From Harness to Hardware: Investigating a Robotic Guide Dog Prototype

Maia Briggs

4th Year Project Report
Computer Science
School of Informatics
University of Edinburgh

2025

Abstract

Navigating the world presents significant challenges for blind and visually impaired (BVI) individuals, with traditional aids like white canes and guide dogs offering limited obstacle detection or requiring intensive care. This project explores the potential of a robotic guide dog prototype to bridge this gap, combining stereo vision, voice interaction, navigation and path planning, to investigate a dynamic, user-centric mobility aid.

An interview was conducted with a BVI participant to determine requirements for a potential robot guide dog. Four requirements were determined, of which three were investigated during the development of the system prototype. The system was developed on a TurtleBot3 platform using ROS2, featuring a 3D printed handle for physical guidance, a stereo vision node for user and obstacle tracking and an audio feedback system for environmental awareness. Two navigation approaches, NAV2 integration and potential fields, were implemented to maintain an optimal guiding distance while adapting to the movements of the user. The prototype also included a voice interaction system, allowing users to query the system for verbal feedback about surrounding obstacles.

While object detection and audio subsystem demonstrated promising functionality, navigation proved challenging due to hardware limitations and algorithmic constraints. These shortcomings highlight the delicate balance required in assistive robotics, emphasising the need to prioritise user trust in order to prevent abandonment of assistive devices.

Though the prototype fell short of fully replicating the adaptability of a real guide dog, the project provides valuable insights into the technical and usability challenges of navigation aids. Future work should focus on improving navigation architectures, upgrading sensor integration, and including BVI users in all aspects of the design process to ensure reliability and intuitive interaction. Ultimately, this work reinforces that assistive technologies must prioritise seamless usability and trustworthiness alongside technical innovation to truly empower the independence of BVI individuals.

Research Ethics Approval

This project obtained approval from the Informatics Research Ethics committee.
Ethics application number: 830542
Date when approval was obtained: 2024-11-15
The participants' information sheet and a consent form are included in the appendix.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Maia Briggs*)

Acknowledgements

I would like to express my deepest gratitude to all who provided invaluable support and guidance through this project.

First and foremost, my sincerest thanks go to my supervisor, James Garforth, for his unwavering patience, expertise and encouragement. His feedback and support throughout this project were instrumental in shaping its direction and ensuring its completion.

I am also immensely grateful to Garry Ellard, Ciarán Johnson and the rest of the Informatics MakerSpace for their technical assistance, hardware support and troubleshooting expertise. Their generosity with time and resources made much of this work possible.

To Olivia and Amelia, thank you for your endless moral support, laughter and general encouragement throughout this year. You both have kept me motivated even when challenges seemed insurmountable.

Table of Contents

Chapter 1

Introduction

1.1 Motivation

Globally, at least 1.1 billion people live with vision loss, including 43 million who are blind and 295 million with moderate-to-severe impairment, a figure projected to rise 55% by 2050 IAPB. While white canes and guide dogs provide critical mobility support for blind and visually impaired (BVI) individuals, their limitations, such as canes missing overhead obstacles, Attia and Asamoah [2020], and guide dogs require intensive care Craigon et al. [2017] have spurred rapid advancements in the development of technological alternatives.

Yet, while recent advancements in deep learning and perception demonstrate improved environmental perception, (e.g. RoboGuide Uni and PeopleLens Hughes [2022], most solutions still fall short of replicating the nuanced support of traditional navigation aids. For instance, devices like the WeWALK Smart Cane prioritise obstacle detection over holistic usability, relying on smartphone integration and static haptic feedback that can fail to replicate the dynamic guidance of a human or guide dog WeW [2024].

Similarly, assistive devices are often underused or completely avoided by those who need them. Of 286 patients questioned in Gothwal and Sharma [2023], 22% of patients had abandoned at least one of their prescribed vision devices, with the majority reporting device-related issues that frustrated them when using the devices. This trend exists across the board with assistive technologies, Brunes et al. [2024] suggests abandonment rates of up to 78% for hearing aids, expressing that abandonment can be due to a change in the needs of users, trouble changing settings and simply that a better device has become available. It is not just assistive technologies that struggle to keep their user base. In Brunes et al. [2024] states that only 68% of their visually impaired participants used at least one type of mobility aid, with the most frequently used type being door-to-door transport (52%). Brunes et al. [2024] shows that the use of mobility aids that facilitate independence in BVI individuals is significantly lower, with white canes being at 38%, guide dogs at 13% and GPS technologies at 12%.

The persistent gaps in both traditional mobility aids and emerging assistive technologies reveal a critical need for solutions that prioritise intuitive usability alongside functional reliability, reassurance and safety for their BVI users. While innovations like WeWALK demonstrate the potential of AI-driven assistance, their minimal adoption in BVI communities, reflect a disconnect between technological capability and the wants of the user base. This disparity is further exacerbated by the projected rise in global vision loss, which demands adaptive tools that empower BVI individuals without creating cognitive load or imposing physical burdens. Ultimately, the success of assistive technologies hinges not only on their technical prowess but on their ability to integrate seamlessly into the lived experience of BVI individuals, promoting independence while respecting the diverse ways in which we navigate the world around us.

1.2 Objectives and Contributions

This project investigates the challenges of developing dynamic, user-centric navigation systems by investigating assistive navigation aids and a potential robotic guide dog prototype. Key contributions include:

1. **Defining an ideal robotic guide dog architecture by:**
   - Synthesising functional and non-functional requirements from a BVI interview and gaps in literature.
   - Proposed a multi-sensor system to address current limitations with assistive guiding aids
2. **Prototyping and evaluating critical subsystems, including:**
   - A stereo vision node for user and obstacle tracking
   - Two different navigation approaches to maintain an ideal guiding distance
   - A voice interaction system for users to locate obstacles in the environment.
3. **Assessed current technological readiness:**
   - Evaluated performance gaps between theoretical capabilities of subsystems and real-world usability.
4. **A roadmap for future work in robotic navigation guides prioritising:**
   - Hybrid navigation solutions.
   - Sensor upgrades and BVI co-design in development cycles.

To evaluate this ideal architecture, a prototype was developed on the TurtleBot3 platform using ROS2, integrating three key subsystems: (1) a stereo vision node for user and environmental tracking using two Raspberry Pi3 Cameras, (2) LiDAR-based obstacle detection and (3) an audio feedback node for verbal interaction. Two different approaches were implemented to combat the unique problem of guidance while staying proactive about the user's current position. The first approach involved integration with ROS2 NAV2 Navigation, which relied on NAV2's behaviour tree for global path planning and would use local navigation goals to position itself in front of the user. The second approach used the potential fields method adapted from Ge and Cui [2002] to maintain an optimal guiding system distance, whilst leading the user and navigating obstacles. To enable these approaches, a stereo vision node processed RGB camera data to track the user's position. However, environmental lighting decisions, inaccurate calibration, and untrustworthy disparity maps revealed persistent challenges in maintaining reliable depth information. To bridge this gap between the user and the robot, the system incorporated an audio feedback node, which provided vocal feedback on objects detected in the environment, allowing the user to interact with the robot verbally.

The development of this prototype has made it clear that technologies designed to empower BVI individuals must prioritise adaptive usability and fully consider the holistic experience of using the product. The challenges encountered in sensor calibration and integration with path planning highlight how quickly any sense of doubt in the system's validity can erode trust in the robot as a whole. This aligns with the participant's insistence on intuitive communication and physical reassurance, suggesting that future work must treat reliability and user agency as non-negotiable constraints, not an afterthought.

Chapter 2

Background

2.1 Guide Dogs

For years, guide dogs have played an essential role in facilitating BVI individuals' independence, mobility, and safety in their day-to-day lives. Research conducted by the RNIB states that a quarter of BVI persons feel they do not leave their homes as much as they would like, with 11% saying they cannot participate in hobbies, physical activities and social events RNIB. [2022]. As well as this, the BVI experience of the world is drastically different from that of a seeing person, with BVI individuals reporting more intense feelings of loneliness, alienation and social inadequacy in a society that is not geared to accommodate them Rokach et al. [2021].

As trained, working service animals, guide dogs assist their handlers in interacting with the world. Guide dogs can do more than help with daily tasks; they can provide a means of interacting with the world at large. From navigating traffic and unforeseen obstacles to providing companionship and easing loneliness, these dogs can be integral in increasing a BVI individual's perceived quality of life Glenk et al. [2019].

However, training these dogs to handle these complex environments is a complexity in itself. Aspiring guide dogs must be assessed against extensive requirements to ascertain their suitability for a working role. The International Guide Dog Federation provides a list of these requirements, IGDF, with evaluations focusing on a dog's physical health, behavioural attributes and physical characteristics. Approximately 70% of dogs that fail to qualify as a guide dog are disqualified based on behavioural traits. Though qualification rates vary from country to country, the number of dogs able to graduate as working dogs are significantly less than their demand ARATA et al. [2010].

This assessment, however, is limited in assessing the capability of these dogs, with many having a nuanced effect dependent on their specific owners' needs and daily life. Dogs who qualify may not find a perfect partner on the first try. Not all guide dog matches are successful, and incorrect matches may actually lead to a severe decrease in quality of life Lloyd et al. [2016]. For instance, a dog's personal traits may make it extremely attentive when not actively working as a guide dog (off harness). This could be seen as a positive trait to a BVI individual requiring more support and companionship, but as unwanted neediness to others. As seen in Craigon et al. [2017], guide dogs are both working dogs and pets, coming with considerable upkeep, inconsistent behaviour, and personal interests that could develop into a problematic distraction later on.

Thus, it is unsurprising that not all BVI individuals use guide dogs in their day-to-day lives. In 2023, Guide Dogs for the blind reported that there are only around 2900 active guide dog partnerships in the UK GDFB [2023]. There are many reasons why a BVI individual may choose not to have a guide dog. For instance, many BVI persons feel they cannot financially support a dog's lifestyle and the environment. They also often feel there are social barriers to getting a guide dog, considering themselves having "too much vision for a dog" AFB [2024]. Thus, applicable individuals with low vision believe that they would not benefit and should not qualify despite their eligibility.

The end of a guide dog partnership also comes with its own challenges. Guide dog owners often report having good experiences with their first dog and tend to experience higher emotional distress when this relationship comes to an end, especially if this end came abruptly or due to a sudden death Jill Nicholson and Griffiths [1995]. This can greatly impede a BVI individual's connection with their subsequent dog, with many needing to develop new skills and routines, whilst facing the feelings of guilt and sadness that may come from the end of a guide dog partnership.

2.2 White Canes

The other primary way in which BVI individuals will navigate the world is through the use of white canes. These are typically used to aid a BVI individual in scanning their environment for obstacles and help find orientation clues Scotland [2024]. However, there are many other types of mobility canes that assist BVI individuals with more than navigation. For instance, shorter white canes are often used as a "symbol cane" to make others in the environment aware of the user's visual impairment UK [2024].

Despite this, white canes have their own disadvantages. Whilst guide dogs can avoid obstacles without their handler needing to be aware of them, white canes allow for the handler themselves to detect the obstacle, and choose how best to manoeuvre around it. This may lead to situations in which the user accidentally trips over an obstacle because the white cane did not make contact with it GDv [2024]. The attention a white cane brings can also be a disadvantage. Many sighted individuals may not know how to act when encountering a BVI individual, leading to unwanted assistance that could develop into a safety hazard Ortiz [2024]. White canes cannot assist with traversing traffic in the same way a guide dog can. Whilst a BVI individual may be aware of the surrounding traffic to an extent, street crossings, especially in areas without proper crossing points, can be dangerous with only a white cane as assistance Attia and Asamoah [2020].

2.3 Visual Assistive Technology

Despite guide dogs and white canes being the simplest and most affordable navigation tools for the BVI, they do not always provide BVI individuals with all information and features for safe mobility Lloyd et al. [2016]. In the 1960s, many researchers began realising the potential that sensor-based technologies have in assisting BVI individuals in their daily lives Hersh and Johnson [2008]. Visual assistive technology is typically divided into three categories: vision enhancement, vision substitution and vision replacement. Guide dogs, white canes, and sensor-based assistive technologies all fall into the vision substitution category, providing an alternative, non-visual way to interpret the environment. Elmannai and Elleithy [2017].

2.3.1 Robot Guide Dog Assistive Technologies

With the development of robotic guiding solutions, gaining attention in recent years, researchers have aimed to create systems that can assist BVI individuals in their daily lives by replicating the functionality of a living guide dog. These projects focus on supporting an individual's orientation and mobility (O&M) skills, mainly addressing navigation, path planning and obstacle avoidance. The most common approach to this is to develop a mobile robot which guides the user by an attached leash or cane, as seen in Xiao et al. [2021], Lu et al. [2021] and Chuang et al. [2018].

