University of Modena and Reggio Emilia

Department of Engineering "Enzo Ferrari"

Master degree
Electronics Engineering

Design and implementation of a Low-cost Wearable Eye Tracker: A Sensor Fusion Approach Combining Electro-OculoGraphy and Pupil Tracking

Tutor:
Prof. Agostino Gibaldi

Co-Tutor:
Prof. Giovanni Gibertoni

Master Thesis of:
Athanas Kafia
Matricola 185596

Academic Year 2024/2025

Abstract

Eye-tracking technology has seen rapid development in recent years, becoming integral to various applications such as human-computer interfaces (HCI), gaming, augmented and virtual reality (AR/VR), and automotive driver alertness monitoring. The standard approach to gaze estimation relies on video-based systems. While video-oculography guarantees high accuracy, it suffers from high computational costs and elevated latency, limiting its efficiency in real-time use. Conversely, Electro-OculoGraphy (EOG) offers minimal latency and low computational demand. However, consumer-grade, EOG is fundamentally limited by low absolute accuracy, non-stationary baseline drift, and signal non-linearities at large eccentricities.

This thesis presents the design and implementation of a low-cost, hybrid eye-tracking system that mitigates these hardware limitations through customized mechanical design and multimodal sensor fusion. Primary EOG acquisition is achieved using a modified AD8232 analog front-end. To reduce operator-dependent placement variability, the system utilizes electrodes integrated with a custom-molded, wearable thermoplastic mask.

To resolve the stochastic drift inherent to AC-coupled amplification, the architecture introduces a hybrid sensor-fusion approach combining EOG with video-based pupil tracking. An OV2660 camera is integrated into the wearable interface to provide absolute spatial referencing. Utilizing a saccade-triggered state machine executed via Psychtoolbox in MATLAB, the system leverages the high temporal resolution of the EOG to monitor rapid saccadic kinematics by continuously measuring the corneo-retinal potential. Simultaneously, the optical data captures the pupil position and acts as a ground-truth coordinate. This absolute spatial referencing automatically resets the EOG baseline and nullifies accumulated integration errors without requiring explicit user recalibration.

By fusing the computational efficiency and low latency of EOG with the absolute spatial accuracy of video-oculography, this research demonstrates a robust, low-cost, and stable interface suitable for assistive Human-Computer Interaction.

Riassunto

La tecnologia di eye-tracking ha registrato un rapido sviluppo negli ultimi anni, affermandosi come componente chiave in numerose applicazioni, tra cui le interfacce human-computer (HCI), il gaming, la realtà aumentata e virtuale (AR/VR) e il monitoraggio dello stato di attenzione del conducente in ambito automotive. L'approccio convenzionale alla stima della direzione dello sguardo si basa su sistemi video. Sebbene la video-oculografia garantisca un'elevata accuratezza, essa è caratterizzata da elevati costi computazionali e da una latenza significativa, che ne limitano l'impiego in applicazioni real-time. Al contrario, l'Elettro-Oculografia (EOG) offre una latenza trascurabile e un ridotto carico computazionale; tuttavia, nei dispositivi di fascia consumer, presenta limitazioni intrinseche quali bassa accuratezza assoluta, deriva del segnale e non linearità a grandi eccentricità.

In questo lavoro si presenta la progettazione e l'implementazione di un sistema di eye-tracking ibrido a basso costo, finalizzato a mitigare tali limitazioni hardware mediante una progettazione meccanica e tecniche di fusione multimodale dei sensori. L'acquisizione primaria del segnale EOG è realizzata mediante un front-end analogico AD8232 opportunamente modificato. Al fine di ridurre la variabilità nel posizionamento degli elettrodi, il sistema impiega elettrodi integrati in una maschera realizzata ad hoc.

Per compensare la deriva stocastica intrinseca all'EOG, l'architettura proposta introduce un approccio di sensor fusion ibrido che integra l'EOG con il tracciamento della pupilla basato su video. Una camera OV2660 è incorporata nell'interfaccia indossabile al fine di fornire un riferimento spaziale assoluto. Mediante una macchina a stati attivata da saccadi, implementata tramite Psychtoolbox in MATLAB, il sistema sfrutta l'elevata risoluzione temporale dell'EOG per il monitoraggio della cinematica saccadica, attraverso la misura continua del potenziale corneo-retinico. Parallelamente, i dati ottici consentono di stimare la posizione della pupilla, fornendo una coordinata di riferimento (ground truth). Tale riferimento spaziale assoluto consente il ripristino automatico della linea di base dell'EOG e la compensazione degli errori di integrazione accumulati, senza necessità di ricalibrazione esplicita da parte dell'utente.

Combinando l'efficienza computazionale e la bassa latenza dell'EOG con l'accuratezza spaziale assoluta della video-oculografia, il sistema proposto dimostra la realizzazione di un'interfaccia robusta, stabile e a basso costo, idonea ad applicazioni di interazione human-computer assistiva.

Contents

1 Background and Objectives **5**

1.1 The Challenge of Robust Gaze Estimation . . . 5
1.2 The Cost-Performance Gap . . . 5
1.2.1 Video-Based Tracking . . . 5
1.2.2 Electro-Oculography (EOG) . . . 6
1.3 Motivation . . . 6
1.4 Objectives . . . 6

2 Materials and Methods **9**

2.1 The Electro-Oculogram (EOG) . . . 9
2.1.1 Oculomotor Dynamics and Anatomy . . . 9
2.1.2 Physiological Origin: The Corneo-Retinal Potential (CRP) . . . 10
2.1.3 Mathematical Model of EOG . . . 10
2.2 Acquisition Hardware: AD8232 . . . 11
2.2.1 Device Architecture . . . 11
2.2.2 Standard ECG Configuration vs. EOG Requirements . . . 11
2.2.3 Hardware Adaptation for EOG . . . 12
2.3 The Electrode Interface . . . 14
2.3.1 Electrode Placement: Bipolar Derivation . . . 14
2.4 Microcontroller and Vision Processing: XIAO ESP32-S3 Sense . . . 14
2.4.1 Power Supply and Signal Isolation . . . 15
2.4.2 System Schematic and Wiring Architecture . . . 15
2.4.3 Deterministic Sampling Architecture . . . 16
2.4.4 Optical Integration: OV2660 Camera Module . . . 16
2.4.5 Distributed Processing and Communication Architecture . . . 16
2.5 Wearable Electrode Interface . . . 17
2.5.1 Mechanical Structure . . . 17
2.5.2 Sensor Integration and Mechanical Decoupling . . . 18

3 System Implementation **19**

3.1 State of the Art: Strategies for EOG Signal Processing . . . 19
3.1.1 Physical Modelling Approaches . . . 19
3.1.2 Sensor Fusion and Continuous Calibration . . . 19
3.1.3 Non-Linear Signal Estimation . . . 20
3.1.4 Justification for Proposed Methodology . . . 20
3.2 Data Acquisition and Pre-Processing . . . 20
3.2.1 The Experimental Protocol . . . 20

3.2.2 Optical Tracking Algorithm . . . 21
3.2.3 Signal Conditioning and Saccade Extraction . . . 22
3.3 Mathematical Modelling and Calibration . . . 23
3.3.1 EOG Calibration: Bilateral Polynomial Model . . . 23
3.3.2 The Numerical Solver and cost function . . . 24
3.3.3 Robust Initialization Strategy . . . 25
3.3.4 Optical Calibration: Absolute Linear Model . . . 25
3.4 Multimodal Drift Correction: Stateless Optical Anchoring . . . 26
3.4.1 The Integration Trap . . . 26
3.4.2 The Sensor Fusion Algorithm . . . 26
3.5 Real-Time Multimodal System Implementation . . . 27
3.5.1 Online vs. Offline Processing . . . 27
3.5.2 The Real-Time Multimodal Loop . . . 27

4 Results **29**
4.1 Experimental Validation Framework . . . 29
4.1.1 Physical Setup and Environmental Control . . . 29
4.1.2 Lighting Conditions and Environmental Robustness . . . 30
4.1.3 Visual Stimulation and Protocol Design . . . 30
4.1.4 Performance Metrics . . . 30
4.2 Signal Characterization and Feasibility Analysis . . . 31
4.2.1 Drift Dynamics: The "Random Walk" Behaviour . . . 31
4.2.2 The Multimodal Paradigm Shift: Optical Anchoring . . . 32
4.2.3 Vertical Hysteresis and Eyelid Artifacts . . . 32
4.2.4 Channel Independence (Crosstalk Analysis) . . . 33
4.3 Phase I: Single-Subject System Validation . . . 33
4.3.1 Horizontal Gaze Estimation . . . 33
4.3.1.1 Calibration Performance . . . 33
4.3.1.2 Sensor Fusion and Estimation Accuracy . . . 34
4.3.1.3 Temporal Stability . . . 35
4.3.2 Vertical Gaze Estimation . . . 35
4.3.2.1 Calibration Performance . . . 35
4.3.2.2 Sensor Fusion and Estimation Accuracy . . . 35
4.3.2.3 Temporal Stability . . . 36
4.3.3 2D Oblique Gaze Estimation . . . 37
4.3.3.1 Calibration Performance in 2D Space . . . 37
4.3.3.2 Component-Level Sensor Fusion Accuracy . . . 37
4.3.3.3 2D Spatial Accuracy and Directional Bias . . . 38
4.3.3.4 Temporal Stability and Sensor Fusion Efficacy . . . 38
4.4 Phase II: Multi-Subject Validation . . . 39
4.4.1 Spatial Error Translation and HCI Implications . . . 41

5 Conclusions **43**
5.1 Summary of Contributions . . . 43
5.1.1 Characterization and Isolation of the AD8232 . . . 43
5.1.2 The Bilateral Polynomial Model . . . 43
5.1.3 Multimodal Sensor Fusion for Drift Correction . . . 44

5.2 Discussion of Results . . . 44
5.2.1 Single-Subject System Validation . . . 44
5.2.2 Multi-Subject Generalization and Hardware Limits . . . 45
5.2.3 Implications for Human-Computer Interaction (HCI) . . . 46
5.3 Limitations of the System . . . 46
5.3.1 Hardware-Specific Calibration and Tuning . . . 46
5.3.2 Electrode Mechanics and Mask Rigidity . . . 46
5.3.3 Bandwidth and Antenna Constraints . . . 46
5.4 Future Works . . . 47
5.4.1 Custom EOG Analog Front-End . . . 47
5.4.2 Flexible Wearable Architecture . . . 47
5.4.3 Advanced Optical Hardware and Tracking Algorithms . . . 47
5.5 Final Conclusion . . . 47

Bibliography **50**

List of Tables

4.1 Summary of Single-Subject Estimation Accuracy (RMSE) 40

List of Figures

2.1 Muscles of the human eye . . . 9
2.2 Corneo-Retinal Potential . . . 10
2.3 AD8232 Board . . . 12
2.4 EOG Placement . . . 15
2.5 Circuit Schematics . . . 16
2.6 Photo of the mask during its initial tests . . . 17

4.1 Drift trend . . . 31
4.2 Hysteresis of the vertical channel . . . 32
4.3 Crosstalk analysis between horizontal and vertical channels . . . 33
4.4 Horizontal Calibration . . . 34
4.5 Horizontal Estimation Performance . . . 34
4.6 Cyclic RMSE Performance of Horizontal Movements . . . 35
4.7 Vertical Calibration . . . 36
4.8 Vertical Estimation Performance . . . 36
4.9 Cyclic RMSE Performance of Vertical Movements . . . 37
4.10 The Calibration of the two Diagonal Channels . . . 38
4.11 Oblique movements divided into two orthogonal components . . . 38
4.12 Gaze Accuracy Map and Direction Error Bias . . . 39
4.13 Cyclic RMSE Performance of the Oblique Movements . . . 39
4.14 Multi-Subject RMSE . . . 40

Introduction

The accurate measurement of human eye movements is a critical challenge in biomedical engineering, with profound implications for neurological assessment, assistive technologies, and future applications in Human-Computer Interaction (HCI). While clinical-grade video trackers and DC-coupled Electro-Oculography (EOG) amplifiers offer high precision, they impose significant cost and proprietary hardware barriers, limiting their availability outside specialized research environments. Conversely, low-cost consumer optical solutions are often highly susceptible to environmental noise, lighting variations, and head pose changes, severely limiting their reliability. Electro-Oculography (EOG) provides a robust, computationally lightweight alternative by measuring the standing corneo-retinal potential to estimate gaze coordinates. However, capturing the quasi-static biopotentials necessary for absolute position tracking typically requires expensive, specialized instrumentation amplifiers capable of DC coupling. Low-cost biopotential sensors, such as those designed for electrocardiography (ECG), are inherently AC-coupled to reject DC electrode offsets. This architecture introduces a severe time-dependent signal decay, or baseline drift. Theoretically, this temporal instability renders such accessible hardware unsuitable for maintaining the absolute 2D spatial coordinates required for continuous gaze tracking.This thesis focuses on overcoming the fundamental limitations of low-cost, AC-coupled biopotential acquisition through algorithmic compensation and multimodal sensor fusion. Utilizing the AD8232—an analog front-end originally designed for cardiac monitoring — this research proposes a shift in design philosophy from hardware complexity to intelligent signal processing. To counteract the stochastic integration drift of the AC-coupled sensor and the geometric non-linearities of the eye's dipole, this work integrates the high-speed oculomotor dynamics of EOG with absolute optical positioning derived from a standard, low-cost camera. The primary scope of this research is the engineering and validation of this hybrid tracking architecture. The proposed system is designed to fulfil three core technical objectives:

1. **Hardware Adaptation:** The modification of the AD8232's active filter topology to optimize the bandwidth specifically for EOG acquisition.
2. **Spatial Linearization:** The compensation of signal saturation at large eccentricities ($\pm 30°$) and orbital asymmetries via a 2nd-Order Bilateral Polynomial Model.
3. **Multimodal Sensor Fusion:** The development of a hybrid integration pipeline that utilizes discrete optical pupil coordinates as a ground-truth anchor to dynamically reset the EOG baseline, effectively neutralizing AC-coupled drift.

By fusing the high temporal resolution of low-cost EOG with the absolute spatial stability of optical tracking, this research aims to demonstrate that high-fidelity, drift-free 2D gaze estimation can be achieved using highly accessible, consumer-grade hardware.

Thesis outline

Chapter 1 — Background and Objectives

This chapter outlines the limitations of current eye-tracking technologies to establish the motivation and specific objectives for developing a low-cost, multimodal gaze-tracking interface.

- Section 1.1: The Evolution of Human-Computer Interaction — This section traces the shift toward natural user interfaces, highlighting eye gaze as a highly efficient, intent-predictive input modality essential for assistive technologies and spatial computing, despite current accessibility and performance barriers.
- Section 1.2: The Cost-Performance Gap — This section analyses the trade-off between accuracy and accessibility in existing eye-tracking technologies, highlighting the limitations that hinder large-scale adoption.
- Section 1.3: Motivation — Driven by the goal of accessible, high-fidelity tracking, this research aims to overcome the signal drift inherent in low-cost, adapted biopotential sensors by dynamically fusing their high-speed electrical data with absolute spatial anchors from an integrated optical camera.
- Section 1.4: Objectives — The primary goal is to build and validate a real-time, multimodal eye-tracking interface through specific milestones, including hardware modification, optical pipeline synchronization, non-linear mathematical calibration, algorithmic drift compensation, and rigorous performance testing.

Chapter 2 — Materials and Methods

This chapter details the hardware, physiological principles, and system architecture used to construct a multimodal gaze-tracking interface, specifically detailing the adaptation of biopotential sensors and their integration with a computer vision pipeline.

- Section 2.1: The Electro-Oculogram (EOG) — This section explains the physiological basis of EOG measuring the eye's standing electrical dipole to track saccades and fixations—and establishes a mathematical model that necessitates a non-linear calibration approach for accurate gaze tracking at larger angles.
- Section 2.2: Acquisition Hardware: AD8232 — This section describes how the AD8232 biopotential amplifier, originally designed for dynamic ECG signals, was physically modified by tuning its high-pass and low-pass filters to capture the quasi-DC signals required for tracking static eye positions.
- Section 2.3: The Electrode Interface — This section outlines the bipolar placement of electrodes around the eyes to measure horizontal and vertical EOG signals, emphasizing the transition from disposable patches to reusable gold-cup electrodes for consistent sensor geometry.
- Section 2.4: Microcontroller and Vision Processing: XIAO ESP32-S3 Sense — This section details the distributed processing architecture, explaining how a central microcontroller manages deterministic, high-speed EOG sampling alongside wireless optical streaming to a host PC, where a Python pipeline extracts pupil coordinates for real-time sensor fusion.
- Section 2.5: Wearable Electrode Interface — This section describes the design and construction of a custom, rigid thermoplastic head-mount that securely houses the EOG electrodes and wiring, ensuring a stable, repeatable, and comfortable sensor-to-skin connection across experimental sessions.

