Biomedical Physics &
Engineering Express

Endorsed by IPEM Institute of Physics and Engineering in Medicine

PURPOSE-LED PUBLISHING™

PAPER • **OPEN ACCESS**

VEOS: vision-based vertical electrooculography inference from monocular periocular video for ocular artefact suppression in EEG

To cite this article: Peter Redmond *et al* 2026 *Biomed. Phys. Eng. Express* **12** 045018

IOP Publishing Biomed. Phys. Eng. Express 12 (2026) 045018 https://doi.org/10.1088/2057-1976/ae83bb

Biomedical Physics & Engineering Express

OPEN ACCESS

RECEIVED
20 March 2026

REVISED
8 June 2026

ACCEPTED FOR PUBLICATION
29 June 2026

PUBLISHED
13 July 2026

PAPER

VEOS: vision-based vertical electrooculography inference from monocular periocular video for ocular artefact suppression in EEG

Peter Redmond[*], Andrew Fleury and Tomás Ward

Dublin City University, Dublin 9, Ireland

[*] Author to whom any correspondence should be addressed.

E-mail: peter.redmond6@mail.dcu.ie

Keywords: EEG, EOG, ocular artefact suppression, biosignal processing, video-based sensing, wearable EEG, computer vision, temporal convolutional network

Abstract

Objective. Ocular artefacts from blinks and eye movements remain a persistent obstacle to reliable electroencephalography (EEG), particularly when dedicated electrooculography (EOG) electrodes are unavailable or undesirable. We investigate whether monocular periocular video can provide a vertical EOG (VEOG)-like surrogate for ocular artefact suppression in EEG. *Approach.* We present VEOS, a video-to-VEOG inference pipeline based on monocular periocular tracking, canthus-defined geometric normalisation, engineered eyelid and iris features, and temporal modelling with a temporal convolutional network. The inferred VEOG is used as an auxiliary reference in blink-window lagged ridge-regression subtraction of ocular artefacts from EEG. Evaluation used leave-one-subject-out validation on two multimodal datasets: an in-house development dataset (VEOS-Dev; 5 participants with synchronised EEG, EOG and 120 Hz video) and the public Eye–brain–computer interface (BCI) dataset (31 subjects, 63 sessions; high-speed video, EEG and ocular measurements). For centred-window models, a $\pm\ell$ boundary margin prevented temporal leakage between training, validation and test data. *Main results.* On VEOS-Dev, VEOS achieves median Pearson correlation $r = 0.81$ to ground-truth VEOG and median blink-onset timing error of 18 ms. On Eye–BCI, the video-only model achieves median $r = 0.74$, outperforming an eye-tracker baseline. For blink-window cleaning, VEOS suppresses approximately 50% of blink peak-to-peak amplitude on frontal EEG channels, compared with 66% for true VEOG. Cleaning preserves task-relevant EEG structure, including motor imagery (MI) mu/beta rhythms, steady-state visual evoked potentials (SSVEP) spectral signal-to-noise ratio, and P300 morphology. At the participant level, VEOS yields a modest improvement in MI, leaves SSVEP unchanged, and approaches true-VEOG cleaning for P300 spelling (90% vs 91%). *Significance.* These results show that camera-derived VEOG can act as a useful ocular reference for EEG artefact suppression without additional periocular electrodes. The validation was conducted on controlled recordings and should be interpreted as a best-case demonstration rather than a full validation in unconstrained wearable settings.

1. Introduction

Electroencephalography (EEG) remains one of the most practical non-invasive sensing modalities for brain–computer interfaces (BCIs), neuroergonomics and physiological monitoring because it offers millisecond temporal resolution and can be deployed with comparatively lightweight hardware. At the same time, EEG is fragile: neural signals in the microvolt range are readily contaminated by non-neural sources, among which ocular artefacts from blinks and eye movements are among the largest and most systematic [1–4]. These artefacts are particularly problematic in frontal and fronto-central channels and can degrade both signal interpretation and downstream decoding performance.

In laboratory recordings, ocular activity is commonly monitored using periocular electrooculography (EOG), which provides a direct auxiliary reference for regression-based correction and for identifying ocular components in decomposition-based pipelines [1, 4–7]. In many applied settings, however, dedicated EOG electrodes are undesirable or unavailable. Additional sensors increase setup time, reduce comfort, and are difficult to integrate into dry-electrode, low-channel or consumer-form-factor systems.

At the same time, eye and face video are increasingly available in experimental and applied systems. Commodity cameras are now common in laptops, tablets, head-mounted displays and multimodal monitoring platforms, while modern computer-vision methods can recover facial landmarks, pupil position and eyelid motion from ordinary video. These measurements are not electrical, but they are driven by the same underlying ocular kinematics that generate the EOG and its associated contamination in frontal EEG.

This motivates the central question addressed here: can synchronised periocular video provide an EOG-like surrogate that is accurate enough to support ocular artefact suppression in EEG? More specifically, can such a surrogate serve as an auxiliary reference within established EEG-cleaning pipelines, so that ocular artefact handling remains possible even when no physical EOG electrodes are recorded?

We address this question with VEOS (Vision-based EOG Signal inference), a monocular periocular video pipeline that maps landmark-derived ocular features to a VEOG-like waveform and then uses the inferred signal as a reference in blink-window regression-based EEG cleaning. The aim is not to replace all aspects of dedicated EOG, nor to infer cognitive state directly from video, but to test whether camera-derived ocular signals can provide a practically useful surrogate for EEG artefact handling.

VEOS is evaluated at three levels. First, we assess whether monocular periocular tracking yields stable pupil and eyelid features. Second, we test whether these features can be mapped to VEOG using linear and temporal models. Third, we evaluate whether the inferred signal improves ocular artefact suppression and preserves task-relevant EEG structure in downstream analyses. Validation is performed on a small in-house development dataset (VEOS-Dev) and on the public Eye–BCI dataset. Because both datasets were recorded under controlled conditions with research-grade instrumentation, the present work should be interpreted as a feasibility and best-case validation rather than as proof of performance in unconstrained wearable use.

1.1. Contributions

This work makes five main contributions:

- A monocular periocular video pipeline for waveform-level VEOG inference, based on dense face-mesh landmarks, explicit landmark definitions, canthus-based geometric normalisation, and engineered eyelid and iris features.
- A comparison of VEOG inference models ranging from static ridge regression to temporally contextual models, including a temporal convolutional network (TCN) and a hybrid formulation that combines a linear baseline with a learned temporal residual.
- A sensor-minimal ocular artefact suppression framework in which the inferred VEOG is used as an auxiliary reference in blink-window lagged ridge-regression subtraction, allowing established reference-based EEG-cleaning methods to be applied without dedicated periocular electrodes.
- A dual-dataset evaluation on a controlled in-house development dataset (VEOS-Dev) and the independently collected public Eye–BCI dataset, allowing assessment of signal-level performance, dataset dependence and generalisation across hardware and recording conditions.
- A multi-level validation strategy spanning waveform agreement, event-level timing and amplitude metrics, frontal blink suppression, preservation of task-relevant EEG features, and downstream decoding outcomes in motor imagery (MI), steady-state visual evoked potentials (SSVEP) and P300 paradigms.

2. Related work

2.1. Ocular artefact suppression in EEG

Ocular artefacts are among the largest non-neural contaminants in EEG and are particularly problematic in frontal electrodes, where blink- and saccade-related activity can exceed the neural signals of interest by an order of magnitude [1, 2, 4]. Classical reference-based approaches use recorded EOG channels to estimate ocular contributions and subtract them from EEG, with the Gratton–Coles method and related lagged-regression variants remaining widely used because they are simple, interpretable and compatible with modest channel counts [4, 5, 8]. These methods are especially attractive when the goal is not full source separation but stable attenuation of ocular transients with predictable failure modes.

