Response to Final Report Review Comments

ECE 445 Senior Design Laboratory

Team #45

AI Navigation Glasses for the Visually Impaired

Submitted by:
Shengnan Cai, Junchen He, Yi Su, Mingyan Gao

May 27, 2026

1. Overview

We sincerely thank the reviewers for their careful evaluation and constructive feedback. We have carefully revised our final report and supplementary materials to address all concerns.

The major revisions include:

- Enlarged and redesigned system architecture figures for readability
- Added complete physical prototype photographs
- Added detailed hardware power supply and voltage regulation design
- Expanded verification scenarios with real testing evidence and quantitative results
- Added detailed subsystem validation descriptions

This document provides a point-by-point response to each reviewer comment.

2. Response to Reviewer 1

2.1. Introduction (Visual Aid / Major Subsystems)

Reviewer Comment: *Provide visual aid including all major subsystems and block diagram.*

Response: We agree with the reviewer that the original Introduction did not provide a sufficiently explicit top-level architectural visualization of the complete system.

To address this, we added a new large-format system overview diagram to the Introduction. The revised diagram presents the full end-to-end architecture of the proposed wearable assistive navigation system and explicitly illustrates the major subsystem decomposition and their functional interconnections.

Specifically, the new visual aid shows:

- the real-world sensing context and visually impaired user interaction;
- the wearable hardware layer, including the ESP32-S3 glasses unit, camera, microphone, Wi-Fi module, and speaker;
- the wireless communication layer for video streaming, audio upload, and bidirectional WebSocket communication;
- the backend processing layer centered around the FastAPI server;
- the AI perception modules, including vision processing, obstacle detection, crosswalk detection, traffic-light recognition, and speech processing;
- the assistive decision layer, including the navigation state machine and hazard-awareness logic;
- the final feedback/output layer that generates directional audio guidance, crossing prompts, obstacle alerts, and live system feedback.

This revised block diagram provides a clearer subsystem-level abstraction of the complete perception-to-feedback pipeline and better supports the subsystem overview presented in the Introduction.

Revision Implemented: A new introduction-level system architecture figure was added to explicitly present all major subsystems and their interconnections.

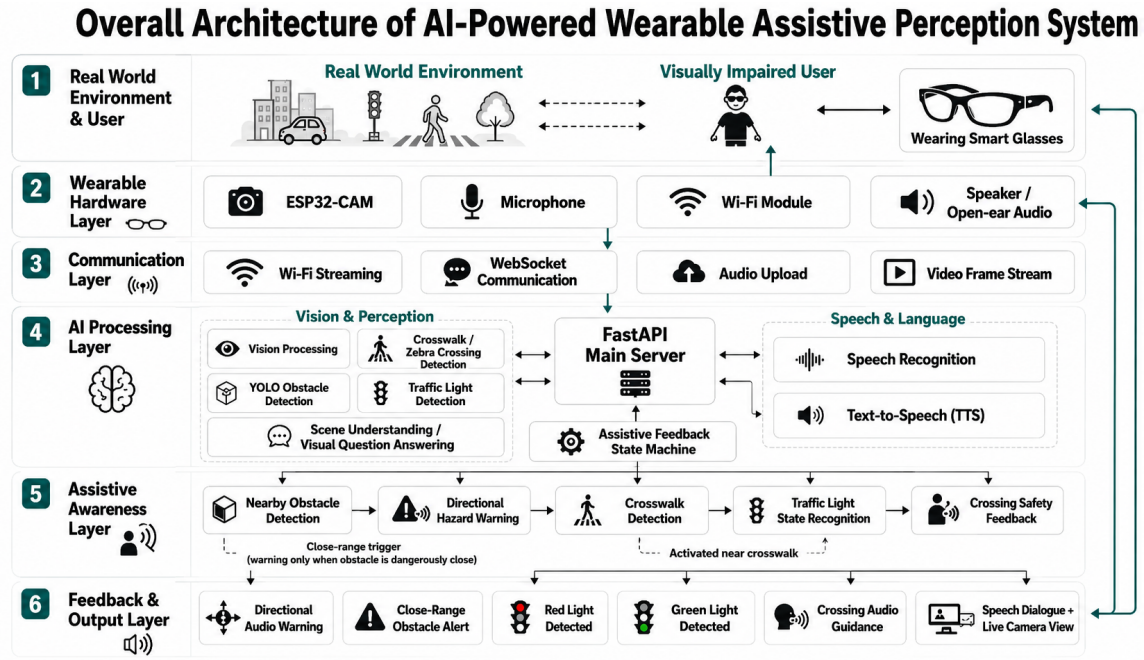

Figure 1: Overall architecture of the AI-powered wearable assistive perception system, showing the hardware, communication, backend processing, AI perception, assistive decision-making, and feedback/output subsystems.

2.2. Design (Physical Design / Power Supply)

Reviewer Comment: *Provide detailed physical design. Describe power supply including battery and voltage regulation circuits.*

Response: We agree that the original report did not provide enough detail about the physical hardware implementation and power architecture. The previous Design section focused mainly on the system architecture, vision pipeline, audio pipeline, and navigation logic, but it did not sufficiently explain how the wearable prototype was physically assembled, how the electronics were protected, or how the final prototype was powered.