Chapter 3 — System Implementation

This chapter details the software architecture, signal processing pipelines, and mathematical models developed to acquire, calibrate, and seamlessly fuse electrical and optical tracking data into a real-time, drift-resistant gaze estimation system.

- Section 3.1: State of the Art: Strategies for EOG Signal Processing — By reviewing current physical modeling and sensor fusion paradigms, this section justifies the selection of a non-linear "Direct Gaze Estimation" method that relies on differential saccadic voltage steps to overcome AC-coupled signal drift without requiring complex anatomical calibration or auxiliary inertial hardware.
- Section 3.2: Data Acquisition and Pre-Processing — This section outlines the synchronized experimental protocol, detailing both the robust computer vision pipeline used to extract absolute pupil coordinates and the EOG signal conditioning algorithms required to isolate pure saccadic voltage changes while filtering out noise and blink artifacts.
- Section 3.3: Mathematical Modelling and Calibration — To accurately translate raw sensor data into physical gaze coordinates, this section formulates a weighted, bilateral 2nd-order polynomial model optimized via the Levenberg-Marquardt algorithm to handle EOG non-linearities, alongside a simpler absolute linear model for the optical data.
- Section 3.4: Multimodal Drift Correction Stateless Optical Anchoring — This section introduces the core sensor fusion algorithm, which resolves the cumulative baseline drift of the EOG by dynamically calculating and subtracting integration errors using the optical camera's absolute spatial data as stateless anchors.
- Section 3.5: Real-Time Multimodal System Implementation — The final section describes the translation of these offline models into a low-latency, real-time MATLAB interface that manages asynchronous dual-stream data parsing, analytic mathematical inversion, and dynamic complementary filtering to provide continuous, robust gaze tracking.

Chapter 4 — Results

Presents the experimental setup, signal characterization, and subsequent validation results of the multimodal gaze-tracking system, proving through single and multi-subject testing that fusing optical anchors with EOG signals successfully neutralizes baseline drift and achieves sufficient accuracy for target-based HCI.

- Section 4.1: Experimental Validation Framework — This section details the controlled, two-phase experimental setup—including physical geometry, environmental robustness, visual stimulation protocols, and performance metrics (RMSE)—used to rigorously evaluate the multimodal system's accuracy across both single and multiple subjects.
- Section 4.2: Signal Characterization and Feasibility Analysis — This section analyses the raw EOG signal, proving that its unpredictable baseline drift and mechanical vertical hysteresis cannot be corrected with simple maths, thereby justifying the necessity of integrating an absolute optical anchor and treating the orthogonal horizontal and vertical channels independently.
- Section 4.3: Single-Subject System Validation — This section details the single-subject validation of the multimodal system, demonstrating how optical sensor fusion successfully calibrates anatomical non-linearities and eliminates temporal signal drift across horizontal, vertical, and complex diagonal eye movements.

- Section 4.4: Multi-Subject Validation — This section evaluates the multimodal system's robustness across multiple users to assess inter-subject variability and translates the observed spatial tracking variance into practical design constraints for gaze-driven graphical interfaces.

Chapter 5 — Conclusions

This chapter summarizes the development and calibration of the low-cost multimodal gaze-tracking system, evaluating its practical accuracy for assistive interfaces, and identifying current hardware limitations to propose specific avenues for future engineering improvements.

- Section 5.1: Summary of Contributions — Highlights the project's core achievements, specifically the physical optimization of the AD8232 amplifier for isolated EOG extraction, the formulation of a bilateral polynomial model to correct anatomical and geometric non-linearities, and the successful implementation of a UDP-based optical sensor fusion pipeline to dynamically neutralize AC-coupled signal drift.
- Section 5.2: Discussion of Results — Interpret both individual and group validation data to demonstrate the efficacy of the sensor fusion and calibration models, identify inherent hardware limitations, and establish practical design constraints for gaze-driven user interfaces.
- Section 5.3: Limitations of the System — Outlines the engineering constraints of relying on off-the-shelf components, noting that the camera requires manual tuning for lighting changes, the rigid thermoplastic mask introduces mechanical noise at the electrode interface, and Wi-Fi bandwidth bottlenecks restrict the optical tracking frame rate.
- Section 5.4: Future Works — To overcome current operational bottlenecks, this section proposes next-generation improvements including the design of a custom DC-coupled analog front-end, the transition to a flexible textile-based mask for better ergonomic compliance, and the integration of hardware-cropped imaging and machine-learning algorithms to maximize optical robustness.
- Section 5.5: Final Conclusion — Concludes the research, affirming that multimodal sensor fusion successfully transforms low-cost, unstable EOG hardware into a highly accurate and accessible 2D gaze-tracking interface.

Chapter 1

Background and Objectives

1.1 The Challenge of Robust Gaze Estimation

The accurate measurement of human eye movements is a complex biomedical engineering challenge. While ocular tracking has historically enabled advances in cognitive science, neurology, and diagnostics [1], its integration into robust, real-time wearable systems remains constrained by hardware limitations.

Eye gaze provides a high-bandwidth, low-latency data stream characterized by rapid, ballistic movements (saccades) and static stabilization periods (fixations). Capturing these dynamics with high spatial and temporal fidelity traditionally requires specialized, intrusive, or highly expensive instrumentation. The fundamental engineering bottleneck lies in balancing system accessibility—encompassing cost, portability, and computational load—with tracking reliability in dynamic, unconstrained environments. Consequently, while the demand for continuous gaze estimation is growing across various technical domains, widespread adoption is severely constrained by the trade-off between hardware accessibility and signal integrity.

1.2 The Cost-Performance Gap

Existing eye-tracking technologies are categorized into two distinct classes, each presenting specific limitations for broad deployment. Research-grade systems provide high precision but are associated with high costs and proprietary restrictions [2]. Conversely, consumer-grade solutions utilize standard imaging hardware, resulting in susceptibility to environmental noise [2].

1.2.1 Video-Based Tracking

Video-Oculography (VOG) is the dominant tracking technology, utilizing infrared imaging of the pupil and corneal reflection (Purkinje images) to derive gaze coordinates [3].

- **Research-Grade Systems:** High-end trackers (e.g., Tobii Pro, EyeLink) achieve precision below 0.5° and sampling rates exceeding 300 Hz. However, these systems rely on specialized hardware and closed-source software, limiting their accessibility for clinical deployment or independent development [4].
- **Consumer-Grade Systems:** Entry-level trackers provide a lower-cost alternative but typically rely on visible-light sensors. These systems are sensitive to ambient lighting variations, head pose changes, and occlusion by eyewear, which degrades tracking accuracy and reliability.

1.2.2 Electro-Oculography (EOG)

Electro-Oculography (EOG) relies on the measurement of the Corneo-Retinal Potential (CRP), a standing electrical dipole between the cornea (positive) and retina (negative). Ocular rotation generates a change in the electrical field detected by peri-orbital electrodes [5]. EOG presents distinct operational characteristics compared to VOG:

- **Computational Efficiency:** The signal consists of 1D voltage streams, requiring significantly less processing power than real-time image analysis.
- **Environmental Robustness:** EOG is unaffected by lighting conditions, blink artifacts, or physical obstructions such as ptosis (drooping eyelids).
- **Linearity:** For angular displacements within $\pm 30°$, the relationship between voltage and gaze angle is approximately linear.

However, traditional EOG implementation is limited by instrumentation requirements. The acquisition of microvolt-level (μV) potentials necessitates high-precision bio-amplifiers with low noise floors, which are typically bulky and expensive [6]. This study addresses these limitations by validating a gaze tracking interface based on the AD8232, a low-cost integrated circuit designed for biopotential acquisition.

1.3 Motivation

The primary motivation of this research is to evaluate the feasibility of high-fidelity eye tracking using cost-effective, off-the-shelf hardware. The study investigates the hypothesis that the physical limitations of stand-alone consumer-grade sensors can be compensated for via multimodal signal processing and sensor fusion. The electrical component of this project utilizes the AD8232 [7], an integrated circuit designed for heart rate monitoring. While the chip employs high-pass filtering unsuitable for static fixation measurement, this work aims to modify the hardware bandwidth to enable oculomotor estimation. To overcome the non-stationary drift inherent to the AD8232's AC-coupled amplification, this study introduces a secondary optical modality: a low-cost, 25 Hz camera.

By fusing the high-speed temporal resolution of the EOG (300 Hz) with the absolute spatial coordinates of the computer vision pipeline, the system can dynamically correct electrical drift without requiring continuous user recalibration. The engineering challenge is therefore threefold:

1. **Hardware Adaptation:** Modification of the analog front-end to extend bandwidth into the near-DC range required for EOG.
2. **Optical Tracking Integration:** Development of a low-latency, Python-based computer vision pipeline to extract continuous pupil coordinates, utilizing UDP network bridging for cross-platform synchronization.
3. **Multimodal Drift Management:** Algorithmic compensation for EOG baseline drift by employing the optical data as absolute spatial anchors.

1.4 Objectives

The objective of this study is to design, implement, and validate a real-time, multimodal Electro-Oculography and Video-Oculography architecture. This goal is achieved through the following specific objectives:

- **Hardware Adaptation:** Analysis of the AD8232 architecture and implementation of circuit modifications to extend the lower cut-off frequency, preserving quasi-DC EOG components.
- **Optical Pipeline Development:** Implementation of a computer vision algorithm for robust pupil extraction and glint-distance filtering, synchronized with the electrical data acquisition via a customized master clock architecture.
- **Signal Modelling and Calibration:** Formulation of a mathematical model to map raw voltage potentials and pixel coordinates to visual angles, utilizing a 2nd-order bilateral polynomial calibration to compensate for non-linear saturation at large eccentricities ($\pm 25°$)
- **Drift Compensation:** Evolution of semantic 'Zero-Reset' protocols into a continuous Multimodal Sensor Fusion algorithm, leveraging absolute optical anchors to mitigate EOG integration error.
- **System Validation:** Quantification of the multimodal system's performance against traditional isolated tracking methods, evaluating spatial accuracy (RMSE), temporal stability across prolonged usage, and isotropic response.
- **Real-Time Implementation:** Development of a low-latency, fully synchronized data-acquisition and estimation framework to prove the viability of real-time sensor fusion on resource-constrained hardware.

Chapter 2

Materials and Methods

2.1 The Electro-Oculogram (EOG)

Electro-Oculography (EOG) relies on the measurement of the standing Corneo-Retinal Potential (CRP). Unlike dynamic muscle action potentials (EMG), the CRP provides a steady-state dipole vector that correlates with ocular orientation [8]. This study utilizes the EOG signal to estimate gaze coordinates relative to the head frame of reference.

2.1.1 Oculomotor Dynamics and Anatomy

The rotation of the eye is controlled by three pairs of extraocular muscles, which function as an antagonistic system. For the 2D gaze tracking implemented in this thesis, two specific muscle pairs are relevant (Fig. 2.1):

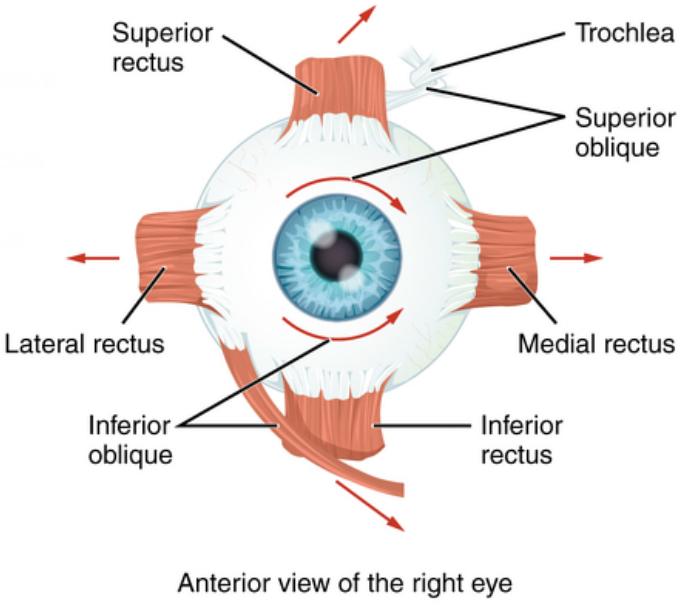


Figure 2.1: Muscles of the human eye

- **Horizontal Movement:** Controlled by the Lateral Rectus (abduction) and Medial Rectus (adduction).
- **Vertical Movement:** Primarily controlled by the Superior Rectus (elevation) and Inferior Rectus (depression).

From a signal processing perspective, eye movements are generally categorized into distinct classes. This research focuses on the two primary types relevant to Human-Computer Interaction:

- **Saccades:** Rapid, ballistic jumps of the eye that abruptly change the point of fixation. Saccadic velocities can exceed $400^{\circ}/s$, appearing in the EOG signal as near-vertical step functions [9].
- **Fixations:** Periods where the gaze is held relatively stable on a target. In the EOG signal, these manifest as DC plateaus. The stability of these plateaus is critical for the proposed system, as the drift correction algorithm (discussed in Chapter 3) relies on detecting these static intervals to reset the integrator error [9].

2.1.2 Physiological Origin: The Corneo-Retinal Potential (CRP)

The eye is modelled electrically as a spherical dipole (Fig. 2.2) with a persistent hyperpolarisation between the anterior cornea (positive) and the posterior retina (negative) [9]. The magnitude of this Corneo-Retinal Potential typically ranges from 6 to $10mV$, subject to diurnal variations and luminance adaptation [9]. As the eye rotates within the orbit, the projection of this dipole onto the facial skin surface changes, allowing for the derivation of angular displacement via surface electrodes.

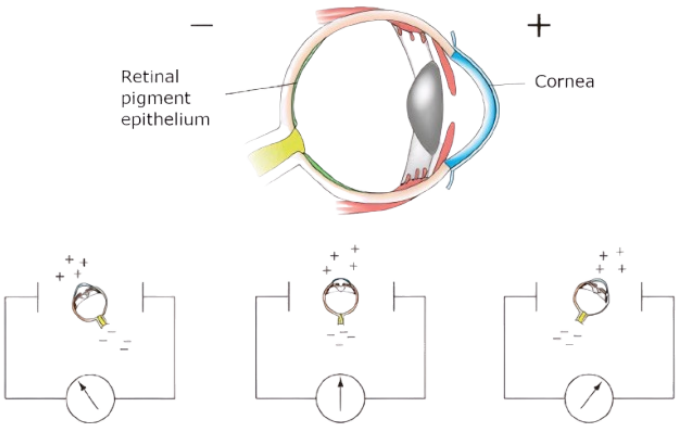


Figure 2.2: Corneo-Retinal Potential

2.1.3 Mathematical Model of EOG

The relationship between the measured EOG voltage (V) and the angular displacement of the eye (θ) is governed by the projection of the dipole vector onto the electrode plane. Theoretically, this relationship follows a sine law:

$$V = V_{\max} \cdot \sin(\theta) \tag{2.1}$$

where V_{max} is the maximum potential of the dipole. However, for gaze angles within the typical range of human interaction ($\theta < 30^{\circ}$), the small-angle approximation ($\sin(\theta) \approx \theta$) holds true, allowing the relationship to be modelled as linear:

$$V \approx k \cdot \theta \tag{2.2}$$

where k represents the sensitivity of the system (typically $15 - 20\mu V/^\circ$) [10].

Linearity vs. Saturation:

While the linear model is sufficient for small movements, the EOG signal exhibits non-linear behaviour at larger eccentricities ($> 30^\circ$). This "saturation" effect occurs due to the geometry of the volume conductor (the skull and skin) and muscle artifacts.

In this work, the hardware captures the full range of movement, but the linear assumption is rejected in favour of a 2nd-order polynomial calibration model (3.3.1). This decision was made to explicitly account for the non-linear saturation observed in the vertical channel, ensuring high accuracy even at the edges of the screen ($\pm 25^\circ$ horizontal, $\pm 16^\circ$ vertical).

2.2 Acquisition Hardware: AD8232

Signal conditioning is performed by the AD8232, an integrated analog front-end originally designed for biopotential acquisition. The component was selected for its integrated instrumentation amplifier architecture and high common-mode rejection ratio (CMRR) [7].

2.2.1 Device Architecture

The AD8232 is a single-lead heart rate monitor front end that acts as a comprehensive analog-to-analog signal chain. Unlike generic instrumentation amplifiers (e.g., INA128), the AD8232 integrates specialized filtering topology directly into the silicon to reject motion artifacts and electrode half-cell potentials [7]. Its internal architecture consists of three primary stages:

1. **Instrumentation Amplifier (In-Amp):** A custom preamp with a fixed gain of $100V/V$, designed to extract small differential bio-potential signals while rejecting common-mode interference (such as $50/60Hz$ mains hum) [7].
2. **High-Pass Filter:** Implemented via an integrator feedback loop that restores the DC level. This is critical for rejecting the significant DC offsets caused by electrode-skin impedance mismatches (which can be up to $\pm 300mV$) [11].
3. **Output Amplifier (Op-Amp):** A final gain stage that allows for adjustable gain and low-pass filtering to condition the signal for the microcontroller's ADC.

