VergeIO: Depth-Aware Eye Interaction on Glasses

XIYUXING ZHANG[†*], BNRist, Tsinghua University, China
DUC VU[†], Carnegie Mellon University, USA
CHENGYI SHEN, Carnegie Mellon University, USA
ZHIKAI QIN, Carnegie Mellon University, USA
RUIQI HONG, Wuhan University, China
YUNTAO WANG, Tsinghua University, China
YUANCHUN SHI, BNRist, Tsinghua University, China
JUSTIN CHAN, Carnegie Mellon University, USA

There is growing industry interest in unobtrusive designs for electrooculography (EOG) sensing of eye gestures on glasses (e.g. JINS MEME and Apple eyewear). We present *VergeIO*, an EOG-based glasses system that enables depth-aware eye interaction by sensing vergence with a glasses-compatible electrode layout and smart glass prototype. It can distinguish between four depth-based eye gestures with 97% accuracy on unseen users without any calibration in a user study across 20 users and 1,520 gesture instances. To reduce false detections, we incorporate a motion artifact detection pipeline and a preamble-based activation scheme. The system uses dry sensors without any adhesives or gel and operates in real time with 3 mW power consumption by the analog sensing front-end.

CCS Concepts: • **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Additional Key Words and Phrases: vergence, eye sensing, electrooculography

ACM Reference Format:
Xiyuxing Zhang, Duc Vu, Chengyi Shen, Zhikai Qin, Ruiqi Hong, Yuntao Wang, Yuanchun Shi, and Justin Chan. 2026. *VergeIO*: Depth-Aware Eye Interaction on Glasses. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 10, 3, Article 192 (September 2026), 31 pages. https://doi.org/10.1145/3831627

1 INTRODUCTION

Given their proximity to the eyes, glasses are uniquely positioned to leverage eye movements for interaction. A less-explored movement is vergence: the inward or outward motion of the eyes that occurs naturally as humans shift their gaze between objects at different depths (Fig. 1) [39]. Unlike fast, reflexive saccades or lateral gaze shifts, vergence is slower and more deliberate, making it well-suited for intentional, hands-free interaction [22, 47, 89].

[†]Co-primary authors.
[*]Corresponding author.

Authors' Contact Information: Xiyuxing Zhang (corresponding author), BNRist, Tsinghua University, Beijing, China, zxyx22@mails.tsinghua.edu.cn; Duc Vu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, ducv@andrew.cmu.edu; Chengyi Shen, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, chengyis@andrew.cmu.edu; Zhikai Qin, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, qinzk23@mails.tsinghua.edu.cn; Ruiqi Hong, Wuhan University, Wuhan, China, rachel@whu.edu.cn; Yuntao Wang, Tsinghua University, Beijing, China, yuntaowang@tsinghua.edu.cn; Yuanchun Shi, BNRist, Tsinghua University, Beijing, China, shiyc@tsinghua.edu.cn; Justin Chan, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, justinchan@cmu.edu
[*]For correspondence, please contact Justin Chan and Yuntao Wang.



This work is licensed under a Creative Commons Attribution 4.0 International License.
© 2026 Copyright held by the owner/author(s).
ACM 2474-9567/2026/9-ART192
https://doi.org/10.1145/3831627

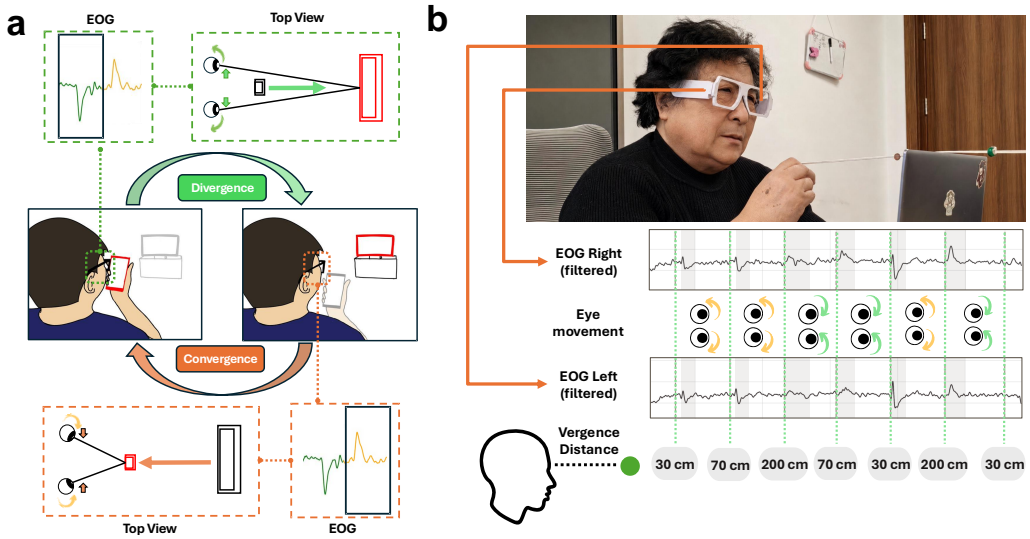


Fig. 1. **Overview of the *VergeIO* system. (a)** *VergeIO* senses eye vergence on glasses using electrooculography (EOG) by detecting gaze shifts between objects at different depths. The eyes diverge when shifting focus from a near to a far object, and converge when focusing from a far to a near object. **(b)** Our system can distinguish different vergence shifts. Each shift corresponds to a unique EOG pattern that can be classified by the proposed algorithms.

Glasses that can sense and respond to these depth-related eye movements can enable new forms of hands-free interaction across key depth cues (Fig. 2):

- *Ambient depth cues.* Varifocal glasses can dynamically refocus between a nearby book and a distant TV, detect abrupt gaze depth shifts and flag attention lapses.
- *Virtual depth cues.* Augmented reality glasses can leverage vergence to select or focus on different virtual windows.
- *Approximate physical depth cues.* Remote screening for ocular disorders (e.g., convergence insufficiency, strabismus [64]) can be performed using one's thumbs and a distant object as approximate depth markers.

Electrooculography (EOG) is a promising modality for vergence (gaze depth) sensing on glasses, with growing industry interest in integrating EOG electrodes into glasses. Notably, these sensors have already been integrated into the nose pads and bridge of commercial-grade glasses including JINS MEME [11], and more recently, Apple was granted a patent [6] to embed EOG sensors into glasses.

Unlike cameras, EOG is robust to varying ambient lighting conditions as glasses are worn across a range of environments. EOG is also low-power compared to commercial-grade cameras. In contrast to mmWave [58] and acoustic [55] systems that place sensors in front of the lenses and can obstruct the user's vision, EOG electrodes have been integrated unobtrusively into the frames of the glasses. It is for these reasons that prior work [34, 45, 49, 53, 71, 77] has investigated EOG for capturing eye movements such as blinks and coarse gaze shifts (e.g., left, right, up, down) to support the understanding of interaction intentions.

The challenge, however, is that current EOG systems for glasses are not designed to robustly detect vergence. During vergence, the eyes move in opposing directions to adjust viewing depth, which creates *opposing electrical signals* that can interfere and become difficult to detect at the sensing electrodes [85].

We present *VergeIO*, an EOG-based glasses system that expands the range of detectable eye movements to include vergence gestures. We identify two design objectives for robust vergence sensing: preserving the



Fig. 2. **Applications of *VergeIO*. (a)** Varifocal lenses adjust focal depth as the user shifts gaze between ambient reference points: a nearby phone and a distant TV. **(b)** Smart glasses enable gaze-based selection between virtual objects at different depths. **(c)** Remote screening of ocular disorders using approximate reference points: thumbs and a distant object.

separability of signal contributions from the left and right eyes to limit interference, and maintaining wide sensor spacing to support a high signal-to-noise ratio (SNR). Accordingly, we use a two-channel design that records signals from the left and right eyes on separate channels using a shared reference electrode on the nose bridge and electrodes on the temples. We further implement this configuration in a hardware prototype compatible with the form factor of everyday glasses.

Using this hardware setup, we show that our system can classify four to six vergence gestures that produce distinguishable EOG signals. These gestures involve changing gaze depth between three distances that span everyday viewing distances: reading distance (30 cm), social interaction distance (70 cm), and far-field visual attention (200 cm).

However, in real-world use, deploying precise depth references is not practical, so our system evaluates two strategies for ad hoc depth references on glasses. **For glasses with displays** such as augmented reality smart glasses, our system can render distance markers using on-screen visual overlays. Although these markers are displayed on a fixed physical screen, they simulate depth using binocular disparity, which causes users to exhibit vergence when gazing between different depths. **For glasses without displays**, users can use their outstretched thumbs as *approximate* depth references: one at full arm's length (≈70 cm) and the other at half arm's length (≈30 cm), for example resting on the opposite elbow, along with any arbitrary far-field target. This approach extends to basic audio-only frames like Amazon Echo Frames [3] or instrumented varifocal glasses.
Our contributions are as follows:

- We present a depth-aware glasses system that distinguishes four vergence gestures with 97.01% accuracy when evaluated on unseen users without calibration in an IRB-approved study with $N = 20$ users. The system is also compatible with traditional EOG sensing gestures including blinks and coarse gaze shifts and achieves 96.10% accuracy across an 11-class set in a feasibility evaluation.
- A smart glasses prototype to sense vergence with flexible electrode contacts to accommodate different users. The prototype has an analog front-end power consumption of 3 mW and can run in real-time on a lightweight microcontroller. The total weight of our system is 75 g.
- To support robust performance in real-world scenarios, we design a motion artifact detection pipeline to discard unwanted events caused by facial and body movements. *VergeIO* achieved an accuracy of 98.77% with a false rejection rate of 1.08% and a false positive rate of 1.23% on a dataset collected from $N = 6$ participants.
- We design a preamble-based activation scheme, where a brief eyebrow raise is performed to activate or deactivate vergence sensing. This reduces the false positive rate across several activities from 0.50–2.43% to nearly zero, and ensures that depth-based interactions are only initiated when intended.

Table 1. Comparison of existing eye sensing systems on glasses. *VergeIO* is explicitly designed for vergence sensing using an EOG electrode configuration for glasses. (DNS = did not specify).

Reference	Sensing Modality	Vergence Sensing	Robust to Ambient Light	Unobstructed View	Glasses Form Factor	Power
CIDER [62]	Cameras	✓	✗	✗	✓	40 µW
Tobii [18]	Cameras	✓	✗	✓	✓	DNS
FocusFlow [89]	Cameras	✓	✓	✓	✗	DNS
Li et al. [56]	Photodiodes	✓	✗	✓	✓	395 µW
GazeTrak [55]	Acoustic	✓	✓	✗	✓	16.4 mW
mmET [58]	mmWave	✓	✓	✗	✓	6 mW
JINS MEME [11]	EOG	✗	✓	✓	✓	DNS
Mishra et al. [64]	EOG	✓	✓	✓	✗	DNS
ElectraSight [76]	EOG (hybrid)	✗	✓	✓	✓	3.3 mW
Lee et al. [53]	EOG	✗	✓	✓	✓	DNS
GAPses [34]	EOG/EEG	✗	✓	✓	✓	6.2 mW
***VergeIO* (ours)**	**EOG**	✓	✓	✓	✓	**3 mW**

- We open-source our hardware design and software code[1] for the community to build on.

We envision that the sensing capability introduced by our approach can broaden the scope of eye interaction available on EOG-based glasses, enabling applications in consumer, health, and safety-monitoring domains.

2 BACKGROUND AND RELATED WORK

2.1 Eye Movement Sensing on Headworn Devices

There has been increased interest in eye movement sensing on head-mounted devices. Camera-based approaches [15, 18, 30, 61, 62, 74], while accurate, are power-hungry and are typically ineffective in the dark without infrared illumination. Photodiode-based systems [56], though lower in power, can be saturated by ambient sunlight in outdoor settings. Emerging acoustic [55] and mmWave systems [58] offer a promising low-power alternative, but can obstruct the user's line of sight.

Electrode-based approaches have made their way into VR headsets [68], goggles [27], smart glasses [34, 44, 45, 48, 49, 53, 71, 76], and earable devices [46, 54, 57, 59]. They are often able to simultaneously record electroencephalography (EEG), electrooculography (EOG), and electromyography (EMG) signals.

Dry electrodes have been integrated into several commercial smart-glasses platforms such as JINS MEME [11] and AttentivU [49] due to their small size, weight, and low power-consumption. They have achieved reliable blink detection and horizontal/vertical saccade tracking [85], facial-expression recognition [71], midair gesture and context recognition [87], fatigue estimation [81], and mediated attention monitoring with real-time biofeedback [50]. We build on this rich body of prior work by enabling vergence and depth-awareness as a key interaction primitive for EOG-based glasses.

2.2 EOG-Based Sensing Systems

EOG has been used to detect various eye movements, including vergence, saccades, blinks, gaze shifts, and fixations [24, 25]. However, it is important to note that EOG is generally not suitable for high-precision tracking. Despite a spatial resolution limited to approximately 1.5° [88], Stevenson et al. [80] demonstrated fixation depth

[1] https://github.com/zx-explorer/VergeIO

estimation using an eight-channel wet-electrode EOG setup, reporting an average fixation distance error of 10.3 ± 10.0% (5.7 ± 4.7 cm). This suggests that obtaining stable absolute depth estimates remains challenging.

Prior systems have leveraged EOG-detected eye movements as an input for assistive interfaces for users with physical disabilities [69], robotic control [28], and hands-free operation of wearable computers for caregivers [65]. Beyond eye movement, EOG has also been used to detect facial expressions (e.g. movements of the cheek, brow, and nose) [71] and to recognize activities such as typing, reading, eating, and talking [41].

As EOG only requires small dry electrodes, there has been growing interest to incorporate them into glasses with academic prototypes including GAPses [34], E-Glasses [53] and ElectraSight [76], and notable commercial efforts including JINS MEME [11], and Apple's recently approved patent [6] to incorporate EOG electrodes into glasses. However, currently proposed electrode configurations are either not able or not focused on sensing vergence, and are limited to blinks, winks, saccades and directional gaze shifts; or require a large number of electrodes that can be uncomfortable. In contrast, *VergeIO* expands the set of eye movements that can be detected on glasses to include four to six different vergence gestures using a glasses-compatible EOG configuration designed to preserve the opposing electrical polarity signals produced by vergence.

2.3 Vergence as an Interaction Gesture

Vergence has emerged as a reliable eye-based input modality in extended reality environments [22, 89] due to its ability to mitigate the "Midas touch" problem, which refers to the difficulty of separating deliberate gaze commands from involuntary natural exploration as the eyes move around a scene [42]. Prior work has used eye-tracking cameras to measure vergence in VR systems for selecting and activating virtual objects [22, 89], controlling linear actuators [72], and studying visual attention in children with reading difficulties [43]. EOG-based systems include configurations that are not compatible with a glasses form factor [82] and customized flexible-electrode designs that require specialized fabrication [64], making them more challenging to scale. In contrast, our approach focuses on detecting vergence using an EOG-based electrode configuration for glasses using conventional Ag/AgCl dry electrodes.