To address this comment, we added a new Design subsection titled "Physical Prototype and Power Architecture." This subsection now describes the final hardware implementation of the wearable glasses prototype. The final prototype uses an OV5640 camera, a Seeed Studio XIAO ESP32-S3 Sense controller, a MAX98357 I2S audio amplifier, and a 3718 enclosed 8 Ω 2 W speaker. The IMU was considered in the early design stage, but it is not used in the current final prototype; therefore, the revised Design section explicitly removes the IMU from the implemented hardware list to avoid overstating the final implementation.

The revised physical design also explains the component placement on the glasses frame. The PCB, ESP32-S3 Sense board, MAX98357 amplifier, and signal wiring are placed inside a transparent upper electronics enclosure. This enclosure improves physical protection for the electronics and provides partial protection from rain or splashes while keeping the PCB and wiring accessible for debugging. The OV5640 camera is mounted near the front of the glasses to approximate the user's forward field of view. The speaker is mounted near the user's ear so that spoken navigation prompts can be heard clearly during testing. The camera and battery are placed below or outside the main electronics enclosure,

which helps shield them under the housing while also avoiding heat accumulation from placing the battery inside the same closed box as the PCB and controller.

We also expanded the power supply description. The final prototype is powered by a replaceable 5 V lithium battery pack rated at 12580 mWh. The battery output is connected to the XIAO ESP32-S3 Sense board through a physical power switch, allowing the prototype to be turned on and off directly. The ESP32-S3 Sense board receives the 5 V input and uses its onboard voltage regulation circuitry to generate the regulated logic rails required by the microcontroller and attached modules. In other words, the final prototype does not use a separate external regulator board; the voltage regulation stage is implemented by the onboard regulator of the XIAO ESP32-S3 Sense, which converts the 5 V battery input into the regulated logic supply required by the ESP32-S3 and attached low-voltage modules. The OV5640 camera and MAX98357 amplifier are powered from the ESP32-S3 Sense board power rails, and all modules share a common ground. No separate external charging module is used in the final prototype because the battery pack is replaceable rather than recharged inside the device.

To make the physical and electrical design clearer, the revised report adds a color-coded wiring diagram, a PCB routing figure, and close-up photographs of the assembled electronics enclosure. The wiring diagram explains the module-level physical connections among the ESP32-S3 Sense, peripheral modules, power, ground, and audio output path. The PCB routing figure shows how the power, ground, and signal traces are organized to reduce loose jumper wiring and improve physical integration. The added front and top-view prototype photographs show that the PCB, controller, amplifier, and wiring are physically integrated inside the transparent enclosure, while the camera and battery are mounted below the enclosure to support rain protection and heat management.

Revision Implemented: Added a new Design subsection titled "Physical Prototype and Power Architecture." This section now documents the final hardware list, component placement, wearable mounting rationale, transparent electronics enclosure, 5 V battery supply, physical power switch, onboard voltage regulation through the XIAO ESP32-S3 Sense board, shared-ground wiring, PCB routing, physical wiring diagram, and final enclosure photographs. The revised text also clarifies that the IMU is not used in the current final prototype.

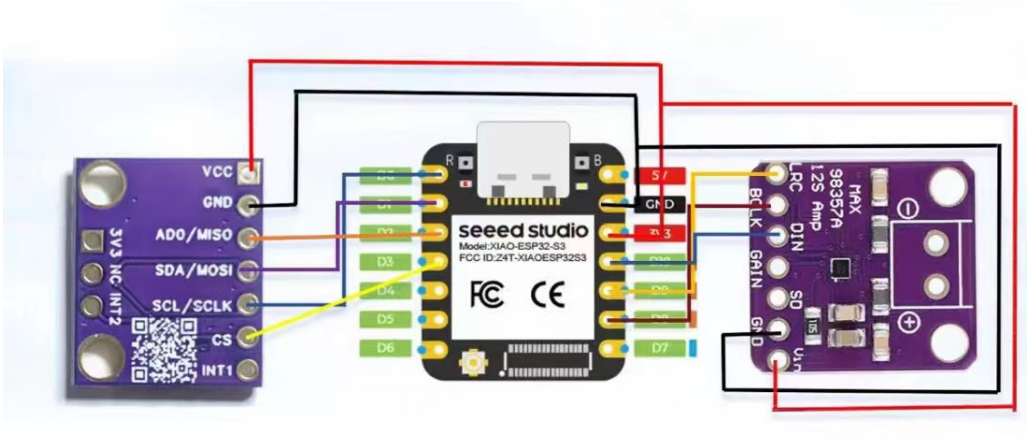

Figure 2: Color-coded physical wiring diagram for the wearable prototype. The XIAO ESP32-S3 Sense serves as the central controller. The OV5640 camera provides the visual input, while the MAX98357 I2S amplifier drives the 3718 enclosed 8 Ω 2 W speaker for spoken navigation prompts. Power and ground are distributed from the ESP32-S3 Sense board to the attached modules.

