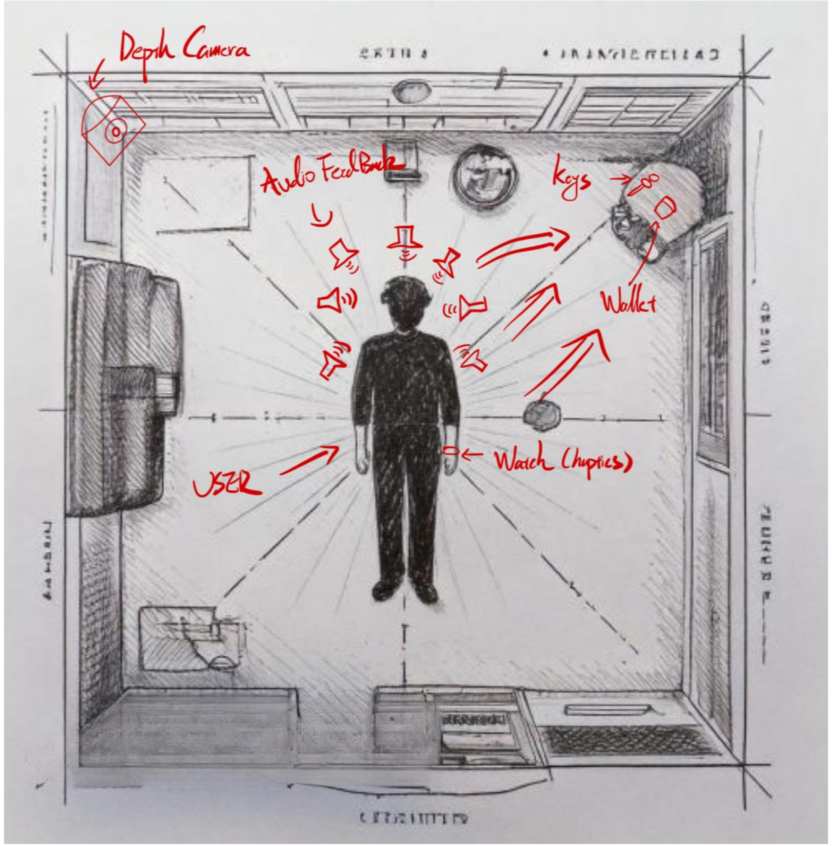
Visually Impaired Accessible Room

Alex Hafemeister, Arshia Dariban, Dhruv Sirohi, Ethan Jiang

System Overview



Motivation & Objectives

- Visually impaired people face highly uncertain environments when traveling.
- A smart room which can be installed quickly and easily in existing public establishments.
- Seamless guidance for the user in a new environment using haptic and audio feedback.

Relatively simple layout, but many moving parts.

Short term stay means user does not have time to acclimate.

Technical Approach and Novelty

Current Methods and Limitations:

Head-Mounted Depth Cameras:

- Used for mapping and object detection
- Limitations: Uncomfortable for prolonged use, lacks intuitive feedback

Haptic Feedback in Object Retrieval:

- Guides precise object grasping
- Limitations: Minimal integration with real-time spatial navigation

Speech Recognition for Navigation:

- Enables hands-free interaction
- Limitations: Prone to errors in noisy environments

Moving Object Detection:

- Uses RGB-D data to track moving objects

Room Scanning and 3D Object Detection:

- Integrates RGB and depth data for precise 3D mapping
- Limitations: High computational requirements

Limitations of Current Practice:

- Single-modality feedback (visual, audio, or haptic alone)
- Head-mounted devices are often inconvenient
- Speech recognition affected by background noise

Our Approach and Novelty:

Stationary Depth Camera (RealSense L515):

- Mounted at an optimal height for stable, real-time room mapping

Multi-Modal Feedback Integration:

- Synchronizes haptic (Apple Watch) and auditory cues for intuitive guidance

Hands-Free Speech Commands:

- Uses Whisper OpenAI for robust speech recognition and context-aware instructions

What’s New:

Combines stationary depth sensing, multimodal feedback, and natural language processing for an intuitive experience that enhances user independence in object retrieval and navigation tasks.

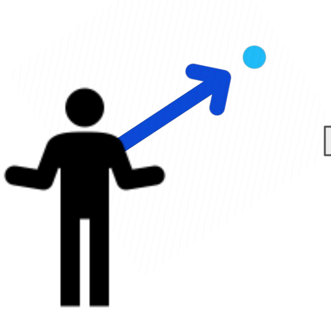
Methods

User request:
I need coffee,
find my keys etc.

User pose detected, vector to destination

Haptic directional guidance via apple watch

Audio guidance via airpods



Speech to text to capture object of interest

Server uses camera data to return vector to the app

Evaluation and Metrics

Response Time: Speed from user request to successful object retrieval.

Accuracy: Rate of correctly identified and located objects.Track the frequency of incorrect object detections or misdirections, especially in environments with multiple similar objects or obstacles.

Ease of Use: Measure how intuitive the system is for visually impaired users, based on the time taken to learn and use the device effectively.

Adaptability: Measures how well the system adjusts to different environmental conditions (e.g., cluttered spaces, varying light levels, different room layouts). This metric assesses the system's flexibility and effectiveness in real-world scenarios beyond controlled environments.

User Satisfaction: Feedback from visually impaired users testing the prototype.

Reliability: System consistency across different environments and conditions.

Current Status and Next Steps

Current status

- Haptic feedback
- Depth map from camera
- Hand tracking and object distance detection
- Body pose landmarks and face mesh
- Face pose computation relative to camera
- Lightweight object detection

Next steps

- Speech to text recognition and command interpretation
- Spatial audio feedback
- Pose and object distance actuation feedback computation
- Actuation command communication to watch + earphones