3 SYSTEM DESIGN

3.1 Vergence Distance Selection

Given the spatial-resolution limitations of EOG discussed in Sec. 2.2, we define the depth-aware task for EOG as detecting transitions between discrete distances rather than performing continuous depth regression. Therefore, a key system design parameter is the selection of vergence distances which is the physical distance between the eye and a fixation distance. These distances must satisfy two requirements: First, they should span the range of realistic, everyday viewing distances, and second, changes in vergence between these distances should result in discernible EOG signals that can be classified.

Since EOG measures eye movements and directly reflects *changes in vergence angle*, we choose distances such that the resulting changes in vergence angle are uniformly spaced. Fig. 3a shows how vergence angle (θ_1, θ_2, θ_3) is defined as the angle between the lines of sight of both eyes as they converge on a fixation point.

Binocular vergence angle for a target at distance d is given by $\theta(d) = 2\arctan\left(\frac{\text{IPD}}{2d}\right)$. Within our interaction range ($d \geq 30$ cm, according to ISO 9241-303 [40]) and an upper-bound adult IPD of 75 mm [12, 31], $\frac{\text{IPD}}{2d} \leq 0.125 \ll 1$. Therefore, $\arctan(x) \approx x$, yielding $\theta(d) \approx \frac{\text{IPD}}{d}$. As an individual's IPD is fixed, the vergence shift between depths simplifies to $\Delta\theta_{ij} \approx \text{IPD}\left|\frac{1}{d_i} - \frac{1}{d_j}\right|$. We therefore select viewing distances of 30, 70, and 200 cm, which produce pairwise vergence angle changes ($\Delta\theta$) in an approximate 1:2:3 ratio. As EOG responses can be regarded linear for gaze angles within ± 30° [33], our selection increase the separability among all possible vergence gestures in practice.

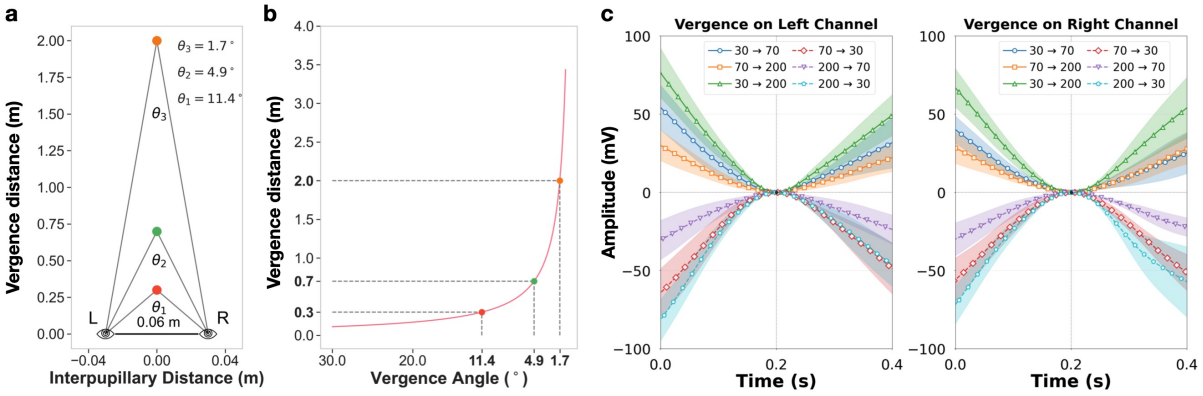


Fig. 3. **Geometric characterization and signal validation for vergence distance selection. (a)** *VergeIO* measures *changes* in vergence angle (θ) using EOG signals, which vary proportionally with these changes. Interpupillary distance of 60 mm is used for illustration. **(b)** Our system can distinguish changes in vergence angles corresponding to three interaction distances: 30, 70, and 200 cm. **(c)** Mean EOG signal for each of the six vergence gestures for the left and right eye, with each trace's extremum aligned to (0.2 s, 0 mV) for comparison. Shaded region represents one standard deviation from the mean.

As shown in Fig. 3b, because the relationship between vergence angle and distance is non-linear, these vergence angles correspond to three vergence distances at key interaction distances: near (30 cm), mid (70 cm), and far (200 cm), which align with focal zones outlined by the American Optometric Association (AOA) [60].

We show in Fig. 3c the mean EOG curves for each of the six gestures when performed using our proposed electrode placement by one participant. The plot shows that the direction of depth shift is separable due to EOG signals with opposite polarities generated by convergence and divergence. For gestures moving in the same depth direction, the largest angular shift (30 ↔ 200 cm) and the smallest angular shift (70 ↔ 200 cm) are visually distinct. However, we do note 30 ↔ 70 cm transitions exhibit overlap with other gestures, making them challenging to separate visually. This motivates our decision to focus on a 4-gesture set as the primary use case for our system.

3.2 Placement of EOG Electrodes

Electrode placement is a critical consideration in our design as it directly impacts the types of eye movements that can be detected. Existing EOG sensing systems on glasses generally rely on two design paradigms. The first employs compact electrodes around the nose pads and bridge, as represented by commercial systems like JINS MEME [11] and academic prototypes such as GAPses [34]. The second relies on wider sensor placements near two temples, as represented by E-Glasses [53]. However, neither paradigm is specifically designed to effectively capture vergence.

JINS MEME places both positive electrodes on the nose pads and the negative electrode on the nose bridge with minimal spatial separation (Fig. 4a). This compact configuration primarily captures conjugate changes (e.g. blinks, saccades) rather than the asymmetric signals required to detect vergence. With limited spatial separation, the opposing electrical dipoles generated during vergence tend to interfere, cancel out, or simply become indistinguishable. This limitation was explicitly confirmed by prior work authored by JINS MEME researchers [85].

E-Glasses [53] (Fig. 4b) increases spatial separation by placing electrodes near two temples and records a single differential channel across the two sides, which is effective for sensing conjugate movements (e.g., saccades) where eyes rotate in the same direction. However, the two eyes rotate in opposite directions during vergence, resulting in the electrical potentials at both temples rising or falling together, which can result in partial cancellation. This

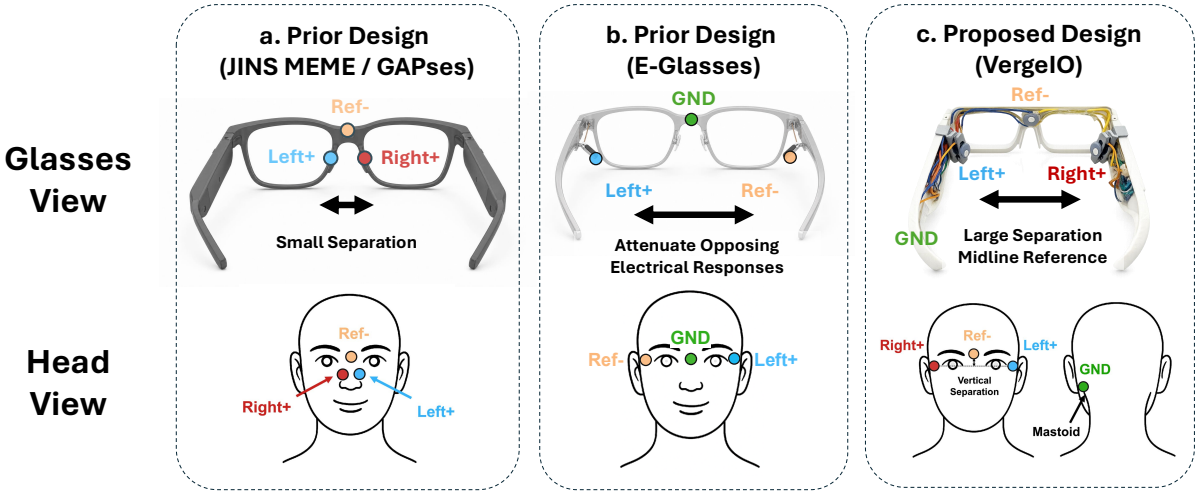


Fig. 4. **Representative glasses-based EOG electrode configurations. (a)** Compact nose-pad and bridge configuration represented by JINS MEME [11] and GAPses [34]. **(b)** Wider temple configuration represented by E-Glasses [53], which records a single differential channel across the two sides. **(c)** *VergeIO* combines wide sensor separation with a shared nose-bridge reference, capturing vergence with high SNR while preserving the ability to sense non-vergence eye movements.

limitation was confirmed by Mishra et al. [64] in a vergence classification task, where a conventional differential setup achieved only 34% accuracy.

To address these limitations, our system is designed to support vergence sensing while being compatible with a glasses form factor. Our design (Fig. 4c) places positive electrodes at each temple (left and right channels), a shared negative electrode on the nose bridge, and a ground electrode at the mastoid. This configuration offers a twofold advantage: the lateral sensing electrodes provide sufficient spatial separation to capture vergence responses with high SNR, while the shared nose-bridge reference preserves the opposing-polarity signals generated by convergence and divergence. Additionally, by introducing a vertical separation between the temple and nose-bridge electrodes, our design preserves the ability to sense non-vergence eye movements (e.g., saccades and blinks). Together, *VergeIO* expands vergence sensing capability while maintaining compatibility with conventional horizontal and vertical EOG-based eye-movement sensing, which was validated in Sec. 6.1.

To compare the signal quality of our approach with the configurations discussed above, we conduct a pilot study to measure the signal-to-noise ratio (SNR) of vergence gestures captured by *VergeIO* against the two paradigms (JINS MEME [11] and GAPses [34]; E-Glasses [53]). In this study, one participant simultaneously wore electrodes required for these layouts: two electrodes on the temples, two on the nose pads, a shared electrode on the nose-bridge as the reference, and a ground electrode at the mastoid.

We recorded four-channel EOG data over ten rounds, where each round consisted of six vergence gestures between the three distances we selected in Sec. 3.1. Synchronized data were sampled at 500 Hz and processed in sequence with a third-order zero-phase low-pass filter at 10 Hz, a notch filter, and a third-order Savitzky–Golay filter [75] (window size is 0.5 second). We manually segmented vergence events from the filtered signals of our two channels. Since distinguishable event boundaries were not visible in the JINS MEME channels, we applied the onset and offset times identified from the *VergeIO* channels to all configurations. For each event, we extracted the two 0.2-second signals immediately before and after the event as noise segments. SNR was then computed as the ratio of the root-mean-square amplitude of the event segment to that of the concatenated noise segments.

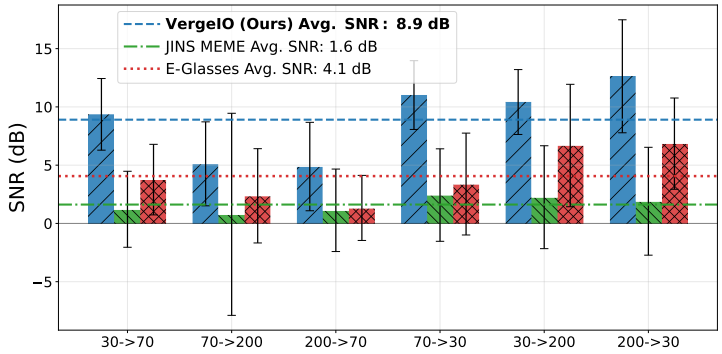


Fig. 5. **SNR comparison across six vergence gestures.** In this pilot study, *VergeIO* outperforms the electrode layouts from prior designs in JINS MEME glasses [11] and E-Glasses [53]. We show the left channel for *VergeIO* and JINS MEME. E-Glasses outputs a single channel only.

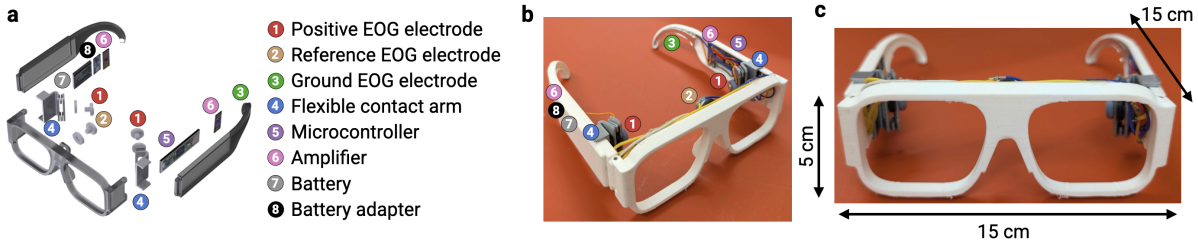


Fig. 6. **Hardware design. (a)** Exploded CAD design illustrating the different components of the hardware design. **(b)** Fabricated design annotated with location of different components. **(c)** Front-view of glasses with dimensions.

As shown in Fig. 5, the average SNR across all six vergence gestures was 8.9 dB with our system, compared to 1.6 dB with the JINS MEME layout and 4.1 dB with the left-right differential layout represented by E-Glasses. These results indicate stronger EOG responses during vergence with our system.

We note that while prior work [66] on EOG sensing incorporates vertical electrodes above and below each eye, vergence is primarily a horizontal movement, and vertical channels only provide modest performance gains. Additionally, we found in our evaluation that (a) participants reported that the additional vertical electrodes caused discomfort (Sec. 5.1) and (b) our electrode configuration is still able to classify between different vertical eye movements (Sec. 6.1).

3.3 Hardware Design

We create a glasses hardware prototype to ensure reliable skin contact of our chosen electrode placement (Fig. 6):

• **Electrodes.** We use dry electrodes for practicality, similar to other glasses with biosignal electrodes [11, 49]. We selected SoftPulse [29] which has been used for EOG acquisition [29, 34, 37], and is biocompatible with the skin (ISO 10993-5/10).

– *Positive electrodes.* These are coupled to flexible contact arms at the user's temple for good skin contact. These two arms can slide along the glasses' temples to accommodate different users.

– *Reference electrode.* This is mounted at the nose bridge to ensure direct contact with the user's skin.

Table 2. Dataset collected from $N = 6$ participants for evaluating the artifact and noise removal pipeline.

Category	Description	#Recordings	#Events	Duration (min)
Positive	Vergence	60	360	25.20
Noise	Blinking	30	180	13.31
Noise	Brow raising	30	180	12.80
Noise	Saccades	30	180	12.99
Motion	Head tilt	30	180	12.83
Motion	Head turn	30	180	12.83
Motion	Stand/Sit	30	180	12.80
Total	**All**	**240**	**1440**	**102.76**

– *Ground electrode.* Curved coating at the end of the arm to maintain good contact with the user's mastoid.

• **Amplifiers.** We chose the BioAmp-EXG-Pill [4] for its compact analog front end ($2.5 \times 1 \times 0.1$ cm) with a 1000× signal gain, and its ability to interface with any 3.3 and 5 V MCU via ADC.

• **Microcontroller.** We selected the Teensy 4.1 [16], which integrates 12-bit ADC and high-performance Cortex-M7 core, enabling precise, high-resolution bio-signal digitization and real-time processing.

• **Battery.** We used the same 570 mAh rechargeable lithium battery as in Google Glass [9]. As discussed in Section 6.6, this battery enables sensing for approximately 197 hours (8.2 days). We also integrated the Adafruit Micro Lipo Charger 4410 [2] into the glasses to enable on-board charging of the sensing system.

