Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Model-based comparison of latency estimation methods for the pupillary light reflex

Marcel Schepelmann

marcel.schepelmann@hci.uni-hannover.de

Leibniz University Hannover

Hans Georg Krojanski

Leibniz University Hannover

Article

Keywords:

Posted Date: May 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9209962/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on September 2nd, 2026. See the published version at https://doi.org/10.1038/s41598-026-68921-9.

Model-based comparison of latency estimation methods for the pupillary light reflex

Marcel Schepelmann[1,*] **and Dr. Hans Georg Krojanski**[1]

[1]Leibniz University Hannover, Human-Computer Interaction, Hanover, 30167, Germany
[*]marcel.schepelmann@hci.uni-hannover.de

ABSTRACT

Accurate estimation of pupillary light reflex (PLR) latency is important in clinical and research settings, yet reported results are often difficult to compare due to differences in hardware, sampling rates, and analysis methods. This study introduces a systematic and reproducible benchmark for evaluating PLR latency estimation algorithms using synthetic data with known ground-truth latency.

Synthetic pupillograms were generated using an established dynamic model of the PLR, extended to handle short light impulses. Five commonly used latency estimation methods were evaluated under these conditions, with 1000 traces simulated per configuration. Performance was assessed using the mean absolute error (MAE) between estimated and ground-truth latency. Across most conditions, a method developed by Bergamin and Kardon, combining filtering, interpolation, and analysis of the first and second derivatives achieved the lowest MAE, although its performance deteriorated under high noise and low-intensity stimuli. In contrast, a simple piecewise linear fit showed consistent, moderate performance across configurations. Threshold-based detection performed well for strong stimuli but degraded for weak responses, while derivative-based methods were sensitive to noise and exponential curve-fitting performed inconsistently across stimulus conditions.

Introduction

The PLR is the autonomic response of the pupil to light. Following an exposure to a light stimulus, the pupil constricts and gradually re-dilates once the stimulus is removed[1]. Pupillometry, the quantitative measurement of this response, is increasingly used in clinical diagnostics, neurophysiological research, and mobile health applications[2–5]. Among the measurable PLR features, latency—the delay between stimulus activation and the constriction onset—is of particular importance. Pathologic latency has been associated with outer retinal dysfunction and afferent damage[6,7]. Latency of the PLR was also reported as biomarker for multiple sclerosis[8,9], Leber hereditary optic neuropathy[10], amblyopia[11,12], diabetes[13] and concussions[2,4].

Despite its significance, reliable latency estimation remains a nontrivial challenge[6,7]. Several factors contribute to this difficulty. First, commercially pupillometers or labor prototypes operate at varying sampling rates ranging from 5 Hz to 2000 Hz[14–16], determining the temporal precision with which small deflections in pupil diameter can be detected. Second, detection-based noise introduces variability that may obscure the early slope of the constriction onset. Third, physiological oscillations such as hippus can impose slow drift on the diameter signal[17–19]. These effects combine to produce uncertainty in determining the true latency, even under controlled experimental conditions. In a study by Capó-Aponte et al.[2] involving 20 participants with concussions and a control group of 20 participants, a statistically significant difference was found between the latency of the two groups. However, the difference in the mean values of both groups was less than 30 ms. This further demonstrates that high accuracy is important. Furthermore, Bos states that a difference of over 20 ms for different estimation methods is unacceptable[15].

A variety of methodological strategies have been proposed for latency estimation, including linear model fitting[20,21], velocity deflections from zero[15], maximal negative acceleration[6] and curve-fitting using a second-order model[15]. However, comparing these methods objectively is challenging, as real PLR data doesn't provide an exact ground truth value for latency. Furthermore, experimental setups differ widely in illumination, device characteristics, and measurement noise. As a result, performance comparisons across studies remain inconsistent, and it is unclear how various latency estimation approaches behave under reduced frame rates, increased noise levels, or atypical physiological responses.

To address these issues, we propose a model-based framework that enables controlled comparison of latency estimation methods under reproducible conditions. By generating synthetic PLR signals with known ground-truth latency and configurable noise, drift, and sampling properties, the framework provides a systematic way to evaluate how different algorithms perform across a broad range of measurement scenarios. This approach not only introduces a new benchmarking but also reveals the conditions under which specific latency estimation strategies fail or remain robust.

Related Work

Pupillary Light Reflex Modeling

Mathematical models of the PLR can be categorized into equilibrium-state models and dynamic models[22]. Equilibrium-state models describe the steady-state relationship between pupil size and luminance and are typically derived from experimental measurements. Although they lack temporal dynamics, they are frequently incorporated into dynamic models[22]. Since latency estimation requires explicit temporal behavior, this review focuses on dynamic PLR models.

Longtin and Milton introduced one of the earliest dynamic PLR models based on a nonlinear delay differential equation[23]. Their model incorporates physiological latency by introducing a delayed retinal light signal and nonlinear iris mechanics, enabling the simulation of delayed and oscillatory pupil responses.

Usui and Hirata proposed a detailed biomechanical model of the human pupillary muscle plant, representing the sphincter and dilator muscles as viscoelastic elements[24]. While physiologically accurate, the resulting system of 19 differential equations is computationally expensive and difficult to parameterize, limiting its applicability for large-scale simulations[25].

Pamplona and Oliveira combined the dynamic formulation of Longtin and Milton with the experimentally validated equilibrium pupil model of Moon and Spencer[26], resulting in a hybrid model that captures both realistic steady-state behavior and temporal dynamics[27]. We follow this modeling strategy in our simulations; further details are provided in the subsection Model Implementation.

To improve simulation accuracy for short light pulses, Fan and Yao proposed a second-order delayed differential equation model that reduces complexity compared to earlier approach from Usui and Hirata while retaining viscoelastic muscle behavior[25]. More recently, Zandi and Khanh combined the Fan and Yao model with a feed-forward neural network to enhance modeling flexibility[28].

Latency Estimation in Pupillometry

Early pupillometry studies reported substantial variability in PLR latency, but often failed to specify how latency was determined. For example, Ellis[1] reported latencies ranging from 220 ms to 500 ms and observed a decrease in latency with increasing stimulus intensity; however, the method used to define response onset was not described.

Bos explicitly addressed this lack of methodological transparency and noted that only a minority of studies reported their latency detection approach[15]. He compared manual and automated latency estimation at different sampling rates and concluded that, at high sampling rates (≈200 Hz), detecting deviations in pupil velocity from zero is sufficient, whereas at lower sampling rates (≈50 Hz), curve-fitting approaches are more appropriate.

