

Article

A Single-Point Adaptive Gaze-to-Cursor Correction Pipeline for Low-Burden Dwell-Based Eye-Tracking Interfaces: Design and Human Evaluation

Paweł Krowicki ¹, Fryderyk Gajdzik ¹, Roman Olejniczak ², Zofia Tomala ³ and Grzegorz Żurek ^{3,*}

¹ Department of Mechanical Engineering, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; pawel.krowicki@pwr.edu.pl (P.K.); fryderykgajdzik@gmail.com (F.G.)

² Neurorehabilitation Clinic, 54-530 Wrocław, Poland; roman.olejniczak@gmail.com

³ Department of Biological Principles of Physical Activity, Wrocław University of Health and Sport Sciences, 51-612 Wrocław, Poland; zosia.tomala@gmail.com

* Correspondence: grzegorz.zurek@awf.wroc.pl; Tel.: +48-71-347-3366

Abstract

Background. Eye-controlled interfaces can support communication and selection tasks when physical interaction is difficult, but conventional multi-point calibration may be burdensome for users who fatigue quickly or have severe motor limitations. This study operationally evaluates a low-burden single-point adaptive gaze-to-cursor correction pipeline for a dwell-based selection interface. **Methods.** The pipeline combines central single-point offset estimation, robust median/median absolute deviation (MAD) sample handling, recorded-stream jump rejection, exponential moving average filtering, dwell-based selection, and post-selection offset adaptation. It was evaluated using application-level gaze-to-cursor coordinates from 43 participants across 129 sessions. The primary endpoint was target-proximity Root Mean Square (RMS) during automatically logged dwell-active selection periods. Because these periods were defined by the pipeline's online state, the endpoint was algorithm-conditioned. Whole-quiz RMS was analyzed as a secondary full-trajectory descriptor rather than as independent validation. **Results.** Averaged across three repeated measurements, target-proximity RMS during dwell-active periods decreased from 193.40 pixels (px) (approximately 3.15°) for the recorded uncorrected coordinates to 56.02 px (0.91°) for the complete pipeline output. The mean reduction was 137.38 px (95% confidence interval (CI), 116.59–158.17 px; Cohen's $d_z = 2.03$), and all 43 participants showed a reduction. **Conclusions.** The results support the practical feasibility of the complete correction pipeline for target localization in the evaluated dwell-based interface. They do not independently validate native eye-tracker accuracy or establish superiority over conventional multi-point calibration.



Academic Editor: Hongying Meng

Received: 24 June 2026

Revised: 15 July 2026

Accepted: 16 July 2026

Published: 19 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: eye tracking; single-point calibration; gaze correction; robust calibration; exponential moving average; target-proximity RMS; dwell selection

1. Introduction

Gaze-based interaction systems require calibration to map measured gaze-related observations to locations on a display. Conventional procedures commonly use several spatially distributed fixation targets to estimate this mapping before gaze estimates are used for interaction. Although such procedures can improve spatial accuracy, they also

require time and user cooperation, which may be limiting in practical human–computer interaction settings and assistive interfaces [1–3].

In assistive communication contexts, including settings involving post-stroke or neurologically impaired users with severe motor limitations, calibration burden is not only a technical issue but also a usability constraint. A procedure that requires many fixation points, long cooperation, or repeated physical adjustments may be impractical when the user fatigues quickly, has limited attention, or cannot easily restart an interaction.

This motivation is supported by prior work showing that eye-tracking or eye-controlled systems can be used to assess cognition, consciousness-related visual responses, and language functions in non-verbal or severely impaired patients with disorders of consciousness or brain injury [4–7]. In particular, Kujawa et al. reported language-function tasks performed with an eye-controlled alternative communication tool in six patients diagnosed with disorders of consciousness, supporting the relevance of gaze-based interaction as an assistive assessment and communication context [6]. These studies motivate low-burden gaze-to-cursor correction procedures, but they do not constitute clinical validation of the present experimental interface.

Reduced calibration procedures are therefore attractive when the interaction task, display geometry, and expected error structure permit a simpler correction. A single central point cannot estimate a full spatial transformation across the screen without additional assumptions, but it can provide an offset estimate when a substantial part of the observed error is translational. Such a calibration step can be combined with robust sample handling and temporal filtering to reduce jitter, abrupt jumps, and transient instability in recorded gaze-position coordinates [8–11].

The intended use of the present method is therefore narrower: an eye tracker may already provide a basic gaze-to-cursor stream, and the application adds a fast single-point correction immediately before a dwell-based task. This should be understood as task-specific recalibration or correction, not as a general replacement for full multi-point calibration.

Target-proximity root mean square (RMS) was used as an application-level evaluation metric rather than as a full validation of eye-tracker accuracy. The primary endpoint described the RMS distance to the active target center during dwell-active stable segments. These segments were defined automatically by the online algorithm when the filtered gaze-to-cursor point was inside an active target and dwell accumulation for that target was in progress. A secondary whole-quiz RMS endpoint evaluated the full recorded interaction, including transitions, reading, visual search, and other task-related behavior. Three tracker position configurations were tested to check whether the observed effect was consistent across repeated measurements.

Robust calibration sample handling can reduce the influence of transient outliers during calibration. Median-based estimation and median absolute deviation (MAD) rejection are well suited to this purpose because they are less sensitive to extreme samples than mean-based estimates [12,13].

Exponential moving average filtering provides a simple recursive smoothing mechanism that can reduce frame-to-frame instability while retaining online operation [10,11].

Dwell-based target selection uses temporal stability as an interaction criterion: a target is selected only after the estimated gaze remains within its region for a configured duration. This can reduce accidental selections in gaze-controlled interfaces, although it introduces a trade-off between responsiveness and selection stability [14].

Accordingly, this study aimed to develop and operationally evaluate a low-burden single-point adaptive gaze-to-cursor correction pipeline for a specific dwell-based selection interface. We assessed whether the complete pipeline reduced application-level target-

proximity RMS during dwell-active selection periods and whether the direction of change was consistent across three repeated acquisition configurations. The study was not designed to validate native eye-tracker accuracy, isolate the effects of individual pipeline components, or compare the method with conventional multi-point calibration.

2. Materials and Methods

2.1. Participants

A total of 129 sessions were analyzed, obtained from 43 participants who underwent three repeated measurements. Each participant contributed one session to each of the three controlled acquisition configurations (Configuration 1, 2, and 3), with every session consisting of eight target selection tasks. This repeated-measures structure enabled a robust participant-level evaluation of the target-proximity reduction achieved by the proposed pipeline.

Data collection took place between December 2024 and March 2025 across two university sites. Eligibility criteria were: (1) age of at least 18 years, (2) provision of informed consent, (3) no severe visual impairment, and (4) at least one eye that could be successfully tracked during calibration.

