NTU Binary Hackathon 2026

Multimodal Assistive System for Hearing & Visually Impairment Individuals

VAB (Visual Audio Buddy)

By Team Trinary

1. Background & Problem Statement

1.1 Background

Technology has significantly transformed how people communicate, learn, and navigate the world. However, its potential to support persons with disabilities, particularly those with sensory impairments, remains underutilised. Most assistive technologies today are designed around a single mode of interaction, such as speech, text, or vision. This design assumption excludes users who cannot reliably access that modality.

Research by the **World Health Organization (WHO)** and the **World Wide Web Consortium (W3C)** consistently shows that single-modality information delivery is a primary cause of digital and environmental exclusion for people with sensory impairments. This limitation is especially severe for individuals with dual sensory loss (deaf-blindness), whose daily independence depends on accessing contextual and environmental information through alternative and redundant channels.

1.2 Prevalence & Scale of the Problem

Sensory impairment in Singapore is significantly under-represented by formal disability registration figures, largely due to its strong association with aging. The following evidence-based statistics demonstrate the scale of the challenge:

Hearing Impairment

- An estimated **500,000 people in Singapore** live with some degree of hearing loss[1]
- Approximately **50% of individuals aged 60+** experience hearing loss, rising to **up to 95% among those in their 80s**[2]
- For births, approximately **0.1% of newborns** have severe or profound hearing loss, and around **0.5%** have milder hearing loss detectable at birth[3]

Vision Impairment

- Over **40,000 individuals** in Singapore are living with blindness or vision impairment, with this number expected to rise sharply as the population ages[4]
- Among adults aged 60 and above, approximately **180,000 people** experience some form of visual impairment[5]

---

[1] Hearing Partners Singapore. (2024). *Hearing loss in Singapore*. https://www.hearingpartners.com.sg/hearing-loss-singapore/

[2] Singapore Association for the Deaf (SADeaf). (2024). *About deafness: Introduction*. https://sadeaf.org.sg/about-deafness/introduction/

[3] Singapore Association for the Deaf (SADeaf). (2024). *About deafness: Introduction*. https://sadeaf.org.sg/about-deafness/introduction/

[4] Guide Dogs Singapore Ltd. (2024). *White Cane Day 2024*. https://guidedogs.org.sg/whitecane2024/

[5] Guide Dogs Singapore Ltd. (2024). *White Cane Day 2024*. https://guidedogs.org.sg/whitecane2024/

- Population-based studies estimate that about **0.2%–0.6% of adults** have bilateral blindness, with a larger percentage experiencing visual impairment or low vision[6]

Dual Sensory Loss (Deaf-Blindness / Multiple Sensory Impairment)

- While less visible, dual sensory loss is increasingly common among older adults
- Studies indicate that **up to 1 in 5 older adults in Singapore** may experience multiple sensory impairments (MSI)[7]

Key Insight: These figures demonstrate that combined hearing and vision loss is not a niche condition, but a growing accessibility challenge in Singapore's aging society. Many individuals with dual sensory loss do not identify as fully deaf or blind, yet still face significant barriers when tools and environments rely on a single sensory channel.

1.3 Core Problem Statement

In everyday public environments, such as convenience stores, clinics, transport hubs, and public buildings, critical information is primarily conveyed visually or audibly. Existing assistive technologies tend to be:

- **Single-modality:** Assuming the availability of either sight or hearing
- **Fragmented:** Requiring users to switch between multiple tools
- **Poorly adapted:** Not optimized for fast-paced, real-world environments

For sensory impaired individuals, these limitations make it difficult to independently understand their immediate surroundings, identify objects, access written information, or navigate safely. As a result, many rely on caregivers or assistance from others for routine daily activities, reducing independence, dignity, and confidence.

1.4 Research Question

How can we help deaf and blind individuals independently understand their immediate surroundings and access essential information in everyday public environments through a multimodal assistive system?