Cost. The off-the-shelf cost is $140, with the optimized cost $73. The cost of the ELDRY electrode coating (4 cm^2, 0.5 mm thickness) is $20; Li-Po rechargeable battery is $15; Li-Po micro charger is $6; microcontroller development board is $30; and two BioAmp-EXG-Pill amplifier modules is $69, which can be reduced to $2 when using the discrete components.

3.4 Motion Artifact and Noise Removal

Given that EOG signals are highly susceptible to facial and body movements [71], the identification and removal of such artifacts are imperative to prevent false positives. Since the resulting EMG (muscular) responses overlap with the EOG signals in frequency, they cannot be easily isolated in the frequency domain.

For this reason, we leverage a temporal, data-driven segmentation approach to identify and remove segments affected by user movement. Our noise removal pipeline is designed to operate on streaming EOG data, and operates on a moving window of 2 s with a step size of 0.1 s. We chose 2 s as it is ample time to perform a vergence movement, which typically takes less than 1 s [84].

We collected a dataset (Table 2) comprising different motion and noise signals from six participants (Fig. 7). Participants were instructed to perform **(1)** vergence movements; **(2)** body and head motions (e.g., turning, head tilt, standing/sitting); **(3)** facial and eye movements typically considered noise (e.g., brow raising, saccades, blinking). Participants were cued with an audible beep every 3 seconds to perform every movement.

Signal Conditioning. All signals were preprocessed using a third-order low-pass Butterworth filter with a 10 Hz cutoff, which was chosen to attenuate high-frequency noise while preserving the dominant spectral content of the EOG signal [64]. A notch filter working at 60 Hz was also adopted to suppress powerline interference.

Segmentation and Silence Removal. To support real-time processing of streaming EOG data, the filtered signal is segmented using a 2-s moving window with a 0.1-s stride. We then apply a dynamic amplitude threshold to minimize computational overhead and prevent false positives from static noise, ensuring that only windows exceeding this threshold are processed. Specifically, the threshold is computed by applying a median absolute deviation (MAD) filter over the first 2 seconds of each channel, which are assumed to be static, and multiplying this by a heuristic value of 4.5.

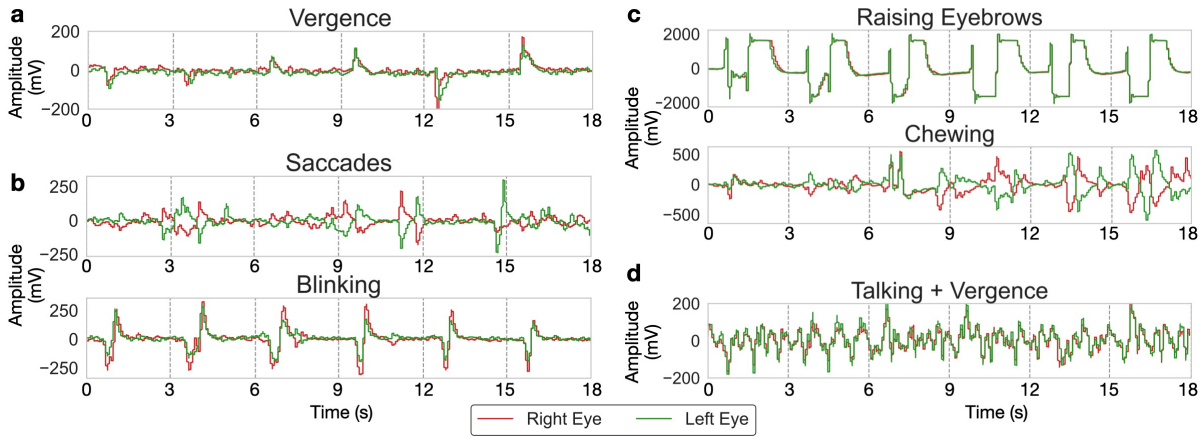


Fig. 7. **EOG signals recorded from the left and right eye during vergence, motion and noise events. (a)** Six vergence gestures. **(b)** Vergence is distinct from eye movements like saccades and blinking. **(c)** Facial movements produce higher amplitude signals than vergence alone. **(d)** When combined with facial movements, vergence is dominated by facial activity, and is treated as noise.

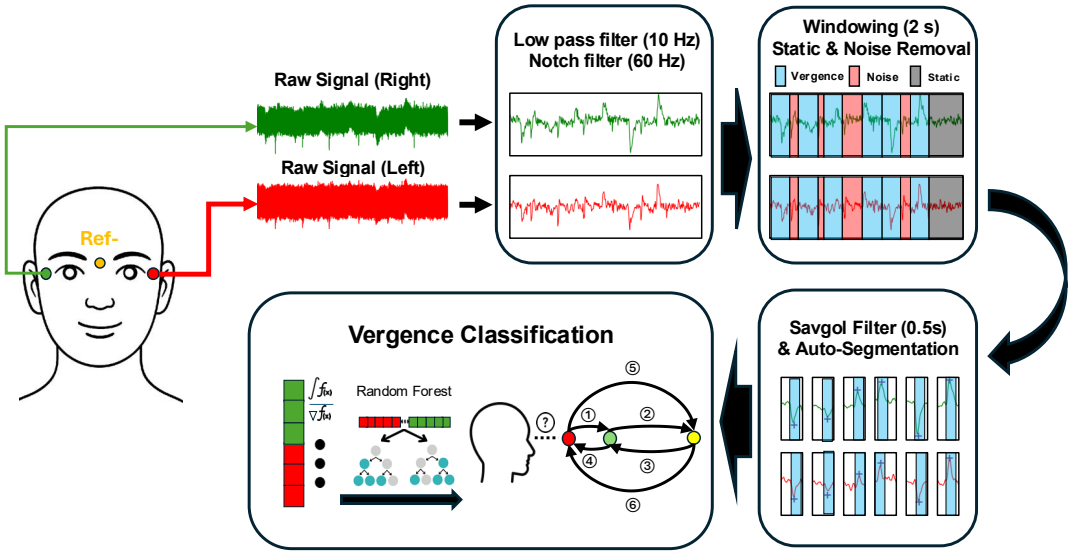


Fig. 8. *VergeIO* processing pipeline.

Artifact Detection and Removal. For each 2-second window we computed statistical features for both left and right EOG channels, including signal mean, extrema, band power, wavelet energy, variance, RMS, peak-to-RMS ratio, trapezoidal integral, maximum derivative, and minimum derivative. We trained a logistic regression classifier to distinguish between vergence and non-vergence (motion/noise) segments and remove the latter. We evaluated the pipeline in Sec. 6.4.

Algorithm 1 Vergence Gesture Classifier

procedure ClassifyVergence(eog)
  $L \leftarrow []$
  for all $seg \in$ Segment(eog, 2s) **do**
    if not IsArtifact(seg) **then**
      $seg \leftarrow$ SavitzkyGolayFilter(seg, 250)
      for all $p \in$ DetectPeaks(seg, 30mV) **do**
        $g \leftarrow seg[p-0.5, p+0.5]$
        $f \leftarrow$ ExtractFeatures(g)
        $l \leftarrow$ RandomForest.Predict(Z-Score Normalize(f))
        append l to L
      end for
    end if
  end for
  return L
end procedure
function ExtractFeatures(g)
  Split g into left and right halves; for each half compute
    $[\max - \min,\ \int g,\ (g_e - g_s)/0.5,\ \overline{\Delta g},\ \mathrm{Var}(\Delta g)]$
  return concatenated 10-dim vector
end function

3.5 Vergence Classification Pipeline

Automatic Segmentation. After noise-artifact removal, each 2 s signal window was smoothed using a Savitzky–Golay filter [75] with a window length of 250 samples (0.5 s), which can clean the signal while preserving the features. Compared with moving averages or wavelet denoising, the Savitzky–Golay filter preserves edges more faithfully and is computationally efficient for real-time use [20, 21]. To precisely determine the onset and offset of vergence gestures in each 2 s window, we employed the peak-detection algorithm in the detecta library [32] with 30 mV as the minimum amplitude threshold, which we determined empirically from the pilot study in Sec. 3.2. Since vergence movements typically range less than 1s [29], we then extracted a window centered on the detected peak, extending 0.5 seconds before and after the peak, yielding a 1-second segment.

Feature Extraction. Each 1 s segment can be regarded as two halves split by the peak we detected. From each half, we extracted five features: amplitude range, definite integral, end-to-end slope, mean of the first derivative, and variance of the first derivative. This results in ten features per EOG channel. Each gesture is represented by a 20-dimensional feature vector, comprising ten features extracted from each of the two channels. Additional details on feature selection are provided in Appendix A.

Vergence Classification. Prior to model training, all features were standardized via z-score normalization to compensate for inter-user variability. The normalization also mitigates data drift due to shifts in electrode placement before and after eyeglass remounting. Using the PyCaret library [23], we automated model selection and determined that the random forest was the most effective algorithm for classifying vergence gestures.

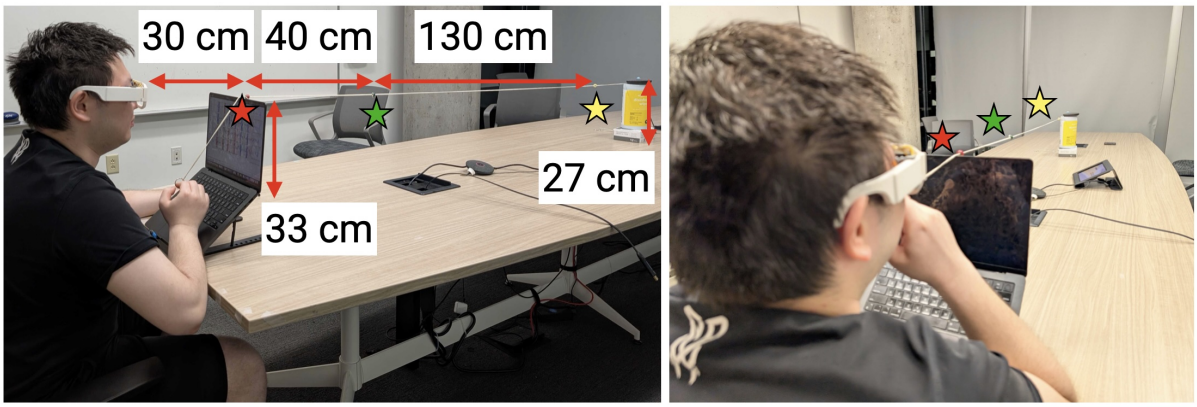


Fig. 9. **Experimental setup for user study.** The stars denote the location of colored beads tied to the Brock String and used as distance markers for the vergence gestures.

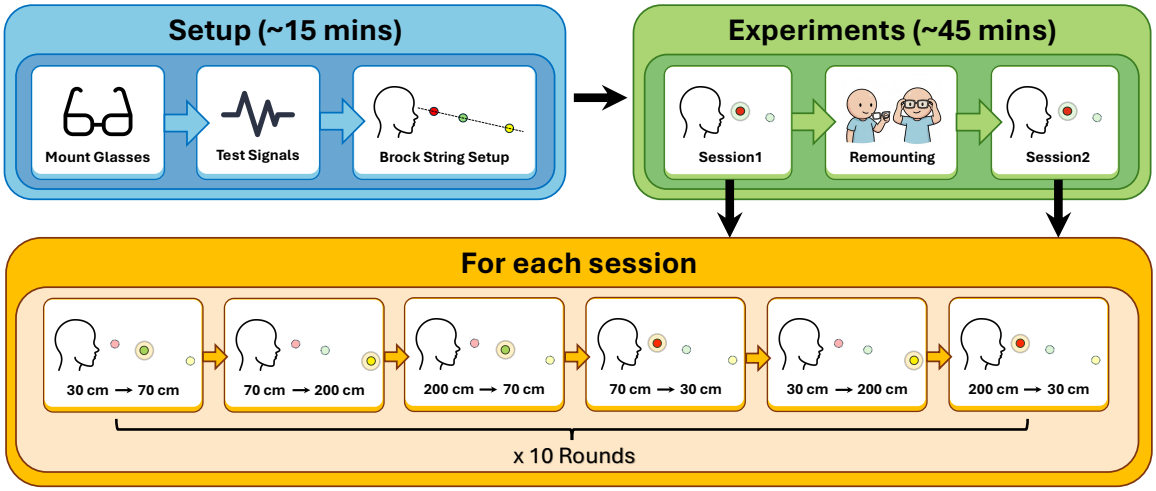


Fig. 10. **Experiment timeline.** Participants first underwent a setup phase to ensure the glasses were properly mounted and received training on performing vergence movements along the Brock String. During the experimental phase, they completed 10 rounds, each consisting of the six vergence gestures. After completing the rounds, participants removed and remounted the glasses before repeating the procedure.

4 STUDY DESIGN

4.1 Experimental Setup

This study was approved by our Institutional Review Board (STUDY2025_00000189). We recruited 20 participants (aged 25.75 ± 11.1), including 4 with normal vision and 16 with common refractive errors (12 myopia, 2 hyperopia, and 2 presbyopia). No participants reported other ocular-related disorders or convergence insufficiency.

All the experiments were conducted in a meeting room (Fig. 9). To trigger the six vergence gestures, we used a Brock String [36], a clinical tool commonly used to assess binocular vision and diagnose convergence

insufficiency. Our string includes three colored beads positioned at the designated vergence distances: red at 30 cm, green at 70 cm, and yellow at 200 cm.

In our setup (Fig. 9), participants sat in front of a laptop. The distal end of the Brock String was secured to a fixed anchor on the table at 27 cm above its surface, while the proximal end passes through a metal ring that the user holds taut at the bridge of their nose, positioned approximately 33 cm high to align with the participant's line of sight. The string is tensioned such that it rests against the top edge of the laptop screen, helping to stabilize it throughout the task. Elevating the string in this way prevented any visual overlap of the beads, ensuring that each target remained clearly distinguishable.

4.2 Data Collection Protocol

We began by fitting each participant with our glasses prototype. The flexible contact arm was adjusted to ensure secure, comfortable electrode contact with the skin over the temples. We positioned the Brock String so that participants could maintain a forward gaze aligned with all three beads.

Following the Brock String exercise procedure [78], participants were instructed to perform vergence gestures such that the target bead remained in sharp focus while the non-target beads appeared blurred. Our experimental protocol (Fig. 10) consisted of two sessions, each comprising ten rounds. In each round, participants performed all six vergence gestures once, cued by an audible beep, with a three-second window to complete each movement. After ten rounds, participants removed the glasses for a ten-minute break before remounting the device. For each participant, we disinfected all recording apparatus in accordance with institutional hygiene protocols.

5 EVALUATION

This section presents the evaluation results of *VergeIO* under the controlled Brock String protocol.

5.1 Vergence Classification

Within-Session Performance. We performed a within-session evaluation for each of the 20 participants in our study (Fig. 11a, Table 3). Based on the observation in Sec. 3.1, we first evaluate our system on four gestures which have a high degree of separability: (30 → 200 cm, 70 → 200 cm, 200 → 30 cm, and 200 → 70 cm). *VergeIO* achieved an average accuracy of 97.99 ± 1.96% across all participants at classifying between the four different vergence gestures. From the per-class perspective presented in Fig. 12(a), our system achieved near-perfect discrimination as shown in the confusion matrix.

