UNIVERSITÀ DEGLI STUDI DI MILANO

CORSO DI DOTTORATO
INFORMATICA - XXXVII Ciclo

DIPARTIMENTO DI INFORMATICA
"Giovanni Degli Antoni"

TESI DI DOTTORATO DI RICERCA

Assistive technologies supporting blind and low vision people during navigation

Candidato: **Gabriele Galimberti**

Supervisore: **Prof. Sergio Mascetti**
Co-Supervisore: **Prof. Dragan Ahmetovic**

Coordinatore di dottorato: **Prof. Roberto Sassi**

A.A.
2023/2024

Abstract

The thesis investigates navigation systems for blind and low vision people, proposing three main contributions. First, the thesis proposes two sonification techniques to provide guidance instructions to blind and low vision people during navigation. The proposed solutions are compared with a state-of-the-art technique showing that they improve the precision with which users move while navigating, in particular during rotations. Second, the thesis identifies a set of data management problems that have mostly been overlooked in the scientific literature, such as how to model the space. The contribution consists of designing innovative solutions for these problems and in showing that they significantly improve the navigation experience in terms of safety and clarity. The third contribution explores the problem of screen reader intelligibility in the presence of high levels of acoustic noise in the environment. Noise adaptation techniques are proposed and experimentally evaluated, showing that they significantly improve the screen reader intelligibility while also not distracting the user.

La borsa di dottorato è stata cofinanziata con risorse del
Programma di finanziamento PON REACT-EU
Azione IV.5 – Dottorati su tematiche green
Questa tesi riflette solo i punti di vista e le opinioni degli autori, né l'Unione Europea né la Commissione Europea possono essere considerate responsabili per essi.

The doctoral scholarship was co-funded with resources in the frame of the
PON REACT-EU financing Program
Action IV.5 – Doctorates on green topics
This thesis reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Author's Publications

This thesis is founded upon the publications authored throughout my three-years of doctoral degree.

Journals

- Ahmetovic, D., Avanzini, F., Baratè, A., Bernareggi, C., Ciardullo, M., Galimberti, G., Ludovico, L. A., Mascetti, S. & Presti, G. (2023). *Sonification of navigation instructions for people with visual impairment.* International Journal of Human-Computer Studies, 177, 103057. [3]

- Ahmetovic, D., Galimberti, G., Avanzini, F., Bernareggi, C., Ludovico, L. A., Presti, G., Vasco, G., & Mascetti, S. (2023). *Enhancing screen reader intelligibility in noisy environments.* IEEE Transactions on Human-Machine Systems, 53(4), 771-780. [8]

- Ahmetovic, D., Coughlan, J., Galimberti, G., & Mascetti, S. (2024). *NavGraph: a modular navigation system architecture to enhance blind travelers' navigation experience.* Submitted to Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.

International conferences

- Galimberti, G. (2023, April). *On the independent and sustainable mobility of people with visual impairments.* In Proceedings of the 20th International Web for All Conference (pp. 158-161).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Context

Independent navigation is a major challenge for blind and low vision (BLV) people. Standard Orientation and Mobility (O&M) techniques using the long cane and guide dog are effective in helping BLV people navigate safely [198], and ambient sounds in the environment such as traffic sounds and accessible pedestrian signals (audible signals installed at certain traffic intersections indicating when the walk light is lit) can help them following certain paths and finding points of interest. However, BLV people usually need navigation assistance to follow a path to a desired destination, especially in an unfamiliar environment, as they may lack access to the visual cues that sighted people use to change direction, follow paths and avoid obstacles and other hazards.

Navigation assistance is often provided through smartphone navigation apps such as Google Maps and Apple Maps. These apps supply real-time turn-by-turn directions, useful for travelers visiting an unknown environment. Specifically, they commonly provide visual feedback, or rely on the user's ability to match instructions with the environment's layout through visual confirmation. To address this issue, accessible solutions have been proposed in prior literature. Such services rely on different localization techniques, such as Global Navigation Satellite Systems (GNSS)[1] [106], beacon networks [147], or visual positioning [47], coupled with maps and other information of the environment. Navigation instructions are usually provided through verbal messages associated with a visual representation of the route and the current state of the travel. In particular, the instructions commonly follow a turn-by-turn schema [9], in which the user is instructed to turn at specific decision points and then proceed along a straight path.

Navigation instructions, as well as navigation assistance services, are conveyed through mobile device speakers using accessibility services such as text-to-speech technology to transform text information into audio [1]. Text-to-speech technology is also employed by screen readers [195] which is an accessi-

[1]GNSS is an umbrella term that encompasses different global satellite localization systems, currently: the US-based GPS, the European Galileo, the Russian GLONASS and the Chinese BDS.

bility service commonly available on smartphones. These accessibility services augment touch screen interactions with verbal descriptions of the explored screen content, enabling non-visual access to Graphical User Interface (GUI) elements [103].

1.2 Problem

While several commercial mobile applications offer assistive navigation services for BLV people, current solutions still fall short in providing adequate support throughout the navigation experience. For BLV people, localization and navigation accuracy in navigation assistance services needs to be higher than for people who are not BLV, who can integrate navigation instructions with their own visual inspection of the environment to ensure that the path they are taking is correct. Indeed, prior work has investigated ways to accurately localize and navigate BLV users within an environment [120, 147, 149]. A second challenge is that the navigation assistance needs to be provided through accurate non-visual instructions that the BLV users can follow unambiguously and precisely. The reason is that, even when the localization and the navigation are accurate, a small degree of imprecision in following the instructions can lead the user to wrong or dangerous routes [13]. Thus, prior work has also explored methods to accurately guide BLV users through verbal instructions [167], non-verbal sounds [139] or haptic feedback [19].

Despite the numerous efforts to improve localization and navigation accuracy, significant uncertainty remains about which solutions are most effective across varying environments. The confusion is compounded by the fact that prior research on accessible navigation mainly focuses on localization technology, and on the interactions with users, without considering other important aspects that contribute to bridge these two parts. Indeed, few papers address the impact of the routing algorithm, the data structure model of the environment, and the logic under the choice of navigation instructions. Specifically, the literature lacks a framework that allows direct comparison of the distinct components that underpin navigation systems. The lack of such a framework makes it difficult to systematically compare the performance of different navigation systems and to draw conclusions on how each component should be designed to optimize navigation performance.

Another challenge during mobility is listening to instructions from navigation services in presence of background noise. The intelligibility of the screen reader feedback can be made difficult by the presence of environmental sounds, such as traffic noise or people's voices [205]. Some users address this problem by manually increasing the screen reader volume when needed [1]. This solution only partially mitigates the problem because, in case of sudden noise, the user may not have time to adapt the volume. Also, changing the volume can distract the user from the current activity. Keeping the screen reader at its maximum volume is not a solution as well, because the screen reader feedback could be perceived as intrusive, it may spotlight the user's disability [175], or it could cover important sound cues from the environment, making mobility more difficult or even endangering the user [205]. Another possible solution is to reduce the screen reader speech rate to improve speech understanding in presence of noise [33]. However, this approach also reduces the information throughput of the screen reader which might not be desired.

1.3 Problem decomposition

We propose to structure the problem of BLV navigation assistance into four separate layers: *Localization*, *Routing*, *Guidance* and *Interface* (Fig. 1.1). This structure reflects a *separation of concerns* design approach for the overall problem of navigating a BLV person and is useful for two reasons. First, it is a form of problem conceptualization that can be used to analyze existing solutions and classify them depending on the sub-problem(s) they address (see Chapter 2). Second, it improves software modularization as we show in Section 4.3. We introduce the sub-problems addressed by each layer (see Chapter 4 for more details):

- *Localization*, that is, computing the user's pose (position and orientation), usually in terms of a 2D position and 1D orientation.
- *Routing* is the problem of deciding the *next target* toward which the user should be directed.
- *Guidance* involves selecting the instructions to guide the user toward the next target.
- *Interface*, that is, conveying the guidance instructions to the users, using a combination of audio, haptic, and visual information.

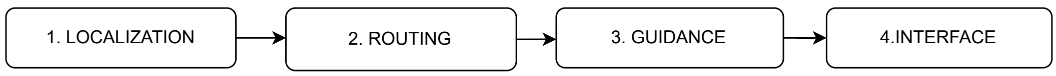


Figure 1.1: Decomposition of the navigation problem.

1.4 Contribution

The thesis involves 3 research activities, resulting in the following contributions. The first research activity addressed a problem related to *Interface* design. During navigation, a small imprecision in following navigation instructions can significantly impact the navigation experience, potentially leading a BLV person into dangerous situations. Thus, our goal was to design an interaction approach to reduce such imprecisions. To achieve this, we propose two sonification techniques that augment navigation instructions with continuous feedback, improving the user's accuracy in following navigation instructions. In a real-world setting, we compare the two sonification techniques and navigation without sonifications, highlighting accuracy improvements when sonifications are used. This research activity is described in Chapter 3 and it's based on the following publication:

- Ahmetovic, D., Avanzini, F., Baratè, A., Bernareggi, C., Ciardullo, M., Galimberti, G., Ludovico, L. A., Mascetti, S. & Presti, G. (2023). *Sonification of navigation instructions for people with visual impairment.* International Journal of Human-Computer Studies, 177, 103057. [3]

The second research activity addressed the design of navigation systems, resulting in the proposal framework called *NavGraph*. Firstly, we structure the navigation problem of BLV people into four layers (Localization, Routing, Guidance, Interface), reflecting a separation of concerns design approach (see Section 1.3). We present a design for two of these layers (*Routing* and *Guidance*), which has been overlooked in the literature, and adopt a solution for the other two layers (*Localization* and *Interface*) previously proposed in literature. We experimentally show the importance of the proposed two layers by comparing a state-of-the-art solution with a proposed system *NavGraph* that implements the designed solution. The results show that *NavGraph* improves the navigation experience and, in particular, is safer, less distracting, and overall more appreciated. This research activity is described in Chapter 4 and it's based on the following publication:

- Ahmetovic, D., Coughlan, J., Galimberti, G., & Mascetti, S. (2024). *NavGraph: a modular navigation system architecture to enhance blind travelers' navigation experience.* Submitted to Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies.

