

Enhancing the Travel Experience for People with Visual Impairments through Multimodal Interaction: NaviGPT, A Real-Time AI-Driven Mobile Navigation System

He Zhang

hpz5211@psu.edu

College of Information Sciences and Technology, Penn State University
University Park, Pennsylvania, USA

Rui Yu

rui.yu@louisville.edu

Department of Computer Science and Engineering, University of Louisville
Louisville, KY, USA

Nicholas J. Falletta

njf5380@psu.edu

College of Information Sciences and Technology, Penn State University
University Park, Pennsylvania, USA

Jingyi Xie

jzx5099@psu.edu

College of Information Sciences and Technology, Penn State University
University Park, Pennsylvania, USA

Sooyeon Lee

sooyeon.lee@njit.edu

Ying Wu College of Computing, New Jersey Institute of Technology
Newark, NJ, USA

Syed Masum Billah

sbillah@psu.edu

College of Information Sciences and Technology, Penn State University
University Park, Pennsylvania, USA

John M. Carroll*

jmc56@psu.edu

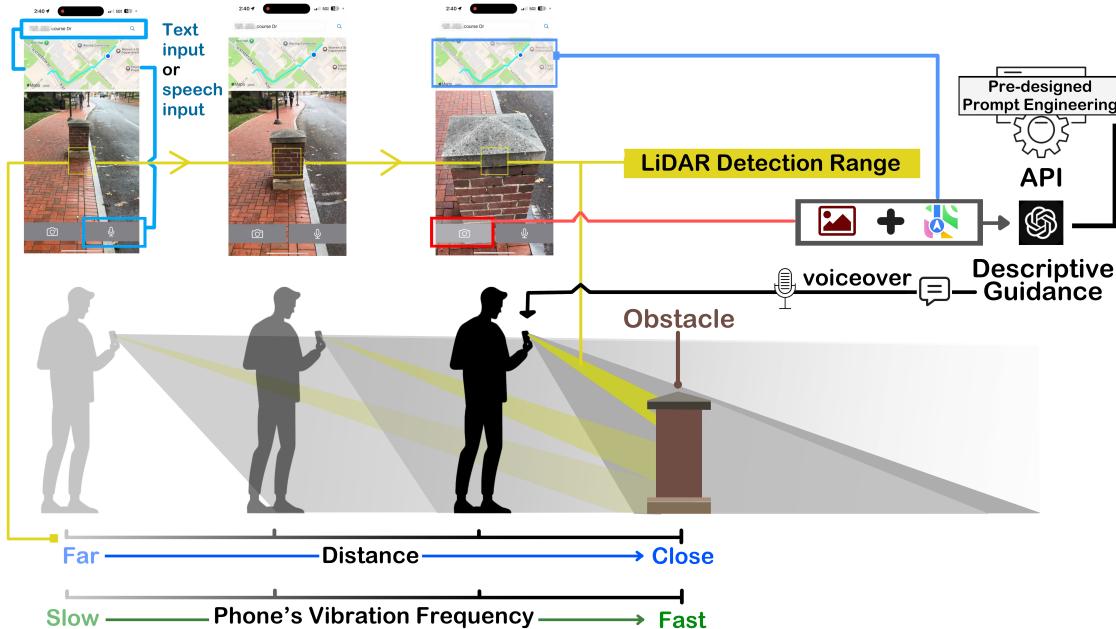
College of Information Sciences and Technology, Penn State University
University Park, Pennsylvania, USA

Figure 1: Workflow of NaviGPT.

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or

republis, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Group '25, January 12–15, 2025, Hilton Head Island, South Carolina, USA

© 2024 Copyright held by the owner/authors. Publication rights licensed to ACM.
ACM ISBN 978-1-4503-XXXX-X/18/06
<https://doi.org/XXXXXX.XXXXXXXX>

Abstract

Assistive technologies for people with visual impairments (PVI) have made significant advancements, particularly with the integration of artificial intelligence (AI) and real-time sensor technologies. However, current solutions often require PVI to switch between multiple apps and tools for tasks like image recognition, navigation, and obstacle detection, which can hinder a seamless and efficient user experience. In this paper, we present NaviGPT, a high-fidelity prototype that integrates LiDAR-based obstacle detection, vibration feedback, and large language model (LLM) responses to provide a comprehensive and real-time navigation aid for PVI. Unlike existing applications such as Be My AI and Seeing AI, NaviGPT combines image recognition and contextual navigation guidance into a single system, offering continuous feedback on the user's surroundings without the need for app-switching. Meanwhile, NaviGPT compensates for the response delays of LLM by using location and sensor data, aiming to provide practical and efficient navigation support for PVI in dynamic environments.