Candidates were excluded if they: (1) had a refractive error exceeding ± 3 D, or (2) began pharmacological treatment during the observation period that could affect visual function.

This research was conducted in accordance with the 1975 Declaration of Helsinki and its later amendments, ensuring strict adherence to ethical standards for human participation. The study protocol was formally reviewed and approved by the Senate Committee on Research Ethics at the Wrocław University of Health and Sport Sciences, Poland (approval no. 5/2021).

2.2. Device Used in the Study

Measurements were performed on an HP OMEN 17-cb1004nw laptop (HP Inc., Palo Alto, CA, USA) equipped with Tobii Eye Tracker 5 (Tobii AB, Stockholm, Sweden). The laptop has a nominal 17.3-inch in-plane switching (IPS) display and a nominal resolution of 1920 by 1080 pixels. Participants used the same chair and desk, and the setup was kept consistent across sessions. The approximate face/eye distance from the screen was 70 cm. The vertical distance from the seat surface to the laptop was 29.5 cm.

The recorded sample files show that the coordinate system used by the experimental application was 1920 by 1061 pixels (px) in all 129 analyzed sessions. This matched the recorded runtime application window size. It differs from the nominal 1920 by 1080 px panel resolution and from 13 session metadata records that listed a 1080 px screen height. Therefore, RMS calculations used the sample-level application coordinate system of 1920 by 1061 px. Approximate angular conversions used the physical pixel pitch derived from the nominal 17.3-inch, 1920 by 1080 display geometry.

The three acquisition configurations were defined by controlled positioning of the eye-tracker assembly relative to the participant's face. A cross laser mounted on the eye-tracker setup was used to reproduce the device position across participants and sessions. In Configuration 1, the laser cross was aimed between the participant's eyebrows along the nose axis. In Configuration 2, the laser cross was aimed at the participant's nose. In Configuration 3, the laser cross was aimed at the participant's chin. These configurations provided repeated controlled measurements and a supporting consistency check; they were not used as the primary experimental factor.

2.3. Design Objective and System Assumptions

The proposed system was designed as a lightweight correction pipeline for dwell-based gaze-to-cursor interaction. Its purpose was not to estimate a full spatial calibration model for the display, but to provide a fast correction mechanism for a visual target selection interface.

The central assumption was that a single central calibration point can compensate primarily for a translational offset in the recorded gaze-to-cursor coordinates. The pipeline therefore combines robust central-point offset estimation with online coordinate correction, jump rejection, temporal stabilization, dwell-based selection logic, and limited post-selection adaptation. The human evaluation examined the complete implemented pipeline during interaction with the dwell-based interface.

The translational model was used because the evaluated interface was not intended to estimate a full screen-wide gaze mapping from one point. Instead, it addressed the narrower situation in which the operating-system gaze-to-cursor stream is already available and a substantial part of the residual error during a short interaction session can plausibly arise from a global displacement caused by posture, tracker placement, or session-specific gaze-to-cursor bias. Under this assumption, a central point can estimate only an offset, not scale, rotation, or spatially varying distortion. The model is therefore justified as a lightweight application-level correction for a dwell interface with large targets, and not as a general substitute for multi-point spatial calibration.

Formally, if residual error contains scale, rotation, or spatially varying components, a one-point estimate cannot identify them. The proposed pipeline can reduce only the component expressible as an approximately constant displacement, together with temporal instability handled by filtering and rejection.

2.4. Proposed Single-Point Adaptive Gaze-to-Cursor Correction Pipeline

2.4.1. Processing Pipeline

The pipeline receives normalized gaze-to-cursor coordinates and produces a final filtered coordinate used for target selection. Its purpose is to compensate for a systematic displacement estimated from one central calibration point, suppress transient jumps and jitter, and emit a selection only after the estimated gaze-to-cursor point remains stable on a target for the required dwell duration.

The evaluated pipeline comprises central-point calibration, robust median/MAD-based outlier rejection, translational offset correction, jump rejection in the recorded cursor stream, exponential moving average filtering, dwell-based target selection, and post-selection offset adaptation. The evaluation compares recorded uncorrected gaze-to-cursor coordinates with the final filtered output on the same time indices. The resulting estimates describe the complete applied pipeline rather than the isolated effect of any single internal operation. The overall processing flow is summarized in Figure 1.

The modules have distinct roles. Central calibration estimates translational bias, MAD rejection reduces calibration outliers, jump rejection removes abrupt cursor-coordinate discontinuities, Exponential Moving Average (EMA) filtering attenuates high-frequency jitter at the cost of latency, dwell logic converts temporal stability into selection, and post-selection adaptation uses successful selections as weak online supervision.

Normalized coordinates were constrained to the unit interval:

$$\text{clamp}_{01}(v) = \begin{cases} 0, & v < 0, \\ 1, & v > 1, \\ v, & \text{otherwise} \end{cases}$$

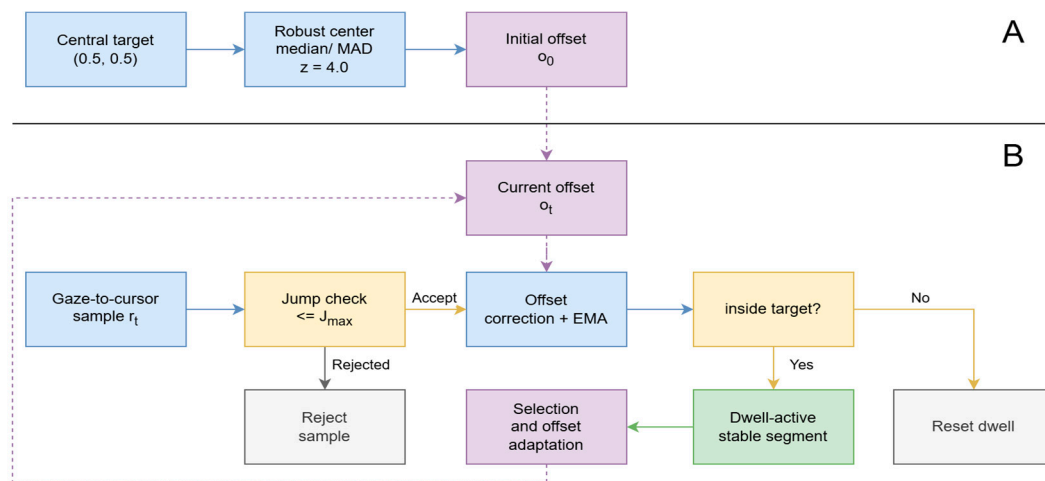


Figure 1. Schematic of the single-point adaptive gaze-to-cursor correction pipeline. Panel (A) shows the initial central-point calibration and offset estimation. Panel (B) shows the online correction, dwell-state handling, and offset-adaptation loop used during interaction. Blue boxes denote input or primary processing operations, yellow boxes denote decision points, the green box denotes the dwell-active stable state, purple boxes denote offset estimation, storage, and adaptation, and gray boxes denote sample rejection or dwell-state reset. Solid arrows indicate the main processing and state flow, whereas purple dashed arrows indicate offset initialization, use, and feedback adaptation.