2.2.2 Standard ECG Configuration vs. EOG Requirements

In its standard configuration, the AD8232 is tuned for Heart Rate Monitoring. An ECG signal is characterized by rhythmic, rapid voltage spikes (the QRS complex). The diagnostic information is contained in the AC components ($0.5Hz - 40Hz$), and the absolute DC voltage is irrelevant [11]. Therefore, the standard application circuit implements a strict High-Pass Filter with a cut-off frequency (f_c) around $0.5Hz$. This configuration is intrinsically incompatible with Electro-Oculography for two reasons:

1. **Loss of Position Data:** EOG fixations are, by definition, static DC events. If a user looks to the right and holds their gaze, the EOG signal should produce a sustained step in voltage. A standard ECG filter would treat this step as "baseline wander" and aggressively filter it out, causing the signal to

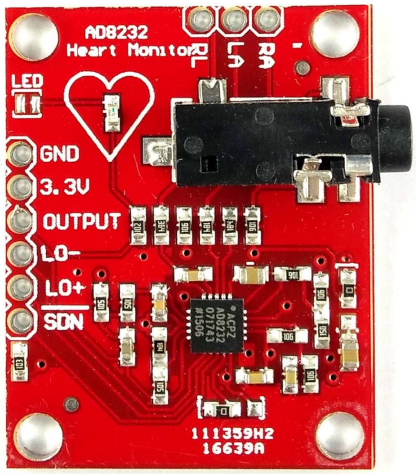


Figure 2.3: AD8232 Board

decay back to zero within seconds. This would result in a system that detects movement (velocity) but loses track of position.

2. **Signal Distortion**: The high-pass filter introduces phase non-linearity, distorting the shape of saccades and introducing overshoot artifacts that complicate the estimation of gaze amplitude.

2.2.3 Hardware Adaptation for EOG

To repurpose the AD8232 for gaze tracking, the frequency response of the acquisition circuit was physically modified to extend the bandwidth toward DC while aggressively filtering high-frequency noise. The AD8232 uses an external Resistor-Capacitor (RC) network to set the cut-off frequencies of its internal active filters.

High-Pass Filter Modification (Drift Reduction)

The most critical adaptation involved lowering the high-pass cut-off frequency (f_{cHP}) to allow the "step" function of a saccade to persist longer before decaying. The target cut-off was set to $0.05Hz$. The cut-off frequency for the two-pole high-pass filter in the AD8232 topology is governed by the relation:

$$f_{cHP} = \frac{10}{2\pi\sqrt{R_1 R_2 C_1 C_2}} \tag{2.3}$$

Assuming a symmetrical configuration where $R_1 = R_2 = R$ and $C_1 = C_2 = C$, the equation simplifies to:

$$f_{cHP} \approx \frac{10}{2\pi RC} \tag{2.4}$$

To achieve the sub-Hertz target, the standard components were replaced with high-impedance values:

- Resistors (R_1, R_2): $10\ M\Omega$
- Capacitors (C_1, C_2): $3.3\ \mu F$

Substituting these values yields the new cut-off frequency:

$$f_{cHP} = \frac{10}{2\pi\left(10 \times 10^{6}\right)\left(3.3 \times 10^{-6}\right)} \approx 0.048Hz \tag{2.5}$$

Additionally, the AD8232 requires a compensation resistor (R_3) at the output of the instrumentation amplifier to ensure stability for this second-order topology. This value is calculated as a ratio of the filter resistor:

$$R_3 = 0.14 \times R_1 = 0.14 \times 10M\Omega = 1.4M\Omega \tag{2.6}$$

Low-Pass Filter Modification (Noise Suppression)

To eliminate high-frequency electromyographic artifacts and mains hum ($50Hz$), the low-pass filter (f_{cLP}) was tuned to a target of approximately 10 Hz. This preserves the primary saccadic velocity profile while rejecting environmental noise. The cut-off frequency is determined by:

$$f_{cLP} = \frac{1}{2\pi\sqrt{R_4 R_5 C_3 C_4}} \tag{2.7}$$

Using $R = 180\ k\Omega$ and $C = 100\ nF$:

$$f_{cLP} = \frac{1}{2\pi\left(180 \times 10^3\right)\left(100 \times 10^{-9}\right)} \approx 8.84Hz \tag{2.8}$$

Thus, the components selected were:

- Resistors (R_4, R_5): $180\ k\Omega$
- Capacitors (C_3, C_4): $100\ nF$

Output Gain Configuration

The final stage of the AD8232 allows for additional signal amplification. The gain of the low-pass filter stage (G_{LP}) is set by the ratio of resistors R_6 and R_7:

$$G_{LP} = 1 + \frac{R_6}{R_7} \tag{2.9}$$

To maximize the dynamic range of the Arduino's ADC (0–5 V) without causing rail-saturation during large eye movements, a gain of 11 was targeted. Selecting $R_7 = 18k\Omega$, we calculate R_6:

$$10 = \frac{R_6}{18k\Omega} \Rightarrow R_6 = 180k\Omega \tag{2.10}$$

These modifications collectively tune the AD8232 to the specific spectral characteristics of EOG, balancing the need for signal stability (low HPF) with noise immunity (low LPF).

Feasibility of Direct DC Coupling

Ideally, an EOG acquisition system should be DC-coupled to prevent any signal decay during fixations. To test the feasibility of achieving this with the AD8232, an experiment was conducted where the internal high-pass filter was bypassed completely by short-circuiting the filter capacitors (C_{HP}).

To evaluate the feasibility of DC coupling, the internal high-pass filter was bypassed. It was hypothesized that removing the capacitive element would extend the bandwidth to $0Hz$. However, experimental results indicated immediate saturation to the supply rails due to electrode half-cell potentials. The interface between the Ag/AgCl electrodes and the skin generates a standing DC voltage (offset) that can range from ± 20 mV to over ± 300 mV depending on skin impedance and electrode quality.

Since the AD8232 has a fixed internal gain of $100V/V$ in its instrumentation stage, a typical offset of just $30mV$ results in an amplified output of:

$$30 \text{ mV} \times 100 = 3.0 \text{ V}$$

This massive DC offset consumes almost the entire dynamic range of the ADC, leaving no headroom for the actual EOG signal (which is typically only $0.5 - 2.0mV$).

In conclusion, this experiment confirmed that AC-coupling is not merely a design choice but a hardware necessity when using high-gain amplifiers without active DC-offset suppression (servo loops). Consequently, the design strategy shifted from "eliminating" the filter to "tuning" it—specifically, increasing the time constant τ to make the decay slow enough to be corrected in software (as detailed in Chapter 3).

2.3 The Electrode Interface

The interface between the biological source and the acquisition hardware is a critical determinant of signal integrity. During the development of the wearable architecture, both reusable gold-cup electrodes and disposable Ag/AgCl adhesive electrodes were evaluated.

While gold-cup electrodes successfully acquired a higher absolute voltage amplitude, their integration into a rigid, head-mounted mask introduced severe mechanical vulnerabilities. Facial contortions during large diagonal gaze shifts caused the rigid mask to act as a lever, physically shearing the gold-cup electrodes away from the epidermis. This stretching of the conductive paste dramatically spiked the contact impedance, resulting in big baseline drift and a signal that does not follow the movements of the eyes correctly in dynamic scenarios.

Consequently, the final integrated system exclusively utilizes disposable, self-adhesive Ag/AgCl electrodes. Although they produce a slightly lower raw amplitude, their direct adhesion to the skin provides exceptional mechanical isolation from the rigid chassis, absorbing sheer forces and yielding a highly stable, flat baseline essential for drift-correction algorithms.

2.3.1 Electrode Placement: Bipolar Derivation

The system utilizes a Bipolar Derivation configuration, measuring the potential difference between two active sites rather than referencing a single site to a distant ground. This common-mode rejection strategy is vital for eliminating power line interference. The montage consists of five electrodes:

1. **Horizontal Channel** ($E_{h+} - E_{h-}$): Placed at the outer canthi. Looking Right brings the positive cornea closer to the right electrode, creating a positive voltage peak.
2. **Vertical Channel** ($E_{v+} - E_{v-}$): Placed supra-orbitally and infra-orbitally aligned with the dominant eye. Looking Up generates a positive potential. [12]
3. **Reference Electrode (Ref):** Placed on a neutral bony prominence (the center of the forehead). This sets the common-mode voltage for the AD8232's "Right Leg Drive" (RLD) circuit.

To mitigate skin-electrode impedance mismatches, a strict preparation protocol involving epidermal abrasion and degreasing with alcohol was maintained prior to the application of the conductive paste.

2.4 Microcontroller and Vision Processing: XIAO ESP32-S3 Sense

Digitization, camera control, and data transmission are managed by a Seeed Studio XIAO ESP32S3 Sense development board. This dual-core, 32-bit microcontroller acts as a hybrid acquisition hub, seamlessly integrating the analog EOG sampling with the digital Video-Oculography (VOG) stream without compromising the strict temporal requirements of the electrical data.

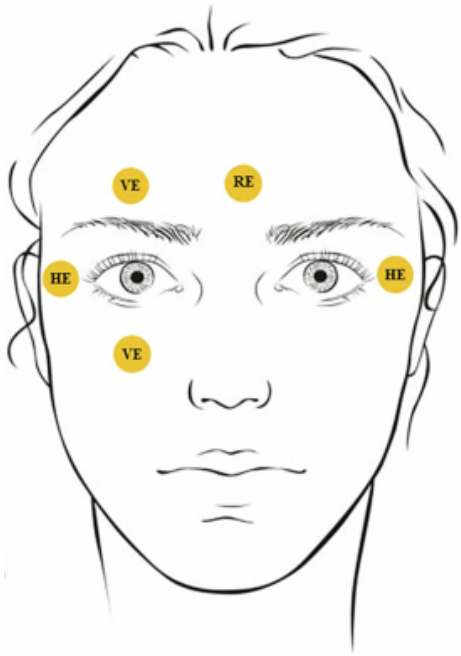


Figure 2.4: EOG Placement

2.4.1 Power Supply and Signal Isolation

To ensure portability and mitigate power-line interference, the wearable system operates entirely untethered, powered by a single-cell 3.7 V, 850 mAh Lithium-Polymer (LiPo) battery. This stand-alone power architecture intrinsically eliminates ground-loop artifacts associated with USB or mains power supplies, significantly improving the Signal-to-Noise Ratio (SNR) of the microvolt-level EOG acquisition.

2.4.2 System Schematic and Wiring Architecture

The hardware integration of the EOG acquisition modules and the central microcontroller is detailed in the system schematic (Figure 2.5). The architecture centres around the XIAO ESP32-S3 Sense, which manages power distribution, analog-to-digital conversion, and peripheral control.

Two modified AD8232 analog front-ends are utilized for independent channel acquisition. Both modules are powered directly by the ESP32's regulated 3.3 V and ground (GND) pins to maintain a unified logic level and common ground reference.

- **Horizontal Channel:** The left and right canthus electrodes (Electrodes L and R) are connected to the Right Arm (RA) and Left Arm (LA) inputs of the first AD8232. Its conditioned analog output is routed to the microcontroller's 12-bit ADC via pin D0.
- **Vertical Channel:** The supra-orbital and infra-orbital electrodes (Electrodes U and D) are connected to the RA and LA inputs of the second AD8232. Its analog output is routed to pin D1. This module is also responsible for driving the Reference Electrode via the Right Leg (RL) pin to actively suppress common-mode interference.

To support the optical tracking pipeline in variable ambient lighting, the circuit includes a dedicated white LED. Driven by digital pin D9 through a 220 ohm resistor, this LED provides the constant, controlled corneal illumination required to generate the primary glint for the computer vision algorithm.

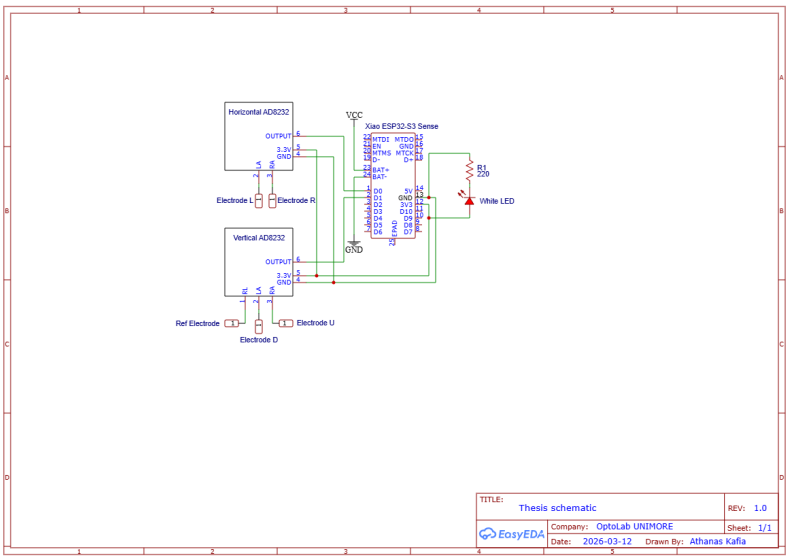


Figure 2.5: Circuit Schematics

2.4.3 Deterministic Sampling Architecture

Precise timing is a prerequisite for accurate sensor fusion. To prevent timing jitter caused by network communication or optical processing overhead, the system utilizes a hardware timer interrupt. The timer triggers an Interrupt Service Routine (ISR) exactly every 3333 μs, resulting in a strictly deterministic sampling frequency (f_s) of 300 Hz. Inside the ISR, the analog voltages from the Horizontal and Vertical channels are read via the microcontroller's 12-bit ADC.

2.4.4 Optical Integration: OV2660 Camera Module

To support the hybrid sensor-fusion architecture, the ESP32S3 Sense utilizes its integrated Digital Video Port (DVP) connected to an OV2660 camera module. Due to the computational constraints of edge-microcontrollers, performing robust, geometry-filtered pupil tracking directly on-chip results in severely degraded frame rates. Consequently, the ESP32 is configured strictly as a high-speed optical streaming node. The camera captures the ocular region in VGA resolution, compresses the frames into JPEG format on-the-fly, and prepares them for wireless transmission, outsourcing the heavy computer vision calculations to the host machine.

2.4.5 Distributed Processing and Communication Architecture

To ensure that the heavy payload of video streaming does not interrupt the deterministic 300 Hz sampling of the EOG, the system abandons physical tethering in favor of a dual-channel, entirely wireless communication protocol over IEEE 802.11 Wi-Fi:

1. **The Electrical Stream (UDP):** Batched EOG data and Psychtoolbox synchronization markers are transmitted directly to the primary MATLAB host via the User Datagram Protocol (UDP). This ensures instantaneous, low-latency delivery of the biopotentials while granting the user unrestricted freedom of movement.

2. **The Optical Stream (HTTP/UDP):** Concurrently, the JPEG frames are streamed wirelessly to a dedicated Python script running in the background of the host PC.

The Python Vision Pipeline: The Python node receives the wireless frames at 15-25 FPS (contingent on local Wi-Fi bandwidth). It establishes a dynamic Region of Interest (ROI) anchored by the primary corneal reflection (glint), which is digitally removed via a Telea in-painting algorithm to prevent thresholding artifacts. To robustly isolate the pupil from the iris across varying lighting conditions, the system applies dynamic spatial thresholding based on local intensity peaks. Finally, the pipeline leverages massive morphological reconstruction, Convex Hull boundary wrapping, and geometric contour filtering (evaluating area, circularity, and anchor proximity) to extract the precise pupil centroid while rejecting tracking anomalies.

The Localhost UDP Bridge: Once the Python pipeline computes the absolute X, Y pixel coordinates of the pupil, it silently broadcasts these coordinates over the host PC's internal localhost network via UDP. The primary MATLAB experiment script, while managing the psychophysical stimulus and reading the incoming Wi-Fi EOG data, actively listens to this internal UDP port. Because UDP is stateless, MATLAB instantly pulls the absolute optical coordinates into its arrays without any blocking delays, successfully fusing the two asynchronous wireless sensor streams onto a single master timeline.

2.5 Wearable Electrode Interface

To mitigate variability in electrode placement and minimize setup time, a custom head-mounted interface was fabricated to replace standard disposable electrodes. This rigid structure ensures consistent sensor geometry relative to the ocular dipole across multiple experimental sessions.