We also evaluate our system on all six gestures supported by the three selected distances. Our evaluation (Table 3) shows that the system achieves a within-session classification accuracy of 82.68 ± 4.29%. There is virtually no misclassification between the 30 → 200 cm and 70 → 200 cm transitions, or between the 200 → 70 cm and 200 → 30 cm transitions, which is similar to the results for four gestures. However, the 30 → 70 cm movement creates an EOG waveform that can overlap with the 30 → 200 cm and 70 → 200 cm movements, and the 70 → 30 cm movement likewise can overlap between the 200 → 70 cm and 200 → 30 cm movements.

Importantly, for both the 4- and 6-gesture evaluations the near-to-far and far-to-near vergence movements are never confused, consistent with the observation in Sec. 3.1.

Cross-Session Performance. When the user remounts the glasses, slight shifts in electrode placement introduce signal drift. To evaluate the robustness of our system against remounting, we trained a model on one of the user's data collection sessions and tested it on the other unseen session. We then reversed the training and testing sessions and averaged the results. Our evaluation (Fig. 11b, Fig. 12b, Table 3) shows that system performance remained stable at 97.78% for the four-gesture task, validating the robustness of our system against signal drift induced by glasses remounting. However, for the 6-gesture detection task, the system performance decreased from 82.68% to 77.78% after remounting.

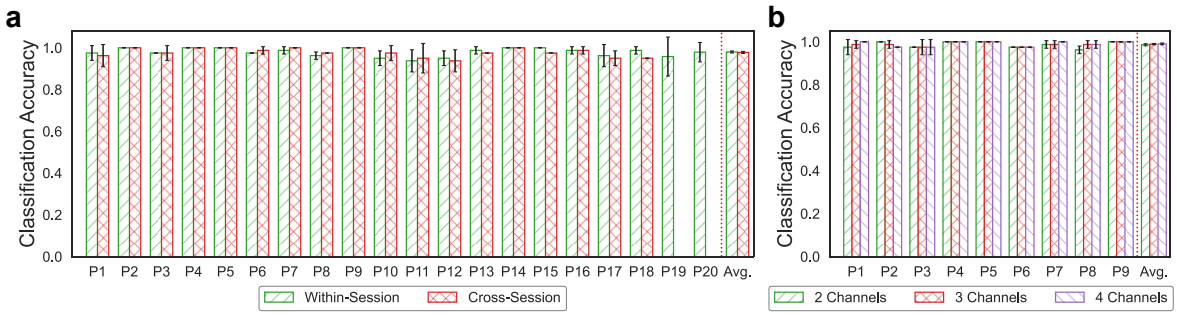


Fig. 11. **(a)** Classification accuracy ($n = 20$) for within-session (training and testing on data collected without remounting the glasses) versus cross-session (training on one session and testing on the unseen session after re-mounting glasses). P19 and P20 did not attend session 2. **(b)** Within-session 5-fold cross-validation accuracy per participant ($n = 9$) for 2-, 3- and 4-channel configurations.

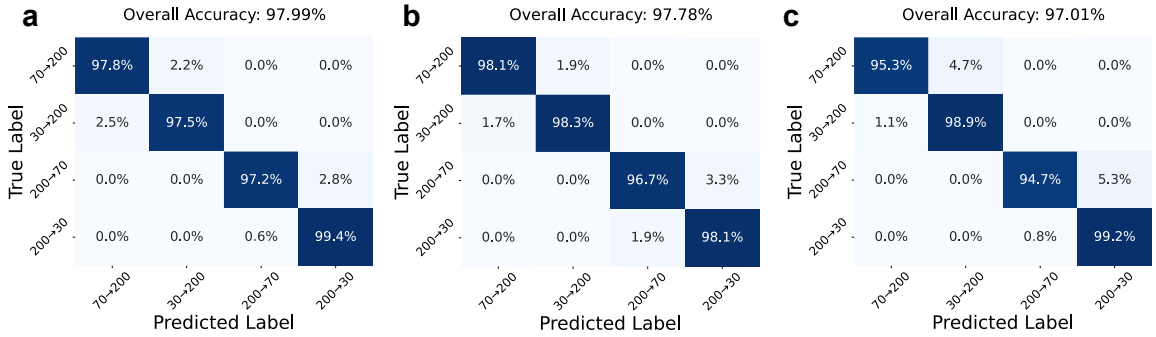


Fig. 12. **(a)** Confusion matrix of within-session 4 gestures classification across all the participants and sessions. **(b)** Confusion matrix of cross-session 4 gestures classification. **(c)** Confusion matrix of cross-user 4 gestures classification.

Table 3. Classification performance (mean ± stdev) for 4- and 6-gesture classifier. The primary use case for our system targets a 4-gesture classifier.

	Within Session	**Cross Session**	**Cross User**
4-gesture accuracy	97.99 ± 1.96%	97.78 ± 2.02%	97.01 ± 2.06%
6-gesture accuracy	82.68 ± 4.29%	77.78 ± 6.17%	77.36 ± 6.17%

Cross-User Performance. We evaluate the performance of a pretrained model in a leave-one-user-out setup across the 20 participants in our study. Results in Table 3 show that our system achieves an accuracy of 97.01% and 77.36% without any new-user enrollment for the 4-gestures and 6-gestures tasks respectively.

Effect of Vertical Channels. We assessed the effect of vertically placed electrodes on system performance across 9 participants. As shown in Fig. 11b, 5-fold cross-validation shows that our two-channel configuration achieved 98.61% accuracy across four gestures, while the three- and four-channel configurations result in modest performance gains of 0.28% and 0.42% and accuracies of 98.89% and 99.03% respectively. These modest gains are likely because vergence is primarily, though not exclusively, a horizontal eye movement. However, all participants

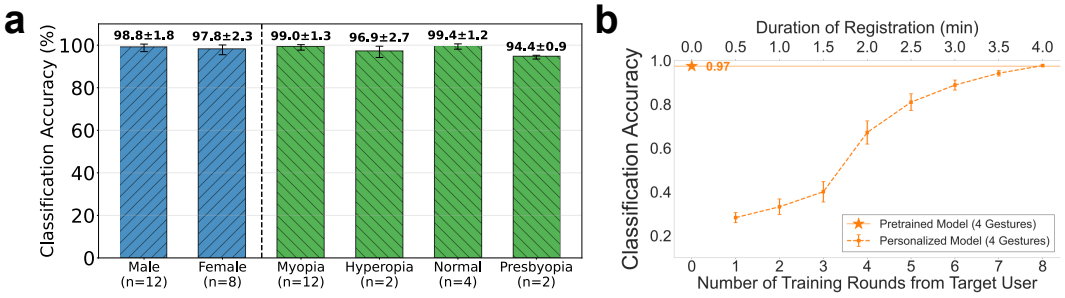


Fig. 13. **(a)** Subgroup analysis on within-session classification accuracy for 4-gestures. **(b)** Effect of training duration on classification accuracy.

reported discomfort with the 3- and 4-channel configurations during data collection. Given these persistent ergonomic concerns, we excluded these configurations from the protocol for the remaining 11 participants. These findings further justify our decision to prioritize a two-channel design for the final system.

5.2 Subgroup Analysis

To evaluate the generalization and robustness of *VergeIO* across diverse user groups, we conducted a subgroup analysis on the within-session data across 2 sessions. We assessed classification performance across demographics and physiological conditions, as shown in Fig. 13a.

Sex. When comparing the classification performance by sex, the female participants ($n = 8$) achieved a mean accuracy of 97.81 ± 2.29% for 4-gesture task and 81.56 ± 4.67% for 6-gesture task. The male participants ($n = 12$) achieved 98.75 ± 1.77% and 83.54 ± 3.98% for 4-gesture and 6-gesture tasks, respectively.

Vision conditions. Participants who normally wear corrective lenses were instructed to wear them with our *VergeIO* glasses on top of them during the study to ensure they could clearly focus on vergence targets at 30, 70 and 200 cm. The mean classification accuracies for the tasks of 4-class and 6-class, respectively, were as follows: 98.96 ± 1.29% and 84.03 ± 4.92% for myopia ($n = 12$); 99.38 ± 1.25% and 82.91 ± 1.98% for normal vision ($n = 4$); 96.88 ± 2.65% and 79.58 ± 0.60% for hyperopia ($n = 2$); and 94.38 ± 0.88% and 77.90 ± 1.77% for presbyopia ($n = 2$).

5.3 Comparison Against Baseline Electrode Layouts

We perform a comparison against two baseline EOG sensing systems. First, we compare our system with Mishra et al.'s work [64], which is focused on vergence sensing using custom aerosol-jet–printed, stretchable, skin-like biopotential electrodes using a 2-, 3-, and 4-channel configurations. Their 2-channel configuration differs from ours and has a horizontal channel across both eyes, a vertical channel above and below the right eye, and a ground electrode obtrusively placed on the forehead, and reports an accuracy of 34%. In contrast, our system uses an unobtrusive 2-channel configuration using standard Ag/AgCl electrodes with the ground on the mastoid and achieves an accuracy of 82.68%. While their 3-channel and 4-channel configurations have increased accuracies of 87 and 91% respectively, additional electrodes bring about discomfort, which negatively affects the user experience.

Second, we compare our system's performance against the EOG electrode layout in the commercial JINS MEME glasses [11] (Fig. 4) and left-right differential layout represented by E-Glasses [53]. To ensure a fair comparison, one participant (P4) wore a combined electrode setup encompassing the positions of both our system and the JINS MEME. We then collected data across ten rounds, each including all six vergence gestures. Evaluating each configuration independently using five-fold cross-validation, our system achieves an accuracy of 100% in

Table 4. Comparison between our system and the baseline proposed by Mishra et al. [64]

	Mishra et al. [64]	***VergeIO* (ours)**
Number of electrodes	5 - 7	4
Number of channels	2 - 3	2
In-session accuracy	34% - 91%	82.68%
Cross-session accuracy	Not specified	77.78%
Cross-user accuracy	Not specified	77.36%
Commodity electrodes	✗	✓
Motion artifact removal	✗	✓
Ground electrode	Forehead (obtrusive)	Mastoid (unobtrusive)
Glasses Compatibility	✗	✓
Angular difference tested	1–3	1.35–4.05

4-gesture task, outperforming the JINS MEME layout (67.5 ± 15%) and the E-Glasses channel (72.5 ± 9.35%). In the more complex 6-gesture task, our system achieved 88.33 ± 1.13% accuracy, whereas JINS MEME and the derived E-Glasses channel achieved only 50.0 ± 1.29% and 53.33 ± 8.50%, respectively. These results are in line with our prior SNR analysis, which demonstrated that our vergence waveforms exhibit an average SNR improvement of 7.3 dB over JINS MEME and 4.8 dB over the E-Glasses configuration.

5.4 Comparison of Classifier Models

We compared our feature-based Random Forest pipeline with 1D-CNN and LSTM models operating directly on filtered two-channel EOG segments. The model architectures and tuning details are provided in Appendix B. As shown in Table 5, Random Forest achieved the highest within-session (82.68 ± 4.29%) and cross-session accuracy (77.78 ± 6.17%), suggesting that the handcrafted features effectively represent the distinguishable patterns of gesture. In the cross-user setting, the random forest and 1D-CNN had comparable results.

Table 5. Comparison of different classifiers for 6 vergence gestures.

Model	**Within-session**	**Cross-session**	**Cross-user**
Random Forest	**82.68 ± 4.29%**	**77.78 ± 6.17%**	77.36 ± 6.17%
1D-CNN	77.13 ± 5.89%	73.56 ± 6.90%	**77.64 ± 4.82%**
LSTM	74.63 ± 6.66%	65.79 ± 7.51%	73.66 ± 5.96%

6 EVALUATIONS TOWARD REAL-WORLD DEPLOYMENT

6.1 Compatibility with Non-Vergence Gestures

Prior smart-glasses EOG interfaces typically support a vocabulary of conventional eye movements, such as blinks, winks, and directional gazes. We therefore evaluate whether our system preserves this compatibility while enabling vergence sensing. We collect a dataset including the four vergence gestures with high separability, along with seven additional eye movements: left and right winks, blinks, and four directional gazes (up, down, left, right). We collected 14 recordings for each class, and performed 5-fold cross-validation. As shown in Fig. 14, our system achieved an accuracy of 96.1 ± 2.4%.

A few misclassifications were observed between upward and downward gaze, which is expected as our design intentionally omits vertical electrode channels to avoid user discomfort. However, our system still captures some vertical eye movements as we intentionally incorporate vertical separation between positive electrodes and reference electrode, as shown in Fig. 4c. These results show that our electrode layout can effectively detect

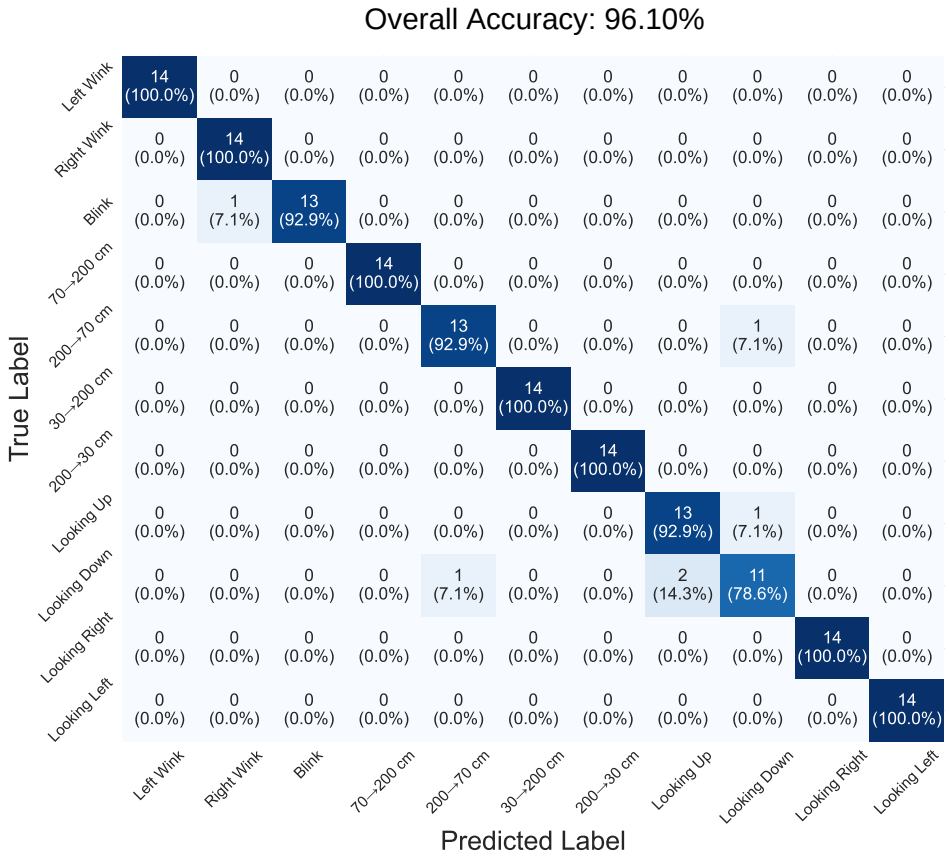


Fig. 14. Confusion matrix classifying between different vergence and non-vergence movements across 11 classes.

eye movements supported by prior smart-glasses EOG systems alongside vergence gestures, expanding the interaction vocabulary available on smart glasses.