Xiao et al. [2021]'s guide dog solution utilised a quadrupedal robot, Mini Cheetah Katz et al. [2019], and a physical leash. By using a leash with dynamic tension, Xiao et al. are able to change the robot's path planning based on how the user is moving. They assume that the robot will have two modes of function, taut mode and slack mode. Whilst in taut mode, Xiao et al. assume that the human will be guided by the robot and that the human will be guided in the direction the robot is moving via the force provided by the leash. When the leash is slack, the human would not feel any force in the leash and would not move. Their hybrid system focuses on capturing the switches between these two leash states, constructing a leash tension model by capturing the relationship between the robot speed and the leash tension, using a linear regression model. This is then fed into their path planner, where they add leash tension as a constraint to their local path planner. Xiao et al. make clear that it is imperative to know the states of both the robot and the human, and thus use an occupancy grid produced by a 2D LiDAR and Depth-RGB camera to estimate the position and depth of the guided person.

Xiao et al.'s robotic dog was, in fact, successful at guiding three blindfolded participants to a given goal without collision, taking roughly 75s. However, whilst their work was successful in directing the participant to the given goal pose (position and orientation), they did not provide any insight into how intuitive their solution was for the guided user, nor any further information on how their dynamic leash system improved the experience for the guided user.

Chuang et al. [2018]'s solution used a trail following mobile robot, with an attached cane-like rod. Users would grasp the cane as if using a white cane, and then be guided by the robot, which was designed to follow man-made trails. There were two hand grasps Chuang et al. considered when designing the cane: a forehand grasp, where the cane would be extended out to allow for more exploration and an upright grasp, where the robot would be pulled closer to the user, allowing for navigation within crowded environments. These two grasps could be likened to how a BVI individual would use a long cane and the shorter symbolic cane, with long canes similarly held extended for exploration, whilst symbolic canes existed to help with close obstacle detection and provide visual stimuli for their disability to others. To ensure the robot kept following the man-made trails, they used three cameras, which would detect if the robot needed to turn left, right or keep moving forward. This behaviour reflex was trained in a deep CNN model with the direct perception of the robot's lateral distance and heading. The system was trained on man-made yellow-blue trails and the textures of the Boston Freedom Trail in both real and virtual environments.

Chuang et al. conducted user studies with 10 BVI participants, with nine being blind and one heavily visually impaired. They found that some participants found the system reliable and trustworthy for navigation, however, wanted there to be stronger interaction between the user and the robotic guide dog, with many saying the robot moved at too slow a speed. Participants also found the suggested price of robot dog of USD 500 acceptable.

Lu et al. [2021] also suggested a combined guide dog robot and cane solution consisting of a mobile robot base with an attached rigid handle. Their robot was designed to follow preselected UWB beacon waypoints while avoiding both static and dynamic obstacles using a sensor tower that included a 3D LiDAR and depth camera. The robot would communicate with the human user through speakers integrated into the UWB beacons and the robot's handle. The beacons would provide more situational awareness of the environment to the users, whilst the handle would give voice feedback including travel information, response to queries for assistance and requests for speed changes. The handle also had a sliding button to stop and adjust the robot's velocity throughout the journey. Lu et al.'s path planning approach used deep reinforcement learning to collate the data from the UWB beacons with the environment's state information retrieved from the sensor tower. This successfully allowed the user to navigate between obstacles; however, could cause unnecessary movement and weaving. Thus, Lu et al. proposed a soft update technique which would take into consideration how the robot's movement would affect the user's comfort throughout navigation, producing a comfort map of the environment. This allowed them to plan trajectories that minimised the user's discomfort level, alleviating the weaving and creating a better user experience.

Lu et al.'s robot dog successfully guided the 8 BVI user study participants around the designated route. The participants liked the handle but found that the speed was too slow for them and that the speed adjustment was not beneficial. They did experience some unnecessary slewing and speed changes. Participants also reported the vocal feedback from the handle as helpful and informative. The participants thought that this would be a good way to navigate unfamiliar environments, where they had not built a clear mental map. They suggested improving the device by providing rough terrain traversability. The participants also suggested that the vocal feedback be gentler, or on prompt to avoid auditory overload, and that the robot be multifunctional as to assist in other tasks such as carrying groceries when shopping.

These studies provide a clear picture of the current landscape of robotic guide dogs, however, all fall short in fully encompassing the capabilities of an assistance animal. Whilst, Lu et al. [2021] does gain insightful feedback from their participants about the features a BVI individual would want in a robotic navigation guide dog, many of the studies do not address the user experience and relationship between the BVI individual and the dog. Lu et al. study makes it clear that BVI individuals desire more from their robotic companion than just simple navigation.

2.3.2 Other Assistive Technologies

Robotic guide dogs are not the only assistive technology for the blind and visually impaired. Many researchers have looked into ways in which we can innovate to assist BVI individuals in navigating their daily lives and tasks. One of the most prominent solutions is the Smart Cane Wahab et al. [2011], which attempts to extend upon the functionality of the white cane, by adding various sensors and buzzers to provide more information about the environment. There are commercial Smart Canes currently being developed and sold by WeWalk, which add more functionality such as ChatGPT integration and live public transport information WeW [2024].

These innovative improvements to the traditional white cane tend not to be widely adopted by BVI individuals. An interesting study by Rezylle et al., Milallos et al. [2022], actually explored the potential reasons why these smart devices have such low adoption rates. They found that only 19.15% of the individuals they interviewed had used a smart cane before, with most users being tech-experienced or generally comfortable dealing with technology. Most interestingly, they found that many of their participants did not want to use a smart cane due to the social stigma of using a white cane. Many felt that using a white cane made them feel 'different' from others, and that a smart one, with a typically more unusual and bulky design, could attract more attention than a white cane would. Beyond this, many participants thought it was complicating an issue that already had existing solutions, and that the benefits did not outweigh the added cognitive load, lack of portability and cost of the smart canes currently available. As well as this, many BVI individuals felt they could not trust smart canes, due to a lack of BVI individuals in trials for products, a lack of specific training for the devices, and a lack of endorsement from orientation and mobility (O&M) instructors, who are widely trusted in BVI communities.

Non-cane assistive technologies often involve eye substitution, typically, in the form of glasses, a headset or a helmet, the most prevalent of these being Microsoft's work modifying a HoloLens for BVI individuals Hol [2024]. This provided valuable information about the user's surroundings, including the individual known people around the user. This allowed BVI individuals to identify and recognise their friends and family in the environment, when they would have otherwise struggled to do so.

While assistive technologies for the BVI individuals have made significant strides in recent years, the scope of assistive technology extends beyond that of tools for the blind and visually impaired. For instance, PARO, an assistive pet robot, helps stimulate social interaction and communication for people with dementia Yu et al. [2015]. This takes the form of a seal robot that would respond to different stimuli and voices to improve mood, social interaction and neuron activity. Though outside reception on PARO was mixed, with some considering using a doll on the elderly as an infantilising and demeaning Sharkey and Wood [n.a], there is scope for assistive robots to help with more than navigation for BVI individuals.

Chapter 3

System Design

3.1 Requirements Gathering

To develop a robotic guide dog that genuinely meets the needs of visually impaired users, it is imperative to ground the design process in lived experience. This study employs qualitative methods to elicit first-hand insights from a BVI individual, focusing on their current navigation strategies and main challenges with assistive technologies. We contacted Guide Dogs UK, Sight Scotland, and the RNIB for potential interviewees, but only the RNIB responded. These insights were collected through an interview with a BVI individual to identify their preferences and unmet needs regarding assistive technology for the visually impaired.

3.1.1 Interview

A semi-structured interview was conducted with a BVI individual to explore their lived experiences, current use of assistive technologies, and expectations for robotic navigation aids. The participant, a volunteer with the RNIB, whose role involved teaching other visually impaired persons Braille and spreading awareness about its importance, provided valuable insights into the daily realities of navigating with visual impairment. The interview centred on the following key areas:

3.1.1.1 Current Navigation Methods

The participant predominantly relies on a sighted guide to navigate their surroundings, emphasising the limitation of alternative methods. While the participant occasionally uses a long cane, they noted their inefficacy in providing full situational awareness, as it is predominantly helpful at detecting ground-level obstacles. They emphasised a long cane's limitation in terms of situational awareness, and reaffirmed that they did not feel as though they had a clear mental picture of their surroundings whilst using the cane. This partial awareness, coupled with fatigue, led them to favor human guidance due to verbal reassurance, adaptability, and the mental load of remembering where in the environment they were, could be left to the human guide. Due to this, and other associated factors such as general fatigue, they stated they prefer to rely on sighted guides as they can offer verbal affirmations that they are safe and not lost in their surroundings. As a result, the participant remains largely housebound outside of RNIB activities, constrained by the limited availability of sighted guides and a reluctance to impose upon family members.

The participant also expressed an extreme reluctance to use guide dogs, due to past experiences with animals and concerns about the dog's ability to follow commands consistently. They emphasised that whilst a human can communicate and provide reassurance, Guide Dogs require an intrinsic trust, especially when they could have unpredictable behaviour. The participant expressed that if they wanted to be more independent, they would like a technology that helped them feel safe and more independent in navigating outside of their usual environment; however, they thought this would be difficult for them at the current time in their life.

3.1.2 Other Assistive Technologies

The participant, however, currently uses various assistive technologies, relying heavily on screen readers and specialised mobile apps to assist her with daily tasks. Much of this technology compensates for the widespread absence of Braille in modern society; without Braille signage on doors, signs, and the other textual elements often taken for granted by sighted individuals, the participant would not be able to navigate daily life. Thus, the participant depends on assistive apps and screen readers to interpret written information in daily life.

Interestingly, the participant noted that some of the most valuable technologies they use day-to-day are not explicitly marketed as assistive tools, but rather, as everyday household devices with accessible design features. For instance, the participant frequently relies on a one-cup kettle, a hot-water dispenser that will dispense the perfect amount of water for one cup and allow them to hear when their cup is filled, preventing accidental burns. This highlights a crucial point: well-designed, intuitive technology benefits not just those with visual impairments, but society as a whole. However, when considering the design of new technologies, accessibility is not always at the forefront of the designer's mind, with many technologies opting to include the bare minimum regarding accessibility rather than considering the full scope of who their product could benefit. The participant reaffirmed this in their interview, mentioning how few accessible technologies there are, and how often the ones made do not fully consider the needs and requirements of the user base. Accessibility should not be an afterthought but an inherent aspect of good design, ensuring that assistive functionality is seamlessly integrated into mainstream products.

However, the participant also expressed frustration with the many assistive technologies that rely on touchscreen interfaces, which lack tactile feedback and can make navigation difficult. A recurring theme throughout the discussion was the importance of clear communication through verbal or tactile feedback, with auditory cues playing a fundamental role in the participant's interaction with their environment.

3.1.3 Ideal Guiding Technology

When asked about their ideal guiding technology, the participant described a system offering verbal guidance, alerting them to obstacles, people and their surroundings. This aligns with a desire for more situational awareness, including alerts for environmental features and verbal navigation similar to the functionality of a navigation device such as a SatNav or Google Maps. However, the participant expressed an extreme dislike for vibrations, stating that haptic feedback can quickly become overwhelming, especially when a machine is buzzing and you are unaware of why, emphasising a strong preference for verbal feedback. Furthermore, they expressed a need for a physical guide they could hold onto for stability and reassurance, while stressing the importance of security and reliability in any system they consider using. The participant envisioned a technology that was more than just a guide dog substitute; it was a technology that actively enhanced independence, integrating essential functions like screen reading and minimising reliance on a smartphone in public spaces.

3.1.4 Analysis

The participant's insights underscore critical considerations in the design of assistive robotic technologies. Their preference for verbal over haptic feedback highlights the need for auditory-based navigation cues, suggesting constant vibration may lead to an overload of uninformative and meaningless information. Their reliance on a physical guide emphasises that a tangible point of contact is vital for security and reassurance, as feeling grounded in the environment through touch can help the user root themselves in the environment and help build a mental map of the environment. Moreover, their reluctance to trust guide dogs, stemming from concerns about unpredictability and prior experience with animals, raises important questions about the psychological barriers to adopting robotic alternatives. The interview findings indicate that for such a system to be successful, it must offer superior navigational assistance and address deeper concerns of trust, agency, and intuitive usability.

3.1.5 System Requirements

The insights gained from the interview revealed a gap between the current landscape of assistive technologies and the actual needs of visually impaired individuals. Specifically, it became evident that there must be both auditory feedback and physical guidance to ensure the user feels entirely comfortable, safe and secure while being guided. Additionally, the user needs enough situational awareness to make decisions about their movements and form a clear mental map of their surroundings. To meet these needs, a comprehensive robotic guide dog system must address both functional and non-functional requirements, as outlined below:

3.1.5.1 Functional Requirements

1. **Physical Interaction** - The robot must have a sturdy handle that allows the user to physically interact with the robot and feel secure about their direction of travel.

2. **Responsive Movement** - The robot must remain responsive to user movements, providing real-time, secure and adaptive physical guidance.

3. **Environmental Awareness** - The robot must be able to detect objects in the surrounding environment, providing verbal cues when necessary to enhance situational awareness and safety.

4. **Text Recognition for Navigation** - The system should also be capable of identifying and reading out text from signs to aid navigation, especially in unfamiliar environments.