Component-based methods such as independent component analysis (ICA) and related blind-source-separation approaches can also isolate ocular activity [6, 7]. However, they generally require sufficient channel density, careful component selection, and can be harder to deploy robustly in low-channel, mobile or consumer-grade systems [1, 3]. In such settings, reference-based approaches remain attractive, provided that a suitable ocular reference is available [1, 4, 5]. More recently, deep-learning approaches have been applied directly to EEG artefact removal. Convolutional, recurrent and generative-adversarial architectures have been trained on benchmark corpora such as EEGdenoiseNet to separate ocular and muscular contamination from neural activity [3, 9–11], and broader reviews have surveyed the rapidly growing literature on deep learning for EEG processing [12]. These methods can learn flexible non-linear mappings, but they typically still require sensor-based measures of ocular activity for either training supervision or runtime reference, and their performance in low-channel wearable settings remains incompletely characterised. VEOS is complementary to this line of work: by producing a camera-derived ocular reference, it can in principle be used either as an input to a deep artefact-removal model or, as here, as a reference in a classical regression-based pipeline.

2.2. Video, eye tracking and ocular surrogates

Eye trackers and cameras have long been used to estimate gaze direction, fixations and blink events [13]. More recently, dense landmark models and appearance-based computer-vision pipelines have enabled robust estimation of eyelid and iris motion from commodity cameras [14, 15]. These developments are relevant because the same eye and eyelid kinematics that are visible in video also drive the EOG and associated frontal EEG artefacts.

Most existing camera-based ocular pipelines, however, stop at event detection, gaze estimation or behavioural descriptors rather than reconstructing an EOG-like waveform suitable for downstream EEG artefact modelling. Prior work related to VEOS has shown the feasibility of inferring EOG-related information from video or of using computer vision for artefact tagging [16, 17], but the literature remains sparse on waveform-level surrogate generation explicitly targeted at reference-based EEG cleaning.

2.3. Positioning of VEOS

VEOS sits between two established approaches. At one end are sensor-based methods that rely on recorded EOG and offer strong ocular references at the cost of extra electrodes. At the other are purely EEG-based methods that attempt to identify ocular contamination without any dedicated reference channel. VEOS instead asks whether a camera-derived signal can provide a middle ground: lower burden than periocular electrodes, but more informative for ocular artefact handling than EEG alone.

The contribution is therefore not a fundamentally new artefact-removal theory. Rather, it is a practical and evaluated system for waveform-level VEOG inference from monocular periocular video, coupled to a conventional regression-based EEG-cleaning pipeline. The central gap addressed is whether such a camera-derived surrogate is accurate enough, not merely to detect blinks, but to support meaningful ocular artefact suppression while preserving relevant EEG structure.

3. Materials and methods

3.1. System overview

Figure 1 summarises VEOS: monocular periocular video is converted into a VEOG-like reference signal, which is then used for blink-window reference subtraction in EEG and evaluated in BCI pipelines.

3.2. Datasets

3.2.1. VEOS-Dev in-house dataset

VEOS-Dev is an in-house multimodal dataset collected for method development and internal validation. It contains recordings from five participants (3 women, 2 men; ages 21–55), each completing 1–3 sessions. Sessions include follow-the-dots gaze calibration, a game-based engagement task with increasing difficulty, and ERP-style visual blocks designed to elicit fixation, blink suppression, rebound blinks, and gaze shifts. Each session contains synchronised 32-channel EEG from an ANT Neuro amplifier, bipolar horizontal and vertical EOG, 120 Hz 720p face video, and an event-marker stream. The development platform uses a common hardware trigger and a shared LSL time base so that EEG, EOG, video timing and task markers remain aligned across modalities. No dedicated eyelid EMG channels or hardware eye tracker are present in VEOS-Dev; periocular muscle activity is therefore observed only indirectly through its effects on EOG and EEG.

3.2.2. Eye–BCI public dataset

Eye–BCI is an independently collected public multimodal dataset that we use as an external validation set [18]. It contains recordings from 31 healthy volunteers (11 women, 20 men; mean age approximately 29 years) acquired over 63 sessions, amounting to roughly 46 hours of usable data. Each session includes four BCI paradigms in randomised order: MI, motor execution, SSVEP, and P300 spelling. Recordings were obtained in an electromagnetically shielded, sound-attenuated chamber. Participants sat approximately 80 cm from a 23 inch display with an integrated Tobii TX300 eye tracker operating at 300 Hz, with head position stabilised by a chin rest. Electrophysiology was recorded with a 65-channel Quik-Cap connected to a SynAmps2 system, providing EEG together with VEOG and eyelid EMG. A high-speed Phantom camera recorded a single periocular view at 150 fps. Because Eye–BCI provides only the left-eye periocular view, we use the left eye as the tracked eye throughout the paper for cross-dataset consistency.

3.2.3. Ethics

VEOS-Dev was collected under institutional ethics approval and all participants provided written

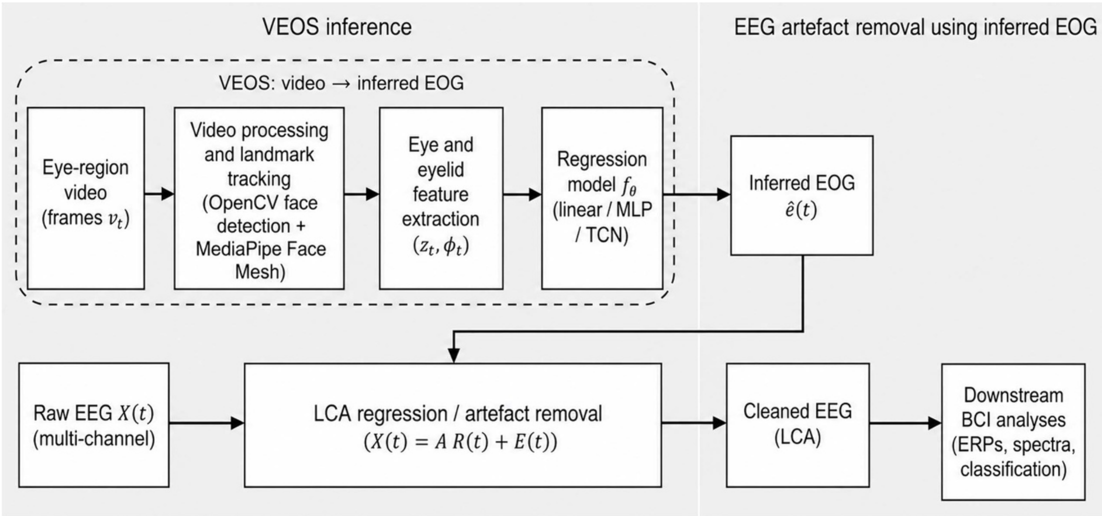

Figure 1. VEOS overview: periocular video is processed into landmark-derived features, mapped to a VEOG-like reference, and used for blink-focused reference subtraction in EEG before downstream BCI decoding. All results reported in this paper use the reference-subtraction (LCA) path. An optional extension to high-density EEG via ICA is discussed in section 5.4.

informed consent. Data collection was conducted under the authors' prior affiliation; analysis and manuscript preparation were completed at the current affiliation.