[6] Chua, J., Baskaran, M., Ong, P. G., Zheng, Y., Wong, T. Y., Aung, T., & Cheng, C.-Y. (2018). Prevalence, risk factors, and visual features of undiagnosed glaucoma: The Singapore Epidemiology of Eye Diseases Study. *Scientific Reports, 8*, 11847. https://www.nature.com/articles/s41598-018-30004-9

[7] Gupta, P., Chan, A., Man, R. E. K., Fenwick, E. K., Aravindhan, A., Ng, J. H., Cheng, C.-Y., & Sabanayagam, C. (2024). *Prevalence, associated risk factors, and patient and economic impact of multiple sensory impairment in a multi-ethnic elderly population in Singapore: the PIONEER study.* BMC Public Health, 24, 1102. https://link.springer.com/article/10.1186/s12889-024-18635-2?utm_source=chatgpt.com

2. Context: Singapore

Singapore has made notable progress in disability inclusion through agencies such as SG Enable and community organizations supporting persons with hearing and visual impairments. Macro-level accessibility, such as barrier-free transport and mobility training, has improved substantially.

Key Initiatives:

- **SG Enable:** Government agency focused on disability support and inclusion
- **Disabled People's Association:** Represents the community through policy discussions, inclusion training, and public outreach
- **SADeaf & Guide Dogs Singapore:** Partner organizations providing mobility training and support

However, sensory impaired individuals continue to face micro-level accessibility barriers in dense, fast-paced urban settings. Tasks such as identifying nearby objects, reading labels, interpreting spatial layouts, or understanding contextual cues remain difficult without reliable visual or auditory input. Currently, no single, integrated solution addresses these challenges holistically in everyday public environments.

3. Target Audience & User Analysis

3.1 Target Users

Primary Target Users: Sensory Impaired individuals. These includes:

- Individuals with hearing impairment
- Individuals with visual impairment

Secondary Users: Deaf-blind individuals

- People with dual sensory loss (combined hearing and visual impairments), particularly older adults.

These secondary users face similar barriers when information is delivered through a single modality and may benefit from the same multimodal design principles.

3.2 Shared User Traits (Evidence-Informed)

Research across disability studies, inclusive design, and human-computer interaction identifies several shared traits among users with hearing and/or visual impairments:

1. **Reliance on non-primary sensory channels** to access information
2. **Difficulty accessing environmental and contextual information,** beyond simple text or speech
3. **High cognitive load** caused by fragmented assistive tools
4. **Reduced independence** in noisy, crowded, and time-pressured public environments

These traits indicate that effective assistive solutions must be redundant, adaptable, and multimodal by design, rather than optimized for a single impairment.

3.3 Key User Pain Points

Based on user research and disability studies literature, we have identified five critical pain points:

Pain Point	Description	Real-World Example
1. Environmental & Contextual Awareness	Users struggle to understand what is around them, what objects they are holding, or what information is being presented in their environment	Cannot tell if someone is approaching, what product is on the shelf, or what signage says
2. Basic Daily Tasks	Tasks such as reading signboards, identifying products, checking ingredient lists, or viewing the time often require assistance due to inaccessible formats	Shopping: Cannot differentiate between similar drink flavors in identical packaging, cannot read ingredient lists to check for allergens

3. Communication Barriers	Users face difficulties conversing with strangers or service staff unfamiliar with sensory impairments, leading to missed or misunderstood information	Asking a salesperson for help but struggling to communicate needs; missing emotional cues (sarcasm, frustration) in responses
4. Micro-Navigation ("Last 10 Metres")	Existing navigation tools address macro-navigation but fail at the 'last few metres,' where obstacle detection and spatial awareness are critical	GPS gets user to the building, but they cannot navigate inside to find the correct counter or avoid obstacles in the hallway
5. Cognitive Overload	Switching between multiple single-purpose assistive tools increases mental effort and reduces usability	Need separate apps for reading text, identifying objects, navigation, and communication—all while trying to accomplish a simple task like buying groceries

4. Current Solutions & Limitations

While various assistive technologies exist, each addresses only specific aspects of the challenges faced by sensory impaired individuals. The following analysis examines existing solutions and their critical limitations:

Category	Examples	Main Modality	What It Addresses	Critical Limitations
Traditional Mobility Aids	White cane, guide dogs	Tactile, canine assistance	Basic obstacle detection & wayfinding	No information about surroundings, no text reading, no communication support
Smart Canes	WeWALK Smart Cane 2	Haptic + audio	Enhanced obstacle detection, some navigation	Relies on audio (unusable for sensory impaired), limited to navigation only
Wearable Haptic Bands	Sunu Band, Wayband	Haptic only	Obstacle detection, direction cues	Limited range, no information access, requires phone pairing
Wayfinding Apps	NaviLens, accessible maps	Visual + audio	Outdoor/indoor route guidance	IDK send help

Communication Devices	Braille displays, captioning tools	Tactile (braille), sometimes audio	Communication, alerts, information	Expensive, requires training, separate devices for each function

Critical Gap: No existing solution integrates navigation, information access, and communication into a single, multimodal platform that works without relying on vision or hearing.