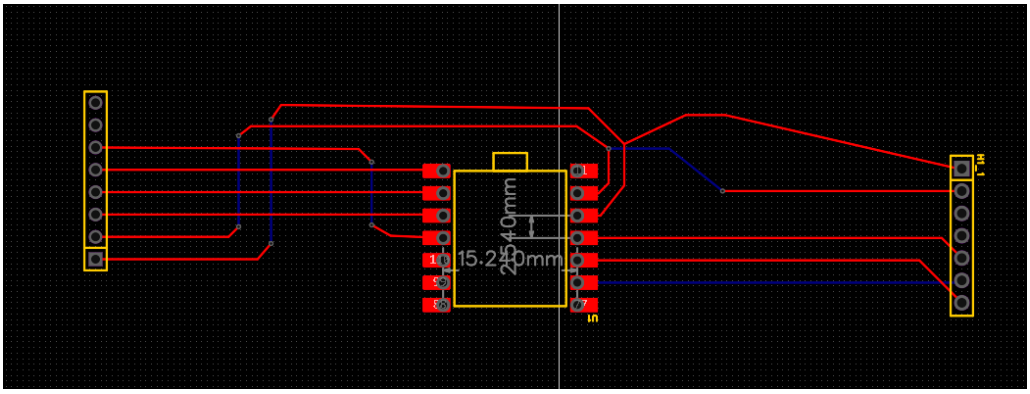

Figure 3: PCB routing layout used for the prototype hardware integration. The routing organizes the power, ground, and signal traces between the ESP32-S3 Sense board and the external connectors, reducing loose jumper wiring and improving the mechanical reliability of the wearable prototype.

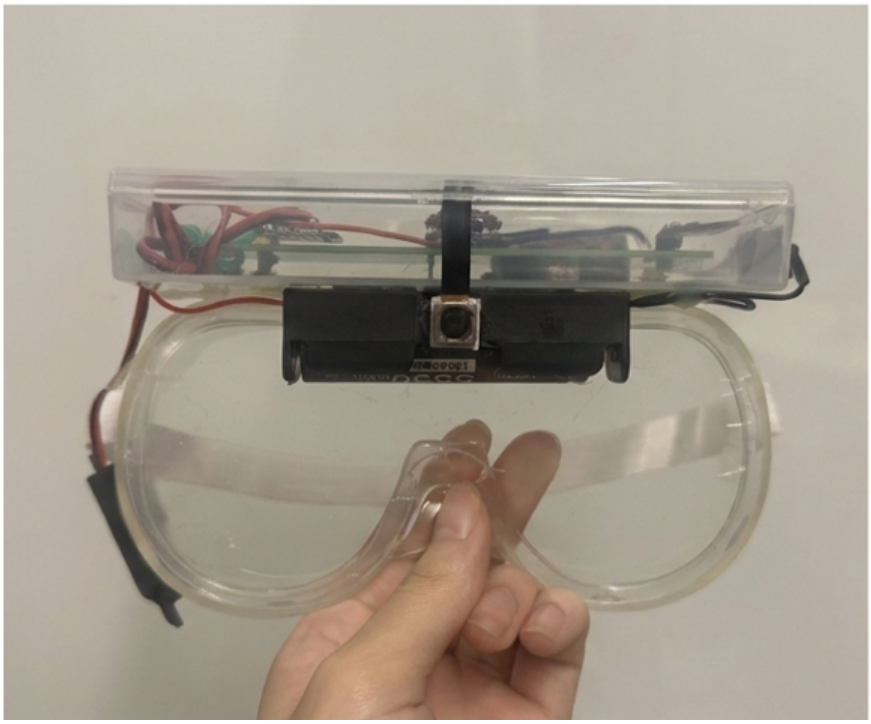

Figure 4: Front view of the final assembled prototype. The transparent upper enclosure contains the PCB, ESP32-S3 Sense board, MAX98357 amplifier, and wiring, which improves physical protection and partial rain shielding. The OV5640 camera and battery are mounted below the enclosure so that the camera remains aligned with the user's forward view while the battery is separated from the main electronics box for better heat management.