3.3. Video processing, landmarks and normalisation

3.3.1. Periocular region of interest (ROI) and tracking

A face/eye ROI is obtained using a Haar cascade detector (OpenCV) [19, 20]. Face-mesh landmarks are then estimated on the ROI for every frame. The ROI is re-detected every $N_{\text{redetect}} = 30$ frames, and its bounding-box coordinates are exponentially smoothed between detections with smoothing factor $\alpha = 0.3$ (that is, $\text{ROI}_t = \alpha\,\text{ROI}_t^{\text{raw}} + (1-\alpha)\,\text{ROI}_{t-1}$) to reduce jitter. Frames are marked invalid if no face is detected or if geometric plausibility checks fail (section 3.3.4). When both eyes are visible, bilateral landmarks are used only to stabilise ROI localisation. All downstream processing, including normalisation, feature extraction and VEOG inference, uses a single tracked eye. We use the left eye throughout for two reasons: Eye–BCI provides only a left-eye periocular view, and using the same eye across both datasets avoids introducing an additional eye-selection confound. This choice does not assume perfect bilateral symmetry. Rather, any asymmetric blink behaviour or unilateral tracking error appears as modelling error and is treated as part of the method's present limitation.

3.3.2. Face mesh model and landmark indices

We use the MediaPipe Face Mesh model with refined landmarks enabled to obtain iris landmarks [14, 15]. For reproducibility, VEOS uses the following landmark indices:

- **Canthi:** $(33, 133)$ and $(362, 263)$ (used for ROI initialisation when visible).
- **Tracked eye eyelid points for scaled EAR:** $[33, 159, 158, 133, 153, 145]$ (or the mirrored set for the other eye).
- **Tracked eye (left) iris landmarks:** we use the five-point iris set corresponding to the tracked eye (MediaPipe provides two sets: 468–472 and 473–477); the iris centre is the mean of the five points.

Figure 2 illustrates the retained landmarks and derived geometry.

3.3.3. Similarity normalisation

To reduce sensitivity to camera placement and head motion, we normalise 2D landmark coordinates using a similarity transform defined by the tracked eye's canthal geometry. Let $p_i(t) \in \mathbb{R}^2$ be a landmark coordinate at frame t. Let $p_{\text{out}}(t)$ and $p_{\text{in}}(t)$ denote the outer and inner canthi of the tracked eye. We define the tracked-eye canthal mid-point

$$c(t) = \tfrac{1}{2}\left(p_{\text{out}}(t) + p_{\text{in}}(t)\right), \tag{1}$$

and a tracked-eye scale

$$d(t) = \|p_{\text{in}}(t) - p_{\text{out}}(t)\|_2. \tag{2}$$

We remove in-plane rotation by aligning the canthal axis to the x-axis: let $\phi(t) = \text{atan2}(\Delta y, \Delta x)$ for $\Delta = p_{\text{in}}(t) - p_{\text{out}}(t)$, and let $R(-\phi)$ be the 2×2 rotation matrix. The normalised coordinate is then

$$\tilde{p}_i(t) = R(-\phi(t))\,\frac{p_i(t) - c(t)}{d(t)}. \tag{3}$$

All geometric features are computed in this normalised coordinate system.

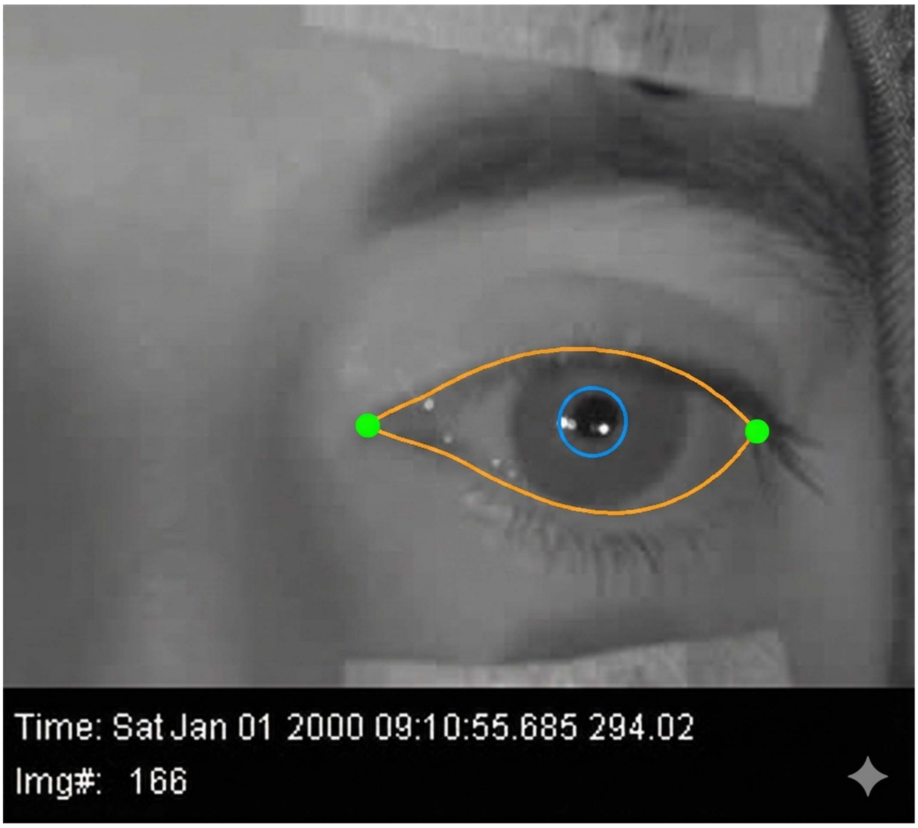

Figure 2. Periocular landmarks used by VEOS (example frame). Eye corners, eyelid landmarks and iris landmarks define pupil displacement and scaled eye aperture features. Image is from Eye–BCI dataset.

3.3.4. Failure handling

A frame is considered *low-confidence* if any of the following holds: (i) no face mesh is returned; (ii) $d(t)$ is outside $[\mu_d - 3\sigma_d,\ \mu_d + 3\sigma_d]$ for the session (implausible scale); (iii) the scaled eye aperture (section 3.4) is outside a plausible range. Where available, we additionally compute a per-frame confidence score using the MediaPipe `NormalizedLandmark.presence` field; if not provided by the runtime, this term is omitted and confidence is based purely on (i)–(iii). Short gaps ($\leqslant$5 frames) are linearly interpolated; longer gaps are excluded from training and evaluation. In the retained analysed recordings, no material tracking dropouts were observed, so invalid-frame exclusion and short-gap interpolation were negligible in practice (table 1).

3.4. Feature engineering

VEOS derives interpretable ocular features from normalised landmarks:

1. **Pupil displacement.** Let $u(t)$ be the tracked-eye iris centre (mean of the tracked-eye 5-point iris landmark set). Define a canthus-based eye coordinate system with origin at the canthal mid-point and vertical axis aligned with the image axes. We use the vertical displacement

$$y(t) = \tilde{u}_y(t), \tag{4}$$

and optionally the horizontal displacement $\tilde{u}_x(t)$ for saccade-related dynamics.

2. **Scaled eye aperture (scaled EAR)** [21]**.** For an eye with landmarks $(p_1, p_2, p_3, p_4, p_5, p_6)$ corresponding to $(33, 159, 158, 133, 153, 145)$ (or the mirrored set), we compute

$$\mathrm{EAR}(t) = \frac{\|\tilde{\mathbf{p}}_{p_2}(t) - \tilde{\mathbf{p}}_{p_6}(t)\|_2 + \|\tilde{\mathbf{p}}_{p_3}(t) - \tilde{\mathbf{p}}_{p_5}(t)\|_2}{2\|\tilde{\mathbf{p}}_{p_1}(t) - \tilde{\mathbf{p}}_{p_4}(t)\|_2}. \tag{5}$$

Because equation (3) is a similarity transform, the EAR is invariant to the choice of $d(t)$.