2.4.2. Robust Single-Point Calibration

During calibration, samples were collected while the user was instructed to keep gaze on the central target. Let the calibration samples be

$$\{(x_i, y_i)\}_{i=1}^N.$$

The initial robust center estimates were the coordinate-wise medians:

$$m_{x,0} = \text{median}(x_i), \quad m_{y,0} = \text{median}(y_i).$$

Coordinate-wise median absolute deviations were computed as

$$\text{MAD}_x = \text{median}(|x_i - m_{x,0}|), \quad \text{MAD}_y = \text{median}(|y_i - m_{y,0}|).$$

The MAD values were rescaled by 1.4826:

$$\sigma_x = 1.4826 \text{MAD}_x, \quad \sigma_y = 1.4826 \text{MAD}_y,$$

with a small numerical lower bound if MAD was nearly zero.

A calibration sample was retained if

$$|x_i - m_{x,0}| \leq z\sigma_x \quad \text{and} \quad |y_i - m_{y,0}| \leq z\sigma_y,$$

where z is the MAD rejection multiplier. The multiplier was set to $z = 4.0$. If enough samples remained after rejection, the retained medians m_x and m_y were used; otherwise, the initial medians were used.

2.4.3. Offset-Based Gaze Correction

The calibration target was the center of the screen:

$$c_0 = (0.5, 0.5).$$

The translational offset was

$$\mathbf{o} = (o_x, o_y) = (0.5 - m_x, 0.5 - m_y).$$

For each recorded uncorrected gaze-to-cursor sample $\mathbf{r}_t = (x_t, y_t)$, the corrected coordinate was computed as

$$\mathbf{q}_t = (\text{clamp}_{01}(x_t + o_x), \text{clamp}_{01}(y_t + o_y))$$

2.4.4. Jump Rejection in the Recorded Cursor Stream

Let the display width and height be W and H pixels. Consecutive recorded cursor-coordinate samples were compared in pixel units:

$$\Delta_x = (x_t - x_{t-1})W, \quad \Delta_y = (y_t - y_{t-1})H.$$

A sample was accepted if

$$\Delta_x^2 + \Delta_y^2 \leq J_{\max}^2,$$

where $J_{\max}$ is the maximum allowed frame-to-frame jump in the recorded cursor stream. The threshold was set to $J_{\max} = 180$ px. This value was used as a predefined engineering threshold to reject abrupt cursor-coordinate displacements that were unlikely to represent useful dwell evidence, while still allowing intentional movement between large screen targets. It was retained as an implementation parameter rather than optimized post hoc for the statistical endpoint. If a recorded cursor-coordinate sample failed this test, the update returned without applying offset correction, EMA filtering, dwell-state updates, or target selection.

Because $J_{\max}$ is distance-based rather than velocity-normalized, its numeric value is tied to the realized sampling and update characteristics of this system. A more general implementation should express this criterion relative to elapsed time or report threshold sensitivity and rejection rates across devices.

Across the 129 analyzed sessions, the jump rejection rule rejected 5555 recorded cursor-coordinate samples, while 286,341 samples with evaluated validity were accepted. Rejected samples therefore accounted for 1.90% of samples with evaluated validity; across sessions, the rejection percentage had a mean of 1.94%, median of 1.89%, and range of 1.24% to 3.03%.

2.4.5. Exponential Moving Average Filtering

The final filtered coordinate was updated using an exponential moving average:

$$\begin{aligned} f_{x,t} &= \alpha q_{x,t} + (1 - \alpha)f_{x,t-1}, \\ f_{y,t} &= \alpha q_{y,t} + (1 - \alpha)f_{y,t-1}. \end{aligned}$$

The smoothing factor was set to $\alpha = 0.18$, as specified in the implementation configuration and retained in the session metadata. This value was selected before the analysis as a practical compromise between suppressing high-frequency cursor jitter and preserving sufficient responsiveness for dwell-based interaction. Smaller values would increase temporal lag, whereas larger values would pass more short-lived instability into the dwell logic.

The EMA introduces a first-order response lag; therefore, the chosen α affects target acquisition time, false dwell risk, and the interaction between smoothing and the 1500 ms dwell threshold. These temporal consequences were not isolated experimentally in the present study.

For $\alpha = 0.18$, the approximate number of update steps required for a first-order EMA step response to reach 90% of a new value is

$$n_{90} = \frac{\ln(0.1)}{\ln(1 - \alpha)} \approx 11.6$$

If the effective gaze-to-cursor update rate is 10 Hz, this corresponds to approximately 1.16 s. This should be interpreted as an approximate temporal characterization only, because the application logger ran at a higher and variable rate and could store repeated cursor coordinates between true gaze-to-cursor updates.

2.4.6. Dwell-Based Selection and Post-Selection Adaptation

Given active targets $T_j = (c_{x,j}, c_{y,j}, R_j)$, the nearest target was selected by minimizing the distance from the filtered point to each target center.

$$d_j^2 = (f_{x,t} - c_{x,j})^2 + (f_{y,t} - c_{y,j})^2.$$

The filtered point was inside target j when its pixel distance from the target center was less than or equal to the target hit radius R_j .

$$[(f_{x,t} - c_{x,j})W]^2 + [(f_{y,t} - c_{y,j})H]^2 \leq R_j^2.$$

If the filtered point was inside the nearest active target and no dwell target was currently assigned, the algorithm entered a dwell-active stable segment for that target and set the dwell timer to zero. If subsequent logged samples remained inside the same nearest target, the dwell timer accumulated. If the point left the target region, no active target was available, or another target became the dwell target, the segment was reset or restarted for the new target. The same target had to remain active for the dwell interval before the algorithm emitted a selection. When the dwell criterion was met, the algorithm emitted a selection and adapted the offset according to

$$\begin{aligned} o_x &\leftarrow o_x + k(c_{x,j} - f_{x,t}), \\ o_y &\leftarrow o_y + k(c_{y,j} - f_{y,t}). \end{aligned}$$

The post-selection adaptation rate was set to $k = 0.05$. After this adaptation, the dwell state was reset and a post-selection cooldown was started.

With $k = 0.05$, each completed dwell selection updates the translational offset by 5% of the residual target center error. This limits abrupt changes but assumes that the selection reflects the intended target; an incorrect selection may therefore reinforce an erroneous offset. The present evaluation did not include a formal convergence analysis or an adaptive-versus-non-adaptive ablation.

2.4.7. Implementation in the Dwell-Based Interface

The pipeline was implemented in a custom in-house gaze-based quiz and visual selection application developed specifically for this study. Each session contained eight tasks: four shape selection tasks, three digit selection tasks using a 3-by-3 grid, and one sorting task involving red and blue shapes and two baskets. The application first presented a central calibration phase and then started the quiz after calibration completion.