6.2 Generalizing to Virtual Depth Cues

Beyond physical Brock String targets, a practical smart-glasses interface must operate with rendered content. We therefore evaluate whether *VergeIO* can detect vergence induced by stereoscopic disparity. This simulates AR/VR applications where users focus on virtual objects rendered at different depths on a fixed-depth display.

To test this, we developed a prototype application on the Meta Quest 2 [13], replicating the Brock string setup described in Sec. 4 within a Unity-based VR environment. Beads were virtually placed at 30, 70, and 200 cm from the user's eye center (Fig. 15a). We then invited five participants to perform the same vergence gestures following the procedure described in Sec. 4, yielding 300 virtual-depth test gestures in total.

We trained our model only on the Brock String dataset, while excluding Brock String sessions from the same target participant when applicable. *VergeIO* achieved an accuracy of 99.58 ± 0.77% and 84.00 ± 4.42% for four- and six-gesture classification respectively (Fig. 15b). This result shows that **(1)** virtual depth overlays can reliably induce vergence, even though they are rendered on a fixed display, and **(2)** a model trained on physical Brock String depth cues can successfully generalize to virtual depth cues.

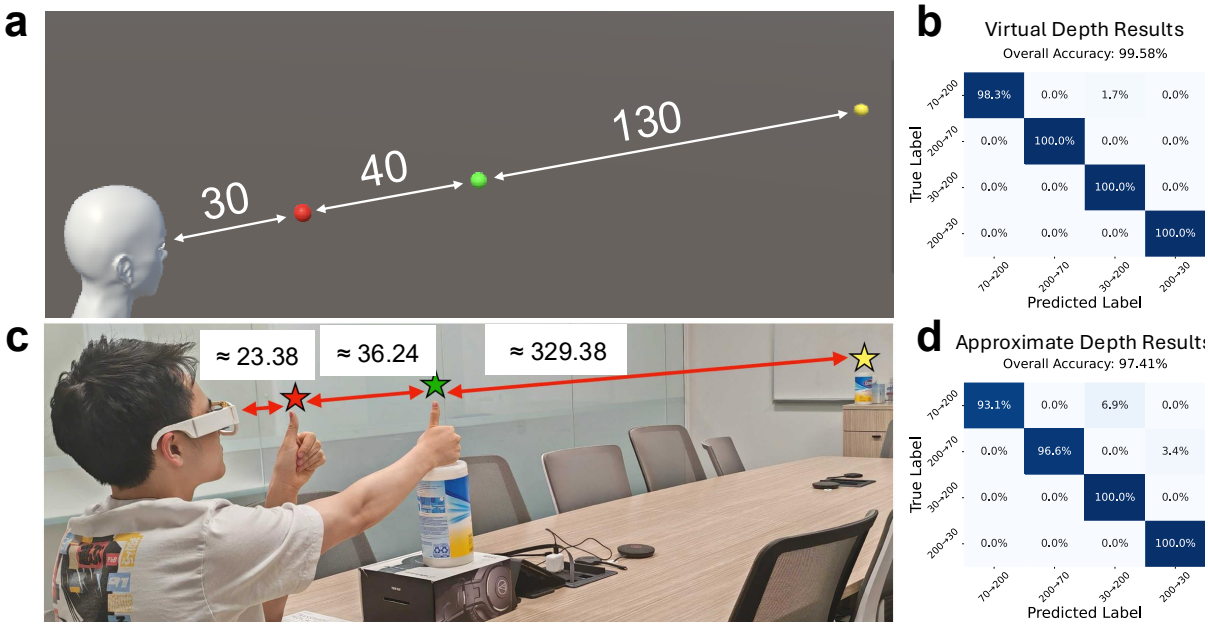


Fig. 15. Experimental setup to evaluate generalization to **(a)** virtual depth cues and **(c)** approximate depth cues. Results are shown as confusion matrices across five participants for the four-gesture task as **(b)** and **(d)** respectively. Distances are in cm.

6.3 Generalizing to Approximate Depth Cues

We next evaluate if our system can work in scenarios where precise physical targets are unavailable. To do this, we propose a body-referenced interaction method where the user fully extends one arm, rests the other hand at the elbow of the extended arm, and aligns both thumbs with a far-field object. This setup approximates the distances of 30 and 70 cm used in the Brock string experimental setup, and corresponds to a half and full arm's length for an average adult [70]. While the Brock string setup used 200 cm as the far-field depth cue, in this scenario the exact depth of the far-field cue is less critical, as vergence-induced EOG signal changes diminish with increasing distance (Fig. 3).

We evaluate generalizability for four gestures as the primary use case, where the user alternates focus between the thumbs and a far-field object regarding far-field target as an anchor. For richer input, this can be extended to the six gesture case using both thumbs and a far-field object as reference points. We had five participants perform the vergence gestures following the protocol in Sec. 4. We note that due to differences in arm length across participants, this led to a natural variation in the depth of the near and middle cue, which averaged 23.38 ± 5.03 cm (ranging from 15.6 to 29.1 cm) and 59.62 ± 6.48 cm (ranging from 51.5 to 68.0 cm), respectively. We used objects at diverse distances as far cues and were at distances averaging 389.00 ± 110.43 cm (ranging from 255.0 to 492.0 cm). We trained our model only on the Brock String dataset, while excluding Brock String sessions from the same target participant when applicable.

VergeIO achieved 84.00 ± 4.42% accuracy on the six-class task. For the four-gesture task, the model achieved 97.41 ± 4.04% accuracy over the 10 rounds data (Fig. 15d). When collapsing predictions into a binary direction-level task: near-to-far vs. far-to-near, the model achieved 100% accuracy, indicating that opposite-direction depth transitions were not confused. These results demonstrate that our system generalizes to approximate physical depth interactions without requiring the physical cues to exactly match the original Brock String distances.

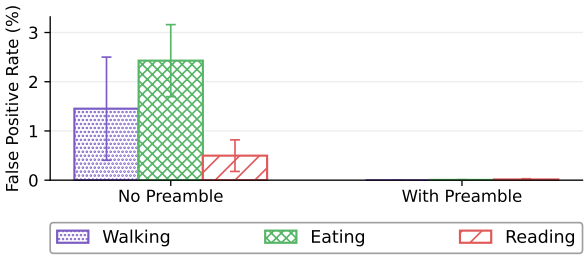


Fig. 16. False positive rates across everyday activities: eating, walking, and reading. A preamble consisting of an eyebrow raise to activate the system reduces the false positives to nearly zero across all three conditions.

6.4 Motion Artifact and Noise Removal Performance

While previous evaluations focused on vergence detection, real-world deployment poses a challenge in rejecting artifacts caused by facial, eye, and body movements. We therefore evaluate whether our artifact-removal pipeline can identify and reject these non-vergence events while reliably preserving vergence inputs.

We evaluated the motion artifact and noise removal pipeline on a dataset collected from $n = 6$ participants (Table 2). The evaluation employs a 2-second moving-window pipeline. We split training and test sets since neighboring windows are highly correlated. This prevents data leakage by ensuring that overlapping windows from the same recording never appear in both sets.

We used a personalized artifact-removal model where participants would perform 6 minutes of calibration: 3 minutes for six rounds of vergence data and 3 minutes for one artifact round for each of the six motion/noise activities. This resulted in an accuracy of 98.77 ± 1.26% across six participants, with a false positive rate of 1.23 ± 1.26%. Detailed calibration-budget ablations and cross-user adaptation results are provided in Appendix C.

6.5 False Positives from Non-Vergence Activity

We further conducted an online false-positive evaluation during non-vergence sessions where participants wore the prototype while performing daily activities. The artifact-removal pipeline ran in real-time on the prototype during the evaluation. Six participants completed three tasks: eating while using a phone or watching a movie (10.52 ± 0.41 minutes), walking while conversing (11.12 ± 0.45 minutes), and reading a document on a computer screen (10.83 ± 0.53 minutes). To broaden evaluation environments, eating and reading were conducted in separate indoor rooms, while walking was conducted outdoors.

In this evaluation, a false positive is defined as a non-vergence sliding window falsely accepted by the system. As shown in Fig. 16, using the frozen classifier trained on the data described in Sec. 6.4, we obtained a false positive rate of 2.43 ± 0.73% during eating. Walking while conversing and reading on a computer produced lower false positive rates of 1.45 ± 1.05% and 0.50 ± 0.32%, respectively. To further reduce the false positive rates, we introduced a brow-raise preamble gesture to explicitly activate and deactivate the system. The motion is distinct from other movements, and results in an EOG amplitude that is significantly higher than vergence (Fig. 7), and it is easy and quick to perform. After incorporating this preamble, the false positive rate dropped to nearly zero across all three activities: 0.005% ± 0.005% for eating, 0% for walking, and 0.015% ± 0.015% for reading.

6.6 Power Consumption

We measured the power consumption of our system using the Monsoon High Voltage Power Monitor [14] at 3.7 V, over 5 runs and averaged them, with each run lasting 1 minute at the lowest clock speed of 24 MHz.

The power consumption of the two amplifier front-ends was 3.3 ± 0.004 mW. This is lower than other eye tracking systems, in particular mmWave radios [58], and microphones and speakers [55], though we note that our system is purpose-built for the specific task of vergence gesture sensing not eye-tracking.

The code for data acquisition (digitization, reading samples into buffers) consumed 4.84 ± 0.02 mW; while data processing (filtering, segmenting, classification) consumed 2.63 ± 0.02 mW.

With a 570 mAh rechargeable lithium-ion battery as in Google Glass [8], the system would operate for approximately 197 hours (8.2 days). This power draw is effectively negligible, as the overall battery life of Google Glass is dominated by other processes that result in a typical battery life of 8 hours [9]. Further, we note that a productized version of the system would consolidate the sensing, data acquisition and processing into a low-power ASIC or SoC. However, the design of an integrated system is out of scope for this work.

We note that other standby processes on the microcontroller consume on average 138.2 ± 0.0741 mW. However, we do not consider this as part of the power consumption of our proof-of-concept system, as a commercial EOG-based system would leverage a low-power ASIC or SoC, which we leave for future work.

6.7 System Latency

We evaluate whether the processing pipeline can support real-time interaction by measuring the wall-clock time of our data processing pipeline, which continuously filters and classifies the streaming EOG signal 2-second rolling windows with a 0.1-second overlap. The average latency across 2500 windows was 15 ± 0.002 ms, which is less than the step duration of 0.1 seconds and demonstrates our system's ability to run in real-time.

6.8 User Experience Survey

Finally, we assess subjective aspects of the user experience including electrode comfort, facial-movement interference, perceived fatigue, accidental-trigger concern, and social acceptability. We evaluated the user experience during the experiment with a survey. We asked the following questions:

(1) How easy is it to keep the electrodes on? (1 = Hard, 5 = Very easy to keep on)
(2) How easy was it to move your facial muscles (e.g. raising your eyebrows, closing your eyelids)? (1 = Very difficult, 5 = Not Difficult)
(3) How concerned are you about accidentally performing these eye gestures in your daily life? (1 = Very concerned, 5 = Not at all)
(4) To what extent did you feel physical or mental fatigue when performing an eye gesture with the system? (1 = High fatigue, 5 = No fatigue)
(5) How comfortable would you feel performing these eye gestures in a public setting? (1 = Very uncomfortable, 5 = Very comfortable)
(6) How comfortable would you feel performing these eye gestures in a private setting? (1 = Very uncomfortable, 5 = Very comfortable)

The survey results are summarized in Fig. 18. Participants (n = 20) had an average age of 25.75 ± 11.1 years and height of 171.83 ± 7.65 cm. Participants reported that the electrodes remained securely in place (3.85 ± 0.93) and that facial movements remained easy to perform (4.10 ± 0.98). Comfort was higher in private settings (4.75 ± 0.55) than in public settings (3.75 ± 0.91), while fatigue rating was 3.55 ± 0.89. Additionally, users also reported confidence in in performing the movement intentionally on command, with the concern about accidentally performing the gesture in daily life received an average score of 3.75 ± 1.25.

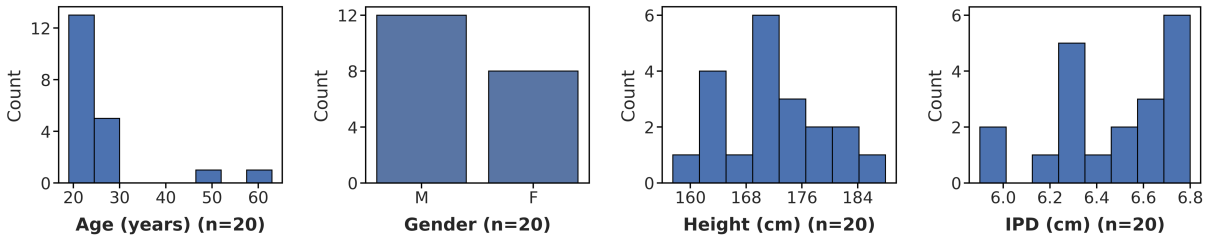


Fig. 17. **Demographic summary of participants in the user study.** Survey data was collected on $n = 20$ participants in the data collection procedure.

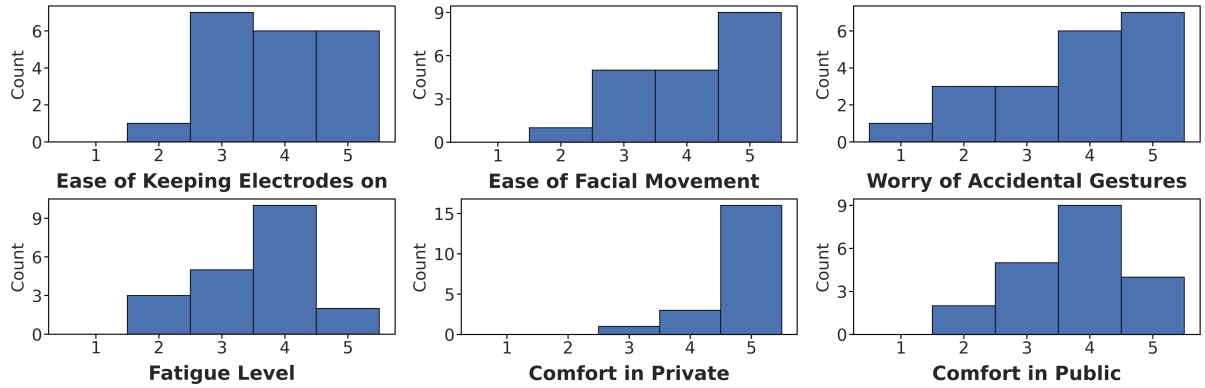


Fig. 18. **User experience survey ($n = 20$).** Histograms summarize user perceptions across (a) ease of maintaining electrode contact, (b) ease of facial muscle movement while wearing the prototype, (c) concern regarding accidental gesture triggers, (d) interaction-induced fatigue, and the social acceptability and comfort of the interaction in (e) private and (f) public settings.