3. **Temporal derivatives.** We compute first and second differences of each feature using a centred finite-difference scheme on a smoothed version (Savitzky–Golay, window 11 frames, polynomial order 2) [22].

The final feature vector $x_t \in \mathbb{R}^d$ concatenates tracked-eye features and their temporal derivatives.

3.4.1. Video-only blink detection and windowing

To define blink-centred inference and cleaning windows without using recorded EOG, VEOS uses a video-only blink detector based on tracked-eye scaled EAR [21]. We compute $\mathrm{EAR}(t)$ for the tracked eye and form a binary closure indicator

$$I(t) = \mathbb{1}\{\mathrm{EAR}(t) < \tau\}. \tag{6}$$

Table 1. Summary of low-confidence and interpolated frames across the retained analysed recordings (median per session, with interquartile range in brackets). Low-confidence frames are those failing the criteria of section 3.3.4; short-gap interpolated frames are sequences of five or fewer consecutive low-confidence frames reconstructed by linear interpolation.

Dataset	Low-confidence frames (%)	Interpolated frames (%)
VEOS-Dev	<0.5 [0.1–0.9]	<0.3 [0.0–0.5]
Eye–BCI	<0.2 [0.0–0.4]	<0.1 [0.0–0.2]

A candidate eye-closure event is a contiguous run of frames where $I(t) = 1$. We classify such an event as a blink if it represents a transient thresholded closure, that is, its duration lies within a short range: at least $T_{\min} = 3$ frames and not exceeding a maximum duration $T_{\max} = 15$ frames (approximately 100–125 ms at 120–150 Hz). Here, duration refers to the interval for which the scaled EAR remains below threshold, not to the full blink from onset to complete reopening. The upper bound was chosen to reject prolonged eye closures and gross tracking failures while retaining the compact low-EAR portion of ordinary blinks. Blink onset is defined at the first downward threshold crossing of the detected run. Thresholds are set per session using the 10th percentile of $\mathrm{EAR}(t)$ in non-blink calibration segments:

$$\tau = Q_{0.10}\left(\mathrm{EAR}\right). \tag{7}$$

Blink-centred windows of length $W_{\text{blink}} = \pm 250$ ms around detected onsets define (i) blink-window VEOG inference metrics and (ii) the windows used to train the EEG-cleaning regression model. This duration was selected to cover the main blink transient and immediate lid-recovery phase while excluding unrelated distant activity: typical spontaneous blinks have a full closure-to-reopen duration of approximately 200–400 ms [21], so a ± 250 ms window centred on onset captures the dominant transient plus a short margin on either side. Preliminary sweeps during development confirmed that performance was stable to moderate changes in this value; a systematic ablation of W_{blink} was beyond the scope of the present study and is noted as future work in section 5.3.

3.5. VEOG target construction

Bipolar VEOG is high-pass filtered at 0.1 Hz and low-pass filtered at 15 Hz (4th-order Butterworth, zero-phase) [23, 24] to remove drift and high-frequency noise. The signal is resampled to the video frame timestamps using polyphase anti-aliasing resampling, yielding a target $v(t)$ at the video rate. For learning, targets are standardised per participant; original units are restored when reporting error metrics in physical units.

3.6. VEOG inference models

We evaluate four models mapping feature sequences to VEOG.

3.6.1. Baselines

In what follows, $v(t)$ denotes the filtered, resampled VEOG target defined in section 3.5, and $\hat{v}(t)$ denotes the model prediction.

(i) **Constant mean:** $\hat{v}(t) = \bar{v}$.

(ii) **Static ridge regression:** $\hat{v}(t) = \mathbf{w}^\top \mathbf{x}_t + b$, fit with ridge penalty [25].

(iii) **Windowed multilayer perceptron (MLP):** a feedforward network operating on a centred temporal feature window. The MLP input is the concatenated feature vector $\mathbf{z}_t = [\mathbf{x}_{t-\ell}^\top \dots \mathbf{x}_{t+\ell}^\top]^\top$, using the same centred context length as the temporal comparison models unless otherwise stated. A typical configuration uses one or two fully connected hidden layers with ReLU activations (for example 128 and 64 units), optional dropout, and a single linear output neuron for scalar VEOG prediction.

3.6.2. TCN model (VEOS-TCN)

The main model is a TCN [26] operating on a centred window of length $L = 41$ frames (≈ 0.34 s at 120 Hz; ≈ 0.27 s at 150 fps). This duration was chosen to cover the principal short-term dynamics of blinks and vertical gaze transients while keeping the model local in time. The odd window length ensures a symmetric centred target frame. The architecture is fixed as:

- 5 residual blocks with dilations $\{1, 2, 4, 8, 16\}$;
- 1D convolution kernel size $k = 5$;
- 64 channels per block;
- weight normalisation, dropout $p = 0.1$ [27], and ReLU activations [28].

The output is the predicted VEOG at the centre frame.

3.6.3. Hybrid linear baseline + TCN residual (VEOS-Hybrid)

To improve interpretability and separate coarse subject-independent structure from subject-specific residual dynamics, we also evaluate a hybrid formulation:

$$\hat{v}(t) = \underbrace{a_0 + a_1 y(t)}_{\text{linear baseline}} + \underbrace{f_\theta(\mathbf{x}_{t-\ell:t+\ell})}_{\text{TCN residual}}, \tag{8}$$

where $y(t)$ is the vertical iris displacement, and f_θ is the same TCN as above predicting a residual. In the broader VEOS framework, the linear baseline can also be interpreted as the pooled subject-independent forward model, while short subject-specific scale and offset calibration can be applied where calibration is acceptable. On VEOS-Dev, VEOS-Hybrid provides the best median performance. On Eye–BCI, VEOS-TCN is retained as the default video-only model because it avoids per-user baseline calibration.

3.6.4. Training recipe

Models are trained with mean-squared error loss using the Adam optimiser [29] with early stopping on validation loss. Unless otherwise stated, we use learning rate 3×10^{-4}, batch size 64, weight decay 10^{-4}, and patience 10. Learning rate and regularisation strength are selected on validation data only, with gradient clipping applied where needed for stable temporal-model training. Inputs are z-scored per feature channel using statistics computed on the training partition only. The temporal models reported here were trained once per fold; a dedicated multi-seed stability analysis was not part of the present study and remains for future work.

3.7. Splits, validation and leakage control

For VEOG inference, we evaluate:

- **VEOS-Dev:** leave-one-subject-out (LOSO) over 5 participants, yielding 5 folds.
- **Eye–BCI:** LOSO over 31 participants, yielding 31 folds.

In each fold, the held-out participant is used for testing only and is never used for hyperparameter selection, normalisation, early stopping, or model selection. The remaining participants form the development portion of the fold. Within that development portion, validation data are created without subject overlap with the test participant. Runs or sessions from the training participants are split into training and validation subsets on a per-participant basis, with contiguous temporal blocks used where required for centred-window models, and hyperparameters are selected using validation performance only.

For models that use centred temporal windows, leakage can arise if neighbouring samples straddle train, validation, or test boundaries. To prevent this, we discard any window whose centre lies within ℓ frames of a split boundary, where $2\ell + 1 = L$. Thus, no prediction window can contain samples from more than one partition.

All feature standardisation constants are computed from training data only within each fold. For EEG cleaning and downstream decoding, all cleaning parameters, including ridge weights and any normalisation constants, are fit on training data only and then applied unchanged to held-out test trials. When considering cross-dataset transfer, no Eye–BCI data are used for training except where explicitly noted for short calibration of hybrid baseline parameters.