3.1.5.2 Non-Functional Requirements

1. **Consistency and Predictability** - The robot should provide consistent and predictable behaviour to instil user trust.

2. **Efficient Sensor Processing** - The system must process sensor data efficiently to avoid delays in navigation feedback that could undermine the user's confidence in the robot's guidance.

3. **User-Friendly Interface** - The user interface must be intuitive to visually impaired users.

4. **Durability and Reliability** - The design is lightweight, but durable and sturdy enough to provide trustworthy guiding information.

Chapter 4

Robot Implementation

Having described an ideal system architecture (*Chapter* 3), we now evaluate its feasibility through prototyping. Thus, this chapter outlines the implementation of some key sub-systems and their real-world performance.

The diagram below, 4.1, outlines the potential subsystems required for an ideal robotic guide dog. These include an audio processing subsystem to ensure the user is aware of their surrounding environment, a Stereo Vision Subsystem to calculate the depth of obstacles in the environment and track the user's pose, and a path planning subsystem for guiding, whilst maintaining a good position in regard to the user. The diagram below experiments with two different potential path planners, NAV2 and Potential Fields, outlined in 7

4.1 Base Hardware Design

The robotic guide dog was built upon the foundation of the TurtleBot3 Burger mobile robot provided by the University of Edinburgh Informatics MakerSpace. TurtleBots offer a simple, modular robotic base that can be easily integrated with ROS (Robot Operating System) and various open-source software, providing a robust yet accessible framework for developing advanced robotics systems. The base system comprises two Dynamixel XL430 Motors, which provide differential drive as well as PID control, handled by the OpenCR control board. The TurtleBot3 is equipped with a 360° LiDAR module, allowing for Simultaneous Localisation and Mapping (SLAM) and capability for integrated navigation and path planning.

4.1.1 Base Hardware Evaluation

A few motor tests were run to ensure the base TurtleBot3 was functioning as expected and that velocity commands could be successfully sent to the robot. A linear velocity ramp test was run on the robot, testing the robot's joint velocities at $0.00ms^{-1}$, $0.04ms^{-1}$, $0.07ms^{-1}$, $0.11ms^{-1}$ and $0.15ms^{-1}$. The robot accurately tracked all commanded velocities with minimal steady-state error, confirming the motor system would operate as intended for navigation tasks.

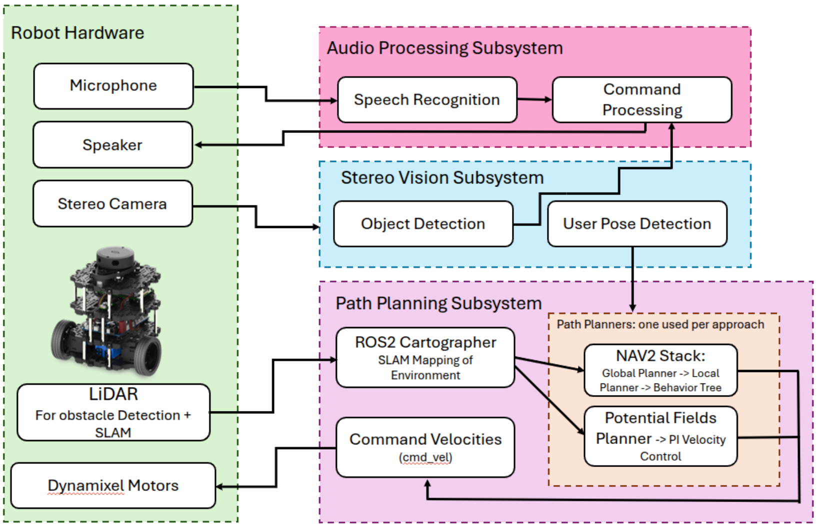


Figure 4.1: Robot Guide Dog Architecture

4.2 Base Software Architecture

The robot was developed using ROS2 (Robot Operating System 2). ROS2 is an open-source framework designed for the development of robotic systems. ROS2 serves as the backbone for handling the communication between the different components of the robot guide dog, providing an efficient means of managing the many capabilities of the robot, such as sensor integration, motor control, mapping, and navigation. ROS2 is built on a modular structure, where each software component, known as a node, performs specific tasks, such as controlling the motors or processing images received from the cameras. Nodes communicate asynchronously using topics, messages, and services, which allow data to flow between the different parts of the system. This allows the whole robot system to be flexible and scalable, as nodes can be modified, added to, or replaced as requirements change. ROS2 Humble was used for this project, due to the version being the latest version of ROS2 with TurtleBot3 compatibility.

4.3 Hardware and Compatibility Upgrades

The base TurtleBot3, provided by the MakerSpace, only contained a Raspberry Pi3 B+ processing unit. Though originally powerful, the Raspberry Pi3 is somewhat outdated. It would not have been able to process LiDAR and Camera data fast enough to provide valuable real-time guidance, whilst also controlling the robot and performing other necessary background tasks. Thus, the central processing unit of the Turtlebot3 was upgraded to a Raspberry Pi 5 (RPi5), to allow for quicker and smoother processing of LiDAR information as well as in-robot processing of image data, avoiding lengthy delays due to bandwidth, thus ensuring efficient sensor processing.

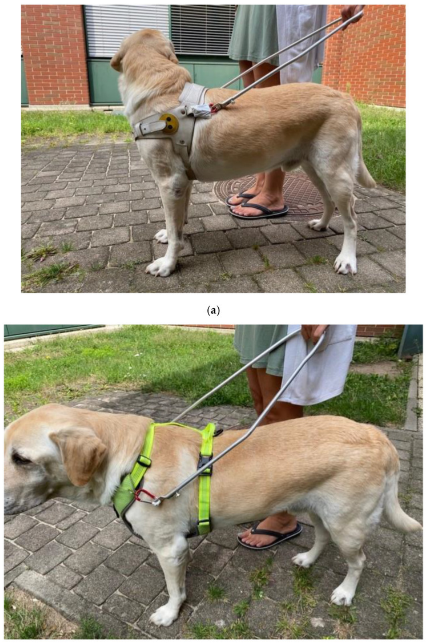
However, several modifications were required to ensure the Raspberry Pi 5 was compatible with the TurtleBot3 framework and ROS2. Since the Raspberry Pi 5 supports only Ubuntu 24.04 and later, while ROS2 Humble is limited to Ubuntu 22.04, Docker created a compatible development environment. The Dockerfile was configured with an Ubuntu 22.04 base image and included the necessary development tools, locales, and ROS2 dependencies. Additionally, it installed all required drivers for the TurtleBot3 hardware, including the LiDAR. Further adjustments were made to ensure the Docker had direct access to the robot's hardware components and sensors by ensuring Docker had access to the dev folder on the system, as well as giving it the proper permissions for use with the RPi5's drivers.

4.4 Physical Interaction

One of the most fundamental aspects of human mobility assistance is tactile interaction, providing users with a stable and intuitive means of guidance. For individuals with visual impairments, physical contact with a guiding tool plays a crucial role in spatial awareness, balance and security; the vast majority of spatial information is conveyed through multiple sensory channels, especially when visual information is lost Giudice [2018] This principle is evident in traditional mobility aids such as the long cane and guide dog harness, offering users a continuous physical connection to their surroundings. The interview findings reaffirm this as a necessity, with the participant expressing a strong preference for a physical guide that could offer stability and reassurance while navigating unfamiliar environments. A custom-designed handle was developed and integrated into the robotic guide dog system to address requirement 1 *(physical interaction)*. This handle serves as the primary interface between the user and the robot, allowing them to maintain a constant point of contact with the robot.

Traditional guide dog handles consist of a handle attached to the dog via its harness. There are two most commonly used handles and harnesses for Guide Dogs. For harnesses, the most frequently used are the Y-harness and the Norwegian Harness, and for handles, the straight handle and the curved handle Weissenbacher et al. [2022]. Whilst few studies have been done on the effectiveness of these handles, Weissenbacher et al. [2022] evaluated the efficacy of both the harnesses and handles on portraying the biomechanics of the dog, particularly the directional force the dog was exerting on its human companion. They found that the shape of the handle did not show significant differences in force portrayed between the dog and the user. Most notably, they suggest that the point of connection between the handle and the harness could significantly influence the transmission of force. For instance, a taller user may end up with a steeper angle of pull, resulting in a change in the required stride length of the user, per distance travelled by the dog. In their study, Weissenbacher et al. [2022] suggested that this be an important consideration when matching the size of dog to it's human companion, however as the robots cannot be matched to their user individually, the difference in height of various users must be considered in a differently. Thus, the handle was modelled after a traditional, straight guide dog handle, with two poles extending to a U-shaped grip the user would hold on to. These poles were made to be thicker than the traditional guide dog handle, to allow for the potential housing of additional electronic components (such as a speaker or lights). The thickness of the poles also allowed for a larger surface area to grip. The handle was split into small pieces, each with a screw-end, so that the various parts of the handle could be screwed together, allowing

Figure 4.2: 3D-Printed Robot Handle



(a)

(b)

Figure 4.3: Side by side comparison of straight and curved handle, retrieved fromWeissenbacher et al. [2022]

for easy modification of the different parts of the handle. However, this also allowed for additional parts to be easily 3D-printed and screwed into the already existing handle, resulting in a cost-effective way of adjusting the handle length for varying user heights. The whole handle doesn't need to be remade in case of breakage, as replacement parts could also be easily 3D-printed and added back into the original handle. The handle was printed out of PLA plastic, which is relatively lightweight and would hopefully allow for more long-term use without arm fatigue, and 3D printed on the Bambu 3D printers. The handle was then attached to the robot using a hinge attachment. This enabled easy adjustment of the pull angle, allowing the user to choose the angle and position to walk with the robot. As the TurtleBot3 Burger is a relatively small robot, these hinge attachments, as well as the entire handle design, went through multiple iterations, to make sure the handle could securely be attached to the robot, as well as allow for each part of the handle to fit together securely, and ensure the user felt secure when the robot was guiding them.

Chapter 5

Stereo Vision

To implement requirements 2 (responsive movement) and 3 *(environmental awareness)*, the robot system needs to be able to identify and place obstacles and the user in the environment. However, in order to achieve this, an active depth estimation for each object/user detected is required. In robotics, most depth estimation is achieved using stereoscopic depth cameras, as seen in Ghosh and Gallego [2024], which come with software packages and on-chip calibration that can quickly and accurately determine depth in indoor and outdoor environments. While monocular depth estimation methods have been proposed, as seen in Bhoi [2019], estimating depth from 2D images remains ambiguous, imprecise and potentially computationally complex. This is since a single camera cannot capture true depth information and can only refer to depth relative to the camera. TurtleBot3's do not natively support depth estimation, with the ROS2 version of the TurtleBot3 having no native support for any camera built into its software architecture. As no stereoscopic depth cameras could be acquired for this project, a stereoscopic solution was developed using two PiCamera3 modules.

5.1 Stereo Camera

Thus, a stereo camera mount was designed and modelled in Fusion 360 to connect the two PiCamera3 modules to the Turtlebot3. This mount was measured and designed so that the camera modules would fit precisely into specific camera slots, eliminating the need for adhesives or screws. After two iterative prototypes, the final mount was 3D-printed on Bambu Lab printers and attached to the robot's second tier, ensuring the LiDAR remained unobstructed. The baseline distance between each camera lens was fixed at 63 mm, a measurement later used to calibrate depth calculations. By rigidly aligning the cameras, the mount guaranteed consistent spatial relationships between the sensors, ensuring accurate coordinate frame transformations in ROS2. The position of the stereo mount was measured relative to the robot's base_link, the coordinate frame position defined as the lowest point of the TurtleBot3's chassis directly beneath its centre of mass. This allowed for the creation of a camera link coordinate frame offset from the base_link to represent the stereo camera's position with respect to the robot. However, to translate depth information from the camera's perspective into actionable navigation

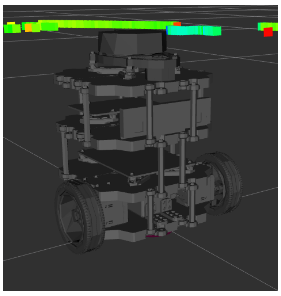


Figure 5.1: Updated Robot Model in the RVIZ2 Visualiser

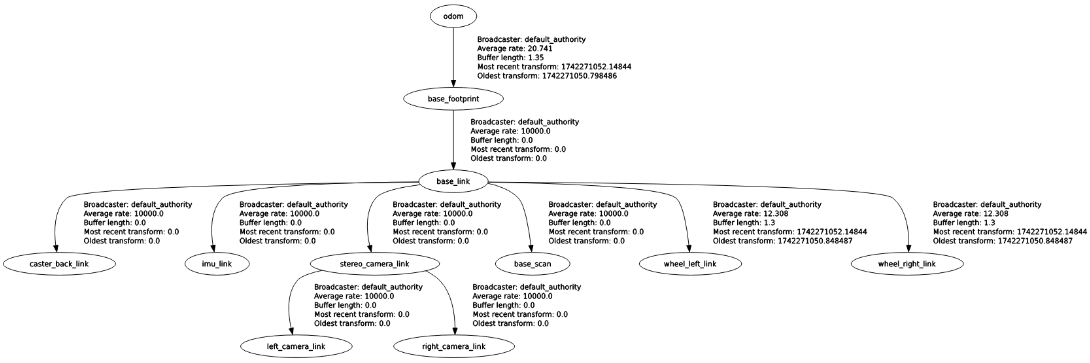


Figure 5.2: Robot Guide Dog's TF Tree

commands and both local and global coordinate information, every component of the robot must be outlined in ROS2's transform tree hierarchy (TF tree).