5. Proposed Solution

5.1 Solution Overview

We propose a multimodal assistive application designed to help sensory impaired users independently understand their surroundings and access essential information in public spaces. The system integrates multiple interaction modalities into a single, coherent experience, reducing reliance on fragmented tools and caregivers.

Core Focus Areas:

- **Environmental awareness** — understanding immediate surroundings
- **Contextual understanding** — accessing written information and object identification
- **Low-effort interaction** — minimizing cognitive load and learning curve

Important: This solution complements rather than replaces existing assistive devices such as braille displays, guide dogs, or interpreters.

5.2 Key Features & How They Address User Needs

Feature	Pain Point Addressed	How It Works
Speech-to-Text	Communication Barriers	Transcription for spoken communication into text for hearing impaired individuals
Object & Text Recognition (using Optical Character Recognition & Text-to-Speech) + Real-time AI alignment feedback	Environmental Awareness & Daily Tasks	**Optical Character Recognition (OCR):** Uses computer vision to identify and differentiate similar items (drinks, produce), read ingredient lists, detect allergens, and recognise objects in the environment. **Text-to-Speech (TTS):** Converts written text (signboards, labels, screen content) into audio output for users with visual impairment **Real-time AI alignment feedback:** Audio guidance in helping the visually impaired individuals position the item correctly. Examples include: **1) Target not detected** → Cues like "Item not detected", "please aim accurately at item" **3) Text readable** (confidence threshold met) → "Item Captured" **Future extension:** Include haptic cues

Micro-navigation assistance	Environmental Awareness	Inspired by NaviLens' focus on last-metre accessibility, we provide marker-free environmental awareness using real-time object detection, and hazards with haptic or audio feedback
Sound Event Detection	Environmental Awareness	Detects important sound events such as doorbells and alarms. When a doorbell is detected, the phone vibrates and sends a notification to inform the user that someone is at the door. For loud alarms (e.g., fire alarms), the flashlight blinks continuously to provide a clear visual alert.

5.3 Extension Features

Feature	How It Works
Integrate Emotion Detection for STT feature	Detects emotional cues (tone, sarcasm) to provide fuller context during conversations.
Voice Recognition/Input	Allows hands-free, voice-activated interaction (e.g., 'What time is it?', 'Read this sign'), reducing physical and cognitive effort
Braille Keyboard Integration	Supports integration with external and built-in braille keyboards, enabling seamless input for users who are blind or have low vision. This allows users to type, navigate, and interact with the app efficiently using their preferred assistive technology. By ensuring compatibility with system-level accessibility tools (e.g., iOS and Android braille input support), the app maintains a fully inclusive and multimodal user experience.

5.4 Design Principles (Accessibility-by-Design)

Our solution is built on evidence-based, user-centered design principles that prioritize accessibility from the ground up:

1. **Mode-Based Simplification:** The app dynamically adapts its interface after profile selection (Blind / Deaf / Deaf-blind individuals), ensuring users only see features relevant to their abilities. This reduces interface clutter and prevents cognitive overload.

2. **Structured, Tactile Interaction:** Tap-based profile selection and confirmation provide a consistent and recoverable interaction pattern. Long-press reset ensures users can safely return to the main page without complex navigation.
3. **Multimodal by Necessity:** Each mode activates only the appropriate sensory outputs:
    a. Blind Mode → Audio guidance + haptic feedback
    b. Deaf Mode → Visual transcript display
4. **Minimal Interaction Friction:** Blind users do not need to press shutter buttons, live OCR scanning with AI feedback enables hands-free alignment.
5. Hearing-impaired users control speech-to-text recording clearly through start/stop interaction, allowing structured conversation flow.
6. **Low Cognitive Load:** Large buttons, simple gesture logic, and clearly segmented workflows ensure that users do not need to navigate complex menus or remember multi-step processes.
7. **Elderly-Friendly Design:** High-contrast visuals, large fonts, bilingual support (English & Chinese), and uncluttered screens ensure accessibility for older adults experiencing progressive sensory decline.