7 EXAMPLE APPLICATIONS

7.1 Vergence-Driven Varifocal Glasses

Varifocal glasses adjust optical focus to match the user's viewing distance [67], helping billions of people with impaired accommodation, including presbyopia, focus across near and far distances [38]. Commercial efforts including IXI [10], ViXion01 [19], and Elcyo [5] demonstrate growing interest in controlling varifocal glasses automatically, but these systems rely on camera-based sensing, which can be power-intensive for continuous wear. In contrast, we present the first attempt to leverage EOG signals as a low-power, lightweight modality for future automatic varifocal control. Importantly, although the eye lens' focusing ability is impaired, individuals can still perform vergence movements betweeen near and far distances [67]. We have also validated this in Sec. 5.2, where our system achieved 94.38 ± 0.62% 4-gesture accuracy for individuals with presbyopia.

Our prototype (Fig. 19a) instruments an existing pair of varifocal glasses [7] with a small servo motor [1] (22.8 mm×12.2 mm×28.5 mm) which is coupled to a knob to adjust the focal length. The motor weighs 29 g, which is relatively modest compared to commercial eye-tracking glasses such as the Tobii Pro Glasses 3 (388 g) [18].

To show that our adjustable varifocal glasses can support focusing at distances of 30, 70, and 200 cm, we conducted a laser-based experiment to measure the effective focal lengths of the lens. The goal was to determine the focal length corresponding to different rotations of the glasses' adjustment knob. In the setup (Fig. 19b), a laser pointer was aimed at a distant wall with the varifocal lens placed in between. Let d be the distance from the

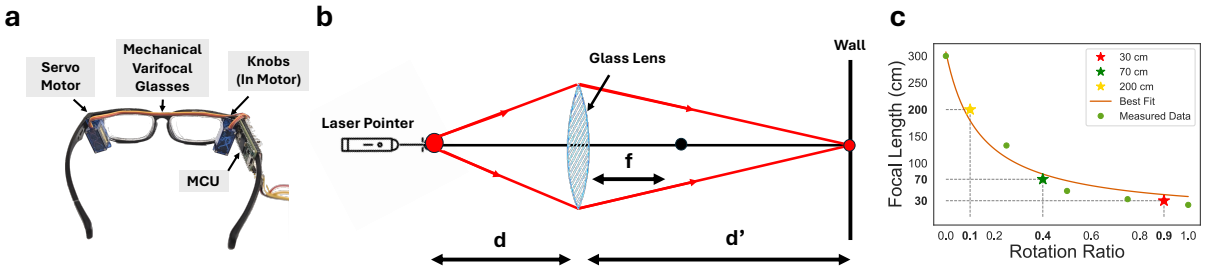


Fig. 19. **(a)** Prototype of mechanically adjustable varifocal glasses with an integrated servo motor for actuation. **(b)** Experimental setup to determine the effective focal length at each varifocal lens setting. **(c)** Effective focal length of the varifocal glasses for different knob rotation ratios.

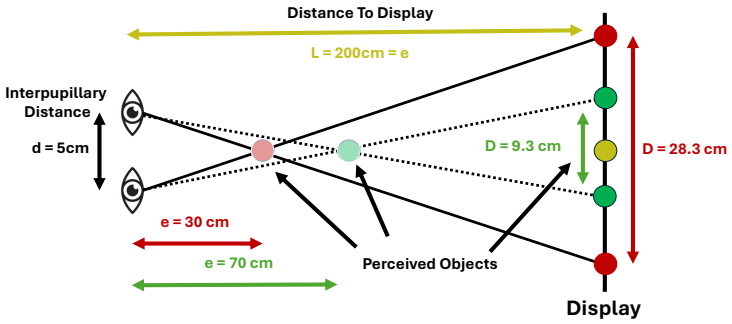


Fig. 20. Stereoscopic depth rendering is achieved by offsetting the left and right images to simulate the three depths (30, 70, 200 cm) supported by our system.

laser to the lens and d' the distance from the lens to the wall. We adjusted d' by moving the varifocal glasses until the laser spot appeared in focus, then computed the effective focal length f using the lens equation [17]: $f = \frac{d \cdot d'}{d + d'}$. The total distance between the laser and the wall was 12 m, which allowed us to measure focal lengths up to 3 m.

This procedure was repeated for five normalized knob rotations ranging from 0.0 to 1.0. As shown in Fig. 19c, the effective focal length spans approximately 25.58 cm to over 300 cm, covering the key distances required for near-to-far visual accommodation in individuals with presbyopia.

7.2 Stereoscopic Depth Rendering of Virtual Objects

Modern AR displays often render virtual objects at varying depths using stereoscopic rendering, which displays slightly different images to each eye to create a sense of depth from binocular disparity [26, 86]. Vergence sensing can be used to align virtual content with the user's depth of focus. Importantly, although all virtual content is rendered on a screen located at a fixed physical distance, stereoscopic rendering can still induce vergence. Fig. 20 illustrates the stereoscopic depth estimation formula [73]: $D = d \cdot \left(\frac{L}{e} - 1\right)$ which computes the required separation D between the left and right projected images to make an object appear at a desired virtual depth e. Here, L is the distance between the eyes and the screen, and d is the interpupillary distance.

In the example, the user is positioned $L = 200\ cm$ from the screen, with an interpupillary distance of $d = 5\ cm$. To render an object at $e = 30\ cm$ and $e = 70\ cm$, the images would need to be separated by $D = 28.3\ cm$ and $D = 9.3\ cm$ respectively. While to render it at $e = 200\ cm$, which matches the physical distance to the screen, the images should overlap, resulting in zero separation.

8 DISCUSSION, LIMITATIONS, AND FUTURE WORK

Interaction value of vergence. Existing EOG glasses can detect blinks, winks, dwell-based gaze, and directional gaze for generic commands or 2D angular selection. In contrast, vergence offers two complementary interaction benefits. First, it provides a natural mapping between a shift in gaze depth and an action associated with the corresponding depth cue. Furthermore, vergence can also serve as an intentional cue that has been leveraged to mitigate the "Midas touch" problem [22, 89], in which natural eye movements detectable by prior EOG systems are mistaken for deliberate commands [42].

Learnability of vergence. Vergence is a learnable and controllable skill even without visual reference points, enabling users to perform intentional depth-based eye gestures with training in a few minutes [35, 47, 51, 63, 72, 89]. An example voluntary control is perceiving a 3D autostereogram [83] using cross-eyed or wall-eyed viewing.

Practical vocabulary size for vergence gestures. Evaluation results show that our system reliably supports 4-gesture vergence interaction, achieving over 97% accuracy both after remounting and when generalizing to new users without calibration. In contrast, our system is less reliable when detecting all six gestures across near, mid, and far distances. Most errors come from the 30 $\leftrightarrow$ 70 cm transitions, as the signals overlap with other gestures. This issue may come from both sensor limitations and variability in the oculomotor response. From a hardware perspective, prior work shows that even with high-quality wet electrodes, tracking both eyes has an uncertainty of about 1.5° for each eye [88], resulting in a 3° uncertainty. In our setup, the 2.7° binocular angular difference is inherently smaller than this theoretical threshold, making it difficult to separate reliably using dry electrodes. From a physiological perspective, the human vergence response consists of a fast open-loop initiation followed by a slower, visually guided closed-loop feedback phase [79]. The resulting inter-subject variability in execution speed and trajectory makes gesture boundaries even harder to distinguish. In summary, fine-grained vergence sensing on EOG glasses remains challenging and we recommend using four-gesture as the default interaction vocabulary.

Vergence with gaze direction changes. Prior work shows that simultaneous changes in gaze direction and depth can alter vergence dynamics [52]. Such changes were unavoidable in our body-referenced study in Sec. 6.3, yet our system achieved 97.41% accuracy. Future work should systematically examine the joint effects of gaze direction and depth with the help of high-precision eye trackers.

Depth granularity. Our choice of depth distances is grounded in the practical system design discussed in Sec. 3.1. The selected near, mid, and far distances not only cover common viewing ranges but also align with the focal zones used in optical systems. However, our current study does not explore distances at a finer granularity. For example, we have not investigated the minimum angular difference required to distinguish two gestures. It also remains unclear whether the same angular change produces similar EOG responses at near and far viewing distances. A deeper study of depth would require more distance levels, finer spacing between targets, and synchronized ground-truth eye tracking to analyze how vergence angle, movement dynamics, and EOG responses vary across various distances. We view this as an important direction for future work.

Adapting to different anatomies. A fundamental challenge for EOG-based eye movement sensing systems is balancing robust electrode-skin contact with user comfort. Our design addresses this balance through optimized structural features including a flexible nose-bridge and curved mastoid support. Future work could also leverage signal quality and user feedback to dynamically guide device positioning or trigger automated micro-adjustments.

Preamble mechanism to mitigate accidental triggers. To support system reliability against daily facial activities (e.g. chewing, speaking), our current design incorporates raising eyebrows as an explicit preamble for activation. Although this mechanism effectively mitigates false positives, requiring a deliberate physical gesture may be intrusive in social contexts and interrupt natural interaction flows. We aim to evolve towards implicit activation mechanisms in future iterations, potentially by fusing EOG with lightweight IMU data or leveraging context-aware algorithms to seamlessly discern interaction intent without conscious user effort.

New user enrollment. Although *VergeIO* can distinguish four vergence gestures on unseen users without calibration, we found that artifact rejection remains user-dependent as electrode-skin contact, facial anatomy, and motion patterns differ across users. *VergeIO* therefore requires a new user to perform six repetitions of each gesture and approximately three minutes of motion/noise activities to personalize artifact rejection.
Long-term usage. Our evaluation was conducted primarily in controlled settings. For future work, longitudinal in-the-wild studies are needed to assess system performance in the presence of perspiration, electrode displacement, skin-impedance drift, spontaneous facial movements, as well as user comfort over hours or days.

9 CONCLUSION

This paper presented *VergeIO*, the first EOG-based glasses system that enables robust, depth-aware eye interaction by sensing vergence. With a glasses-compatible electrode layout with positive electrodes on the temples and a shared negative electrode on the nose bridge, *VergeIO* preserves the opposing-polarity signals from the two eyes and captures high-SNR vergence responses. Our evaluation across 20 users demonstrated that *VergeIO* can distinguish four vergence gestures on unseen users with 97.01% accuracy without calibration. With a front-end power consumption of only 3 mW and a processing latency of 15 ms, the system supports low-power, real-time vergence sensing. We envision that the sensing capability introduced by *VergeIO* will broaden the scope of eye interaction available on EOG-based glasses, serving as a lightweight and privacy-preserving primitive that enables a new generation of depth-aware applications in consumer electronics, healthcare, and safety-monitoring domains.

10 LLM USE CLARIFICATION

Gemini-3 Pro was used for generating human characters in Fig. 10, as well as for grammar correction and text refinement throughout this paper.

References

[1] 2026. 2PCS SG90 Servo Motor Micro Servo 9G Servo Motor for RC Robot Arm Helicopter Airplane Remote Control (2X). (2026). https://www.amazon.com/Servo-Helicopter-Airplane-Remote-Control/dp/B09L525KDT/ref=asc_df_B09L525KDT.
[2] 2026. Adafruit Micro-Lipo Charger for LiPoly Batt with USB Type C Jack. (2026). https://www.adafruit.com/product/4410.
[3] 2026. Amazon Echo Frames. (2026). https://www.amazon.com/b?ie=UTF8&node=117191383011
[4] 2026. BioAmp EXG Pill. (2026). https://www.digikey.com/en/products/detail/upside-down-labs/0047/19525716.
[5] 2026. Elcyo. (2026). https://en.elcyo.com/
[6] 2026. Eye detection methods and devices. (2026). https://patents.google.com/patent/US20240134449A1/en?oq=US20240134449
[7] 2026. Flex Focal Adjustable Glasses Dial Vision Reading Glasses 0D to 350D Zoom Eyeglasses Men and Women. (2026). https://www.amazon.com/PUDOEN-Adjustable-Glasses-Reading-Eyeglasses/dp/B0GD5PP3XV/.
[8] 2026. Google Glass review. (2026). https://www.techradar.com/reviews/gadgets/google-glass-1152283/review/7.
[9] 2026. Google Glass smart eyewear returns. (2026). https://www.bbc.com/news/technology-40644195.
[10] 2026. IXI – The world's first autofocus eyewear. (2026). https://ixieyewear.com/
[11] 2026. JINS MEME. (2026). https://jinsmeme.com/en/.
[12] 2026. Learn about IPD and lens spacing on Meta Quest. (2026). https://www.meta.com/help/quest/261777072346131/.
[13] 2026. Meta Quest 2. (2026). https://www.meta.com/quest/products/quest-2/.
[14] 2026. Monsoon High Voltage Power Monitor. (2026). https://www.msoon.com/high-voltage-power-monitor.
[15] 2026. Pupil Lab Neon Eye Tracking Glasses. (2026). https://pupil-labs.com/products/neon.
[16] 2026. Teensy 4.1. (2026). https://www.sparkfun.com/teensy-4-1.html.
[17] 2026. Thin Lens Equation. (2026). https://openstax.org/books/university-physics-volume-3/pages/2-4-thin-lenses.
[18] 2026. Tobii Pro Glasses 3. (2026). https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3.
[19] 2026. ViXion01. (2026). https://vixion.jp/vixion01/?wovn=en
[20] Aws M Abdullah, Ali R Ibrahim, Ammar A Al-Hamadani, Mohammed K Al-Obaidi, and Anas F Ahmed. 2023. A review for filtering techniques of the Electrooculography (EOG) signals. *Glob. J. Eng. Technol. Adv* 16 (2023), 163–171.

[21] Deepshikha Acharya, Asha Rani, Shivangi Agarwal, and Vijander Singh. 2016. Application of adaptive Savitzky–Golay filter for EEG signal processing. *Perspectives in Science* 8 (2016), 677–679. Recent Trends in Engineering and Material Sciences. doi:10.1016/j.pisc.2016.06.056

[22] Sunggeun Ahn, Jeongmin Son, Sangyoon Lee, and Geehyuk Lee. 2020. Verge-it: Gaze interaction for a binocular head-worn display using modulated disparity vergence eye movement. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–7.

[23] Moez Ali. 2020. *PyCaret: An open source, low-code machine learning library in Python*. PyCaret version 1.0. https://www.pycaret.org

[24] Rafael Barea, Luciano Boquete, Manuel Mazo, and Elena López. 2002. System for assisted mobility using eye movements based on electrooculography. *IEEE transactions on neural systems and rehabilitation engineering* 10, 4 (2002), 209–218.

[25] Chama Belkhiria, Atlal Boudir, Christophe Hurter, and Vsevolod Peysakhovich. 2022. Eog-based human–computer interface: 2000–2020 review. *Sensors* 22, 13 (2022), 4914.

[26] Laurent Blondé, Didier Doyen, and Thierry Borel. 2010. 3D stereo rendering challenges and techniques. In *2010 44th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 1–6.

[27] Andreas Bulling, Daniel Roggen, and Gerhard Tröster. 2009. Wearable EOG goggles: eye-based interaction in everyday environments. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*. 3259–3264.

[28] Yingxi Chen and Wyatt S Newman. 2004. A human-robot interface based on electrooculography. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, Vol. 1. IEEE, 243–248.

