electronics

MDPI

Article

Eyeglass-Type Switch: A Wearable Eye-Movement and Blink Switch for ALS Nurse Call

Ryuto Tamai [1], Takeshi Saitoh [1,*], Kazuyuki Itoh [2] and Haibo Zhang [1]

[1] Faculty of Computer Science and Systems Engineering, Department of Artificial Intelligence, Kyushu Institute of Technology, Fukuoka 820-8502, Japan; haiboz@ai.kyutech.ac.jp (H.Z.)
[2] Research Institute, National Rehabilitation Center for Persons with Disabilities, Saitama 359-8555, Japan; itoh-kazuyuki-0923@rehab.go.jp
* Correspondence: saitoh@ai.kyutech.ac.jp

Abstract

We present the eyeglass-type switch, an eyeglass-mounted eye/blink switch designed for nurse-call operation by people with severe motor impairments, with a particular focus on amyotrophic lateral sclerosis (ALS). The system targets real-world bedside constraints—low illumination at night, supine posture, and network-independent operation—by combining near-infrared (NIR) LED illumination with an NIR eye camera and executing all processing on a small, GPU-free computer. A two-stage convolutional pipeline estimates eight perioc-ular landmarks and the pupil center; eye-closure is detected either by a binary classifier or by an angle criterion derived from landmarks, which also skips pupil estimation during closure. User intent is determined by crossing a caregiver-tunable "off-area" around neutral gaze, implemented as rectangular or sector shapes. Four output modes—single, continuous, long-press, and hold-to-activate—are supported for both oculomotor and eyelid inputs. Safety is addressed via relay-based electrical isolation from the nurse-call circuit and audio feedback for state indication. The prototype runs at 18 fps on commodity hardware. In feature-point evaluation, mean errors were 2.84 pixels for landmarks and 1.33 pixels for the pupil center. In a bedside task with 12 healthy participants, the system achieved $F = 0.965$ in single mode and $F = 0.983$ in hold-to-activate mode; blink-only input yielded $F = 0.993$. Performance was uniformly high for right/left/up and eye-closure cues, with lower recall for downward gaze due to eyelid occlusion, suggesting camera placement or threshold tuning as remedies. The results indicate that the proposed switch provides reliable, low-burden nurse-call control under nighttime conditions and offers a practical input option for emergency alerts and augmentative and alternative communication (AAC) workflows.

Keywords: ALS; assistive technology; eye-tracking; blink detection; nurse-call interface; pupil-center localization; two-stage CNN

check for updates

Academic Editors: Bipin Indurkhya, Eleuda Nuñez and Marie-Monique Schaper

Received: 17 September 2025
Revised: 18 October 2025
Accepted: 22 October 2025
Published: 27 October 2025

Citation: Tamai, R.; Saitoh, T.; Itoh, K.; Zhang, H. Eyeglass-Type Switch: A Wearable Eye-Movement and Blink Switch for ALS Nurse Call. *Electronics* **2025**, *14*, 4201. https://doi.org/10.3390/electronics14214201

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In severe motor impairments such as amyotrophic lateral sclerosis (ALS), multiple system atrophy (MSA), and muscular dystrophy, voluntary limb movements—the primary means for intentional communication—are restricted, while oculomotor function is often relatively preserved. In medical and caregiving settings, establishing safe and reliable meth-ods for nighttime nurse calls is an urgent challenge, and there is a demand for input–output interfaces that impose minimal burden and are robust to false activations even under low-light conditions. In our previous work, we reported a nighttime calling system that assumes

nighttime use and achieves stable eye-image capture with near-infrared (NIR) LEDs and an NIR camera, mitigates head-motion effects with an eyeglass-mounted wearable camera, and realizes bedside, network-independent operation using a small computer [1].

Recently, various wearable sensing technologies—such as photonic smart wristbands for cardiorespiratory monitoring and biometric identification—have demonstrated the potential of compact and noninvasive devices for continuous physiological measurement. These developments illustrate the growing importance of wearable systems that integrate sensing and interaction for healthcare and assistive applications. The present work extends this trend to eye-based interfaces, focusing on communication support for users with severe motor impairments.

That system employs a two-stage convolutional neural network (CNN) to (i) classify eye open/closed states and (ii) estimate the pupil center, and it outputs a call signal when the pupil position exceeds a predefined threshold. In experiments, the mean error of pupil-center estimation was 1.17 pixels, and tests with five healthy participants yielded a precision of 0.83, recall of 1.00, and F-score of 0.91, demonstrating feasibility. Subsequently, however, evaluations conducted with patient collaborators and failure cases caused by blinks revealed remaining issues that must be addressed for clinical deployment.

Building on this prior work [1], the present paper introduces a comprehensive framework that newly integrates system design, algorithms, and interaction schemes tailored to real clinical usage and requirements. Our specific contributions are as follows:

- Integrated eyeglass-type design for low-light operation: A near-infrared (NIR) imaging system and lightweight CNN-based estimator are embedded in an eyeglass-type form factor to achieve robust pupil and eyelid tracking even in dark, bedside environments.
- Configurable caregiver-centered interaction framework: We introduce a switch operation model that allows caregivers to flexibly adjust geometric and temporal thresholds (e.g., off-area shape, duration parameters) to match individual user conditions, thereby improving safety and personalization.
- Comprehensive validation across real-world settings: The system's performance was verified through controlled experiments with healthy participants and actual use by multiple ALS patients, as well as public demonstrations including the Osaka–Kansai Expo, confirming its robustness and usability.

These contributions collectively distinguish the proposed system from prior camera-based or EOG-based interfaces by offering a practical, low-burden, and tunable eye-based switch suitable for clinical and home environments.

The remainder of this paper is organized as follows: Section 2 reviews related work; Section 3 describes the system architecture; Section 4 details the algorithms and interaction design; Section 5 presents the evaluation; and Section 6 concludes the paper.

2. Related Research and Products

The nighttime nurse-call interface based on oculomotor and eyelid movements addressed in this study belongs to the lineage of augmentative and alternative communication, that is, AAC, for people with severe motor impairments. The related landscape can be grouped into four areas: first, alternative input interfaces; second, algorithms for detecting gaze, pupil position, and blinks; third, wearable infrared tracking devices that operate under low illumination; and fourth, integration with nurse-call systems and safety design. An overview follows.

2.1. Alternative Input Interfaces

AAC systems that use eye gaze are widely available as both commercial products and open-source software. For example, Tobii Dynavox software allows per-button dwell-

time adjustment to suppress accidental selections while maintaining selection speed [2]. OptiKey runs even with low-cost eye trackers or standard webcams and natively supports dwell selection together with external switches [3]. Research has also reported on reducing workload and erroneous selections by dynamically optimizing dwell time and predicting user intent [4]. Complementing interface-side optimizations, large language models (LLMs) can accelerate eye-gaze text entry by providing context-aware prediction and completion; recent work with users who have ALS shows reduced selection effort and increased communication rate [5].

Communication devices in practical use include Den-no-Shin [6], OriHime eye+Switch [7], CYIN (welfare model) [8], and Fine-Chat [9]. Den-no-Shin integrates text composition, speech output, and external device control around a single-switch scanning user interface. OriHime eye+Switch combines eye-gaze input with switch input to enable text entry and speech output. CYIN detects weak voluntary biosignals and converts them into switch inputs with a multichannel sensor configuration. Fine-Chat is designed for scanning selection and word prediction with minimal movement and supports multiple input methods.

