University of San Agustin
General Luna St., 5000 Iloilo City, Philippines
www.usa.edu.ph

COLLEGE OF TECHNOLOGY

World Navigation Hat - Development of a Wearable Navigation Aid using AIoT for the Visually Impaired

An Undergraduate Design Project Presented to Faulty of the Computer Engineering Department of College of Technology University of San Agustine

In Partial Fulfillment of the Requirement of the Course

CPE 413 - CpE Practice and Design I

By:

D'Souza, Jason C.

Pabito, Ethel Herna C.

Wang, ChenLin

Daywan, Vince Ginno B.

Submited to:

Engr. Glenda S. Quanzon

November 11, 2025

Table of Contents

List of Figures

List of Tables

Introduction

Background of the study

Visual impairment is a global health problem that significantly affects individual's daily lives by impeding independent navigation, social interaction, and overall quality of life (Theodorou et al., 2023). There is an estimation of 2.2 billion people globally who are identified as visually impaired and this number could still increase to 2.5 billion by 2050 as stated by the World Health Organization (World Health Organization, 2023). In the Philippines, 2.17 million Filipinos are identified as visually impaired as quantized by reports from the Philippines Eye Research Institute (PERI) and Department of Health (DOH) (Shinagawa Lasik & Aesthetics, 2025). Visual impairment does not pertain to total blindness. According to World Health Organization, 2023, visual impairment can be identified and categorized based on the presenting visual acuity

(1) Mild Vision Impairment, visual acuity is better than 6/18

(2) Moderate Vision Impairment, visual acuity is worse than 6/18 but better than 6/60

(3) Severe Vision impairment, visual acuity is worse than 6/60 but better than 3/60

(4) Blindness, visual acuity is worse than 3/60

  (4.a) Blindness with light perception, individuals can only perceive light

  (4.b) Total Blindness, individuals who have no light perception

Other common types include astigmatism, near-sightedness (myopia) and far-sightedness (hyperopia), which account for a large share of impairments, alongside conditions like cataracts, glaucoma, and age-related macular degeneration (AMD), with AMD alone affecting 8.06 million people globally in 2021 .

cite

Navigation for visually impaired individuals increasingly relies on assistive devices that utilize sensory substitution, converting visual information into auditory or tactile cues to enhance spatial awareness (Skulimowski, 2025). These devices often incorporate IoT sensors to capture real-time environmental data, providing navigational assistance in complex environments (Real & Araújo, 2023) (Mohamadi et al., 2024). In addition, visually impaired individuals also face safety concerns and challenges such as accidents, falls, and collisions, as well as difficulties with navigation, including road crossings and destination location (Ikram et al., 2024; Gao et al., 2025; Muhsin et al., 2023). Beyond these practical difficulties, social isolation and reduced access to information further compound the challenges, often limiting educational and employment opportunities (Arvind, 2023). Current existing solutions or aid for visually impaired individuals are guide dogs, white canes, and electronic travel aids (Muhsin et al., 2023).

Sensory substitution devices (SSDs) allow users to perceive their environment by converting sensory information from one modality to another, particularly aiding those with visual impairments (Mishra et al., 2025). These devices help with obstacle detection, navigation, and object recognition, enhancing independence for users relying on traditional aids (Tokmurziyev et al., 2025). However, global adoption is limited by issues such as cognitive overload, extensive training needs, ergonomic discomfort, and the lower processing bandwidth of non-visual senses compared to vision (Hou et al., 2025).

Rationale

Visual Impairment encompasses a spectrum of conditions, from moderate vision loss to complete blindness, which significantly impacts individual's ability to perceive their environment and perform daily activities (Kumar et al., 2025). This condition also affects the individual's social interaction and overall quality of life (Que et al., 2025). More than 2.17 million people in the Philippines live with visual impairment, which restricts their perception, mobility, navigation, and accessibility to information—depriving them of work opportunities and independence (Chavarria et al., 2025). To mitigate the challenges encountered by the VIP, assistive technologies and rehabilitation strategies have been developed (Skulimowski, 2025). Traditional assistive aids like white canes offer limited feedback, initiating the development of electronic travel aids that deliver richer environmental information and improve autonomous navigation (Chandra et al., 2025) (Kim, 2024). Another solution is a wearable SSD, an assistive technology that an individual can wear on their body to help users navigate both their physical surroundings and digital applications, and emerging as a potential avenue to restore or gain a sense of spatial perception or awareness and navigational ability (Skulimowski, 2025). Additionally, SSDs promotes social interaction that aligns with one of the United Nations' Sustainable Development Goals for health and equality (Xue et al., 2025).

In previous studies, SSDs have been limited in their applicability, as existing devices often perform poorly in real-world situations, encountering cognitive overload problems that overwhelm users with excessive information, making navigation difficult (Casanova et al., 2025). Another identified concern is that the SSD design is uncomfortable, as it is bulky and requires extensive training, which limits its practicality (Olaosun et al., 2024). In addition, many of these devices lack digital integration as capabilities, focusing mainly on obstacle detection and not connecting well with digital tools (Makati et al., 2024). These deficiencies underscore the pressing need for higher SSDs that can outperform and enhance existing devices. Therefore, a need for development of optimized smart navigation device improves the functional efficiency and performance of SSDs without overlooking the previous concerns like overloading data processing.

In this study the researchers aim to address these matters through the development of a wearable navigation hat integrated with AIoT, which introduces a smart navigation hat that uses 3D point cloud data to give audio feedback and connects to IoT services for navigation in both physical and digital environments. Furthermore, it enhances sensory processing, focusing on reducing cognitive overload in the system and making navigation simpler for VIP users. Additionally, the hat features customizable options, including a modular operating system that allows users to choose between gesture or voice commands.