The third research activity is focused on dynamically adapting the screen reader output based on environmental noise. The goal is to improve the screen reader intelligibility, without increasing the distraction from the surrounding soundscape. To this end, we designed and evaluated three audio compensation techniques: **Rate** that reduces the speed of the speech, **Vol** that increases the overall volume of the speech, and **Eq** that adjusts the volume of specific frequency bands of the speech. Through an initial user study, we evaluate screen reader speech intelligibility in the presence of background sound, as well as the distraction from the background sound. The BLV participants commented that making the speech both louder and slower would make it easier to comprehend. Thus, in a second evaluation we included conditions obtained as combinations of speech slowing with the other two compensations (**Rate**+**Vol** and **Rate**+**Eq**). Four of the proposed compensations, in particular those that pair speech rate reduction with volume or equalization adaptation, significantly improve screen reader's speech intelligibility in all the considered scenarios, and the compensations do not have a significant impact on the user's ability to pay attention to ambient sounds. This research activity is described in Chapter 5 and it's based on the following publication:

- Ahmetovic, D., Galimberti, G., Avanzini, F., Bernareggi, C., Ludovico, L. A., Presti, G., Vasco, G., & Mascetti, S. (2023). *Enhancing screen reader intelligibility in noisy environments.* IEEE Transactions on Human-Machine Systems, 53(4), 771-780. [8]

1.5 Methodology

To conduct research on the topic of assistive technologies for BLV people, this thesis adopts methodologies from three distinct research fields: Human-Computer Interaction (HCI), Mobile Data Management, and Music Informatics. Following the well-established research methods in HCI outlined by Lazar et al. [123], we produced empirical, theoretical, and artifact-based contributions. Empirical contributions required to collect data through experimental design, questionnaires, interviews, sensor data acquisitions, observational studies and navigation tests. Empirical data were collected using artifact contributions such as interfaces and developed prototypes. For theoretical contributions, we followed a user-centered design approach for model conceptualizations and design of solutions.

The development of prototypes and artifacts in this thesis necessitated the advancement of specific technologies within the field of Mobile Data Management. In particular, implementing effective software components for mobile navigation, augmented reality techniques, and the integration of data from mobile device sensors required specialized technological innovations. These contributions leveraged and extended existing Mobile Data Management methodologies to support the thesis's goals, underscoring the technical foundation essential for building and refining interactive systems in this research.

The Music Informatics field played a crucial role in the thesis, particularly through the application of sound design techniques for audio feedback interfaces, advanced audio processing methods, and precise audio speaker calibration. These techniques were essential for creating an immersive and responsive user experience, highlighting the significance of integrating Music Informatics methodologies to enhance the effectiveness of the interactive systems developed in this thesis.

All studies in this thesis begin with a user-centered design phase to develop and propose a solution to a research problem. In a manner common to all the studies in this thesis, we developed a unique artifact that served as the basis for the subsequent phases. Following the development of these artifacts, a recruiting phase was conducted to engage volunteer participants from the BLV community. Then, an experimental design was conducted to gather data through observational studies, experimental evaluation [5, 9], comparison with state-of-the-art solutions [4], as well as questionnaires and interviews[1, 201]. For the evaluation of navigation system components, we conducted navigation tests, while for interface components not related to navigation, we conducted listening tests. Once all objective and subjective data were collected, they were analyzed using statistical methods. Each study is approved by the Ethics Committee of the University of Milan.

1.6 Thesis organization

The thesis is structured into several key chapters, each contributing to the exploration of assistive technology solutions for BLV individuals. This structure provides a logical progression from reviewing related literature, through methodological development and innovative solutions, to broader reflections on the impact and future directions of the research.

The second chapter, titled *Related Work*, provides a thorough review of existing research and literature pertinent to the thesis topics. This chapter explores previous studies on navigation systems, sonification techniques, and assistive technologies designed for people with BLV. By analyzing these works, the chapter aims to place the current research within a broader academic context and to identify potential gaps or areas that require further exploration.

The third chapter, *Sonification of Navigation Instructions*, focuses on the process of augmenting navigation instructions with auditory cues and sonification. This chapter describes how sonification can guide a BLV person during the navigation along an unknown path. It also analyzes the effectiveness of different sonification techniques and their potential applications in real-world scenarios.

The fourth chapter, *Modular navigation system architecture*, introduces innovative design for a turn-by-turn navigation system, dividing the architecture of the system in multiple independent layers. The design is aimed to improve the mobility of people with BLV, offering a more efficient navigation with fewer instructions, compared to a state-of-the-art design.

The fifth chapter, *Noise Adaptation Techniques*, presents various techniques for adapting audio feedback in presence of noise, thus improving the intelligibility of screen reader speech in different scenarios for BLV users.

Finally, the last chapter, *Conclusion*, summarizes the key contributions of the thesis. It revisits the main findings and their significance, considering the broader implications for the future of accessible navigation systems. This chapter also highlights the potential for further development in this area, with a focus on improving accessibility for visually impaired individuals.

Chapter 2

Related work

After an initial review of the research background in the field, this chapter describes the related work by identifying, for each relevant paper in the literature, its main contribution(s) according to the four layers identified in Section 1.3. Building on these contributions, we identify and list the main applicative scenarios and use cases enabled by state-of-the-art navigation assistance applications and research work aimed at improving autonomous mobility for BLV individuals. Table A.1 in the Appendix A presents the main contributions in the literature addressing at least one of the four layers.

2.1 Background

2.1.1 Orientation and Mobility by BLV People

Orientation and mobility (O&M) by BLV people has been largely studied in the field of cognitive neuroscience. In these studies, the O&M strategies adopted by blind people [170] and by low vision people [16] are often addressed separately.

For blind people, three principal O&M models have been proposed. The **Spatial model** argues that blind people develop a mental representation of the spatial layout of their surroundings, organized around their body [176]. This representation is based on non-visual sensory information, such as sound, touch, and proprioception. According to the **Route model** [64], blind people learn a route in order to navigate through the environment. During navigation, tactile landmarks and signals acquired through echolocation are used to orient themselves along the route [78]. Based on the **Survey model**, blind individuals create a mental representation of the spatial layout of their surroundings, similarly to the spatial model. Unlike the spatial model, this mental representation is not based on the individual's body, but rather on a global understanding of the surroundings through non-visual cues [184]. Recent work argues that these three different O&M models are not used independently, but they are actually integrated by blind people while navigating [78].

As far as people with low vision are concerned, the role of residual sight is dominant in O&M strategies [16]. Nonetheless, integration of residual sight with other non-visual strategies to acquire information about the environment (*e.g.* echolocation, use of sleep shades to learn strictly non-visual techniques) is essential also for people with low vision [81, 89] and it can be trained in specialized courses [59].

2.1.2 Mobile Device Usage by BLV People

Mobile devices are convenient for accessing information ubiquitously, and, for this reason, they are frequently used by many people with disabilities [104]. They can also be used to provide assistive capabilities, including access to visual cues detected with computer vision techniques [14], support in social networking [20], and in particular mobility and navigation assistance on-the-go [98, 157, 167]. Due to these functionalities, almost the entirety of BLV people in developed countries use mobile devices, and over 80% of them use smartphones for mobility assistance [137].

2.1.3 Sonification

Data sonification can be defined as "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation" [115, 191]. The use of non-verbal sound may be advantageous over speech for reduced cognitive load and linguistic differences [52].

Prior work has proposed taxonomies of sonification functions and techniques [191]. With specific regard to techniques, some of the most commonly employed ones include audification, model-based, and parameter mapping sonification. The latter in particular [82] amounts to representing changes in some data dimensions through corresponding changes in one or more auditory dimensions. One of its distinguishing features is the *mapping function* between data and auditory dimensions. In this thesis, we employ parameter mapping sonification with a simple, one-to-one mapping.

Auditory dimensions can be clustered into five high-level categories: Pitch-related, Timbral, Loudness-related, Spatial, and Temporal. An extensive review [53] surveyed the mapping strategies used in the related literature to associate physical quantities (particularly those related to kinematics, such as distance, orientation, velocity, etc.) to auditory dimensions. The authors showed that pitch is the most often used dimension. An evaluated sonification technique in this thesis (The **IS** sonitifcation presented in the section 3.3.2) adopts temporal auditory dimensions instead. Higher-level musical features may also be used, such as tonality and polyphony; in this case, the term *musification* is more appropriate [46]. Musification, namely the description of information through music objects, may also enhance engagement in addition to mere data representation. Another sonification technique evaluated in this thesis (The **MS** sonification presented in the section 3.3.3) provides an example of musification.

2.2 The *Localization* problem

Many prior works in the field of assistive navigation for BLV people address the problem of how to accurately compute the user's position. This is because assistive technologies supporting the mobility of BLV people require high localization accuracy [9, 40, 80]. This section summarizes the variety of technologies applied to solve the *Localization* problem, emphasizing the main contributions aimed at enhancing localization accuracy, as shown in Table 2.1.