6. Technical Architecture

6.1 System Overview

The application is implemented as a mobile-first platform with the following architecture:

Input Modalities:

- Camera: rear and front-facing for object recognition, text reading, obstacle detection
- Microphone: voice commands, speech-to-text)
- Touch interface: simple gestures, fallback interaction

Output Modalities:

- Audio feedback: Text-to-Speech (TTS) for users with visual impairment
- Haptic vibrations: patterns for alerts, confirmations, directional cues
- Text display: large, high-contrast text for users with partial vision

6.2 Core Technologies

- **Computer Vision (CV):** Object detection, text recognition (OCR), scene understanding
- **Natural Language Processing (NLP):** Speech-to-text, text-to-speech, emotion detection from tone and context
- **Machine Learning (ML):** Product identification, ingredient scanning, personalized user preferences
- **Sensor Fusion:** Combining camera, microphone, and accelerometer data for context-aware assistance

6.3 Technology Stack (Implementation)

To ensure feasibility within the hackathon scope while maintaining extensibility, the system leverages **cross-platform development and cloud-based AI services**, prioritising reliability and rapid prototyping over custom model training.

Frontend (Mobile Application)
- **Flutter (Dart)**
    - Cross-platform development for Android and iOS
    - Strong support for accessibility features and rapid UI iteration
- **Key Flutter Plugins**
    - `camera` – real-time image capture
    - `speech_to_text` – voice command input
    - `flutter_tts` – audio output via text-to-speech
    - `permission_handler` – runtime permission management

Backend & Cloud Infrastructure
- **Firebase**
    - Firebase Authentication (optional): lightweight user management
    - Cloud Functions: backend logic and API orchestration

○　Firestore: storing user preferences (e.g. output mode, verbosity level)

This lightweight backend architecture reduces infrastructure complexity while allowing the system to scale conceptually.

AI & Multimodal Services
- **Computer Vision: Google ML Kit / Google Cloud Vision API**
  - Optical Character Recognition (OCR)
  - Basic object detection and classification
- **Speech & Language**
  - **Google Speech-to-Text API:** Transcription of spoken language in noisy public environments
  - **Google Text-to-Speech API:** Natural-sounding audio output with adjustable speech rate

These services were selected for their robustness, multilingual support, and ability to handle real-world conditions such as accents and background noise.

6.4 Design Rationale

The chosen technology stack reflects the project's guiding principles:

- **Multimodal by necessity**: Each technology supports a specific sensory channel, ensuring redundancy rather than reliance on a single mode.
- **Hackathon-feasible**: Cloud-based AI services allow the team to focus on user experience and system integration rather than model training.
- **Extensible**: The architecture supports future enhancements such as on-device inference, advanced haptics, or personalised ML models.
- **User-centred**: Technical decisions prioritise low latency, reliability, and accessibility over experimental complexity.

6.5 Scope & Limitations (Technical)

For the purposes of this hackathon:

- Advanced real-time obstacle avoidance and full scene understanding are **out of scope**
- Emotion detection is limited to basic inference and treated as a **future enhancement**
- The system focuses on demonstrating **core multimodal interaction** rather than exhaustive coverage of all assistive scenarios

7. Usability & Accessibility Considerations

7.1 Universal Design Principles

- **Sensory redundancy:** Every critical interaction provides feedback through multiple channels (audio + haptic, or visual + haptic)
- **Motor accessibility:** Large touch targets, voice control, minimal required gestures
- **Cognitive accessibility:** Consistent interaction patterns, clear feedback, progressive disclosure of complexity

7.2 Intuitive User Experience

- **Low learning curve:** Natural voice commands ("What is this?", "Read the label"), no complex menus
- **Adaptability:** Users can customize output preferences (audio volume, haptic intensity, speech rate)
- **Error prevention:** Confirmations before critical actions, clear cancel options