Bergamin and Kardon investigated the clinical relevance of PLR latency and highlighted the absence of a generally accepted estimation method[6]. They proposed an algorithm based on filtering, interpolation, and analysis of the first and second derivatives of the pupil signal. The method was developed using recordings from seven healthy subjects and evaluated on a larger cohort under the assumption that latencies in the left and right eyes are identical. The correlation coefficient between both eyes over several measurements was used as a metric to measure reliability. The authors further noted that increased spatial resolution may be more beneficial than very high sampling rates.

More recently, Jendritza et al.[7] compared three different methods for determining PLR latency using pupillometry data from 150 healthy subjects. The evaluated approaches were based on linear fits to the baseline pupil diameter and the descending part of the pupillogram, using different numbers of data points. Test–retest reliability was used as the primary evaluation metric. Among the tested approaches, a method using a reduced fitting window during the pupillary constriction phase proved to be the most robust, particularly for small constriction amplitudes and noisy data.

Across these studies, latency estimation methods are typically evaluated using indirect assumptions, such as inter-eye symmetry[6] or test–retest consistency[7], because the true latency is unknown. This limitation motivates the use of synthetic PLR data with known ground-truth latency, enabling direct and quantitative comparison of latency detection methods under controlled conditions.

Methods

Model Implementation

For the purpose of generating synthetic PLR signals under varying conditions (e.g., sampling rate, noise level, and inter-individual variability), a dynamic model that is both computationally efficient and physiologically plausible is required. The biomechanical model by Usui and Hirata[24] was deemed too complex for this purpose, consistent with the assessment by Fan and Yao[25]. The model by Fan and Yao is specifically tailored to short light pulses[22], while the hybrid deep-learning approach by Zandi and Khanh[28] offers limited interpretability and flexibility for systematic parameter variation. Therefore, the model proposed by Pamplona and Oliveira[27] was selected as the basis for synthetic data generation.

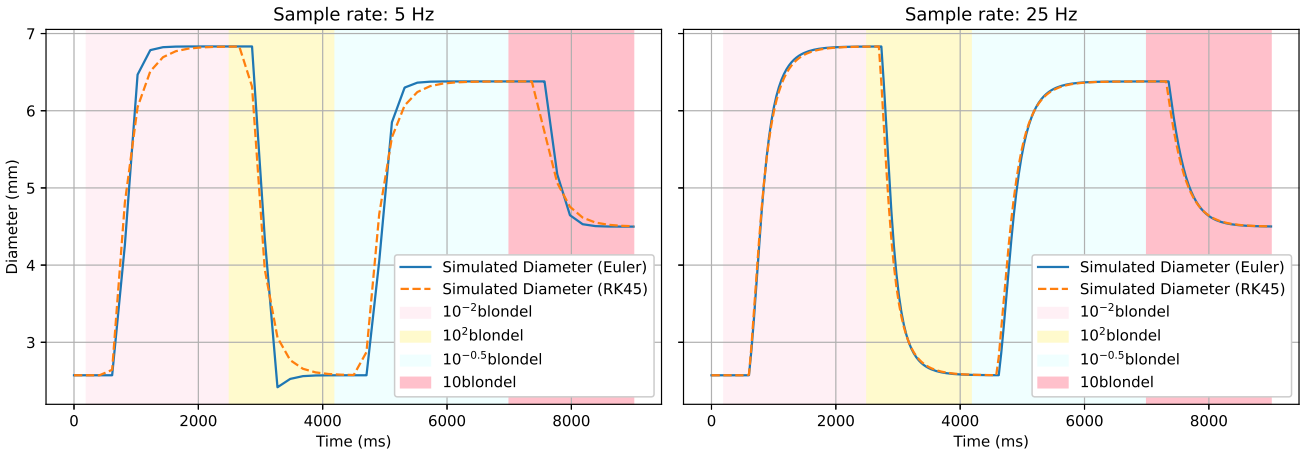


Figure 1. Comparison of the results of our solver (solid line) with the explicit Runge-Kutta method of order 5(4) from SciPy (dotted line). The same illumination levels as those used by Pamplona and Oliveira (10^2, 10^2, $10^{-0.5}$, and 10 blondels) were chosen for this simulation. On the left side, the simulation was performed with a sample rate of 5 Hz, and on the right side with a sample rate of 25 Hz.

The governing differential equation of the Pamplona model is given by

$$\frac{dM}{dD}\frac{dD}{dt} + 2.3026 \operatorname{atanh}\left(\frac{D-4.9}{3}\right) = 5.2 - 0.45 \ln\left[\frac{\phi(t-\tau)}{\tilde{\phi}}\right], \tag{1}$$

with

$$M(D) = \operatorname{atanh}\left(\frac{D-4.9}{3}\right), \tag{2}$$

where D is the pupil diameter in mm, $\phi(t)$ the stimulus intensity in lumens at time t, $\tilde{\phi}$ is the light level threshold below which there is no longer any pupil response, and τ the latency in milliseconds. The function $M(D)$ models the elastic response of the iris as a function of pupil diameter. This equation combines a dynamic first-order differential system with the equilibrium-state model of Moon and Spencer[26], ensuring realistic steady-state pupil diameters across varying luminances.

Pamplona and Oliveira account for the asymmetry between pupil constriction and re-dilation, with constriction occurring approximately three times faster than re-dilation[1,29]. This asymmetry is implemented via different integration step sizes δt_c and δt_d respectively:

$$\delta t_c = \frac{T_i - T_{i+1}}{S}, \quad \delta t_d = \frac{T_i - T_{i+1}}{3S}, \tag{3}$$

where T_i and T_{i+1} denote consecutive simulation time steps in milliseconds. The parameter S is an individual-specific constant that governs the velocity of pupil constriction and dilation.

We implemented a custom Euler-based solver to handle these adaptive steps[30]. While Euler integration is less accurate than higher-order methods[31], comparison with SciPy's[32] explicit Runge–Kutta 5(4) solver (RK45) confirmed acceptable accuracy, with a mean squared error (MSE) below 0.01 mm^2 for sampling rates above 25 Hz (Fig. 1, 2).

A known limitation of the Pamplona and Oliveira model is its handling of short light pulses, as previously noted by Fan and Yao[25]. The issue originates from the delayed stimulus term $\phi(t-\tau)$ during numerical integration. The latency τ is computed according to Stark's empirical model[33]:

$$\tau(R, L_{fL}) = 253 - 14 \ln(L_{fL}) + 70R - 29R \ln(L_{fL}), \tag{4}$$

where L_{fL} denotes luminance in foot-Lamberts and R the stimulation frequency in Hz.