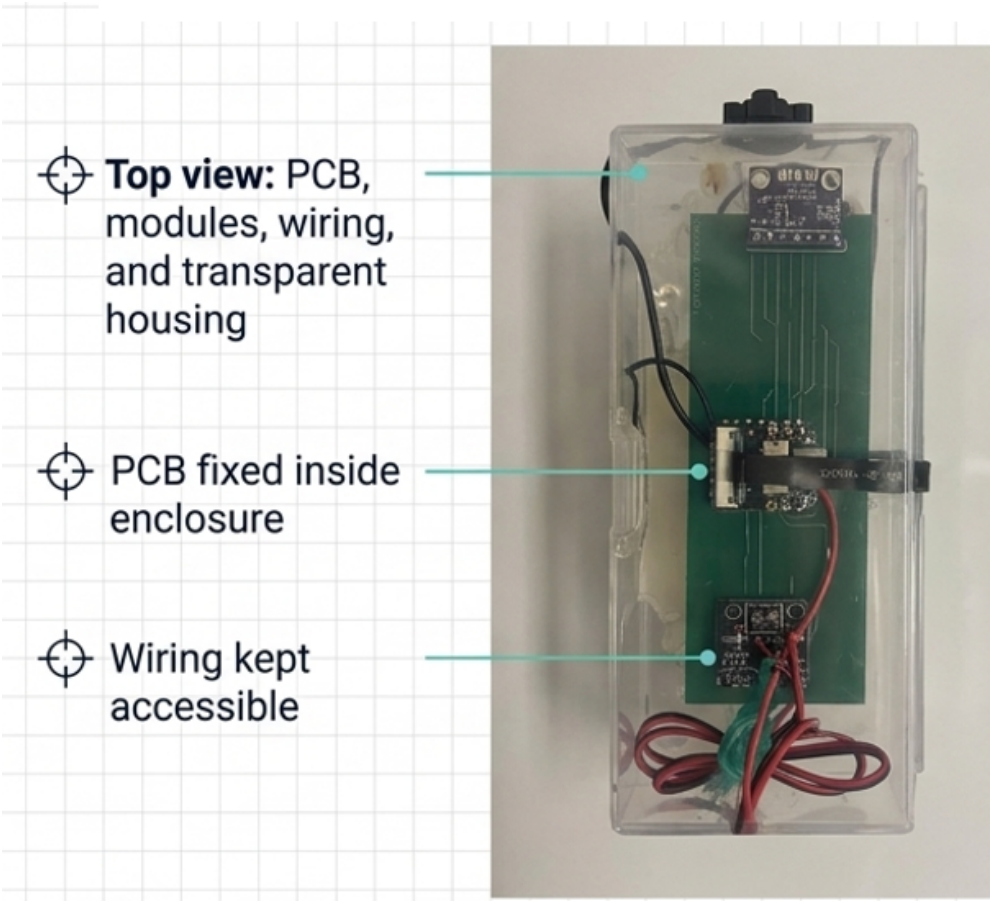

Figure 5: Top view of the electronics enclosure and internal PCB assembly. The PCB is fixed inside the transparent housing, while the wiring remains accessible for debugging and maintenance. This physical layout keeps the ESP32-S3 Sense, amplifier, PCB, and wiring organized inside the enclosure instead of relying on loose external jumper connections.

2.3. Requirements and Verification

Reviewer Comment: *Describe verification scenarios involving user and results.*

Response: We agree with the reviewer that the original Requirements and Verification section emphasized subsystem-level implementation evidence but did not sufficiently present the verification process from the user's operational perspective.

To address this issue, we substantially revised the section by reorganizing the verification around user-centered interaction scenarios. The revised version now explicitly describes the user action, the expected system response, the observed output, and the quantitative evidence supporting each requirement. This revision makes the verification procedures more reproducible and directly connects measured results to real-world usage scenarios.

The revised report now includes end-to-end verification scenarios covering live camera streaming, voice-command input, spoken feedback playback, blind-path following, obstacle warning, crosswalk assistance, traffic-light-gated crossing, and latency-sensitive interaction performance.

For obstacle-awareness verification, the report now clarifies the implemented decision logic from the user's perspective. The obstacle detector first identifies candidate obstacles and compares each obstacle

mask against the detected navigable-path region. If an obstacle sufficiently overlaps the navigable-path region, it is treated as a forward hazard and triggers a high-priority warning prompt. Otherwise, the system classifies the obstacle as a left-side or right-side object relative to the walking direction. This ensures that objects visible in the scene but not directly blocking the user's walking path do not generate unnecessary collision alerts.

The revised traffic-light verification section now includes manually reviewed outdoor validation results. Across representative traffic-light test frames, the detector achieved an overall state-agreement rate of 93.3% against manual inspection. Agreement rates across representative conditions were 94.6% under sunny daylight, 91.8% under rainy or cloudy conditions, 92.1% under low-light scenes, and 93.0% under strong illumination or glare. These results demonstrate reliable traffic-light-state recognition under representative outdoor conditions.

The no-light condition is now explicitly documented as a conservative fallback scenario rather than an unresolved failure case. When no stable traffic-light state is detected for repeated frames, the workflow enters a fallback branch that prevents unsafe immediate crossing decisions while maintaining workflow continuity during testing and demonstration.

The revised report also adds a quantitative latency-impact analysis to explain which components most affect the user's real-time experience. The dominant latency contributors were identified as ESP32-to-server camera uplink transmission and local TTS queueing/playback, while server-side ASR and visual-model inference contributed comparatively smaller delays. This addition better connects system timing behavior to practical user responsiveness.

Revision Implemented: The Requirements and Verification section was fully rewritten to include user-centered verification scenarios, quantitative traffic-light validation, clarified obstacle-warning logic, explicit no-light fallback behavior, and latency-impact analysis tied directly to end-user system responsiveness.