Depending on the condition, users may employ push-button switches or switches based on piezoelectric sensors; however, as the disease progresses, movements become extremely small, and operating these switches themselves often becomes difficult. Approaches that map blinks and eye movements directly to word selection are advancing; for instance, a method that uses only a smartphone camera and facial landmark detection has been proposed to select fixed phrases from combinations of left, right, up, and blink [10]. Binary input via inhalation and exhalation, known as the sip-and-puff method, has also been used for many years.

Electrooculography (EOG) is a camera-free, compact, and low-power alternative that operates in darkness and even with the eyes closed. Recent graphene-textile implementations showed close agreement with clinical electrodes and high detection rates for interaction tasks [11,12]. Compared to our NIR camera-based eyeglass-type switch, EOG reduces optical sensitivity but introduces electrode placement, skin impedance drift, and hygiene/comfort considerations.

2.2. Algorithms for Detecting Gaze, Pupil, and Blinks

Gaze estimation and pupil-center detection using only video have been extensively explored, ranging from remote setups to near-field eyeglass-mounted configurations. Remote eye tracking based on the simultaneous detection of high-speed stroboscopic reflections and the pupil center has been reported at the 500 Hz level, providing high temporal resolution [13]. WebGazer, which operates in a browser environment, estimates gaze position with standard webcams by leveraging self-calibration and interaction logs, thereby broadening experimental and application possibilities [14].

Advances in deep learning have led to higher accuracy in pupil-center detection. Using images acquired from commercial eye-tracking devices, convolutional neural networks have been shown to estimate the pupil center and, thereby, boost accuracy even with off-the-shelf hardware [15]. Furthermore, in visible-light wearable settings, a deep-learning approach that integrates pupil-center detection and tracking has been proposed, together with design guidelines robust to headset misalignment and illumination changes [16]. Chinsatit and Saitoh also proposed a two-stage CNN that detects the pupil center with high accuracy [17]. Taken together, these studies complement our choice of near-infrared imaging and substantiate the validity of the two-stage estimation pipeline and region of interest (ROI) design adopted in this work.

For blink detection, recent surveys of deep learning have organized a wide range of architectures, including CNNs and time-series models [18]. In addition, ultra-low-power

wearable devices that treat blinks as events using capacitive or electrostatic sensors have been reported, attracting attention as camera-independent input solutions [19].

2.3. Wearable Infrared Tracking for Low-Light and Nighttime Use

To obtain stable imagery at night or in dim hospital rooms, combining near-infrared illumination with an eye camera is effective. As an eyeglass-type research platform, Tobii Pro Glasses 3 (Tobii AB, Stockholm, Sweden) integrates a near-eye camera, a scene camera, and an inertial measurement unit, and has been reported to deliver high accuracy in everyday contexts, including walking [20]. Pupil Labs Invisible (Pupil Labs, Berlin, Germany) and Neon embed NIR eye cameras (Pupil Labs, Berlin, Germany) and IMUs in the frame and provide high-frequency eye data and gaze points. A recent review has surveyed near-infrared, near-eye tracking for health applications, noting the expanding clinical use of blink and eyelid-movement measurements [21].

2.4. Integration with Nurse-Call Systems and Safety Design

When integrating with nurse-call systems, safety design is essential, including electrical isolation via relays, limits on maximum continuous output time, and the insertion of a dead time to prevent repeated activations. To reduce false triggers, it is effective to design spatial thresholds using both rectangular and sector-shaped regions, to incorporate temporal conditions such as dwell, long-press, and hold-to-activate, and to introduce hysteresis and debouncing.

In our prior work, we presented a calling method based on pupil-center estimation and reported, under an average illuminance of 9.9 lux, a precision of 0.83, a recall of 1.00, and an F-score of 0.91 for the calling task, together with a mean pupil-center error of 1.17 pixels [1].

2.5. Positioning of This Work

Taken together, progress in this field can be organized around three pillars that have advanced in parallel: the diversification of input methods, the maturation of wearable sensing suited to low-light environments, and the safe electrical integration with medical circuits. Building on this trajectory, the present study proposes a practice-oriented framework specifically designed for nurse-call operations under the constraints of low illumination, supine posture, and network independence. The approach utilizes eyeglass-based infrared imaging and a two-stage estimation pipeline—landmark regression combined with pupil-center regression and a closed-eye branch—employing an interaction design that integrates spatial and temporal thresholds, treats blink-based input on an equal footing, and provides visualization and parameter settings that enable caregivers to rapidly retune the system. These aspects constitute the main contributions of this work.

3. System Architecture

3.1. Design Rationale and Hardware Configuration

The goal of the system is to enable reliable nurse-call operation using only ocular and eyelid movements under three constraints: nighttime use, supine posture, and operation without network dependence. Because imaging is performed in the dark, no visible illumination is used; instead, stable eye images are acquired with near-infrared (IR) LEDs and an IR camera. To reduce the influence of head motion, we adopt an eyeglass-mounted wearable camera. All processing is completed on a small bedside computer in a standalone configuration. For safety and false-activation suppression, the nurse-call circuit is electrically isolated via a relay controller.

Figure 1 shows the system overview. The components are an eyeglass-mounted IR camera, a small bedside computer, a relay controller, a numeric keypad for caregiver opera-

tion, and, when needed, a bell-splitting device. For nighttime use, imaging is performed with IR illumination and an IR camera, and the eyeglass form factor maintains a stable relative position between the camera and the eye. Additionally, we utilize lightweight CNN models that operate in real-time on commodity PCs, eliminating the need for a GPU and, thereby, minimizing deployment and maintenance costs.

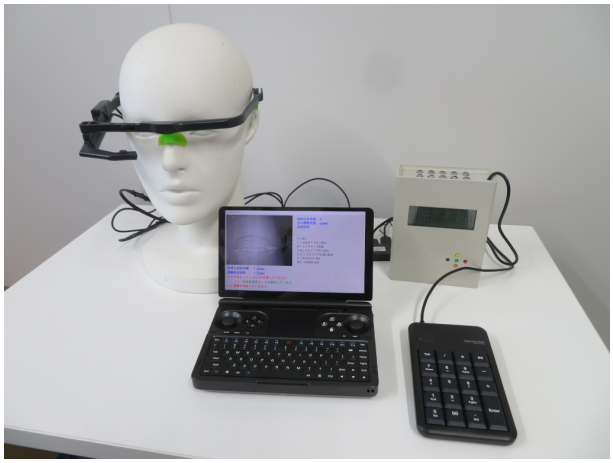


Figure 1. System overview of the eyeglass-type switch. From left to right: wearable near-infrared eyeglass camera; compact notebook PC running the software (main window shown); relay controller providing electrically isolated output to the nurse-call circuit (**upper right**); external numeric keypad for caregiver operation (**lower right**).