General Objective

To develop an AIoT visual sensory substitution hat device that converts real-time environment information through emulating human sensory processes into audio cues for the visually impaired people (VIP)

Specific Objectives

(1) To develop the main pipeline to represent environmental information to audio cues

- (1.a) Working photogrammetry (2D to 3D) system from the cameras
- (1.b) Uses help from the IMU to create a virtual environment
- (1.c) Toggable modes to prioritize what information is cued into audio

(1.d) Allows insertion anywhere in the pipeline for more custom preferences

(2) To develop the device to be portable and comfortable

(2.a) Uses battery instead of needing direct connection to power

(2.b) Can be charged through UPS instead of disposable batteries

(2.c) All packed into a hat to avoid putting strain on sensitive areas like eyes, nose, and ears

(3) To develop an interface that is not visually dependent nor over stimulating

(3.a) Command and instructions are derived from hand guestures through hand recognition

(3.b) The commands and instructions are easy to understand and navigate

(3.c) The audio output is controlled minimizing unessary information

(4) To test if users develop unsubconscious visual senses of the environment

(4.a) They can navigate indoor through rooms/doors/furnitures

(4.b) They can detect movement and even catch trown objects

Significance of the study

The proposed AIoT navigation hat has the potential to escalate the performance of existing assistive devices by:

(1) Enhancing independence and mobility

(2) Improving cognitive comfort and ease of use

(3) Bridging physical and digital accessibility participation

(4) Promoting Social Inclusion and Employment Opportunities for the VIP

In addition, this system innovates the research for Assistive Technology in the Philippines.

(1) VIP can navigate safely and confidently with less assistance

(2) The system simplifies understanding spatial information, reducing mental strain (cognitive overload) and allowing quick interpretation without long training

(3) With AIoT integration, users can access navigation apps and online services hands-free using voice or gesture commands

(4) As VIP gain independence, they can engage more in education, work, and community activities, boosting their self-esteem and reducing reliance on others

The project's success may inspire further studies on brain-inspired computing and wearable assistive devices, paving the way for more innovative tools. This project device will specifically benefit to the following:

(1) **Visually Impaired and Blind Individuals** - They are the primary beneficiaries, gaining increased mobility, safety, and access to digital and social spaces through sensory substitution and IoT assistance.

(2) **Families and Caregivers** - With the device promoting user independence, caregivers will experience reduced physical and emotional strain while maintaining peace of mind about the user's safety.

(3) **Educators and Accessibility Advocates** - Teachers and advocates working with visually impaired students can integrate the technology into learning environments to enhance inclusivity and accessibility. Such as allowing them to access resources and infrastructure that did not take the visually impaired into account (Sites, Media, etc.)

(4) **Medical and Rehabilitation Specialists** - Eye health professionals and rehabilitation centers can use the system as a tool for sensory training and mobility rehabilitation programs, especially in fill in aspects the patent lacks or have trouble with.

(5) **Researchers and Engineers in Assistive Technology** - The project offers a novel framework that combines virtual world modeling, IoT, and sensory emulation, providing a foundation for future innovations in human-computer interaction and embedded systems.

(6) **Government and NGOs for Disability Support** - Organizations involved in disability welfare can adopt or fund similar low-cost solutions to support national accessibility programs and meet inclusive development goals.