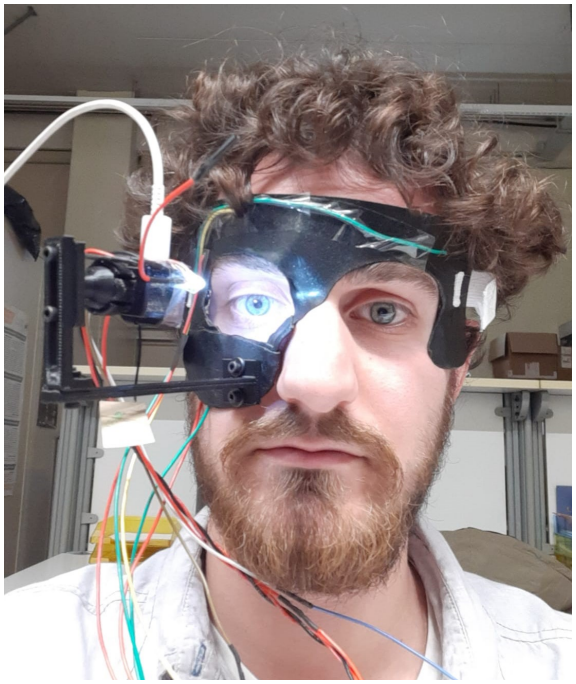


Figure 2.6: Photo of the mask during its initial tests

2.5.1 Mechanical Structure

The headset chassis is constructed from Black Worbla, a thermoplastic polymer chosen for its malleability and rigidity upon cooling. The frame was moulded directly to the facial contours to ensure a conformal fit. The design adopts a peri-orbital geometry, covering the supraorbital and infra-orbital

regions while leaving the central and peripheral visual fields distinctively unobstructed. Mechanical retention is provided by an adjustable elastic strap.

2.5.2 Sensor Integration and Mechanical Decoupling

To resolve the physical conflict between the rigid thermoplastic frame and the elasticity of human skin, a mechanical decoupling strategy was implemented. Standard disposable Ag/AgCl electrodes are adhered directly to the face. Rather than mounting the electrodes to the mask, only the wires that are connected to the electrodes are fixed on the mask, leaving a couple of centimetres of wire to be free and not block the electrodes with the rigid mask. This strain-relief mechanism completely isolates the adhesive electrodes from the mechanical movement of the mask. If the user shifts their head or flexes facial muscles, the "spare" wire absorbs the physical shock, leaving the electrode-skin interface perfectly undisturbed. Figure 2.6 is an image taken during the initial tests with the camera mounted to the mask.

Chapter 3

System Implementation

3.1 State of the Art: Strategies for EOG Signal Processing

The design of Electro-Oculography (EOG) interfaces addresses two primary constraints: baseline drift and signal non-linearity. While the electro-oculographic response is theoretically linear at small angles ($< 15°$), practical implementation is limited by the time-dependent signal decay of AC-coupled amplification and the geometric saturation of the electrical dipole at large eccentricities. Literature approaches to these challenges can be categorized into three paradigms: Physical Modelling, Sensor Fusion, and Non-Linear Signal Estimation.

3.1.1 Physical Modelling Approaches

A rigorously theoretical approach is to treat the eye not merely as a signal source but as a geometric dipole within a volume conductor. Barbara et al. (2019) [13] proposed a "Battery Model" of the eye, moving away from data-driven "black-box" regression toward a "white-box" physical model. Their method explicitly models the EOG potential V as a function of the specific distances between the electrode placement and the centres of the cornea and retina (d_c and d_r).

The primary advantage of this approach is its anatomical validity; by estimating the true position of the electrodes relative to the eye, it can theoretically decouple the gaze angle from artefactual variance. However, the practical limitation of this method lies in its calibration complexity. To function correctly, the system must accurately estimate subject-specific parameters—such as the exact 3D location of the electrodes and the geometric centre of rotation of the eye—using complex optimization algorithms (Levenberg-Marquardt) before gaze estimation can even begin. For a low-cost system intended for rapid deployment, this requirement introduces a significant barrier to entry, as it relies on precise anatomical assumptions that may not hold for a user with a simple, non-customized electrode mask.

3.1.2 Sensor Fusion and Continuous Calibration

Addressing the issue of long-term stability, particularly for wearable devices, Nezvadovitz and Rao (2022) [14] explored the potential of sensor fusion. Their work acknowledges that EOG signals are inherently unstable over long durations due to sweat, skin impedance changes, and retinal light adaptation. To counter this without requiring the user to constantly look at calibration targets, they utilized the Vestibulo-Ocular Reflex (VOR).

By integrating a gyroscope (IMU) with the EOG system, their algorithm correlates head rotation with compensatory eye movements. Using a Hidden Markov Model (HMM), the system continuously

re-calibrates the EOG scaling factors in the background whenever the user moves their head. While this is a highly innovative solution for "free-living" mobile eye tracking, it imposes a hardware penalty: it necessitates an Inertial Measurement Unit (IMU) and requires the user to be active. For a stationary Human-Computer Interaction (HCI) setup—such as a paralyzed patient controlling a computer screen—head movements may be infrequent or impossible, rendering the VOR-based calibration strategy ineffective.

3.1.3 Non-Linear Signal Estimation

A third paradigm focuses on exploiting the intrinsic signal characteristics of saccades to resolve drift mathematically. Manabe et al. (2015) [15] proposed a "Direct Gaze Estimation" technique based on the nonlinearity of the EOG signal.

Manabe's key insight addresses the drift problem by changing what is measured. Instead of tracking the absolute DC voltage level (which drifts due to electrode half-cell potentials and AC coupling), the method estimates the absolute eye angle based on the voltage difference (ΔV) observed during a saccade. By modelling the voltage-to-angle relationship as a polynomial function $p(\theta)$ rather than a linear constant, the algorithm can infer the starting and ending positions of the eye purely from the magnitude of the jump, independent of the drifting baseline. This approach effectively turns the non-linearity of the signal (usually considered a defect) into a feature that helps uniquely identify the gaze angle.

3.1.4 Justification for Proposed Methodology

After evaluating these approaches, the 'Direct Gaze Estimation' method proposed by Manabe et al. (2015) [15] was selected as the framework for this study. This selection is based on three design criteria:

1. **Hardware Minimalism:** The method relies exclusively on bio-potential acquisition, eliminating the need for auxiliary inertial sensors required by VOR-based fusion.
2. **Computational Efficiency:** The polynomial model can be resolved using standard least-squares optimization, avoiding the complex geometric parameter estimation associated with volume conductor models.
3. **Drift Resilience:** By analysing differential voltage steps (ΔV) rather than absolute DC levels, the algorithm mitigates the impact of AC-coupled signal decay inherent to the AD8232 architecture.

3.2 Data Acquisition and Pre-Processing

To bridge the gap between the physiological domain and the digital processing environment, a rigorous acquisition and synchronization pipeline was implemented to handle the asynchronous electrical and optical data streams.

3.2.1 The Experimental Protocol

Experimental validation was conducted using a fixed-geometry setup to ensure correspondence between induced eye rotation and ground-truth visual angles. The subject was positioned 51.7 cm from a display (59.7×33.4 cm), subtending a horizontal visual field of $\pm 30^{\circ}$ and vertical field of $\pm 16^{\circ}$. Head stability was maintained via a chin rest to isolate oculomotor rotation from vestibular artifacts.

Visual Stimulation (The Calibration Pattern):

Visual targets are generated using the Psychophysics Toolbox in MATLAB. To capture the full dynamic range of the AD8232 and map the operational Field of View (FOV) of the camera, a "Step-Ramp" protocol is utilized:

1. **Horizontal Sequence:** The target displaces from the centre ($0°$) to eccentricities of: $\pm 4°, \pm 8°, \pm 12°, \pm 16°$.
2. **Vertical Sequence:** Displacements from the centre ($0°$) up to $\pm 12°$ (limited by the vertical physiological range).
3. **Diagonal Sequence:** Combined angular displacements.
4. **Timing:** Each fixation is held for 1.5 seconds. This duration allows the EOG signal to settle into a stable plateau and provides the camera pipeline sufficient frame depth to extract a robust median pixel coordinate.

3.2.2 Optical Tracking Algorithm

While the EOG is acquired via deterministic hardware interrupts, the optical pupil coordinates are extracted continuously via a dedicated computer vision pipeline. The algorithm processes the wireless JPEG frames at approximately 25 FPS using the following sequential logic:

1. **Static Glint Anchoring:** To prevent the tracking algorithm from erroneously locking onto peripheral corneal reflections or tear ducts (the "canthus trap"), the primary LED glint is identified within a strictly constrained bounding box centred in the camera frame. This leverages the mechanical rigidity of the mask, assuming the glint remains relatively stationary while the pupil rotates around it.
2. **In-painting and Pre-Processing:** A dynamic Region of Interest (ROI) is cropped around the detected glint. To prevent the high-intensity LED reflection and its surrounding diffraction halo from corrupting the binarization maths, the glint mask is dilated and digitally removed using a Telea in-painting algorithm.
3. **Dynamic Binarization (Smoothed Otsu):** Rather than relying on a fixed greyscale threshold, the system utilizes Otsu's method [16] to dynamically calculate the optimal bimodal threshold of the in-painted ROI. To specifically isolate the dark pupil from the medium-toned iris, a scaling ratio is applied to the raw Otsu value. Furthermore, an infinite impulse response (IIR) temporal smoothing filter is applied to this threshold across sequential frames, effectively eliminating high-frequency binarization flicker.
4. **Morphological Reconstruction:** When the LED is perfectly collinear with the pupil, residual diffraction halos can cause the binarized pupil to appear as a hollow ring (the "donut effect"). A massive morphological closing operation using an elliptical structural element (21×21 pixels) is applied to the binary mask to aggressively fill internal topological holes before a gentle erosion clears external noise.
5. **Convex Hull and Geometric Filtering:** Morphological contours are extracted from the binary mask, and the largest candidate is selected. To repair boundary defects where the glint may have overlapped the pupil edge, this contour is wrapped in a Convex Hull—a mathematical boundary that

bridges missing segments to guarantee a closed, convex polygon. This finalized hull is then evaluated against strict physical limits, including Minimum/Maximum Area and maximum allowable distance from the glint anchor.

6. **Coordinate Extraction:** An ellipse is mathematically fitted to the Convex Hull vertices to derive the precise geometric centroid (P_X, P_Y) of the pupil. If the camera loses sight of the eye (e.g., during a blink or an extreme ocular rotation beyond the field of view), the pipeline gracefully outputs a `-1` flag. This allows the downstream MATLAB synchronization algorithm to intercept the missing data and inject `NaN` placeholders, preventing mathematical corruption during sensor fusion.

3.2.3 Signal Conditioning and Saccade Extraction

Raw EOG signals are inherently non-stationary and contaminated by artifacts. To reliably extract the voltage differences (ΔV) required for the Manabe calibration model, a multi-stage signal processing pipeline was implemented in MATLAB. This pipeline transforms the raw ADC stream into precise saccadic events.

1. **Spectral Filtering (Band-Pass):** The raw signal is first conditioned using a 2nd-order Butterworth Band-Pass Filter with cut-off frequencies of $0.1Hz$ and $10Hz$.
   - Low-Cut ($0.1Hz$): Removes the DC baseline drift and slow-changing skin potential artifacts. While this removes absolute position data, it stabilizes the signal for velocity estimation.
   - High-Cut ($10Hz$): Suppresses high-frequency electromyographic (EMG) noise and mains hum ($50Hz$), isolating the spectral energy of eye movements (typically $< 10Hz$).
2. **Velocity-Based Event Detection:** Saccades are detected using a velocity-thresholding algorithm based on the method proposed by Engbert and Kliegl (2003) [17].
   - Velocity Estimation: The continuous velocity profile $v(t)$ is calculated as the first derivative of the filtered signal multiplied by the sampling frequency (F_s).
   - **Adaptive Threshold:** To account for variable noise levels across trials, the detection threshold is dynamic rather than fixed. It is defined as:

$$thresh_{vel} = \lambda \cdot \sqrt{\left\langle (v - \langle v \rangle)^2 \right\rangle} \tag{3.1}$$

     Where $\lambda = 6$ (an empirically selected multiplier) and the variance term is calculated using the median rather than the mean to ensure robustness against outliers.
3. **Artifact Rejection (Blink Detection):** Blinks produce high-amplitude spikes that can mimic vertical saccades. A specific rejection module distinguishes these artifacts by analysing the high-frequency residual of the signal.
   - **Robust Scaling:** The signal's deviation is normalized by the local median absolute deviation (MAD).
   - **Rejection Logic:** Periods where the normalized residual exceeds a Z-score of 6.5 are flagged as blinks. Saccade candidates detected during these windows are automatically discarded to prevent calibration errors.

4. **Feature Extraction (ΔEOG Calculation):** Once a valid saccade is identified (start time t_{on} and end time t_{off}), the system calculates the exact voltage change (ΔEOG) corresponding to that eye movement. To ensure the measurement is not corrupted by the immediate signal overshoot or the subsequent AC decay, a Windowed Median approach is used instead of simple point subtraction:

   - **Pre-Saccade Value:** Calculated as the median voltage in a short window ($N = 15$ samples) ending just prior to the saccade onset.
   - **Post-Saccade Value:** Calculated as the median voltage in a short window starting immediately after the saccade peak stabilizes.

   $$\Delta EOG = \text{median}\left(V_{post}\right) - \text{median}\left(V_{pre}\right) \tag{3.2}$$

   This differential measurement captures the pure saccadic step amplitude while ignoring the absolute DC drift, providing the stable input required for the polynomial calibration model described in the next section.

3.3 Mathematical Modelling and Calibration

Once the EOG signal has been conditioned and segmented into discrete saccadic events, the system must translate these voltage differences (ΔV, measured in μV) into physical gaze coordinates ($\Delta\theta$ measured in degrees). While many low-cost eye trackers assume a linear relationship ($V = k \cdot \theta$), experimental data collected during the pilot phase of this work demonstrated that a linear model is insufficient for the full range of ocular motion ($\pm 25°$). To achieve high accuracy, this system implements a Bilateral 2nd-Order Polynomial Model optimized via a non-linear least squares solver for the electrical channel, complemented by an absolute linear model for the optical channel.

3.3.1 EOG Calibration: Bilateral Polynomial Model

The relationship between the corneo-retinal potential change and the eye angle is governed by the geometry of the volume conductor (the skull and skin).

Non-Linearity Compensation:

Experimental data indicates that linear regression models ($y = mx+q$) produce systematic undershoot errors at eccentricities beyond $\pm 20°$ due to dipole saturation. To compensate for this characteristic, a 2nd-order polynomial model is implemented:

$$V\left(\theta\right) = c_2 \cdot \theta^2 + c_1 \cdot \theta + c_0 \tag{3.3}$$

The quadratic term (c_2) accounts for the geometric saturation observed in both vertical and horizontal channels.

Bilateral Asymmetry:

Anatomically, the electrical field of the eye is not perfectly isotropic. The nasal bone structure differs from the temporal bone structure, and electrode placement is rarely perfectly symmetrical relative to the eye's dipole centre. Consequently, a single polynomial cannot fit both leftward (downward) and rightward (upward) movements with equal precision. To resolve this, the system employs a Bilateral Model, splitting the calibration space into two independent domains based on the direction of movement:

- **Positive Domain (p_{pos}):** Active when $\Delta V > 0$ (Right/Up).
- **Negative Domain (p_{neg}):** Active when $\Delta V < 0$ (Left/Down).

This results in four distinct sets of coefficients (Horizontal Left/Right, Vertical Up/Down), ensuring that anatomical asymmetries do not compromise tracking accuracy.

3.3.2 The Numerical Solver and cost function

To determine the optimal coefficients (p_{pos}, p_{neg}) that map the raw voltages to the known target angles, the system formulates an optimization problem. Instead of simple algebraic curve fitting (e.g., polyfit), the system utilizes the Levenberg-Marquardt algorithm, implemented via the MATLAB solver lsqnonlin.

The Optimization Loop:

The solver iteratively adjusts the sequence of estimated angles $\hat{\theta}$ to minimize the difference between the "Forward Model" (what voltage the model predicts for a given angle) and the "Observed Data" (the actual voltage measured by the AD8232).

Cost Function:

The objective function $J(\theta)$ minimizes the weighted residual sum of squares:

$$J(\theta) = \sum_{k=1}^{N} w_k \cdot \left(\Delta V_{observed}[k] - \text{Model}(\Delta\theta_k)\right)^2 \tag{3.4}$$

Where:

- $\Delta V_{observed}$: The saccadic voltage step extracted by the pre-processing pipeline.
- $\text{Model}(\Delta\theta)$: The predicted voltage calculated using the current polynomial coefficients (p_{pos} or p_{neg}).
- w_k: A confidence weight assigned to each fixation. High-quality fixations (stable, low noise) are assigned higher weights, forcing the model to prioritize clean data over noisy artifacts.

The weight $w_k \in [0, 1]$ for the k-th fixation is calculated as the product of four penalty factors derived from the signal segment $x[n]$:

1. **Variance Penalty (Q_{mad}):** Based on the Median Absolute Deviation (MAD). Fixations with high variance (noise) are penalized.

$$Q_{mad} = \frac{1}{1 + \text{MAD}(x)} \tag{3.5}$$

2. **Stability Penalty (Q_{vel}):** Based on the 95th percentile of the intra-fixation velocity. High internal velocity indicates drift or micro saccades.

$$Q_{vel} = \frac{1}{1 + \frac{v_{95}}{4 \cdot \text{MAD}(x)}} \tag{3.6}$$

3. **Amplitude Penalty (Q_{amp}):** Penalizes unphysiologically large excursions within a fixation window (indicative of blinks).