ROS2's TF tree manages the spatial relationships between all 3D coordinate frames in a robotic system. Each coordinate frame, defined by an origin and three orthogonal axes (X, Y, Z), represents the spatial location of a component of the robot, such as the LiDAR, the wheels, etc, relative to the frame of the other components in the TF tree. This allows each component to reason about the positions and orientations of various entities relative to one another by chaining matrix transformations to align data from all the different components of the robot into a common coordinate system. This tree structure ensures all sensor data and motor commands share a unified spatial context, even as the robot moves. To integrate the stereo camera with the ROS2 framework, camera ROS, an open-source ROS2 package developed by Christian Rauch, was used to publish valid camera information. Whilst Camera ROS provides native support for image capture and basic parameter tuning, significant modifications were required to ensure that both cameras could be interfaced with the Turtlebot3 and Pi5. Camera ROS depends on the libcamera library to provide support for a variety of cameras, including the Raspberry Pi Cameras. However, the base version of libcamera does not support ARM-based systems like the Raspberry Pi. Because of this, a different version of libcamera, explicitly made for the Raspberry Pi, was used instead. This version, however, does not have official support for Raspberry Pi setups that do not use the Raspberry Pi's native Operating System, Raspbian, causing issues with the cameras not being detected by the libcamera software. This was fixed by editing the DockerFile so that the Docker container used by the robot used an older version of Ubuntu 22.04, where the kernel device driver was able to be directly accessed by libcamera, allowing for the package to function as expected.

Camera ROS is designed for single-camera setups. This meant that, for both camera feed topics to be launched in parallel, a launch file had to be designed to containerise both topics. Though Camera ROS provides support for this natively, there was an issue with launching this specifically with the two RPi Cameras on the RPi5, resulting in incorrect pixel formatting and colour channels not being detected. Thus, when launched, the correct settings had to be specified, allowing the colour channels to switch from blue-green-red (BGR) to red-green-blue (RGB), which ROS2 expects.

Camera ROS was then used to publish two camera nodes, left-camera-node and right-camera-node, each publishing their respective camera's video feed on a left and right Camera ROS topic. Both these topics had to be defined with respect to a frame in the TF tree. However, Camera ROS does not provide functionality to change the name of the frame in which its topics are produced. Thus, the Camera ROS package had to be modified so that the frame of each camera topic could be parametrised at launch. This allowed each topic to have its own, individual frame, allowing each camera to have a different spatial position in the TF tree hierarchy, *figure* 5.2.

The two cameras were calibrated using the ROS2 camera calibration package, a tool designed to compute the intrinsic (lens distortion, focal length) and extrinsic (baseline distance, relative orientation) parameters of the camera setup. The calibration involved capturing synchronised image pairs of a 7x9 checkerboard pattern at varying orientations, positions and distances. The camera calibration nodes use corner detection to identify the intersections of the checkerboard in both camera feeds, solving for lens distortion coefficients and translation and rotation quaternions between the two cameras. The resulting calibration matrices were then stored in YAML files.

Camera Calibration allows us to correct minor misalignments and ensure sub-pixel geometric consistency between the two RPi Cameras. It also eliminates any potential barrel/pincushion effects that may occur in the camera image, producing a potentially more accurate depth estimation.

Next, a stereo processing node was created to process the two camera topics and convert them into a disparity map. In this greyscale image, each pixel's intensity represents the horizontal positional difference between matching points in a stereo image pair. The node subscribes to the left and right raw image topics published by Camera ROS and loads the intrinsic and extrinsic parameters produced by camera calibration. Upon receiving synchronised image pairs, the node first corrects lens distortion using OpenCV's undistort function and the calibration information. The rectified images are then converted to greyscale.

OpenCV's StereoSGBM matcher was used to compute the disparity map based on the semi-global matching (SGBM) method proposed in Hirschmüller [2008]. Unlike simpler block matching algorithms, which will attempt to match a small block of pixels in the left image, to the most similar, according to a cost function, block of pixels in the right image and record the horizontal shift (disparity) between them, SGBM

Figure 5.3: Side by side comparison between left camera image and disparity map

will aggregate costs along multiple paths in the image, considering more global pixel neighbourhoods whilst smoothing the disparity map and penalising disparity changes. The resulting disparity map was not as accurate as expected, as can be seen in *figure* 5.3, where lighter regions are supposed to represent closer objects, which is not the case in this example. However, this disparity map was then normalised to an 8-bit range for the depth calculation and visualisation and published to the stereo camera topic, in the stereo camera frame.

The pixel's depth on the disparity map could then be calculated using the focal length and the baseline distance between the two cameras:

$$z = \frac{x - x'}{Bf}$$

Where z is the depth of the pixel, B is the baseline distance between the two camera's O and O', f is the focal length of the camera and $x - x'$ is equivalent to the disparity at pixel x.

5.2 Object Detection

To enable real-time obstacle and pedestrian detection for the robotic guide dog, *requirement 2*, a pre-trained YOLOv8n (You Only Look Once) model was used Jocher et al. [2023]. YOLOv8n is a lightweight convolutional neural network pre-trained on Google's Open Images V7 dataset Kuznetsova et al. [2020]. This dataset contains over 1.9 million images across 600 object categories. The dataset was chosen for its diversity because it spanned a wide range of object categories, particularly labels for critical navigation-related classes, i.e. shoe, person, vehicle, across a wide range of environments.

The YOLOv8n architecture is specifically optimised for speed on less powerful devices such as the Raspberry Pi, allowing for computational efficiency whilst maintaining a high mean average precision. For this application, the pre-trained model was used without fine-tuning, prioritising speed over domain-specific accuracy.

The object depth detector node uses this model. This node subscribes to the left camera image and the disparity map topic produced by the stereo processor node. Object detection is performed on the left camera image when an image and a disparity map are received. If an object of an appropriate class is detected, the object detection node compares the pixel region of the detected bounding box with that of the disparity map. The mean disparity of this central region is then computed. The mean depth of this region is then calculated using the triangulation method outlined above.

5.3 Evaluation

5.3.1 Object Detection

The object detection system was evaluated by measuring its ability to reliably identify everyday household objects (shoes, glasses, clothing, toys) and a person at distances ranging from 50cm to 200cm. Ground truth was established by manually checking object positions in the camera frame and verifying that the entire object was visible in the frame. Each object was left in front of the camera at each depth for 60 seconds. If after 60 seconds, the object had not been detected, it had failed to be detected.

The system demonstrated strong performance for medium and large objects, with 100% detection rates up to 200cm, while struggling with smaller/textureless items. The flowers failed detection at all distances except for 75cm, suggesting lighting or focal issues. The toy cat, the smallest object, showed a sharp cut-off beyond 100cm, highlighting the system's pixel resolution limits. Notably, glasses were undetected beyond 100cm, as well, indicating similar challenges with low-contrast, transparent items. The system, however, demonstrated robust performance in detecting shoes and human subjects across all tested distances, achieving consistent detection rates. This reliability suggests the pipeline can accurately locate a user's pose, provided the target remains within the operational range.

5.3.2 Stereo Vision

The stereo vision system's depth estimation performance was evaluated by measuring its ability to detect the most reliable objects from the tests above, and a person between 50cm and 200cm at known distances. Ground truth was measured via a tape measure from the robot's stereo camera to the closest point of the object to it. As shown in Table 5.4, the system demonstrated varying accuracy levels depending on distance and object type. For instance, the smallest object, the toy cat, proved most accurate at distances closer than 100cm, however, was not detected by the system at distances greater than 100cm. Clothing proved to be the most reliable detection, maintaining relatively consistent error margins across all tested distances, aside from 175cm.

Several key patterns emerge from the recorded data. The reliability of the depth detections decreased significantly for most of the objects beyond 100cm, with the exception of the Toy Bear, which seemed to have its most accurate detections beyond 100cm, maintaining relatively consistent error margins of 5-20cm between the 100 and 175cm tests. Lighting also had a considerable effect on the depth reading, though not

Figure 5.4: Average Error of depth values(cm) versus ground truth for various detected objects. n/d indicates the system failed to detect the object

Object	**Shoe**	**Person**	**Toy Bear**	**Clothing**	**Toy Cat**
50cm	40	17	49	24	9
75cm	2	3	70	56	5
100cm	4	5	25	2	10
125cm	19	5	5	4	n/d
150cm	27	18	4	7	n/d
175cm	6	75	1	17	n/d
200cm	36	53	39	9	n/d

recorded here; these tests were taken during the daytime using natural light, which on average produced more accurate depth detections than when used at night with unnatural light.

However, it remains that despite all depth detections lying in a relatively similar range to the ground truth, the results do not reach near the level of accuracy that would be required in order to make a trustworthy robot guide dog. This could be due to a couple of reasons. The first is the field of view of the cameras used. To make detections, only one camera topic was used, meaning that the field of view was reduced to that of one camera, rather than both. This resulted in two main observations: for small objects such as the toy cat, the object could be accurately detected at close ranges, <=100cm, as the whole object could be seen clearly in the camera view, meaning a more accurate disparity map could be generated. However, at further distances, the Toy Cat was too small for the pixel quality of the PiCam3, meaning that it was too small and unclear for the model to detect. The opposite is true for larger objects like the toy bear, which filled the entire camera view at closer distances, meaning that more accurate disparity maps could be created when the bear was positioned at further distances, resulting in more accurate detections.

Another reason for the poor depth accuracy is the disparity map itself. As can be seen in Figure5.3, the disparity map is not accurate, and does not show darker and lighter regions based on depth as would be expected, producing a practically nonsensical map. This was most likely due to the camera calibration, which tended to be unreliable and gave drastically different parameters on each calibration. However, camera quality would've also had an impact on how the camera was calibrated, and subsequently the SGBM matcher, resulting in the blocky, uninformative map shown in Figure 5.3

Further work could be done to improve the reliability and accuracy of these depth detections, elaborated on in Chapter 8.1. However, it is worth noting that most depth cameras are specially manufactured, often calibrated in the factory, and come with calibration packages, on-board depth processing and increasingly commonly, time-of-flight sensors Whyte [2022]. Thus, commercial cameras tend to have way higher levels of accuracy than most custom-built stereo cameras. It is strongly recommended that any robotic guide dog use one of these cameras instead, but due to their high cost, it was out of the scope of this project.

Chapter 6

Audio Processing

To provide intuitive environmental feedback and enable voice interaction with the robotic guide dog *(Requirement 3, environmental awareness and 4, text recognition for navigation*), the system had to be able to process vocal commands from the user and give vocal feedback based on detected objects in the surrounding environment. This was achieved using two primary components: a voice command service node that processes the detected object information into spatialised verbal cues, and a speech recognition client that enables hands-free voice control. This bidirectional audio interface would allow the user to receive both environmental updates and provide environmental information through natural language.

The TurtleBot3 was augmented with dedicated audio hardware to enable voice interaction capability. A USB-powered speaker module was added to the TurtleBot3, connected via the ALSA driver stack for Linux, configured for low-latency audio output with the Raspberry Pi 5. The Docker container permissions were modified to grant access to the /dev/snd interface, allowing the RPi 5 to access sound devices and achieve real-time playback. For voice input during development, a USB microphone was connected to a companion laptop running the speech recognition node. This was because better speech recognition required a device with an active WiFi connection, which the RPi5 does not have on campus, and due to the lack of accessible USB ports on the RPi5. This also separated computationally intensive audio processing from the robot's onboard resources, providing sufficient audio fidelity and functionality for testing and prototype validation.

The voice command node subscribes to the topic of detected objects published by the object detection system, maintaining a time-filtered buffer of recent object detections. Each object entry includes its semantic class (”person”, ”chair”, etc), depth from the robot, the horizontal offset from the robot and the timestamp. Prevents multiple node threads from accessing the buffer simultaneously, whilst a clean-up timer periodically removes stale detections older than 0.25 seconds, so that only relevant environmental information is kept. This node maintains a ROS2 service interface. This runs constantly during robot runtime, accepting natural language queries and returning an appropriate verbal response based on the current object detections.