Screen targets were represented by an identifier, a normalized center (c_x, c_y) , and a hit radius in pixels. During the task, the pipeline estimated the filtered gaze-to-cursor point and compared it with active targets. A dwell-active stable segment began when the filtered gaze-to-cursor point entered the hit region of the nearest active target; while the

same target remained active and the filtered point stayed inside its hit region, the dwell timer accumulated. A target was selected only when this same-target dwell state persisted for the configured dwell interval. The dwell interval was 1500 ms and the post-selection cooldown was 300 ms.

The recorded input stream consisted of application-level gaze-to-cursor coordinates obtained through Tobii Eye Tracker 5, Tobii Experience (version 1.69.32600.0) running on Windows 11, version 25H2, and the operating-system cursor interface [15–17]. The data path was Tobii Eye Tracker 5 → Tobii Experience → operating-system cursor coordinates → application logger → offline analysis. The experimental application did not read native Tobii Pro Software Development Kit (SDK) or Tobii Pro Lab data streams; it logged the currently available cursor coordinates in its own processing loop, which ran at a higher and variable realized rate than the 10 Hz gaze-to-cursor trigger. Across the analyzed samples.csv files, consecutive stored records could therefore contain repeated cursor coordinates between true gaze-to-cursor updates. The analyzed recorded uncorrected baseline is an application-level cursor stream, not native timestamped research-grade Tobii gaze samples.

Consequently, the baseline represents the recorded uncorrected cursor coordinates available to the application after Tobii Experience and operating-system processing. Tobii Eye Tracker 5 is a consumer eye-tracking device used with Tobii Experience software; Tobii (version 1.69.32600.0) describes it as primarily intended for interactive use in games and not compatible with Tobii Pro SDK or Tobii Pro Lab for collection of Eye-Tracking Data [15,16]. The recorded uncorrected baseline may therefore include vendor- or operating-system-level processing such as smoothing, latency, throttling, coordinate quantization, and duplicated cursor coordinates.

2.4.8. Task Phases and Rationale for the Dwell-Active Endpoint

Each task was interpreted as comprising three phases: visual search/exploration, intentional dwell selection, and emitted selection with optional post-selection adaptation. During visual search, reading, and comparison of possible answers, the participant's intended gaze target is not known with sufficient certainty; target center RMS to a particular answer is therefore not a valid primary reference for that phase. During intentional dwell selection, the interface requires the user to keep the gaze-to-cursor point on one selectable target until the dwell criterion is met. Target proximity to that active target is therefore operationally meaningful for the behavior the interface is designed to support.

At the same time, dwell-active samples are not independent fixation-validation windows. They are entered when the online filtered coordinate satisfies the dwell-entry condition, so the endpoint is algorithm-conditioned. For this reason, the stable segment analysis is paired with whole-quiz RMS as a secondary descriptor of the full trajectory, including visual search, transitions, reading, and non-dwell periods.

The practical test sequence was organized to mirror the intended use of the interface. Each session began with a short central single-point correction phase in which the participant was instructed to keep the gaze-to-cursor point on the central target until the calibration progress bar was completed. The quiz then presented eight tasks: four shape selection tasks with four choices each, three digit selection tasks using a 3-by-3 grid, and one color-sorting task. In the shape and digit tasks, the participant selected an answer by maintaining the gaze-to-cursor point inside one target region until the 1500 ms dwell criterion was met. In the sorting task, items were first selected by dwell and then assigned to the corresponding color basket. These representative task screens are shown in Figure 2 and illustrate why the task naturally separates into visual search/exploration, intentional dwell selection, and emitted selection with optional post-selection adaptation.

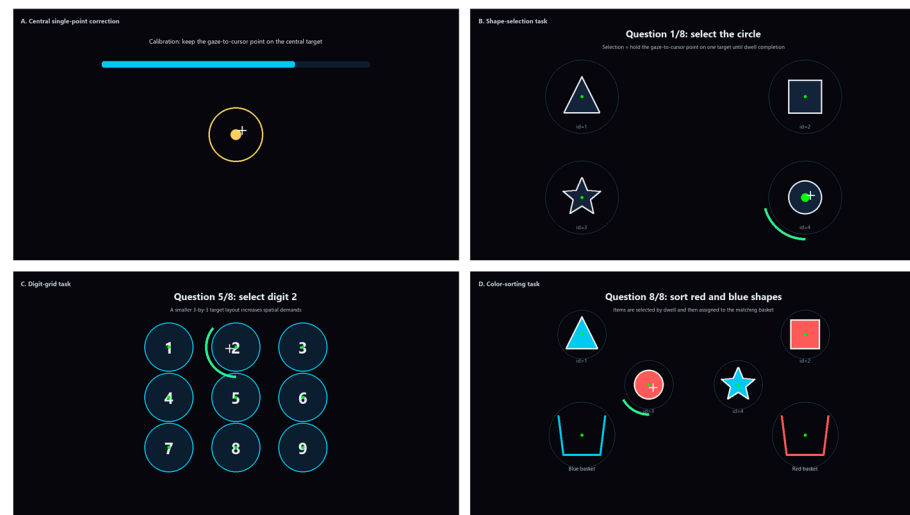


Figure 2. Representative screens of the experimental quiz interface. (A) Central single-point correction; (B) four-choice shape selection; (C) 3-by-3 digit selection; (D) color-sorting task. The panel is included to clarify the test procedure and task phases; it is not an additional outcome analysis.

2.4.9. Application-Level Data Quality Summary

The analyzed dataset comprised 43 participants and 129 included sessions. The application logs contained 1935 target selections during the dwell-based tasks. This number exceeded the 1032 administered tasks (129 sessions $\times$ 8 tasks) because the sorting task required multiple item and basket selections. For jump rejection in the recorded cursor-coordinate stream, 5555 samples were rejected and 286,341 samples with evaluated validity were accepted. Rejected samples therefore accounted for 1.90% of samples with evaluated validity. Across sessions, the rejection percentage had a mean of 1.94%, median of 1.89%, and range of 1.24% to 3.03%.

Because the application logged the cursor-based stream provided through Tobii Experience and the operating-system cursor interface rather than native Tobii Pro SDK or Tobii Pro Lab data, the logs did not contain native binocular or monocular tracking status, pupil validity, device confidence values, high-precision eye-tracker timestamps, or blink and tracking-loss markers [18,19]. The logs also do not provide a separate native record of failed acquisitions or missed dwell attempts outside the implemented application-level states. Missing or invalid data could therefore only be represented at the application level through the recorded cursor-coordinate stream, the jump rejection flag, dwell-state fields, and emitted target selections.