4. **Duration Penalty (Q_{len}):** Penalizes fixations that are too short (< 100 ms) to be reliable.

During the calibration phase, the final weight is raised to a power ($w_i^{1.8}$) to strictly suppress non-ideal data points. The power was determined empirically. Standard Least Squares regression assumes that all data points are equally valid. However, EOG calibration data is often contaminated by blinks, unstable fixations, or fleeting saccades. The reason for introducing the weights into the cost function is to prevent these poor-quality points from skewing the model.

3.3.3 Robust Initialization Strategy

A major risk when inverting polynomial models numerically is the "Phantom Root" problem. Since a quadratic function is parabolic, it maps two different angle values to the same voltage level (one physical root and one non-physical "phantom" root on the other side of the vertex).
If the Levenberg-Marquardt solver is initialized with a naive guess (e.g., $\theta_{init} = 0°$), local gradients may push the solution toward the phantom root or get stuck in a local minimum, producing nonsensical gaze estimates (e.g., predicting $150°$ for a $10°$ movement).

Sensitivity-Scaled Initialization

To guarantee convergence to the correct physical root, the solver requires an initial guess that is strictly within the "basin of attraction" of the true solution. The system implements a Sensitivity-Scaled Initialization routine. Before the non-linear optimization begins, an initial search vector $\hat{\theta}_{init}$ is generated using a robust linear approximation:

$$\hat{\theta}_{init}[k] = \frac{\Delta V[k]}{S_{nom}} \tag{3.7}$$

Where S_{nom} is a fixed Nominal Sensitivity Factor.
Through empirical analysis of the AD8232 gain and electrode impedance, this factor was set to:

$$S_{nom} = 3.5 \mu V/° \tag{3.8}$$

This value was chosen as a "safe conservative estimate"—it is slightly higher than the typical vertical sensitivity (≈ 2.5) and horizontal sensitivity (≈ 3.0), which intentionally prevents the initial guess from overshooting the target. By seeding lsqnonlin with this realistic approximation, the solver is constrained to the linear region of the manifold, ensuring it refines the accuracy rather than searching for the root blindly.

3.3.4 Optical Calibration: Absolute Linear Model

While the electrical channel leverages differential steps to combat AC-decay, the complementary optical camera acts as a direct spatial sensor. It does not suffer from AC-coupled decay, nor does the pixel displacement saturate non-linearly within the operational $\pm 25°$ tracking limit. Consequently, the complex integration and non-linear optimization models required for the EOG are unnecessary for the optical channel. The optical calibration utilizes a stateless, 1st-order linear regression:

$$Angle = m \cdot P_{abs} + q \tag{3.9}$$

Where P_{abs} is the absolute median pixel coordinate extracted during the stable fixation window. By mapping the absolute stable pixel values directly to the known target angles, the camera pipeline is rendered entirely immune to cumulative error. If a blink occurs during a movement, the system simply waits

until the eye reopens and instantly reads the correct absolute position. This mathematical independence between the non-linear, differential EOG model and the linear, absolute optical model forms the robust foundation necessary for the subsequent multimodal sensor fusion phase.

3.4 Multimodal Drift Correction: Stateless Optical Anchoring

The most significant challenge in adapting the AD8232 for oculography is its inherent instability for static measurements. The internal high-pass filter creates an exponential signal decay ($V_{out}(t) = V_{step} \cdot e^{-t/\tau}$) during static fixations. This decay manifests as Integration Drift: as the Joint Angle Estimation (JAE) algorithm cumulatively sums the predicted angular jumps to determine the absolute eye position, the baseline slides continuously toward infinity.

3.4.1 The Integration Trap

Early iterations of this system attempted to fuse the sensors by calculating the pixel differences of the camera data and integrating them alongside the EOG. This approach suffered from the "Integration Trap." If the user blinked during a saccadic movement, the optical camera was temporarily blinded and missed the jump. Because the displacement was integrated via cumulative sum (cumsum), that missed amplitude was permanently omitted from the spatial timeline, corrupting the remainder of the tracking session.

3.4.2 The Sensor Fusion Algorithm

To definitively resolve baseline drift, the system abandons optical integration in favour of continuous Absolute Anchoring. The Multimodal Sensor Fusion algorithm operates as follows:

1. **Raw Estimation:** The EOG voltage differences are passed through the JAE algorithm, producing an estimated gaze trajectory ($\hat{\theta}_{raw}$). As expected, this trajectory diverges from the true targets over time due to AC-coupled drift.

2. **Absolute Anchoring:** Concurrently, the optical pixels are passed through the linear calibration model, generating an array of absolute estimated angles ($Anchor_{opt}$).

3. **Dynamic Error Subtraction:** For every fixation index k, the algorithm evaluates the optical anchor. If the camera successfully tracked the pupil (i.e., the value is not NaN), the instantaneous integration drift is calculated as the difference between the EOG's drifting estimation and the camera's absolute truth:

$$Drift_{curr} = \hat{\theta}_{raw}[k] - Anchor_{opt}[k] \tag{3.10}$$

4. **Correction:** This calculated drift is subtracted from the current and subsequent EOG position estimates:

$$\hat{\theta}_{corr}[k] = \hat{\theta}_{raw}[k] - Drift_{curr} \tag{3.11}$$

System Resilience:

This architecture provides robust fail-over redundancy. If the optical camera loses tracking due to an extended blink, extreme physical occlusion, or movements outside the camera's FOV, the $Anchor_{opt}$ evaluates to NaN. The algorithm immediately ceases updating the $Drift_{curr}$ variable and holds the last

known drift value constant. This allows the high-speed EOG to seamlessly take over, tracking the eye purely electrically until the camera recovers its lock on the pupil and re-anchors the signal.

3.5 Real-Time Multimodal System Implementation

To demonstrate the practical viability of the sensor fusion architecture, a live multimodal gaze-tracking interface was developed in MATLAB. Unlike the offline analysis, which prioritizes statistical rigour and bidirectional filtering, the real-time system prioritizes low latency, asynchronous data handling, and immediate responsiveness. Crucially, to ensure complete galvanic isolation, eliminate 50 Hz mains powerline noise, and allow untethered user mobility, the EOG acquisition hardware is fully battery-powered and communicates entirely over Wi-Fi.

3.5.1 Online vs. Offline Processing

The system architecture distinguishes between the static calibration phase and the dynamic real-time tracking loop.

- **Calibration (Iterative Optimization):** Hardware coefficients are derived offline using the Levenberg-Marquardt algorithm (`lsqnonlin`), which iteratively minimizes the residual error across the full multimodal dataset.
- **Real-Time Tracking (Analytic Execution):** To minimize latency during active tracking, the system utilizes direct analytic conversions. The EOG voltages are mapped via the quadratic formula, and the optical pupil coordinates are mapped via the calibrated linear models. This reduces the computational complexity of the coordinate transformation to O(1) per sample.

3.5.2 The Real-Time Multimodal Loop

The core of the application is a high-frequency `while` loop that manages dual-stream wireless communication, adaptive sensor fusion, and graphical rendering.

1. **Asynchronous Stream Parsing (Wi-Fi & UDP):** The system simultaneously listens to two distinct wireless pipelines. High-frequency EOG data (300 Hz) is parsed via a Wi-Fi stream from the untethered, battery-powered microcontroller. Concurrently, the optical pupil coordinates (P_X, P_Y) are received asynchronously via an IPv4 UDP datagram bridge from the Python computer vision pipeline. To prevent optical latency from compounding, the UDP buffer is flushed at each iteration, extracting only the most recent valid coordinate pair.
2. **EOG Conditioning and Analytic Inversion:** To suppress electrical noise without introducing the lag of a high-order digital filter, a small Moving Average Buffer (N=5) is applied to the raw EOG voltages. Given the calibrated coefficients (a, b, c), the relative EOG gaze angle θ_{eog} is derived instantly using the quadratic formula:

$$a\theta^2 + b\theta + (c - V_{curr}) = 0 \tag{3.12}$$

   To ensure system stability during massive artifacts where the term under the square root might become negative, a linear safety fallback ($V_{curr}/3.5$) is implemented to prevent mathematical execution crashes.

3. **Optical Absolute Anchoring:** When valid pupil coordinates are received via the UDP bridge, they are immediately mapped into absolute visual degrees using the pre-calibrated subject-specific linear equations. These serve as the absolute ground-truth anchors for the fusion algorithm.

4. **Adaptive Complementary Sensor Fusion:** Continuous drift management is handled automatically via a dynamic Innovation Filter. The algorithm predicts the current eye position by integrating the high-frequency EOG displacement ($\Delta\theta_{eog}$) with the previously known position.

   If the camera is actively tracking the pupil, the system calculates the geometric disagreement between the EOG prediction and the optical anchor. A dynamic weight (W_{dyn}) is calculated based on a predefined rejection limit (e.g., 6.0 degrees):

   $$\theta_{fused} = W_{dyn} \cdot (\theta_{prev} + \Delta\theta_{eog}) + (1 - W_{dyn}) \cdot \theta_{pupil} \tag{3.13}$$

   If the sensors agree, the EOG provides high-frequency smoothing to the optical signal. If they wildly disagree (e.g., during a vertical eyelid artifact), W_{dyn} scales to 0, rejecting the EOG artifact. Conversely, if the camera goes blind (e.g., during a blink or extreme rotation), the system forces $W_{dyn} = 1$, relying 100% on the EOG to seamlessly bridge the optical blind spot.

5. **Visualization and Threading:** To visually validate the fusion algorithm, the GUI simultaneously renders three independent cursors: the raw drifting EOG, the absolute optical anchor, and the final fused output. To prevent the MATLAB graphics engine from blocking the high-speed data acquisition, the plot update command uses `drawnow limitrate`. This decouples the display latency from the background processing, updating the screen at 17 to 25 FPS while allowing the mathematical fusion loop to run unhindered.

Chapter 4

Results

4.1 Experimental Validation Framework

System performance was evaluated within a controlled experimental environment designed to isolate oculomotor variables. The setup minimizes external noise sources, specifically vestibular artifacts (head movement) and retinal potential drift caused by luminance variations. The validation process was conducted in two distinct phases:

- **Phase I: Single-Subject System Validation** utilized the custom-moulded multimodal mask on a primary subject to extensively validate the sensor fusion algorithm, characterizing its performance across Horizontal, Vertical, and Diagonal (2D) modalities.
- **Phase II: Multi-Subject Validation** evaluated the final hardware and software architecture across a group of five subjects ($N = 5$) to quantify inter-subject variability, ensuring the subject-specific calibration models and the fusion architecture generalize across different facial geometries and physiological profiles.

4.1.1 Physical Setup and Environmental Control

The subject is positioned at a fixed viewing distance of 51.7 cm from a 27-inch LCD monitor (1920×1080 pixels; physical dimensions 59.7×33.4 cm). This geometry establishes the maximum visual stimuli boundaries for the experiment:

- **Screen Horizontal Range:** A half-width of 29.85 cm subtends a visual angle of $\pm 30.0°$.
- **Screen Vertical Range:** A half-height of 16.7 cm subtends a visual angle of $\pm 17.9°$.

While the corneo-retinal dipole measured by the EOG is theoretically unconstrained by these limits and can track extreme ocular excursions, the functional limits of the multimodal fusion system are bounded by the optical camera's field of view and the physical constraints of the mask assembly (such as eyelid occlusion and anatomical self-shadowing). The reliable tracking boundaries for the optical pupil pipeline were experimentally determined as:

- **Optical Horizontal Range:** $[-16°, +16°]$.
- **Optical Vertical Range:** $[-12°, +12°]$.

When a subject's gaze exceeds these optical limits, the camera starts to lose sight of the pupil (the probability of finding the pupil falls below 90%), and the multimodal architecture seamlessly falls back to 100% EOG integration to bridge the blind spot until the pupil re-enters the valid optical tracking window.

Subject Stabilization

Head stabilization is maintained via a chin rest with forehead support. This constraint isolates ocular rotation from head translation, mitigating vestibular artifacts in the absence of inertial compensation.

4.1.2 Lighting Conditions and Environmental Robustness

Unlike traditional strict-laboratory setups, data collection was conducted across various times of day and ambient lighting conditions (ranging from natural daylight to dim artificial evening light) to validate the real-world robustness of the multimodal architecture. The system's tracking accuracy remained consistent regardless of the ambient environment due to two key design factors:

- **Optical Invariance:** The implementation of temporally smoothed, dynamic bimodal thresholding (Otsu's method) ensured that the pupil binarization pipeline continuously adapted to its local Region of Interest, making the computer vision algorithm independent of macroscopic room lighting variations.
- **Electrical AC-Coupling:** While the absolute magnitude of the Corneo-Retinal Potential (CRP) naturally fluctuates with light-dark adaptation (known physiologically as the Arden Ratio), the EOG acquisition hardware utilizes AC-coupling and band-pass filtering. This isolates the high-frequency saccadic kinematics, effectively rendering the displacement estimation immune to slow, light-induced baseline voltage drifts.

Furthermore, it was empirically observed that dimmer ambient lighting environments (such as during evening sessions) subjectively improved the participants' ability to comfortably focus on the backlit LCD monitor. This enhanced visual contrast naturally reduced subject fatigue and spontaneous blink rates, ultimately yielding cleaner kinematic data with fewer high-frequency artifacts.

4.1.3 Visual Stimulation and Protocol Design

Visual stimuli were generated using the Psychophysics Toolbox (Psychtoolbox) within MATLAB. To ensure microsecond-precision synchronization without inducing serial communication bottlenecks, a unidirectional "timestamp snapping" architecture was implemented. Upon every optical screen update, MATLAB instantaneously captures the most recent hardware timestamp from the continuous EOG data stream broadcasted by the custom wearable mask. This perfectly aligns the visual stimulus onset with the microcontroller's internal hardware clock, completely eliminating the need for bidirectional event markers and preventing serial buffer overflows.

The Step-Ramp Protocol

The quantitative validation presented in this chapter focuses exclusively on Discrete Step Saccades using a Return-to-Centre protocol (e.g., $0^\circ \rightarrow Target \rightarrow 0^\circ$). This protocol was selected because it tests the system on diverse saccade amplitudes and directions simultaneously. It guarantees the system is rigorously tested on movements initiating from, and terminating at, varying eccentricities.

4.1.4 Performance Metrics

The primary metric for system accuracy is the Root Mean Square Error (RMSE), calculated between the estimated gaze angle $\hat{\theta}[n]$ and the ground-truth target angle $\theta_{target}[n]$ over the stable fixation windows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(\hat{\theta}[n] - \theta_{target}[n])^2} \tag{4.1}$$

This metric aggregates both systematic bias (calibration error) and stochastic noise (jitter) into a single value, providing a comprehensive measure of system reliability. Throughout this chapter, RMSE will be used to benchmark the raw EOG drift against the final Fused System output.

4.2 Signal Characterization and Feasibility Analysis

Before evaluating the final multimodal gaze estimation accuracy, specific characterization experiments were performed on the raw EOG signals to validate the design assumptions of the sensor fusion pipeline. These tests quantified the drift dynamics, the mechanical hysteresis of the vertical channel, and the cross-talk between axes, ultimately justifying the introduction of the optical pupil tracker.

4.2.1 Drift Dynamics: The "Random Walk" Behaviour

To determine if the baseline drift of the AC-coupled EOG could be modelled and subtracted using a traditional linear or polynomial function, a drift trend analysis was conducted. The system calculated the estimation error at every "Return-to-Centre" event over a sequence of continuous saccades.

A linear regression model ($y = mx + q$) and a high-order polynomial (20th-order) were fitted to these error points to test for predictability. The results demonstrated a severe failure of "blind" signal processing:

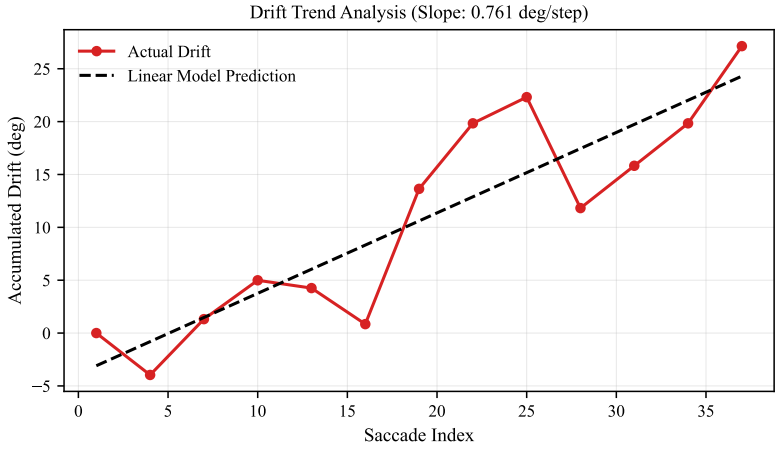


Figure 4.1: Drift trend

1. **Stochastic Fluctuations:** The actual drift deviated violently from both linear and polynomial predictions, exhibiting severe stochastic fluctuations (e.g., drifting $+20°$ in one direction before randomly reversing).