The problem can be illustrated by considering a very short light pulse with a duration of δ as shown in Fig. 3. According to Stark's formula, there are then two latencies for the entire process. A short latency τ_1 during the light pulse and a longer latency τ_2 before and after the light pulse. Since the light intensity for the light pulse is known during simulation, equation (4) can be

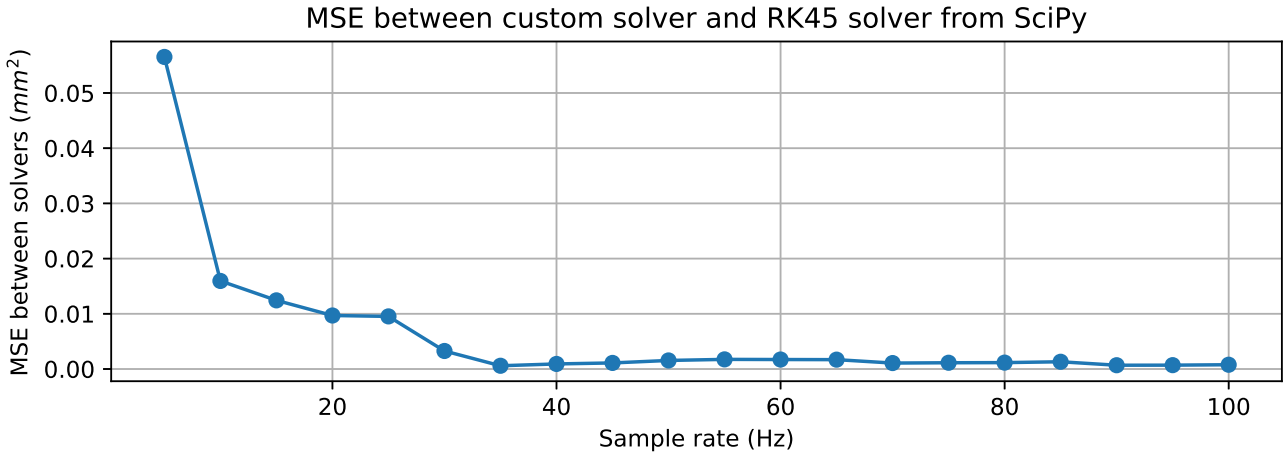


Figure 2. MSE between our custom solver, which uses the Euler method, and an RK45 solver from SciPy. The MSE is evaluated over sample rates between 5 and 100 over a step size of 5.

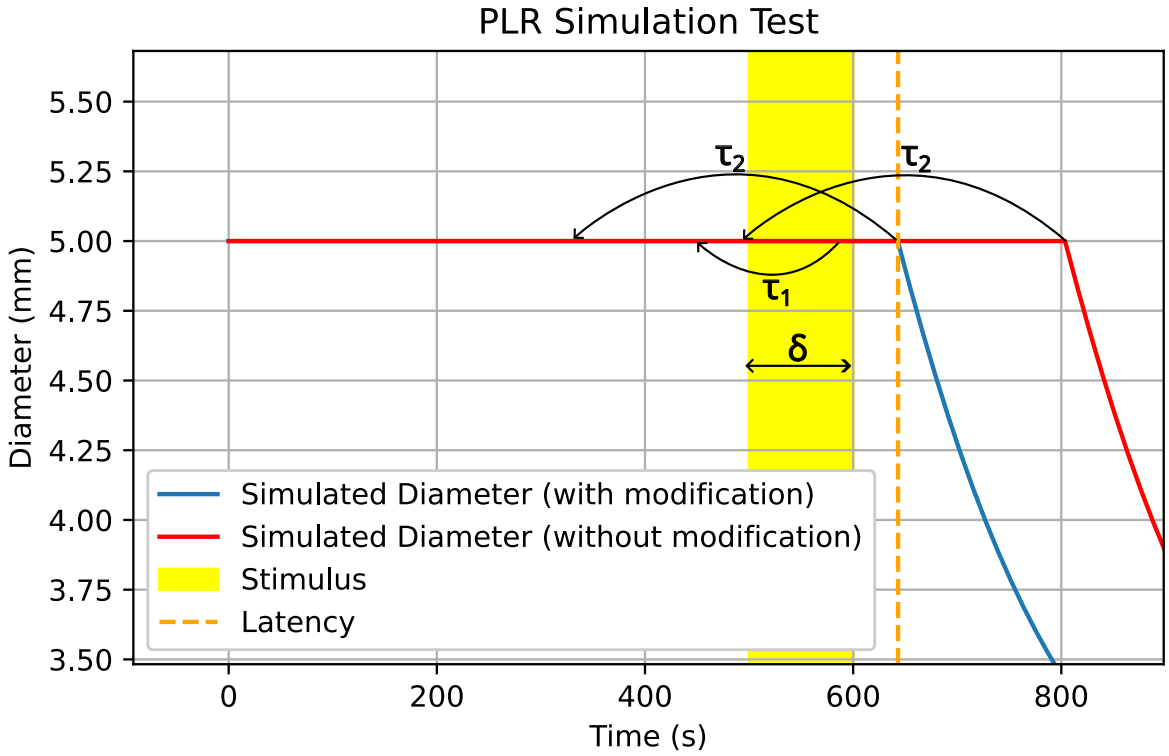


Figure 3. Illustration of default latency handling for a short light pulse and the effect of the proposed modification. A brief stimulus of duration δ (yellow region) induces a short latency τ_1 during the pulse and a longer latency τ_2 before and after the pulse. The arrows τ_1 and τ_2 indicate that at time t, $\phi(t-\tau)$ acts on the pupil. The dashed orange line indicates the model-predicted onset of the pupil response according to equation (4). Without modification (red curve), the response is delayed because the longer latency τ_2 is active during numerical integration, causing the delayed stimulus evaluation to reference pre-stimulus luminance values. With the proposed modification (blue curve), the response onset aligns with the predicted latency τ_1, resulting in a single continuous pupil response.

used to calculate when the pupil response should begin (dashed orange line). However, since τ_2 is already active again at this time step during numerical integration, a value is calculated for $\phi(t-\tau)$ that lies before the onset of the light pulse. As a result, the pupil's response may be delayed or even fragmented whenever $\delta < \tau_2$.

To address this issue, the delayed stimulus evaluation in equation (1) was modified to ensure that short light pulses elicit a single, continuous response. The model maintains an active latency τ_{active} and the time t_{active} at which it became effective. When a stimulus change occurs, shorter latencies are adopted immediately, while longer latencies are deferred until the preceding shorter latency has elapsed.

Formally, the delayed stimulus is defined as

$$\phi_{\text{delayed}}(t) = \begin{cases} \phi_{\text{hold}}, & t - \tau_{\text{active}} < t_{\text{active}}, \\ \phi(t - \tau_{\text{active}}), & t - \tau_{\text{active}} \geq t_{\text{active}}, \end{cases} \tag{5}$$

where ϕ_{hold} denotes the stimulus value immediately before the change that activated the current latency. This modification ensures that brief light impulses with durations shorter than the calculated latency produce a physiologically plausible, uninterrupted pupil response.