The eyeglass-type switch supplies its own NIR LED illumination and runs the eye camera with fixed exposure/gain, thereby reducing dependence on ambient visible light. In line with this design, our prior study [1] confirmed correct pupil-center detection at 3.92 lx, indicating robustness to low-illuminance conditions. In the present evaluation (dark room with room lights off), reproducibility was primarily limited by eyelid occlusion and initial camera placement rather than illumination changes. Future experiments will quantify performance across controlled illuminance levels and spectra, and assess pose-aware sector thresholds and exposure auto-lock windows to stabilize measurements further.

3.2. Software Architecture

From each camera frame, the system detects the pupil-center coordinates. When the detected position exceeds a predefined threshold, the computer outputs a signal to the relay controller using either a single-shot mode or a continuous mode.

3.3. Eye-Image Processing Pipeline

1. ROI extraction: To suppress background intrusion specific to wearable imaging, the caregiver manually extracts the periocular ROI.
2. Two-stage CNN estimation: We consider two alternatives for the estimation pipeline (Figure 2). Our default pipeline estimates eight eyelid/eye contour landmarks with CNN1 and the pupil center with CNN2. Eye closure is gated either by a binary open/closed classifier (CNN3) or by a threshold on the angle θ_{eye} computed from the CNN1 landmarks.
3. User interface (UI) and visualization: Figure 3 shows the main window of the eyeglass-type switch. The upper-left pane overlays the IR eye image with estimated landmarks and the pupil center, allowing the caregiver to verify tracking quality in real time. The upper-right area presents system status (current user, binary eye state, and transmission flag). Below it, a key-hint menu lists caregiver operations such as mode switching, placement/toggling of the off-area map, ROI setting, and user

switching. Along the bottom of the window, time thresholds (e.g., long-press and closure times) and runtime messages are displayed for rapid tuning (t_{long}, t_{keep}, t_{close}). All on-screen text is presented in Japanese to suit the intended patient/caregiver population; localization to other languages is straightforward. This layout prioritizes caregiver-centered parameter adjustment during setup and bedside operation.

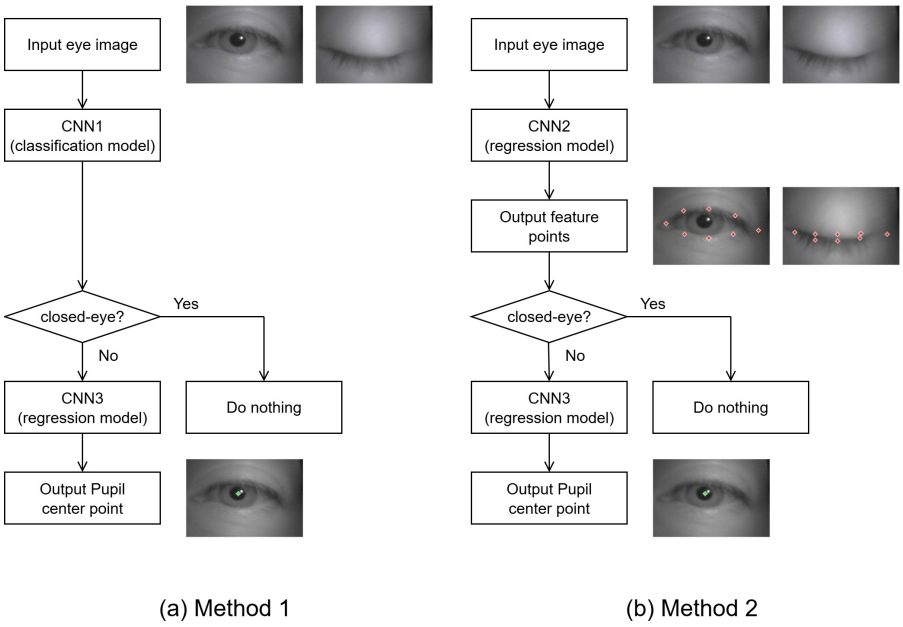


Figure 2. Estimation pipeline alternatives for pupil-center detection and closed-eye gating. (**a**) Method 1: classification-first pipeline—CNN1 performs open/closed classification; if open, CNN3 regresses the pupil center. (**b**) Method 2: landmark-first pipeline—CNN2 regresses periocular feature points; eye closure is judged from the landmark-based angle θ_{eye}; if open, CNN3 regresses the pupil center. Example frames (open/closed) and landmark overlays are shown.

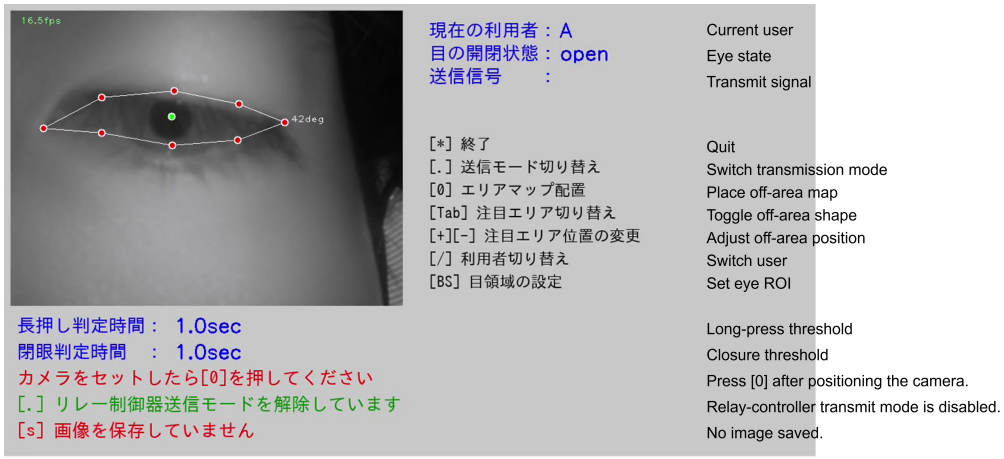


Figure 3. Main window of the eyeglass-type switch. (**Upper-left**): IR eye image with overlay of eight contour landmarks and the pupil center. (**Upper-right**): system status (user, open/closed eye state, transmit flag). Middle-right: key-hint menu for caregiver operations. (**Bottom**): time thresholds and runtime messages. On-screen text is shown in Japanese for the target patient/caregiver use. In the deployed system, all on-screen text is presented in Japanese; English labels are used here for publication purposes.

3.4. Thresholds, Output Modes, and Operation

During initial setup, the caregiver configures the threshold region (off-area) and the closed-eye decision angle θ_{eye} for the individual user, with subsequent fine-tuning as

needed. The off-area can be either rectangular or sector-shaped to provide an operating space robust to camera tilt and inter-individual variability. Its angle, radius, and size are manually adjustable, and the settings persist until reconfigured. When the pupil position crosses the off-area, the switch state is set to ON and a signal is sent from the computer to the relay controller. For eyelid-based input, an analogous decision is made using a threshold on the eye-closure duration.

The system supports four output modes—single, continuous, long-press, and hold-to-activate—applicable to both oculomotor and eyelid inputs. By selecting modes based on symptoms and operational characteristics, users can strike a balance between suppressing false activations and maintaining responsiveness.

When users cannot view the screen, the system provides state feedback via ON/OFF sound effects and audio messages for mode changes. The operation centers on a numeric keypad, reducing the caregiver's PC interaction burden.