[29] Yun-Hsuan Chen, Maaike Op de Beeck, Luc Vanderheyden, Evelien Carrette, Vojkan Mihajlović, Bernard Grundlehner, Stefanie Gadeyne, Paul Boon, and Chris Van Hoof. 2014. Soft, Comfortable Polymer Dry Electrodes for high Quality ECG and EEG Recording. *Sensors (Basel, Switzerland)* 14, g014. doi:10.3390/ecsa-1-g014

[30] Chul Woo Cho, Ji Woo Lee, Kwang Yong Shin, Eui Chul Lee, Kang Ryoung Park, Heekyung Lee, and Jihun Cha. 2012. Gaze Detection by Wearable Eye-Tracking and NIR LED-Based Head-Tracking Device Based on SVR. *Etri Journal* 34, 4 (2012), 542–552.

[31] Neil A Dodgson. 2004. Variation and extrema of human interpupillary distance. In *Stereoscopic displays and virtual reality systems XI*, Vol. 5291. SPIE, 36–46.

[32] Marcos Duarte. 2021. *detecta: A Python module to detect events in data*. doi:10.5281/zenodo.4598962

[33] B. Estrany and Pilar Fuster-Parra. 2022. Human Eye Tracking Through Electro-Oculography (EOG): A Review. In *Cooperative Design, Visualization, and Engineering*, Yuhua Luo (Ed.). Springer International Publishing, Cham, 75–85.

[34] Sebastian Frey, Mattia Alberto Lucchini, Victor Kartsch, Thorir Mar Ingolfsson, Andrea Helga Bernardi, Michael Segessenmann, Jakub Osieleniec, Simone Benatti, Luca Benini, and Andrea Cossettini. 2024. GAPses: Versatile smart glasses for comfortable and fully-dry acquisition and parallel ultra-low-power processing of EEG and EOG. *IEEE Transactions on Biomedical Circuits and Systems* (2024).

[35] Eugene Hamori, Roy D Broad, and Calvin E Reed Jr. 1982. Study of unaided cross-eyed stereopsis. *Perception* 11, 3 (1982), 297–304.

[36] Layali Ibrahim Hassan, Samira Mohamed Ibrahim, and Mustafa Abdu. 2017. Efficacy of home-based vision therapy for convergence insufficiency in secondary schools' students. *Sudanese Journal of Ophthalmology* 9, 1 (2017), 5–9.

[37] Patrick Heijden, Camille Gilbert, Samira Jafari, and Mattia Lucchini. 2024. Multi-Channel Soft Dry Electrodes for Electrocardiography Acquisition in the Ear Region. *Sensors* 24 (01 2024), 420. doi:10.3390/s24020420

[38] Brien A Holden, Timothy R Fricke, S May Ho, Reg Wong, Gerhard Schlenther, Sonja Cronjé, Anthea Burnett, Eric Papas, Kovin S Naidoo, and Kevin D Frick. 2008. Global vision impairment due to uncorrected presbyopia. *Archives of ophthalmology* 126, 12 (2008), 1731–1739.

[39] Mark S. Hughes. 1991. Dictionary of Eye Terminology. *Archives of Ophthalmology* 109, 9 (09 1991), 1208–1208. arXiv:https://jamanetwork.com/journals/jamaophthalmology/articlepdf/639252/archopht_109_9_015.pdf doi:10.1001/archopht.1991.01080090032015

[40] International Organization for Standardization. 2011. ISO 9241-303:2011: Ergonomics of Human-System Interaction—Part 303: Requirements for Electronic Visual Displays. https://www.iso.org/standard/57992.html.

[41] Shoya Ishimaru, Kai Kunze, Yuji Uema, Koichi Kise, Masahiko Inami, and Katsuma Tanaka. 2014. Smarter eyewear: Using commercial EOG glasses for activity recognition. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. 239–242.

[42] Robert J. K. Jacob. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '90)*. Association for Computing Machinery, New York, NY, USA, 11–18. doi:10.1145/97243.97246

[43] Elizabeth Carolina Jiménez, August Romeo, Laura Pérez Zapata, Maria Solé Puig, Patricia Bustos-Valenzuela, José Cañete, Paloma Varela Casal, and Hans Supèr. 2020. Eye vergence responses in children with and without reading difficulties during a word detection task. *Vision Research* 169 (2020), 6–11.

[44] S. Kanoh, S. Ichi-nohe, S. Shioya, K. Inoue, and R. Kawashima. 2015. Development of an eyewear to measure eye and body movements. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2267–2270. doi:10.1109/EMBC.2015.7318844

[45] Minsol Kim, Serena Pei, Samiksha Singh, Selam Brook, and Nataliya Kosmyna. 2024. Privacy-Preserving Eye Movement Classification with Camera-Free, Non-Invasive EOG-EEG Glasses. In *2024 IEEE 20th International Conference on Body Sensor Networks (BSN)*. IEEE, 1–4.

[46] Tobias King, Michael Knierim, Philipp Lepold, Christopher Clarke, Hans Gellersen, Michael Beigl, and Tobias Röddiger. 2025. earEOG via Periauricular Electrodes to Facilitate Eye Tracking in a Natural Headphone Form Factor. *arXiv preprint arXiv:2506.07193* (2025).

[47] Dominik Kirst and Andreas Bulling. 2016. On the verge: Voluntary convergences for accurate and precise timing of gaze input. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1519–1525.

[48] Nataliya Kosmyna, Karim Adl, and Minsol Kim. 2024. Wearable Pair of EEG, EOG and fNIRS Glasses for Cognitive Workload Detection. 1–4. doi:10.1109/BSN63547.2024.10780518

[49] Nataliya Kosmyna, Caitlin Morris, Utkarsh Sarawgi, Thanh Nguyen, and Pattie Maes. 2019. AttentivU: A wearable pair of EEG and EOG glasses for real-time physiological processing. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 1–4.

[50] Nataliya Kosmyna, Utkarsh Sarawgi, and Pattie Maes. 2018. AttentivU: Evaluating the feasibility of biofeedback glasses to monitor and improve attention. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. 999–1005.

[51] Shinya Kudo, Hiroyuki Okabe, Taku Hachisu, Michi Sato, Shogo Fukushima, and Hiroyuki Kajimoto. 2013. Input method using divergence eye movement. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 1335–1340.

[52] Arun N Kumar, Yanning Han, Louis F Dell'Osso, Dominique M Durand, and R John Leigh. 2005. Directional asymmetry during combined saccade–vergence movements. *Journal of Neurophysiology* 93, 5 (2005), 2797–2808.

[53] Joong Hoon Lee, HanSeop Kim, Ji-Young Hwang, Jinmook Chung, Tae-Min Jang, Dong Seo, Yuyan Gao, Junhyun Lee, Haedong Park, Seungwoo Lee, Hong Moon, Huanyu Cheng, Sang-Hoon Lee, and Suk-Won Hwang. 2020. 3D Printed, Customizable and Multi-functional Smart Electronic Eyeglasses (E-glasses) for Wearable Healthcare Systems and Human-Machine Interfaces. *ACS Applied Materials Interfaces* (04 2020). doi:10.1021/acsami.0c03110

[54] Philipp Lepold, Tobias Röddiger, Tobias King, Kai Kunze, Christoph Maurer, and Michael Beigl. 2024. OpenEarable ExG: Open-Source Hardware for Ear-Based Biopotential Sensing Applications. In *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 916–920.

[55] Ke Li, Ruidong Zhang, Boao Chen, Siyuan Chen, Sicheng Yin, Saif Mahmud, Qikang Liang, François Guimbretière, and Cheng Zhang. 2024. GazeTrak: Exploring Acoustic-based Eye Tracking on a Glass Frame. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*. 497–512.

[56] Tianxing Li and Xia Zhou. 2018. Battery-free eye tracker on glasses. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 67–82.

[57] Peichen Liu, Sadasivan Puthusserypady, I Scott MacKenzie, Cihan Uyanik, and John Paulin Hansen. 2025. EarEOG: Using Headphones and Around-the-Ear EOG Signals for Real-Time Wheelchair Control. *Proceedings of the ACM on Human-Computer Interaction* 9, 3 (2025), 1–16.

[58] Ruichun Ma, Yasuo Morimoto, John S Ho, Sam Shiu, and Jiang Zhu. 2025. mmET: mmWave Radar-Based Eye Tracking on Smart Glasses. In *Proceedings of the 23rd ACM Conference on Embedded Networked Sensor Systems*. 30–42.

[59] Hiroyuki Manabe, Masaaki Fukumoto, and Tohru Yagi. 2013. Conductive rubber electrodes for earphone-based eye gesture input interface. In *Proceedings of the 2013 International Symposium on Wearable Computers*. 33–40.

[60] Gary L Mancil, Ian L Bailey, Kenneth E Brookman, J Bart Campbel, Michael H Cho, Alfred A Rosenbloom, and James E Sheedy. 2011. Optometric clinical practice guideline: Care of the patient with presbyopia. *St. Louis, MO: American Optometric Association* (2011).

[61] Addison Mayberry, Pan Hu, Benjamin Marlin, Christopher Salthouse, and Deepak Ganesan. 2014. iShadow: design of a wearable, real-time mobile gaze tracker. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. 82–94.

[62] Addison Mayberry, Yamin Tun, Pan Hu, Duncan Smith-Freedman, Deepak Ganesan, Benjamin M Marlin, and Christopher Salthouse. 2015. CIDER: Enabling robustness-power tradeoffs on a computational eyeglass. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 400–412.

[63] LN McLin and CM Schor. 1988. Voluntary effort as a stimulus to accommodation and vergence. *Investigative ophthalmology & visual science* 29, 11 (1988), 1739–1746.

[64] Saswat Mishra, Yun-Soung Kim, Jittrapol Intarasirisawat, Young-Tae Kwon, Yongkuk Lee, Musa Mahmood, Hyo-Ryoung Lim, Robert Herbert, Ki Jun Yu, Chee Siang Ang, et al. 2020. Soft, wireless periocular wearable electronics for real-time detection of eye vergence in a virtual reality toward mobile eye therapies. *Science advances* 6, 11 (2020), eaay1729.

[65] F Mizuno, T Hayasaka, K Tsubota, S Wada, and T Yamaguchi. 2003. Development of hands-free operation interface for wearable computer-hyper hospital at home. In *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, Vol. 4. IEEE, 3740–3743.

[66] Barath S Narayan, Sreyas Janamanchi, Achintya Harsha, Prashanth Jonna, Madhav Rao, et al. 2024. EOG-Assist: Real-Time EOG-Based Eye Tracking for Assistive Systems. In *2024 18th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 795–800.

[67] Nitish Padmanaban, Robert Konrad, and Gordon Wetzstein. 2018. Autofocals: gaze-contingent eyeglasses for presbyopes. In *ACM SIGGRAPH 2018 Emerging Technologies*. 1–2.

[68] Yun Suen Pai, Marsel L. Bait, Juyoung Lee, Jingjing Xu, Roshan L Peiris, Woontack Woo, Mark Billinghurst, and Kai Kunze. 2022. NapWell: An EOG-based Sleep Assistant Exploring the Effects of Virtual Reality on Sleep Onset. *Virtual Real.* 26, 2 (June 2022), 437–451. doi:10.1007/s10055-021-00571-w

[69] David W Patmore and R Benjamin Knapp. 1998. Towards an EOG-based eye tracker for computer control. In *Proceedings of the third international ACM conference on Assistive technologies*. 197–203.

[70] DF Roberts, KA Provins, and RJ Morton. 1959. Arm strength and body dimensions. *Human Biology* 31, 4 (1959), 334–343.

[71] Soha Rostaminia, Alexander Lamson, Subhransu Maji, Tauhidur Rahman, and Deepak Ganesan. 2019. W!nce: Unobtrusive sensing of upper facial action units with eog-based eyewear. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 1 (2019), 1–26.

[72] Lingyan Ruan, Bin Chen, and Miu-Ling Lam. 2018. Human-computer interaction by voluntary vergence control. In *SIGGRAPH Asia 2018 Posters*. 1–2.

[73] John T. Rule. 1941. The Geometry of Stereoscopic Projection. *J. Opt. Soc. Am.* 31, 4 (Apr 1941), 325–334. doi:10.1364/JOSA.31.000325

[74] Wayne J Ryan, Andrew T Duchowski, and Stan T Birchfield. 2008. Limbus/pupil switching for wearable eye tracking under variable lighting conditions. In *Proceedings of the 2008 symposium on Eye tracking research & applications*. 61–64.

[75] Abraham Savitzky and Marcel JE Golay. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry* 36, 8 (1964), 1627–1639.

[76] Nicolas Schärer, Federico Villani, Aishwarya Melatur, Steven Peter, Tommaso Polonelli, and Michele Magno. 2024. ElectraSight: Smart Glasses with Fully Onboard Non-Invasive Eye Tracking Using Hybrid Contact and Contactless EOG. *arXiv preprint arXiv:2412.14848* (2024).

[77] Nicolas Schärer, Federico Villani, Aishwarya Melatur, Steven Peter, Tommaso Polonelli, and Michele Magno. 2025. Electrasight: Fully onboard eye tracking for smart glasses with hybrid eog (heog). *IEEE Internet of Things Journal* (2025).

[78] M. Scheiman and B. Wick. 2008. *Clinical Management of Binocular Vision: Heterophoric, Accommodative, and Eye Movement Disorders*. Wolters Kluwer Health/Lippincott Williams & Wilkins. https://books.google.com/books?id=jGGROHBFYt8C

[79] John L Semmlow, Chang Yaramothu, and Tara L Alvarez. 2019. Dynamics of the disparity vergence slow (fusion sustaining) component. *Journal of eye movement research* 12, 4 (2019), 34.

[80] Cory Stevenson, Tzyy-Ping Jung, and Gert Cauwenberghs. 2015. Estimating direction and depth of visual fixation using electrooculography. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 841–844.

[81] Benjamin Tag, Andrew W. Vargo, Aman Gupta, George Chernyshov, Kai Kunze, and Tilman Dingler. 2019. Continuous Alertness Assessments: Using EOG Glasses to Unobtrusively Monitor Fatigue Levels In-The-Wild. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3290605.3300694

[82] Domenico Tringali, Dorian Haci, Federico Mazza, Konstantin Nikolic, Danilo Demarchi, and Timothy G Constandinou. 2021. Eye Accommodation Sensing for Adaptive Focus Adjustment. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*. 7460–7464. doi:10.1109/EMBC46164.2021.9629844

[83] Christopher W Tyler and Maureen B Clarke. 1990. Autostereogram. In *Stereoscopic displays and applications*, Vol. 1256. SPIE, 182–197.

[84] Christopher W Tyler, Anas M Elsaid, Lora T Likova, Navdeep Gill, and Spero C Nicholas. 2012. Analysis of human vergence dynamics. *Journal of Vision* 12, 11 (2012), 21–21.

[85] Yuji Uema and Kazutaka Inoue. 2017. JINS MEME algorithm for estimation and tracking of concentration of users. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 297–300.

[86] Ming Wan, Nan Zhang, Huamin Qu, and Arie E Kaufman. 2004. Interactive stereoscopic rendering of volumetric environments. *IEEE Transactions on Visualization and Computer Graphics* 10, 1 (2004), 15–28.