2. **Over-fitting of Saccadic Steps:** Because saccades manifest as low-frequency "steps," high-order polynomials cannot distinguish between unwanted AC-decay and intentional eye movements. The polynomial attempts to subtract the user's actual gaze shifts, resulting in severe signal distortion.

This result definitively invalidates standard continuous de-trending strategies. The drift is confirmed to be non-stationary, non-linear, and mathematically unpredictable, sharing identical spectral characteristics with the saccadic movements themselves.

4.2.2 The Multimodal Paradigm Shift: Optical Anchoring

The failure of blind mathematical de-trending necessitates a semantic paradigm shift. To distinguish true drift from true movement, the system must have access to an absolute, un-drifting reference point.

This physiological limitation was the primary catalyst for the development of the Multimodal Sensor Fusion architecture. By integrating a high-speed optical camera into the mask, the system generates "Absolute Optical Anchors." When processed through the Adaptive Complementary Filter developed in Section 3, these absolute optical coordinates dynamically correct the integrated EOG displacement on the fly. This architecture ensures that the EOG is only utilized for what it does best (high-frequency kinemetics and extreme eccentricities), while the optical tracker eliminates the unpredictable random walk.

4.2.3 Vertical Hysteresis and Eyelid Artifacts

To investigate the non-linearities of the vertical channel, a specific "Hysteresis Check" was performed, comparing Top→Centre movements against Centre→Bottom movements.

The analysis revealed significant mechanical hysteresis. Upward movements yielded a consistent voltage response, reflecting the smooth relaxation of the Levator muscle. Conversely, downward gaze exhibited severe variance and signal saturation. This discrepancy is mechanically attributed to the upper eyelid descending and acting as a sliding dielectric over the electrically positive cornea (Bell's Phenomenon). This physical artifact justifies two critical algorithmic choices:

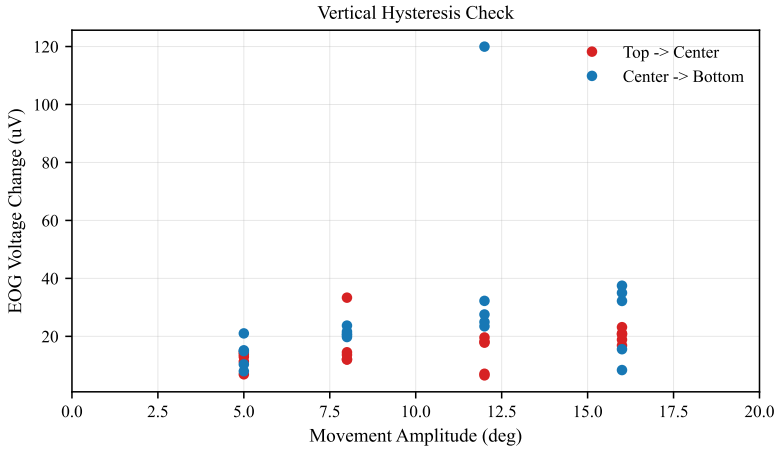


Figure 4.2: Hysteresis of the vertical channel

- **Bilateral Polynomial Modelling:** A single symmetric calibration model would fail to capture this position-dependent sensitivity, necessitating the split (Positive/Negative) quadratic calibration approach.
- **Adaptive Filter Innovation Rejection:** Because eyelid artifacts can cause sudden, massive spikes in vertical EOG voltage, the Complementary Filter's dynamic weight (W_{dyn}) was programmed with an "Innovation Rejection Limit." If the EOG predicts a massive vertical jump but the optical camera verifies the eye did not move, the algorithm automatically rejects the eyelid artifact, keeping the cursor perfectly stable.

4.2.4 Channel Independence (Crosstalk Analysis)

To justify the decision to model the Horizontal and Vertical channels as independent 1D systems (rather than a computationally expensive coupled 2D mixing matrix), a Crosstalk Analysis was performed. The artefactual voltage induced on the Vertical Channel ($\Delta V_{leakage}$) was measured during purely Horizontal Saccades.

A linear regression ($\Delta V_{leakage} = \alpha \Delta V_{signal}$) yielded a crosstalk factor of $\alpha = -0.005$. This corresponds to a signal leakage of just 0.5%. Because the crosstalk is negligible ($< 1\%$), the Horizontal and Vertical components can be mathematically treated as orthogonal, validating the use of two separate, highly efficient 1D solvers for real-time tracking.

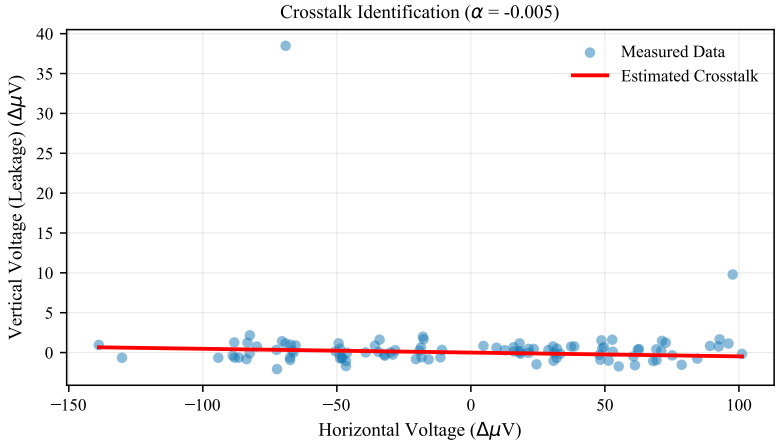


Figure 4.3: Crosstalk analysis between horizontal and vertical channels

4.3 Phase I: Single-Subject System Validation

This phase focuses on the quantitative validation of the fully integrated multimodal mask on a primary subject. The primary objective is to demonstrate how the continuous fusion of high-frequency EOG kinematics with absolute optical anchors overcomes the fundamental limitations of each stand-alone sensor.

4.3.1 Horizontal Gaze Estimation

The quantitative validation of the horizontal axis evaluates the system's ability to model the corneo-retinal dipole and correct AC-coupled baseline drift over an extended sequence of saccadic movements. The analysis is divided into model calibration, independent estimation, and temporal stability.

4.3.1.1 Calibration Performance

The horizontal calibration utilized the 2nd-order Bilateral Polynomial Model. As illustrated in Figure 4.4, the rightward (positive) tracking data exhibits a highly linear response, while the leftward (negative) data demonstrates the expected dipole saturation curve at larger eccentricities ($\geq 15°$). The application of the bilateral model successfully captured this anatomical asymmetry. The regression fit yielded a Raw Sigma of 16.2 ADC units. Relative to the maximum signal amplitude, this translates to a Normalized Sigma of 0.044 (4.4% residual error). Due to this low variance, the algorithm assigned a maximum fusion weight of 1.00 to the horizontal EOG channel for the subsequent estimation phase.

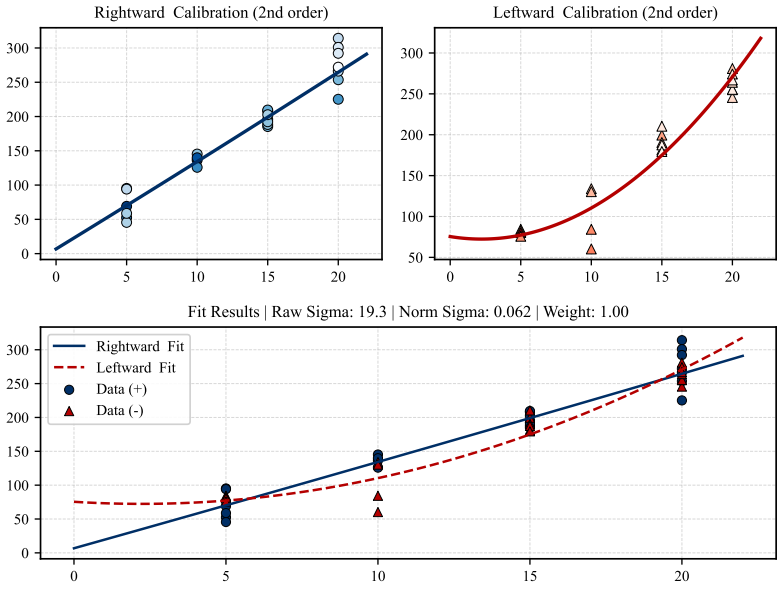


Figure 4.4: Horizontal Calibration

4.3.1.2 Sensor Fusion and Estimation Accuracy

To evaluate generalization, the calibrated model was applied to an independent estimation dataset comprising approximately 50 target transitions. Figure 4.5 presents the comparison between the uncorrected EOG signal, the optical pupil anchors, and the final fused output. Driven by the inherent properties of AC-coupled amplification, the raw EOG signal exhibited severe baseline wandering, accumulating an Absolute Mean Error of $1.22°$ and a Root Mean Square Error (RMSE) of $13.39°$. The discrete optical anchors provided by the camera module maintained a stable absolute reference with an RMSE of $2.13°$. The multimodal sensor fusion pipeline successfully synchronized these inputs, utilizing the optical anchors to dynamically reset the EOG integration baseline. The resulting Fused Output closely tracked the true target coordinates, achieving a final Absolute Mean Error of $1.84°$ and an RMSE of $2.51°$.

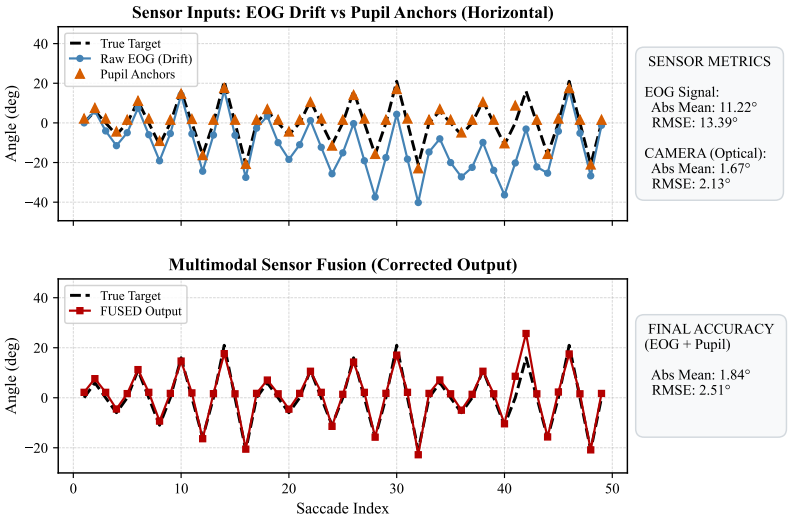


Figure 4.5: Horizontal Estimation Performance

4.3.1.3 Temporal Stability

The efficacy of the drift correction is further demonstrated by analysing system accuracy over time. The estimation sequence was segmented into three distinct repetition cycles to measure cumulative error (Figure 4.6). The raw EOG channel demonstrates a predictable degradation, with the RMSE increasing progressively across the three cycles ($5.2° \rightarrow 14.7° \rightarrow 16.9°$) as integration drift accumulated. Conversely, the fused output maintained a flat error profile, recording RMSE values of $1.8°$, $1.8°$, and $3.5°$ across the respective cycles. This confirms that the sensor fusion architecture effectively arrests continuous temporal drift, maintaining consistent tracking accuracy over extended operational periods.

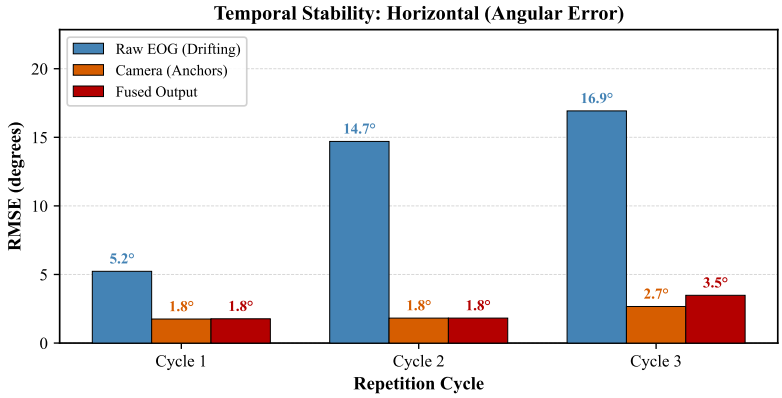


Figure 4.6: Cyclic RMSE Performance of Horizontal Movements

4.3.2 Vertical Gaze Estimation

The quantitative validation of the vertical axis evaluates the system's ability to model the corneo-retinal dipole and correct the pronounced AC-coupled baseline drift associated with vertical ocular movements. The analysis is divided into model calibration, independent estimation, and temporal stability.

4.3.2.1 Calibration Performance

The vertical calibration utilized the 2nd-order Bilateral Polynomial Model. As illustrated in Figure 4.7, both the upward (positive) and downward (negative) tracking data demonstrate non-linear dipole saturation curves at larger eccentricities. The application of the bilateral model successfully captured this geometry. The regression fit yielded a Raw Sigma of 16.6 ADC units. Relative to the maximum signal amplitude, this translates to a Normalized Sigma of 0.094 (9.4% residual error). Due to this acceptable variance, the algorithm assigned a maximum fusion weight of 1.00 to the vertical EOG channel for the subsequent estimation phase.

4.3.2.2 Sensor Fusion and Estimation Accuracy

To evaluate generalization, the calibrated model was applied to an independent vertical estimation dataset comprising approximately 50 target transitions. Figure 4.8 presents the comparison between the uncorrected EOG signal, the optical pupil anchors, and the final fused output. Driven by the inherent properties of AC-coupled amplification and mechanical eyelid artifacts, the raw EOG signal exhibited severe baseline wandering, accumulating an Absolute Mean Error of $12.62°$ and a Root Mean Square Error (RMSE) of $16.22°$. The discrete optical anchors provided by the camera module maintained a stable absolute reference with an RMSE of $2.53°$. The multimodal sensor fusion pipeline successfully synchronized

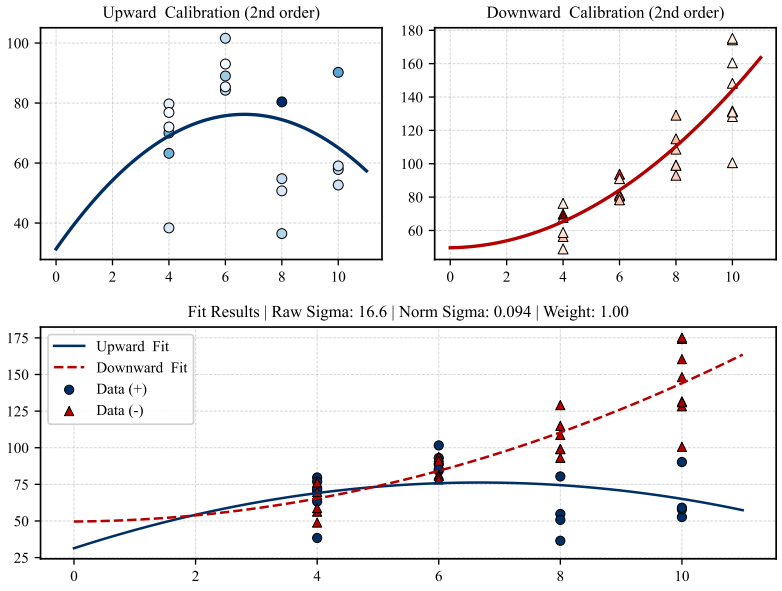


Figure 4.7: Vertical Calibration

these inputs, utilizing the optical anchors to dynamically reset the EOG integration baseline. The resulting Fused Output closely tracked the true target coordinates, achieving a final Absolute Mean Error of 2.24° and an RMSE of 3.92°.

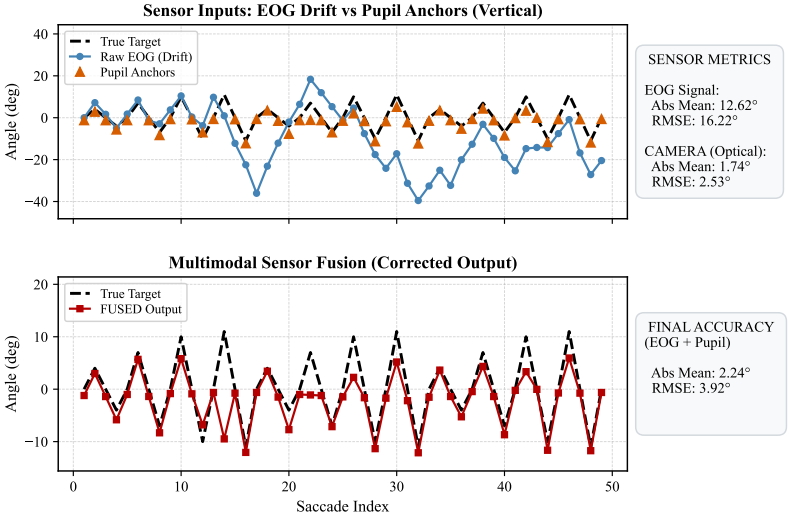


Figure 4.8: Vertical Estimation Performance

4.3.2.3 Temporal Stability

The efficacy of the drift correction is further demonstrated by analysing system accuracy over time. The estimation sequence was segmented into three distinct repetition cycles to measure cumulative error (Figure 4.9). The raw EOG channel demonstrates significant degradation, with the RMSE increasing progressively across the three cycles ($10.5° \rightarrow 20.3° \rightarrow 19.3°$) as integration drift accumulated. Conversely, the fused output maintained a highly stable error profile, recording RMSE values of $5.2°$, $3.4°$,

and 2.3° across the respective cycles. This confirms that the sensor fusion architecture effectively arrests continuous temporal drift on the vertical axis, maintaining functional tracking accuracy over extended operational periods.