Table 1: User-centered verification scenarios and observed results

User Scenario	Expected Behavior	Observed Result	Status
Wearing the glasses during live operation	Continuous visual sensing and backend perception updates	The ESP32-S3 camera stream successfully reached the backend through `/ws/camera`, and processed frames were visualized through the browser monitor.	Pass
Issuing hands-free voice commands	Navigation modes should respond to spoken user commands	The microphone uplink delivered audio through `/ws_audio`, and recognized commands correctly triggered navigation-state transitions.	Pass
Receiving spoken navigation feedback	Audio prompts should provide real-time guidance	The backend generated navigation prompts and streamed playback through `/stream.wav`; prompts were reproduced through the wearable I2S speaker output.	Pass
Blind-path following	Directional correction guidance should be produced	The system detected tactile-path geometry and generated left, right, or straight guidance according to centerline and heading estimation.	Pass
Obstacle encountered along the walking route	Immediate warning should override routine guidance	Obstacle detections overlapping the navigable path triggered high-priority hazard alerts, while non-overlapping detections were classified as lateral objects.	Pass
Approaching a crosswalk	Crosswalk assistance should guide transition into waiting mode	The workflow successfully entered crosswalk-seeking and traffic-light waiting states after satisfying geometric proximity thresholds.	Pass
Traffic-light-controlled crossing	Crossing should occur only after repeated green confirmation	The system required repeated stable green detections before entering crossing; red and yellow states consistently blocked crossing.	Pass

Table 2: Traffic-light user-scenario verification results

Scenario	Agreement	Verification Basis	Observed Behavior	Status
Sunny daylight	94.6%	Manual frame review	Stable red/green classification with consistent temporal voting.	Pass
Rainy / cloudy outdoor	91.8%	Manual frame review	Reduced contrast occasionally delayed stabilization, but repeated-frame voting preserved correct state gating.	Pass
Low-light scenes	92.1%	Manual frame review	Valid traffic-light-state recognition remained stable when signals were visible.	Pass
Bright-light / glare	93.0%	Manual frame review	Temporal stabilization maintained robust classification under illumination variation.	Pass
No-light / unclear signal	Conservative fallback	Workflow inspection	The system entered the documented fallback branch and avoided unsafe immediate crossing decisions.	Pass
Overall	93.3%	Aggregated manual validation	Reliable traffic-light-state agreement across representative outdoor scenarios.	Pass

Table 3: Latency bottlenecks and user impact

Latency Source	Measured Evidence	User Impact	Optimization Direction
Camera uplink	131 ms (chat), 230–287 ms (obstacle), 248–288 ms (traffic-light)	Largest fixed visual-path latency component	Reduce frame size, JPEG quality, or transmission frequency
Local TTS queueing/playback	Queue wait up to 1896 ms; playback 1833–2912 ms	Most noticeable user-facing delay	Allow prompt interruption and prioritize newest safety-critical guidance
Vision inference	58 ms (obstacle), 32 ms (traffic-light)	Moderate backend processing delay	Reduce obstacle-detection frequency and optimize scheduling
JPEG decode/encode	16–30 ms per stage	Secondary visual-processing overhead	Lower viewer quality or disable viewer during latency-sensitive testing
Voice input path	0.1–0.3 ms transmission; 8–12 ms ASR callback	Negligible compared to visual/audio output latency	Lower optimization priority

3. Response to Reviewer 2

3.1. Introduction Readability

Reviewer Comment: *Increase fonts in Figure 1. Too small to be seen.*

Response: We agree with the reviewer that the visual elements in the Introduction section were not

sufficiently readable in the original submission.

To address this concern, we improved the readability of the subsystem overview by revising both the overall system architecture diagram and the subsystem summary table.

The architecture figure was redrawn at higher resolution with enlarged labels, simplified routing arrows, increased spacing between functional blocks, and reduced visual clutter.

The subsystem summary table was also reformatted with larger font size, increased row spacing, and more consistent column alignment. Several descriptions were condensed to improve readability while preserving all subsystem-level information.

Together, these revisions provide a clearer and more accessible introduction-level presentation of the system architecture and subsystem interconnections.

Revision Implemented: The introductory visual aids were revised for readability. Figure 1 was replaced with a higher-resolution large-font architecture diagram, and the subsystem summary table was reformatted with larger text and improved spacing.