With this configuration, the system provides a reproducible nurse-call interface under low-light, supine, and network-independent conditions. The accuracy of estimation, the behavior of the output modes, and the safety mechanisms are examined in the following sections.

4. Algorithms and Interaction Design

This section describes feature estimation from eye images—pupil center, contour landmarks, and open/closed state—and the decision and temporal-control design that leads to switch output, including the output modes.

4.1. Input and ROI Extraction

The system input is an eye image captured by a wearable IR camera. Because the camera is positioned close to the eye, the eye's location and scale within the image vary with individual facial morphology and eye size. Although the camera arm can be adjusted to position the camera appropriately, fine adjustments can still be challenging in some cases, which may compromise the CNN's estimation accuracy. While automatic determination of an appropriate ROI from the captured image would be desirable, we adopted a manual ROI-setting policy based on caregiver feedback. Figure 4 illustrates examples; the rectangle denotes the ROI. In Figure 4 (left), an improperly set ROI leads to poor detection of landmarks near the outer canthus, whereas in Figure 4 (right), proper ROI adjustment enables correct landmark detection.

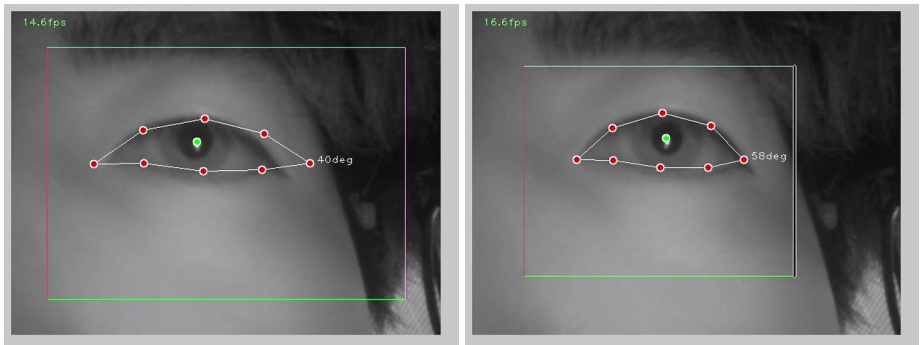


Figure 4. Effect of ROI placement on landmark detection. (**left**) A suboptimal ROI leads to missed landmarks near the outer canthus. (**right**) A corrected ROI enables accurate contour landmark detection.

4.2. Two-Stage CNN for Feature Estimation

A two-stage CNN estimates (i) eye contour landmarks and (ii) the pupil center. The first-stage network (CNN1) regresses the landmarks; the second-stage network (CNN3)

regresses the pupil-center coordinates. As an implementation example, CNN1 uses VGG16 with dropout, and CNN3 uses Xception, yielding eight contour landmarks and the pupil center P from a single eye image.

Two alternatives are supported for eye-closure detection. In the classification approach (CNN2), the eye image is classified as open or closed. In the geometric approach, we compute the canthus angle θ_{eye} from the CNN1 landmarks and declare a closed eye when $\theta_{eye} < T_\theta$, where T_θ is the threshold angle for closure. The geometric approach is robust to individual differences and mounting offsets; when the eye is judged closed, the pupil-center estimation step is skipped.

4.3. Design of the Threshold Region (Off-Area)

Let P_0 denote the pupil center during neutral fixation. Around P_0, we place an off-area as shown in Figure 5. When the current pupil position P crosses this boundary, we interpret it as user intent and transition to a switch-ON decision. Two shapes are implemented: the rectangular regions for the four directions (Figure 5, left) and a sector defined by center angle and radius (Figure 5, right). To accommodate camera tilt and inter-individual variability, a caregiver can manually adjust angle, size, and radius. Rectangles alone are sensitive to mounting angle, whereas combining them with sectors yields an operating space that is robust to device pose.

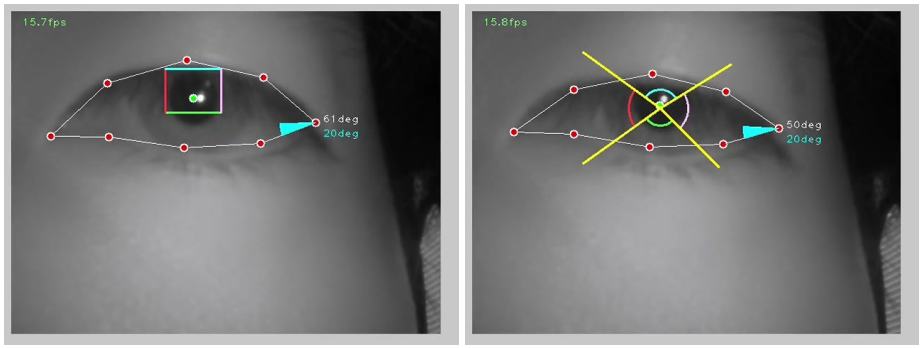


Figure 5. Threshold region ("off-area") configurations around the neutral pupil position P_0. (**left**) Rectangular regions for the four directions (up, down, left, right). (**right**) A sector-shaped region defined by center angle and radius. Angle, size, and radius are caregiver-tunable to accommodate camera tilt and inter-individual variability; combining rectangles with sectors yields an operating space robust to device pose.

4.4. Decision Logic for Switch Output

In general, switch behavior falls into two types: (1) momentary, which turns ON only while the actuator is engaged and returns to OFF when released; and (2) alternate (toggle), which flips between ON and OFF each time it is actuated. The proposed system adopts the momentary type: The ON state corresponds to an asserted serial signal from the computer to the relay controller, and the OFF state corresponds to no signal. The system provides multiple transmission modes for signal output; details are described below. By default, we use a normally open (NO) configuration—when the user is not moving the eyes or eyelids, the switch remains OFF, and it turns ON only when an intentional movement is detected.

4.4.1. Oculomotor-Based Decision

When P crosses the off-area, one of four output modes is invoked and selected according to symptoms and operational characteristics:

- Single: Transmit exactly once at the moment of crossing; further transmissions are inhibited until the gaze returns.

- Continuous: Transmit periodically for as long as the gaze remains beyond the boundary; holding gaze alone suffices for operation.
- Long-press: Transmit when the dwell beyond the boundary exceeds t_{long}, then transmit once more upon return. Combined with relay latching, this realizes long-press behavior. t_{long} is adjustable from 1.5 s to 3.5 s in 0.5 s steps.
- Hold-to-activate: Transmit exactly once when the boundary crossing persists longer than t_{keep}; no additional transmission occurs on return. t_{keep} is adjustable from 1.5 s to 3.5 s in 0.5 s steps. This mode is provided to prevent false activations that can arise in the single-shot mode due to pupil-center jitter or incidental, unintended glances.

Figure 6 presents the time charts for all modes. The caregiver sets $t^* \in \{t_{long}, t_{keep}\}$ as appropriate.