2.5. Evaluation Metrics and Statistical Analysis

The primary endpoint was application-level target-proximity RMS during dwell-active selection periods [20,21]. These periods were defined from the online state recorded by the application: a stable sample was any quiz sample with a non-empty `stable_target_id`, assigned when the filtered gaze-to-cursor point was inside the nearest active target's hit region and the dwell timer for that target was active. This endpoint evaluates the operational phase for which the interface was designed, but it is dwell-conditioned and algorithm-conditioned rather than an independent ground-truth measure of spatial gaze accuracy.

Operationally, a dwell-active stable segment began at the first logged sample after the filtered point entered a target's dwell region. At that point, the algorithm stored the target identifier in `stable_target_id` and set `stable_time_ms` = 0. The segment continued while the same target remained active. It ended when the filtered point left the target

region, when another target became active, or immediately before a dwell-based selection reset the dwell state. Stable samples therefore represent the application's dwell-entry and dwell-accumulation periods during the task, including all logged samples in that state. Samples outside this state were retained for the secondary whole-quiz endpoint.

This design was chosen because the intended target is ambiguous during search and reading, whereas during dwell selection, the active target is the user's likely intended selectable object. The resulting metric is therefore task-relevant but not independent of the pipeline that defines dwell entry.

For each participant, stable recorded-coordinate RMS and stable filtered RMS were averaged across the three repeated measurements. This participant-level aggregation respects the repeated-measures structure and avoids treating 129 sessions as independent participants.

No main paired inference treated the 129 sessions as independent observations; the participant was the unit of analysis for the primary statistical comparison.

Target-proximity improvement was defined as

$$\Delta_{\text{RMS}} = \text{RMS}_{\text{raw}} - \text{RMS}_{\text{filtered}},$$

so positive values indicate lower RMS after filtering. The primary paired comparison was performed using the Wilcoxon signed-rank test. A paired t -test was reported as a supplementary analysis. Cohen's d_z , the 95% confidence interval (CI) for the mean improvement, and the number of participants showing improvement were also reported. Offline data processing and statistical analyses were performed using Python 3.11.9 with NumPy 2.2.6, pandas 2.3.3, and SciPy 1.17.0. Figures and spreadsheet outputs were prepared using Matplotlib 3.10.3 and openpyxl 3.1.5, respectively.

For a given stage $s \in \{\text{raw}, \text{corrected}, \text{filtered}\}$ let $\mathbf{p}_i^{(s)} = (p_{x,i}^{(s)}, p_{y,i}^{(s)})$ denote the stage-specific screen coordinate in pixels for stable sample i . Let $\mathbf{c}_i = (c_{x,i}W, c_{y,i}H)$ denote the center of the same target that was active in the dwell state for that sample, with W and H the screen width and height. In the analyzed sample files, `stable_target_id` and `nearest_target_id` matched for stable samples, so $\mathbf{c}_i$ was obtained from the corresponding `nearest_target_cx` and `nearest_target_cy` fields. Recorded, corrected, and filtered RMS values were computed on the same set of stable sample time indices; only the coordinate stage $\mathbf{p}_i^{(s)}$ differed between comparisons. Session-level stable RMS is defined as

$$\text{RMS}_{s,\text{stable}} = \sqrt{\frac{1}{N_{\text{stable}}} \sum_{i=1}^{N_{\text{stable}}} \|\mathbf{p}_i^{(s)} - \mathbf{c}_i\|_2^2}.$$

The same formula applies to whole-quiz RMS when the sample set includes all quiz samples.

Approximate angular RMS was computed from the same sample-level distances rather than by converting only aggregated pixel RMS values. The physical display size was derived from the nominal 17.3-inch, 1920 by 1080 panel geometry, giving a display width of 382.99 mm, a display height of 215.43 mm, and a pixel pitch of approximately 0.199 mm/px. For each sample, the pixel displacement from the relevant target center was converted to millimeters and then to visual angle using

$$\theta_i = \arctan \left(\frac{\sqrt{[(p_{x,i}^{(s)} - c_{x,i}) 0.199]^2 + [(p_{y,i}^{(s)} - c_{y,i}) 0.199]^2}}{700} \right),$$

with θ_i expressed in degrees. Approximate angular RMS was then computed as the root mean square of the sample-level angular distances. Because the eye-to-screen distance was approximately 70 cm rather than measured continuously, all angular values are reported as approximate.

Stable segment RMS therefore represents application-level target proximity during the interface's own dwell-active periods. It should not be interpreted as general eye-tracker accuracy or independent fixation accuracy. Whole-quiz RMS was retained as a secondary descriptor of the full interaction trajectory, not as independent validation. It is less specific than the dwell-active endpoint because visual search and reading phases do not provide an unambiguous intended target, but it includes visual search, reading, transitions, target exploration, non-dwell periods, and dwell-active periods, thereby reducing the risk that interpretation relies only on dwell-conditioned samples. The three acquisition configurations were analyzed only as a supporting consistency check of whether recorded-to-filtered target-proximity reduction was observable across the controlled acquisition protocol.

3. Results

Table 1 summarizes the participant-level comparison of target-proximity RMS during dwell-active stable segments, and Table 2 reports the corresponding paired statistical evidence. These segments comprise all logged quiz samples in which the application had entered the dwell state for a target, as indicated by non-empty `stable_target_id`. Filtered coordinates had substantially lower RMS distance to the active target center than recorded uncorrected gaze-to-cursor coordinates. Mean stable RMS decreased from 193.40 px to 56.02 px, corresponding to a mean improvement of 137.38 px.

Table 1. Descriptive participant-level target-proximity root mean square (RMS) summary for dwell-active stable segments. Pixel values are reported in pixels (px); angular values are approximate degrees of visual angle.

Unit	Metric	Recorded Uncorrected RMS	Pipeline Output RMS	Improvement
px	Mean	193.40	56.02	137.38
px	SD	68.71	7.83	67.57
px	Median	181.70	56.53	124.10
px	95% CI for mean improvement	—	—	116.59–158.17
°	Mean	3.149	0.915	2.235
°	95% CI for mean improvement	—	—	1.896–2.573

Table 2. Paired participant-level statistical evidence for target-proximity RMS reduction during dwell-active stable segments.

Statistic	Result
Improved participants	43/43
Wilcoxon signed-rank test	$p < 10^{-8}$
Paired <i>t</i> -test	$t = 13.33, p < 10^{-8}$
Cohen's d_z	2.03

The improvement was observed in all 43 participants. The Wilcoxon signed-rank test was significant ($p < 10^{-8}$), and the paired effect size was large (Cohen's $d_z = 2.03$). Using

the approximate display geometry, participant-level stable RMS decreased from 3.149° for the recorded uncorrected coordinates to 0.915° for the complete pipeline output, with a mean reduction of 2.235° (95% CI, 1.896 – 2.573°). The paired participant-level changes are shown in Figure 3, and the corresponding reduction distribution is shown in Figure 4.