Conceptual Framework

Figure 1: IPO Model of the Conceptual Frame of the Study

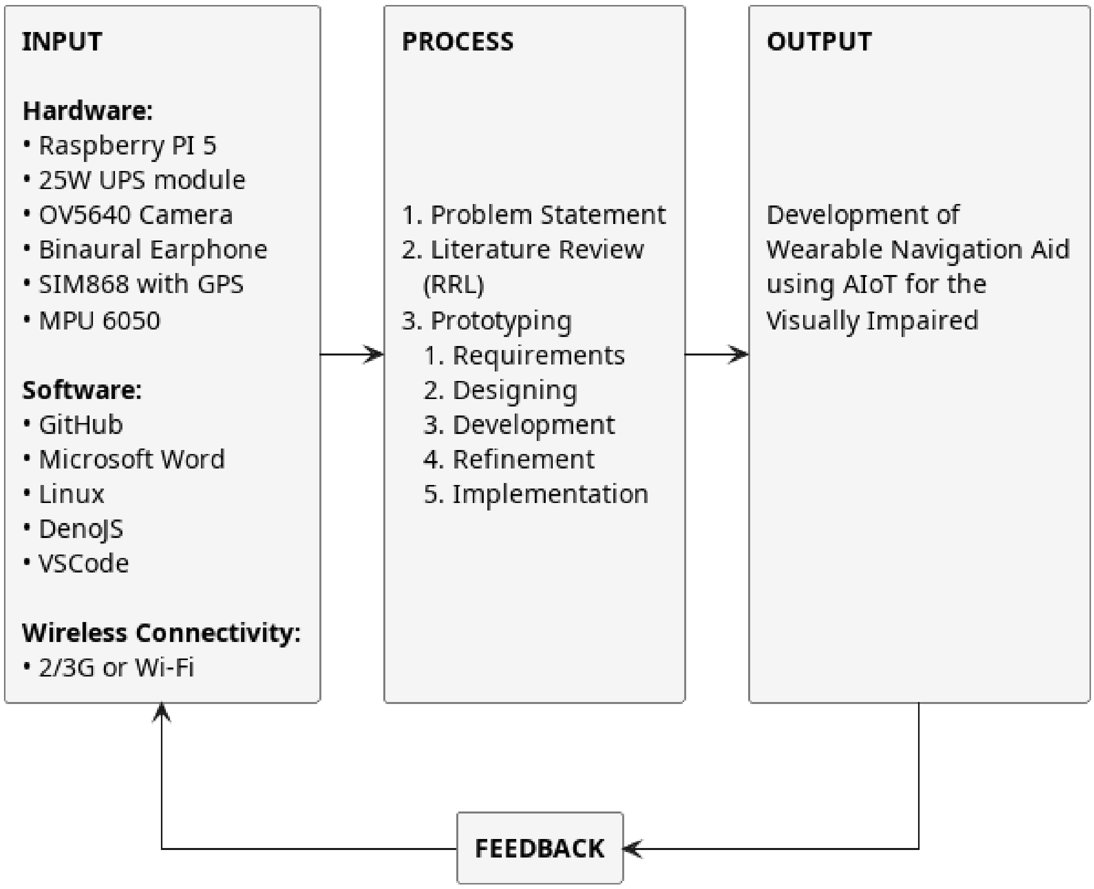


Figure 1 states the conceptual framework of our research in a form of an Input-Process-Output (IPO) model. We start with three kinds of required input such has Hardware for our Microcontroller, Sensors, Actuators, and SIM Module, Software for our Project management, Client-side device, and Server-side device, and the Wireless Connectivity used in our device which can be either 2/3G which is used on deployment doors, or Wi-Fi which is used deployment indoors or development. Our method of research mimics typical development cycle with addition to literature review and research, such has literature review, prototyping, designing, etc. By the end of each iterative development we are to expect to be closer to the end goal, with a feedback from either advisor or mentor to start the cycle again until our end goal of an wearable navigation aid is developed.

Theoretical Framework

This research attempts to solve an issue facing most Visual Sensory Substitution Devices which is primarily about the issue on translating a high bandwidth visual information to low bandwidth audio information without causing overstimulation, our novel solution is the emulate the human sensory mechanism to carry the processing load from the user and ensure that substituted information is minimal and nessary to the user. The theoretical principles includes:

(1) **Focus and Spatial Resolution (Foveation)** - The human vision has a tiny high-resolution fovea covering around 1 to 2 degrees of the visual field but accounts for around half of visual cortex (Krantz, 2012), outside this region is a much coarser peripheral vision where acuity drops rapidly (Iwasaki & Inomata, 1986).

Our device attempts to mimic this by using a focus point system where the user can contain the focus radius indicanting that area of the environment the user wants to be translated into audio and at what detail.

(2) **Selective Attention and Daliency** - Has the visual system cannot process all details at once, it selectively attends to salient or task-relavant features, this is done by implementing a bottom-up saliency and top-down goals to filter out redundant/irrelevant information (Kristjánsson et al., 2016). This is to mean that people focus on key objects ignoring uniform backgrounds as the nervous system ”tunes out” repeated stimuli and amplifies novel/focused ones (Gershman, 2024).

Has the device translates environmental information, any stagnating/constant information must be tuned out over time leaving behind changes indicating motion.

(3) **Scene Gist and Gestalt Organization** - Human vision rapidly extract the gist (background) and figures (foreground) from a scene within the first fixation (36ms) with around 80% accuracy (Loschky, 2025). This process is done through Gestalt principles that groups elements to simplify complex images such as by similarity, proximity, common region, continuity, etc. (UserTesting, 2024)

This indicates that the device should be able to generally/primitively seperate the environmental data into background and foreground categories where foreground can then be seperated into groups, this is information to indicate the translated audio.

(4) **Parallel Motion vs. Detail Pathways** - Human visual systems process motion and detials in two seperate channels, the **magnocellular pathway** for fast motion and size but in less detail, and the **parvocellular pathway** for static object in more detail (Zeki, 2015).

This suggest that when our device switches to motion detection it should prioritize the speed, size, and direction of that motion.

(5) **Temporal Dynamics and Scanning** - Vision is not a singlular static snapshot but rather continuous sample of the world via eye movements has humans typically shift gaze 3-4 times a second (Kristjánsson et al., 2016).

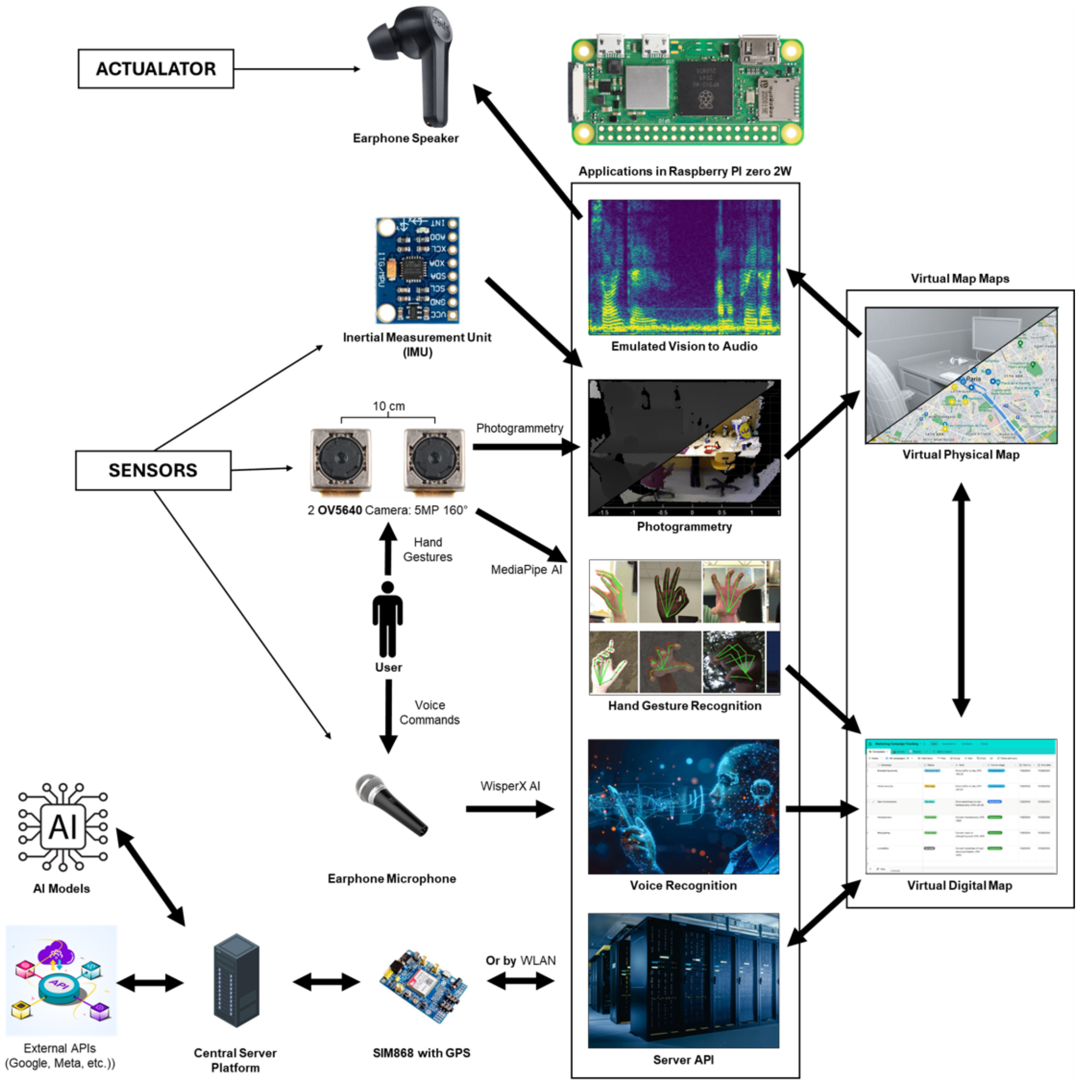
Thus our device should rapidly provide updating audio to allow temporal integration rather than one large static soundscape all at once, this also solves the issue of cognitive overload and sensory fautigue.