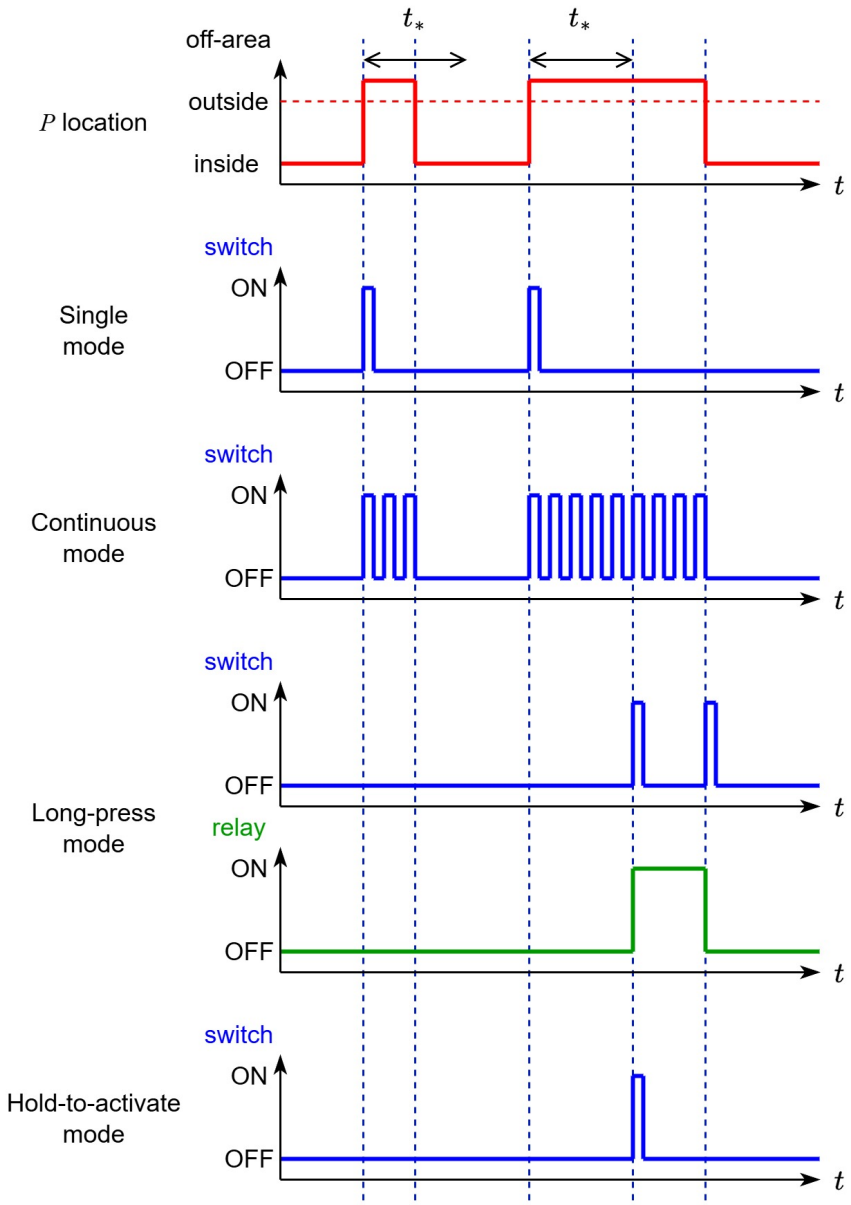


Figure 6. Time charts for oculomotor-based signal transmission. Four output modes are illustrated: Single (one shot at boundary crossing), Continuous (periodic transmission while beyond the boundary), Long-press (transmit after dwell $> t_{long}$ and once more on return), and Hold-to-activate (transmit once after dwell $> t_{keep}$, no transmission on return). The caregiver selects $t^* \in \{t_{long}, t_{keep}\}$ as appropriate.

4.4.2. Eyelid-Based Decision (Blink Input)

For users who struggle with oculomotor control, we also implement input based on the eye-closure duration t_{close}. Using the open/closed estimator, a signal is transmitted when the closure exceeds t_{close}. The same four output modes are supported, and t_{close} is caregiver-adjustable. Figure 7 shows the time charts; the design mirrors the oculomotor case.

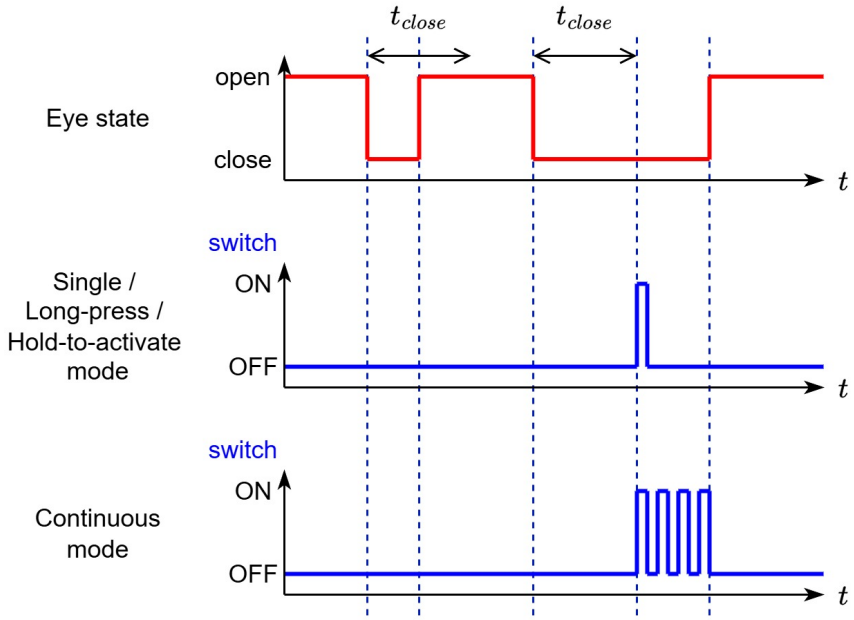


Figure 7. Time charts for blink-based (eyelid) signal transmission. Eye closure is detected by the open/closed estimator, and a signal is issued when the closure duration exceeds the threshold $> t_{close}$. The same four output modes as the oculomotor case are illustrated: Single (one shot at threshold crossing), Continuous (periodic transmission while the eye remains closed beyond the threshold), Long-press (transmit after closure $> t_{long}$ and once more on reopening), and Hold-to-activate (transmit once after closure $> t_{keep}$ with no transmission on reopening). Caregivers set t_{close} and select $t^* \in \{t_{long}, t_{keep}\}$ as appropriate.

4.5. Temporal Stabilization and Operational Feedback

Switch ON/OFF commands are sent from the computer to the relay controller, which drives the dry-contact inputs of the nurse-call system or external devices. This maintains a safely isolated control path while preventing unintended continuous outputs.

Caregivers adjust parameters—off-area shape, angle, radius, and the timings t_{long}, t_{keep}, t_{close}—via a numeric keypad. For the user, system state is indicated by ON/OFF tones and audio messages for mode changes, enabling operation even when the screen is not visible.

4.6. Handling Alignment Error and Individual Differences

With eyeglass-mounted imaging, the eye's position and scale in the frame vary across users and wearing positions. Robust operation is achieved by combining (i) contour regression with angle-based closure decisions, (ii) a configurable off-area, and (iii) manually adjustable parameters, which together accommodate mounting offsets and differences in natural eyelid opening.

Taken together, these algorithms and interaction designs enable switch output that balances clarity of intent interpretation, through the off-area and four output modes; robust detection, through contour regression and angle-based logic; and operational adaptability, through caregiver tuning; even under the constraints of low illumination, supine posture, and network independence.