3.8. Blink-window reference subtraction in EEG

We use the inferred VEOG as a reference in a lagged, ridge-regularised regression model, in the style of Gratton–Coles, to estimate ocular contribution in each EEG channel; we refer to this family of reference-subtraction methods as linear component analysis (LCA) [1, 4, 5]. Let $\mathbf{r}(t) \in \mathbb{R}^K$ be a vector of K time-lagged samples of the reference (VEOG), spanning $\pm$ 100 ms. For each EEG channel $e_c(t)$ we fit

$$e_c(t) \approx \mathbf{w}_c^\top \mathbf{r}(t) + b_c \tag{9}$$

using only blink-centred windows defined by the video-only detector. The estimated ocular component $\hat{o}_c(t) = \mathbf{w}_c^\top \mathbf{r}(t) + b_c$ is subtracted to obtain cleaned EEG $\tilde{e}_c(t) = e_c(t) - \hat{o}_c(t)$. Ridge regularisation strength is selected once per dataset on training folds.

3.9. BCI decoding

We evaluate decoding using standard pipelines: (MI; band-pass filtering, common spatial patterns (CSP) and linear discriminant analysis (LDA)) [30, 31], steady-state visual evoked potentials (SSVEP; canonical correlation analysis (CCA)) [32], and P300 spelling (event-related potential features and LDA) [33].

For each paradigm and cleaning condition, decoding accuracy was computed separately for each held-out participant in the LOSO evaluation. Where a participant contributed multiple sessions, session-wise accuracies were aggregated within participant before group summarisation. Reported summary values are medians across participants, with interquartile ranges.

4. Results

4.1. Blink detector performance

On Eye–BCI, the video-only blink detector (tracked-eye scaled-EAR thresholding with a duration constraint) [21] achieves median frame-wise accuracy 0.95, precision 0.90, recall 0.92 and $F1$ 0.91 across subjects; 94% of detected blink onsets fall within ± 20 ms of VEOG-defined onsets. On VEOS-Dev we observe qualitatively similar behaviour using the same detector, but refrain from over-interpreting quantitative blink-detector statistics due to the small sample size.

4.2. VEOG inference quality

4.2.1. VEOS-Dev (LOSO)

Table 2 compares VEOG inference models on VEOS-Dev. Temporal context is essential: windowed MLPs substantially outperform static models, and TCNs further improve performance. VEOS-Hybrid provides the best overall metrics on VEOS-Dev.

4.2.2. Eye–BCI (LOSO)

Table 3 reports VEOG inference on Eye–BCI. Video-only VEOS-TCN achieves strong correlation and outperforms an eye-tracker-only baseline. Fusing video and eye tracking provides a modest additional gain, while eyelid EMG provides an upper bound (not sensor-minimal).

Table 2. VEOG prediction on VEOS-Dev (median over subjects, LOSO). r: Pearson correlation; nRMSE: normalised RMSE. Bold indicates best results.

Model	r (all)	r (non-blink)	r (blink win.)	nRMSE
Constant mean	0.00	0.00	0.00	1.00
Static linear (video)	0.41	0.45	0.28	0.77
Windowed MLP (video)	0.74	0.79	0.55	0.54
VEOS-TCN (video)	0.79	0.83	0.60	0.49
VEOS-Hybrid (baseline + TCN residual)	**0.81**	**0.85**	**0.63**	**0.47**

Table 3. VEOG prediction on Eye–BCI (median over subjects, LOSO). Video-only models correspond to VEOS-TCN. Bold indicates best results.

Input / Model	r (all)	r (non-blink)	r (blink win.)	nRMSE
Constant mean	0.00	0.00	0.00	1.00
Eye-tracker only (MLP)	0.66	0.72	0.45	0.58
Video only (VEOS-TCN)	0.74	0.79	0.56	0.51
Video + eye tracker (TCN)	0.77	0.82	0.58	0.48
Video + eyelid EMG (upper bound)	**0.84**	**0.88**	**0.71**	**0.40**

4.3. Sensitivity to temporal context and model structure

Although this study was not designed as a full ablation benchmark, the progression from static linear regression to windowed MLPs, TCNs, and the hybrid linear-plus-residual model provides a limited sensitivity analysis of two central design choices: the use of temporal context and the value of modelling residual non-linearity.

On VEOS-Dev (table 2), performance improves substantially when moving from the static linear model to the windowed MLP, indicating that instantaneous geometry alone is insufficient to capture the full structure of VEOG. The further improvement from the MLP to the TCN suggests that local temporal organisation, including blink envelopes and short-lived ocular transients, is better modelled when the architecture operates directly on feature sequences rather than on a flattened temporal window. The best performance is obtained with VEOS-Hybrid, consistent with the idea that a simple linear relation between vertical iris displacement and VEOG captures the coarse subject-independent structure, while the residual model accounts for non-linear and subject-specific corrections.

This pattern is consistent with the qualitative interpretation of the VEOS framework. The similarity-normalised geometric features provide a stable representation of eye and eyelid motion, but the electrical ocular signal is not determined by static pose alone. Blink morphology, in particular, depends on short-term temporal evolution and on factors such as eyelid closure dynamics and subject-specific amplitude scaling. The observed ranking of models therefore supports the choice of a temporally contextual main model and motivates the use of the hybrid formulation when a small amount of calibration is acceptable.

A fuller ablation of feature subsets, normalisation choices, and blink-window parameters remains for future work. However, the present comparison indicates that the dominant performance gain in VEOS arises from introducing temporal context, while the additional but smaller gain from VEOS-Hybrid reflects the benefit of separating a physically interpretable baseline from residual non-linear correction.

4.3.1. Feature and normalisation ablation

To give a minimal indication of the contribution of two further design choices, we retrained the VEOS-TCN model on Eye–BCI under two ablation conditions: (a) without canthus-based similarity normalisation (raw pixel-coordinate landmarks with per-session standardisation only), and (b) without temporal derivative features (pupil displacement and EAR only, no Savitzky–Golay-smoothed differences). Results are reported in table 4.

Both ablations produce measurable degradations, supporting the inclusion of similarity normalisation and temporal derivatives as standard components of the VEOS feature pipeline.

4.4. Lag and amplitude agreement

Correlation alone does not capture lag and amplitude mismatches that can affect cleaning. Table 5 reports event-level agreement (median across subjects): blink-onset timing error and blink peak-to-peak amplitude ratio.

4.5. EEG cleaning with inferred VEOG

Figure 3 illustrates blink suppression on a representative Eye–BCI session for frontal EEG (Fp1) using different references. Across sessions, as detailed in

Table 4. Minimal ablation of VEOS-TCN on Eye–BCI (median over subjects, LOSO). Removing similarity normalisation or temporal-derivative features reduces VEOG prediction quality relative to the full model.

Configuration	r (all)	nRMSE
VEOS-TCN (full)	0.74	0.51
− similarity normalisation	0.67	0.58
− temporal derivatives	0.70	0.55

Table 5. Event-level agreement between inferred VEOG and ground-truth VEOG on Eye–BCI (median across subjects). Timing error is the absolute onset difference (ms). Amplitude ratio is inferred/true blink peak-to-peak.

Method	Blink onset error (ms)	Blink amplitude ratio
Video-only (VEOS-TCN)	18	0.72
Eye-tracker only	20	0.58

table 6, VEOS reference subtraction reduces blink peak-to-peak amplitude by approximately 50% on Fp1 (true VEOG: 66%). Figure 4 shows the distribution of blink peak-to-peak reduction across sessions.

4.6. Preservation of task-relevant EEG features

To assess whether cleaning preserves relevant neural features, we evaluate the task-defining signal components reported in standard BCI decoding pipelines for each paradigm: (i) MI mu/beta band power and CSP patterns, which underlie sensorimotor-rhythm decoding [30, 34]; (ii) SSVEP spectral signal-to-noise ratio (SNR) around stimulus frequencies, which underlies CCA-based decoding [32]; (iii) P300 ERP morphology (amplitude and latency), which underlies ERP-feature LDA decoding [33]. Here, SNR denotes SNR and ERD/ERS denotes event-related desynchronisation/synchronisation. Figure 5 presents representative results, and table 7 summarises median relative changes in key task features.