Global Navigation Satellite System (GNSS) is an umbrella term that encompasses different global satellite localization systems, currently: the US-based GPS, the European Galileo, the Russian GLONASS and the Chinese BDS. Among others, commercial applications like *Google Maps* [1], *Apple Maps* [2], *GoodMaps*[3] and *BlindSquare*[4], and research applications [40, 80, 128, 129, 136, 174, 200, 211, 106] adopt this solution. Various techniques have been studied in the literature to enhance the position accuracy of GNSS with hybrid approaches, including the combination of GNSS with visual techniques [40] and inertial sensor positioning [200, 174]. These hybrid approaches enhance navigation robustness in complex scenarios where satellite signals may be disrupted.

Radio-signal-based solutions have been adopted for indoor positioning due to the low accuracy of GNSS systems in indoor environments. WiFi localization [2, 29, 67, 91, 160] leverages the availability of WiFi networks to compute the position of a device within an environment. Other common techniques used for wireless localization use Radio Frequency Identification (RFID) tags [42] or Bluetooth Low Energy (BLE) beacons [15, 41, 51, 84, 94, 111, 147]. Fingerprinting-based solutions have also been proposed to increase the accuracy of radio-signal-based approaches [9, 146, 167]. In case of Ultra Wide Band technologies the Two-Way Ranging method was proposed to compute the distance based on the time it takes for the signal to travel between the transmitter and the receiver [163]. Beacon-based approaches rely on an environment densely populated with transmitter installations, where deploying and maintaining these transmitters represents a significant logistical and operational challenge. To address this issue, crowdsourcing solutions were proposed [79].

Inertial odometry [66, 162, 188, 210] uses the data from Inertial Measurement unit (IMU) sensors (magnetometer, accelerometer and gyroscope) to compute how user's location changes in time through the Pedestrian Dead Reckoning (PDR) [96, 202]. The accuracy of this family of techniques can be increased by using visual-inertial odometry, which combines features extracted from images captured by camera sensors with data from IMU sensors [69, 49, 106, 72].

Since odometry and visual-inertial odometry are subject to drift problems

[1] `https://www.google.com/maps/`
[2] `https://www.apple.com/maps`
[3] `https://goodmaps.com`
[4] `https://www.blindsquare.com`

(position accuracy decreases over time) they are also used in combination with other techniques that provide sporadic position fixes [49, 69]. These techniques can also be conveniently implemented by relying on existing augmented reality libraries [3, 7, 43, 206] and computer vision algorithms [85, 94, 95, 118, 125, 145].

Some papers in the literature use LiDAR scanners to recreate a 3D representation of the environment [85, 118, 164]. In particular, the structure-from-motion (SfM) technique can be used to compute the pose from a point cloud representation of the environment using LiDAR sensors. An example of this approach is used in the Cabot project [85] which use the LiDAR to provide precise environmental mapping and localization in a Simultaneous Localization and Mapping (SLAM) process. Commercial apps such as *Apple Maps* [5], *Google Maps* [6], *GoodMaps* [7], and *NaviLens* [8] have introduced features that use a smartphone's camera and eventually the LiDAR sensor to enhance localization accuracy. This is achieved through visual-inertial odometry, possibly combined with LiDAR sensor [85, 118], datasets of outdoor or indoor environments [164, 49, 151] used to update the position, and GNSS positioning [40] to correct outdoor position based on the GNSS signal.

Many solutions also adopt sensor fusion and filtering techniques to combine multiple positioning sources for a more accurate localization. For example, Kalman filter has been used to combine IMU sensor data with radio-based signal [73]. In particular, Particle Filtering has been used to improve localization accuracy in many different works [49, 70, 87, 91, 146, 162, 188]. Particle Filtering techniques are sequential probabilistic methods used for estimating the state of a system by representing the state distribution with a set of particles, which are updated based on observations and resampled to improve accuracy. Deep learning methods have also been proposed to implement sensor fusion of multiple localization sources [152]. These solutions may also model the environment constraints (*e.g.*, areas where the user is not expected to go) to achieve more robust localization [69, 147, 162].

Table 2.1: Related papers with a proposed solution for the *Localization* problem.

Main Localization techniques	**Paper**	**Proposed approach**
GNSS	[106]	GNSS positioning combined with visual landmarks
	[40]	GNSS positioning combined with visual information
	[174] [200]	GNSS positioning combined with Inertial sensor positioning

[5]https://www.apple.com/maps
[6]googlemaps
[7]https://goodmaps.com
[8]https://www.navilens.com/

Main Localization techniques	**Paper**	**Proposed approach**
Radio signal	[9] [41] [51] [84] [111]	Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons
	[167]	Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons, Semantic features
	[15]	Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons, Inertial sensor positioning
	[146], [147]	Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons, Inertial sensors positioning, Particle filter
	[2]	Combination of WiFi positioning, beacon positioning and Radio Frequency Identification positioning
	[73] [91]	Combination of WiFi Fingerprinting positioning and Inertial sensor positioning
	[79]	Crowdsourcing-based Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons
	[94]	Received Signal Strength Indication (RSSI) of BLE beacons combined with images in Deep Learning using a CNN
	[163]	Received Signal Strength Indication (RSSI) of BLE beacons combined with Ultra Wide Band beacons using Two-Way Ranging
	[42]	Radio Frequency Identification positioning
	[29] [67] [160]	WiFi fingerprinting
Inertial sensors	[210]	Crowdsourcing-based inertial sensors segments, Acoustic-based localization
	[162] [188]	Inertial sensors combined with a Particle filter
LiDAR	[118]	LiDAR scanning, Floormap detection and understanding
	[85]	LiDAR scanning, Scene understanding using CNN
	[148]	LiDAR scanning, Inertial sensor positioning

Main Localization techniques	Paper	Proposed approach
Visual-inertial odometry	[49] [70]	Visual inertial-odometry, Visual landmarks, Particle filter
	[72] [145] [125]	Combination of Visual inertial-odometry with Inertial sensor positioning
	[7] [69]	Visual-inertial odometry using Augmented reality, Visual landmarks
	[95]	Visual-inertial odometry using existing street cameras
	[187]	Visual landmarks
Hybrid positioning	[152]	Scene understanding using CNN, Recurrent Neural Networks for inertial sensors

2.3 The *Routing* problem

Table 2.2 shows the main contributions of relevant papers in the state of the art that propose solutions for the *Routing* problem. In the turn-by-turn navigation paradigm, the **routing** problem is the problem of computing a route, a set of waypoints, from a starting position to a destination. Although routing is not usually described in detail in previous contributions, turn-by-turn navigation paradigm is adopted by many existing solutions [3, 9, 91, 95, 200] Route computation addressing the routing problem adopting a *space representation*, a data structure to model the environment. Beyond space representation, another task associated with the routing problem is the computation of the *next target*, which represents the intermediate waypoint relative to the user position. During the navigation, the actual trajectory may deviate from the planned route by more than a set threshold distance. This deviation is a potential problem called *veering* which can be solved with a veering compensation approach.

Space modeling

A common approach in turn-by-turn navigation is to represent space using a graph-based model. Graph-based models represent the turning points as nodes and the straight paths along which the user is expected to walk (*e.g.*, an aisle) as edges between the nodes [3, 9, 15, 41, 43, 58, 91, 95, 118, 167, 188, 200]. Graph-based model makes it possible to compute the shortest route between any pair of nodes on the graph by weighting each edge with the geographical distance of its two nodes. Different edge weights can model different interpretations of *best* route, for example, balancing distance with other factors such as potential hazards along the route.

An alternative space representation model is space discretization. Various solutions are adopted to achieve this including non-overlapping polygonal tiles of different size and shape [162], and grids in which all cells have the same size [2, 42, 127]. In particular, the grid-based representation is frequently used to model obstacles and the environment [2, 125], as well as the arrangement of RFID tags in the environment [42]. Using the grid-based approach, the route is represented as a sequence of cells that starts from the cell containing the user's position and ends at the cell containing the destination position. Since the grid can contain a large number of cells, the route computation can be computationally demanding. Therefore, an approach used in the literature is to create a topological graph from the grid to reduce the computational effort required to compute a route [127].

Discrete and graph-based models can be fused into a hybrid model. For instance, Kim et al. [111] propose a hybrid model composed of a graph-based model with nodes represented by beacons, which are used to compute the route, and a discrete model that divides the space into areas, which provide information about the environment. An alternative hybrid model proposed by Yoon et al. [206] employs waypoints to create the graph representation of the space, and an rectangular area (called Check-off zone) representing an approximation of a waypoint (see, for example Figure 2.1).

Next target computation

In turn-by-turn navigation, the user is directed along a straight line to reach a waypoint, and, once it is reached, the user is directed toward the following one. Previous solutions update the target when the user is closer than a given threshold to the turning point [3, 9, 41, 42, 79, 85, 91, 95, 118, 128, 162, 188, 200].

In the solution adopting graph-based models, the next target is selected as the next node on the route computed using a graph search algorithm like Dijkstra's. Furthermore, some papers propose ad hoc algorithm that considers one or more criteria for updating the weights of graph edges, such as the number of turn instructions required to reach the destination [56, 58, 119].

In discrete space models, the next target is computed using an ad hoc algorithm based on the space representation, or graph search algorithm by converting the discrete model into a graph. An instance of ad hoc algorithm is adopted using the tile representation [162]. In the tile representation, the

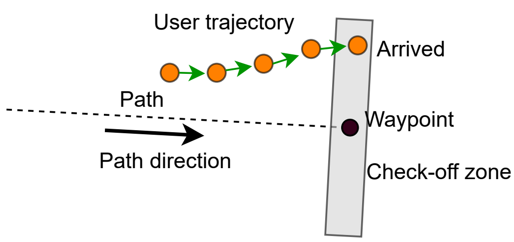


Figure 2.1: Example of a user trajectory entering a check-off zone. According to this model, the user is considered to have arrived to the waypoint even if there is a lateral deviation.

route is represented by a sequence of tiles. Since each tile has two reference points used for route computation, the target becomes the closest reference point in the next tile when the user is in proximity (less than a threshold distance) to that tile. There are two approaches to compute the next target using the grid representation. The first approach is to compute a route using the A^* algorithm. In this case, the route is a sequence of cells and the target is located at the center of the first cell in the sequence. The second approach is to convert the grid into a graph, simplifying the computation of routes using a graph search algorithm, as cells can be viewed as nodes in a graph. The route becomes a sequence of nodes instead of a sequence of cells and the target is the center of the cell represented by the first node in the sequence. Extensions of these approaches include ad hoc solutions that define the costs of cells in the grid or edges in the graph based on criteria such as the presence of obstacles, accessibility, and spatial awareness [2, 44, 125].