Here, “after pipeline application” refers to the complete implemented pipeline rather than the EMA filter alone. The comparison uses identical algorithm-conditioned dwell-active time indices. These statistics should not be interpreted as showing that failed acquisitions, missed dwell events, or unsuccessful target selections were eliminated, because the primary window excludes samples outside the dwell-active state.

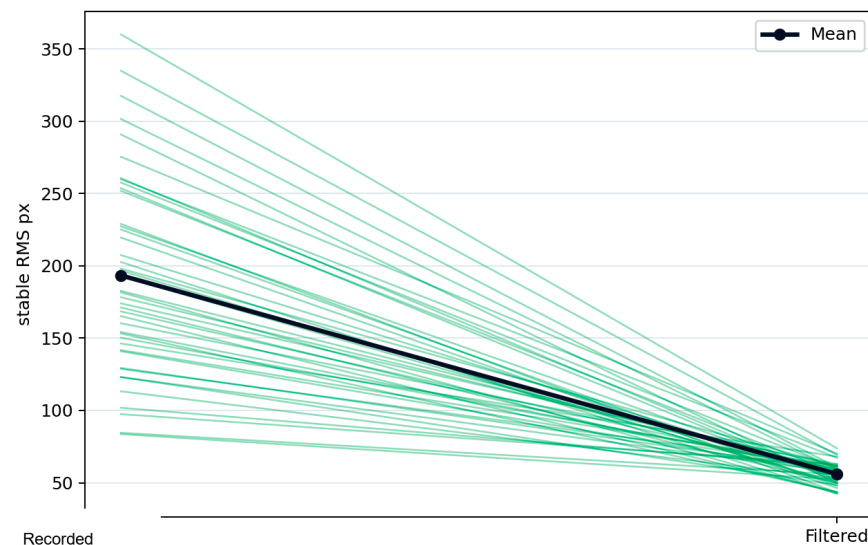


Figure 3. Participant-level target-proximity RMS during dwell-active periods for the recorded uncorrected coordinates and the complete pipeline output. Thin lines show individual participants, and the dark line shows the group mean.

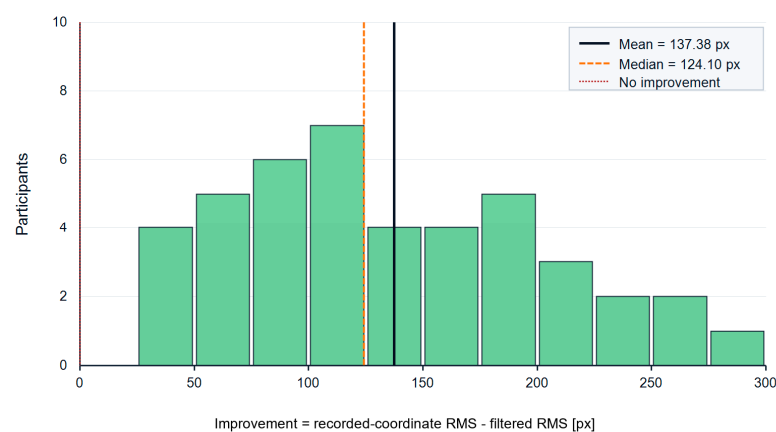


Figure 4. Distribution of participant-level target-proximity RMS reduction during dwell-active periods ($\Delta\text{RMS} = \text{RMS}(\text{recorded}) - \text{RMS}(\text{pipeline})$). Positive values indicate lower RMS for the complete pipeline output.

3.1. Whole-Task Target-Proximity Analysis

Whole-quiz RMS was analyzed as a secondary endpoint because it includes the full recorded task trajectory: transitions, reading, visual search, target exploration, non-target periods, and dwell-active periods. Mean whole-quiz RMS decreased from 325.35 px for the recorded uncorrected coordinates to 247.66 px for the complete pipeline output, with a mean reduction of 77.70 px (95% CI, 62.88–92.51 px). All 43 participants showed a reduction

(Wilcoxon $p < 10^{-8}$; Cohen's $d_z = 1.61$). This full-trajectory descriptor complements the dwell-active primary endpoint but is less specific because the intended target was not externally defined during visual search, reading, and transitions.

3.2. Target Selection Timing and Example Trajectory

Across 1935 logged target selections, the median time from task onset to target selection was 5.17 s (interquartile range (IQR): 2.83 to 14.68 s; mean: 10.20 s). When timing was measured from the first availability of the selected target in the active target set, the median was 2.79 s (IQR: 1.78 to 7.91 s; mean: 6.61 s). The final continuous dwell segment immediately preceding selection lasted a median of 1.51 s (IQR: 1.51 to 1.52 s), consistent with the configured 1500 ms dwell threshold. Figure 5 shows an example single-target trajectory, illustrating the recorded cursor path, filtered trajectory, target hit region, and selection point for one representative target selection.

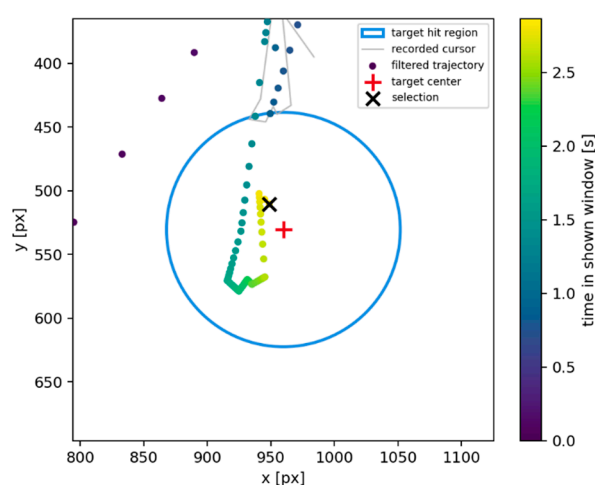


Figure 5. Example single-target selection trajectory for one representative target selection.

3.3. Consistency Across Controlled Acquisition Configurations

As a supporting consistency check, recorded-to-filtered target-proximity RMS reduction was observed in each of the three controlled acquisition configurations. This indicates consistency within the acquisition protocol, but it should not be interpreted as evidence that one tracker position setting had a performance advantage over another. Mean stable RMS improvement was 124.76 px in Configuration 1, 134.99 px in Configuration 2, and 152.40 px in Configuration 3, with improvement observed in all 43 participants within each configuration (Table 3).

Table 3. Stable RMS improvement across controlled acquisition configurations. Configuration 1: laser cross aimed between the eyebrows along the nose axis; Configuration 2: laser cross aimed at the nose; Configuration 3: laser cross aimed at the chin.