CCS Concepts

- Human-centered computing → Ubiquitous and mobile computing systems and tools; Accessibility systems and tools; Contextual design.

Keywords

People with visual impairments, prototype, AI-assisted tool, accessibility, llm, disability, navigation, multimodal interaction, mobile application

ACM Reference Format:

He Zhang, Nicholas J. Falletta, Jingyi Xie, Rui Yu, Sooyeon Lee, Syed Masum Billah, and John M. Carroll. 2024. Enhancing the Travel Experience for People with Visual Impairments through Multimodal Interaction: NaviGPT, A Real-Time AI-Driven Mobile Navigation System. In *Proceedings of The 2025 ACM International Conference on Supporting Group Work (Group '25)*. ACM, New York, NY, USA, 7 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

People with visual impairments (PVI) face challenges in daily life, especially when traveling [11]. They may find themselves in difficult situations due to their inability to effectively grasp their surroundings and changes within them, particularly when in unfamiliar environments, which increases the risk [4]. Even with some assistive tools, such as guide dogs, white canes, tactile paving, and human assistance, a range of issues may still arise, including limitations in usability, range, and interactivity [27]. To reduce the limitations of a single assistive method, it is common for PVI to use multiple assistive tools simultaneously [21]. This combination of assistive technologies can effectively enhance their perception of the environment and improve their daily life experiences, such as the use of both a white cane and tactile paving [33], or a guide dog in conjunction with a white cane [38]. Additionally, multimodal and perception-enhancing methods, such as screen magnifiers and screen readers [23], can effectively improve their perceptual efficacy. In this context, developing more usable assistive tools with multiple capabilities for PVI is essential.

With the development of technologies such as computer vision (CV) [37] and natural language processing (NLP) [20], applications that replace human vision by using electronic cameras or digital conversions to obtain information are becoming increasingly prevalent. These include functions like object recognition, person detection, text extraction, text reading, and voice assistants. In recent years, with the rise of general artificial intelligence and large language models, these models, such as GPT, are being deployed in more practical applications, bringing benefits to PVI. These models allow PVI to access more content through various interactive means. Applications like ChatGPT¹, Seeing AI², Envision AI³, and Be My AI⁴ utilize a combination of image, text, and voice interactions to provide PVI with descriptions of real-world scenes. This efficient, accurate, and detail-rich descriptive capability, combined with user-friendly natural language interactions, offers tremendous support for accessibility for PVI [41].

However, although these applications have made significant contributions to the development of assistive tools for PVI, at present, they still lack design considerations specifically tailored for PVI in terms of functionality and interaction methods, or their interactive capabilities are confined within the scope of these applications, lacking interaction with the external environment and various other needs. Moreover, these applications have insufficient support for certain specific scenarios, such as navigation during travel for PVI.

Taking ChatGPT, one of the most popular LLM applications today, and Be My AI, an application specifically designed for visually impaired users, as examples, researchers observed the challenges these applications face in such scenarios [41]. In the case of ChatGPT, it primarily operates via text-based dialogue (prompt engineering). Although it allows interaction through multimodal data like uploaded images, users often need to manually write prompts, which can be complex and time-consuming [45], in particular for PVI [26]. Even though ChatGPT has introduced voice interaction capabilities, this is limited to a single modality and does not effectively support the navigation needs of PVI.

For the current version of Be My AI (as of Aug, 2024), it is designed to better meet the needs of PVI (through camera and voice interaction) and enhances their independence. However, past researches have reported some shortcomings in its use, particularly issues such as delays, lengthy feedback, and the requirement for users to actively initiate interactions [reference], which are considered disadvantages in navigation scenarios for PVI.

To address these limitations and improve the navigation experience of PVI during daily travel, we integrated the LLM (GPT-4) with Apple Maps and developed a high-fidelity prototype called NaviGPT, providing a novel approach to enhancing the PVI experience. By utilizing real-time map navigation information and offering contextual feedback through location data, the LLM delivers a more dynamic and context-aware experience. Unlike interactions initiated by the user for a specific purpose, this system offers an "introduction" feature based on location, allowing PVI to interact with their surroundings in a smoother and more natural way. In this system, Apple Maps provides key information such as

¹<https://openai.com/chatgpt/>

²<https://www.seeingai.com/>

³<https://www.letsenvision.com/>

⁴<https://www.bemyeyes.com/blog/introducing-be-my-ai>

nearby landmarks, routes, and the user's current position, while the LLM interprets these crucial details and offers visual information through the camera during navigation. This enhances both the independence and safety of PVI during travel.

