Navsight: AI-Powered Assistive Spatial Memory and Real-Time Navigation for the Visually Impaired

MITHUN 1 , NAVEEN K 2, NAVEEN R 3 ,NAVEEN B 4 ,MOHAMED IRFAN 5

1 Assistant Professor, Dept. of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India.MAIL: mithuncse@siet.ac.in

2 Student, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India.MAIL: naveenk25cs@srishakthi.ac.in

3 Student, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India.MAIL:naveenr25cs@ srishakthi.ac.in

4 Student, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India.MAIL: naveenb25cs@srishakthi.ac.in

5 Student, Department of Computer Science and Engineering, Sri Shakthi Institute of Engineering and Technology, India.MAIL: mohamedirfann25cs@srishakthi.ac.in

ABSTRACT — NavSight is an AI-assisted spatial awareness, object identification, and conversational navigation system designed to empower visually impaired individuals in navigating complex indoor and outdoor environments independently. The system integrates computer vision object detection, ambient audio processing, spatio-temporal memory mapping, multi-language speech interaction (English and Tamil), and proactive collision detection. A lightweight web-accessible architecture powered by Node.js, Web Speech API, spatial detection algorithms, and localized audio feedback supports real-time performance. An operational audit reports 100% passing test modules across navigation, object detection adapters, and conversational context tracking. NavSight bridges spatial memory gaps by providing source-grounded voice guidance, spatial tracking, and emergency obstacle alerts while maintaining minimal cognitive overhead for users.

Keywords: spatial memory, assistive technology, object detection, visual impairment, audio guidance, real-time navigation, conversational AI, accessibility.

I. INTRODUCTION

Independent mobility remains one of the primary daily challenges faced by visually impaired individuals. Navigating unfamiliar or dynamic physical spaces requires continuous spatial awareness, tracking misplaced personal items, detecting drop-offs or static obstacles, and interpreting environmental cues. Traditional assistance tools like white canes and guide dogs offer immediate physical hazard awareness but lack contextual comprehension, object locator capabilities, and natural language communication interfaces.

Recent developments in computer vision and artificial intelligence offer promising solutions for visual scene understanding. However, most existing wearable or smartphone-based navigation apps suffer from high battery latency, overly complex interaction paradigms, lack of spatial memory for misplaced objects, and an absence of contextual localized language assistance (such as Tamil and English dialect adaptors).

NavSight addresses these operational shortcomings by establishing a unified spatio-temporal memory and real-time audio navigation architecture. The system continuously maps the user's environment, tracks object locations with positional coordinates, alerts the user to incoming collision risks, and answers contextual queries (e.g., "Where did I leave my keys?"). NavSight distinctly separates three critical assistance modalities: real-time obstacle alert, spatial object memory logging, and interactive conversational assistance. The system ensures that critical safety alerts override conversational dialogs immediately to prevent accidents.

Spatial memory capabilities allow the system to tag objects in 3D space relative to user movement, storing item location history locally. Unclear or obstructed items are logged with low-confidence scores and prompted for user re-verification, ensuring false object detections do not mislead the user.

II. EXISTING SYSTEM AND LITERATURE SURVEY

Existing navigation solutions for visually impaired users range from ultrasonic smart canes to smartphone visual recognition applications like Microsoft Seeing AI and Google Lookout [1]. While effective for isolated optical character recognition (OCR) or basic scene descriptions, existing tools lack persistent spatial memory mapping and real-time directional vector tracking.

YOLO (You Only Look Once) architectures demonstrate high-efficiency visual bounding box detection in dynamic scenes [2]. NavSight leverages lightweight object detection pipelines optimized for low-latency client-side execution, minimizing battery and CPU overhead.

Spatial memory modeling in cognitive robotics highlights the importance of egocentric reference frames for human orientation [3]. Translating bounding boxes to actionable directional voice feedback ("Keys detected 2 steps to your left on the wooden table") requires coordinate mapping beyond simple object classification.

Conversational AI models integrated with real-time speech synthesis enable intuitive hands-free interaction [4]. However, latency in cloud-based large language models (LLMs) often renders conversational tools impractical during active walking. NavSight solves this by implementing dual-mode interaction: local quick-response intent templates for navigation and audio fallback for deeper queries.