Configuration	Recorded Uncorrected Mean [px]	Pipeline Output Mean [px]	Mean Improvement [px]	Improved
1	181.98	57.22	124.76	43/43
2	190.07	55.08	134.99	43/43
3	208.16	55.77	152.40	43/43

4. Discussion

This study presented a custom single-point adaptive gaze-to-cursor correction pipeline designed for dwell-based eye-controlled interaction and evaluated it in human interaction

sessions. The proposed solution combines rapid central-point offset estimation with robust calibration handling, translational correction, jump rejection in the recorded cursor stream, temporal stabilization, dwell-based selection logic, and online post-selection adaptation. This design follows the rationale that calibration, outlier handling, and temporal filtering address distinct sources of gaze-to-cursor error in eye-tracking interaction [1–3,8–13].

The main finding is that the complete correction pipeline reduced application-level target-proximity RMS during operational dwell-active selection periods. The primary endpoint is intentionally tied to the dwell behavior required by the interface: after visual search and exploration, the user must keep the gaze-to-cursor point on one answer to select it. Thus, stable segments are not treated as arbitrary post hoc success cases, but as the task phase in which target proximity is behaviorally meaningful. However, because entry into this phase is determined by the online filtered coordinate, the endpoint remains algorithm-conditioned and does not provide independent validation of spatial gaze accuracy against externally cued fixation targets.

The secondary whole-quiz endpoint showed the same direction of target-proximity reduction, although with a smaller mean reduction because it included transitions, reading, visual search, target exploration, and periods before or outside target dwell. It provides a broader full-trajectory descriptor that complements, but does not replace, the more task-specific dwell-active endpoint.

The findings therefore concern target localization in the implemented interface and do not demonstrate improved native eye-tracker accuracy or superiority over conventional multi-point calibration.

The contribution of the work is twofold. First, it specifies and implements a complete lightweight gaze-to-cursor correction pipeline for a dwell-based interface. Second, it provides human data showing that the complete pipeline reduced target-proximity RMS in the operational dwell-active state of that interface. The evaluation does not determine which individual component caused the observed reduction, because central offset estimation, jump rejection, temporal smoothing, dwell logic, and post-selection adaptation were evaluated together as one applied pipeline. The dwell-based operating context is also important, because gaze selection performance depends on the balance between response speed and selection stability [14].

Accordingly, the term correction is used in an operational sense: the pipeline modifies cursor coordinates to improve dwell-based target localization in this interface, not to recover an independently verified gaze point.

The present study did not include a conventional multi-point calibration baseline or externally cued fixation-validation targets; these limitations and the corresponding requirements for future validation are described in the Limitations section.

The reduction was observed in all three controlled acquisition configurations. This observation is used only as a consistency check for the protocol, not as an analysis of differences between tracker position settings or as evidence for the advantage of any single tracker position. Further validation should also measure eye-to-screen geometry more precisely if angular target-proximity RMS is a major endpoint.

Limitations

Several limitations should be considered when interpreting the present findings.

First, the recorded input was acquired indirectly through the Tobii Experience gaze-to-cursor function and operating-system cursor interface rather than through native research-grade Tobii Pro SDK or Tobii Pro Lab streams. This acquisition path may affect temporal resolution, quantization, smoothing, throttling, coordinate duplication, and latency. Gaze-to-cursor updates were triggered at 10 Hz, while the application logged the currently

available cursor position at a higher and variable realized rate. Consecutive stored records could therefore contain repeated coordinates between actual gaze-to-cursor updates. The stored trajectory should be interpreted as the application-level recorded cursor-coordinate stream rather than as native research-grade gaze-coordinate records.

Potential preprocessing may include vendor calibration, smoothing, latency, throttling, coordinate quantization, and duplication before the application logger receives the cursor position. Because native Tobii Pro SDK or Tobii Pro Lab streams were not recorded, the logs did not contain native binocular or monocular tracking status, pupil validity, device confidence values, high-precision eye-tracker timestamps, or blink and tracking-loss markers. Application-level missing-data representation was therefore limited to the recorded cursor-coordinate stream, jump rejection flag, dwell-state fields, and emitted target selections.

The observed jump rejection rate was low in the analyzed dataset: 5555 recorded cursor-coordinate samples were rejected and 286,341 samples with evaluated validity were accepted. Rejected samples therefore accounted for 1.90% of samples with evaluated validity. The session-level rejection percentage averaged 1.94% (median 1.89%, range 1.24% to 3.03%). These values characterize the present implementation and acquisition path, not a device-independent rejection rate.

Second, the study evaluated a specific device, measurement setup, software path, and experimental protocol. The results should not be generalized automatically to other eye trackers, native gaze application programming interfaces (APIs), display geometries, calibration procedures, or measurement conditions without additional validation. In particular, the present findings should not be interpreted as evidence about native Tobii Pro SDK data quality, because those research-grade data streams were not logged.

Third, stable RMS was reported for dwell-active stable segments identified using the online logic of the evaluated pipeline. This means the primary endpoint used all logged samples for which the application itself had entered target dwell, rather than manually selected windows. However, the experimental task did not include a separate, independent validation block with unambiguously cued fixation points. Therefore, the present dataset does not allow a fully independent validation of spatial accuracy relative to a known target that the participant was explicitly instructed to fixate upon.

Fourth, angular RMS values are approximate. The display pixel pitch was derived from the nominal laptop panel geometry, and the recorded sample coordinate system was the 1920 by 1061 px application window. The eye-to-screen distance was approximately 70 cm, and the position of the eyes relative to the active screen area was not measured per participant.

Fifth, the clinical and assistive relevance is motivational. Although the prior literature supports the relevance of eye-controlled interaction for patients with limited verbal or motor communication channels [4–7], the present experimental sample and protocol do not establish clinical effectiveness in post-stroke or severely motor-impaired users.

Sixth, no ablation separated the contributions of central offset estimation, jump rejection, EMA filtering, dwell logic, and post-selection adaptation. The study also did not compare adaptive and non-adaptive versions of the algorithm, so the specific contribution and long-term stability of the adaptation rule remain unresolved. Thus, the present results do not support causal module-level attribution.

A future ablation should compare at least offset-only correction, offset plus jump rejection, offset plus EMA filtering, the full pipeline without post-selection adaptation, and the full adaptive pipeline. The study does not include an ablation analysis of the individual components of the correction pipeline. Consequently, it is not possible to isolate the contribution of specific algorithmic elements to the observed performance improvements.

A future validation study should combine externally cued pre- and post-task fixation targets, a direct conventional multi-point calibration comparator, native research-grade gaze recording where available, and a predefined ablation of the pipeline modules. It should also assess parameter sensitivity, cumulative offset drift, rejected-sample proportions, acquisition time, and false dwell events in the intended user population.

5. Conclusions

This work implemented a complete low-burden single-point gaze-to-cursor correction pipeline in a dwell-based selection interface. The pipeline combined central offset estimation, MAD-based calibration outlier handling, jump rejection in the recorded cursor stream, EMA filtering, dwell logic, and post-selection adaptation.