5. Experiments

5.1. Feature-Point Detection Experiment

Prior to the system-level evaluation, we assessed the detection accuracy of 17 feature points consisting of one pupil center and 16 periocular contour landmarks. We constructed a bespoke dataset using two types of wearable cameras, denoted D1 and D2. D1 comprised 400 images each from four facility staff and three individuals with ALS (2800 images in total). D2 comprised 750 images each from six healthy participants (4500 images in total). All images were normalized to 120×80 pixels and labeled for eye state (open/close). Training employed data augmentation—scaling, translation, rotation, and luminance adjustment—and fine-tuning (FT) in which models were first trained on D2 and then refined on D1. We compared VGG16, VGG16 with dropout (VGG16-D), VGG16 with batch normalization (VGG16-BN), and Xception.

All models were implemented using the PyTorch [22] deep-learning framework. Training was performed for 100 epochs with a batch size of 50 and a learning rate of 0.001, using the Adam optimizer and mean squared error (MSE) loss function for both contour landmark and pupil-center regression. These parameters were chosen to balance convergence stability and generalization across different participants and imaging conditions.

For quantitative evaluation, ground-truth coordinates were manually annotated for the eye state and for all 17 feature points. The eye state was labeled open when approximately $\geq$50% of the pupil region was visible and closed otherwise.

The experiment covered 48 conditions formed by the Cartesian product of four CNN models, three training regimes (D1 only; D1 plus D2; fine-tuning with D2→D1), and four landmark-set sizes. We employed a leave-one-subject-out protocol, in which one patient served as the test subject, while the remaining two patients, along with staff and healthy participants, formed the training set. The Euclidean error between detected and ground-truth feature points measured performance. Table 1 summarizes the results; for brevity, only the five best-performing conditions are listed. The best condition used Xception with simultaneous detection of eight contour points plus the pupil center (nine points in total) under the fine-tuning regime, yielding a mean contour-point error of 2.84 pixels.

Table 1. Landmark–pupil estimation accuracy across model and training configurations. We evaluated 48 conditions combining CNN backbones, training regimes (D1 only; D1+D2; fine-tuning D2→D1), and landmark set sizes. The table lists the five best conditions. Errors are mean Euclidean distances in pixels.

CNN Model	Number of Points	Training Method	Error [Pixel]
Xception	9	FT	2.84
Xception	17	FT	2.98
Xception	8	FT	3.44
VGG16	9	FT	5.04
VGG16-D	8	FT	5.05

For eye open/close decisions derived from angles computed using the detected landmarks, the accuracies were 91.8% for open, 58.6% for close, and 83.8% overall. Typical error sources included images with lower-field contour occlusion and motion blur caused by blinks.

Pupil-center detection (CNN3) was evaluated under 36 conditions. Table 2 reports the results, again listing only the top five conditions. The best performance was obtained with a VGG16-BN single-point model trained on D1+D2, achieving a mean error of 1.33 pixels ($\approx$0.58 mm). A prior study using a ConvNet and 17,820 images reported 1.17 pixels; the small gap is plausibly due to our smaller dataset size, and the 0.17-pixel difference is expected to have a negligible impact in practice.

Finally, replacing the open/close criterion from "visibility of the pupil region" to "angle-based information derived from landmarks" produced a slight decrease in accuracy but improved the explainability of failure modes.

Our system minimizes ambient-light dependence through self-supplied NIR and fixed exposure/gain; residual errors were primarily due to eyelid occlusion during downward gaze and the initial camera pose. EOG offers different trade-offs—no camera and small form factor, but sensitivity to electrode placement and long-term skin dynamics [11,12]. We will focus on improving downward-gaze robustness in the current pipeline while recognizing EOG as a viable complementary option.

Table 2. Pupil-center detection accuracy under 36 conditions. The table reports the five best configurations; errors are mean Euclidean distances in pixels.

CNN Model	Number of Points	Training Method	Error [Pixel]
VGG16-BN	1	D1+D2	1.33
VGG16-D	1	D1	1.33
VGG16	1	FT	1.41
Xception	1	D1+D2	1.44
Xception	1	FT	1.46

In addition to landmark detection, we compared two methods for eye-state (open/close) classification: CNN1 (classification model) and CNN2 (the geometric angle–based estimation). In our previous study [1], the CNN1 classifier achieved an accuracy of 82.1%. In the current system, the angle-based method (CNN2) achieved an accuracy of 83.8%, demonstrating comparable or slightly improved performance, with no statistically significant difference. Although the CNN1-based method offers higher flexibility for unseen conditions, the geometric approach provides a lower computational cost and better real-time stability. Therefore, the system allows caregivers to select either CNN1 or CNN2 depending on the user's condition and environment, enabling flexible adaptation to diverse patient needs.

5.2. Switch Operation Experiment

The developed system used the following components: a wearable IR device (GPE3, Gazo Co., Ltd., Niigata, Japan), a compact general-purpose computer (GPD WIN Mini 2024, TENKU Co., Ltd., Tokyo, Japan), and a relay controller (UG-R10, Uchiyama Technical Laboratory, Fukushima, Japan). The main program was implemented in C++, and CNN training was conducted with the Keras neural-network library. The average processing speed was approximately 18 fps, confirming that the operation is practical even without a GPU.

The experiment involved 12 healthy student participants lying supine on a bed in a dark room with the room lights switched off. A monitor was placed in front of each participant to display visual output, and participants were instructed to keep their gaze on the monitor and not to move their eyes except when prompted. Five auditory cues were used—"up," "down," "left," "right," and "close [eyes]"—with both timing and content randomized. Two student experimenters served as caregivers for the system. Before each session, the operators configured, for each participant, the off-area shape and the threshold angle T_θ for eye-closure judgment. During the experiment, we set $t_{keep} = t_{close} = 1.0$ s.

After the session began, participants performed the actions specified by the auditory cues. In each set, every cue was presented twice (eight prompts total), and one set lasted about 4 min. Each participant—per caregiver operator—completed two sets, one in single-shot mode and one in hold-to-activate mode; the cumulative number of participant-sessions was 22. Because the oculomotor action required to operate the switch is identical in the

Continuous and Long-press modes, only these two transmission modes were tested. To simulate nighttime use, all sessions were conducted in a dark room with the room lights off under near-infrared (NIR) illumination. Ambient illuminance was not remeasured in this study; the room, NIR lighting, and fixed exposure/gain camera settings matched those of our prior study [1], which reported 3.92∼16.7 lx.

Let TP be the number of correct responses to auditory cues, FN the number of misses, and FP the number of false activations in the absence of a cue. We computed recall $R = TP/(TP + FN)$, precision $P = TP/(TP + FP)$, and F-score $F = 2PR/(P + R)$.

Table 3 summarizes the results. The top two rows report #1 denotes the single-shot mode and #2 denotes the hold-to-activate mode. The lower five rows show results for each cue. Although the metrics differ between the modes, the differences are modest. By instruction category, the downward direction showed the lowest performance with $R = 0.909$, $P = 0.882$, and $F = 0.896$. Likely causes include reduced contrast due to upper-eyelid descent and eyelash occlusion during downward gaze, interference between the threshold regions and facial geometry or the eyeglass frame, and mismatches between the off-area design and camera pose. For the eye-closure cue alone, performance was $R = 1.000$, $P = 0.992$, and $F = 0.996$, indicating that reliable switch operation is achievable even when oculomotor control is limited. Overall, the performance gap between modes is acceptable and can be managed through operational policy.

