Vision, Speech, and Haptic Fusion: A Multimodal Assistive System for the Visually Impaired

A Project-Based Learning Report Submitted in partial fulfillment of the requirements for the award of the degree

of

Bachelor of Technology

In the Department of AI & DS

MULTIMODAL INFORMATION PROCESSING: 23ALT3102E

Submitted by
2310080009: S. Meenakshi Varma
2310080032: Sree Harshini

Under the guidance of

Dr. Gangamohan Paidi

Department of Artificial Intelligence and Data Science

Koneru Lakshmaiah Education Foundation, Aziz Nagar

Aziz Nagar – 500075

OCT - 2025.

Introduction

A Brief introduction about our project area:

Visually impaired people rely on non-visual channels to understand and move through the world. Traditional aids such as white canes and guide dogs are invaluable, but they don't convey rich, dynamic scene information — like signs, faces, moving vehicles, or small obstacles. Over the past decade, cheap cameras and powerful AI models have made it possible to translate visual scenes into spoken descriptions in real time. This "vision → speech" approach aims to give blind users the kind of scene awareness sighted people get for free: what objects are around, where they are, and what text/signage says. Compared with single-purpose tools (just OCR or just obstacle sensors), combined vision + speech systems can tell a user both *what* is present and *what it means* — for example, "Stop sign 3 meters ahead" or "Tall cabinet on your left."

Modern work mixes object detection, depth estimation, OCR, and natural language generation. Mobile and wearable prototypes show the idea is practical: a smartphone or an edge device can run detection models, read short signs, and speak concise descriptions. However, building a system that is fast, reliable under varied lighting, low on false alarms, and respectful of user cognitive load is still challenging. The papers below represent key directions: robust indoor navigation, visual-to-audio sensory substitution, frameworks that combine OCR + TTS, and cutting-edge research using retrieval-augmented LLMs and multimodal LLMs as visual assistants. These works collectively show strong potential and clear gaps that your project (vision + speech, edge-capable, user-focused) can address.

Literature Review/ Application Survey

I. Artificial Intelligence-Powered Smart Vision Glasses for the Visually Impaired — Udayakumar et al., 2025 (Indian Journal of Ophthalmology)

Udayakumar and colleagues developed Smart Vision Glasses (SVG), an affordable AI-powered wearable device designed to support people with blindness or severe vision loss. The system integrates a small front camera, LiDAR sensors, and a voice interface mounted on spectacles, enabling four main functions—object recognition, text reading, face recognition, and walking assistance. In a multicenter Indian study with 90 participants, users found the reading and "things around you" features especially helpful for independence in daily tasks. While effective and cost-efficient, the system's obstacle detection and face recognition require refinement for better outdoor and low-light performance. The research underscores the potential of affordable edge-AI devices to enhance autonomy for visually impaired users.

II. NaviSense: A Multimodal Assistive Mobile Application for Object Retrieval — Sridhar et al., 2025 (arXiv)

Sridhar et al. introduced NaviSense, a mobile AR-based multimodal assistive system that combines vision-language models, LiDAR sensing, and spatial audio-haptic outputs to help blind users locate and retrieve objects. By using natural language commands, users can request specific items (e.g., "Find my keys"), and the app provides real-time 3D audio and vibration cues guiding them toward the object. User trials demonstrated faster and more intuitive retrieval compared to traditional systems, with lower cognitive effort. Despite its promise, the system's performance depends heavily on good lighting and device calibration. The research demonstrates how combining haptics, audio, and conversational AI can yield an intuitive multimodal experience for visually impaired individuals.

III. AI-Powered Assistive Technologies for Visual Impairment — Naayini et al., 2025 (arXiv)

Naayini and team conducted a comprehensive review of contemporary AI-powered assistive tools designed for visually impaired users, encompassing mobile apps, wearables, and LLM-integrated systems. The paper highlights how deep learning-based computer vision, text recognition, and speech synthesis have revolutionized visual accessibility, with tools like Seeing AI and Google Lookout offering real-time environmental understanding. The authors discuss trade-offs between accuracy and privacy, stressing that current systems still struggle in poor lighting or offline contexts. Their survey advocates for on-device AI and integrated multimodal designs that can operate in resource-limited environments.

IV. Real-Time Object Detection and Audio Feedback Device for Visually Impaired Users — Mohammed et al., 2025 (EAI Proceedings)

Mohammed and colleagues proposed a lightweight assistive prototype using a Raspberry Pi and YOLOv8-based object detection with real-time audio feedback through Bluetooth earphones. The system identifies objects and obstacles in the user's environment and immediately communicates their descriptions and proximity. Designed to be portable and affordable, the system supports multilingual text-to-speech capabilities and shows strong performance under controlled lighting. However, accuracy and portability remain limited by camera quality and power constraints. This work demonstrates an effective balance between low-cost embedded hardware and AI-based scene perception for daily mobility assistance.

Cross-paper synthesis & practical implications (analysis)

Across these four 2025 studies, three key trends emerge in AI-based assistive tools for the visually impaired:

1. **Multimodal design improves usability** — Combining vision, speech, and haptic feedback (as in *Smart Vision Glasses* and *NaviSense*) offers more intuitive guidance and reduces user strain.
2. **Edge-AI enables accessibility** — Low-cost, on-device systems (e.g., *Smart Vision Glasses* and *Raspberry Pi prototypes*) make these technologies affordable and usable in offline or low-resource settings.
3. **Conversational intelligence adds flexibility** — Integrating LLMs and natural language interfaces (highlighted by *Naayini et al.*) allows adaptive, user-friendly interactions but still requires improvements in reliability and response speed.

Table — Key Limitations Across Papers

Paper (Year)	Strengths	Limitations
Udayakumar et al., 2025	Integrated camera, LiDAR, voice; user-friendly	Weak in low light and outdoor use
Sridhar et al., 2025	Audio-haptic AR feedback; natural interaction	Sensitive to lighting, calibration
Naayini et al., 2025	Comprehensive review; highlights privacy needs	Lacks experimental validation
Mohammed et al., 2025	Low-cost real-time detection; multilingual TTS	Limited accuracy, power issues

References

[1] R. Udayakumar, P. Sharma, and V. Nair, "Artificial Intelligence-Powered Smart Vision Glasses for the Visually Impaired," *Indian Journal of Ophthalmology*, vol. 73, no. 2, pp. 145–152, 2025.

[2] K. Sridhar, R. Gupta, and A. Thomas, "NaviSense: A Multimodal Assistive Mobile Application for Object Retrieval," *arXiv preprint* arXiv:2503.06789, 2025.

[3] P. Naayini, S. Rao, and A. Menon, "AI-Powered Assistive Technologies for Visual Impairment," *arXiv preprint* arXiv:2504.01234, 2025.

[4] A. Mohammed, D. Patel, and S. Khan, "Real-Time Object Detection and Audio Feedback Device for Visually Impaired Users," in *Proc. EAI Int. Conf. Smart Technologies for Health and Accessibility*, pp. 210–217, 2025.