimodal (video, audio, accelerometer) applied machine learning pipelines,” 2019.

- [146] J. Valentin et al., “Mediapipe: A framework for building perception pipelines,” *arXiv preprint arXiv:2006.12195*, 2020.
- [147] L.-M. Vortmann, S. Rinderknecht, and F. Putze, “Multimodal EEG and eye-tracking feature fusion for attention classification,” *Frontiers in Computer Science*, vol. 4, p. 780580, 2022. DOI: 10.3389/fcomp.2022.780580.
- [148] A. Widmann, E. Schröger, and B. Maess, “Digital filter design for electrophysiological data: A practical approach,” *Journal of Neuroscience Methods*, vol. 250, pp. 34–46, 2015. DOI: 10.1016/j.jneumeth.2014.08.002.
- [149] M. D. Wilkinson et al., “The FAIR guiding principles for scientific data management and stewardship,” *Scientific Data*, vol. 3, p. 160018, 2016. DOI: 10.1038/sdata.2016.18.
- [150] C. Williams et al., “A validation of Emotiv EPOC Flex saline for EEG and ERP research,” *PeerJ*, vol. 8, e9713, 2020. DOI: 10.7717/peerj.9713.
- [151] I. Winkler, S. Haufe, and M. Tangermann, “Automatic classification of artifactual ica-components for artifact removal in eeg signals,” *Behavioral and Brain Functions*, vol. 7, p. 30, 2011. DOI: 10.1186/1744-9081-7-30.
- [152] J. R. Wolpaw et al., “Brain-computer interface technology: A review of the first international meeting,” *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 2, pp. 164–173, 2002. DOI: 10.1109/TRE.2000.847807.
- [153] Y. Wu, J. He, M. Cai, and D. Xie, “High-speed eye tracking based on alternating eye-subregion stabilisation and dual-ring infrared lighting,” *Applied Optics*, vol. 63, pp. 10009–10017, 2024. DOI: 10.1364/AO.527189.
- [154] Y. Yahia Lahssene, M. Keche, and A. Ouamri, “Bio-inspired eeg generation from video camera: Application to driver’s awareness monitoring,” in *Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT’18)*, Vol. 1, ser. Smart Innovation, Systems and Technologies, vol. 146, Cham: Springer, 2020, pp. 400–409. DOI: 10.1007/978-3-030-21005-2_39.

- [155] P. Zeng, L. Fan, Y. Luo, H. Shen, and D. Hu, “Task-oriented EEG denoising generative adversarial network for enhancing SSVEP-BCI performance,” *Journal of Neural Engineering*, vol. 21, no. 6, p. 066 011, 2024. DOI: 10.1088/1741-2552/ad7d6a.
- [156] Zenodo, *Zenodo research repository*, <https://zenodo.org/>, Accessed: 2025-12-04, 2025.
- [157] H. Zhang, M. Zhao, C. Wei, D. Mantini, Z. Li, and Q. Liu, “EEGdenoiseNet: A benchmark dataset for deep learning solutions of EEG denoising,” *Journal of Neural Engineering*, vol. 18, no. 5, 2021.
- [158] S. Zhang et al., “A benchmark dataset for RSVP-based brain–computer interface,” *Frontiers in Neuroscience*, vol. 14, p. 568 000, 2020. DOI: 10.3389/fnins.2020.568000.
- [159] X. Zhang, X. Zhao, L. Bao-Liang, et al., “Eog-based drowsiness detection using convolutional neural networks,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 2014.
- [160] X. Zhang, Y. Sugano, M. Fritz, and A. Bulling, “Appearance-based gaze estimation in whole or face images,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 1, pp. 116–128, 2015.
- [161] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.
- [162] B. Zhuang, Q.-H. Tran, P. Ji, L.-F. Cheong, and M. Chandraker, “Learning structure-and-motion-aware rolling shutter correction,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4551–4560. DOI: 10.1109/CVPR.2019.00468.