In 43 participants and 129 sessions, the complete pipeline reduced application-level target-proximity RMS during operational dwell-active selection periods. Participant-level mean stable RMS decreased from 193.40 px to 56.02 px, with a mean improvement of 137.38 px and a large paired effect size. In approximate angular units, stable RMS decreased from 3.15 degrees to 0.91 degrees. Whole-quiz RMS also decreased as a secondary full-trajectory descriptor that includes visual search, reading, transitions, target exploration, non-dwell periods, and dwell-active periods.

These findings support the practical feasibility of a low-burden operational correction layer for the evaluated gaze-to-cursor interface. They do not establish independent spatial gaze accuracy, native eye-tracker accuracy improvement, clinical effectiveness in assistive populations, or superiority over conventional multi-point calibration. Further studies should address these design limitations through independent validation, comparative calibration, and module-level analysis.

Author Contributions: Conceptualization, P.K., F.G., R.O. and G.Ž.; methodology, P.K. and F.G.; software, P.K. and F.G.; validation, P.K., F.G., R.O. and G.Ž.; formal analysis, P.K., F.G. and G.Ž.; investigation, P.K., F.G., Z.T. and G.Ž.; resources, P.K., F.G., R.O., Z.T. and G.Ž.; data curation, P.K., F.G., Z.T. and G.Ž.; writing—original draft preparation, P.K., F.G. and G.Ž.; writing—review and editing, P.K., F.G., R.O. and G.Ž.; visualization, P.K. and F.G.; supervision, P.K.; project administration, P.K. and G.Ž. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Senate Committee on Research Ethics at the Wrocław University of Health and Sport Sciences, Poland (approval no. 5/2021).

Informed Consent Statement: Informed consent was obtained from all participants involved in the study.

Data Availability Statement: Data and methods of analysis are available to qualified researchers upon request.

Acknowledgments: We sincerely thank all participants for their contribution to the study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Morimoto, C.H.; Mimica, M.R.M. Eye gaze tracking techniques for interactive applications. *Comput. Vis. Image Underst.* **2005**, *98*, 4–24. [[CrossRef](#)]
2. Nyström, M.; Andersson, R.; Holmqvist, K.; van de Weijer, J. The influence of calibration method and eye physiology on eyetracking data quality. *Behav. Res. Methods* **2013**, *45*, 272–288. [[CrossRef](#)] [[PubMed](#)]
3. Drewes, H.; Pfeuffer, K.; Alt, F. Time- and space-efficient eye tracker calibration. In Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, Denver, CO, USA, 25–28 June 2019; pp. 1–8. [[CrossRef](#)]

4. Zurek, G.; Kryś-Noszczyk, K.; Kunka, B.; Binder, M. Eye Tracking as a Novel Approach to Cognitive Assessment in DOC Patients Based on Evidence from a Multicenter Clinical Trial. *Sci. Rep.* **2025**, *15*, 44205. [CrossRef] [PubMed]
5. Zurek, G.; Binder, M.; Kunka, B.; Kosikowski, R.; Rodzen, M.; Karas, D.; Mucha, G.; Olejniczak, R.; Goraczko, A.; Kujawa, K.; et al. Can Eye Tracking Help Assess the State of Consciousness in Non-Verbal Brain Injury Patients? *J. Clin. Med.* **2024**, *13*, 6227. [CrossRef] [PubMed]
6. Kujawa, K.; Zurek, G.; Kwiatkowska, A.; Olejniczak, R.; Zurek, A. Assessment of Language Functions in Patients with Disorders of Consciousness Using an Alternative Communication Tool. *Front. Neurol.* **2021**, *12*, 684362. [CrossRef] [PubMed]
7. Kujawa, K.; Zurek, A.; Goraczko, A.; Olejniczak, R.; Zurek, G. Monitoring Eye Movements Depending on the Type of Visual Stimulus in Patients with Impaired Consciousness Due to Brain Damage. *Int. J. Environ. Res. Public Health* **2022**, *19*, 6280. [CrossRef] [PubMed]
8. Villanueva, A.; Cabeza, R. A novel gaze estimation system with one calibration point. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2008**, *38*, 1123–1138. [CrossRef] [PubMed]
9. Hornof, A.J.; Halverson, T. Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behav. Res. Methods Instrum. Comput.* **2002**, *34*, 592–604. [CrossRef] [PubMed]
10. Stampe, D.M. Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behav. Res. Methods Instrum. Comput.* **1993**, *25*, 137–142. [CrossRef]
11. Špakov, O. Comparison of eye movement filters used in HCI. In Proceedings of the Symposium on Eye Tracking Research and Applications, Santa Barbara, CA, USA, 28–30 March 2012; pp. 281–284. [CrossRef]
12. Leys, C.; Ley, C.; Klein, O.; Bernard, P.; Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **2013**, *49*, 764–766. [CrossRef]
13. Rosengren, W.; Nyström, M.; Hammar, B.; Stridh, M. A robust method for calibration of eye tracking data recorded during nystagmus. *Behav. Res. Methods* **2020**, *52*, 36–50. [CrossRef] [PubMed]
14. Majaranta, P.; MacKenzie, I.S.; Aula, A.; Rähkä, K.-J. Effects of feedback and dwell time on eye typing speed and accuracy. *Univers. Access Inf. Soc.* **2006**, *5*, 199–208. [CrossRef]
15. Tobii Help Center. Specifications for Eye Tracker 5. Available online: <https://help.tobii.com/hc/en-us/articles/360012483818-Specifications-for-Eye-Tracker-5> (accessed on 8 July 2026).
16. Tobii Gaming. Tobii Eye Tracker 5. Available online: <https://gaming.tobii.com/product/eye-tracker-5/> (accessed on 8 July 2026).
17. Tobii Help Center. Windows Interaction Features for ET5. Available online: <https://help.tobii.com/hc/en-us/articles/360011075918-Windows-Interaction-features-for-ET5> (accessed on 8 July 2026).
18. Tobii Developer Zone. Tobii Pro SDK. Available online: <https://developer.tobii.com/tobii-pro-sdk/> (accessed on 8 July 2026).
19. Tobii Pro SDK Python API. GazePoint and PupilData Class References. Available online: https://developer.tobiipro.com/tobii-research/python/reference/1.6.2.37-alpha-gcca303b/namespacetobii__research.html (accessed on 8 July 2026).
20. Holmqvist, K.; Nyström, M.; Mulvey, F. Eye tracker data quality: What it is and how to measure it. In Proceedings of the Symposium on Eye Tracking Research and Applications, Santa Barbara, CA, USA, 28–30 March 2012; pp. 45–52. [CrossRef]
21. Feit, A.M.; Williams, S.; Toledo, A.; Paradiso, A.; Kulkarni, H.; Kane, S.; Morris, M.R. Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 1118–1130. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.