Inter-individual variability is modeled following Pamplona and Oliveira by scaling the simulated pupil diameter D between subject-specific lower and upper bounds:

$$D_{\text{final}} = C_{bD}(D) + \big(C_{tD}(D) - C_{bD}(D)\big) r_I, \tag{6}$$

where $r_I \in [0,1]$ denotes an individual index. The functions $C_{tD}(D)$ and $C_{bD}(D)$ define the upper and lower envelopes of pupil diameter as a function of luminance, derived from the experimental data of Moon and Spencer[26]. Since the original fifth-degree polynomials reported by Pamplona and Oliveira did not yield realistic values in our simulations, new polynomial approximations were fitted to the mean pupil diameter curve obtained from Moon and Spencer's data:

$$C_{tD}(D) = -0.015D^5 + 0.390D^4 - 3.913D^3 + 18.452D^2 - 38.770D + 31.368, \tag{7}$$

$$C_{bD}(D) = -0.022D^5 + 0.537D^4 - 4.935D^3 + 21.440D^2 - 43.234D + 34.305. \tag{8}$$

The parameter D is the pupil diameter calculated using equation (1).

Data Generation

Hippus, the small involuntary oscillations of pupil diameter, is introduced in our simulations to reflect natural variability. While Pamplona and Oliveira[27] modeled hippus as fluctuations in stimulus brightness, we add it retrospectively to the pupil diameter in order to isolate the response to illumination. The oscillations are implemented as a sinusoid with a frequency between 0.05 Hz and 0.3 Hz[34] with an amplitude of 0.2 mm[35].

Synthetic PLR traces were generated parametrically. For a given set of generation parameters—sampling rate, noise level, and stimulus intensity—a population of pupil responses was simulated. Each population consisted of 1000 independent traces to ensure statistically representative error estimates. The ground-truth latency was known for every trace and used for subsequent evaluation.

Stimulus intensity was controlled indirectly by specifying an average baseline diameter and an average minimum diameter reached after the light pulse; the corresponding stimulus intensity was then determined by the model. Low-intensity stimuli were represented by baseline and minimum diameters of 5 mm and 4 mm, respectively, while high-intensity stimuli used 7 mm and 3 mm. Inter-individual variability was introduced using the individual index r_I (6), sampled uniformly for each trace, which adjusts the final pupil diameters and ensures that the actual response deviates from these representative values.

The stimulus duration was fixed at 200 ms for all simulations. Sampling rate and noise level were treated as configurable parameters and varied systematically in the experiments described in the next section.

The underlying delayed differential equation was solved using a fixed internal integration step size of 1 ms for all simulations. The pupil diameter time series obtained from this high-resolution integration was subsequently sampled at the target sampling rates used in the experiments. This two-stage procedure reduces discretization-induced shifts of the response onset that may arise when integrating directly at low sampling rates, and ensures consistent temporal accuracy across all experimental conditions, including low-rate configurations.

Experiments

Five latency estimation methods were evaluated in our experiments. The first method is a linear intersection approach inspired by Jendritza et al.[7], representing methods based on the intersection of a linear fit through the baseline and a linear fit through the contraction phase. The second method corresponds to Bos's[15] preferred approach for high sampling rates, which identifies velocity deflections from zero, while the third method evaluates Bos's approach for low sampling rates, where curve fitting is recommended. The fourth method is that of Bergamin and Kardon[6], and a simple threshold method was included as a baseline.

For the simple threshold method, latency was estimated as the time when the pupil diameter first fell below three times the standard deviation of the baseline. For linear-fit-based methods, SciPy's least-squares solver[32] was used. To avoid heuristic choices regarding the number of points for fitting, a piecewise linear model was implemented in which both lines and the breakpoint are estimated simultaneously under a continuity constraint. This concept was proposed by Bos for his second-order curve-fitting method, which models the pupil response using two exponential functions[15], and was here extended to the linear intersection method. For Bos's high-sample-rate method, the maximum negative velocity was used instead of a zero-crossing due to noise-induced fluctuations.

For Bergamin and Kardon's method[6], we followed the procedure described in the original publication. The pupil signal was first smoothed using a Savitzky-Golay filter (window size of 5 time steps, second-order polynomial), then interpolated to 300 Hz. The first derivative was calculated and further smoothed with a Gaussian filter ($\sigma = 25$) to obtain velocity, followed by a second derivative to compute acceleration. Latency was defined as the point of maximum negative acceleration[6].

The performance of latency estimation methods was evaluated using synthetic pupillometry data with known ground-truth latency. For all experiments, the MAE between the estimated and true latency was used as the evaluation metric. Reported MAE values represent averages over populations of 1000 independently generated PLR traces per experimental configuration.

Two complementary experiments were conducted. In the first experiment (Fig. 4), the sampling rate was varied, while the measurement noise was held constant. Sampling rates between 25 Hz and 250 Hz were evaluated, covering the range of modern image-based pupillometers. Lower sampling rates were not considered, as the numerical solver accuracy deteriorates below this range, while higher rates were excluded due to limited relevance for camera-based systems and increased computational cost. The noise standard deviation was fixed at 0.01, 0.02 and 0.03 to illustrate interactions between noise and temporal resolution. The mean value of the noise was always set to 0 for all configurations and all experiments. For each sampling rate, latency estimation methods were applied, and the MAE was computed. In the corresponding plots, the sampling rate is shown on the x-axis, MAE on the y-axis, and each curve represents a different latency estimation method.

In the second experiment (Fig. 5), the noise level was varied while the sampling rate was fixed. Gaussian noise is assumed for the experiments, but can easily be replaced by other noise behaviors. Standard deviations between 0 mm and 0.03 mm were applied to simulate measurement inaccuracies. Steinhauer et al.[36] report pupil diameter measurement errors in the range of approximately 0.01–0.05 mm for video-based pupillometry systems. Although Gaussian noise is unbounded and individual samples may exceed ±0.05 mm, the selected noise levels reflect the scale of reported pupil diameter measurement errors. Rare larger deviations were tolerated to avoid artificial truncation of the noise distribution and to assess robustness under slightly adverse conditions. The sampling rate was fixed at 30 Hz and 90 Hz, reflecting commonly used and high-frequency video-based pupillometry systems, respectively. Additional results at 250 Hz are reported as a reference to illustrate interactions between noise and temporal resolution. As before, MAE values were computed over populations of 1000 traces for each noise level. In these plots, the noise standard deviation is shown on the x-axis, MAE on the y-axis, and curves correspond to latency estimation methods.

All experiments were performed separately for low- and high-intensity stimuli, as defined in the Data Generation subsection, to assess the robustness of latency estimation methods under varying constriction amplitudes. Ground-truth latency was identical across methods within each condition.

In Fig. 4 and 5, MAE values below 20 ms are highlighted in green. This threshold is based on Bos' statement that latency differences of more than 20 ms are considered unacceptable[15], as discussed in the Introduction.