2 Related Work

2.1 The Current State of Assistive Technologies for PVI

The well-being of PVI has been a major focus for researchers, leading to the continuous development of assistive technologies. Traditional aids such as white canes, guide dogs, and tactile paving have long been used to assist PVI in their life. Currently, updated versions of traditional assistive tools that integrate technology are emerging one after another, such as PVI assistance robots [6] and smart white canes [18, 19]. In addition, with the advent of the internet, mobile technology and wearable device, more advanced solutions have emerged, such as remote volunteer services [39, 43], where sighted volunteers provide real-time assistance, and specialized assisting systems [8] designed to enhance mobility and understanding for PVI. In recent years, the integration of CV and AI has significantly advanced assistive technologies. An increasing number of smart assistive tools are being developed, and in addition, researchers are paying more attention to the collaboration between PVI and these tools and systems, as well as the user experience [40]. In this paper, we mainly focus on navigation tasks for PVI.

2.2 Navigation Systems for the People with Visual Impairments

Researchers have developed various prototypes to support PVI in navigating both outdoors and indoors [30]. These aids typically include two crucial features for independent mobility: obstacle avoidance and wayfinding [29]. Obstacle avoidance ensures that PVI can safely navigate their environment without encountering obstacles, often using traditional methods such as guide dogs and white canes. Wayfinding, in contrast, helps PVI to identify and follow a path to a specific location, requiring an understanding of their surroundings through digital or cognitive maps [36], and accurate localization to track their movement within these maps.

The advent of smartphone-based applications like Google Maps [1], BlindSquare [7], and others has significantly enhanced outdoor navigation using Global Positioning System (GPS) and mapping services such as the Google Maps Platform [2] and OpenStreet Map [3]. However, the accuracy of GPS can falter by up to ± 5 meters [17], which poses challenges, especially in "last-few-meters" navigation [32]. Indoor environments exacerbate these challenges due to poor GPS reception and the absence of detailed indoor mapping [24, 31].

To address these challenges, researchers have suggested integrating GPS with other smartphone built-in sensors like Bluetooth [34] and Near-field communication (NFC) [16], and creating rich indoor maps to capture environmental semantics [13]. Despite the potential of these technologies, they require significant initial investment and ongoing maintenance [5, 14, 28] to be effective and also depend on users carrying additional devices such as IR tag readers [22].

In recent advancements, the application of CV technologies has emerged as a cost-effective approach for enhancing indoor navigation [10, 42]. Using smartphones, CV-based systems can interpret visual cues like object recognition [46], color codes, and significant landmarks or signage [15, 32]. These systems can also process various tags like barcodes, RFID, or vanishing points for better navigation support [12, 25, 35]. However, the reliability of solely using CV for precise navigation for PVI remains insufficient [32]. Our prototype integrates CV, particularly utilizing LiDAR, with a LLM to jointly provide assistance for PVI during their travels.

3 Prototype Design and Implementation

3.1 System Architecture

NaviGPT's architecture is designed to seamlessly integrate various components, creating a robust and responsive navigation system for visually impaired users. The system consists of 7 primary modules: (1) **User Interface (UI) Layer**: A simplified, accessible interface optimized for voice and text inputs. (2) **Navigation Engine**: Built on the Apple Maps API for accurate location and routing services. (3) **LiDAR Module**: Utilize the LiDAR sensor of the device (iPhone 12 Pro and advanced models) to detect the distance information between the camera and the object. (4) **Vibration Module**: Invoke the device's vibration module and use different vibration frequencies to inform the user about the proximity between the device and the object. The closer the object, the higher the vibration frequency; the farther the object, the lower the vibration frequency. (5) **Image Capture Module**: Utilizes the device's camera to capture environmental images. (6) **LLM Integration**: Incorporates OpenAI's GPT-4 for image description, information processing, and content generation. (7) **Data Integration Layer**: Combines inputs from various sources to provide comprehensive navigation assistance.

3.2 User Interface Design

The UI of NaviGPT is designed with simplicity and intuitiveness in mind, catering specifically to the needs of visually impaired users. The interface consists of four main components strategically positioned for easy access (as shown in Figure 3): **Map Interface**: Located at the top of the screen, this simplified map view provides a visual reference for sighted assistants or users with partial vision. **Voice Interaction Button**: Positioned in the bottom right of the screen, this prominent microphone button allows users to easily activate voice commands and receive audio feedback. **Camera Interaction Button**: Placed at the bottom left of the screen, this button enables users to quickly capture images of their surroundings for AI analysis. **LiDAR Detection**: The LiDAR detection feature activates along with the always-on camera, with the detection area being the yellow square at the center. It can measure the distance between the camera and the object selected within the yellow square.

This minimalist design approach ensures that visually impaired users can interact with the system efficiently through touch and voice, reducing cognitive load and enhancing usability.