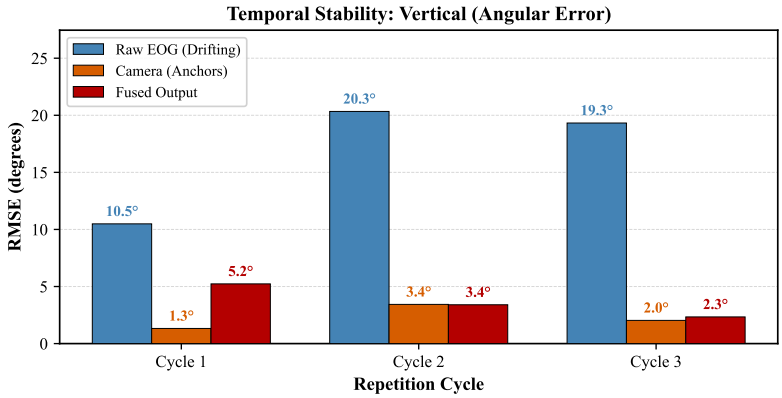


Figure 4.9: Cyclic RMSE Performance of Vertical Movements

4.3.3 2D Oblique Gaze Estimation

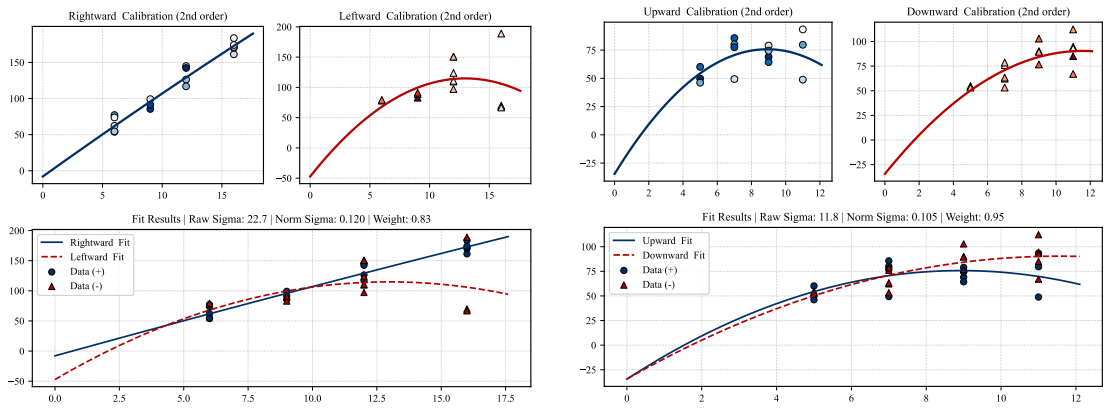
The quantitative validation of 2D oblique (diagonal) movements represents the most rigorous test of the system. Diagonal saccades require the simultaneous activation of multiple extraocular muscles, inducing complex mechanical shear on the skin-electrode interface and physiological cross-talk between the orthogonal acquisition channels.

4.3.3.1 Calibration Performance in 2D Space

To accurately map diagonal vectors, the horizontal and vertical channels were calibrated simultaneously using the 2nd-order Bilateral Polynomial Model. As shown in Figure 4.10a and Figure 4.10b, the calibration successfully linearized the dipole saturation across both axes. The horizontal channel regression yielded a Normalized Sigma of 0.120 (12.0% residual error), assigning an estimation weight of 0.83. The vertical channel regression yielded a Normalized Sigma of 0.105 (10.5% residual error), assigning an estimation weight of 0.95. These metrics indicate that despite the mechanical complexity of oblique movements, the bilateral model effectively captured the underlying geometric properties of the EOG signal.

4.3.3.2 Component-Level Sensor Fusion Accuracy

The efficacy of the sensor fusion algorithm during diagonal estimation is detailed in the temporal movement progression (Figure 4.11). Breaking the 2D movement down into orthogonal components reveals distinct channel behaviours. The fused horizontal component achieved a highly accurate Root Mean Square Error (RMSE) of 2.11°. Conversely, the fused vertical component exhibited a higher RMSE of 3.37°. As anticipated in the physiological characterization, the vertical channel inherently suffers from lower Signal-to-Noise Ratio (SNR) due to passive eyelid interference. However, the multimodal pipeline successfully utilized the optical pupil anchors to continuously bind the drifting EOG baseline across both independent axes, generating a stable composite output.



(a) Diagonal Calibration of the Horizontal Channel (b) Diagonal Calibration of the Vertical Channel

Figure 4.10: The Calibration of the two Diagonal Channels

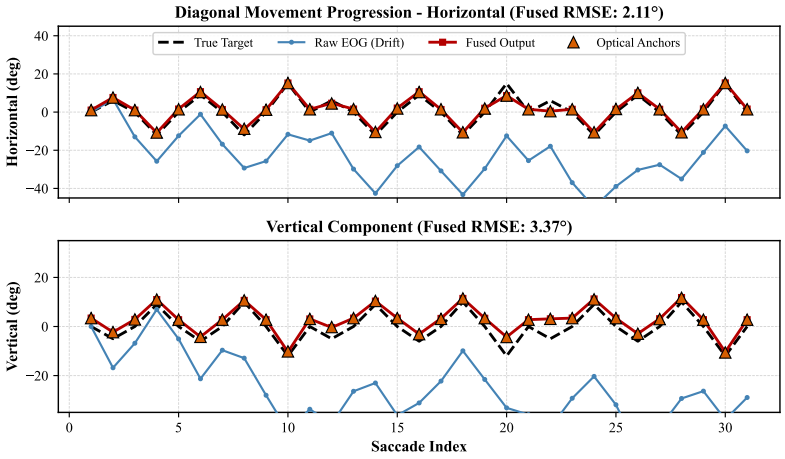


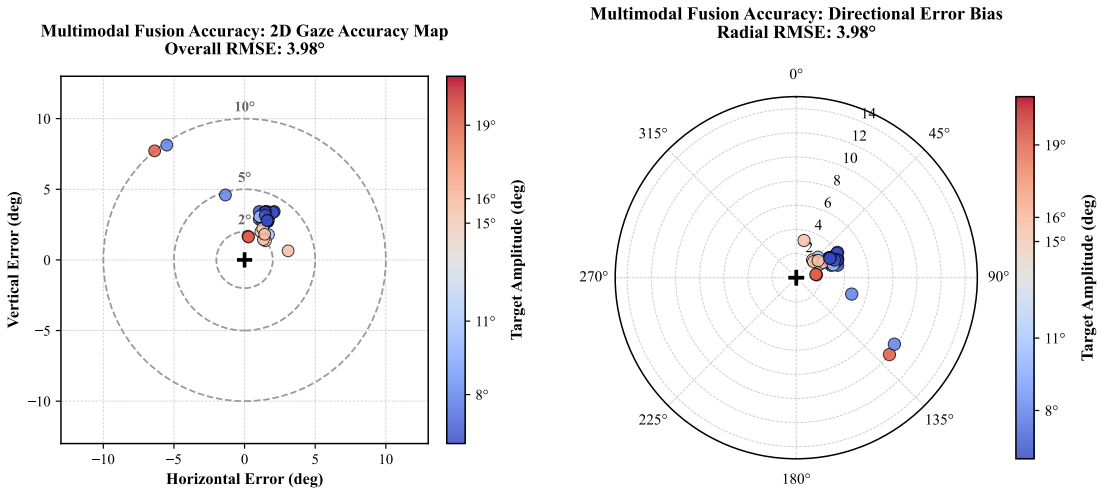
Figure 4.11: Oblique movements divided into two orthogonal components

4.3.3.3 2D Spatial Accuracy and Directional Bias

To evaluate the true performance of the interface, the horizontal and vertical components were combined to analyse absolute spatial error. Figure 4.12a presents a 2D Gaze Accuracy Map, illustrating the distribution of the fused estimation coordinates relative to the zero-error origin. The plot demonstrates tight spatial clustering, with the vast majority of estimation points falling within a 5° radius of the true target. This spatial accuracy is further contextualized in the Directional Error Bias plot (Figure 4.12b). The polar representation maps the absolute error magnitude against the angular direction of the saccade. The system achieved a total Radial RMSE of 3.98°. The distribution confirms an absence of severe directional bias; the optical drift correction maintained proportional accuracy across all four quadrants of the visual field, independent of the target amplitude.

4.3.3.4 Temporal Stability and Sensor Fusion Efficacy

The necessity of the optical sensor fusion architecture is proven conclusively by the cyclic stability analysis (Figure 4.13). During diagonal movements, the combined mechanical shear acting on the electrodes caused the uncorrected, AC-coupled EOG signal to fail completely. The Raw EOG drift compounded exponentially across the testing cycles, producing unusable spatial errors ($24.7^\circ \rightarrow 39.5^\circ \rightarrow 45.2^\circ$).



(a) Gaze Accuracy Map (Dartboard Plot)

(b) Directional Error Bias Plot (Polar Plot)

Figure 4.12: Gaze Accuracy Map and Direction Error Bias

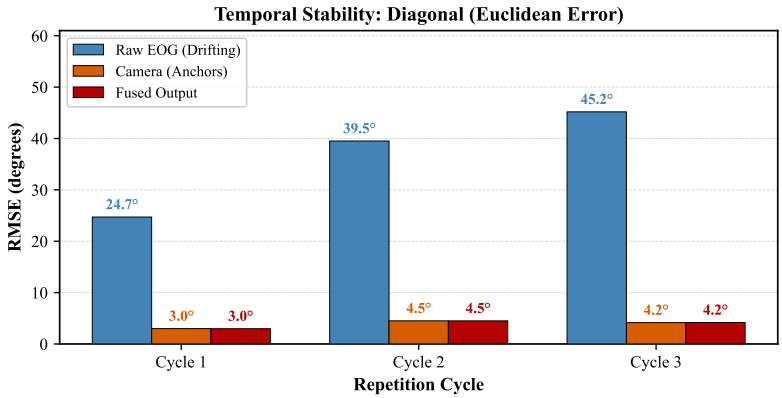


Figure 4.13: Cyclic RMSE Performance of the Oblique Movements

By integrating the continuous spatial ground truth of the pupil tracking module, the multimodal system completely neutralized this drift. The Fused Output maintained a stabilized, flat error profile across all three cycles ($3.0° \rightarrow 4.5° \rightarrow 4.2°$). This stark contrast unequivocally validates the core hypothesis of the study: while low-cost, AC-coupled EOG hardware alone cannot sustain 2D absolute interaction due to baseline instability, the integration of optical sensor fusion bridges the gap, yielding a stable, functional Eye-Computer Interface.

4.4 Phase II: Multi-Subject Validation

To evaluate the generalization of the sensor fusion architecture across varying facial morphologies and skin-electrode impedances, the system was tested on a group of five independent subjects. Figure 4.14 illustrates the fused Root Mean Square Error (RMSE) across horizontal, vertical, and 2D oblique tracking paradigms.

Group Performance and Consistency

The multi-subject validation confirms the robustness of the multimodal pipeline. The group achieved an

Table 4.1: Summary of Single-Subject Estimation Accuracy (RMSE)

Movement Axis	Raw EOG RMSE	Optical Anchors RMSE	Fused Output RMSE
Horizontal	$15.61°$	$3.67°$	$\mathbf{3.64}°$
Vertical	$16.22°$	$2.53°$	$\mathbf{3.92}°$
Diagonal (2D Radial)	$> 45.0°$ (Unstable)	-	$\mathbf{3.98}°$

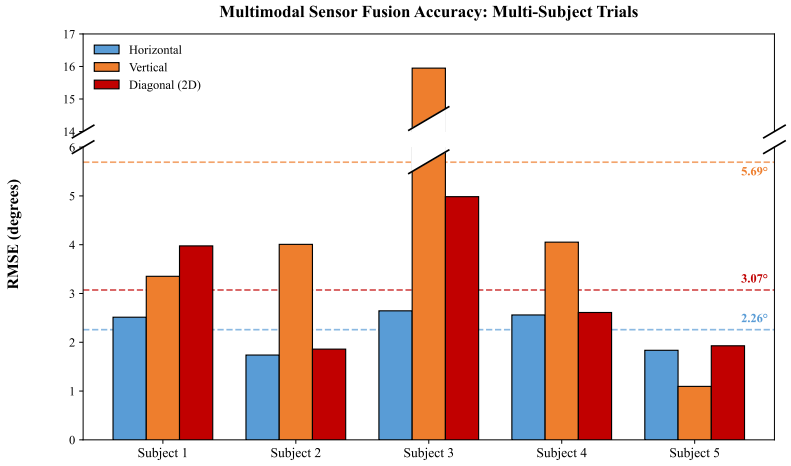


Figure 4.14: Multi-Subject RMSE

average Horizontal RMSE of $2.26°$ ($\sigma \approx 0.42°$) and an average Diagonal RMSE of $3.07°$ ($\sigma \approx 1.39°$). Comparing these metrics to the initial single-subject baseline (Horizontal: $3.64°$, Diagonal: $3.98°$), the average system performance actually improved. The low standard deviation across the horizontal axis confirms that the optical anchor extraction and dynamic integration reset mechanisms are highly agnostic to inter-subject variability.

Vertical Axis Variance and Outlier Analysis

The average Vertical RMSE for the group was recorded at $5.69°$. However, this elevated average is heavily skewed by a single severe outlier. While Subjects 1, 2, 4, and 5 maintained vertical errors between $1.1°$ and $4.1°$, Subject 3 experienced an isolated tracking degradation, yielding a vertical RMSE of approximately $16.0°$. This anomaly highlights the physiological vulnerability of vertical EOG: specific eyelid kinematics or poor supra-orbital electrode coupling on certain subjects can degrade the analog Signal-to-Noise Ratio (SNR) beyond the corrective capacity of the optical pipeline. Excluding this single outlier, the adjusted vertical average for the remaining subjects drops to a highly functional $3.15°$.

Theoretical Lower Bound (Ideal Case Scenario)

It is critical to note that during isolated trials within the group testing, the multimodal system sporadically achieved tracking errors as low as $0.8°$. While this exceptional accuracy was not constantly sustained throughout extended saccadic sequences, it establishes the theoretical "ideal case" resolution limit of the hardware when electrode impedance, lighting conditions, and algorithmic synchronization align perfectly.

4.4.1 Spatial Error Translation and HCI Implications

To contextualize the final group tracking accuracy (2.26° Horizontal, 3.07° Diagonal, and $\sim 3.15^\circ$ adjusted Vertical) for practical interface design, the angular error must be translated into absolute spatial screen coordinates (pixels). Assuming a standard desktop HCI environment with a user-to-screen distance (D) of 600 mm and a standard 24-inch 1080p monitor (pixel pitch ≈ 0.276 mm/pixel), the spatial error (E_{mm}) is calculated geometrically:

$$E_{mm} = D \cdot \tan(\theta_{RMSE}) \tag{4.2}$$

Applying the cohort averages:

- **Horizontal Error (2.26°)**: ≈ 23.7 mm, which translates to an on-screen radial error of ≈ 86 **pixels**.
- **Vertical/Diagonal Error ($\approx 3.10^\circ$)**: ≈ 32.5 mm, which translates to an on-screen radial error of ≈ 118 **pixels**.

From an engineering perspective, these metrics define the minimum viable target resolution for a Graphic User Interface (GUI) driven by this multimodal system. To guarantee reliable gaze-based selection without continuous miss-clicks, interactive elements (such as icons or virtual keys) must be designed with a bounding box of at least 120×120 pixels, spaced appropriately to accommodate the system's residual 2D spatial variance.

Chapter 5

Conclusions

5.1 Summary of Contributions

This study investigated the absolute performance limits of consumer-grade biomedical hardware for Electro-Oculography (EOG) and culminated in the development of a low-cost, multimodal sensor fusion eye-tracking system. The primary objective was to characterize and optimize the AD8232 amplifier to extract the highest possible accuracy for EOG. Subsequently, to overcome the inherent drift of AC-coupled amplification, the system architecture was expanded to integrate optical pupil tracking, providing a continuous spatial ground truth for real-time baseline correction.