The target using Hybrid models is computed similarly to the graph-based model, as all waypoints are represented as nodes in a graph. However, the method for assigning the next target varies depending on the model. For instance, Yoon et al. [206] compute the route using the graph but determine when to assign the next target based on proximity to the discrete model (check-off zone) containing the target. In this model, the next target is the node in the route that corresponds to the center of the next check-off zone. Conversely, the hybrid model proposed by Kim et al. [111] assigns the next target using the graph model alone, without utilizing the discrete model, which is used for announcing points of interest.

Veering compensation

When there is veering, recalculating the route to correct the deviation can help identify a new target. Many papers address this unintended deviation by performing *rerouting*, which involves recomputing the route to update it [42, 9, 127, 125, 148, 162]. Other solutions do not involve rerouting; instead, an instruction is provided rather than reassigning the target [91, 95, 188, 200]. This approach models scenarios where the user should not or cannot deviate significantly from the route. One possible solution, briefly mentioned in the literature, is to provide an instruction when the user veers more than a threshold distance from the intended route, rather than updating the target [3].

Table 2.2: Related papers with a proposed solution for the *Routing* problem.

Main Routing approach	**Paper**	**Proposed approach**
Graph representation	[3]	Turn-by-turn navigation, veering correction, update the target at a given threshold to the waypoint

Main Routing approach	Paper	Proposed approach
	[9] [91] [95] [188] [200]	Turn-by-turn navigation, veering correction
	[41] [118]	Turn-by-turn navigation
	[43]	Turn-by-turn navigation, the connection between anchors represents a straight path, anchors as nodes
	[79]	Turn-by-turn navigation, veering correction, paths organized as a network of routes between the beacons
Grid	[2]	The grid cells indicate where it is possible to go, graph representation derived from the grid representation
	[42]	Graph representation derived from the grid representation, turn-by-turn navigation, veering correction
	[125]	The grid cells indicate where it is possible to go or unknown areas, recompute route based on obstacle on the trajectory
	[127]	The grid cells indicate where it is possible to go or unknown areas, graph representation derived from the grid representation, recompute route based on obstacle on the trajectory
	[148]	The target is the next gait point, recompute route based on obstacle on the trajectory, 3D grid
Multiple representation: Checkoff zones, Graph representation	[206]	Update the target when user enter in the check-off zone
Multiple representation: Hierarchy of areas, Graph representation	[111]	Update the target at a given threshold to the point of reference

Main Routing approach	Paper	Proposed approach
Tile representation	[162]	Update the target at a given threshold to the next tile, veering correction

2.4 The *Guidance* problem

Most existing solutions do not provide an explicit and complete procedure for generating announcements. Generally, an announcement can be either a single statement or a combination of statements. Common types of announcements proposed in research papers include: navigation instructions, Point of Interest (PoI) information, security announcements and updates about mobility.

The main contributions of the papers addressing the *Guidance* problem are summarized in Table 2.3.

Navigation instructions. These announcements relate to actions required during navigation. The most common instructions in assistive navigation typically involve directives to rotate or walk. These directives are given based on missing distance, veering, or an angular error in orientation relative to the required direction. Other navigation instructions are provided based on the purpose outlined in the paper, such as changing the floor or stopping before a pedestrian crossing. Some papers briefly mention the *Guidance* problem, for example, reporting that a rotation instruction is generated when there is a need to correct the user's orientation or when the angular error is larger than a threshold [3, 4, 5, 9, 12, 66, 84, 95, 125, 148, 188, 203, 204, 211]. However, most of these papers do not specify which instruction is given when a rotation is completed. Instead, other papers report the need to inform the user that the rotation was completed [3, 4, 5, 9, 10] and others focus on the reaction time to navigation instructions [12, 153]. Many papers adopting a turn-by-turn navigation paradigm for the *Routing* layer provide a navigation instruction when the user arrives at the target [9, 10, 31, 66, 148, 162, 204, 206, 209]. Furthermore, BLV people tend to veer off straight path. Many papers discuss strategies to compensate this unintended deviation by providing an instruction when the user veers beyond a threshold distance from the route [3, 9, 10, 62, 84, 91, 95, 188, 200, 204, 209, 211]. Another veering correction is proposed by Ahmetovic et al. [3] providing a side-step instruction to return inside the threshold distance from the route.

Point of Interest (PoI) information. These announcements convey details about environmental features, which may include information about the texture of the pavement, the significance of certain environmental elements, or connections between different parts of the environment. Solutions suggest to provide an announcement when the user is in the proximity of points of interest [9, 84, 167, 162].

Security announcements. During navigation, BLV people may encounter situations where instructions are provided for safety reasons. These announcements are made to inform about potential dangers. One example is the obstacle alert, which indicates the proximity of an obstacle. Common issues addressed include pedestrian crossings, stairs, and curbs. Indeed, obstacles may be present on the route [10, 84, 203], or there may be challenging points where it is difficult to understand what to do, such as encountering a moving obstacle while crossing the street at a crosswalk [95].

Updates about mobility. These announcements are related to the journey itself. This category includes anticipatory navigation instructions, such as information about upcoming actions and their distance, or variations in the route during navigation, such as route changes or interruptions. Additionally, some studies provide navigation instructions that should be followed when the user is before the target so that a BLV traveler knows beforehand which action to perform at the target [10, 31, 207, 209].

Table 2.3: Related papers with a proposed solution for the *Guidance* problem. PoI = Point of Interest information, Security = Security information

Paper	**Main Guidance approach**			
	Navigation Instructions	**PoI**	**Security**	**Updates about Mobility**
[3], [4] [5], [12] [125] [148], [204], [206], [211]	X			
[9], [162] , [167]	X	X		
[10], [31], [66], [188],[207]	X			X
[95], [203], [209]	X		X	X
[62]	X		X	
[84]	X	X	X	X

2.5 The *Interface* problem

In their research, scholars and developers addressed the three main levels of spatial knowledge [197], *i.e.*, knowledge about a point in space (*e.g.*, a landmark, an obstacle, a destination), knowledge about a sequence of points (a route to a destination, or "route knowledge"), and integrated knowledge about the environment (*i.e.*, cognitive-map like knowledge, or "survey knowledge").

To convey these types of knowledge to a BLV traveler, different Interface solutions exist. The proposed solutions adopt audio feedback, haptic feedback or graphical[9] feedback. These modalities can be used individually or combined

[9]Graphical feedback approaches are not covered in this thesis, as they do not align with the focus of this thesis.

to obtain multimodal feedback. Table 2.4 presents the contributions of each relevant paper in the literature addressing the *Interface* problem.

Table 2.4: Related papers with a proposed solution for the *Interface* problem.

Main Interface Approach	**Paper**	**Proposed approach**
Audio	[3]	App transforming environmental data into sounds to facilitate spatial understanding.
	[7]	Navigation system employing spatialized audio cues to guide users.
	[9]	Indoor navigation system using sound and spoken instructions for orientation.
	[29]	Device providing spoken information about the surrounding environment to facilitate mobility.
	[51]	App guiding users through predefined routes using voice commands.
	[67]	App offering real-time spoken information about traffic and public transport conditions.
	[69]	Smartphone app combining audio cues and spoken instructions for indoor navigation.
	[70]	Mobile application using audio feedback for indoor orientation.
	[106]	Device using 3D sounds to provide orientation cues.
	[110]	Interface delivering spoken instructions for orientation in complex environments.
	[128]	Portable device providing audio feedback for navigation.
	[129]	Personal navigation system employing spatialized audio to indicate direction.
	[136]	Vocal announcement system informing passengers in public transport stations.
	[167]	Smartphone app offering sonification and spoken instructions to guide users.
	[174]	System verbally communicating details about public transportation and stops.
	[187]	Vision-based localization system providing spoken instructions for navigation.
	[200]	Portable system converting navigation information into auditory signals.
Haptic	[19]	Smartphone vibration features used for conveying simple navigational instructions.
	[85]	Robot using haptic feedback to signal obstacles and directions.

Main Interface Approach	Paper	Proposed approach
	[101]	Vibrotactile bracelets that provide directional cues for navigation.
	[107]	Haptic strap used to enhance spatial awareness and provide directional guidance.
	[109]	Utilizes natural-sound-like vibrations to simulate environmental cues for navigation.
	[125]	Glass-mounted RGB-D camera and haptic feedback through vibration motors for obstacle detection and guidance.
	[135]	Head-mounted haptic sensors designed to aid navigation for visually impaired individuals.
	[148]	Electronic white cane using vibrations to signal obstacles and directions.
	[168]	Remote-controlled vibrotactile device guiding users through tactile feedback.
	[203]	Underfoot haptic cues in a virtual environment for intuitive movement guidance.
Multimodal	[95]	Wearable device providing tactile and auditory feedback for pedestrian guidance.
	[111]	System using audio and vibration feedback to assist with urban navigation.
	[162]	Integrated device combining sonification, spoken instructions, and haptic feedback for navigation.
	[211]	Bracelet offering vibrations and audio signals to assist runners with direction and pace.