3.3 LLM Integration with Navigation and Image Data

The core of NaviGPT's intelligent navigation system lies in its unique integration of OpenAI's GPT-4 with navigation (from Apple Maps) and image data (from user's input):

Map Data Processing: Apple Maps API provides real-time location data, routing information, and points of interest (destination).

Context Generation: The system combines map data with photos taken by the camera and submits them to GPT-4 via an API. The map provides rich information for navigation tasks, such as the current location of BVI, the navigation route, and the destination, while the photos provide information about the current environment. This information offers the LLM extensive contextual details, enabling it to provide feedback based on the user's purpose and current situation. **Intelligent Response Generation:** GPT-4 processes the query and generates natural language responses, providing navigation instructions, environmental descriptions, and safety alerts. Since each time of user input is dynamic, GPT-4 does not provide highly formatted feedback like that of traditional navigation systems (e.g., "In 200 feet, turn left"). However, we have used prompt engineering method to enhance the formatting of the LLM's responses. This helps the LLM focus more effectively on handling tasks within the navigation context, allowing it to deliver targeted and efficient feedback.

Response Refinement: The system filters and refines the LLM's output to ensure relevance and accuracy before presenting it to the user. As we mentioned before, we use preset prompt engineering methods to control the output. Unlike some LLM-powered applications such as Be My AI or ChatGPT, NaviGPT does not provide detailed descriptions of the images provided by BVI. This avoids lengthy responses, making navigation, an inherently dynamic task that requires some degree of real-time feedback, more concise and efficient.

3.4 Technical Implementation and Interaction Flow

NaviGPT's workflow (see in Fig. 1) is designed to provide a seamless and intuitive experience:

Initialization: Upon launch, the application will request the necessary permissions (see in Fig. 2), including GPS location (for map navigation), camera (for photo interactions), photo library (for storing travel photos), microphone (for voice interactions), and speech recognition (for converting speech to text). It is important to note that these permissions are required to fulfill the app's interaction and functionality needs, and the developers will not access this data in any way.

Destination Input: Users can input their destination via voice command or text. The speech is converted to text and processed by Apple Maps to extract the address. Apple Maps API is queried to validate the address and create a walking route. This interaction is similar to the process when the user individually uses Apple Maps. After this step, the user will continuously receive general navigation feedback from Apple Maps based on their current location. This feedback will include turn-by-turn directions, real-time updates on their route progress, and alerts for upcoming turns or changes in the path. The navigation will adjust dynamically as the user moves,

ensuring they stay on the correct route, with the system providing appropriate guidance for reaching the destination.

Real-time LiDAR Detection and Dynamic Vibration Frequency's Feedback: The LiDAR will remain active after NaviGPT starts, continuously and in real-time detecting the distance between the objects in the camera's field of view (within the central yellow square) and the mobile device. The device will provide ongoing vibration feedback based on the proximity of the object. A dynamic vibration frequency curve is set according to the distance, with a usable threshold. Specifically, when an object is detected at 10 meters or more, the device will vibrate at the slowest frequency (once every 3 seconds). When an object is detected at 30 cm or less, the device will vibrate at the fastest frequency (5 times per second). Between 10 meters and 30 cm, the vibration frequency decreases as the distance increases, and increases as the distance decreases.

Environmental Data Capture: The user captures their surroundings using the camera button. In this step, when the user wants to explore their surroundings or notices an approaching obstacle through LiDAR detection and vibration feedback, they can quickly take a picture using the camera button in the UI. The photo will be automatically saved to their device's photo album, and further processing will be carried out via an API call to GPT-4 for enhanced insights or contextual understanding.

Context Integration: The system aggregates: a) The captured image. b) Current location (from Apple Maps). c) Destination and next navigation step (from route planning of Apple Maps). This integrated data is combined and then transmitted to GPT-4 all at once.

LLM Processing and Response Generation: GPT-4 processes the multimodal prompt (examples are shown in Fig. 3b, and 3c). It generates a natural language response that includes: (a) Description of the captured image. (b) Current location description. (c) Safety assessment of the immediate environment. (d) Next navigation instruction. (e) Any relevant warnings or additional information. If the captured photo has a poor angle or does not detect content related to navigation (such as in Fig. 3a), NaviGPT will still provide a response to the user's input but will additionally prompt the user to retake a photo that is more suitable for navigation purposes. Notably, in this scenario, the GPT feedback audio will take priority over the general navigation prompts. Once the GPT feedback has finished playing, the regular navigation instructions will resume. Importantly, the navigation functionality will not be interrupted during the photo-taking process or while receiving the GPT feedback.

Continuous Monitoring and Updates: The system continuously updates the user's location. It prompts for new image captures at key decision points or at regular intervals. The process repeats from step 3 to provide ongoing navigation assistance.