5.1.1 Characterization and Isolation of the AD8232

A central contribution of this work is identifying the operational boundaries of the AD8232 integrated circuit when repurposed from heart-rate monitoring to EOG. Modifications to the active filter topology extended the lower cut-off frequency to 0.05 Hz, permitting the acquisition of quasi-static fixation potentials and the higher cut-off frequency to 10 Hz. Crucially, the study demonstrated that separating the system from USB tethering and utilizing an isolated DC battery power supply eliminated 50 Hz mains hum and ground loop interference. This electrical isolation dramatically reduced the baseline noise floor, maximizing the Signal-to-Noise Ratio (SNR) achievable with this specific IC.

5.1.2 The Bilateral Polynomial Model

To translate the raw electro-oculographic potentials into gaze coordinates, a robust calibration model was formulated utilizing a center-extremity target pattern. Experimental characterization revealed that the EOG signal is inherently non-linear at large eccentricities and anatomically asymmetric.

- **Saturation Handling:** A simple linear model was proven insufficient, producing undershoot errors at the periphery ($\pm 30^{\circ}$). The implementation of a 2nd-Order Polynomial successfully linearized this saturation, capturing the geometric properties of the corneo-retinal dipole.
- **Anatomical Asymmetry:** The analysis confirmed that the electrical sensitivity of the eye is not isotropic. By splitting the domain into Positive (Right/Up) and Negative (Left/Down) coefficients, the Bilateral Model accommodated the structural differences in the orbital bone and electrode placement, ensuring high fidelity across the full visual field.

5.1.3 Multimodal Sensor Fusion for Drift Correction

The most significant architectural contribution of this study is the integration of wireless pupil tracking to dynamically correct EOG signal drift. AC-coupled hardware inherently suffers from a stochastic "Random Walk" baseline, making absolute gaze position estimation highly unstable over time. By incorporating an ESP32-S3 microcontroller and a camera module, optical pupil coordinates were transmitted alongside batched EOG data via User Datagram Protocol (UDP). This sensor fusion pipeline utilizes the optical data to periodically realign the integration baseline of the EOG signal, effectively neutralizing the drift and yielding a corrected 2D spatial accuracy of $3.07°$.

5.2 Discussion of Results

The experimental validation detailed in Chapter 4 provides critical insights into the feasibility, stability, and fundamental limitations of low-cost Electro-Oculography. Beyond the raw mathematical error metrics, interpreting these results across both single-subject and multi-subject group reveals the practical utility of the proposed sensor fusion system.

5.2.1 Single-Subject System Validation

The quantitative evaluation of the single-subject dataset provided a rigorous baseline for assessing the hardware's capabilities and the efficacy of the multimodal sensor fusion architecture. The results yield several critical engineering insights regarding the strengths and inherent weaknesses of low-cost, AC-coupled EOG.

Efficacy of the Sensor Fusion Architecture
The most significant finding from the validation phase is the absolute necessity and success of the optical sensor fusion pipeline. Across all testing paradigms, the raw AC-coupled EOG signal exhibited integration drift, culminating in failing during 2D oblique movements where the uncorrected error exceeded $45°$. By utilizing the discrete optical pupil coordinates to dynamically reset the integration baseline, the system effectively neutralized this non-stationary behaviour. The multimodal architecture reduced the fused estimation errors to functional levels, achieving an RMSE of $3.64°$ for horizontal tracking, $3.92°$ for vertical tracking, and a radial RMSE of $3.98°$ during complex diagonal saccades. The cyclic stability analysis conclusively demonstrated that this approach arrests temporal drift, maintaining a flat error profile over extended operational periods.

Validation of the Bilateral Polynomial Model
The calibration data validated the hypothesis that a simple linear regression is insufficient for mapping the corneo-retinal dipole at high eccentricities ($\geq 15°$). The implementation of the 2nd-Order Bilateral Polynomial Model successfully accommodated the anatomical asymmetry of the orbital cavity and the non-linear saturation of the dipole. This spatial modelling was crucial for preventing overshoot and undershoot errors at the periphery of the visual field, optimizing the raw signal before it entered the sensor fusion pipeline.

Axis-Specific Limitations and Mechanical Artifacts
Despite the success of the drift correction, the validation highlighted fundamental physical limitations within the EOG acquisition:

- **The Vertical SNR Deficit:** The vertical channel consistently proved to be the weakest link in the

system. The passive descent of the upper eyelid over the cornea acts as a dynamic dielectric shield, drastically altering the electrical impedance and generating higher baseline variance compared to the horizontal axis.

- **Mechanical Shear and Electrode Coupling:** The diagonal validation exposed a critical mechanical vulnerability. Initial tests with rigid gold cup electrodes mounted directly to the thermoplastic mask resulted in massive signal degradation due to mechanical shear lifting the electrodes during complex facial contortions. The system only achieved stability after pivoting to skin-adhered Ag/AgCl electrodes. This indicates that any future wearable housing must mechanically isolate the sensors from the rigid structural frame to preserve the impedance bridge.

Ultimately, the single-subject analysis confirms that while low-cost AC-coupled hardware cannot independently sustain absolute 2D gaze tracking—particularly during extreme vertical or diagonal excursions — it does maintain viable spatial accuracy for purely horizontal or small-amplitude vertical movements. This inherent mechanical and physiological disparity explains the prevailing trend in current literature, where EOG applications are overwhelmingly restricted to horizontal 1D control paradigms. However, the strategic integration of an optical ground-truth overcomes these fundamental bottlenecks, successfully extending the capabilities of low-cost hardware to yield a highly stable 2D interface.

5.2.2 Multi-Subject Generalization and Hardware Limits

To evaluate the generalization of the sensor fusion architecture across varying facial morphologies, inter-pupillary distances, and skin-electrode impedances, the system was tested on a group of five independent subjects.

Group Performance and Consistency

The multi-subject validation confirms the robustness of the multimodal pipeline. The group achieved an average Horizontal RMSE of 2.26° ($\sigma \approx 0.42^\circ$) and an average Diagonal RMSE of 3.07° ($\sigma \approx 1.39^\circ$). The low standard deviation across the horizontal axis confirms that the optical anchor extraction and dynamic integration reset mechanisms are highly agnostic to inter-subject physiological variability.

Vertical Axis Variance and Outlier Analysis

The average Vertical RMSE for the group was recorded at 5.69°. However, this elevated average was heavily skewed by a single severe outlier. While four of the subjects maintained functional vertical errors between 1.1° and 4.1°, one subject experienced an isolated tracking degradation, yielding a vertical RMSE of approximately 16.0°. This anomaly further reinforces the physiological vulnerability of vertical EOG identified in the single-subject tests: specific eyelid kinematics or poor supra-orbital electrode coupling on certain individuals can degrade the analog Signal-to-Noise Ratio (SNR) beyond the corrective capacity of the optical pipeline. Excluding this single outlier, the adjusted vertical average for the remaining subjects drops to a highly functional 3.15°.

Theoretical Lower Bound (Ideal Case Scenario)

It is critical to note that during isolated trials within the single and group testing, the multimodal system sporadically achieved tracking errors as low as 0.75°. While this exceptional accuracy was not constantly sustained throughout extended saccadic sequences, it establishes the theoretical "ideal case" resolution limit of the hardware when electrode impedance, lighting conditions, and algorithmic synchronization align perfectly.

5.2.3 Implications for Human-Computer Interaction (HCI)

To contextualize the final group tracking accuracy (2.26° Horizontal, 3.07° Diagonal, and $\sim 3.15^\circ$ adjusted Vertical) for practical interface design, the angular error must be translated into absolute spatial screen coordinates. Assuming a standard desktop HCI environment with a user-to-screen distance (D) of 600 mm and a standard 24-inch 1080p monitor (pixel pitch ≈ 0.276 mm/pixel), the spatial error (E_{mm}) is calculated geometrically ($E_{mm} = D \cdot \tan(\theta_{RMSE})$). Applying the group averages, the Horizontal error of 2.26° translates to an on-screen radial error of approximately 86 pixels, while the Diagonal/Vertical error of $\approx 3.10^\circ$ translates to a radial error of approximately 118 pixels. From an engineering perspective, these metrics strictly define the minimum viable target resolution for a Graphic User Interface (GUI) driven by this multimodal system. To guarantee reliable gaze-based selection without continuous user frustration or miss-clicks, interactive elements must be designed with a bounding box of at least 120×120 pixels. This confirms that while the system lacks the pinpoint precision required for high-density text reading or cursor-level drawing, it is highly capable of driving grid-based augmentative and alternative communication (AAC) boards or smart-home control interfaces.

5.3 Limitations of the System

While the proposed system successfully meets the cost and accuracy requirements for a proof-of-concept interface, the experimental validation highlights several engineering bottlenecks inherent to the use of generic, off-the-shelf components. The current system accurately defines the limits of the hardware, but these constraints impact setup repeatability and data transmission.

5.3.1 Hardware-Specific Calibration and Tuning

The use of a generic camera module (ESP32-S3 Sense) lacks the dynamic range and auto-exposure capabilities of dedicated infrared eye-tracking sensors. Consequently, the camera cannot automatically adjust to variable ambient lighting or optimize pupil contrast. This necessitates manual intervention during every session; parameters within the Python computer vision back-end must be tuned for each subject to ensure reliable optical tracking.

5.3.2 Electrode Mechanics and Mask Rigidity

While reusable gold cup electrodes provide the superior raw voltage required to maximize the AD8232's performance, they conflict mechanically with the rigid thermoplastic mask. The mask acts as a physical lever against the face. Minor facial muscle twitches push against the rigid frame, slightly altering the Z-axis pressure applied to the gold cups. This mechanical shear introduces baseline wandering and increases the standard deviation of the acquired signal compared to electrodes adhered directly and exclusively to the skin.

5.3.3 Bandwidth and Antenna Constraints

The multimodal system is heavily constrained by wireless transmission bottlenecks. The ESP32 relies on a primitive, built-in Wi-Fi antenna that is highly susceptible to RF interference. During experimental setups utilizing metallic chin rests, the resulting Faraday cage effect significantly attenuated signal strength. Furthermore, transmitting full VGA JPEG frames alongside high-frequency EOG data consumes substantial bandwidth, rendering the camera frame rate entirely dependent on local Wi-Fi network speed and congestion. Under optimal conditions, the frame rate capped at 18-22 FPS, which is pretty good for a

first prototype but for more specific or more extreme uses it limits the temporal resolution of the optical drift correction.

5.4 Future Works

Based on the limitations identified during hardware characterization and sensor fusion, future iterations of this architecture should prioritize purpose-built circuitry, flexible ergonomics, and optimized optical data pipelines.

5.4.1 Custom EOG Analog Front-End

While the AD8232 proved functional, it remains a repurposed ECG amplifier. Future hardware development should focus on designing a custom Analog Front-End circuit tailored specifically for the impedance characteristics and frequency spectrum of the corneo-retinal dipole, potentially incorporating DC-coupled instrumentation amplifiers to eliminate drift at the hardware level.

5.4.2 Flexible Wearable Architecture

To resolve the conflict between rigid mechanical support and facial compliance, future iterations should abandon rigid thermoplastics. The ideal physical architecture would utilize a flexible, breathable fabric (e.g., a modified neoprene mask) integrated with flexible textile or printed conductive electrodes. This approach would allow the sensors to conform naturally to the user's facial geometry, maintaining consistent contact pressure while eliminating mechanical levering.

5.4.3 Advanced Optical Hardware and Tracking Algorithms

To overcome the current Wi-Fi bandwidth limitations, future systems should integrate image sensors capable of hardware-level Region of Interest (ROI) cropping. Similar to high-end clinical systems, capturing and transmitting only a tightly cropped window around the eye drastically reduces the data payload, permitting significantly higher frame rates without sacrificing image quality. Furthermore, the computer vision back-end should be upgraded with more robust, machine-learning-driven pupil detection algorithms capable of maintaining tracking integrity under severe occlusion, extreme gaze angles, or sub-optimal lighting conditions.

5.5 Final Conclusion

This research validates the feasibility of high-fidelity, absolute 2D gaze tracking using low-cost, AC-coupled hardware through the strategic implementation of multimodal sensor fusion. By characterizing the physiological distortions of the Electro-Oculogram —specifically non-linear dipole saturation and severe integration drift — and compensating for them using a 2nd-Order Bilateral Polynomial Model paired with discrete optical pupil anchors, the system successfully bridges the performance gap between consumer electronics and clinical instrumentation. The quantitative multi-subject validation demonstrates that the hybrid architecture effectively neutralizes temporal baseline wandering, achieving a highly stable 2D spatial accuracy between 2.26° and 3.15° across diverse facial morphologies. While physical limitations regarding vertical eyelid interference and mechanical electrode shear remain, the prototype conclusively proves that accessible, continuous Eye-Computer Interfaces can be engineered without reliance on

specialized, expensive DC amplification. Ultimately, the synergistic fusion of high-speed electrical oculomotor dynamics with absolute optical positioning provides a robust, low-latency foundation for the next generation of accessible assistive technologies.

Acknowledgements

As this journey comes to a close, I would like to sincerely thank Prof. Agostino Gibaldi and Prof. Giovanni Gibertoni. Thank you for your invaluable help during my thesis work, for your availability, and for the great overall experience at OptoLab. I also extend my deepest gratitude to Prof. Luigi Rovati, my mentor during the TACC program and other courses, for his continuous support throughout these years.

A very special thank you goes to my family. Thank you for supporting me, allowing me to grow, make mistakes, and ultimately find my way. Thank you for your unconditional love and kindness, for instilling in me a relentless curiosity and determination, and for always caring so deeply about my life and my future. I would not be here without you. Of course, I have to mention Marsi — I've always said he's my gift.

I want to thank those who have supported me throughout my academic and personal journey, from one place to the other, from one country to the next.

I am deeply thankful to Yousra for being there — consistently, genuinely, and in the ways that mattered most; more than words can fully express.

Faleminderit!

Good job, Athanas.

Bibliography

[1] A. P. Morris, "Neuroscience: Tiny eye movements link vision and attention," *Current Biology*, vol. 25, pp. R769–R771, 2015. [Online]. Available: https://doi.org/10.1016/j.cub.2015.07.011

[2] IMotions, "The latest in eye tracker technology," https://imotions.com/products/hardware/eye-tracking/, 2026.

[3] J. Stahl, *Eye Movement Recording*. Elsevier, 2014, pp. 245–247. [Online]. Available: https://doi.org/10.1016/b978-0-12-385157-4.00127-5

[4] Tobii, "Tobii pro spectrum," https://www.tobii.com/products/eye-trackers/screen-based/tobii-pro-spectrum, 2026.

[5] EyeWiki, "Electrooculogram," https://eyewiki.org/Electrooculogram#Principle_of_EOG, 2025.

[6] ADInstruments, "Bio amps," https://www.adinstruments.com/products/bio-amps?srsltid=AfmBOoq2qaSoxQLnig3pMTewfUdmHFjVuNpBq6XaO3Q839Z6f3HQB4OA, 2026.

[7] Analog Devices, "Ad8232 datasheet," https://www.analog.com/media/en/technical-documentation/data-sheets/ad8232.pdf, 2012.

[8] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*. United States: Oxford University Press, 2011.

[9] R. J. Leigh and D. S. Zee, *The neurology of eye movements - 4th Ed.* United States: Oxford University Press, 2006.

[10] J. G. Webster and A. J. Nimunkar, *Medical instrumentation : application and design*. United States: Wiley, 2020.

[11] J. J. A. M. Junior, D. P. Campos, L. C. de Andrade Villela De Biassio, P. C. Passos, P. B. Júnior, A. E. Lazzaretti, and E. Krueger, "Ad8232 to biopotentials sensors: Open source project and benchmark," *Electronics*, vol. 12, p. 833, 2023. [Online]. Available: https://doi.org/10.3390/electronics12040833

[12] A. López, F. Ferrero, M. Valledor, J. Campo, and O. Postolache, "A study on electrode placement in eog systems for medical applications," in *2016 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2016, pp. 1–5.

[13] N. Barbara, T. A. Camilleri, and K. P. Camilleri, "Eog-based gaze angle estimation using a battery model of the eye," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 6918–6921. [Online]. Available: https://doi.org/10.1109/embc.2019.8856323

[14] J. Nezvadovitz and H. Rao, "Using natural head movements to continually calibrate eog signals," *Journal of Eye Movement Research*, vol. 15, 2022. [Online]. Available: https://doi.org/10.16910/jemr.15.5.6

[15] H. Manabe, M. Fukumoto, and T. Yagi, "Direct gaze estimation based on nonlinearity of eog," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 1553–1562, 2015. [Online]. Available: https://doi.org/10.1109/tbme.2015.2394409

[16] LearnOpenCV, "Otsu's thresholding with opencv," https://learnopencv.com/otsu-thresholding-with-opencv/, 2020.

[17] R. Engbert and R. Kliegl, "Microsaccades uncover the orientation of covert attention," *Vision Research*, vol. 43, pp. 1035–1045, 2003. [Online]. Available: https://doi.org/10.1016/s0042-6989(03)00084-1