2.5.1 Haptic Feedback

Haptic feedback is an interface modality that can be used to convey navigation instructions [19, 65, 101, 107, 109, 168] through vibration patters. Different types of hardware devices were experimented in the literature to convey the navigation instructions with tactile feedback, including vibrotactile bracelets[26, 101, 168], haptic strap [107], head-mounted haptic sensors [135, 125, 204], haptic cues underfoot [203] natural-sound-like vibrations [109], smartphone vibration [19], smart cane [24, 60, 74, 92, 112, 148, 194] and belts [65]. The combination of haptic feedback with audio feedback is also explored in research [127, 177, 211].

Since this thesis focuses on pre-existing generic hardware, such as smartphones, without the use of additional hardware, we do not employ haptic feedback.

2.5.2 Audio Feedback

Audio feedback-based interfaces are widely explored in research. In this domain, previous papers have adopted various solutions, including spoken instructions [2, 15, 18, 29, 67, 110, 136, 162, 174], sonification techniques [3, 4, 35, 155, 200], earcons [5, 7, 69, 106], and their combination [5, 9, 31, 43, 62, 95, 111, 167]. Many approaches, in particular, leverage spatialized audio to provide navigation assistance [7, 48, 106, 129, 178, 200].

Regarding the knowledge about a point in space, typical applications use sound to indicate where a PoI is localized. These include obstacle detection [38], identification or avoidance of relevant or potentially dangerous elements (*e.g.*, stairs or pedestrian crossings) [37, 139, 156, 173, 208]. The *Soundscape* project[10] uses spatialized audio and verbal cues to enhance ambient awareness. Through *Soundscape*, BLV people can place an audio beacon in the specific location of a point of interest. During usage, a spatialized 3D sound will be played in the direction of the audio beacon. Another mobile device application called "overTHERE"[11], developed from *Smith-Kettlewell's Virtual Talking Signs* project [48], can be used to determine heading direction of a point of interest. When the device is oriented towards a point of interest, "overTHERE" plays a verbal message to inform what is the point of interest and the distance to the point of interest. Instead, when the device is not oriented an intermittent impulse is spatialized indicating the direction of point of interest, mimicking the sensation of hearing sounds from the direction of the point of interest.

Regarding the knowledge about routes, several approaches have been proposed for the use of sound in wayfinding tasks. Klatzky et al. [113] compare the use of sound instructions to follow a route with those given by a spatial language (*e.g.* left, right). They show that the cognitive load is lower when instructions are given by sounds than by speech. Some approaches use simple panning of stereo sound [139], while others provide immersive 3D sound [178]. The Personal Guidance System [128], in particular, is a seminal work in this field. This system obtains information from a GNSS receiver and uses different types of auditory displays, spatial sound delivery methods, and tracker locations [129]. Usoh *et al.* [189] presented a comprehensive study of the application of non-speech sound for navigation and pathfinding. A more recent review on the use of sound for assisting the mobility of BLV people is provided by Spagnol *et al.* [178], albeit with a focus on spatial audio. The SWAN system employs non-speech auditory beacons that change in timbre and position with increasing levels of the subject's practice: the learning experience of the system improves both the navigation speed and accuracy [190]. The proximity distance to a target direction is the most common information conveyed using sonification. Fiannaca et al. [62] propose Headlock, which provides audio feedback (sonification and speech) to guide a user across an open space to a doorway. NavCog [9, 167] localizes the user through the sensing of the nearby

[10] `https://www.scottishtecharmy.org/soundscape`

[11] `https://www.ski.org/projects/overthere`

Bluetooth beacons and uses a combination of spoken instructions combined with earcons and sonifications to provide guidance to a destination. For a straight path, the user is periodically informed about the distance to the next target with synthesized verbal messages and the speed of an intermittent clicking sound increases as the user approaches the next target. For turns, once the user is aligned correctly to the new required direction (based on gyroscope data), a confirmation sound is played. Different types of sonification have been studied in the literature to better support the user during rotations (which are particularly critical) [4, 5] and have been adapted to guide users along straight paths [3].

Regarding the knowledge about the environment, various applications use sound to help a user form a cognitive map of an unknown environment, either *online* (*i.e.*, while exploring the environment) or *offline* (*i.e.*, through the learning of maps of the environment in advance). Guerreiro *et al.* developed a gesture-based virtual navigation system used to explore a turn-by-turn representation of the environment [83]. Lahav and coworkers [121] performed several studies on map exploration by blind subjects, which led to the development of the audio-haptic BlindAid system [122]. Katz and coworkers developed a system exploiting a 3D audio virtual environment to investigate the structural properties of spatial representations [105]. They used a "ears in hand" metaphor [131], based on an egocentric view of the virtual map in which the ears of the user are virtually placed at the position of the hand or the handheld device used to explore the map. The same metaphor was used by Geronazzo *et al.* [77] in developing a non-visual system for offline exploration of 2D maps. A different approach is used in a sonification-based obstacle detection and avoidance navigation system. The Watchout system proposed by Presti et al. [155] enables obstacle avoidance using sonification to indicate obstacle distance, width and height. Earcons and auditory icons are used to alert the BLV user to specific situations. Constantinescu et al. [45] used auditory icons to provide an awareness of the scene framed by a head-mounted camera. Participants in their study preferred an interface using these auditory icons over speech-based earcons (spearcons), which may be difficult to understand if delivered rapidly.

2.6 Services Provided by Navigation Assistance Systems

Navigation assistance services proposed in literature and commercial mobile apps offer support during mobility of BLV people. These include navigation systems for both indoor and outdoor environments, support for using public transport, environmental descriptions and descriptors, recording and repeating paths, setting and searching for points of interest, and obstacle detection and avoidance systems.

Some papers propose navigation systems that enable BLV people to navigate complex scenarios such as train stations and airports [84, 85, 111]. In addition, the independent mobility of BLV people is enhanced by systems for detecting pedestrian crossings [6, 171], solutions for assisted crossing in presence of traffic lights [39, 138] and traffic intersection [95, 145], and systems that record and repeat routes [43, 66, 188, 206]. Furthermore, obstacle detection and avoidance systems also contribute to BLV people mobility [31, 35, 38, 45, 57, 85, 107, 125, 127, 135, 154, 155, 156, 173, 208].

Many papers in the scientific literature have shown that indoor and outdoor positioning techniques enable a number of services that can effectively support the independent mobility of BLV people [56, 166, 182]. People involved in these studies mentioned that some applications are particularly useful, including: *Ariadne GPS*[12], *BlindSquare*[13], *Clew*[14], *Google Maps*[15], *iMove* [99], *Lazarillo app*[16], *Soundscape*[10], *NavCog* [9], *Nearby explorer*[17], *Seeing Eye GPS*[18], *Step-Hear*[19], *ViaOpta Nav*[20], and *WeWalk*[21].

Some papers have investigated in particular which applications are commonly used to support independent mobility with public transportation [21, 22, 100, 201, 212]. Popular applications cited by participants involved in these studies were *BlindSquare*, *Google Maps*, *Lazarillo app*, *Lyft*[22], *Apple Maps*[23],*Moovit*[24], *Siri*[25], *Step-Hear*, *Uber*[26], and *WeWalk*. Research has found that the ability to create a mental map of the route before moving thanks to software like *Google Maps* reassures BLV people [34, 100, 159, 212]. Many research projects have proposed solutions to make BLV people more autonomous in using public transport, such as bus [29, 67, 165, 174, 207, 210]. Some papers have studied accessible platforms to guide BLV people to choose the preferred public transport and to track the movement of the mentioned means of transport [27, 114, 136, 183, 207].

Many apps and research projects offer solutions to describe an environment. Recent commercial apps dedicated to describing the scene in front of a user are designed for BLV people, such as BeMyEyes [27], Envision [28], and SeeingAI [29]. Many research projects provide descriptions of framed images in the environment through camera sensors [7, 28, 85, 86]. Furthermore, Geographic Information Systems are used to provide more general information about the

[12]https://www.ariadnegps.eu/en/
[13]https://blindsquare.com
[14]https://www.perkins.org/resource/clew-navigation-app-review/
[15]https://www.google.com/maps/
[16]https://lazarillo.app
[17]https://tech.aph.org/ne_info.htm
[18]https://apps.apple.com/us/app/seeing-eye-gps
[19]https://www.step-hear.com
[20]https://www.itcares.it/portfolio/viaoptanav/
[21]https://wewalk.io/
[22]https://lyft.com
[23]https://www.apple.com/maps
[24]https://www.moovit.com/
[25]https://www.apple.com/siri
[26]https://uber.com
[27]https://www.bemyeyes.com/
[28]https://www.letsenvision.com
[29]https://www.seeingai.com/

environment to inform the user, both in research literature [11, 80, 102] and in commercial applications such as *Google Maps*, *Soundscape*, and *BlindSquare*.

A particularly useful feature for aiding mobility is the ability to record and subsequently replay a route in an environment [43, 66, 188, 206, 200]. This technology enables BLV people to navigate spaces more confidently by allowing them to retrace their steps along routes.

A common request among BLV individuals is to locate and set points of interest. This functionality is explored in research [9, 48, 167] and provided by common commercial apps such as *BlindSquare*, *Lazarillo app*, *Ariadne GPS*, *Soundscape*, *ViaOpta Nav*, and*Nearby explorer*.

Table 2.5 provides a review of research works that contribute to mobility services for BLV people.

Table 2.5: Works and applications in literature with a contribution on services during mobility.