(6) **Multisensory Integration** - The brain integrates additional senses like auditory, vestibular, and proprioceptive senses to form a coherent representation of the environment (Kristjánsson et al., 2016).

Thus our device should align with those senses to prevent conflicting senses that could often cause nausea. This could be in a form of aligning the virtual environment with IMU to align with vestibular senses, or lower the output audio when the user is focusing on something else like talking to others.

Aside from these principles our device should be able to assist the user by navigating the physical world and the digitial world. While the principle assists with navigating the physical world like environment, object, etc. we implement AIoT to assist the user navigating the digital world like google maps, weather, social media platforms, etc. This is where the device will act as the client while the AIoT is managed by the server, as seen in figure 2.

Figure 2: AIoT System Architecture



The figure 2 shows the following components of the system:

(1) **Sensors** - The sensors captures stereo images through camera, user's head orientation through IMU, and voice commands through Microphone.

(2) **Microcontroller** - The raspberry pi zero 2w is the main microcontroller that handles client side processes like the five provided applications.

(3) **Actuators** - The Speaker from the earphone is the only major actualor of the

system.

(4) **Applications** - There are five major client side applications that interacts with the virtual maps such as to Emulate Human Vision, Photogrammetry, Hand Guesture Recognition, Voice Recognition, and Server API. Two of these application uses client side AI models:

(4.a) **Hand Gesture Recognition** - **MediaPipe** an Open-source model by Google is used to detect the hand gesture from the user.

(4.b) **Voice Recognition** - **WhisperX** by OpenAI is used to recognize instructions/infomations provided by the user

(5) **Virtual Maps** - There are two kinds of map that stores and manages two kinds of data, data that is plotted into physical space called the Physical Map, and data is not plotted into physical space called Digital Map.

(6) **Server Side** - The server side is responsible for dealing for dealing with client data and interact with external APIs such as:

(6.a) **Web APIs** - Interacts with web APIs such as google maps for directions or social media for communication

(6.b) **Local AIs** - Processes local AI such has Ollama GPT-OSS for assistant without actual humans or Yolo for advance image recognition/segmentation

The system is as follows:

(1) The stereo camera with help from the IMU is processed by the Photogrammetry Application to generate the Physical Map, where the Emulated Vision Application

can process it into audo cues.

(2) The right camera is individually processed by the Hand Gesture Application to detect if the guesture matches any instruction, if it does it records it in the Digital Map.

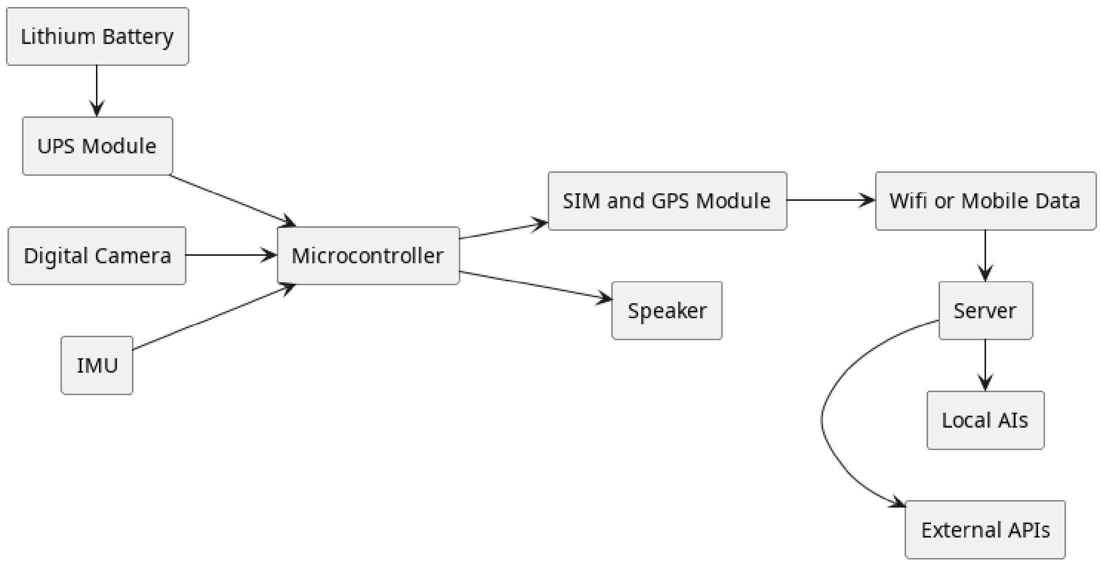
(3) If the Digital Map requests to enable voice commands then the Microphone is processed by the Voice Recognition to be provided to the Digital Map.