4 Advantages of NaviGPT Over Existing Systems

4.1 Comparison with Existing Map Navigation Systems.

NaviGPT, developed based on Apple Maps, goes beyond the standard features of mainstream navigation systems like Google Maps and Apple Maps. While these platforms offer extensive functions,

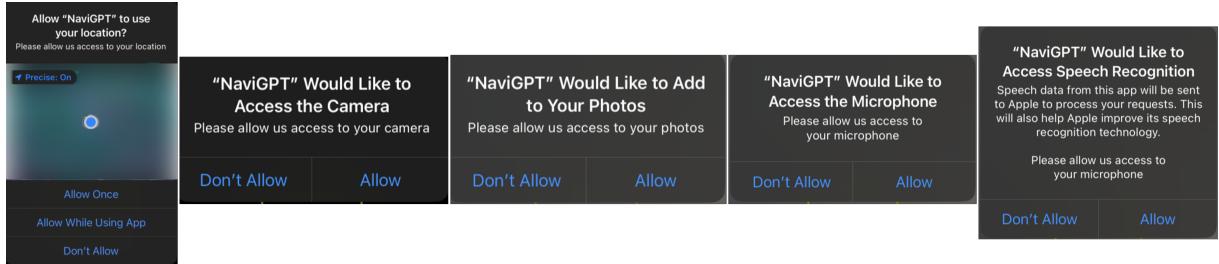


Figure 2: NaviGPT requests permissions.



(a) Example of Bad Photos.



(b) Example of Good Photos.



(c) Example of UI and Good Photos.

Figure 3: Example of Bad and Good Photos

their complexity often overwhelms PVI. Designed primarily for sighted users, these tools require navigating intricate menus and visual prompts, which can hinder PVI users.

NaviGPT, however, is tailored specifically to PVI needs, prioritizing simplicity and accessibility. Instead of relying on visual elements, detailed maps, and text-based instructions, it combines LiDAR, vibration feedback, and contextual responses from a LLM, reducing the dependency on vision-based interactions. This allows

PVI users to focus on navigation without processing complex visual data.

A key advantage of NaviGPT is its simplified UI, offering fast, spoken instructions with minimal distractions. Unlike existing apps that require multiple actions to configure settings or navigate cluttered screens, NaviGPT delivers clear and timely feedback. By providing focused guidance rather than excessive detail, it ensures a smoother, more efficient navigation experience for PVI users, making it easier to access essential information and move safely in dynamic environments.

4.2 Comparison with Existing AI-powered Assistive Systems for PVI.

By comparing NaviGPT with other LLM-integrated applications like Be My AI and Seeing AI, NaviGPT's primary advantage lies in its seamless integration of a navigation system into the LLM interaction flow. It also utilizes LiDAR and vibration feedback to simulate a white cane, providing real-time feedback to PVI. This integration allows users to access key features, AI-based image identification, navigation, and road safety confirmation—without switching between apps, enhancing the overall experience.

Unlike Be My AI, which focuses on detailed image descriptions, NaviGPT is designed for daily travel and navigation, making it more efficient in recognizing and responding to the changing environment. PVI require quick access to safety information, such as detecting obstacles and signage, which NaviGPT provides in real-time. While Be My AI offers more detailed feedback, often exceeding 60 words, it typically requires longer wait and reading out time (over 5 seconds in tests). This level of detail may be unnecessary in dynamic walking scenarios, where rapid feedback is more valuable.

Table 1 compares feedback from Be My AI and NaviGPT. While Be My AI provides more detailed information, NaviGPT's quicker, focused feedback ensures safe and efficient navigation through LiDAR and vibration alerts.

5 Limitations and Future Work

While NaviGPT shows promise, there are several limitations that must be addressed in future iterations. One of the main challenges is the system's reliance on external hardware, such as LiDAR sensors, which are only available on certain advanced devices. This restricts accessibility for users who may not have access to the latest technology [9]. Addressing this issue could involve exploring alternative, more widely available sensors or optimizing the system

Input Picture	Feedback (Words Count)	
	From Be My AI (86 words)	From NaviGPT (28 words)
	<p>The image shows a walkway that appears to be under construction or maintenance. There are orange and white traffic cones placed along the path, with yellow caution tape strung between them, creating a barrier. On the right side, there is a plastic sheet covering something, possibly construction materials or an area under repair. Wooden planks are also visible on the ground near the plastic sheet. In the background, a few people are walking along the path. The ground looks wet, suggesting it might have rained recently.</p>	<p>It's safe to walk but be cautious as there is construction underway with barriers and caution tape indicating a restricted area. Please navigate around the construction area carefully.</p>

Table 1: Example of Feedbacks from Be My AI and NaviGPT.