Results

In most tested conditions, the method proposed by Bergamin and Kardon exhibited the lowest MAE. Interestingly, its performance decreased at higher noise levels under low-intensity stimulus conditions, despite being the only method to explicitly incorporate filtering to mitigate noise. This degradation, however, was slightly compensated by an increased sampling rate, indicating that higher temporal resolution improves robustness. One possible explanation is that the Gaussian noise used in our simulations, while commonly used to approximate measurement noise, may not fully capture all characteristics of noise present in real pupillometry recordings. As a result, the filtering strategy employed by Bergamin and Kardon may not interact with the simulated noise in the same way as with naturally occurring noise. The method using the minimum value of the first derivative was also susceptible to noise. And although this approach was used as representative of the method recommended

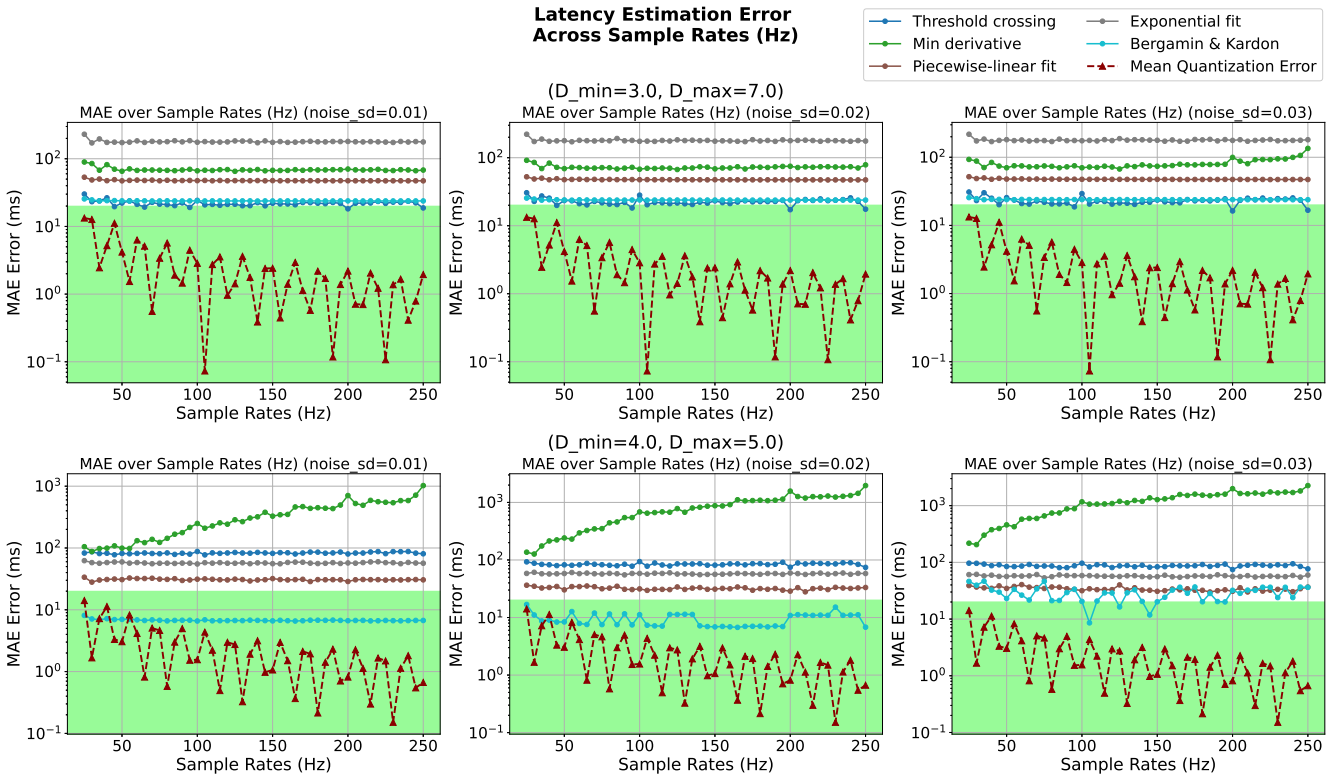


Figure 4. MAE of all evaluated latency estimation methods as a function of sampling rate. Columns correspond to increasing noise levels (σ = 0.01, 0.02, 0.03 mm, left to right). The top row shows results for high-stimulus conditions, while the bottom row shows results for low-stimulus conditions. The x-axis spans sampling rates from 25 Hz to 250 Hz in steps of 5 Hz. The green area indicates MAE below 20 ms.