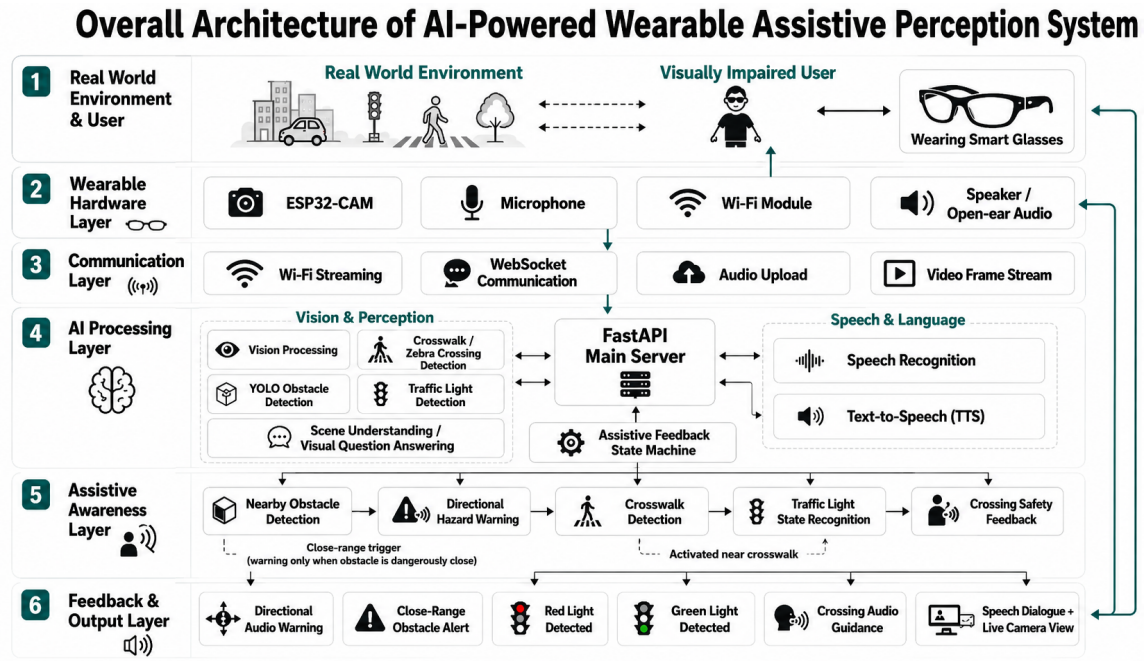

Figure 6: Revised large-font overall system architecture diagram.

Table 4: Revised subsystem overview and interconnect summary

Subsystem	Primary Function	Main Interconnects
Wearable hardware	Provides sensing, local audio playback, and physical user interaction.	ESP32-S3 GPIO, I2S, PDM microphone, camera connector, power rails
Embedded firmware	Manages camera capture, audio streaming, TTS playback, and timing control.	`/ws/camera`, `/ws_audio`, `/stream.wav`, timing metadata
Backend bridge	Receives live streams, maintains frame buffers, and distributes runtime state.	FastAPI WebSockets, HTTP WAV stream, browser monitor sockets
AI perception	Performs blind-path segmentation, obstacle detection, and traffic-light recognition.	YOLO segmentation outputs, YOLO-E obstacle masks, temporal traffic-light states
Navigation logic	Converts perception outputs into finite-state navigation decisions.	Crosswalk geometry, path centerline, obstacle overlap, stable-light counter
Voice interaction	Processes spoken commands and generates concise navigation prompts.	Paraformer ASR, command parser, prompt queue, I2S speaker output

3.2. Prototype Photos

Reviewer Comment: *Provide figures and photos of final prototype.*

Response: We agree that the original report did not provide enough visual evidence of the final physical prototype. To address this comment, we added final prototype photographs to the Design section. The added figures show the actual wearable prototype from the front, top, and side views, including the front-mounted OV5640 camera, the transparent upper electronics enclosure, the routed wiring, the internal PCB, the ESP32-S3 Sense board, the MAX98357 amplifier, the battery placement, and the speaker mounted near the user's ear.

These figures demonstrate that the system was physically integrated into a wearable glasses form factor rather than only described as a block diagram or software pipeline. The revised captions also explain the purpose of the physical layout: the camera is positioned near the front of the glasses to capture the user's forward view, the PCB and electronics are enclosed for protection and physical stability, the battery is separated from the main enclosure to reduce heat buildup, and the speaker is placed close to the ear to improve the audibility of navigation prompts.

Revision Implemented: Added final prototype photographs to the Design section. The new figures document the assembled wearable prototype, including front-view camera and battery placement, top-view PCB/electronics integration, and side-view speaker and wearable placement. The figure captions were expanded to explain how the physical layout supports the intended wearable navigation use case, rain protection, wiring stability, and heat management.

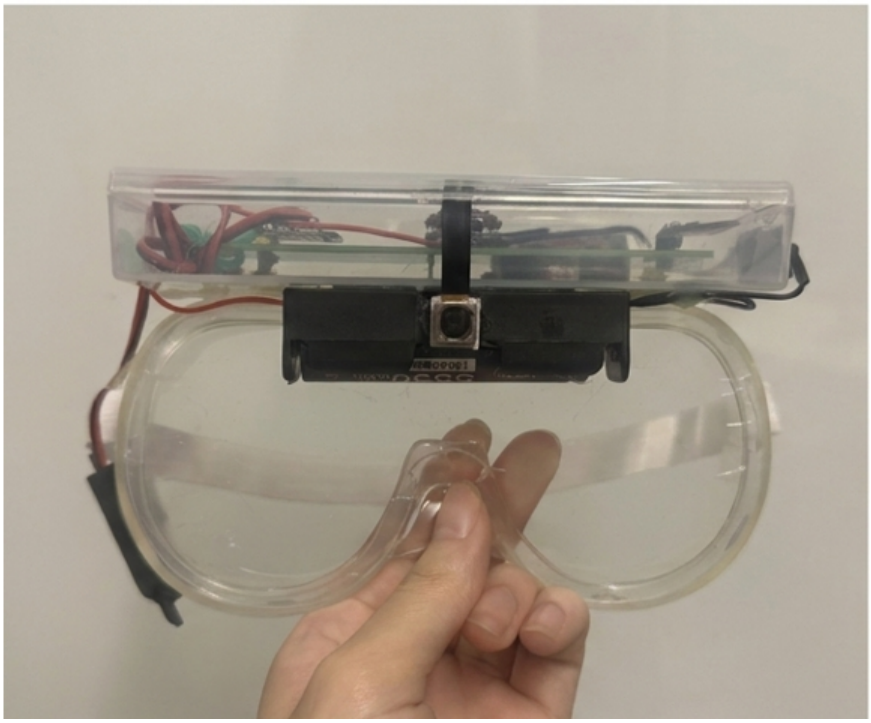

Figure 7: Final prototype front view. The camera is mounted below the transparent electronics enclosure and aligned with the forward walking direction. The PCB, ESP32-S3 Sense board, MAX98357 amplifier, and wiring are placed inside the upper enclosure for improved physical protection, while the battery is kept outside the main electronics box to reduce heat accumulation.