Figure 6.1: Audio Processing Node 'find' responses, taken over 5 trials

Trial	Correct Interpretation	Detected Position	Actual Position
1	Yes	Right	Left
2	Yes	Right	Right
3	No	N/A	Left
4	Yes	Left	Right
5	Yes	N/A	Not in frame

The voice command node supports two primary query types: general environment scanning ("what's in front?") and target object searches ("find the chair"). This is done by slicing each command and checking for relevant keywords to ascertain the user's intended command. For environment scanning, the system filters the closest two objects within a 3-meter range and compiles them into a sentence, containing a description of the object approximate location, e.g, "I can see a person 2.3m away, a chair 1.5m away"). Targeted searches identify the closest instance of a requested object class and will respond with both distance and directional information, e.g., "chair detected 1.5m ahead, to the right"). The pyttsx3 TTS engine converts these responses into audible speech at a reduced speaking rate, 140 words per minute, for slower speech rates, which may improve clarity Tivadar [2017].

Complementing the voice response system, a speech recognition client node provides hands-free control using the SpeechRecognition library. The client continuously monitors a microphone input, processing audio through Google's speech recognition API. This speech recognition module is more accurate but requires internet access. An offline CMU Sphinx backend speech recogniser would be defaulted to as a fallback if an internet connection is unavailable. Recognised commands beginning with the wake word "robot" are forwarded to the voice command service, where responses are spoken and logged for debugging. This architecture enables verbal interaction while maintaining modularity between the speech processing node and response generation.

6.1 Evaluation

Two experiments were performed on the audio feedback node, each testing a different side of the system's functionality. The first experiment involved the user requesting the robot to 'find' the toy bear, which was placed at a different location in the robot's field of view. The robot was judged on its audio response, whether it correctly identified that the object was in its field of view, and if it located the object to the left or the right of the robot.

As seen in table6.1, the system could mostly understand the command and respond with the correct object, even when the requested object was not in the frame. The system could not recognise the requested command on the third trial and did not correctly identify the bear. However, the toy bear's reported location was not accurate, with the system only once detecting the toy bear's position correctly. This is due to the y position calculated by the object detection node, which was wildly inaccurate, jumpy

Figure 6.2: Audio Processing Node 'what's in front' responses, taken over three trials

Trial	Detected Objects Correctly	Detected Order Correct
1	No	No
2	Yes	No
3	Yes	Yes

and unreliable, resulting in a practically random position calculation. A significant oversight of this system is its inability to detect an object directly in front of it. Due to the inaccuracies of the y position and the camera's small field of view, most detected objects would be directly in front of the robot anyway, so this did not seem necessary. However, this led to the conclusion that perhaps calculating the left and right position was, in fact, pointless for the current iteration of the robot, with the robot instead telling the user that an object was directly in front of them, providing the user with more valuable and correct information. However, this does not mean that this approach would not be a helpful feature in further implementations of a robot guide dog, as a camera with more accurate positioning information as well as a wider field of view would be able to identify objects of more varied positioning, perhaps leading to more useful environmental information for the user. Further potential implementations of this will be elaborated on in Chapter 8.1

The second experiment run on the processing node involved asking the robot 'what's in front'. A successful response involved correctly recognising the command as well as correctly identifying the two closest objects, with the nearest object being announced first. Three objects: a toy bear, a shoe and a toy cat were placed in the robot's camera frame, at different depths, and after each trial, the positions of each object were swapped. As shown in Table 6.2, the system detected the closest two objects two of the three times, with the system not detecting the toy cat, despite it being placed further away from the shoe in the frame. However, the system failed to announce the objects in the correct order two of the three times, most likely due to incorrect depth information.

Overall, the audio processing node performed well and correctly interpreted the user's command on all but one occasion. However, it is important to note that the node's verbal feedback was unclear and sometimes hard to interpret. This was especially true when hearing depth information, which the node said very quickly and with little clarity. The clarity of the node was improved upon by changing the way in which the robot phrased it's responses, i.e. changing "detected a toy bear, 0.8m away" to, "detected a toy bear, '–' 0.8 '–' meters away", where '–' are explicitly added pauses. The speech rate of the node was also decreased incrementally to attempt to achieve a better level of clarity. However, speeds lower than 140 words per minute sounded more unnatural, especially with the voice of the pyttsx3's TTS engine combined with the general fuzziness of the robot's speaker.

Chapter 7

User Detection and Navigation System

The robotic guide dog's ability to maintain appropriate positioning relative to its user while navigating environments remains a rather under-investigated topic regarding assistive guiding technologies. This functionality was implemented through an integrated two-node system consisting of a modified version of the object detection node described earlier, *Chapter* 5.2, and adaptive navigation code. Together, these components allow the robot to adjust its own movement, based on the user's position and environmental constraints, creating a natural interaction that closely mimics the behaviour of a real guide dog *(Requirement 2, responsive movement)*.

The user detection node uses an edited version of the object detection node described earlier, *Chapter* 5.2. Due to the robot's camera position, the user's lower body was more likely to be detected than the whole person. Thus, detected objects were filtered by the different human detection classes, with weightings towards lower body parts such as ”shoe”, ”footwear” and ”leg”. This would reduce the number of false positives whilst accommodating a wider variety of footwear styles, which the user might wear. In a crowded environment where multiple individuals were detected, the system would default to using the closest detection, which was most likely to be the user. Similarly to the object detection node, the user detection node subscribes to both the left camera feed and the disparity map generated by the stereo processing system. It processes the synchronised image pairs to determine the object's depth relative to the robot.

The system performs coordinate transformations using ROS2's TF2 library to make this positional data useful for navigation. Each detected user position is recorded as a x, y, z, position where x is the depth of the object from the user, y is the horizontal position of the object and z is the height of the object. In this case, the z position has been set to the camera's height due to how close the detected user would be to the floor. Each detected user position is dynamically transformed from the camera frame to the robot's odometry frame, enabling the robot to update the user's position with respect to the world environment and the robot's recorded map position, even as both parties move through the environment. This transformation pipeline also accounts for temporal factors, with timestamps ensuring position estimates are not mismatched due to different timestamps across the multiple components of the robot.

To generate a map of the surrounding obstacles in the environment, ROS2 Cartographer was used, ROS. ROS2 Cartographer is a system that provides real-time simultaneous localisation and mapping (SLAM) in 2D and 3D, and was used by the TurtleBot3 to produce accurate maps of the environment for both path planners.

Integrating user detection with autonomous navigation presented one of the most significant challenges in developing the robotic guide system. Two different navigation architectures were implemented in an effort to create a reliable guiding behaviour that could maintain appropriate positioning relative to the user while navigating through the surrounding environments. Both approaches sought to combine the robot's existing navigation capabilities with the user detection node, though each employed fundamentally different strategies for achieving this integration.

7.0.0.1 ROS2 NAV2 integration approach

The first implementation sought to integrate the existing TurtleBot3 NAV2 navigation stack with user detection. NAV2, Navigation uses behaviour trees to create customisable and intelligent navigation behaviours using many independent modular servers. These servers will each be responsible for one part of the navigation behaviour tree. For instance, one server may control the robot, one may create the path, and another may smooth this path for a nice, more straightforward trajectory. These servers communicate via ROS2 services, each updating based on different behaviour trees, allowing the robot to perform many unique and complex navigation tasks.

To integrate user detection with NAV2, a user navigation node was created to interact with the existing navigation stack through its action server interface. The aim was to combine the robustness of a reliable navigation system with modified logic to position the robot relative to its user. The node continuously calculates appropriate target positions along the global path based on the user's detected pose. These target positions are computed by calculating the distance between the user position and all path points. Then, the node uses a specified target distance that the robot should stay ahead of the user. This could then be used to calculate the target position, by calculating many how many path points ahead the target position should be:

$$s = \lfloor \frac{d}{r} \rfloor$$

Where s, is the number of path points to the target ahead, d is the desired target distance in meters and r is the path resolution, i.e. the distance between each point in the path, assuming path points are evenly spaced. These target positions are then sent to the NAV2 navigation system as sequential local goals. A validation sub-system was implemented to cross-reference the user detections with nearby LIDAR pointcloud data to confirm the presence of an actual user and account for inaccuracies in the depth detection before issuing navigation commands. This involved many transformations between the camera frame, map frame and LIDAR frame to ensure spatial consistency across all sensors. This architecture would benefit from NAV2's built-in path planning and obstacle avoidance capabilities, while theoretically allowing the robot to adjust to the user's movements through dynamic adjustments to the goal. However, it introduced complexity in managing the interaction between the local goals and the final goal pose.

7.0.0.2 Potential Field Approach

The second navigation approach implemented a custom potential field controller that, instead of using the NAV2 stack, provided direct velocity commands to the robot when a user was in frame, based on the sensor inputs and a costmap provided by the ROS2 cartographer node. This implementation focused on reactive control based on a modified version of the potential fields method Ge and Cui [2002]. As outlined in Ge and Cui [2002], the potential field method is a solution to path planning in dynamic environments, mainly used in situations where there is no exact goal the robot needs to 'land' on, like with a guide dog leading its user. The method uses an attractive potential function, a force that effectively 'pulls' the robot toward the goal position, defined as a function of the relative distance between the goal pose and the robot, where the goal pose is a fixed position in space. This attractive potential force is defined as:

$$\mathbf{F}_{attr}(\mathbf{p},\mathbf{v}) = F_{attr_1}(\mathbf{p}) + F_{attr_2}(\mathbf{v})$$

where:

$$\mathbf{F}_{attr_1}(\mathbf{p}) = m\alpha_p||\mathbf{p}_{tar}(t) - \mathbf{p}(t)||^{m-1}\mathbf{n}_{RT}$$
$$\mathbf{F}_{attr_2}(\mathbf{v}) = n\alpha_v||\mathbf{v}_{tar}(t) - \mathbf{v}(t)||^{n-1}\mathbf{n}_{v\ RT}$$

and:

- $\mathbf{F}_{attr_1}$ and $\mathbf{F}_{attr_2}$ are the attractive potential forces in terms of position and velocity
- m and n are exponents for the nonlinearity of the attractive forces for position and velocity, respectively
- α_p and α_v denote the positional and velocity gain
- $\mathbf{P}_{tar}$ and $\mathbf{V}_{tar}$ are the target position and velocity
- $\mathbf{n}_{RT}$ is the unit vector pointing from the robot to the target as seen in figure 7.1
- $\mathbf{n}_{v\ RT}$ is the unit vector denoting the relative velocity direction of the target with respect to the robot, as seen in figure 7.1

However, an opposing repulsive potential function is calculated to ensure that the robot avoids obstacles in its path. This function generates forces that 'push' the robot away from both static and dynamic obstacles, creating a safety margin that adapts to the robot's velocity, as defined in Ge and Cui [2002]. This function combines two components, a radial repulsion force perpendicular to obstacle surfaces and a tangential steering force for dynamic avoidance:

$$\mathbf{F}_{rep}(\mathbf{p},\mathbf{v}) = \begin{cases} 0, & \text{if } \rho_s(\mathbf{p},\mathbf{p}_{obs}) - \rho_m(\mathbf{v}_{RO}) \geq \rho_0 \text{ or } \mathbf{v}_{RO} \leq 0 \\ \mathbf{F}_{rep1} + \mathbf{F}_{rep2}, & \text{if } 0 < \rho_s(\mathbf{p},\mathbf{p}_{obs}) - \rho_m(\mathbf{v}_{RO}) < \rho_0 \text{ and } v_{RO} > 0 \\ \text{undefined}, & \text{if } \mathbf{v}_{RO} > 0 \text{ and } \rho_s(\mathbf{p},\mathbf{p}_{obs}) < \rho_m(\mathbf{v}_{RO}) \end{cases}$$

where the component forces are calculated as:

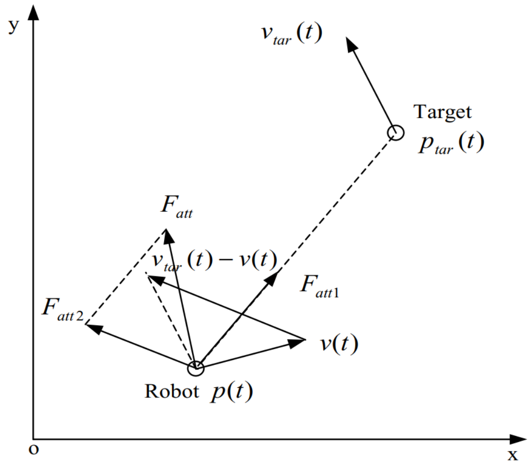


Figure 7.1: Attractive force model in 2D space as seen in Ge and Cui [2002], demonstrating directional pull toward the goal

$$\mathbf{F}_{rep_1} = \frac{-\eta}{\rho_s(\mathbf{p},\ \mathbf{p}_{obs})\ -\ \rho_m(\mathbf{v}_{RO}))^2}\left(1 + \frac{\mathbf{v}_{RO}}{a_{max}}\right)\mathbf{n}_{RO}$$

$$\mathbf{F}_{rep_2} = \frac{\eta\ \mathbf{v}_{RO}\ \mathbf{v}_{RO\perp}}{\rho_s(\mathbf{p},\mathbf{p}_{obs})\alpha_{max}(\rho_s(\mathbf{p},\mathbf{p}_{obs}) - \rho_m(\mathbf{v}_{RO}))^2}\mathbf{n}_{RO\perp}$$

and:

- $\rho_s(\mathbf{p},\mathbf{p}_{obs})$ is the effective obstacle distance with respect to the robot
- $\rho_m(\mathbf{v}_{RO}) = \frac{v_{RO}^2(t)}{2\alpha_{max}}$ is the safety margin, i.e., the distance the robot travels before the robot's velocity reduces to zero.
- $\mathbf{V}_{RO} = \mathbf{V}_{robot} - \mathbf{v}_{obs}$ is the relative velocity vector
- $\mathbf{n}_{RO}$ and $\mathbf{n}_{RO\perp}$ are the unit vectors parallel and perpendicular to $\mathbf{v}_{RO}$
- η controls the repulsive strength
- a_{max} is the robot's maximum deceleration capability

After the calculation of both attractive and repulsive forces, the resultant force can be obtained using:

$$\mathbf{F}_{res} = \mathbf{F}_{attr} + \mathbf{F}_{rep}$$

Using this equation, the attractive force to the goal and repulsive forces from obstacles in the surroundings can be calculated to calculate this resultant force, allowing a mobile robot to move towards the goal without colliding with obstacles.