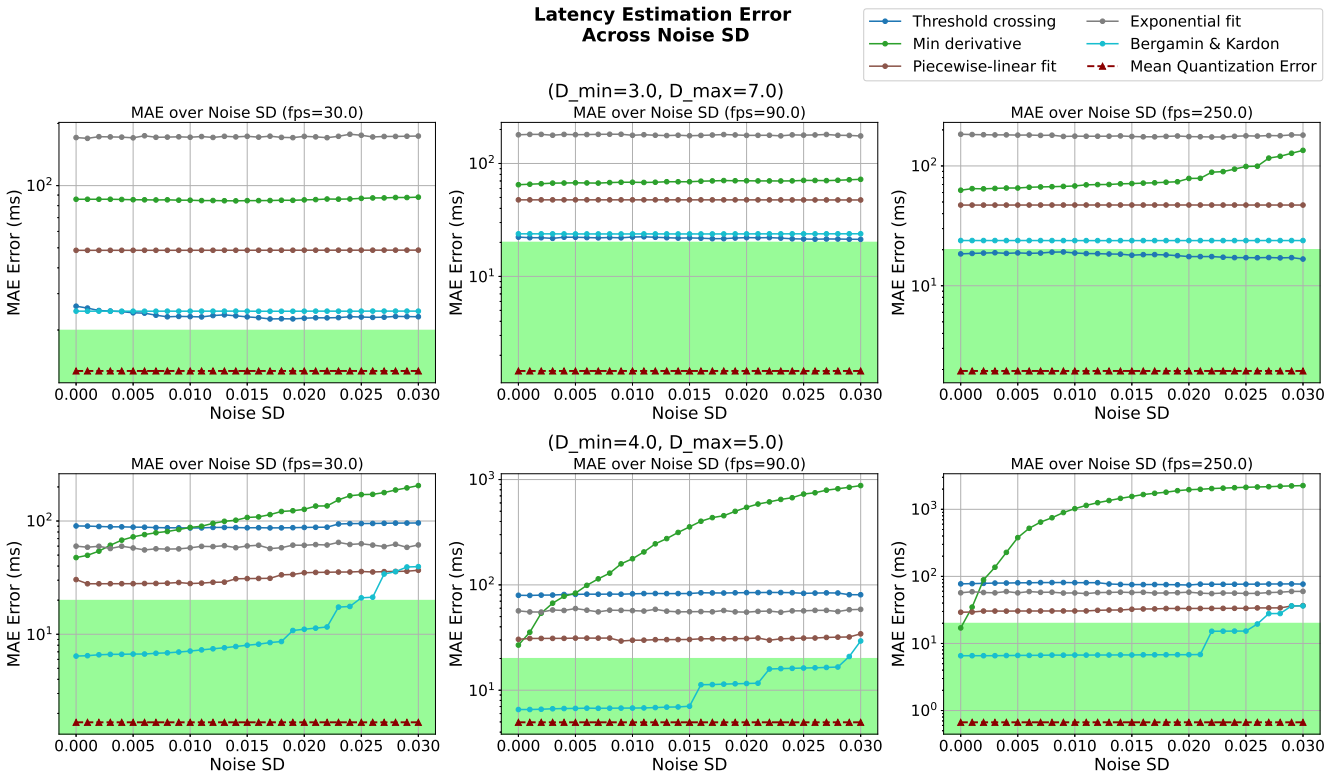


Figure 5. MAE of all evaluated latency estimation methods as a function of noise level. Columns correspond to different sampling rates (30 Hz, 90 Hz, and 250 Hz, left to right). The top row shows results for high-stimulus conditions, while the bottom row shows results for low-stimulus conditions. Gaussian noise standard deviation ranges from $\sigma = 0$ to 0.03 mm in steps of 0.001 mm. The green area indicates MAE below 20 ms.

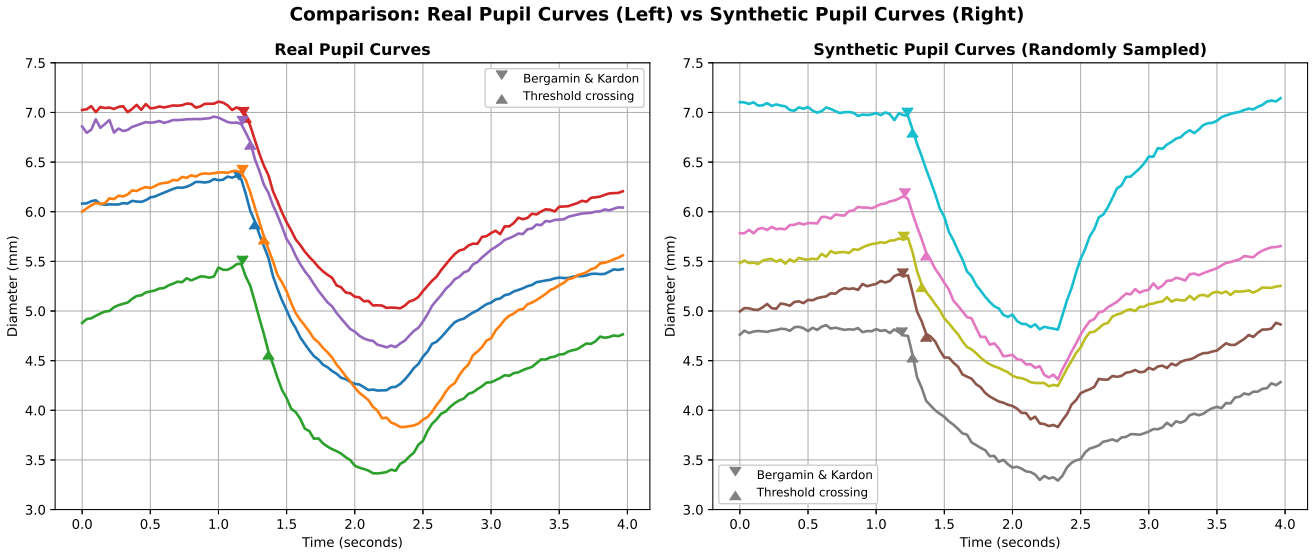


Figure 6. Comparison between empirical and simulated PLR recordings. Left: Five PLR curves recorded using a custom-built laboratory pupillometer. Right: Five randomly selected simulated pupillograms generated with matched sampling rate, noise level, and pupil diameter range (4–6 mm). Markers indicate latency estimates obtained using the Bergamin–Kardon method and the simple threshold-based method.

by Bos for high sampling rates, performance does not improve at higher sampling rates and even deteriorates with higher sampling rates in the low-stimulus experiments. The method recommended by Bos for lower latencies, a second-order curve fit, also does not appear to be affected by the sampling rate and exhibited the poorest performance under high-intensity stimulus conditions. In high-intensity stimulus conditions, the simple threshold method exhibited competitive performance. However, its accuracy decreased under low-intensity stimulus conditions, indicating limited robustness to small pupil constrictions. The piecewise linear fit method demonstrated consistent, average performance across all sampling rates and noise levels, showing little sensitivity to parameter variations. This stability suggests that the approach may be well suited for scenarios with unknown hardware specifications or as a straightforward initial method for latency estimation.

Surprisingly, the predefined quality criterion of an MAE of less than 20 ms is often not achieved. This result is particularly interesting because the difference between the mean values of the control group and subjects with concussion in the study by Capó-Aponte et al.[2] was <30 ms, which is not much higher than the selected threshold value. This also shows how important good spatial and temporal resolution is in pupillometry. If both are not given, we advise against regarding latency as a meaningful biomarker.

For qualitative validation of the benchmark with real measurements, five PLR recordings were acquired from a single subject using a custom-built laboratory pupillometer[37]. This experiment serves as a plausibility check to assess whether the benchmark findings transfer to empirical data. Ethics approval was obtained from the institutional research ethics committee of Leibniz University Hannover (ID: EV LUH 13/2025). All methods were performed in accordance with relevant guidelines and regulations, including the Declaration of Helsinki. Written informed consent was obtained from the participant.

The simulation frame rate and stimulus duration was matched to that of the pupillometer, and the noise level of the synthetic data was adjusted by visual comparison to approximate the measurement noise observed in the empirical recordings. The minimum and maximum pupil diameters in the simulation were set to 4 mm and 6 mm, respectively, to roughly correspond to the diameter range observed in the recorded data.

Under these conditions, the benchmark identified the method proposed by Bergamin and Kardon as yielding the lowest latency estimation error, whereas, for example, the simple threshold-based method performed very poorly for this configuration. Figure 6 compares the five empirically recorded pupillograms with five randomly selected simulated pupillograms generated under the matched conditions. In both real and simulated data, the Bergamin and Kardon method produces latency estimates that are more consistent with the observed pupil response onset than those obtained using the threshold method.

While the simulated traces exhibit a sharper transition between constriction and redilation than the empirical recordings, these differences do not affect latency estimation and are therefore not critical for the scope of this study.