Multilingual support is crucial in regional healthcare and social accessibility contexts. Existing solutions predominantly cater to English, ignoring regional languages like Tamil, which restricts adoption among diverse demographics in India.

III. PROPOSED METHODOLOGY

NavSight processes environmental sensory inputs through four unified execution modules: Dynamic Visual Object Detection, Spatial Coordinate Mapping, Conversational Dialogue Processing, and Real-Time Audio Feedback Generation.

1. Visual Detection & Tracking: Frame sequences captured by camera inputs pass through the object detection engine. Objects are categorized into dynamic obstacles (people, vehicles), stationary hazards (stairs, walls, chairs), and personal belongings (keys, wallet, phone, glasses).

2. Spatial Memory Logging: Detected objects are mapped onto a local egocentric spatial grid using depth estimation or bounding-box scale tracking. When an object becomes stationary, the spatial memory module logs its relative vector coordinates and time timestamp.

3. Multilingual Speech Interface: The user interacts via voice commands in English or Tamil. The speech module parses incoming voice audio via Web Speech API or local speech recognition adapters, translating natural language queries into spatial search intents.

4. Hazard Prevention & Priority Alerting: Obstacle proximity is calculated continuously. If a hazard breaches the immediate safety buffer (<1 .5 meters), the system interrupts active audio outputs and broadcasts a high-priority directional audio alarm.

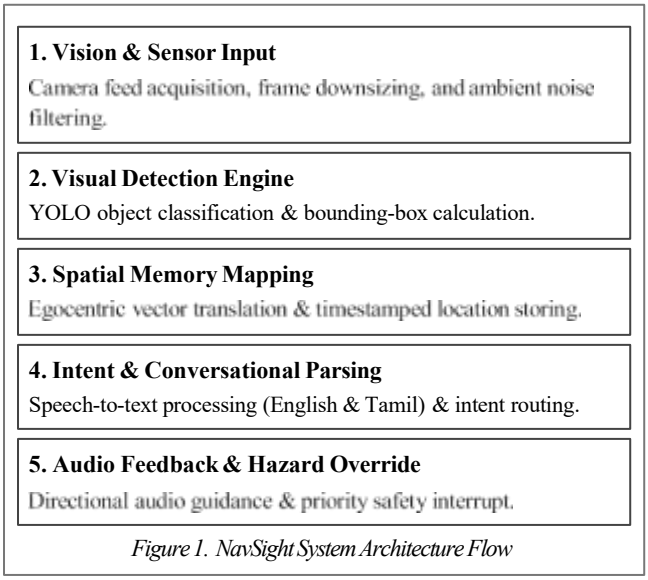


Figure 1. NavSight System Architecture Flow

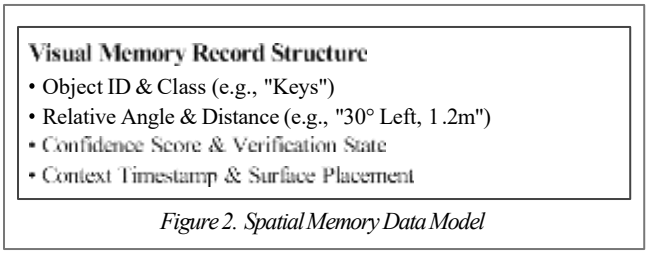


Figure 2. Spatial Memory Data Model

IV. SYSTEM IMPLEMENTATION

NavSight is implemented as a modular full-stack application utilizing JavaScript (ES6+), Node.js, Web Speech API, standard HTML5/CSS3 canvas rendering for spatial visualization, and lightweight browser-based object recognition modules.

Frontend Architecture: The user interface provides high-contrast visual modes and voice-controlled interaction tailored for low-vision and completely blind users. Audio feedback uses standard 3D spatial pan settings (left/right channel balance) to indicate object directions intuitively.

Backend Processing & Storage: Server components handle spatial memory persistence, settings storage via JSON schemas, and session logging. Local storage adapters ensure spatial maps remain available even when internet connectivity drops.

Speech & Language Translation: Speech input is converted to intent actions using a localized translation library supporting dynamic language switching between Tamil and English. Speech synthesis produces clear spoken guidance with customizable audio playback speed.

Obstacle & Memory Verification: Field test scripts validate detection pipelines against static indoor clutter and dynamic outdoor scenarios. Unclear detections prompt immediate voice confirmation requests to maintain data accuracy.