The second navigation approach expands on this method by introducing two new additions to the resultant force equation. While traditional potential fields methods consider only goal attraction and obstacle repulsion Ge and Cui [2002], this implementation introduces a user-centric force $\mathbf{F}_{user}$ that maintains an optimal guiding distance through adaptive attractive and repulsive potential forces. This function combines two components. The first, is piecewise distance force, which will pull the robot towards the user

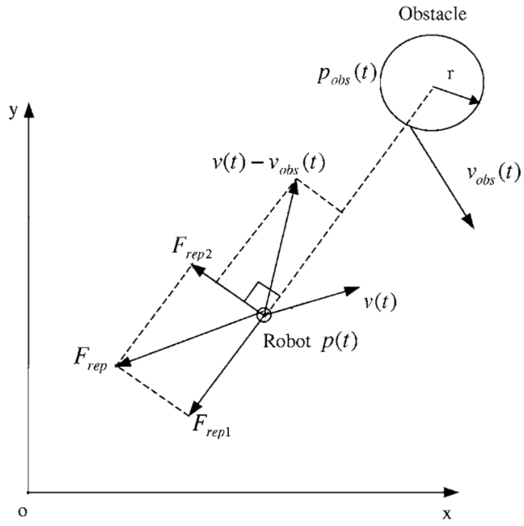


Figure 7.2: Repulsive force model in 2D space as seen in Ge and Cui [2002], directional push away from an obstacle

if the distance between the user and the robot is too large, or repel the robot away from the user, in order to maintain a safe distance in front of them. The second is a velocity tracking force, similar to $\mathbf{F}_{attr_2}$ in Ge and Cui [2002].

$$\mathbf{F}_{user} = \mathbf{F}_{user_1} + \mathbf{F}_{user_2}$$

where:

$$\mathbf{F}_{user_1} = \begin{cases} 2\alpha_{attr} \cdot (u_d - d_{des}) \cdot \hat{\mathbf{n}}_{ru}, & \text{if } u_d > d_{des} \\ 2\alpha_{rep} \cdot |u_d - d_{des}| \cdot (-\hat{\mathbf{n}}_{ru}), & \text{otherwise} \end{cases}$$

$$F_{user_2} = 2\alpha_v \cdot ||\mathbf{v}_u - \mathbf{v}_r|| \cdot \hat{\mathbf{n}}_{\Delta v}$$

and:

- $u_d = ||p_u - p_r||$ is the current distance between the user and the robot
- d_{des} is the desired following distance between the user and the robot
- α_{attr} is the attractive user gain
- α_{repel} is the repulsive user gain
- α_v is the velocity gain
- $\hat{\mathbf{n}}_{ru} = \frac{\mathbf{p}_u - \mathbf{p}_r}{||\mathbf{p}_u - \mathbf{p}_r||}$ is the unit vector from the robot to the user
- $\mathbf{v}_u$ and $\mathbf{v}_r$ are the user and robot velocities, respectively
- $\hat{\mathbf{n}}_{\Delta v} = \frac{\mathbf{v}_u - \mathbf{v}_r}{||\mathbf{v}_u - \mathbf{v}_r||}$ is the unit vector representing the direction of relative motion between the user and the robot.

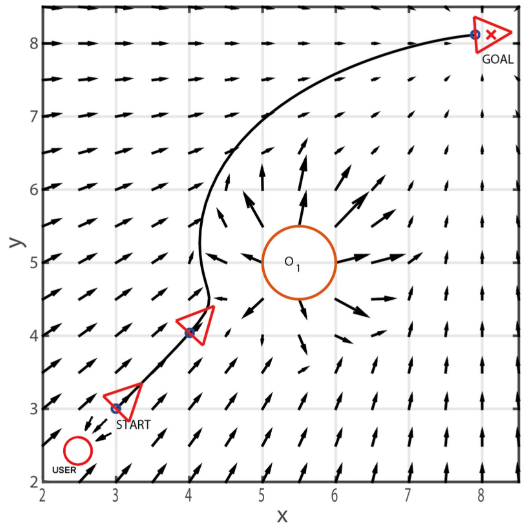


Figure 7.3: Traditional potential field on robot, adapted from Fedele et al. [2018]

The gains α_{attr}, α_{rep}, and α_v balance the responsiveness and stability of the robot's positioning, with respect to the user. The factor of 2 comes from the derivative of the quadratic potential function:

$$U = \alpha \cdot e^2 \text{ where } e = |u_d - d_{des}|,$$

Thus, producing a spring-like force analogous to Hooke's Law:

$$F = -\frac{dU}{de} = -2\alpha \cdot e$$

This formula mirrors that of the linear force-displacement relationship of a physical spring ($F = -kx$), but parametrises stiffness as $k = 2a$ to simplify tuning. We omit the traditional $\frac{1}{2}$ factor in the potential energy definition ($U_{Hooke} = \frac{1}{2}kx^2$) so that the gains directly scale the effective "stiffness" of the virtual spring governing the user-robot interaction.

This allows the robot's motion to adjust to the user's position in a spring-like manner, oscillating toward equilibrium, while still compensating for irregularities in velocities.

The resultant force for the potential fields method can now be defined as:

$$\mathbf{F}_{res} = \mathbf{F}_{attr} + \mathbf{F}_{rep} + \mathbf{F}_{user}$$

However, in order to apply this force to the robot, the force must be converted into velocity commands that can be sent to the motors which control the robot. To achieve this Proportional-Integral (PI) Velocity Control was used Alvarez-Ramirez et al. [1998]

7.0.0.3 Proportional-Integral Control

Proportional-Integral control is a feedback mechanism that adjusts a system's output based on both positional error and the past accumulated steady-state error. For the robotic guide dog, it translates the $\mathbf{F}_{res}$ into smooth velocity commands that the motors can use to drive the robot, ensuring stable and predictable motion. The PI controller operates through two components.

The Proportional Term, P, reacts to the current error between the desired and actual position. This can be written as:

$$\omega_p = K_p^{\theta} \cdot_{error}$$

Where θ_{error} is the angular difference between the robot's heading and the direction of the resultant force. The proportional gain K_p^{θ} determines the strength at which the robot turns toward the resultant force direction.

The Integral Term, I, compensates for steady-state error by accumulating the past error discrepancies over time. This can be written as:

$$\omega_i = K_i^{\theta} \cdot \int_0^t \theta_{error}(\tau) d\tau$$

The integral gain, K_i^{θ} corrects steady-state offsets, which account for real-world environmental discrepancies such as friction, uneven flooring and sensor drift. This ensures that the robot eventually aligns with the desired heading.

For linear velocity, the same principles apply to the magnitude of the resultant force, with the linear velocity scaling proportionally to $||\mathbf{F}_{res}||$.:

$$v = \mathbf{k}_p^d \cdot ||\mathbf{F}_{res}|| + K_i^d \cdot \int_0^t ||\mathbf{F}_{res}(\tau)|| d\tau$$

A more popular error-based control that is commonly used in robotics is Proportional-Integral-Derivative control Willis [1999]. The added derivative term, D, corresponds to the rate of error of change, which damps the system and reduces oscillation. The derivative term was not used for two main reasons: the high-frequency sensor noise from the cameras and LiDAR. This high-frequency noise could be misinterpreted in the derivative term and lead to jerky motion, which would be unsuitable for a guiding robot. The integral term also sufficiently eliminates lingering errors without a potential overshooting risk introduced by the derivative term. Overall, this simplifies system tuning and reduces calibration complexity while allowing for controlled velocity.

To adapt PI control to the robot, velocity was split into its angular and linear components, angular velocity ω_v, and linear velocity v. ω_v was calculated by the angular error θ_{error} between the robot's heading and the resultant force, where:

$$\omega = K_p^{\theta} \cdot \theta_{error} + K_i^{\theta} \cdot \int_0^t \theta_{error}(\tau)\, d\tau$$

The linear velocity is proportional to the magnitude of the resultant force, calculated using the linear velocity equation above 7.0.0.3.

To ensure the PI controller does not send velocity commands greater than the hardware limits of the TurtleBot3, the integral terms were clamped to bounds specified by the TurtleBot3's mechanical limits:

$$\int \theta_{error}\, dt \in \left[-\frac{\omega_{max}}{K_i^{\theta}}, \frac{\omega_{max}}{K_i^{\theta}}\right], \int ||\mathbf{F}_{res}||dt \in \left[0, \frac{v_{max}}{K_i^{d}}\right]$$

where $\omega_{max} = 0.8\ rads^{-1}$ and $v_{max} = 0.15\ ms^{-1}$, ensuring the velocities remain physically achievable by the robot.

These velocities are then converted into ROS2 Twist commands, which are published to the command velocity topic. The robot can then interpret and use these commands to control its motors.

Both navigation approaches were equally unsuccessful at maintaining an appropriate guiding distance from the user. However, this will be elaborated on in Chapter 7.1.

7.1 Evaluation

Unfortunately, as mentioned in Chapter 7, neither implementation of the path planner worked when it came to testing. As well as this, any attempts at tests were significantly stalled due to packet overflow to the Dynamixel Motors on the TurtleBot3, which would prevent new velocities from being sent to the motors until the TurtleBot3 had been switched off for at least an hour. However, the failures of both approaches are outlined below.

7.1.1 NAV2 Evaluation

Tests involved requesting the robot move to a specific location on the robot's map and staying attentive to a following user's detected position while navigating. The robot would move to the requested location, avoiding obstacles, with the robot only starting to move once a detected user was in frame. During the NAV2 test, the user pose was correctly detected, as seen in Figure 7.5. However, the robot would not adjust to the user's position, and would frequently leave the user behind, even when no user position was detected. This was due to the action server and how local goals were sent to the action server. For NAV2's action server to start moving towards these local goals, the global goal was cancelled and stored, and each local goal was instead sent sequentially to the action server. However, even though the action server was able to cancel the global goal, the path planner was not interrupted, and the path was never updated with the local goals. This resulted in the robot moving according to its initial path, and not considering the user's pose. Future work would potentially involve fixing this approach using the Pütz et al. [2018] package. However, this will be elaborated on in Chapter 8.1

7.1.2 Potential Field Evaluation

Tests were similar to the NAV2 approach, where the robot was asked to move to a specific location and stay attentive to the user's detected location. However, the Potential Field approach did not result in the robot moving to the correct position. The user pose was successfully detected, as shown in 7.4, and a map of the surroundings was generated. However, while a user's pose was detected, the robot would take off in a seemingly random direction, lose the user's position, and stop moving. This suggests that the environmental potential field was not being applied to the robot as expected, i.e. 7.3. This made it more attentive to the user's position than with the NAV2 implementation. Thus, due to the constant stopping and starting of the robot's movement, whenever the pose was lost, the robot did not adapt to the user's position as they moved.

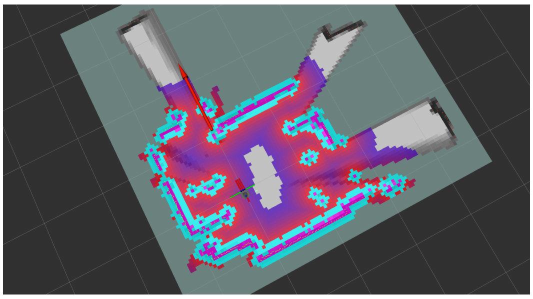


Figure 7.4: Generated map during Potential Field's test, user pose displayed as red arrow

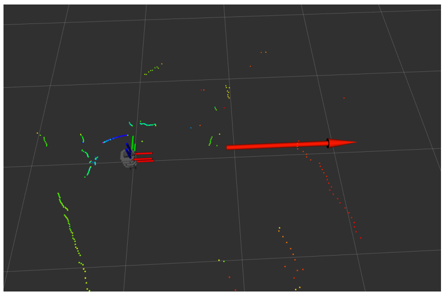


Figure 7.5: Correctly detected user pose, represented by red arrow

Though this approach was much more attentive to the user than the NAV2 approach, it was not a guide that would inspire any sense of trust or reliability in its user. Beyond this, the Potential Fields Method has its limitations when it comes to mobile robotics. Though the spring-like user force, $\mathbf{F}_{user}$, would have been a good way of modelling the robot's position in regard to the user if it had worked, potential fields as a broader method, does not produce the fine-grained paths needed for human navigation, and its not necessarily suitable for environments with many dynamic objects such as a path or a road. However, this will be elaborated on in Chapter 8

Chapter 8

Discussion

The development of assistive robotics demands a delicate balance between technical innovation and real-world usability, with any assistive technology needing to provide a sense of reliability and trust that other developers may not consider, or even need to implement. While the prototype demonstrated functional stereo vision and voice interaction, and despite all its challenges, it was technically able to navigate, its evaluation exposed critical limitations, specifically in reliability, environmental adaptability, and user trust. These shortcomings mirror broader trends in assistive technology, especially where high abandonment rates often stem from gaps between expected performance, practical deployment and user trust.