Table 3. Results of the switch-operation experiment. The top two rows summarize overall performance for #1 single-shot mode and #2 hold-to-activate mode. The lower rows report metrics by instruction category (right, left, up, down, close [eyes]). *TP*: true positives; *FN*: false negatives; *FP*: false positives; *R*: recall; *P*: precision; *F*: F-score.

Mode	*TP*	*FN*	*FP*	*R*	*P*	*F*
# 1	326	4	27	0.988	0.924	0.955
# 2	321	9	1	0.973	0.997	0.985
right	132	0	2	1.000	0.985	0.992
left	131	1	0	0.992	1.000	0.996
up	132	0	9	1.000	0.936	0.967
down	120	12	16	0.909	0.882	0.896
close [eyes]	132	0	1	1.000	0.992	0.996

Figure 8 shows an example of an FN. The participant executed a downward gaze, but the system classified the frame as "closed eye." A human observer would likely make the same judgment. We infer that, for this participant, the upper eyelid moved substantially during downward gaze. Similar tendencies were observed in other FNs. This issue could likely be mitigated by adjusting the camera position. However, because this study involved healthy participants, and eyelid kinematics may differ from those of the ALS users for whom the system is intended, we must verify whether the FNs observed here represent problems that should be fixed or instead reflect a mismatch between the experimental participants and the target user population. Additionally, during downward gaze, we observed cases where the pupil center was detected slightly outside the contour landmarks, resulting in misclassification. Addressing this will require further improvement in pupil-center detection accuracy.

The hold-to-activate mode demonstrated exceptional precision and robust resistance to false activations, making it well suited for operations that prioritize minimizing missed emergency calls. The single-shot mode offers clear intent reflection but tends to show more FPs due to small pupil-position jitter and blinks. The reduced performance for the "down" direction is mainly attributable to upper-eyelid descent and eyelash occlusion, which

lower pupil contrast during downward gaze, together with geometric interference between the off-area placement, facial contour, and the eyeglass frame. To improve robustness, we have added a description of practical adjustments that have been verified through internal testing. Enlarging the inferior off-area angle, slightly reducing its radius, and extending the dwell-time parameter (t^*) by about 0.5 s for downward activation can better accommodate movement trajectories and suppress transient false activations. Caregivers can individually tune these parameters through the interface, allowing adaptation to user-specific eye morphology and usage conditions. These refinements are expected to stabilize downward-gaze detection without adding computational cost.

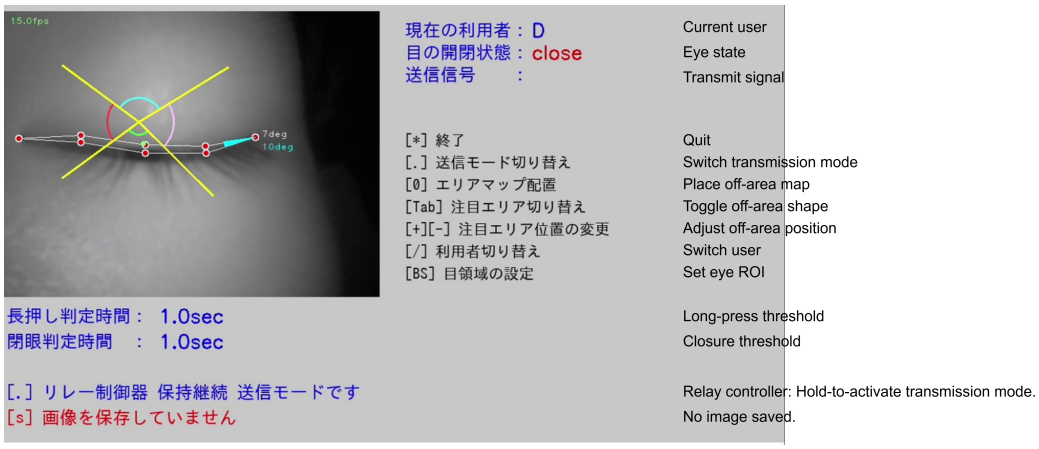


Figure 8. Example of a false negative (downward gaze misjudged as "closed eye"). During downward gaze, the upper eyelid descends and eyelashes occlude the pupil region, lowering contrast and leading the system—as well as human observers—to classify the frame as closed. Such cases suggest mitigation by camera placement and threshold tuning, or by auxiliary off-areas dedicated to downward gaze. In the deployed system, all on-screen text is presented in Japanese; English labels are used here for publication purposes.

Overall, the results indicate that the proposed system achieves both high recall and high precision under low-light, supine, and network-independent conditions. In particular, the hold-to-activate mode and blink-based input contribute to balancing false-activation suppression with operational reliability in practical use.

In addition to the controlled evaluation reported here, the system has been used in real-world settings by multiple ALS patients. It has been trialed by numerous attendees at public exhibitions (including the Osaka–Kansai Expo). Across these deployments, we did not observe any operational problems, which supports the practical viability of the approach. Furthermore, through these real-world uses and public demonstrations, the system maintained stable responsiveness even when the processing rate was reduced to approximately 10 frames per second (fps). This confirms that the operating speed of approximately 18 fps used in this study is sufficient for real-time assistive communication, providing a good balance between responsiveness and computational efficiency on a GPU-less portable computer.

6. Conclusions

Focusing on oculomotor and eyelid movements—which are often relatively preserved as residual functions in people living with ALS—we designed and implemented an improved eyeglass-type switch building on our previous system. By enabling standalone operation on a commodity notebook PC, diversifying switch activation behaviors, and extending signal output options, we increased applicability in real-world environments

and enhanced operational flexibility. These improvements were guided by feedback from patients and caregivers and constitute the core contribution of this work.

On the design side, we adopted a caregiver-centric, easily adjustable approach: manual ROI setting according to imaging conditions; selection and parametric tuning of off-area shapes (rectangular or sector); and implementation of four output modes based on oculomotor input—single, continuous, long-press, and hold-to-activate—together with the same four modes based on eyelid input. This yielded an operating interface adaptable to diverse symptoms and movement characteristics.

In a user study involving 12 healthy participants, the F-scores were 0.965 for the single mode and 0.983 for the hold-to-activate mode, outperforming the prior system and providing quantitative evidence of utility. Analysis of error patterns showed that false negatives were concentrated in the "down" direction, with cases attributable to large eyelid motion and small pupil-center estimation errors. These issues suggest opportunities for improvement through camera placement optimization and further enhancement of pupil estimation accuracy.

Overall, the system developed in this paper addresses the limitations of prior versions while simultaneously providing a caregiver-oriented UI/operability and strong false-activation suppression, thereby increasing the feasibility of deployment in both clinical and home settings. Going forward, we will continue field evaluations by loaning units to patients to further enhance reliability.

In sum, the proposed eyeglass-type switch demonstrates high practicality as an input channel for emergency alerts and AAC systems, and it represents a promising option for supporting communication among people living with ALS.