Main Service	**Paper**	**Proposed Approach**
Point of interest setting and search	[9, 48, 167] BlindSquare [13] Lazarillo app [16] Soundscape [10] ViaOpta Nav [20] Nearby explorer[17]	Allow to search and set the location of a Point of interest with or without auditory feedback
Public Transportation Mobility Support	[21, 22, 100, 136, 201, 212]	Investigation into apps and technologies *Google Maps*[15] and *Moovit*[24] that help BLV individuals navigate public transportation effectively.
Bus Transport Assistance	[29, 67, 165, 174, 207, 210]	Research on solutions to enhance bus transport usability for BLV individuals, including navigation, bus tracking, and accessible information platforms.
Navigation in Complex Environments	[84, 85, 111]	Development of systems to assist BLV individuals in navigating complex environments like airports and train stations.
Pedestrian Crossing	[6, 171]	Technologies for detecting pedestrian crossings and aiding safe crossing activities in traffic environments.
Traffic intersections navigation	[95, 145]	Technologies for the safe navigation of traffic intersections.
Descriptor of Environmental features	[7, 28, 85, 86]	Research that provide real-time description of the surroundings to BLV users, enhancing situational awareness.

Main Service	Paper	Proposed Approach
	BeMyEyes[27], *Envision*[28], *SeeingAI*[29]	Commercial applications that provide real-time description of the surroundings to BLV users, enhancing situational awareness.
Geographic information system	[11, 80, 102]	Utilizes a knowledge-based database to offer context-aware support by providing detailed information about the environments being visited.
Record and repeat the route	[43, 66, 188, 206, 200]	Utilizes technology to record a route and enable users to retrace their steps to and from a starting point.
Obstacle detection and avoidance	[31, 35, 38, 45, 57, 85, 107, 125, 127, 135, 154, 155, 156, 173, 208]	Proposed works to detect an obstacle providing feedback to avoid it.

Chapter 3

Sonification of navigation instructions

Navigation assistance services for BLV people pose the challenge of providing accurate guidance through non-visual navigation instructions. This study proposes two techniques to guide the user during navigation assistance, and in particular during turns and straight paths. Both techniques adopt a combination of speech and non-speech audio with the aim of providing continuous, accurate, and unobtrusive guidance. The two solutions differ in the sonification technique used to generate the non-speech audio. The techniques were evaluated in a real-world environment with nine BLV participants. Experimental results show that one of the proposed techniques is significantly more effective than a baseline approach adopted in existing solutions, in terms of navigation accuracy during turns and straight paths. Thus, it is a practical and effective solution to the problem of non-visual navigation assistance for BLV people.

3.1 Background

Sonification is the use of non-speech audio to represent information [192], and it can be used as a means to accurately guide BLV people during navigation assistance. In particular, sonification of turn instructions is based on the angular error relative to the correct direction to be followed. *NavCog* [9] uses a confirmation sound to indicate when the user is correctly aligned with the target direction while turning, remaining silent when the user is not aligned. However, this can lead a BLV person to turn without realizing they have over-rotated, resulting in a misleading direction after reaching the target orientation [12].

The use of sonification techniques with a continuous feedback during turns can convey continuously the amount of the angular error, effectively communicating how close the user is to the target orientation. In two previous contributions, we have studied how sonification feedback support BLV people during turn instructions.[4, 5]. The turn instructions are provided through simple verbal messages (*e.g.*, "turn right") followed by real-time sonification feedback that conveys the quantity of rotation needed to complete the turn. While the user rotates, the auditory feedback changes to represent the distance to the target angle. In particular, we proposed three different sonification techniques[1]:

IS (Intermittent Sound) – This sonification triggers impulsive "ticking" sounds at a variable rate, which is inversely proportional to the angular distance from the target direction.

AM (Amplitude Modulation) – This approach employs a sinusoidal sound, modulated in amplitude by a low-frequency (sub-audio) sinusoidal signal. The frequency of the modulating signal is inversely proportional to the angular distance, producing a slowly pulsing sound at large angular distances, which becomes stationary when the target is reached.

MS (Musical Scale) – The initial distance from the target angle is subdivided into eight circular sectors, corresponding to as many grades of an ascending major musical scale. While approaching the target angle, the user enters new circular sectors, thus triggering new notes; when the target has been passed, a descending note is produced.

The above sonifications were compared with a baseline approach (**Ping - P**), used in existing solutions [9], which consists of a single impulsive sound triggered when the turning angle is correct. Through empirical experiments in a laboratory, we measured the rotation accuracy, rotation time, and sonification preferences of BLV participants during turns. The studies demonstrated that a baseline solution for indicating direction is not sufficiently accurate, potentially leading a BLV person in the wrong direction. Moreover, the MS sonification demonstrated higher accuracy compared to the baseline. Coupling the proposed sonification approaches with the ping sound when the correct turning angle is achieved further improves the accuracy of all three approaches, without impacting the rotation time. In particular, **IS+P** (obtained by coupling **IS** and **P**) and **MS+P** (obtained by coupling **MS** and **P**) sonifications achieved the highest accuracy. To support the finish of a turn, deep learning approaches can be combined with the turn instructions [12] reducing the angular error during rotation.

3.2 Contribution

Our prior work [5] shows that the proposed sonification approaches can be useful to support BLV people during navigation assistance. However, the investigation considered only a portion of the turn-by-turn navigation assistance process, that is the turn instructions, while travelers also need to be guided during straight path segments. Furthermore, the tests were conducted in a laboratory with seated participants, therefore without considering the effect of users' own movement and external environment influences on the guidance

[1] Examples of the sonifications are available online `https://earcons.netlify.com/`

accuracy. To address these issues, we extends our previous work with the following contributions:

- We generalize the **IS+P** and **MS+P** sonifications, which obtained the best scores in the previous study, to support the users also during straight path segments in turn-by-turn navigation assistance.
- We design and implement a prototype mobile application that provides turn-by-turn navigation assistance and supports the proposed sonification approaches. The prototype relies on accurate visual-inertial localization, based on visual markers positioned in the environment. This solution has been implemented using native mobile Augmented Reality (AR) libraries (see Section 3.4.1).
- We evaluate the accuracy, the required time, and the appreciation of **IS+P** and **MS+P** sonifications, compared to the baseline approach, during navigation assistance tasks in a real-world setting.

Our results confirm that, also in a real-world scenario, the proposed sonification techniques result in significantly smaller rotation errors and rotation times compared to the baseline condition. Instead, on straight paths, **IS+P** achieves lower distance errors than both **MS+P** and the baseline, without impacting the overall walking time. Participants' comments highlight that **IS+P** was also perceived to be more precise and that it would be preferable on unfamiliar routes. Instead, on familiar routes, **P** was deemed more appropriate since it is less invasive and distracting with respect to the background sounds. Indeed, this sonification was also perceived to be less annoying, in particular with respect to **MS+P**. The proposed sonifications could be easily integrated into state-of-the-art navigation assistance solutions for BLV people [167], replacing the baseline approach which is currently used, and resulting in more reliable, accurate, and safe guidance.

3.3 Sonification Design

In the turn-by-turn navigation assistance paradigm, the navigation of a route from a start position to a destination is conceptualized as a sequence of two alternating activities: i) traversal of straight paths connecting a waypoint to another, and ii) turns occurring at each waypoint to head towards the next one. The former activity mainly depends on the linear distance between the traveler and the destination waypoint, the latter on the angular distance between the direction of the traveler and the direction of the next waypoint. Prior solutions for BLV people convey such distances through verbal instructions at the beginning of each activity (*e.g.*, "proceed 10 meters" [167] or "turn to 11 o'clock" [32]), and confirm the completion of the activity with a single impulse **Ping** (**P**) sound [157]. However, the inherent imprecision in following such instructions can lead to navigation errors and potential hazards [13]

With the aim of providing a more accurate navigation assistance for BLV people, we propose the use of sonification to convey the remaining distance (angular or linear) in real-time while the user is turning or traversing a straight path. The core idea of these techniques is to compute in real time a sound that represents the distance to the correct angle (during rotations) or to the next turning point (during straight paths). Specifically, in our prior work [5], for the task of conveying turn instructions, we have investigated three sonification techniques (briefly described in Section 3.1): **Intermittent Sound** (**IS**), **Amplitude Modulation** (**AM**), and **Musical Scale** (**MS**). These approaches, which we refer to as *base sonifications*, were evaluated in comparison with the baseline solution (*i.e.*, **P**) used in existing navigation assistance systems [167]. Moreover, we have also proposed and evaluated three variants in which **P** is added to each of the above sonifications when on target, thus obtaining **IS+P**, **AM+P**, and **MS+P** (*compound sonifications*). This form of reinforcement introduced further improvements in user performance.

In this work [3], we generalize two of the proposed sonification approaches for both navigation activities (turns and straight movements), namely for the representation of both angular and linear distances. The two sonifications are **IS+P** and **MS+P**, which yielded the best accuracy and appreciation results in the previous investigation.

3.3.1 Design Criteria

If data sonification can be thought of as a musical instrument played by the data, sonification assisting users in reaching a goal can be thought of as an instrument that encourages the user to reach those goals by exploiting musical expectations. Different musical backgrounds should also be considered in the design, but this is true for any sonification, which must be tailored to the sonic culture of the target audience [130].

With specific reference to the proposed design, **IS** is composed of short broadband pulses. Taken individually, such bursts mask a large part of the spectrum (maximizing the probability of being audible) for a very short time (minimizing the probability of compromising other sound events). Taken together, as the target approaches, the pulses result in a sort of accelerando that creates a pleasing tension and invites the listener to complete the musical gesture.

This latter aspect is at the core of the **MS** sonification as well. For this sonification, the choice of using an ascending major scale is driven by the fact that it is rooted in Western music theory and culture [23]. Also, the generalizability of this choice is justified by the ubiquity of the major scale in popular music [97] and a musical predisposition in infancy reported in scientific literature [186].