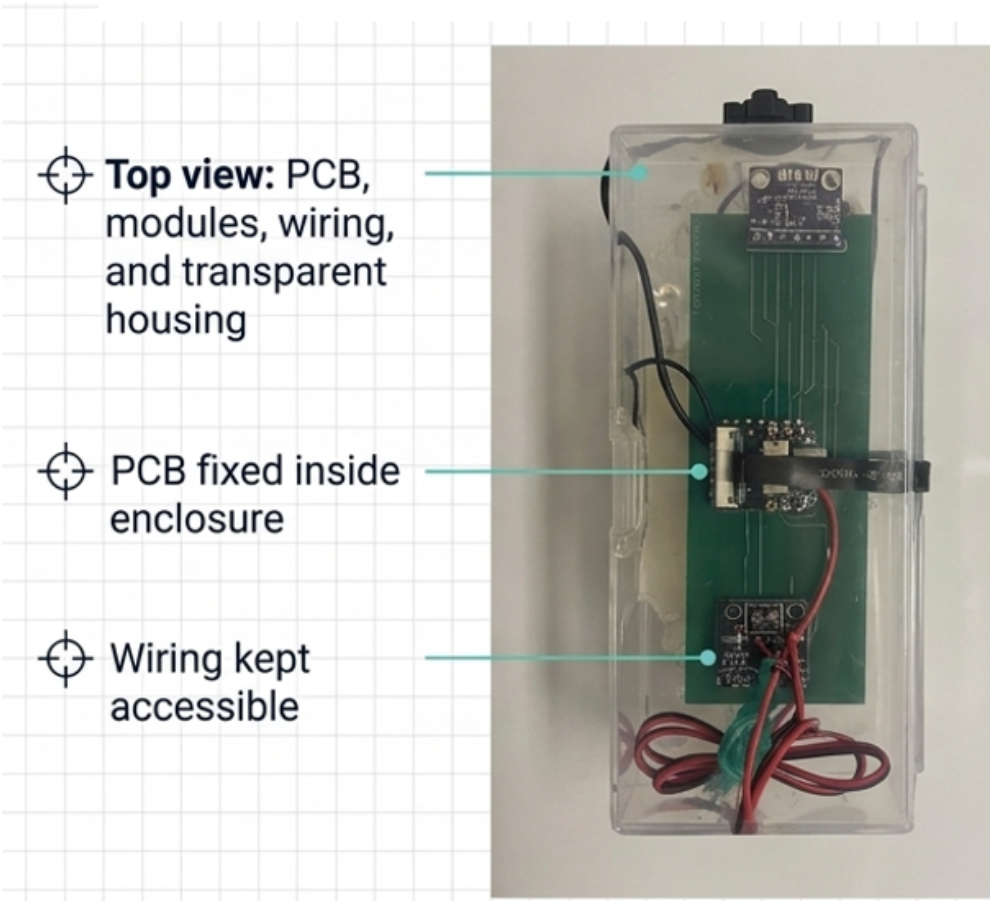

Figure 8: Top view of the final prototype electronics enclosure. The PCB is fixed inside the transparent housing, the modules and wiring are organized inside the box, and the wiring remains accessible for debugging and maintenance.