When measured against the project's original requirements (*Section 1.2* , the prototype achieved partial success:

- **Requirement 1 (Physical Interface)**: The robot handle satisfied structural requirements, providing firm physical support. However, the lack of a reliable path planning system prevented worthwhile user testing, meaning comprehensive tests remain necessary to evaluate the subjective perception of navigational security and directional confidence.
- **Requirement 2 (Responsive Movement):** The robot was responsive to user movements, with the potential fields approach. However, did not dynamically adapt to the user's position or provide secure guidance.
- **Requirement 3 (Environment Awareness):** The robot could detect and identify environmental obstacles, and identify them when requested.
- **Requirement 4 (Text Recognition for Navigation):** The system could provide verbal feedback about the environment. However, this requirement remained mostly unimplemented. However, this functionality could be easily added, as future work, elaborated on in Chapter8.1.

Of the non-functional requirements proposed in 1.2, many were left unaddressed due to the unfinished nature of the prototype and lack of user testing. Notably, sensor processing was highly efficient due to the RPI5 *(Non-Functional Requirement 1)*. Still, without user-testing a complete prototype, we cannot assess whether this would be fast enough for active navigation. However, to properly assess all the non-functional requirements user-testing would be required.

The depth perception results reveal a core tension in robotics, especially assistive robots, that being a compromise between affordability and reliability. While the system achieved varying accuracy, it ultimately did not produce results that would be reliable for any navigation, let alone for guiding. The easiest fix for this, as mentioned in 5.3.2, would be the addition of a depth, or time-of-flight camera, which provides much more accurate and reliable results. However, when so many BVI individuals do not see the point in purchasing expensive assistive technologies Gothwal and Sharma [2023], increasing the price to include such a costly camera may further dissuade a BVI individual from buying that product. However, it remains an essential truth that inconsistencies in performance and reliability do not just erode user trust but also reduce the entire system's safety, which is unacceptable for such important navigation tasks.

The catastrophic failures of both NAV2 and potential field navigation expose how fundamental the user's consideration is to the entire concept of a robot guide dog. While real guide dogs dynamically adjust their pace and path to the location of their handlers, many robot guide dog implementations such as, Lu et al. [2021], do not consider it within their path planning stack. However, effective guide dogs must tightly couple obstacle avoidance with human motion prediction, a feature which proved deceptively difficult to implement. Beyond this, the potential fields method, in itself, is an unreliable path planning method with many limitations when it comes to mobile robotics Pütz et al. [2018]. The potential fields method can lead to trap situations due to local minima, i.e. the robot is pulled into a dead end inside a U-shaped obstacle, is a sub-optimal route, cannot route through closely spaced obstacles, and is prone to many oscillations. These limitations make it particularly useful for environments that are sparse and do not have many dynamic obstacles, such as lunar robotics. However, potential methods would not perform well on pavements or when navigating complex environments like cities or homes. Despite this, applying a potential field to the user is certainly a method of maintaining a position relative to the user, which could be used in future applications, alongside another path planner. However, this will be elaborated on in section 8.1.

Though the robotic guide dog prototype fell short of its original expectation, the development process exemplifies why even theoretically sound navigation methods struggle to find a user base in BVI individuals, often failing to provide a complete, safe experience that would put BVI individuals at ease. Moving forward, this work suggests a shift in focus for the field of assistive technologies, toward an approach that values the experience of the BVI individual over more marketable AI-driven products. It is important that, as an industry, more considerations are given to how we can make our technologies more accessible and trustworthy to the BVI user, investing time into bringing users into the design process, ensuring we create products that do not just facilitate independence in navigation, but in all aspects of a BVI individual's day-to-day life.

It is an indisputable fact that assistive technologies are, and will continue to be, essential to the improvement of society. Their contributions benefit and help BVI individuals navigate the world both securely and comfortably while still inspiring independence in those who feel they are not. However, the failure of this robotic guide dog poses an interesting question: Is an unreliable guide robot better than none at all? The answer is quite clearly not. The robot's performance would not just be unsafe for its users, but would actively harm their confidence in the assistive technologies and potentially in navigating the wider world. As the participant mentioned in the interview 3.1.1, safety and trust were their two most essential requirements, and the fact that they could not trust most forms of navigation significantly influenced their decision not to leave the house without a sighted guide. Beyond this, even functional but impractical robots lead to abandonment Gothwal and Sharma [2023]. Thus, for assistive robotics, reliability is the foundation of user autonomy, where even occasional failures can perpetuate cycles of dependency and erode independence.

8.1 Further Work

The challenges encountered during the development of this prototype have highlighted clear avenues for further work to advance robotic guide dog systems and provide a more cohesive, user-responsive system for assistive guiding technologies. The first would be to improve the accuracy of the depth estimations and the disparity map. This could be achieved either by investing in Depth or time of flight cameras, with on-board depth processing, or by further investigating the calibration and setup of the two cameras on the robot. The latter approach would require a complete exploration of stereo vision, stereo calibration, and the cameras used. However, this would result in a potentially cheaper, accurate way of measuring depth. Another way the stereo vision could be more accurate is by investigating more types of sensor integration. Similar to a time-of-flight camera, an extra IR sensor could be added to the robot to explicitly record the distance between the robot and the detected object, ensuring a more accurate result. As well, the addition of a 3D-LiDAR could exponentially improve accuracy, as well as give more in-depth navigation information through more detailed PointClouds, which would allow for more dynamic navigation and quicker identification of developing hazards in the environment. However, this would once again introduce the problem of cost vs robustness.

Further work would also include the integration of a hybrid navigation architecture. As mentioned earlier 8, potential fields, as a method, suffer from many limitations. However, the idea of a potential field for the user may result in an accurate, user-adaptive navigation system. Future work would not only include fixing these two approaches, but also combining them. This could be done using the Move Base Flex package Pütz et al. [2018]. This package allows for interrupts within the NAV2 behaviour tree, meaning that the robot's path could be interrupted, updated, and changed, potentially allowing for velocities to be updated according to the potential field of the user. However, more research would need to be done on this package to investigate whether this truly is possible. Either way, a hybrid navigation architecture that considers the user's pose is integral to the functionality of a reliable robotic guide dog, a step toward providing the adaptability and attentiveness of a real guide dog.

Future work should enhance the system's audio capabilities by implementing Requirement 4 for text recognition and improving voice synthesis quality. A fully integrated text recognition system would eliminate the need for BVI users to rely on additional devices to interpret environmental text, creating a more seamless and independent navigation experience. Upgrading to a more advanced speech synthesis engine, such as ElevenLabs Fre [2025] could further refine the system's usability by producing more apparent, more natural-sounding responses tailored to user preferences.

Perhaps the most profound lesson from this project lies not in its technical shortcomings but in its isolation from the community it sought to serve. Initially, the intention had been to gather more BVI participants in the hopes of fully representing the wants and needs of the user base. However, this was not possible within this project. It highlights a critical point mentioned by the participant in the interview, that not many BVI individuals get the chance to influence assistive technologies, with many projects being made without full consideration of the BVI community and the best assistive technologies often being those that weren't specifically assistive in the first place. Future work would involve embedding BVI individuals across the vision-loss spectrum in every development stage, as suggested by the BVI participants in Gothwal and Sharma [2023]. As well as this, an actual robot guide dog product should have long-term real-world trials, where users can test with these robots, for long enough to feel comfortable and expose any edge cases. Beyond this, more work would ensure such robots' proper training and advertisement.

Chapter 9

Conclusion

This project aimed to investigate the feasibility of a robotic guide dog prototype, focusing on dynamic, user-centric navigation for blind and visually impaired, BVI individuals. The work was guided by four key requirements derived from an interview with a BVI participant and gaps in existing literature. These were:

1. **Physical Interaction** via a sturdy handle,
2. **Responsive movement** to adapt to user positioning,
3. **Environmental Awareness** through obstacle detection,
4. **Text Recognition for Navigation**, for intuitive communication with the user.

Of these four functional requirements, three were thoroughly investigated, leaving mixed but instructive results:

- A modular 3D-printed handle that was successfully designed and attached to the TurtleBot3.
- A custom stereo camera, which demonstrated partial success in object detection and depth estimation, highlighting the need for commercial-grade sensors.
- A voice interaction system that provided actionable feedback about nearby obstacles, though clarity requires refinement.
- Two navigation approaches, NAV2 integration and potential fields, were implemented, but neither achieved reliable user-following behaviour.. Hardware constraints and algorithmic limitations underscored the complexity of replicating a guide dog's dynamic, adaptive movement.

The prototype's shortcomings, particularly in navigation, highlight how integral it is for functional subsystems to be integrated seamlessly. For BVI users, even minor inconsistencies (e.g., erratic robot movements or delayed depth updates) erode trust, mirroring abandonment trends observed in other assistive technologies. The interview underscored that reliability and predictability are non-negotiable; a system that works ”most of the time” is inadequate for real-world deployment.

Overall, this project was significantly challenging, presenting substantial technical challenges that limited the prototype's final functionality. Developing reliable hardware required persistent effort to address technical issues through repeated rebuilding and recalibration. Despite this, each obstacle encountered provided valuable insights into the complexities of creating effective assistive robotics and into what makes the bond between guide dogs and their handlers so irreplaceable.

As both a Computer Scientist and an admirer of working dogs, the robot's struggles highlighted that true partnership cannot be coded. While it is reassuring that no algorithm can yet replicate the adaptive intelligence of a guide dog. As engineers, we can see this as an invitation to develop assistive technologies that supplement the furry friends that support our society. Though this project has been complex and come with many challenges, it has only strengthened my respect for the important work guide dogs undertake daily; in a world of circuits and code, sometimes the best technology still has a heartbeat.

Bibliography

Ros 2 cartographer — ros 2 workshop documentation. URL https://ros2-industrial-workshop.readthedocs.io/en/latest/_source/navigation/ROS2-Cartographer.html. [Online; accessed 2025-04-03].

University of glasgow - schools - james watt school of engineering - international partnerships - glasgow college, uestc - spring newsletter 2024 - robot guide dog. URL https://www.gla.ac.uk/schools/engineering/international/uestc/newsletter-spring-24/headline_1046424_en.html. [Online; accessed 2025-03-29].

Guide dogs vs. white canes: The comprehensive comparison – clovernook. https://clovernook.org/2020/09/18/guide-dogs-vs-white-canes-the-comprehensive-comparison/, 2024. Accessed: 2024-10-15.

Using ai, people who are blind are able to find familiar faces. https://news.microsoft.com/source/features/diversity-inclusion/project-tokyo/, 2024. Accessed: 2024-10-18.

Product – wewalk smart cane, 2024. URL \url{https://wewalk.io/en/product/}. Accessed: 2024-10-18.

Free text to speech & ai voice generator — elevenlabs, 4 2025. URL https://elevenlabs.io/. [Online; accessed 2025-04-05].

AFB. Why some people choose not to use a guide dog — american foundation for the blind. https://afb.org/role-guide-dogs-joint-afb-gdb-report/why-not-use-guide-dog, 2024. Accessed: 2024-10-14.

Jose Alvarez-Ramirez, America Morales, and Ilse Cervantes. Robust proportional-integral control. *Industrial & engineering chemistry research*, 37(12):4740–4747, 1998.

Sayaka ARATA, Yukihide MOMOZAWA, Yukari TAKEUCHI, and Yuji MORI. Important behavioral traits for predicting guide dog qualification. *Journal of Veterinary Medical Science*, 72(5):539–545, 2010. doi: 10.1292/jvms.09-0512.

Isaac Attia and Daniel Asamoah. The white cane, its effectiveness, challenges and suggestions for effective use: A case of akropong school for the blind. *Journal of*

Education, Society and Behavioural Science, 2020. URL `https://papers.ssrn.com/sol3/Delivery.cfm`.

Amlaan Bhoi. Monocular depth estimation: A survey, 2019. URL `https://arxiv.org/abs/1901.09402`.