Discussion

Our evaluation of latency estimation methods on synthetic pupillometry data highlights differences in robustness and accuracy. The method by Bergamin and Kardon achieved the most promising results, but its performance declined under high noise with low-intensity stimuli, an effect that could be mitigated by higher sampling rates, indicating that temporal resolution remains a critical factor for accurate latency estimation.

The piecewise linear fit method exhibited consistent, average performance across all tested configurations, indicating that it may be well suited for applications with unknown hardware specifications, such as smartphone-based pupillometry.

Limitations of this study include the use of synthetic data, which, while providing ground-truth latency, may not capture all complexities of real physiological responses. Noise was modeled as additive Gaussian, which may differ from real-world measurement noise in pupillometers. Future work should further investigate the transferability of these results to empirical data and explore whether the incorporation of filtering techniques could improve the performance of methods that currently do not employ them, as only the Bergamin and Kardon method explicitly uses filtering.

References

1. Ellis, C. J. The pupillary light reflex in normal subjects. *The Br. J. Ophthalmol.* **65**, 754–759 (1981).
2. Capó-Aponte, J., Urosevich, T., Walsh, D., Temme, L. & Tarbet, A. Pupillary Light Reflex as an Objective Biomarker for Early Identification of Blast-Induced mTBI. *Spine* **2**, 1–4, DOI: 10.4172/2165-7939.S4-004 (2013).
3. Capó-Aponte, J. E., Beltran, T. A., Walsh, D. V., Cole, W. R. & Dumayas, J. Y. Validation of Visual Objective Biomarkers for Acute Concussion. *Mil. Medicine* **183**, 9–17, DOI: 10.1093/milmed/usx166 (2018).
4. Ciuffreda, K. J., Joshi, N. R. & Truong, J. Q. Understanding the effects of mild traumatic brain injury on the pupillary light reflex. *Concussion* **2**, CNC36, DOI: 10.2217/cnc-2016-0029 (2017). Publisher: Future Medicine.
5. Mariakakis, A. *et al.* PupilScreen: Using Smartphones to Assess Traumatic Brain Injury. *Proc. ACM on Interactive, Mobile, Wearable Ubiquitous Technol.* **1**, 81:1–81:27, DOI: 10.1145/3131896 (2017).
6. Bergamin, O. & Kardon, R. H. Latency of the Pupil Light Reflex: Sample Rate, Stimulus Intensity, and Variation in Normal Subjects. *Investig. Ophthalmol. & Vis. Sci.* **44**, 1546–1554, DOI: 10.1167/iovs.02-0468 (2003).
7. Jendritza, R. *et al.* Optimizing latency calculation for robust evaluation of the pupillary light response in chromatic pupillography. *Documenta Ophthalmol.* DOI: 10.1007/s10633-025-10055-w (2025).
8. Jakobsen, J. Pupillary function in multiple sclerosis. *Acta Neurol. Scand.* **82**, 392–395, DOI: 10.1111/j.1600-0404.1990.tb03323.x (1990). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1600-0404.1990.tb03323.x.
9. Van Diemen, H. A. M., Van Dongen, M. M. M. M., Nauta, J. J. P., Lanting, P. & Polman, C. H. Pupillary light reflex latency in patients with multiple sclerosis. *Electroencephalogr. Clin. Neurophysiol.* **82**, 213–219, DOI: 10.1016/0013-4694(92)90170-M (1992).
10. Lüdtke, H., Kriegbaum, C., Leo-Kottler, B. & Wilhelm, H. Pupillary light reflexes in patients with Leber's hereditary optic neuropathy. *Graefe's Arch. for Clin. Exp. Ophthalmol.* **237**, 207–211, DOI: 10.1007/s004170050220 (1999).
11. Kase, M., Nagata, R., Yoshida, A. & Hanada, I. Pupillary light reflex in amblyopia. *Investig. Ophthalmol. & Vis. Sci.* **25**, 467–471 (1984).
12. Barbur, J. L., Hess, R. F. & Pinney, H. D. Pupillary function in human amblyopia. *Ophthalmic Physiol. Opt.* **14**, 139–149, DOI: 10.1111/j.1475-1313.1994.tb00102.x (1994). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1475-1313.1994.tb00102.x.
13. Lanting, P. *et al.* Assessment of pupillary light reflex latency and darkness adapted pupil size in control subjects and in diabetic patients with and without cardiovascular autonomic neuropathy. *J. Neurol. Neurosurgery, Psychiatry* **53**, 912–914, DOI: 10.1136/jnnp.53.10.912 (1990).
14. Nowak, W., Żarowska, A., Szul-Pietrzak, E. & Misiuk-Hojło, M. System and measurement method for binocular pupillometry to study pupil size variability. *BioMedical Eng. OnLine* **13**, 69, DOI: 10.1186/1475-925X-13-69 (2014).
15. Bos, J. E. Detection of the pupil constriction latency. *Med. Biol. Eng. Comput.* **29**, 529–534, DOI: 10.1007/BF02442326 (1991).
16. Andersson, R., Nyström, M. & Holmqvist, K. Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more. *J. Eye Mov. Res.* **3**, DOI: 10.16910/jemr.3.3.6 (2010).
17. Bouma, H. & Baghuis, L. C. J. Hippus of the pupil: Periods of slow oscillations of unknown origin. *Vis. Res.* **11**, 1345–1351, DOI: 10.1016/0042-6989(71)90016-2 (1971).

18. Lisowski, L. *et al.* The diagnostic significance of pupillary reflex pathways: insights from classical examination and advanced pupillometry. *Front. Neurosci.* **19**, DOI: 10.3389/fnins.2025.1677431 (2025). Publisher: Frontiers.

19. Turnbull, P. R. K., Irani, N., Lim, N. & Phillips, J. R. Origins of Pupillary Hippus in the Autonomic Nervous System. *Investig. Ophthalmol. & Vis. Sci.* **58**, 197–203, DOI: 10.1167/iovs.16-20785 (2017).

20. Kelbsch, C. *et al.* Objective Measurement of Local Rod and Cone Function Using Gaze-Controlled Chromatic Pupil Campimetry in Healthy Subjects. *Transl. Vis. Sci. & Technol.* **8**, 19, DOI: 10.1167/tvst.8.6.19 (2019).

21. Jendritza, R. *et al.* Influencing Factors on Pupillary Light Responses as a Biomarker for Local Retinal Function in a Large Normative Cohort. *Investig. Ophthalmol. & Vis. Sci.* **65**, 3, DOI: 10.1167/iovs.65.6.3 (2024).

22. Napieralski, P. & Rynkiewicz, F. Modeling Human Pupil Dilation to Decouple the Pupillary Light Reflex. *Open Phys.* **17**, 458–467, DOI: 10.1515/phys-2019-0047 (2019). Publisher: De Gruyter Open Access.