4.7. BCI decoding outcomes

Table 8 reports median decoding accuracy on Eye–BCI under the participant-level LOSO summary.

The participant-level pattern is straightforward. For MI, true VEOG yields a modest improvement over having no explicit ocular channel (75% versus 71%), and VEOS-based cleaning recovers most of that benefit (74%). For SSVEP, all four conditions perform similarly (93%, 93%, 92%, 92%), indicating that VEOS-based cleaning does not materially alter a paradigm whose narrow-band occipital response is relatively insensitive to frontal blink artefacts. This is consistent with the near-zero median change in the simple SSVEP SNR summary in table 7: that summary is a preservation check at the stimulus frequency, whereas CCA-based decoding depends on broader multichannel trial structure. For P300 spelling, VEOS again tracks the benefit of true VEOG closely (90% versus 91%), consistent with the role of VEOS as a blink-focused vertical ocular surrogate.

Taken together, these results suggest that VEOS is most useful where blink-related contamination is a practical limitation, while remaining neutral rather than harmful in paradigms such as SSVEP where frontal ocular artefacts are less critical to decoding performance.

5. Discussion

VEOS demonstrates that monocular periocular video can provide a useful VEOG-like reference for ocular artefact suppression in EEG without additional skin-contact sensors. The strongest gains occur in frontal blink transients, where VEOS achieves a large fraction of the benefit of true VEOG. Temporal modelling is critical: TCNs capture blink dynamics that are not represented by instantaneous geometry alone. At the same time, the signal-level results show that VEOS does not fully reproduce true VEOG amplitude and morphology, particularly near full eyelid closure, where landmark quality degrades and blink-related ocular dynamics are most difficult to observe directly from video. This explains why VEOS is effective but not equivalent to a recorded ocular channel.

The existing model comparison also provides a limited sensitivity analysis of the VEOS design. The large gap between the static linear model and the temporally contextual models shows that waveform inference depends strongly on short-range temporal structure rather than on framewise geometry alone. The smaller but consistent gain from the hybrid model suggests that much of the signal can be explained by a simple vertical-position baseline, while the remaining error is concentrated in nonlinear and subject-specific residuals. This supports the present modelling choices, even though a fuller ablation of normalisation, feature subsets, and blink-window parameters remains to be carried out.

5.1. Why VEOS helps most in blink-sensitive paradigms

The downstream results are best understood in terms of what each paradigm demands from the ocular reference. For P300 spelling, the main practical problem is often transient blink contamination around stimulus onset and within the ERP latency range. Here, VEOS is well matched to the task: it is a blink-focused, vertical ocular surrogate, and it removes enough of the dominant blink-related structure that the resulting P300 performance is close to true-VEOG cleaning (90% versus 91%).

For MI, VEOS recovers most of the modest benefit obtained with true VEOG. This pattern is consistent with the idea that ocular cleaning helps

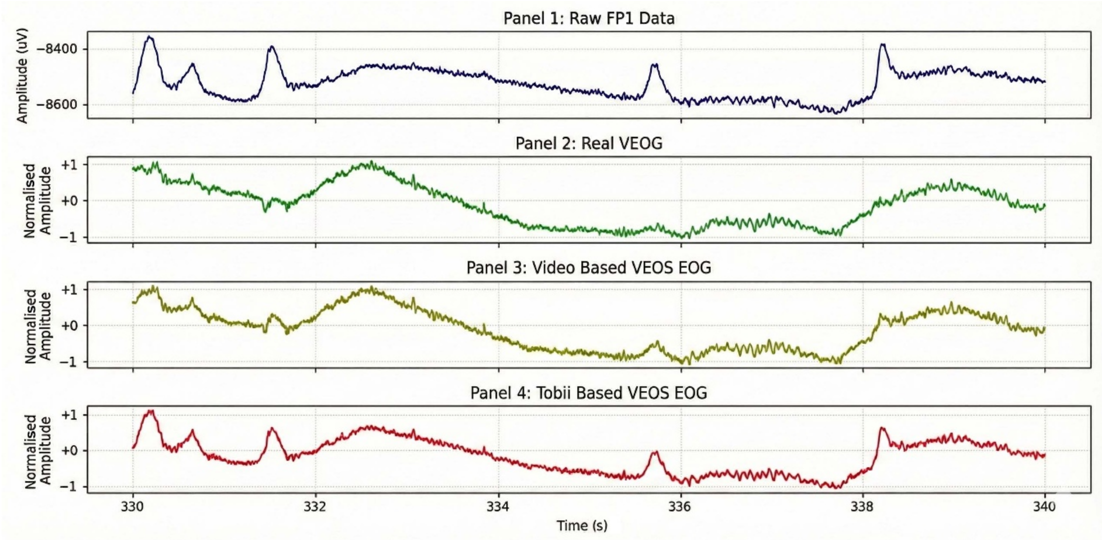

Figure 3. Blink-window reference subtraction on Fp1. **Panel 1:** Uncleaned raw EEG data. **Panel 2:** EEG cleaned using the ground-truth VEOG reference. **Panel 3:** EEG cleaned using the VEOS (video-only) inferred reference. **Panel 4:** EEG cleaned using the eye-tracker inferred reference. Time on the *x*-axis is measured from the start of the recording (shown here for a 330–340 s excerpt).

Table 6. Cleaning performance on Fp1 using different ocular references (median over Eye–BCI sessions). Corr.: correlation between EEG and true VEOG in blink windows; P2P: peak-to-peak amplitude.

	Correlation			Blink P2P (μV)		
Reference	Pre	Post	Δ(%)	Pre	Post	Δ(%)
True VEOG	0.58	0.12	−79	220	75	−66
VEOS (video)	0.58	0.19	−67	220	110	−50
Eye-tracker	0.58	0.46	−21	220	190	−14

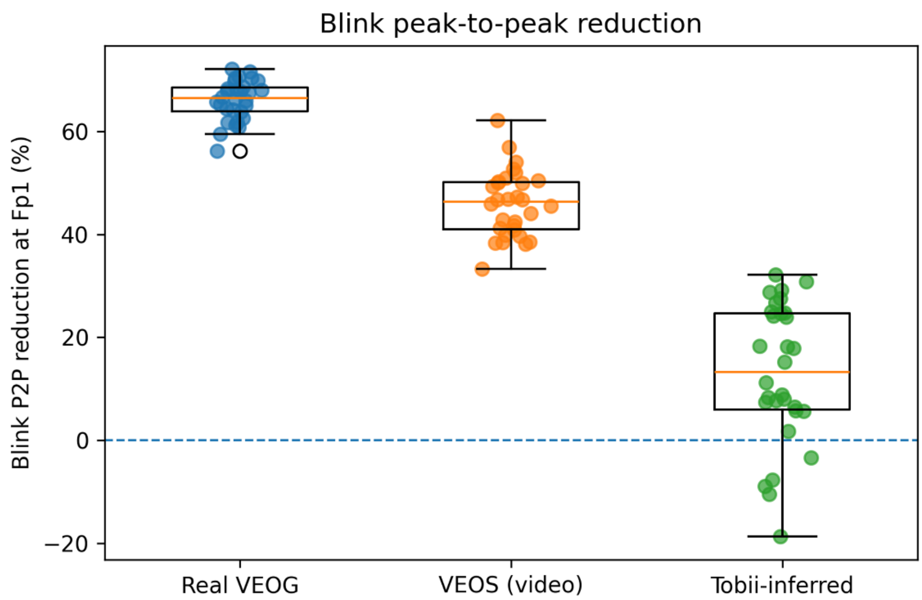

Figure 4. Distribution of blink peak-to-peak amplitudes across sessions (Eye–BCI). VEOS achieves substantial suppression, approaching true VEOG.