7.3 Avoiding New Accessibility Barriers

We are committed to not introducing new barriers:

- No reliance on color alone for information (use patterns, labels, haptics)
- No time-dependent interactions that cannot be paused or extended
- No audio-only critical alerts (always paired with haptic or visual)

8. Scope & Limitations

8.1 Current Technical Limitations

- **Requires smartphone:** Users must have access to a modern smartphone with camera and microphone
- **Internet connectivity:** Some features (advanced object recognition, emotion detection) require cloud processing
- **Lighting conditions:** Camera-based features perform better in well-lit environments
- **Language support:** Initial version supports English; multilingual support (Mandarin, Malay, Tamil) in future iterations

8.2 Privacy & Data Considerations

- All image and audio processing occurs on-device where possible to protect privacy
- Cloud processing (when required) uses encrypted connections and does not store personal data
- Users have full control over what data is shared and can opt out of cloud features

9. Future Work & Extensibility

9.1 Phase 1 – Minimum Viable Product (MVP)

Focus: Core multimodal communication & accessibility functions
Goal: Validate primary user needs and ensure reliability

Key Deliverables:

- **OCR with AI Alignment Feedback (Blind / Low Vision Mode)**
  - Live scanning
  - Real-time alignment guidance (audio cues)
  - Auto-capture based on OCR confidence
- **Speech-to-Text (STT) for Hearing Assist Mode**
  - Real-time transcription
  - Controlled start/stop interaction
- **Text-to-Speech (TTS) Output**
  - Supports two-way communication
  - Enables Deaf/Mute users to respond verbally through device

This phase validates the core value proposition: Reducing interaction friction through multimodal support.

9.2 Phase 2 – Environmental Awareness Expansion

Focus: Micro-navigation and real-world safety
Goal: Enhance independence beyond transactional interactions

Planned Features:

- **Micro-Navigation Assistance**
  - Real-time object detection (people, doors, obstacles)
  - Audio + haptic alerts
  - Inspired by "last 10 metres" accessibility principles
- **Sound Event Detection**
  - Detect critical environmental sounds (alarms, loud noises, name calling)
  - Visual + vibration alerts for Deaf users

This phase extends functionality from communication support to spatial awareness.

9.3 Phase 3 – Intelligent & Adaptive Extensions

Focus: Personalization, intelligence, and deeper multimodal integration
Goal: Increase long-term usability and contextual adaptation

Planned Features:

- **Emotion Detection for STT**
  - Basic tone classification (calm / urgent / neutral)

- ○ Contextual cues to improve conversational understanding
- ● **Voice Recognition & Hands-Free Commands**
  - ○ Activate scanning or conversation mode via voice
  - ○ Reduce dependency on touch interaction
- ● **AI-Powered Personalization**
  - ○ Learn user preferences (speech speed, font size, feedback type)
  - ○ Adapt output modalities dynamically

This phase transitions the system from reactive to adaptive.

9.4 Phase 4 – Ecosystem Integration & Scalability (Long-Term Vision)

Focus: Infrastructure-level integration and broader adoption

- ● **Wearable Device Integration**
  - ○ Smartwatch notifications
  - ○ Haptic bands for discreet feedback
- ● **Public Infrastructure Partnerships**
  - ○ Integration with public transit systems
  - ○ Retail accessibility enhancements
- ● **AR / Spatial Computing Integration**
  - ○ Explore spatial mapping for enhanced environmental interpretation
- ● **Community Accessibility Contributions**
  - ○ User-reported accessibility insights
  - ○ Shared public environment information

10. Conclusion

As Singapore's population ages, dual sensory loss will become increasingly common, making single-modality accessibility insufficient. By adopting a multimodal, user-centered approach, this project demonstrates how assistive technology can restore access to environmental and contextual information, empowering sensory impaired individuals with greater independence, safety, and dignity in everyday life.

Our solution is designed from the ground up with accessibility principles, ensuring that it is genuinely usable by people with disabilities. We do not seek to replace existing assistive tools, but to fill critical gaps that no single solution currently addresses—creating a unified, multimodal platform that reduces cognitive load and restores independence.

Through careful attention to universal design, low learning curves, and dignity-preserving interactions, we aim to set a new standard for inclusive assistive technology in Singapore and beyond.