Finally, **P** is basically an auditory icon [75] resembling a small bell, such as those found at the hotel desk or in restaurant kitchens. This may carry a cultural and conditioning component, eliciting meanings such as “I’ve arrived” or “Ready!”. Even if these symbolic aspects are outside of the cultural space of the listener, it still remains a distinctive sound. Moreover, when used in conjunction with other sonifications, it occurs at the end of a tensional gesture (accelerating pulses for **IS** or an ascending scale for **MS**), thus providing an additional sense of resolution.

3.3.2 IS+P Sonification

In detail, this sonification triggers broadband pulses ("ticking" sounds) at a variable rate. The rate is inversely proportional to the distance, resulting in faster ticking at closer distances. Thus, the sonification vaguely resembles to a Geiger-Müller counter [76]. Our implementation employs a noisy unpitched pulse, which is repeated at rates from 1 to 15 Hz. In addition, as a reinforcement, a final impulsive sound ("Ping") is played when the target is reached.

The mapping between distance (linear or angular) and the sonification parameters (*e.g.*, the sound rate in this case) can adopt a linear or exponential scale. After a preliminary evaluation phase, we resolved to implement an exponential mapping for **IS+P**. As a result, the ticking frequency increases slowly while the user is far from the target, while it rapidly increases in the proximity of the target. This way, the user is prompted to heighten the awareness and slow the movement when approaching the target.

The mapping function $freq_{IS+P}$ of the **IS+P** sonification is defined in Equation 3.1 and represents the pulse frequency in Hz. D represents the remaining distance to reach the target (while walking) or the angular error (while rotating), and T is the total initial distance to the target (while walking) or the target angle (while rotating) at the start of sonification. When D is within a distance threshold $D_{threshold}$ a final impulsive sound ("Ping") is played.

$$freq_{IS+P} = \begin{cases} 15, & D \leq D_{threshold} \\ 1 + 14\left(1 - \left|\frac{D}{T}\right|\right)^4 & D < T \\ \left|\frac{T}{D}\right| & D \geq T \end{cases} \tag{3.1}$$

3.3.3 MS+P Sonification

In this sonification, the initial distance from the target is subdivided into eight parts, corresponding to as many grades of an ascending major scale. While approaching the target, the user enters new parts of the distance, thus triggering new notes; when the target has been passed, a descending note is produced. Our implementation corresponds to a C major scale starting from 261.63 Hz (C_4 in scientific pitch notation) and notes are performed using a piano timbre. Additionally, as in the previous sonification, a final "Ping" sound when the target is reached is used as a reinforcement.

Differently from **IS+P**, the exponential mapping was considered to be misleading for **MS+P**, as even small movements would trigger many notes when in the proximity of the target. A linear response to the distance from the target is expected to be easier to predict in this case, as the eight distance portions would be equal, and moving at a constant speed would trigger notes at a constant pace. Thus, for **MS+P** a linear mapping was chosen. Employing different mapping strategies is coherent with the intrinsic characteristics of the sonification techniques: **MS** is inherently discrete, triggering notes at fixed distance values, whereas **IS** is continuous.

The mapping function $tone_{MS+P}$ of **MS+P** sonification is defined in Equation 3.2 and represents the tone played. D represents the remaining distance to reach the target (while walking) or the angular error (while rotating), and T is the total initial distance to the target (while walking) or the target angle (while rotating) at the start of sonification. When D is included in one of the eight portions, the respective note of the portion is played. If D is greater than T, no tone is played. When D is within a distance threshold $D_{threshold}$ a final impulsive sound ("Ping") is played.

$$tone_{MS+P} = \begin{cases} C4 & \frac{6T}{7} \leq D < T \\ D4 & \frac{5T}{7} \leq D < \frac{6T}{7} \\ E4 & \frac{4T}{7} \leq D < \frac{5T}{7} \\ F4 & \frac{3T}{7} \leq D < \frac{4T}{7} \\ G4 & \frac{2T}{7} \leq D < \frac{3T}{7} \\ A5 & \frac{T}{7} \leq D < \frac{2T}{7} \\ B5 & D_{threshold} \leq D < \frac{T}{7} \\ C5 & D < D_{threshold} \end{cases} \tag{3.2}$$

3.4 Prototype Navigation Assistance Application

In order to experimentally evaluate the proposed sonification techniques in a real-world environment, we developed a prototype mobile navigation assistance application that implements them. The application localizes the user in the environment (see Section 3.4.1), generates the route towards a target destination (see Section 3.4.2), and conveys the navigation instructions to the user through an auditory interface (see Section 3.4.3). User pose and navigation instructions provided by the app are continuously logged. These data are used for the evaluation of the proposed sonification techniques in Section 3.5.

3.4.1 User Localization

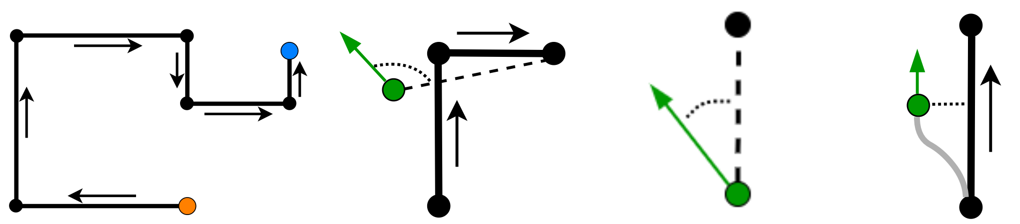
To localize the user, the application employs computer vision techniques available in the native AR library for iOS devices. The advantage of using existing AR libraries for mobile devices is that they are publicly available and well-engineered. Specifically, the app computes and tracks the user's pose in terms of a 2D position and a 1D orientation (*i.e.*, direction).

To compute the user's pose, the app employs visual markers detection [132]. We use generic images as markers that can be detected using the iOS AR library. During the setup phase, each marker's image, its pose in real-world coordinates, and its size are entered into the system. When the application frames a marker, it can compute the relative pose of the camera with respect to the marker image. Since the marker's pose in the environment is also known, the app can compute the device's pose in real-world coordinates as well. Specifically, the computation requires dealing with two reference systems: the *Global Reference System* (GRS) represents the 2D position in the WGS84 format and the orientation angle with respect to the north, while a *Local Reference System* (LRS) is instead generated by the AR library each time it starts an AR session. The mapping between LRS and GRS origins is initially unknown. When the user frames the first marker, the AR library provides that marker's pose in terms of LRS. However, since the marker pose is also known in terms of GRS, the mapping between LRS and GRS can be computed, thus enabling the conversion between poses in the two reference systems.

When a marker is not framed, the device pose is tracked using visual-inertial odometry [93] and Simultaneous Localization and Mapping (SLAM) [55]. These functionalities are also available as part of the native iOS AR library. The AR library keeps track of the device position in terms of LRS and this is converted into the pose in terms of GRS. Clearly, these techniques are subject to a drifting error, causing the mapping between LRS and GRS to degrade. Hence, it is necessary to periodically frame a marker to compute a position fix (*i.e.*, re-compute the mapping between LRS and GRS).

3.4.2 Navigation Strategy

During the setup phase, a set of pre-defined routes is stored in the application. Each route connects a *starting position* to a *destination position* through a set of alternating turning points and straight paths. Figure 3.1a shows an example of a route: the orange and blue dots are the starting and destination positions, respectively, while the black line segments and points are the straight paths and the turning points, respectively. Although the route is pre-defined, the actual navigation instructions are computed considering the user's pose, which is estimated in real-time. As an example, in Figure 3.1b the green dot represents the position of a user who approaches a turning point, and the green arrow represents the user's orientation. The target rotation angle is computed considering the user's position and orientation, with respect to the direction of the next turning point (dashed line), and is depicted as the dotted arc in the figure.



(a) Route (b) Turn (c) Orientation Offset (d) Lateral Offset

Figure 3.1: *Navigation Strategies*

When the marker at the starting position is framed, the user's initial pose is computed. Then the application provides rotation instructions to orient the user toward the next turning point. Upon reaching the correct orientation, the user is instructed to walk straight toward the next turning point. When the next turning point is approached, the navigation procedure iterates. Finally, when the user reaches the destination position, the navigation task is completed.

While this implementation of the marker-based localization strategy was conceived solely as part of the experimental apparatus for the evaluation of the proposed sonification techniques and not intended as a complete localization solution, there are contexts in which a similar approach could indeed be practical. One example is the localization inside art galleries and museums, in which paintings can be used as implicit visual landmarks [7]. Another potential application is outdoor visual localization. For this use case, a solution is already implemented in AR libraries [2].

One important aspect to take into account when designing the navigation strategy is that the user's actual pose may deviate from the one planned in the route. The main reason for this is that, although our aim is to instruct the users precisely, a certain level of imprecision in the user movements is unavoidable. There are three main sources of imprecision that we observed during the experiments: over-rotation errors (during a rotation, the participant rotates more than expected), over-walking errors (when approaching a turning point the participant does not stop immediately when instructed), and veering, which is a known problem in the navigation of BLV people [88] and that happens when the person, instead of walking along a straight path, deviates towards the left or the right.