(4) If the Digital Map requests to access Network data such as GPS or Server APIs then that request is sent to the central Server.

(5) The central server processes the request either by running the AI models or making other external API request to fulfill the request, the response is sent back to the client's Digital Map.

(6) Once the Digital Map want to provide information to the user it is sent to the Physical Map which is then sent the the Emulated Human Application to generate the final Audio.

Figure 3: Block diagram of the System



The connections of between each components allowing this to work could be seen in figure **??**, such as the connections of each materials with relation to the raspberry pi.

Scope and Delimitation

Our research focuses on Sensory Substitution Devices, more specifically on Vision Substitution. Our device aims to fill in what different variety of visually impaired individuals lacks, like if fully blind our device should be able to act as an artificial eye providing information the person wants, or if the person is near sighted it could read text from afar from the user and warn of objects rapidly approaching the user, and so on. It does this by having a modular system running the main pipeline, the main pipeline inputs images from the stereo camera, creates a point cloud, maps it to a virtual physical map, follows an algorithm to emulate human vision processes, and finally generate the necessary audio cues for the individual. This pipeline allows applications to take into account more varied needs such has hand gesture recognition to control the parameters of the main pipeline so the person can navigate the device without seeing anything, image recognition to read text near or far, face recognition to recognize friends and family, and even more personalized interest like keeping track of expenses.

This research is conducted in the University of San Agustin and thus the prototype is initially to be tested by the researchers who only half of them are little bit near sighted, our testing sample is to soon include members of Iloilo blind associations such as the Association of Disabled Persons - Iloilo Inc. The main pipeline is prioritized over other features has it is the central component and partially answers the research question of solving existing sensory substitution issues. Has a result our prototype doesn't take into much consideration the much varied range of visual imparities, and since our study prioritize the main function over the physical design, the prototype would be heavier than what it could be if optimized for weight. However, our main objective is to build the main pipeline and the capability of support applications anywhere in the pipeline (modular operating system) showing it is possible that some of the issues facing most sensory substitution devices could be solved.

Review of Related Literature

Related Studies

Sensory Substitution Device Wearable Navigation Device

Review of Related Literature

Zvorișteanu et al. (2021) presented the **Sound of Vision (SoV) system**. It is a wearable SSD designed to aid VIPs in spatial cognition and navigation, using stereo-vision-based 3D reconstruction to interpret the environment and translate spatial data into audio and haptic cues. The device integrates a stereo camera, an infrared depth sensor, and an inertial measurement unit (IMU) to track the motion of the user's head and the installed camera mounted on the headgear, using these as its inputs. Furthermore, the output audio and haptic feedback are delivered via headphones and a haptic belt. The device processes visual data in real time using GPU-accelerated computation to reconstruct the 3D Environment and identify obstacles via segmentation algorithms based on disparity and histogram analysis. The depth of each pixel is determined by the stereo disparity formula $Z = \frac{f \cdot B}{d}$, where Z is the depth, B is the baseline, f is the focal length, and d is the disparity. The SoV converts environmental data into audio-haptic cues, allowing users to detect obstacles, open spaces, and open areas. In testing the device's performance with the visually impaired participants, it showed a 10% improvement in depth accuracy and an 88.5% task success rate compared to standard stereo vision methods. The results highlight the device's effectiveness in real-world navigation and its potential to enhance mobility and spatial awareness compared to traditional aids such as white canes.

Sami et al. (2025) developed a **smart wearable (hat) assistive device** that integrates object detection with voice-assisted navigation to support VIP in real-time spatial awareness. The device uses an ESP32-CAM module to capture live video and transmit it via Wi-FI to a mobile application that employs a TensorFlow-based deep learning model for efficient object recognition. The camera-detected objects are then processed into audio prompts that send alerts to the user when there are nearby obstacles and provide navigation directions. The programmed system produced a seamless object-detection-to-voice pipeline. The Smart Hat's system architecture demonstrates the integration of an IoT processing method to enable wireless communication between the smart hat and the user interface for remote processing and real-time updates for the user. Testing results show an object detection accuracy of over 91%, highlighting the reliability in dynamic environments. In addition, the Smart Hat device offers compact, low-cost, and user-friendly features. It demonstrates how IoT-enabled computer vision can transform traditional assistive technologies into intelligent, voice-guided navigation tools for the visually impaired.

RealSense AI (2025), in collaboration with Eyesynth, developed a wearable sensory-substitution device designed to enhance spatial awareness and independent navigation for visually impaired users by converting real-time 3D spatial data into bone-conduction audio cues. The device is called **Non-Invasive Image Resynthesis into Audio (NI-IRA)**. The device employs the RealSense D415 depth camera to capture real-time 3D depth data calculated using the stereo disparity formula $Z = \frac{f \cdot B}{d}$. Then it is processed to generate a point cloud representation of the surroundings, where each depth pixel being mapped to 3D coordinates using the equation $X = (u - c_x)\frac{z}{f_x},\ Y = (v - Cy)\frac{Z}{f_y},\ Z = Z$ to provide a complete spatial surrounding rather than a 2D view. The point-cloud data conjoint with the inertial and head motion tracking, is then processed through an onboard ASIC and Simultaneous Localization and Mapping (SLAM) modules to identify object orientation, translating them into audio signals such as pitch (height), volume (distance), and stereo panning (lateral position), allowing users to perceive objects and obstacles up to 5 meters away without external infrastructure. In testing NIIRA, it demonstrated improved independence and navigation in both indoor and outdoor environments, adapting to users' preferences, and displayed high obstacle-detection accuracy, minimal processing latency, and enhanced independent navigation. The participants reported that the bone-conduction audio feedback provided a clear environmental awareness without blocking external sounds.