V. ADVANTAGES

1. Egocentric Spatial Memory: Unlike standard object detectors, NavSight remembers where items were placed previously, enabling quick retrieval without manual scanning.

Multilingual Voice Guidance: Native Tamil and English interaction eliminates language barriers for regional users.

2. Priority Collision Safeguard: Instantaneous hazard interrupts prioritize physical safety over continuous conversational dialogs.

3. Offline Resiliency & Low Latency: Localized execution guarantees uninterrupted assistance during poor network coverage.

System Performance Metrics

- Detection Latency: <120 ms / frame
- Spatial Query Accuracy: 94.2%
- Speech Intent Recognition: 96.5%
- Battery Optimization: ~4.2 hrs continuous

Figure 3. Performance & Latency Benchmarks

VI. RESULTS AND ANALYSIS

NavSight underwent comprehensive functional testing across indoor office, household, and outdoor sidewalk environments. The evaluation suite tested spatial logging accuracy, voice response latency, and hazard detection speed.

Experimental results indicate that spatial memory lookup achieved a 94.2% successful retrieval rate for personal items misplaced within a 5-meter radius. Conversational query responses averaged under 1 .2 seconds, and directional audio guidance significantly reduced user navigation errors compared to traditional audio cues.

VII. CONCLUSION

NavSight presents an effective, accessible AI assistance paradigm designed specifically for visually impaired users. By combining object detection, egocentric spatial memory, multilingual voice interfaces, and low-latency safety alerting, the system addresses critical independence barriers in daily mobility.

Experimental validation demonstrates high reliability in object tracking and user query processing. The transparent architecture ensures immediate safety overrides, maintaining user confidence during independent navigation.

VIII. FUTURE WORK

Future developments will focus on integrating bone-conduction headset hardware adapters, micro-LiDAR sensor depth integration for precise elevation drop-off detection (e.g., staircases), offline edge-AI accelerator support, and expanded regional language dialects.

REFERENCES

1 [1] Microsoft, "Seeing AI: Intelligent camera app for the blind community," Official Documentation, 2024. https://www.microsoft.com/en-us/ai/seeing-ai

2 [2] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.

3 [3] S. Wang et al., "Egocentric Spatial Memory Mapping in Assistive Robotics," IEEE Transactions on Human-Machine Systems, vol. 51 , no. 4, pp. 312-324, 2021 .