We did not pre-stratify participants by eye size or shape in the present study. Instead, we relied on caregiver-tunable ROI placement and adjustable off-area geometry to accommodate inter-individual variation. As future work, we will retrospectively analyze correlations between error metrics and proxy measures of eye size (e.g., ROI dimensions, inter-canthal landmark distance) and design a prospective study with predefined ocular size/shape categories to report stratified performance. We also plan to enhance the robustness of downward gaze detection through improved thresholding and feature integration, conduct larger-scale clinical evaluations with ALS users to validate generalizability, and expand the system to integrate seamlessly with other assistive communication and environmental control interfaces. In parallel, we aim to explore semi-automated calibration methods and adaptive algorithms that allow the eyeglass-type switch to self-adjust to individual variations and long-term use. These directions will contribute to the broader clinical adoption and sustainable deployment of the proposed system in real-world care environments.

Future work will focus on enhancing the robustness of downward gaze detection through improved thresholding and feature integration, conducting larger-scale clinical evaluations with ALS users to validate generalizability, and expanding the system to integrate seamlessly with other assistive communication and environmental control interfaces. Future work will also include formal human factors evaluations and refinement of caregiver-driven error recovery mechanisms to enhance long-term usability and safety in clinical and home environments.

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

Funding: This research was funded by the JSPS KAKENHI Grant Number 23H03787.

Institutional Review Board Statement: This study has been approved by the research ethics committee of the Graduate School of Computer Science and Systems Engineering, Kyushu Institute of Technology (Approval number: 20-02, 6 June 2020), and by the Institutional Review Board of the National Rehabilitation Center for Persons with Disabilities (Approval number: 2020-031, 29 May 2020).

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

Data Availability Statement: The data presented in this study are available on request from the corresponding author due to participant privacy restrictions.

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

References

1. Sakamoto, K.; Saitoh, T.; Itoh, K. Development of Night Time Calling System by Eye Movement Using Wearable Camera. In *HCI International 2020—Late Breaking Papers: Universal Access and Inclusive Design, Proceedings of the 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, 19–24 July 2020*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2020; Volume 12426, pp. 345–357. [CrossRef]
2. Tobii Dynavox. Dwell Time and Selection Options—User Documentation and Support Articles. 2025. Available online: https://www.tobiidynavox.com/ (accessed on 21 October 2025).
3. OptiKey. Optikey V4. Available online: https://www.optikey.org/ (accessed on 21 October 2025).
4. Isomoto, T.; Yamanaka, S.; Shizuki, B. Dwell Selection with ML-based Intent Prediction Using Only Gaze Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *6*, 1–21. [CrossRef]
5. Cai, S.; Venugopalan, S.; Seaver, K.; Xiao, X.; Tomanek, K.; Jalasutram, S.; Morris, M.R.; Kane, S.; Narayanan, A.; MacDonald, R.L.; et al. Using large language models to accelerate communication for eye gaze typing users with ALS. *Nat. Commun.* **2024**, *15*, 9449. [CrossRef] [PubMed]
6. Hitachi KE Systems, Ltd. Den-no-Shin. Available online: https://www.hke.jp/products/dennosin/denindex.htm (accessed on 21 October 2025).
7. OryLab, Inc. OriHime eye+Switch. Available online: https://orihime.orylab.com/eye/ (accessed on 21 October 2025).
8. Cyberdyne, Inc. CYIN. Available online: https://www.cyberdyne.jp/products/cyin.html (accessed on 21 October 2025).
9. Accessyell Co., Ltd. Fine-Chat. Available online: https://accessyell.co.jp/products/fine-chat/ (accessed on 21 October 2025).
10. Ezzat, M.; Maged, M.; Gamal, Y.; Adel, M.; Alrahmawy, M.; El-Metwally, S. Blink-To-Live eye-based communication system for users with speech impairments. *Sci. Rep.* **2023**, *13*, 7961. [CrossRef] [PubMed]
11. Golparvar, A.J.; Yapici, M.K. Graphene Smart Textile-Based Wearable Eye Movement Sensor for Electro-Ocular Control and Interaction with Objects. *J. Electrochem. Soc.* **2019**, *166*, B3184–B3193. [CrossRef]
12. Golparvar, A.J.; Yapici, M.K. Toward graphene textiles in wearable eye tracking systems for human–machine interaction. *Beilstein J. Nanotechnol.* **2021**, *12*, 180–190. [CrossRef] [PubMed]
13. Hosp, B.; Eivazi, S.; Maurer, M.; Fuhl, W.; Geisler, D.; Kasneci, E. RemoteEye: An open-source high-speed remote eye tracker. *Behav. Res. Methods* **2020**, *52*, 1387–1401. [CrossRef] [PubMed]
14. Steffan, A.; Zimmer, L.; Arias-Trejo, N.; Bohn, M.; Ben, R.D.; Flores-Coronado, M.A.; Franchin, L.; Garbisch, I.; Wiesmann, C.G.; Hamlin, J.K.; et al. Validation of an Open Source, Remote Web-based Eye-tracking Method (WebGazer) for Research in Early Childhood. *Infancy* **2024**, *29*, 31–55. [CrossRef] [PubMed]
15. Larumbe-Bergera, A.; Garde, G.; Porta, S.; Cabeza, R.; Villanueva, A. Accurate Pupil Center Detection in Off-the-Shelf Eye Tracking Systems Using Convolutional Neural Networks. *Sensors* **2021**, *21*, 6847. [CrossRef] [PubMed]
16. Ou, W.L.; Kuo, T.L.; Chang, C.C.; Fan, C.P. Deep-Learning-Based Pupil Center Detection and Tracking Technology for Visible-Light Wearable Gaze Tracking Devices. *Appl. Sci.* **2021**, *11*, 851. [CrossRef]
17. Chinsatit, W.; Saitoh, T. CNN-Based Pupil Center Detection for Wearable Gaze Estimation System. *Appl. Comput. Intell. Soft Comput.* **2017**, *2017*, 8718956. [CrossRef]
18. Xiong, J.; Dai, W.; Wang, Q.; Dong, X.; Ye, B.; Yang, J. A review of deep learning in blink detection. *PeerJ Comput. Sci.* **2025**, *11*, e2594. [CrossRef] [PubMed]
19. Liu, M.; Bian, S.; Zhao, Z.; Zhou, B.; Lukowicz, P. Energy-efficient, low-latency, and non-contact eye blink detection with capacitive sensing. *Front. Comput. Sci.* **2024**, *6*, 1394397. [CrossRef]
20. Onkhar, V.; Dodou, D.; de Winter, J.C.F. Evaluating the Tobii Pro Glasses 2 and 3 in static and dynamic conditions. *Behav. Res. Methods* **2023**, *56*, 4221–4238. [CrossRef] [PubMed]

21. Zhu, L.; Chen, J.; Yang, H.; Zhou, X.; Gao, Q.; Loureiro, R.; Gao, S.; Zhao, H. Wearable Near-Eye Tracking Technologies for Health: A Review. *Bioengineering* **2024**, *11*, 738. [CrossRef] [PubMed]
22. PyTorch. PyTorch. Available online: https://pytorch.org/ (accessed on 21 October 2025).

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