Udayakumar et al. (2025) designed a **Smart Vision Glasses (SVG)** that processes environmental input data using AI, LiDAR depth mapping, and computer vision algorithms. The device's front-facing camera captures visual data, while the LiDAR sensor measures real-time depth information using the principle of Time-of-Flight (ToF) – where the distance (D) is calculated using the formula $D = \frac{c \cdot t}{2}$, where c represents the speed of light and t the time taken for emitted light to return after reflection. These depth and image data are then processed through an AI-based recognition model that uses Convolutional Neural Networks (CNNs) to classify objects, text, and faces within the user's field of view. The recognized elements are subsequently analyzed and prioritized based on proximity and relevance using the LiDAR-assisted spatial mapping, which produces a semantic understanding of the scene; implementing Volume of Interest (VOI), a controllable spatial region or focus radius that limits feedback, to prevent sensory overload, enhancing user focus in dynamic environment. This processed information is then transmitted to a smartphone-based application converting the recognized data into voice audio cues through text-to-speech (TTS) engine. The device applies natural language processing (NLP) principles to ensure that the audio output is contextually meaningful and user-friendly. The input-output process follows a sequential pipeline:

(1) capture of the image depth data

(2) AI-driven detection and classification

(3) distance estimation through the ToF equation

(4) conversion of results into voice audio output describing nearby objects, text, or faces

The SVG through hearing supports the four primary modes – “Thighs Around You”, “Reading”, (Walking Assistance), and “Face Recognition”. The SVG is also equipped with gesture and voice control interfaces, allowing users to switch modes or issue commands hands-free, enhancing accessibility and user interaction. A multicenter usability study was conducted across five rehabilitation centers in India, involving 90 participants with a mean age of 23.5 years, to test the device functionality. The participants tested the four primary modes by completing real-world navigation and recognition tasks, and their feedback on the device’s usability and helpfulness was also recorded. Results showed that 72.9% indicated that ”Reading” mode is helpful, 44.7% for “The Things Around You”, 36.5% for “Face Recognition”, and 22.4% for “Walking Assistance”, showing that the device effectively enhances environmental awareness, reading ability, and object identification.

Ruan et al. (2025) developed a multifaceted sensory substitution wearable device that uses an audio-based curb detection to improve real-time awareness and navigation safety for individuals who are blind or have low vision. The device integrates a stereo camera, ultrasonic range sensors, and inertial motion units (IMUs) to identify curbs and ground-level transitions Adapting the stereo disparity formula $Z = \frac{f \cdot B}{d}$, the system calculates the curb height and distance, while the ultrasonic sensors confirm the surface proximity of the environment for redundancy and accuracy. The sensory data are processed and converted into audio and vibrational cues, where the pitch and repetition rate of the sound correspond to the curb’s distance and height, providing intuitive, time-sensitive feedback. In the pilot testing, 12 participants were involved, 8 with blindness and 4 with low vision. The participants tested the navigation feature of the device in both indoor and outdoor test environments, including simulated sidewalks, ramps, and descending curbs. After testing, the performance statistics of the device showed an average curb detection accuracy of 94.3%, with a false-negative rate of 3.7%, and an average alert response time of 1.2 seconds, indicating the system's reliability in detecting and signaling curbs in real time. Participants reported that it also improved confidence in mobility and reduced the risk of missteps or trips. This demonstrates that the audio-based curb detection approach driven by LiDAR-like depth estimation and multimodal feedback effectively enhances real-time alerting and safe navigation for the VIP.

Viancy V et al. (2024) introduced AuralVision, a wearable assistive device that is designed as eyewear to improve navigation for visually impaired individuals by incorporating object detection, scene classification, and reinforcement learning. The device's system architecture includes a camera sensor to capture the user's environment, image-processing and object-detection software to classify obstacles and identify significant objects, and an audio module to translate detection results into auditory cues for navigation. Utilizing a convolution neural network (CNN) approach or deep learning-based vision algorithms trained on the Object Net 3D dataset to interpret the surrounding environment and convert the visual information to auditory cues that convey spatial awareness. AuralVision uses 3D visualization and sound-based feedback to identify objects, classify road scenes, and assist users with obstacle avoidance and pathfinding. The device adapts to dynamic environments through continuous learning, thereby improving navigation decision-making.

Hamilton-Fletcher et al. (2021) developed a mobile sensory-substitution device called SoundSight. This device translates input color, depth, and temperature data into output audio cues, allowing the visually impaired users to sense their environment. The system captures environmental input using a mobile with an RGB-D camera and a thermal sensor. Then it processes the data stream through a multi-feature mapping pipeline before integrating it into a composite audio stream. Depth is calculated using the stereo disparity formula to determine the distance of the object from the user's perspective. At the same time, the RGB camera's color hue values are mapped to sound pitch using the equation $f_c = k_c \times H$, where H represents the hue intensity and scaling constant. Temperature measurements are converted to amplitude modulation through $A_t = k_t \times T$, where T is the sensed temperature and k_t is a sensitivity coefficient. These parallel mappings are integrated using the data fusion algorithms, synchronizing depth, color, and thermal inputs before generating the composite audio streams that encode environmental depth through rhythm, color through pitch, and temperature through loudness. In testing the device's functionality, 15 blind and 10 low vision participants are involved. The device demonstrated rapid user learning and over 85% recognition accuracy, demonstrating that the multi-feature sonification can reliably convey complex environmental information.