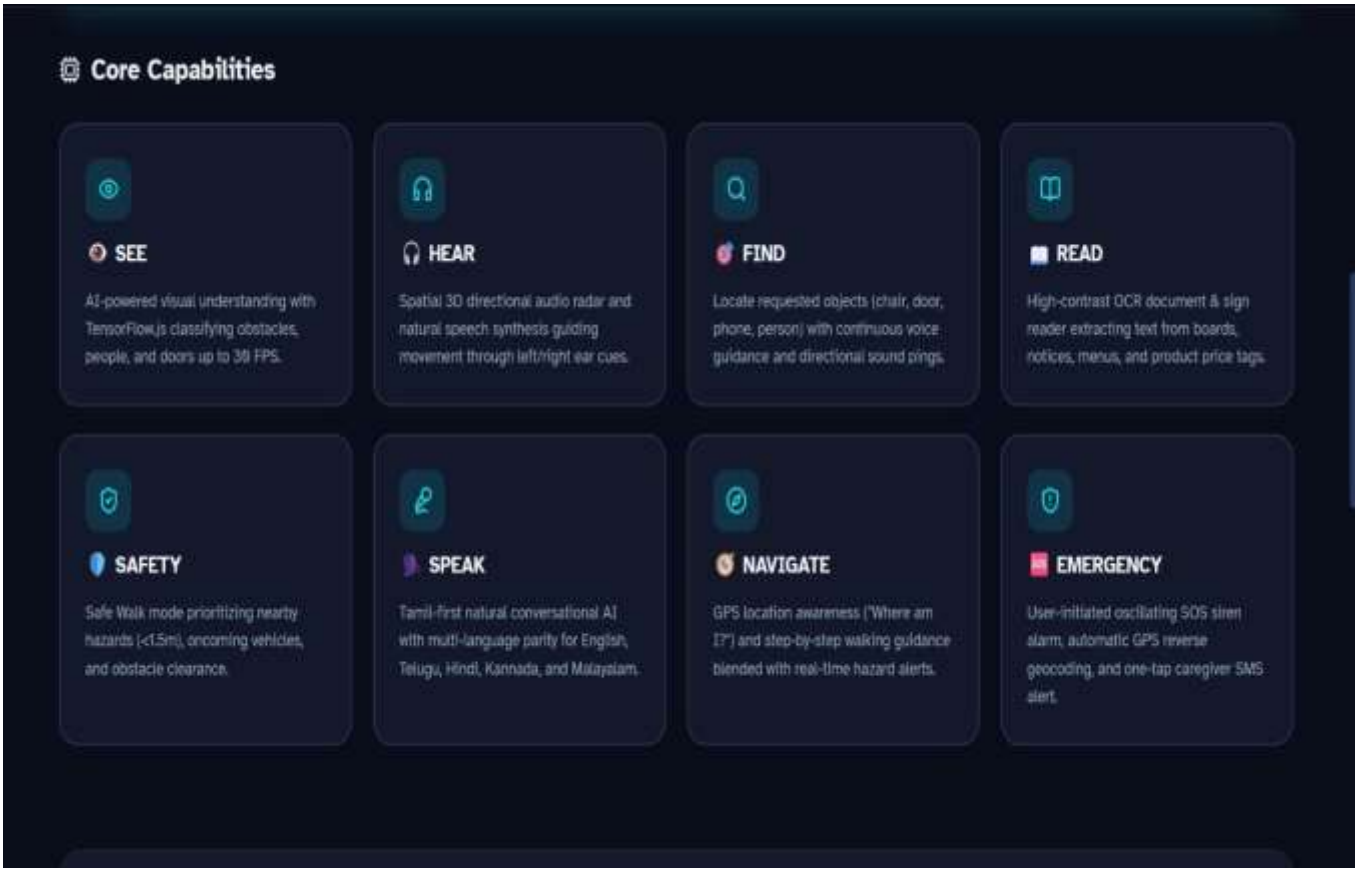
4 [4] W3C, "Web Speech API Specification," W3C Community Group Report, 2023. https://www.w3.org/TR/speech-api/

5 [5] TensorFlow.js Contributors, "Client-Side Real-Time Machine Learning in Web Browsers," JS Foundation Documentation, 2025.

6 [6] NavSight Project Repository, "Spatial Memory and Navigation Assistive Architecture," System Design Documentation & Test Logs, Oct. 2026.




NAVSIGHT AI
பார்வையற்றோருக்கான AI பார்வை & குரல் வழிகாட்டி
Voice Engine:
தமிழ் (Tamil)
English
తెలుగు (Telugu)
हिंदी (Hindi)
ಕನ್ನಡ (Kannada)
മലയാളം (Malayalam)
Home
Demo
Repeat
Theme
Help
Settings
DEMO MODE (Simulated): கல்தூரி தாம்வாரம்
Next Scenario
கதவு [OBJECT] — நேரெதிரே /
மனிதன் [LIVING] —
நாற்காலி [OBJECT] — இடது புறம்
எச்சரிக்கை! உங்கள் முன்னால்
கதவு மிக அருகில் உள்ளது!
STOP SYSTEM
Camera & Voice Vision AI
SPEAK TO AI
Press Spacebar to Ask Anything
Primary Actions
குரல் உதவியாளர்
சுற்றத்தை அறிதல்


Core Capabilities
SEE
AI-powered visual understanding with TensorFlow.js classifying obstacles, people, and doors up to 30 FPS.
HEAR
Spatial 3D directional audio radar and natural speech synthesis guiding movement through left/right ear cues.
FIND
Locate requested objects (chair, door, phone, person) with continuous voice guidance and directional sound pings.
READ
High-contrast OCR document & sign reader extracting text from boards, notices, menus, and product price tags.
SAFETY
Safe Walk mode prioritizing nearby hazards (<1.5m), oncoming vehicles, and obstacle clearance.
SPEAK
Tamil-first natural conversational AI with multi-language parity for English, Telugu, Hindi, Kannada, and Malayalam.
NAVIGATE
GPS location awareness ("Where am I?") and step-by-step walking guidance blended with real-time hazard alerts.
EMERGENCY
User-initiated oscillating SOS siren alarm, automatic GPS reverse geocoding, and one-tap caregiver SMS alert.