23. Longtin, A. & Milton, J. G. Modelling autonomous oscillations in the human pupil light reflex using non-linear delay-differential equations. *Bull. Math. Biol.* **51**, 605–624, DOI: 10.1007/BF02459969 (1989).

24. Usui, S. & Hirata, Y. Estimation of autonomic nervous activity using the inverse dynamic model of the pupil muscle plant. *Annals Biomed. Eng.* **23**, 375–387, DOI: 10.1007/BF02584438 (1995).

25. Fan, X. & Yao, G. Modeling Transient Pupillary Light Reflex Induced by a Short Light Flash. *IEEE transactions on bio-medical engineering* **58**, 36–42, DOI: 10.1109/TBME.2010.2080678 (2011).

26. Moon, P. & Spencer, D. E. On the Stiles-Crawford Effect. *JOSA* **34**, 319–329, DOI: 10.1364/JOSA.34.000319 (1944). Publisher: Optica Publishing Group.

27. Pamplona, V. F., Oliveira, M. M. & Baranoski, G. V. G. Photorealistic models for pupil light reflex and iridal pattern deformation. *ACM Transactions on Graph.* **28**, 106:1–106:12, DOI: 10.1145/1559755.1559763 (2009).

28. Zandi, B. & Khanh, T. Q. Deep learning-based pupil model predicts time and spectral dependent light responses. *Sci. Reports* **11**, 841, DOI: 10.1038/s41598-020-79908-5 (2021). Number: 1 Publisher: Nature Publishing Group.

29. Bergamin, O., Schoetzau, A., Sugimoto, K. & Zulauf, M. The influence of iris color on the pupillary light reflex. *Graefe's Arch. for Clin. Exp. Ophthalmol.* **236**, 567–570, DOI: 10.1007/s004170050122 (1998).

30. Griffiths, D. F. & Higham, D. J. Euler's Method. In Griffiths, D. F. & Higham, D. J. (eds.) *Numerical Methods for Ordinary Differential Equations: Initial Value Problems*, 19–31, DOI: 10.1007/978-0-85729-148-6_2 (Springer, London, 2010).

31. Hairer, E., Wanner, G. & Nørsett, S. P. Runge-Kutta and Extrapolation Methods. In Hairer, E., Wanner, G. & Nørsett, S. P. (eds.) *Solving Ordinary Differential Equations I: Nonstiff Problems*, 129–353, DOI: 10.1007/978-3-540-78862-1_2 (Springer, Berlin, Heidelberg, 1993).

32. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **17**, 261–272, DOI: 10.1038/s41592-019-0686-2 (2020).

33. Link, N. & Stark, L. Latency of the pupillary response. *IEEE Transactions on Biomed. Eng.* **35**, 214–218, DOI: 10.1109/10.1365 (1988).

34. Usui, S. & Stark, L. A model for nonlinear stochastic behavior of the pupil. *Biol. Cybern.* **45**, 13–21, DOI: 10.1007/BF00387209 (1982).

35. Hachol, A. *et al.* Measurement of pupil reactivity using fast pupillometry. *Physiol. Meas.* **28**, 61, DOI: 10.1088/0967-3334/28/1/006 (2006).

36. Steinhauer, S. R., Bradley, M. M., Siegle, G. J., Roecklein, K. A. & Dix, A. Publication guidelines and recommendations for pupillary measurement in psychophysiological studies. *Psychophysiology* **59**, e14035, DOI: 10.1111/psyp.14035 (2022).

37. Schepelmann, M. & Krojanski, H. G. Construction of a Mobile Headset for Automated Pupillary Motor Assessment. In *2025 IEEE/ACS 22nd International Conference on Computer Systems and Applications (AICCSA)*, 1–7, DOI: 10.1109/AICCSA66935.2025.11315195 (2025). ISSN: 2161-5330.

Acknowledgements

All figures were created using Matplotlib version 3.9.2. `https://matplotlib.org/`

Author contributions statement

M.S. designed and implemented the simulation framework, conducted the experiments, analyzed the results, and drafted the manuscript. H.G.K. contributed in an advisory role, provided methodological feedback, and critically revised the manuscript. All authors reviewed and approved the final version of the paper.

Data availability statement

The code used in this research is publicly available at:
https://github.com/PLR-Analyzer/PLR-Latency-Benchmark

Additional information

The authors declare no competing interests.

Figure legends

Figure 1. Comparison of the results of our solver (solid line) with the explicit Runge-Kutta method of order 5(4) from SciPy (dotted line). The same illumination levels as those used by Pamplona and Oliveira (10^2, 10^2, $10^{-0.5}$, and 10 blondels) were chosen for this simulation. On the left side, the simulation was performed with a sample rate of 5 Hz, and on the right side with a sample rate of 25 Hz.

Figure 2. MSE between our custom solver, which uses the Euler method, and an RK45 solver from SciPy. The MSE is evaluated over sample rates between 5 and 100 over a step size of 5.

Figure 3. Illustration of default latency handling for a short light pulse and the effect of the proposed modification. A brief stimulus of duration δ (yellow region) induces a short latency τ_1 during the pulse and a longer latency τ_2 before and after the pulse. The arrows τ_1 and τ_2 indicate that at time t, $\phi(t-\tau)$ acts on the pupil. The dashed orange line indicates the model-predicted onset of the pupil response according to Eq. 4. Without modification (red curve), the response is delayed because the longer latency τ_2 is active during numerical integration, causing the delayed stimulus evaluation to reference pre-stimulus luminance values. With the proposed modification (blue curve), the response onset aligns with the predicted latency τ_1, resulting in a single continuous pupil response.

Figure 4. Mean absolute error (MAE) of all evaluated latency estimation methods as a function of sampling rate. Columns correspond to increasing noise levels (σ = 0.01, 0.02, 0.03 mm, left to right). The top row shows results for high-stimulus conditions, while the bottom row shows results for low-stimulus conditions. The x-axis spans sampling rates from 25 Hz to 250 Hz in steps of 5 Hz. The green area indicates MAE below 20 ms.

Figure 5. Mean absolute error (MAE) of all evaluated latency estimation methods as a function of noise level. Columns correspond to different sampling rates (30 Hz, 90 Hz, and 250 Hz, left to right). The top row shows results for high-stimulus conditions, while the bottom row shows results for low-stimulus conditions. Gaussian noise standard deviation ranges from σ = 0 to 0.03 mm in steps of 0.001 mm. The green area indicates MAE below 20 ms.

Figure 6. Comparison between empirical and simulated PLR recordings. Left: five PLR curves recorded using a custom-built laboratory pupillometer. Right: five randomly selected simulated pupillograms generated with matched sampling rate, noise level, and pupil diameter range (4–6 mm). Markers indicate latency estimates obtained using the Bergamin–Kardon method and the simple threshold-based method.