Han et al. (2026) designed a Multi-Path Sensory Substitution Device (MSSD) that enhances low-vision mobility and virtual navigation of the visually impaired people by integrating real-time depth mapping, IMU-based motion tracking, and depth optimization algorithms. The system processes input from a depth camera and inertial sensors to create a 3D spatial model of the environment, using the stereo disparity equation to calculate object distance and the projection model $X = (u - c_x)\frac{z}{f_x}$, $Y = (v - c_y)\frac{z}{f_y}$ to map each pixel to world spatial coordinates. After acquiring the data from the processed input, it generates a 3D point cloud, which then is converted into an occupancy grid for spatial awareness, while the IMU readings refine user orientation using the equation $\theta_t = \theta_{t-1} + \omega t \Delta t$. A modified A* (A-star) algorithm is used to compute the optimal navigation routes using the cost function $f(n) = g(n) + h(n)$, balancing actual distance and heuristic estimates to avoid obstacles. The selected virtual path is then translated into audio-haptic feedback, where the vibration intensity and sound frequency indicate direction and proximity, using speakers or headphones to create output data directional audio cues and vibration motors/haptic actuators for tactile feedback. To check the device's functionality, the researchers tested it with 20 low-vision participants. The system attained a result of 91% path following accuracy and a 38% reduction in navigation errors. As the device demonstrates a satisfactory functional performance, this indicates that combining 3D mapping, motion sensing, and heuristic path planning effectively supports real-time, safe, and virtual navigation for the visually impaired individuals.

Commère and Rouat (2023) evaluated five depth-to-sound sonification methods to compare their performance and effectiveness. One of the methods evaluated is the LiDAR-to-repetition-rate sonification model. The identified sonification model converts visual information into rhythmic auditory cues, helping visually impaired individuals to develop a sense of spatial awareness through sound. Using the Time-of-Flight (ToF) equation $D = \frac{c \cdot t}{2}$, the model measures the distance of an object and maps it inversely to the beep repetition rate $R = \frac{k}{D}$, where R is the repetition rate (Hz), D is the distance (meters), and k is a scaling constant which is determined experimentally to maintain the perceptible tempo differences between near and far objects which results if an object is closer it generates faster repetition rates or beeps and far detected objects have slower rhythm beeps which the model's spectrogram design reflects this relationship. Whereas the temporal density (number of pulses per second) indicates proximity, while frequency and amplitude encode additional object features such as height and reflection intensity, creating a dynamic "rhythmic depth map" that represents distance through time spacing between sound bursts—in experimental testing with 28 sighted but blindfolded participants, a three-phase protocol consisting of Depth Estimation, Azimuth Estimation, and Retention Assessment confirmed that repetition-rate sonification produces the lowest mean absolute error (MAE) in distance perception and was the most intuitive and accurate among the other mapping methods like frequency or amplitude-based cues. The results finding revealed a strong inverse correlation between the perceived distance and repetition rate ($R \propto \frac{1}{D}$), indicating that LiDAR-driven rhythmic approach effectively translates the spatial depth into comprehensive sound patterns, offering an enhanced basis for real-time auditory sensory substitution systems for the visually impaired.

Zhao et al. (2025) conducted a clinical evaluation of an assistive device, a wearable electronic navigation aid (ENA), for visually impaired individuals through a prospective, non-randomized, single-arm, open-label trial involving 30 participants (each participant was given five trials for each functional task), selected through purposive sampling to ensure that individuals with blindness or severe visual impairment could assess the performance of the assistive device. The study primarily focused on the functional efficacy of the device, assessing how effectively it supports navigation and daily tasks of the visually impaired users. The assistive device used in testing integrates ultrasonic sensors, vibration motors, and audio output components. The study implemented a structured, protocol-guided or functional evaluation framework that combined quantitative measures including accuracy rate ($\frac{Sucessful\ Trails}{Total\ Trails} \times 100$), error rate ($\frac{Navigation\ Errors}{Total\ Trails} \times 100$), reaction time ($RT = T_t response - T_{stimulus}$), task completion time ($TCT = T_{end} - T_{start}$), and performance improvement rate ($PIR = \frac{(Post\ test - Pre\ test)}{Pre\ test} \times 100$) to assess navigation accuracy, responsiveness, and efficiency objectively, supported by qualitative user feedback on comfort, usability, and confidence. The study reveals that the ENA device effectively improved safe navigation and mobility performance for visually impaired users.

Summary of Related Literature

Recent studies (2019–2025) on assistive technologies for the VIP highlight wearable and mobile devices utilizing a sensory substitution approach to convert visual/spatial data into audio, haptic, or multimodal cues, applying stereo vision (e.g., disparity formula $Z = \frac{f \cdot B}{d}$), depth sensors like LiDAR/ToF (e.g., $D = \frac{c \cdot t}{2}$), AI-driven computer vision (e.g., CNNs, TensorFlow), and IMUs. This approach expresses the outcome produced by the device such as real-time obstacle detection, object recognition, and navigation, similar to the Caraiman et al.'s Sound of Vision (SoV), which addresses sensory overloading via a controllable focus radius (VOI) and modular components for partial generalizability; Sami et al.'s Smart Hat for cost-effective voice output; RealSense's NIIRA point-cloud processing to semantic audio; Commère and Rouat's LiDAR sonification for rhythmic depth cues; Hamilton-Fletcher et al.'s SoundSight for mobile timbre/volume mappings; Ruan et al.'s curb detector for 94.3% accuracy; Han and Li's Multi-Path Sensory Substitution Device (MSSD) for 91% path accuracy; Zhao et al.'s electronic navigation aid (ENA); and Udayakumar and Gopalakrishnan's Smart Vision Glasses (SVG) for AI-LiDAR voice modes. However, systems like AuralVision that utilizes ToF lasers for bone-conduction stereo sound, lab-effective but range-limited, Eyesynth's NIIRA object recognition process prone to overload without human emulation, Depth Sonification repetition rates for intuitiveness, SoundSight LiDAR/thermal to audio hwich is app-constrained, Smart Hat that is non-modular, and multi-faceted SSDs with 85% curb accuracy often face issues like sensory overload, training demands, limited real-world adaptability, and lack of digital-physical navigation. These studies proposes an idea that work closely aligns with SoV by tackling sensory overloading through VOI and modular components but advances it by explicitly mimicking human perception via a modes system and a modular IoT pipeline that applications can embed for navigation in both physical and digital worlds, addressing SoV's training needs and real-world performance gaps while incorporating brain emulation for superior overload reduction and versatility, building on these predecessors to offer a more intuitive, scalable solution for VIP independence. In addition, evaluations across studies ranging from 12 to 90 participants showcase 85% to 95% accuracies and qualitative benefits like reduced errors (e.g., 38% in MSSD). These studies introduce an innovative and improved SSD by adding brain emulation, AI, and IoT for advanced overload reduction and versatility.