[87] Hui-Shyong Yeo, Juyoung Lee, Woontack Woo, Hideki Koike, Aaron J Quigley, and Kai Kunze. 2021. JINSense: Repurposing Electrooculography Sensors on Smart Glass for Midair Gesture and Context Sensing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 457, 6 pages. doi:10.1145/3411763.3451741

[88] Laurence R. Young and David Sheena. 1975. Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation* 7, 5 (01 Sep 1975), 397–429. doi:10.3758/BF03201553

[89] Chenyang Zhang, Tiansu Chen, Eric Shaffer, and Elahe Soltanaghai. 2024. FocusFlow: 3D Gaze-Depth Interaction in Virtual Reality Leveraging Active Visual Depth Manipulation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–18.

A FEATURE SELECTION FOR VERGENCE CLASSIFICATION

Prior EOG-based vergence work has shown that compact time-domain features can effectively represent vergence-related EOG windows for real-time classification. For example, Mishra et al. [64] used handcrafted features such as signal integral, amplitude, velocity, and variance, further evaluating candidate features through wrapper-based selection. Motivated by prior work, we designed a compact feature representation tailored to the morphology of the two-channel vergence signals captured by our glasses.

As illustrated in Fig. 3c, vergence EOG segments typically contain a dominant extremum: transitions from near to far produce a valley in both EOG channels, whereas transitions from far to near produce a peak. We therefore split each detected vergence segment at its dominant extremum. This produces two temporal portions corresponding to the approach-to-extremum and recovery phases of the vergence response.

A.1 Proposed Compact Feature Representation

For each half-segment, we extract five temporal features: amplitude range, area under the curve, end-to-end slope, mean of the first derivative, and variance of the first derivative. These descriptors capture the magnitude of the response, the strength of the accumulated signal, the coarse temporal trend, the average local change and the local temporal variability. This yields 10 features per EOG channel.

A.2 Comprehensive Candidate Pool for Validation

To examine whether a richer representation would improve performance, we constructed a broader candidate pool of commonly used time-series descriptors for waveform shape and dynamics. Let $x = [x_1, x_2, \ldots, x_T]$ denote a single-channel waveform segment consisting of T samples, where μ and σ represent the mean and standard deviation of the segment, respectively. Let $p \in \{1, 2, \ldots, T\}$ be the index of the detected peak. The first- and second-order discrete derivatives of the signal are denoted as $\Delta x_i = x_i - x_{i-1}$ and $\Delta^2 x_i = \Delta x_i - \Delta x_{i-1}$.

Specifically, in addition to the 10 features we extracted above for each EOG channel, we extracted the following categories of features to form the complete candidate pool:

- Morphological Descriptors: Segment length T, and the normalized peak position calculated as $p/(T-1)$.
- Distributional and Amplitude Statistics: Mean μ, standard deviation σ, median, minimum $\min(x)$, maximum $\max(x)$, peak-to-peak amplitude $\max(x) - \min(x)$, skewness $\frac{1}{T}\sum_{i=1}^{T}(\frac{x_i-\mu}{\sigma})^3$, and kurtosis $\frac{1}{T}\sum_{i=1}^{T}(\frac{x_i-\mu}{\sigma})^4$.
- Energy Descriptors: Total signal energy $E = \sum_{i=1}^{T} x_i^2$, and the absolute area under the curve AUC $= \sum_{i=1}^{T} |x_i|$. We also compute the left-half energy $E_L = \sum_{i=1}^{p} x_i^2$ and right-half energy $E_R = \sum_{i=p}^{T} x_i^2$, which define the energy asymmetry as $\frac{E_R - E_L}{E_L + E_R}$.
- Derivative Descriptors: To capture local temporal dynamics, we calculate the gradient mean $\mu_{\Delta x}$, gradient standard deviation $\sigma_{\Delta x}$, mean absolute gradient $\frac{1}{T-1}\sum_{i=2}^{T}|\Delta x_i|$, and maximum absolute gradient $\max(|\Delta x|)$. Similarly, for the second-order derivative, we extract the mean absolute value $\frac{1}{T-2}\sum_{i=3}^{T}|\Delta^2 x_i|$ and maximum absolute value $\max(|\Delta^2 x|)$. Finally, we compute the gradient zero-crossing rate (ZCR) of Δx.

Together with the 10 compact descriptors, this comprehensive candidate pool contains 32 features per channel, resulting in a 64-dimensional feature vector for two-channel EOG fusion.

A.3 Feature Evaluation and Selection

To finalize our feature representation, we evaluated our proposed 20-dimensional compact set against both its single-channel variants and the comprehensive 64-dimensional candidate pool. The single-channel baselines extract the 10 compact features from only the left or right EOG channel. In contrast, the expanded representation utilizes the full 32-feature candidate pool per channel, totaling 64 features.

Table 6. Feature selection results for six-gesture vergence classification.

Feature Representation	Features per Channel	Total Dimensions	Accuracy
Single-channel (Left), Compact	10	10	69.03 ± 9.16%
Single-channel (Right), Compact	10	10	66.30 ± 9.43%
Two-channel fusion, Compact	10	20	74.95 ± 7.38%
Two-channel fusion, Full candidate pool	32	64	75.74 ± 7.17%

Our evaluation shows that two-channel fusion improves vergence classification. The compact two-channel representation achieved an accuracy of 74.95 ± 7.38%, outperforming the individual single-channel variants (left channel only: 69.03 ± 9.16%; right channel only: 66.30 ± 9.43%). Furthermore, expanding the feature space to the larger 64-dimensional candidate pool yielded only a modest performance gain, reaching 75.74 ± 7.17%. Consequently, we adopted the 20-dimensional compact representation for our main pipeline. This design choice preserves most of the performance of the richer candidate pool while minimizing feature dimensionality and computational overhead.

B ARCHITECTURE AND IMPLEMENTATION DETAILS OF BASELINE MODELS

The Random Forest baseline based on features uses the 20-dimensional feature vector described in Appendix A. The deep learning baselines use the same filtered two-channel EOG segments directly, where each segment is resampled to 160 time points, producing a 2×160 input tensor.

For within-session evaluation, models were evaluated using five-fold cross-validation within each user session, with folds grouped by collection round. For cross-session evaluation, models were trained in one session and tested in the other session after remounting of the glasses for the same user, with both transfers evaluated. For cross-user evaluation, each target user-session was held out for testing, and models were trained only on sessions from the remaining users. To avoid label leakage, hyperparameters were tuned using only the training data; the held-out test labels were used only for final evaluation.

The 1D-CNN baseline consists of two temporal convolution blocks followed by a classifier. The first block applies a 1D convolution to the two-channel input, followed by batch normalization, ReLU activation, max pooling, and dropout. The second block applies another 1D convolution followed by batch normalization, ReLU activation, and global average pooling. The representation is then passed to a fully connected layer for classification.

To capture longer-term temporal dependencies, the LSTM baseline treats the same 2×160 EOG segment as a temporal sequence with two input features per time step. The model uses one or two LSTM layers, optionally bidirectional, and routes the final time-step output to a linear six-class classifier.

Both deep learning baselines were trained using the AdamW optimizer and cross-entropy loss. For each evaluation split, we executed 50 Optuna Tree-structured Parzen Estimator (TPE) trials to automatically tune the hyperparameters. As shown in Table 7, the shared search space included learning rate, weight decay, batch size, and dropout rates. CNN-specific hyperparameters included the number of convolution channels and kernel sizes, while LSTM-specific hyperparameters included hidden size, number of recurrent layers, and bidirectionality.

C EVALUATION OF THE ARTIFACT-REMOVAL PIPELINE: PERFORMANCE AND CALIBRATION BURDEN

This section provides additional details for the motion artifact removal evaluation described in Sec. 6.4. We investigate the impact of varying training data volumes (vergence and artifact rounds) on system performance and assess the pipeline's cross-user generalizability.

Table 7. Hyperparameter search spaces for the 1D-CNN and LSTM baseline models. Each model was tuned using 50 Optuna TPE trials per evaluation split, using training data only.

Hyperparameter	Search space
Learning rate	Log-uniform from 10^{-4} to 3×10^{-3}.
Weight decay	Log-uniform from 10^{-6} to 10^{-2}.
Batch size	{16, 32, 64}.
CNN channels	First convolution: {8, 16, 32, 64}; second convolution: {16, 32, 64, 96}.
CNN kernel sizes	First convolution: {5, 7, 9, 11}; second convolution: {3, 5, 7}.
CNN dropout	Uniform from 0 to 0.5.
LSTM hidden size	{16, 32, 64, 96}.
LSTM layers	{1, 2}.
LSTM bidirectionality	{unidirectional, bidirectional}.
LSTM dropout	Uniform from 0 to 0.5; applied only when more than one recurrent layer is used.

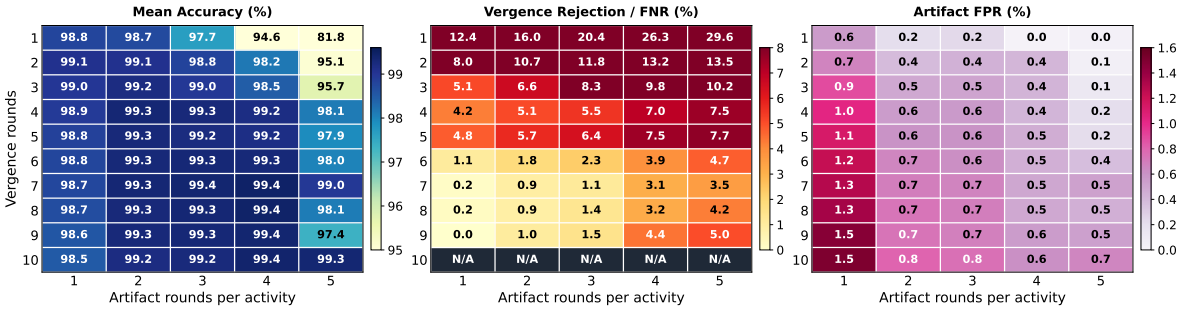


Fig. 21. **Training-data budget analysis for the personalized artifact-removal model.** The mean performance across all six participants for models trained from scratch using a specific number of vergence and artifact rounds.

C.1 Evaluation Protocol and Metrics

Consistent with the real-time scenarios, all evaluations employed a 2-second moving window. To preclude data leakage among highly correlated adjacent windows, we enforced training-test splits at the recording or round level. System performance is quantified using three metrics: overall accuracy, vergence false rejection rate (the proportion of true vergence windows classified as artifacts), and artifact false acceptance rate (the proportion of artifact/noise windows incorrectly identified as vergence).

We investigated two settings. The personalized setting trained the model from scratch exclusively on the target participant's data. The cross-user adaptation setting pre-trained the model on data from all other participants before fine-tuning it with varying amounts of target-user data. Together, these settings demonstrate the registration burden for a personalized model and the feasibility of adapting a generalized model to new users.

C.2 Personalized Artifact-Removal Performance

Fig. 21 illustrates the personalized data-budget analysis. Each cell in the matrix represents a specific training-data budget, defined by the number of vergence rounds (horizontal axis) and artifact rounds (vertical axis). For each configuration, we report the accuracy, false negative rate, and false positive rate.

Our results reveal that increasing the volume of vergence training data primarily mitigates the false rejection of intended inputs. With a fixed budget of one artifact round, scaling the vergence training data from one to six rounds reduced the vergence false rejection rate from 12.36% to 1.08%. Concurrently, the artifact false acceptance

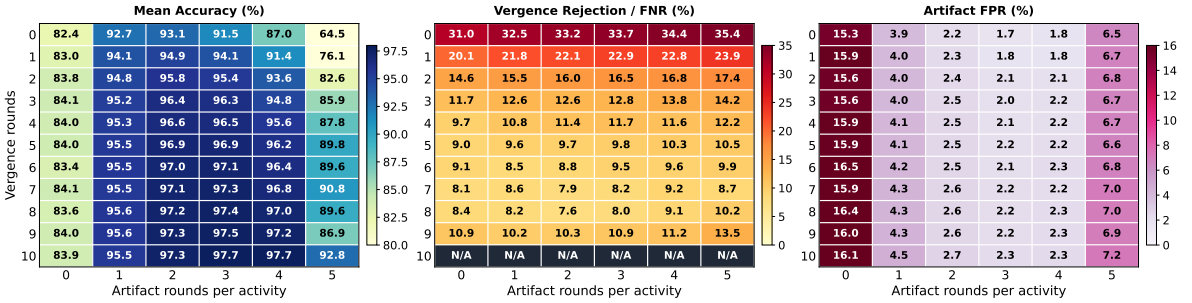


Fig. 22. **Performance of the cross-user artifact-removal model with target-user adaptation.** The model is pre-trained on data from other participants and subsequently adapted using varying amounts of target-user data. Each cell reports the mean performance across all six participants for cross-user models adapted with a specific number of target-user vergence and artifact rounds.

rate remained stable and low, shifting only from 0.56% to 1.23%. This demonstrates that supplementary vergence examples enable the model to better preserve vergence inputs without compromising its ability to reject noise.

Guided by this analysis, we adopted six vergence rounds and one artifact round (V6/A1) as the primary operating point for training our main evaluation model. This configuration strikes a highly practical balance between system robustness and user burden, translating to the concise 6-minute total registration time for new users: 3 minutes for six rounds of vergence data, and an additional 3 minutes to collect one artifact calibration round for each of the six motion/noise activities. It yielded an overall accuracy of 98.77 ± 1.26%, a false rejection rate of vergence of 1.08 ± 1.55% and a false positive rate of 1.23 ± 1.26%. While incorporating additional artifact rounds into the training set could further suppress false acceptances, it would linearly increase the data collection burden. Thus, the V6/A1 setting offers an optimal data collection protocol while maintaining minimal error rates.

C.3 Cross-User Generalization and Target-User Adaptation

Fig. 22 details the performance of the cross-user adaptation setting. Here, the model was pre-trained on the remaining participants and subsequently adapted using variable amounts of data from the target user. In a zero-shot scenario, the model achieved only 82.41% accuracy, severely affected by a 31.02% vergence false rejection rate and a 15.33% artifact false positive rate. This performance drop highlights the inter-user variance in motion and artifact patterns, likely stemming from individual differences in sensor contact, facial anatomy, and motion.

Introducing target-user vergence data alone mitigated false rejections but failed to adequately suppress artifact acceptances. For example, incorporating six vergence rounds for the target-user reduced the false rejection rate of vergence to 9.10%, yet the false positive rate remained 16.51%. This indicates that while vergence help anchor the intended signal, they lack the negative constraints necessary to accurately model user-specific noise profiles.

In contrast, injecting even a single artifact round from the target participant enhanced noise rejection. By combining six vergence rounds with one artifact round from the target user, the adapted cross-user model achieved 95.55% accuracy, with an 8.53% vergence false rejection rate and a 4.16% artifact false positive rate. This result shows the importance of target-user artifact data for noise suppression.

In summary, these evaluations justify the adoption of the personalized V6/A1 setting, where the model was trained from scratch using six rounds of vergence data and one round of artifact data per activity. Fully generalized cross-user models prove insufficiently reliable for real-world deployment, whereas a brief, 6-minute personalized data collection session to train the model from scratch yields reliable accuracy and minimal error across both metrics. The data-budget ablation confirms the V6/A1 configuration as a practical operating point that maximizes algorithmic reliability while respecting strict real-world usability constraints.