to work without advanced hardware. If we rely solely on AI and reduce dependence on sensors, this could result in delays, especially in environments with poor internet connectivity. Due to the weak GPS signal inside buildings or underground, the current version of the navigation feature is suitable for outdoor environments only. In addition, privacy concerns related to the continuous use of LiDAR and camera inputs must be addressed. Users may feel uncomfortable with persistent data collection, so future versions of NaviGPT will need to ensure strong privacy protections, including local data processing options and transparent data use policies. Our future work will focus on several key areas to further enhance NaviGPT and fully explore its potential as a navigation aid for PVI, such as adding object recognition and fast feedback features. Additionally, we plan to conduct comprehensive user research to better understand the specific needs and preferences of PVI. These experiments will provide critical insights into how users interact with the application in real-world environments, enabling us to identify potential pain points and improve the system's design, responsiveness, and user experience.

Another important avenue for future development is the integration of more advanced multimodal systems. By further leveraging AI, we aim to bridge the functional gaps between various assistive technologies and consolidate them into a unified platform. This could include expanding the current functionality to incorporate speech-based interactions, real-time environmental mapping, and even predictive analytics that anticipate the user's next actions, movements and emotions based on advanced models and historical data [44]. Such integrations would result in a more intelligent, cohesive system, enhancing the user's ability to navigate complex and dynamic environments.

In addition to improving the technical capabilities, we also plan to explore how the application could adapt to different user contexts, such as indoor navigation in crowded spaces or specialized outdoor environments (e.g., urban vs. rural settings). This would involve customizing feedback based on situational awareness, ensuring the system remains flexible and effective in a wide range of scenarios.

Acknowledgments

This research was supported by the US National Institutes of Health, and the National Library of Medicine (R01 LM013330).

References

- [1] 2021. Google Maps - Transit & Food. Retrieved February 13, 2021 from <https://apps.apple.com/us/app/google-maps-transit-food/id585027354>
- [2] 2021. Welcome to Google Maps Platform - Explore where real-world insights and immersive location experiences can take your business. Retrieved February 13, 2021 from <https://cloud.google.com/maps-platform/>
- [3] 2021. Welcome to OpenStreetMap! OpenStreetMap is a map of the world, created by people like you and free to use under an open license. Retrieved February 13, 2021 from <https://www.openstreetmap.org/>
- [4] Dragan Ahmetovic, João Guerreiro, Eshed Ohn-Bar, Kris M. Kitani, and Chieko Asakawa. 2019. Impact of Expertise on Interaction Preferences for Navigation Assistance of Visually Impaired Individuals. In *Proceedings of the 16th International Web for All Conference* (San Francisco, CA, USA) (W4A '19). Association for Computing Machinery, New York, NY, USA, Article 31, 9 pages. <https://doi.org/10.1145/3315002.3317561>
- [5] Yicheng Bai, Wenyang Jia, Hong Zhang, Zhi-Hong Mao, and Mingui Sun. 2014. Landmark-based indoor positioning for visually impaired individuals. In *2014 12th International Conference on Signal Processing (ICSP)*. IEEE, 668–671.
- [6] Prajna Bhat and Yuhang Zhao. 2022. "I was Confused by It; It was Confused by Me:" Exploring the Experiences of People with Visual Impairments around Mobile Service Robots. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 481 (Nov. 2022), 26 pages. <https://doi.org/10.1145/3555582>
- [7] BlindSquare. 2020. BlindSquare iOS Application. <https://www.blindsightsquare.com/>.
- [8] Roger Boldu, Denys J.C. Matthies, Haimo Zhang, and Suranga Nanayakkara. 2020. AiSee: An Assistive Wearable Device to Support Visually Impaired Grocery Shoppers. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 4, Article 119 (Dec. 2020), 25 pages. <https://doi.org/10.1145/3432196>
- [9] Fernando HF Botelho. 2021. Accessibility to digital technology: Virtual barriers, real opportunities. *Assistive Technology* 33, sup1 (2021), 27–34. <https://doi.org/10.1080/10400435.2021.1945705>
- [10] Andrius Budronis, Darius Plikynas, Povilas Daniūšis, and Audrius Indrusonis. 2020. Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review. *Assistive Technology* (2020), 1–17.
- [11] Fatma El-Zahraa El-Taher, Luis Miralles-Pechuán, Jane Courtney, Kristina Millar, Chantelle Smith, and Susan McKeever. 2023. A Survey on Outdoor Navigation Applications for People With Visual Impairments. *IEEE Access* 11 (2023), 14647–14666. <https://doi.org/10.1109/ACCESS.2023.3244073>
- [12] Wael Elloumi, Kamel Guissous, Aladine Chetouani, Raphaël Canals, Rémy Leconge, Bruno Emile, and Sylvie Treuillet. 2013. Indoor navigation assistance with a Smartphone camera based on vanishing points. In *International Conference on Indoor Positioning and Indoor Navigation*. IEEE, 1–9.
- [13] Wafa Elmamnai and Khaled M. Elleithy. 2017. Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions. *Sensors (Basel, Switzerland)* 17 (2017).
- [14] Navid Fallah, Ilias Apostopoulos, Kostas Bekris, and Eelke Folmer. 2012. The user as a sensor: navigating users with visual impairments in indoor spaces using tactile landmarks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 425–432.
- [15] Giovanni Fusco and James M Coughlan. 2020. Indoor localization for visually impaired travelers using computer vision on a smartphone. In *Proceedings of the 17th International Web for All Conference*. 1–11.
- [16] Aura Ganz, James M Schafer, Yang Tao, Carole Wilson, and Meg Robertson. 2014. PERCEPT-II: Smartphone based indoor navigation system for the blind. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 3662–3665.