Materials and Methods

Research Design/Methods

This research implements an experimental and iterative prototyping research focused on developing and testing the wearable AIoT-based visual sensory substition device. This is done by following a development methodology which is similar to the engineering design cycle as seen in figure 4.

Figure 4: Development Methodology of the research

Source: openeducationalberta.ca

Materials and Requirements

Materials and Processes

Hardware Material

Software Specification

Library and Board Managers

Other Materials/Equipment/Devices

Prototype Building Procedures

Data Presentation and Analysis

Ethical Considerations

Results and Discussion

Appendices

References

Commère, L., & Rouat, J. (2023). Evaluation of short range depth sonifications for visual-to-auditory sensory substitution.

Gershman, S. J. (2024). Habituation as optimal filtering. *iScience*, *27*(8), 110523.

Hamilton-Fletcher, G., Alvarez, J., Obrist, M., & Ward, J. (2021). Soundsight: A mobile sensory substitution device that sonifies colour, distance, and temperature. *Journal on Multimodal User Interfaces*, *16*(1), 107–123.

Han, Z., Li, S., Wang, X., Hu, X., Higashita, R., & Liu, J. (2026). Multi-path sensory substitution device navigates the blind and visually impaired individuals. *Displays*, *91*, 103200.

Hou, Y., Xie, Q., Zhang, N., & Lv, J. (2025). Cognitive load classification of mixed reality human computer interaction tasks based on multimodal sensor signals. *Scientific Reports*, *15*(1).

Iwasaki, M., & Inomata, H. (1986). Relation between superficial capillaries and foveal structures in the human retina. *Investigative Ophthalmology & Visual Science*, *27*(12), 1698–1705.

Krantz, J. H. (2012). Chapter 3: The stimulus and anatomy of the visual system. In *Experiencing sensation and perception*. Pearson Education.

Kristjánsson, Á., Moldoveanu, A., Jóhannesson, Ó. I., Balan, O., Spagnol, S., Valgeirsdóttir, V. V., & Unnthorsson, R. (2016). Designing sensory-substitution devices: Principles, pitfalls and potential1. *Restorative Neurology and Neuroscience*, *34*(5), 769–787.

Loschky, L. C. (2025, October). Recognizing the gist of a scene [Accessed: 2025-11-06].

Mishra, A., Bai, Y., Narayanasamy, P., Garg, N., & Roy, N. (2025). Spatial audio processing with large language model on wearable devices.

RealSense AI. (2025). Eyesynth: Powering Sonic Vision for the Blind.

Ruan, L., Hamilton-Fletcher, G., Beheshti, M., Hudson, T. E., Porfiri, M., & Rizzo, J.-R. (2025). Multi-faceted sensory substitution using wearable technology for curb alerting: A pilot investigation with persons with blindness and low vision. *Disability and Rehabilitation: Assistive Technology*, *20*(6), 1884–1897.

Sami, M., Agha, D., Baloch, J., Dewani, A., & Maheshwari, K. D. (2025). Efficient object detection and voice-assisted navigation for the visually impaired: The smart hat approach. *Sukkur IBA Journal of Computing and Mathematical Sciences*, *8*(2), 1–12.

Shinagawa Lasik & Aesthetics. (2025, May). Perfect vision and eye health in the philippines.

Theodorou, P., Tsiligkos, K., & Meliones, A. (2023). Multi-sensor data fusion solutions for blind and visually impaired: Research and commercial navigation applications for indoor and outdoor spaces. *Sensors*, *23*(12), 5411.

Tokmurziyev, I., Cabrera, M. A., Khan, M. H., Mahmoud, Y., Moreno, L., & Tsetserukou, D. (2025). Llm-glasses: Genai-driven glasses with haptic feedback for navigation of visually impaired people.

Udayakumar, D., Gopalakrishnan, S., Raghuram, A., Kartha, A., Krishnan, A. K., Ramamirtham, R., Muthangi, R., & Raju, R. (2025). Artificial intelligence-powered smart vision glasses for the visually impaired. *Indian Journal of Ophthalmology*, *73*(Suppl 3), S492–S497.

UserTesting. (2024, April). 7 gestalt principles of visual perception: Cognitive psychology for ux.

Viancy V, N. R. C., M Mouli, M., & S, S. (2024). Aural vision - envisioning the independence of the visually impaired. *2024 International Conference on Inventive Computation Technologies (ICICT)*, 387–392.

World Health Organization. (2023, August). Blindness and visual impairment.

Zeki, S. (2015). Area v5 - microcosm of the visual brain. *Frontiers in Integrative Neuroscience*, *9*.

Zhao, Y., Yang, S., Tao, Y., & Kang, H. (2025). Evaluation of the efficacy of an assistive device for blind people: A prospective, non-randomized, single arm, and open label clinical trial. *Current Eye Research*, *50*(8), 865–869.

Zvorişteanu, O., Caraiman, S., Lupu, R.-G., Botezatu, N. A., & Burlacu, A. (2021). Sensory substitution for the visually impaired: A study on the usability of the sound of vision system in outdoor environments. *Electronics*, *10*(14), 1619.