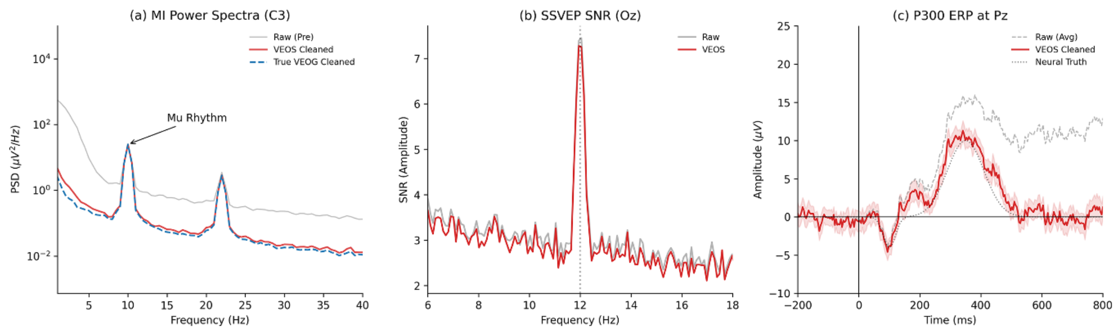

Figure 5. Task-feature preservation checks (representative participant/session examples): (a) MI mu/beta power spectra pre and post-cleaning; (b) SSVEP frequency-domain signal-to-noise ratio (SNR) around the stimulus frequency; (c) P300 grand-average ERP at a representative parietal electrode, including the residual artefact removed by cleaning. VEOS suppresses blink-related transients while preserving task-relevant spectral and temporal structure.

Table 7. Summary of task-feature preservation on Eye–BCI (median across participants). Values are relative changes induced by cleaning compared with uncleaned EEG (smaller is better). Where only bounds were observed, we report '<' thresholds.

Paradigm / Feature	True VEOG cleaning	VEOS cleaning
MI: mu/beta ERD/ERS ratio change	<5%	<5%
SSVEP: spectral signal-to-noise ratio (SNR) change at stimulus frequencies	small (near-zero)	small (near-zero)
P300: peak amplitude change (parietal electrode)	<3%	<3%

mainly by reducing contamination before spatial-filter estimation rather than by dramatically changing the underlying sensorimotor rhythms. The remaining small gap to true VEOG is compatible with residual ocular variance that is only partially represented by the inferred reference.

For SSVEP, performance is essentially unchanged across conditions. This is consistent with the expectation that narrow-band occipital responses are relatively insensitive to frontal blink artefacts, provided that the cleaning stage does not distort the task-relevant spectral structure. In this setting, the main value of VEOS is not a large gain in accuracy but the fact that it does not materially harm the decoder while still providing a practical ocular reference for blink-related contamination.

These paradigm-dependent results help clarify what VEOS is, and what it is not. VEOS should be viewed as a practical surrogate ocular channel that is particularly effective for blink-dominated contamination, not as a full replacement for recorded EOG across all ocular behaviours.

5.2. Practical considerations

VEOS is intended as a sensor-minimal ocular reference rather than as a fully optimised real-time deployment system. Accordingly, the trade-off addressed in this paper is reduced sensor burden rather than demonstrated computational efficiency relative to signal-only ocular cleaning methods. In the present implementation, face-mesh inference is the main computational bottleneck; the downstream landmark normalisation, feature extraction and TCN inference are comparatively lightweight, as the TCN uses only 5 residual blocks of 64 channels on a 41-frame window. We did not perform a formal throughput, latency, power, or mobile-device benchmark in this study, and therefore make no quantitative claim about suitability for online or battery-constrained deployment. We note, however, that published mobile-optimised implementations of MediaPipe Face Mesh run at real-time rates on commodity smartphone hardware [15], which suggests that mobile deployment is plausible but requires direct measurement. The current results should be interpreted as an offline feasibility study under controlled recording conditions. Future work should quantify runtime, memory footprint, robustness to glasses and illumination changes, and the viability of on-device or privacy-preserving implementations. A further practical requirement is the per-session calibration of the blink-detection threshold. As described in section 3.4, the EAR closure threshold τ is set to the 10th percentile of the tracked-eye EAR computed over non-blink calibration segments. In effect, VEOS requires an estimate of the session-specific EAR distribution before blink detection can be performed. This step requires a short period of natural fixation or resting-state video recorded prior to the main task. The calibration data required are video-only; no concurrent EEG or EOG recording is necessary at this stage. This requirement is modest and consistent with

Table 8. Median BCI accuracy (%) on Eye–BCI under different cleaning conditions, summarised across held-out participants in the LOSO evaluation. Bold indicates best results.

Paradigm	None	True VEOG	VEOS (video)	Eye tracker
Motor imagery (L/R)	71	**75**	74	70
SSVEP	**93**	**93**	92	92
P300 spelling	88	**91**	90	86

the baseline periods routinely included in BCI experimental protocols for spatial filter estimation and signal normalisation. The threshold is applied per session rather than per participant, which accommodates session-to-session variation in camera placement and ambient illumination. The sensitivity of detection performance to calibration segment length and to the percentile parameter was not systematically evaluated in the present study and represents a topic for future ablation work.

5.3. Limitations

Several limitations should be acknowledged. First, the present evaluation focuses exclusively on vertical EOG and blink artefacts. While VEOS extracts horizontal iris displacement features that could in principle support HEOG inference and saccade-related artefact suppression, the Eye–BCI dataset provides only VEOG ground truth, precluding quantitative evaluation of horizontal artefact correction. Saccades produce substantial artefacts on lateral frontal channels and may affect paradigms differently than blinks; future work should evaluate HEOG inference on datasets with horizontal EOG references.

Second, the VEOS-Dev dataset comprises only five participants, limiting the generalisability of within-dataset findings. The VEOS-Dev findings should therefore be interpreted as preliminary within-dataset evidence, with the stronger test of external validity coming from the larger independent Eye–BCI evaluation. Although Eye–BCI provides a larger sample (31 subjects), it was collected under controlled laboratory conditions with a high-speed camera (150 fps) and fixed head position. Performance on commodity webcams (30–60 fps), under variable lighting, or with unconstrained head movement remains to be characterised. In addition, we did not perform a dedicated multi-seed stability analysis of the temporal models, so the reported fold-wise results should be interpreted as performance estimates under the stated training configuration rather than as a full optimisation-stability study.

Third, landmark localisation degrades substantially near eyelid closure, which coincides precisely with the blink events VEOS aims to capture.

Finally, this study employed offline processing throughout. In its current form, VEOS uses centred inference windows (section 3.6.2) and a $\pm$ 100 ms lagged reference model for cleaning, which are non-causal; an online implementation would require a causal variant or would incur a fixed algorithmic delay. Real-time capability was not evaluated.

5.4. Extension to high-density EEG and ICA

The VEOS framework can in principle be extended to high-density EEG recordings in which ICA is viable. In such settings, the inferred signal $\hat{e}(t)$ could be used as a correlation reference to score ICA components: components whose activations exhibit high Pearson correlation with $\hat{e}(t)$ across blink windows, and whose scalp topographies are consistent with a frontal or peri-ocular distribution, would be candidates for removal as ocular sources, in the manner described by Jung *et al* [6].