- [17] GPS.gov. [n.d.]. GPS Accuracy. <https://www.gps.gov/systems/gps/performance/accuracy/>.
- [18] Jin Sun Ju, Eunjeong Ko, and Eun Yi Kim. 2009. EYECane: navigating with camera embedded white cane for visually impaired person. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, Pennsylvania, USA) (*Assets '09*). Association for Computing Machinery, New York, NY, USA, 237–238. <https://doi.org/10.1145/1639642.1639693>
- [19] Izaz Khan, Shah Khusrro, and Irfan Ullah. 2018. Technology-assisted white cane: evaluation and future directions. *PeerJ* 6 (2018), e6058.
- [20] Diksha Khurana, Aditya Koli, Kiran Khatter, and Sukhdev Singh. 2023. Natural language processing: state of the art, current trends and challenges. *Multimedia tools and applications* 82, 3 (2023), 3713–3744.
- [21] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2022. Tools and technologies for blind and visually impaired navigation support: a review. *IETE Technical Review* 39, 1 (2022), 3–18. <https://doi.org/10.1080/02564602.2020.1819893>
- [22] Gordon E Legge, Paul J Beckmann, Bosco S Tjan, Gary Havey, Kevin Kramer, David Rolkosky, Rachel Gage, Muzi Chen, Sravan Puchakayala, and Aravindhan Rangarajan. 2013. Indoor navigation by people with visual impairment using a digital sign system. *PLoS one* 8, 10 (2013).
- [23] Barbara Leporini and Fabio Paternò. 2004. Increasing usability when interacting through screen readers. *Universal access in the information society* 3 (2004), 57–70.
- [24] Ki-Joune Li and Jiyeong Lee. 2010. Indoor spatial awareness initiative and standard for indoor spatial data. In *Proceedings of IROS 2010 Workshop on Standardization for Service Robot*, Vol. 18.
- [25] Troy McDaniel, Kanav Kahol, Daniel Villanueva, and Sethuraman Panchanathan. 2008. Integration of RFID and computer vision for remote object perception for individuals who are blind. In *Proceedings of the 2008 Ambi-Sys Workshop on Haptic User Interfaces in Ambient Media Systems*, HAS 2008. Association for Computing Machinery, Inc. 2008 1st Ambi-Sys Workshop on Haptic User Interfaces in Ambient Media Systems, HAS 2008 ; Conference date: 11-02-2008 Through 14-02-2008.
- [26] Hugo Nicolau, Kyle Montague, Tiago Guerreiro, André Rodrigues, and Vicki L. Hanson. 2017. Investigating Laboratory and Everyday Typing Performance of Blind Users. *ACM Trans. Access. Comput.* 10, 1, Article 4 (March 2017), 26 pages. <https://doi.org/10.1145/3046785>
- [27] Sabrina A. Panéels, Adriana Olmos, Jeffrey R. Blum, and Jeremy R. Cooperstock. 2013. Listen to it yourself! evaluating usability of what's around me? for the blind. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (*CHI '13*). Association for Computing Machinery, New York, NY, USA, 2107–2116. <https://doi.org/10.1145/2470654.2481290>
- [28] J Eduardo Pérez, Myriam Arregui, Masatomo Kobayashi, Hironobu Takagi, and Chieko Asakawa. 2017. Assessment of semantic taxonomies for blind indoor navigation based on a shopping center use case. In *Proceedings of the 14th Web for All Conference on The Future of Accessible Work*. 1–4.
- [29] Paymon Rafian and Gordon E Legge. 2017. Remote sighted assistants for indoor location sensing of visually impaired pedestrians. *ACM Transactions on Applied Perception (TAP)* 14, 3 (2017), 19.
- [30] Santiago Real and Alvaro Araujo. 2019. Navigation systems for the blind and visually impaired: Past work, challenges, and open problems. *Sensors (Basel, Switzerland)* 19, 15 (02 Aug 2019), 3404. <https://doi.org/10.3390/s19153404> 31382536[pmid].
- [31] Ranga Rodrigo, Mehrnaz Zouqi, Zhenhe Chen, and Jagath Samarabandu. 2009. Robust and efficient feature tracking for indoor navigation. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39, 3 (2009), 658–671.