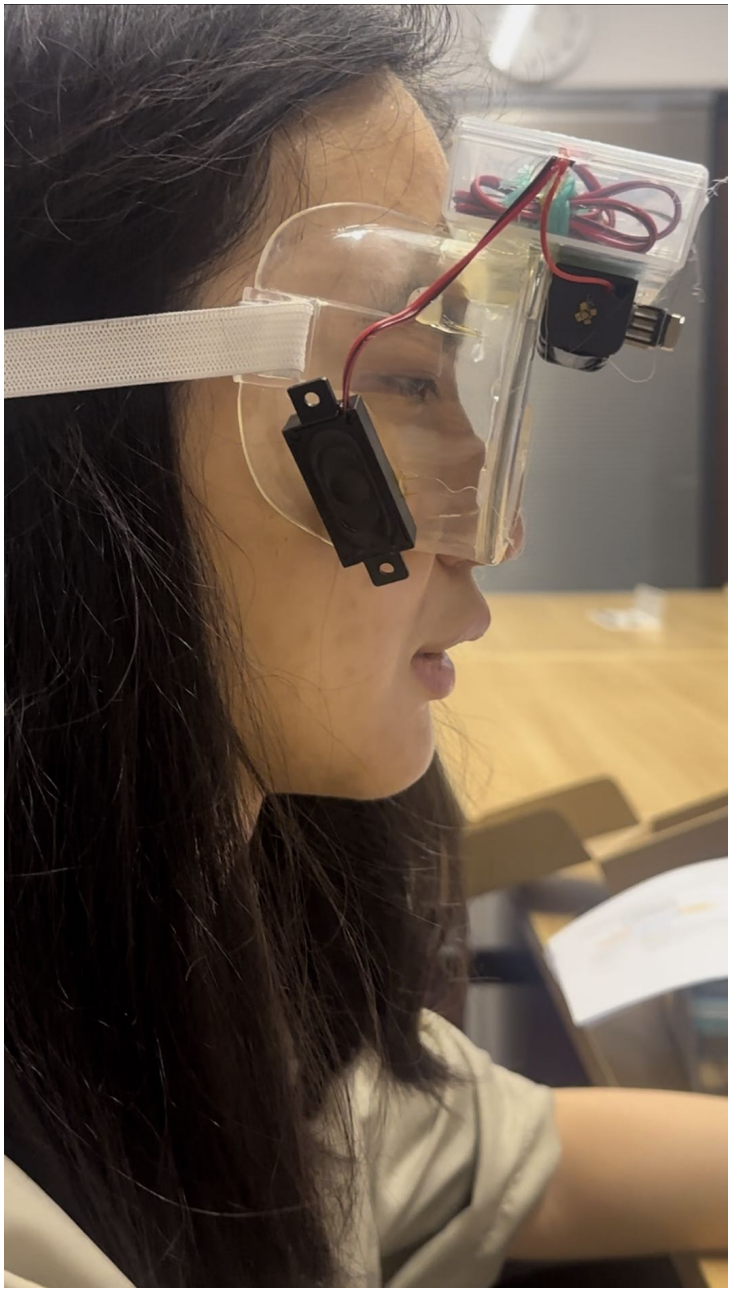

Figure 9: Final assembled wearable prototype during user-side testing. The OV5640 camera is mounted near the front of the glasses frame to approximate the user's forward field of view. The electronics enclosure and wiring are fixed near the upper/front portion of the frame, while the 3718 enclosed speaker is positioned near the user's ear for audible navigation feedback.

3.3. Verification Results Sufficiency

Reviewer Comment: *Verification results are insufficient.*

Response: We agree with the reviewer that the original verification section did not sufficiently connect subsystem-level measurements to realistic user-facing operational scenarios.

To address this concern, we substantially revised the Requirements and Verification section by re-organizing the evaluation around explicit user-centered test scenarios and by adding quantitative perfor-

mance evidence tied directly to practical system behavior.

First, the revised report introduces the traffic-light user-scenario verification table (Table 2), which evaluates system behavior under representative outdoor operating conditions including sunny daylight, rainy or cloudy environments, low-light scenes, bright illumination, and no-light fallback cases. For each scenario, verification was performed through manual inspection of detected traffic-light states during user-operated testing. Across all representative scenarios, the detector achieved an overall state-agreement rate of 93.3%, with individual scenario agreement ranging from 91.8% to 94.6%. This addition provides concrete user-level validation of traffic-light recognition performance under varying environmental conditions.

Second, the revised report adds the latency-impact analysis (Table 3), which quantifies the dominant timing contributors affecting user responsiveness. The analysis shows that the largest end-user delay arises from local TTS queueing and playback, with queue wait reaching up to 1896 ms and playback durations ranging from 1833 ms to 2912 ms depending on operating mode. Camera uplink latency contributes an additional 131–288 ms depending on task mode, while server-side visual inference remains comparatively moderate at approximately 32 ms for traffic-light detection and 58 ms for obstacle detection. The analysis further confirms that voice-input processing introduces negligible delay relative to visual and audio-output latency.

Finally, the revised verification section explicitly links each major requirement to a measurable user scenario, observed system output, quantitative result, or documented remaining limitation. This makes the revised evaluation substantially more reproducible and aligns it more directly with the practical use conditions of the wearable navigation system.

Together, these additions strengthen the verification evidence by moving beyond subsystem functionality checks toward scenario-based quantitative validation of user-facing system performance.

Revision Implemented: The verification section was comprehensively rewritten and expanded with three concrete additions. First, Table 1 was added to map each major user-facing scenario to its expected behavior, observed system output, and pass/fail status, including live camera operation, voice commands, spoken feedback, blind-path following, obstacle warning, crosswalk approach, and traffic-light-controlled crossing. Second, Table 2 was added to provide quantitative traffic-light verification results based on manual frame review, reporting a 93.3% overall state-agreement rate and scenario-specific agreement rates under sunny daylight, rainy/cloudy outdoor conditions, low-light scenes, and bright-light/glare conditions. Third, Table 3 was added to quantify latency bottlenecks from the user's perspective, including camera uplink latency, local TTS queueing/playback delay, vision inference time, JPEG decode/encode overhead, and voice-input latency. The revised section therefore no longer reports only subsystem availability; it now connects each verification claim to a user scenario, measured result, or documented fallback behavior.

4. Closing Statement

We sincerely appreciate the reviewers' valuable feedback. These revisions significantly improved the clarity, completeness, and technical rigor of our final report.

We believe the revised report now better reflects the full engineering scope and implementation quality of the AI Navigation Glasses system.