Audun Brunes, Helle K. Falkenberg, Inger C. Berndtsson, and Trond Heir and. Use and underuse of mobility aids in individuals with visual impairment: a cross-sectional study of a norwegian sample. *Disability and Rehabilitation: Assistive Technology*, 19(2):266–272, 2024. doi: 10.1080/17483107.2022.2081735. URL `https://doi.org/10.1080/17483107.2022.2081735`. PMID: 35713634.

Tzu-Kuan Chuang, Ni-Ching Lin, Jih-Shi Chen, Chen-Hao Hung, Yi-Wei Huang, Chunchih Teng, Haikun Huang, Lap-Fai Yu, Laura Giarré, and Hsueh-Cheng Wang. Deep trail-following robotic guide dog in pedestrian environments for people who are blind and visually impaired - learning from virtual and real worlds. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5849–5855, 2018. doi: 10.1109/ICRA.2018.8460994.

Peter Craigon, Pru Hobson-West, G. England, Chantelle Whelan, Emma Lethbridge, and Lucy Asher. "she's a dog at the end of the day": Guide dog owners' perspectives on the behaviour of their guide dog. *PLOS ONE*, 12:e0176018, 04 2017. doi: 10.1371/journal.pone.0176018.

Wafa Elmannai and Khaled Elleithy. Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions. *Sensors*, 17 (3), 2017. ISSN 1424-8220. doi: 10.3390/s17030565. URL `https://www.mdpi.com/1424-8220/17/3/565`.

Giuseppe Fedele, Luigi D'Alfonso, Francesco Chiaravalloti, and Gaetano D'Aquila. Obstacles avoidance based on switching potential functions. *Journal of Intelligent Robotic Systems*, 90, 06 2018. doi: 10.1007/s10846-017-0687-2.

GDFB. Guide dogs annual report and accounts 2023. 2023.

Shuzhi Sam Ge and Yun J Cui. Dynamic motion planning for mobile robots using potential field method. *Autonomous robots*, 13:207–222, 2002.

Suman Ghosh and Guillermo Gallego. Event-based stereo depth estimation: A survey, 2024. URL `https://arxiv.org/abs/2409.17680`.

Nicholas A Giudice. Navigating without vision: Principles of blind spatial cognition. In *Handbook of behavioral and cognitive geography*, pages 260–288. Edward Elgar Publishing, 2018.

Lisa Maria Glenk, Lucie Přibylová, Birgit Ursula Stetina, Sami Demirel, and Karl Weissenbacher. Perceptions on health benefits of guide dog ownership in an austrian population of blind people with and without a guide dog. *Animals (Basel)*, 9(7):428, July 2019.

Vijaya K Gothwal and Sujata Sharma. What are the reasons for abandonment of low vision devices prescribed in a large tertiary eye care centre? *Ophthalmic Physiol. Opt.*, 43(1):17–24, January 2023.

Marion A Hersh and Michael A Johnson. Disability and assistive technology systems. In *Assistive Technology for Visually Impaired and Blind People*, pages 1–50. Springer London, London, 2008.

Heiko Hirschmüller. Stereo processing by semi-global matching and mutual information. *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(2):328–341, February 2008. URL https://elib.dlr.de/55367/.

Alyssa Hughes. Peoplelens: Using ai to support social interaction between children who are blind and their peers - microsoft research, 3 2022. URL https://www.microsoft.com/en-us/research/blog/peoplelens-using-ai-to-support-social-interaction-between-children-who-are-b [Online; accessed 2025-03-29].

IAPB. Projected change in vision loss 2020 to 2050. https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/. [Online; accessed 12-10-2024].

IGDF. Guide dog assessment and training. URL \url{https://www.igdf.org.uk/about-us/starting-a-guide-dog-organisation/guide-dog-assessment-and-training/}. [Online; accessed 13-10-2024].

Susan Kemp-Wheeler Jill Nicholson and David Griffiths. Distress arising from the end of a guide dog partnership. *Anthrozoös*, 8(2):100–110, 1995. doi: 10.2752/089279395787156419. URL https://doi.org/10.2752/089279395787156419.

Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics yolov8, 2023. URL https://github.com/ultralytics/ultralytics.

Benjamin Katz, Jared Di Carlo, and Sangbae Kim. Mini cheetah: A platform for pushing the limits of dynamic quadruped control. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 6295–6301, 2019. doi: 10.1109/ICRA.2019.8793865.

Alina Kuznetsova, Hassan Rom, Neil Alldrin, Jasper Uijlings, Ivan Krasin, Jordi Pont-Tuset, Shahab Kamali, Stefan Popov, Matteo Malloci, Alexander Kolesnikov, Tom Duerig, and Vittorio Ferrari. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *IJCV*, 2020.

Janice Lloyd, Claire Budge, Steve La Grow, and Kevin Stafford. An investigation of the complexities of successful and unsuccessful guide dog matching and partnerships. *Frontiers in Veterinary Science*, 3, 2016. ISSN 2297-1769. doi: 10.3389/fvets.2016.00114. URL https://www.frontiersin.org/journals/veterinary-science/articles/10.3389/fvets.2016.00114.

Chen-Lung Lu, Zi-Yan Liu, Jui-Te Huang, Ching-I Huang, Bo-Hui Wang, Yi Chen, Nien-Hsin Wu, Hsueh-Cheng Wang, Laura Giarré, and Pei-Yi Kuo. Assistive navigation using deep reinforcement learning guiding robot with uwb/voice beacons and semantic feedbacks for blind and visually impaired people. *Frontiers in Robotics and AI*, 8, 2021. ISSN 2296-9144. doi: 10.3389/frobt.2021.

654132. URL https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2021.654132.

Rezylle Milallos, Vinita Tibdewal, Yiwen Wang, Andre Udegbe, and Tae Oh. An exploratory study on the low adoption rate of smart canes. In Margherita Antona and Constantine Stephanidis, editors, *Universal Access in Human-Computer Interaction. User and Context Diversity*, pages 492–508, Cham, 2022. Springer International Publishing. ISBN 978-3-031-05039-8.

Open Navigation. Nav2 — nav2 1.0.0 documentation. URL https://docs.nav2.org/.

Sam Ortiz. Everything you need to know about white canes. https://lhblind.org/everything-you-need-to-know-about-white-canes/, 2024. Accessed: 2024-10-15.

Sebastian Pütz, Jorge Santos Simón, and Joachim Hertzberg. Move Base Flex: A highly flexible navigation framework for mobile robots. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, October 2018. URL https://github.com/magazino/move_base_flex. Software available at https://github.com/magazino/move_base_flex.

Guide Dogs RNIB., TPT. Vi lives: An in-depth understanding of the experiences of people living with visual impairment (vi) in the uk. 2022.

Ami Rokach, David Berman, and Alison Rose. Loneliness of the blind and the visually impaired. *Frontiers in Psychology*, 12, 2021. ISSN 1664-1078. doi: 10.3389/fpsyg.2021.641711. URL https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2021.641711.

Sight Scotland. Types of mobility canes for vision impairment — sight scotland. https://sightscotland.org.uk/articles/information-and-advice/types-mobility-canes-vision-impairment, 2024. Accessed: 2024-10-15.

Amanda Sharkey and Natalie Wood. The paro seal robot: demeaning or enabling? n.a.

Hotimir Tivadar. Speech rate in phonetic-phonological analysis of public speech (using the example of political and media speech). *Jazykovedny Casopis*, 68(1):37, 2017.

Blind Veterans UK. The different types of white cane - blind veterans uk. https://www.blindveterans.org.uk/sight-loss-resources/the-different-types-of-white-cane/, 2024. Accessed: 2024-10-15.

Mohd Helmy Abd Wahab, Amirul A. Talib, Herdawatie A. Kadir, Ayob Johari, A. Noraziah, Roslina M. Sidek, and Ariffin A. Mutalib. Smart cane: Assistive cane for visually-impaired people, 2011. URL https://arxiv.org/abs/1110.5156.

Anna Weissenbacher, Alexander Tichy, Karl Weissenbacher, and Barbara Bockstahler. Influence of two types of guide harnesses on ground reaction forces and step length of guide dogs for the blind. *Animals*, 12(18):2453, 2022.

Refael Whyte. Comparing depth cameras: itof versus active stereo — by refael whyte — chronoptics time-of-flight — medium, 6

2022. URL https://medium.com/chronoptics-time-of-flight/comparing-depth-cameras-itof-versus-active-stereo-e163811f3ac8. [Online; accessed 2025-04-01].

Mark J Willis. Proportional-integral-derivative control. *Dept. of Chemical and Process Engineering University of Newcastle*, 6, 1999.

Anxing Xiao, Wenzhe Tong, Lizhi Yang, Jun Zeng, Zhongyu Li, and Koushil Sreenath. Robotic guide dog: Leading a human with leash-guided hybrid physical interaction. *CoRR*, abs/2103.14300, 2021. URL https://arxiv.org/abs/2103.14300.

Ruby Yu, Elsie Hui, Jenny Lee, Dawn Poon, Ashley Ng, Kitty Sit, Kenny Ip, Fannie Yeung, Martin Wong, Takanori Shibata, and Jean Woo. Use of a therapeutic, socially assistive pet robot (paro) in improving mood and stimulating social interaction and communication for people with dementia: Study protocol for a randomized controlled trial. *JMIR Research Protocols*, 4(2):e45, May 2015. ISSN 1929-0748. doi: 10.2196/resprot.4189. URL http://dx.doi.org/10.2196/resprot.4189.

Appendix A

Participants' information sheet

Participant Information Sheet - Interviews

Project title:	Developing a system to guide and interact with a visually impaired user.
Principal investigator:	James Garforth
Researcher collecting data:	Maia Briggs
Funder (if applicable):	N/A

This study was certified according to the Informatics Research Ethics Process, reference number 830542. Please take time to read the following information carefully. You should keep this page for your records.

Who are the researchers?

The researcher of this study is Maia Briggs, an undergraduate Computer Science student at the University of Edinburgh. The study is being conducted as a part of the Final Year Honours Project of Maia Briggs. This study will be supervised by James Garforth.

What is the purpose of the study?

The purpose of this study is to gather user requirements for a robotic guide dog aid, by gaining insight into the day-to day life of blind and visually impaired individuals, including their experience with or without a guide dog. Additionally, it seeks to understand whether a robotic guide dog would enhance their quality of life and explore the requirements and features they would value in a robotic guiding solution.

Why have I been asked to take part?

You have been invited to participate in this study because of your experience with visual impairment, or role in supporting visually impaired individuals. Your input is essential for understanding navigation challenges and preferences, that will help us design a system that fully meets the needs and requirements of visually impaired users.

Do I have to take part?

No – participation in this study is entirely up to you. You can withdraw from the study at any time, up until the completion of this interview without giving a reason. After this point, it will no longer be possible to withdraw because we are not collecting any data that would allow us to identify you.

What will happen if I decide to take part?

Upon your decision to take part in this study, you will be invited to participate in an interview. will involve questions about your experiences with visual impairment, including the use of guide dogs and navigating daily life. We will also ask what makes for a good navigation experience, as well as any specific qualities you consider important in a guide dog, or a technological alternative.

The interview will take approximately 30 minutes to complete, scheduled at a time and location most convenient for you. Interviews may also be conducted remotely.

Are there any risks associated with taking part?

There are no significant risks associated with participation.

What data are you collecting about me?

The data we collect for our research is completely anonymous: We are not collecting any information that could, in our assessment, allow anyone to identify you. Your signed participant consent form will be kept separately from your responses and destroyed once work on this thesis has been completed, June 2025.

What will happen to the results of this study?

The results of this study may be summarised in published articles, reports and presentations. Your anonymised data may be published and can also be used for future research.

Who can I contact?

If you have any further questions about the study, please contact the lead researcher, **Maia Briggs, email: s2142737@ed.ac.uk**

If you wish to make a complaint about the study, please contact inf-ethics@inf.ed.ac.uk. When you contact us, please provide the study title and detail the nature of your complaint.

Updated information.

If the research project changes in any way, an updated Participant Information Sheet will be made available on http://web.inf.ed.ac.uk/infweb/research/study-updates.

Alternative formats.

To request this document in an alternative format, such as large print or on coloured paper, please contact **Maia Briggs, email: s2142737@ed.ac.uk**

Appendix B

Participants' consent form

Participant number:______________________

Participant Consent Form - Interview

Project title:	Developing a system to guide and interact with a
Principal investigator (PI):	James Garforth
Researcher:	Maia Briggs
PI contact details:	james.garforth@ed.ac.uk

By participating in the study you agree that:

- I have read and understood the Participant Information Sheet for the above study, that I have had the opportunity to ask questions, and that any questions I had were answered to my satisfaction.
- My participation is voluntary, and that I can withdraw at any time without giving a reason. Withdrawing will not affect any of my rights.
- I consent to my anonymised data being used in academic publications and presentations.
- I understand that my anonymised data will be stored for the duration outlined in the Participant Information Sheet.

Please tick yes or no for each of these statements.

1. I agree to take part in this study.

Yes	**No**

Name of person giving consent	Date dd/mm/yy	Signature
Name of person taking consent	Date dd/mm/yy	Signature

THE UNIVERSITY *of* EDINBURGH
informatics