However, this use of $\hat{e}(t)$ raises a caveat that the present study does not resolve. A correlation-based scoring criterion cannot, on its own, distinguish a genuinely ocular component from a neural component whose time course is systematically coupled to blink timing, for instance a component reflecting blink-suppression activity or attentional modulation associated with the blink event. This ambiguity is not specific to VEOS; it applies to correlation-based ICA labelling in general and is acknowledged in the broader artefact-removal literature [6, 7]. Proper validation of this extension would require a dataset with sufficient channel density for stable decomposition, ground-truth component labels, and a formal evaluation of false-positive rejection rate. This falls outside the scope of the present feasibility study and is noted as a direction for future work. For the low-channel wearable scenario that motivates VEOS, the reference-subtraction (LCA) path remains the recommended approach.

6. Conclusions

We presented VEOS, a camera-based method for inferring a VEOG-like biosignal from monocular periocular video and using it for ocular artefact suppression in EEG. Across an in-house dataset and the public Eye–BCI dataset, VEOS achieves strong waveform-level agreement with true VEOG and enables substantial blink suppression when used as a reference in blink-window regression. In the present Eye–BCI evaluation, VEOS yields a modest improvement in MI, leaves SSVEP performance essentially unchanged, and approaches true-VEOG cleaning for P300 spelling. These findings support camera-derived

VEOG as a useful ocular reference for EEG artefact suppression in settings where additional periocular electrodes are undesirable. Because evaluation was limited to controlled recordings, further work is needed before extrapolating to unconstrained wearable deployment.

Acknowledgments

During the preparation of this manuscript, the authors used AI-assisted tools for language editing and figure drafting. All outputs were reviewed and edited by the authors, who take full responsibility for the content.

Data availability statement

Eye–BCI is publicly available (Synapse ID: syn64005218). VEOS-Dev image data is not publicly available due to ethics restrictions; derived feature streams and trained model weights are available upon reasonable request. Code and shareable artefacts are openly available at https://zenodo.org/records/18256709 [35].

The data that support the findings of this study are openly available at the following URL/DOI: https://dx.doi.org/10.13039/501100001602. The Eye–BCI dataset is publicly available (Synapse ID: syn52510979). VEOS-Dev image data is not publicly available due to ethics restrictions; derived feature streams and trained model weights are available upon reasonable request. Code and shareable artefacts are openly available at: https://zenodo.org/records/18256709.

Conflict of interest

The authors declare no conflict of interest.

Funding

This work was supported in part by Transpoco Ltd and by Science Foundation Ireland (Grant Nos. SFI/12/RC/2289_P2 and 18/SP/5942).

Ethics statement

This study was conducted in accordance with the regulations of the Dublin City University Research Ethics Committee (DCU REC) and was approved under REC reference DCUREC/2021/175. All participants provided written informed consent prior to participation. The Eye–BCI dataset [18] was obtained from its public release and used in accordance with the conditions specified by the original authors.

Author contributions

Andrew Fleury 0009-0003-6916-6770
Funding acquisition (equal),
Resources (supporting), Writing – review & editing (supporting)

Tomás Ward 0000-0002-6173-6607
Funding acquisition (lead), Project administration (lead), Resources (equal), Supervision (lead), Validation (equal), Writing – original draft (supporting), Writing – review & editing (supporting)

References

[1] Urigüen J A and Garcia-Zapirain B 2015 *J. Neural Eng.* **12** 031001
[2] Fatourechi M, Bashashati A, Ward R K and Birch G E 2007 *Clin. Neurophysiol.* **118** 480–94
[3] Brophy E, Redmond P, Fleury A, De Vos M, Boylan G and Ward T 2022 *Front. Neuroergonom.* **2** 805573
[4] Croft R J and Barry R J 2000 *Neurophysiol. Clin./Clin. Neurophysiol.* **30** 5–19
[5] Gratton G, Coles M G H and Donchin E 1983 *Electroencephalogr. Clin. Neurophysiol.* **55** 468–84
[6] Jung T P, Makeig S, Bell A J, Sejnowski T J and Lee T W 2000 Independent component analysis of biomedical signals *Proc. Int. Workshop on Independent Component Analysis and Signal Separation (ICA)*
[7] Hyvärinen A and Oja E 2000 *Neural Netw.* **13** 411–30
[8] Romero S, Mañanas M A and Barbanoj M J 2008 A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: a simulation case *Comput. Biol. Med.* **38** 348-60
[9] Zhang H, Zhao M, Wei C, Mantini D, Li Z and Liu Q 2021 *J. Neural Eng.* **18** 056057
[10] Kalita B, Deb N and Das D 2024 *Sci. Rep.* **14** 24234
[11] Zeng P, Fan L, Luo Y, Shen H and Hu D 2024 *J. Neural Eng.* **21** 066011
[12] Chaddad A, Wu Y, Kateb R and Bouridane A 2023 *Sensors* **23** 9556
[13] Hansen D W and Ji Q 2009 *IEEE Trans. Pattern Anal. Mach. Intell.* **32** 478–500
[14] Lugaresi C *et al* 2019 arXiv:1906.08172
[15] Grishchenko I, Ablavatski A, Kartynnik Y, Raveendran K and Grundmann M 2020 arXiv:2006.10962
[16] Redmond P, Awais A M and Ward T 2023 Inferring EOG signal from video *MEG United Kingdom and Ireland Conf. 2023*
[17] Redmond P, Awais A and Ward T 2023 Automatic tagging of BCI artefacts using computer vision *Proc. 10th Int. BCI Meeting*
[18] Guttmann-Flury E, Sheng X and Zhu X 2025 *Sci. Data* **12** 587
[19] Viola P and Jones M 2001 Rapid object detection using a boosted cascade of simple features *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)* p I–511–I–518
[20] Bradski G 2000 The OpenCV Library *Dr. Dobb's J. Softw. Tools* **25** 120–6

[21] Soukupová T and Čech J 2016 Real-time eye blink detection using facial landmarks *Proc. 21st Computer Vision Winter Workshop (CVWW 2016)*
[22] Savitzky A and Golay M J E 1964 *Anal. Chem.* **36** 1627–39
[23] Widmann A, Schröger E and Maess B 2015 *J. Neurosci. Methods* **250** 34–46
[24] Gustafsson F 1996 *IEEE Trans. Signal Process.* **44** 988–92
[25] Hoerl A E and Kennard R W 1970 *Technometrics* **12** 55–67
[26] Bai S, Kolter J Z and Koltun V 2018 arXiv:1803.01271
[27] Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 *J. Mach. Learn. Res.* **15** 1929–58
[28] Nair V and Hinton G E 2010 Rectified linear units improve restricted Boltzmann machines *Proc. 27th Int. Conf. on Machine Learning (ICML)* pp 807–14
[29] Kingma D P and Ba J 2015 Adam: a method for stochastic optimisation *Proc. 3rd Int. Conf. on Learning Representations (ICLR)*
[30] Ramoser H, Müller-Gerking J and Pfurtscheller G 2000 *IEEE Trans. Rehabil. Eng.* **8** 441–6
[31] Fisher R A 1936 *Ann. Eugen.* **7** 179–88
[32] Bin G, Gao X, Wang Y, Li Y, Hong B and Gao S 2009 *J. Neural Eng.* **6** 046002
[33] Farwell L A and Donchin E 1988 *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23
[34] Pfurtscheller G and Lopes da Silva F H 1999 *Clin. Neurophysiol.* **110** 1842–57
[35] Available at: The Eye–BCI dataset is publicly available (Synapse ID: syn52510979). VEOS-Dev image data is not publicly available due to ethics restrictions; derived feature streams and trained model weights are available upon reasonable request. Code and shareable artefacts are openly available at https://zenodo.org/records/18256709