- [32] Manaswi Saha, Alexander J Fiannaca, Melanie Kneisel, Edward Cutrell, and Meredith Ringel Morris. 2019. Closing the Gap: Designing for the Last-Few-Meters Wayfinding Problem for People with Visual Impairments. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 222–235.
- [33] Laura Šakaja. 2020. The non-visual image of the city: How blind and visually impaired white cane users conceptualize urban space. *Social & cultural geography* 21, 6 (2020), 862–886.
- [34] Daisuke Sato, Uran Oh, Kakuya Naito, Hironobu Takagi, Kris Kitani, and Chieko Asakawa. 2017. NavCog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. 270–279.
- [35] Ender Tekin and James M. Coughlan. 2010. A mobile phone application enabling visually impaired users to find and read product barcodes. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Joachim Klaus, Wolfgang Zagler, and Arthur Karshmer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 290–295.
- [36] Barbara Tversky. 1993. Cognitive maps, cognitive collages, and spatial mental models. In *Spatial Information Theory A Theoretical Basis for GIS*, Andrew U. Frank and Irene Campari (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 14–24.
- [37] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. 2018. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience* 2018, 1 (2018), 7068349.
- [38] Cindy Wiggett-Barnard and Henry Steel. 2008. The experience of owning a guide dog. *Disability and Rehabilitation* 30, 14 (2008), 1014–1026.
- [39] Jingyi Xie, Rui Yu, Kaiming Cui, Sooyeon Lee, John M. Carroll, and Syed Masum Billah. 2023. Are Two Heads Better than One? Investigating Remote Sighted Assistance with Paired Volunteers. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) (*DIS '23*). Association for Computing Machinery, New York, NY, USA, 1810–1825. <https://doi.org/10.1145/3563657.3596019>
- [40] Jingyi Xie, Rui Yu, He Zhang, Sooyeon Lee, Syed Masum Billah, and John M. Carroll. 2024. BubbleCam: Engaging Privacy in Remote Sighted Assistance. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 48, 16 pages. <https://doi.org/10.1145/3613904.3642030>
- [41] Jingyi Xie, Rui Yu, He Zhang, Sooyeon Lee, Syed Masum Billah, and John M. Carroll. 2024. Emerging Practices for Large Multimodal Model (LLM) Assistance for People with Visual Impairments: Implications for Design. arXiv:2407.08882 [cs.HC] <https://arxiv.org/abs/2407.08882>
- [42] Rui Yu, Sooyeon Lee, Jingyi Xie, Syed Masum Billah, and John M. Carroll. 2024. Human–AI Collaboration for Remote Sighted Assistance: Perspectives from the LLM Era. *Future Internet* 16, 7 (2024), 254. <https://doi.org/10.3390/fi16070254>
- [43] Chien Wen Yuan, Benjamin V. Hanrahan, Sooyeon Lee, Mary Beth Rosson, and John M. Carroll. 2017. I Didn't Know that You Knew I Knew: Collaborative Shopping Practices between People with Visual Impairment and People with Vision. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 118 (Dec. 2017), 18 pages. <https://doi.org/10.1145/3134753>
- [44] He Zhang, Xinyang Li, Yuanxi Sun, Xinyi Fu, Christine Qiu, and John M. Carroll. 2024. VRMN-bD: A Multi-modal Natural Behavior Dataset of Immersive Human Fear Responses in VR Stand-up Interactive Games. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. 320–330. <https://doi.org/10.1109/VR58804.2024.00054>
- [45] He Zhang, Chuahao Wu, Jingyi Xie, Yao Lyu, Jie Cai, and John M. Carroll. 2024. Redefining Qualitative Analysis in the AI Era: Utilizing ChatGPT for Efficient Thematic Analysis. arXiv:2309.10771 [cs.HC] <https://arxiv.org/abs/2309.10771>
- [46] P. A. Zientara, S. Lee, G. H. Smith, R. Brenner, L. Itti, M. B. Rosson, J. M. Carroll, K. M. Irick, and V. Narayanan. 2017. Third Eye: A shopping assistant for the visually impaired. *Computer* 50, 02 (feb 2017), 16–24. <https://doi.org/10.1109/MC.2017.36>