We designed two compensation strategies to remedy these errors. First, the application keeps track of the *orientation offset*, which is the difference between the user's current orientation and the direction of the user toward the next turning point. Figure 3.1c shows an example of orientation offset: the user's position and orientation are represented as a green point and a green arrow respectively, the next turning point is the black dot, the black dashed line segment connects the user's position to the turning point, and the dotted arc represents the *orientation offset*. When the *orientation offset* is larger than a given threshold, the user is instructed to rotate. This instruction is identical to the rotation instructions given at turning points. Such compensation can remedy all three sources of error mentioned above. However, in the case of veering, despite a limited *orientation offset*, the user can still move far away from the expected route before the compensation is triggered, which can be potentially dangerous. This is exemplified in Figure 3.1d: the user's movement (green path) can gradually diverge from the planned route, resulting in a *lateral offset* (the dotted line segment) between the user's current position and the planned route. Thus, as a second compensation strategy, if the *lateral offset* is larger than a given threshold, the user is instructed to move sideways towards

[2] See `https://developer.apple.com/documentation/arkit/argeotrackingconfiguration`

the planned route through a verbal message (*e.g.*, "move right"), similarly to prior work [167].

3.4.3 Auditory Interaction Design

The auditory interface consists of a combination of verbal instructions, generated using the mobile device speech synthesizer, and sonifications, which convey quantities associated with these instructions. The sonification technique can be selected as an application setting and, during a single navigation task, a single sonification approach is used. The application provides the following messages:

- "Turn" message[3]: the instruction is "turn left" or "turn right". After the message, the quantity of rotation is conveyed through sonification. This instruction is provided when the *orientation offset* is larger than 30°, which can happen in two scenarios: i) when the user arrives in the proximity (*i.e.*, 0.5 m) of a turning point, and ii) when the user unintentionally rotates while traversing a straight path;
- "Walk straight" message: after a rotation, when the user reaches the right angle, the application reads the message "Proceed straight for x meters". This instruction is followed by the sonification, which quantifies the remaining distance while the user is walking;
- "Move sideways" messages: when the lateral offset is larger than 1 m, the application reads the message "move left" or "move right". The instruction is not followed by a sonification, as typically one step is sufficient for the user to get back on the expected path;
- End message: when the participant reaches the end of the last path segment, the application reads the message "target reached".

3.5 User Study

To understand how the proposed sonification techniques affect the navigation process, we conducted a user study with nine BLV participants. The experimental evaluation included three navigation tasks (one for each sonification technique), and a final questionnaire that collected the participants' subjective feedback.

3.5.1 Experimental Design

The goal of the experimental evaluation was to compare the two proposed sonifications (*i.e.*, **IS+P** and **MS+P**) with one another and with the baseline (*i.e.*, **P**), to evaluate how they impact the navigation precision, speed, and

[3]Henceforth we report the message translation from Italian

user appreciation. To assess this, we computed objective measurements and participants' subjective feedback.

The following objective measurements were collected:

- **Supervisors' interventions**: the number of times the supervisors intervened during the test, either because the participant requested support or to prevent potential hazards to the participant;
- **Over-rotation error**: the angular error between the target angle and the participant's direction at the end of a turn;
- **Over-rotation time**: the time between the moment participants are informed to stop rotating and the moment they end the turn;
- **Rotation velocity**: the average angular speed achieved during the turn;
- **Over-walking error**: on straight paths, the distance between the point where a participant receives the instruction to stop walking and the point where the participant actually stops;
- **Over-walking time**: the time between the instant when a participant is informed to stop walking and when the participant actually stops;
- **Walking speed**; the average walking speed achieved while traversing a straight path.

Except for the **Supervisors' interventions**, which were manually annotated, the measurements were computed from the data logged by the app.

Participants' subjective feedback was collected through five Likert-like scale questions. Possible answers ranged from 1 (disagree) to 7 (agree) and were repeated for each sonification technique. The questions were the following [4] (for each question, we indicate the name we use in the study to refer to that question):

Pleasantness: "The sound[5] was pleasant";

Annoyance: "The sound was annoying";

Precision: "The sound guided me precisely";

Quickness: "The sound guided me quickly";

Appreciation: "I appreciated the sound";

For each question, the participants were also invited to provide comments to motivate their scores. Furthermore, additional opinions, comments, and suggestions on the entire navigation procedure, the sonifications, and the navigation instructions were collected.

[4] The experiments were conducted in Italian, henceforth we report the English translation.

[5] In the questions we avoided using the technical term "sonification" and instead we used the term "sound".

Table 3.1: Participants' demographic information. For impairment: LB=legally blind, TB=totally blind. For route frequency: Fam.=familiar, Unfam.=unfamiliar, D=daily, W=weekly, MO=monthly, R=rarely.

			Impairment		**Expertise**		**Screen reader**		**Route**	
PID	**Age**	**Gender**	**Level**	**Onset**	**Music**	**Mobile**	**Speed**	**Voice**	**Fam.**	**Unfam.**
P1	33	F	LB	age 27	4	5	65-70	F	D	W
P2	22	F	LB	birth	3	5	55	M	D	R
P3	25	F	LB	age 7	4	4	65-75	F	D	R
P4	22	M	TB	birth	3	4	90	M	W	R
P5	50	F	LB	age 25	2	5	75	M	D	MO
P6	33	M	LB	age 12	3	3	70	F	D	MO
P7	21	M	TB	age 6	3	4	60	F	D	MO
P8	59	M	TB	birth	4	4	60	F	D	W
P9	63	M	TB	age 5	2	3	55	F	D	MO

The impact of different sonifications on the participants' navigation performance during turns and straight paths was analyzed for statistical significance using unpaired tests, since each navigation task could have had a different number of actual turns and straight paths (see Section 3.4.2). Specifically, the Mann-Whitney-U test [134] was used when two data series were analyzed, and for more data series we used the Kruskal-Wallis test [116], with the Dunn test for post-hoc comparisons [54]. The target significance levels in multiple comparisons were adjusted using the Benjamini-Hochberg False Discovery Rate method [25]. Instead, subjective Likert-like responses were paired across different sonifications, and therefore Friedman's test [68] was used. The Nemenyi test, which already accounts for familywise error, was used for post-hoc comparisons [150].

3.5.2 Participants

The study involved 9 participants, whose demographic data are reported in Table 3.1. The inclusion criteria for this study required the participants to be BLV people, unable to orient themselves with residual sight (namely people with severe visual impairment), and with no other disabilities. Four participants self-identified as female, the others as male. The age of participants was between 21 and 63 (36.44 ± 15.66)[6]. Four participants were completely blind, two participants were legally blind with light perception (**P1**, **P5**), three participants were legally blind with light and shape perception (**P2**, **P3**, and **P6**).

In addition to the demographic information, we also collected participants' self-reported expertise in music as a Likert-like scale item ranging from 1 (no expertise) to 5 (high expertise). Musical expertise is related to this study because it is known to modulate lower-level auditory function [180] and thus the ability to comprehend sonification techniques. We also collected the partici-

[6]Henceforth we use the notation $a \pm b$ to report mean a and standard deviation b.

pant's self-reported expertise with the use of mobile devices (again, Likert-like scale items ranging from 1 to 5) and details about their usual screen reader settings (see the details in Table 3.1).

Another relevant aspect that we investigated is related to how participants listen to the screen reader in mobility. Two participants reported that they do not listen to the screen reader while walking (P4 and P7). In particular, P7 clarified that he stops when he needs to interact with the screen reader. The other seven participants, instead, use the screen reader while walking: three use the device speaker only (P1, P3, P9), one uses a single earphone only (P5), and two use either the device speaker or a single earphone, depending on the situation (P6, P8). Only one participant (P2) reported using stereo headphones, specifically "pass-through" headphones[7].

Finally, we asked the participants how often they travel independently along familiar and unfamiliar routes. All participants reported traveling known routes frequently (all on a daily basis, except for P4 who travels familiar routes independently on a weekly basis). Instead, participants report traveling unfamiliar routes less frequently: three participants reported traveling unfamiliar routes rarely (*i.e.*, less than once per month), four on a monthly basis, and two on a weekly basis.

3.5.3 Experimental Setting and Apparatus

The study was conducted on the campus of the University of Milan. We defined three routes with the same number of planned turns (*i.e.*, six turns each) and approximately the same length (84m, 96m, and 90m, respectively). For each route, the planned turns were equally divided into left and right turns. Note that, as explained in Section 3.4.2, the number of turns actually performed by the participants can be greater than the number of planned turns for a given route, because there may be additional correction turns to compensate for imprecision. The routes were designed to avoid hazards and objects that the participant can use to support orientation and navigation (like walls) because we wanted the participants to rely on the provided sonification only, without being guided by other objects in the environment; for example, we wanted to avoid that participants could walk along a straight path by following a wall (BLV people are usually trained to walk along walls using the white cane or echolocation). Visual markers (A2 paper size) were displayed along the routes, on walls or stands, in the proximity of each turning point.

The study was conducted using an iPhone 13 Pro, running the application described in Section 3.4. To ensure that the device camera could correctly frame the visual markers, maintain a clear view of the surrounding environment, and have a consistent orientation in the direction of the participant, the device was positioned on the participant's chest, placed on a chest strap mount. The auditory information was provided through the device speaker, without using headphones, in order not to mask the surrounding soundscape, which is important for BLV people to avoid hazards.

[7]Pass-through headphones allow the user to hear the environment soundscape as well.

3.5.4 Experimental Protocol

The study lasted about one hour and was organized into four main phases: a preliminary questionnaire, a study introduction, the navigation phase, and a final questionnaire. The preliminary questionnaire collected data related to participants' demographic, expertise, and habits. The main results of the questionnaire have been reported in Section 3.5.2. Afterward, we briefly introduced our study, provided an overview of the navigation tasks, and explained to the participant how to follow the instructions provided by the application.

During the navigation phase, each participant was asked to complete three tasks, each consisting of an initial tutorial and a navigation trial. Each task was associated with a different sonification. During the tutorial, the supervisor explained how to use the sonification and the participant was asked to try it on a short test route consisting of two short straight paths and one rotation. Then, the navigation trial consisted in walking along one of the three predefined routes. The tasks were counterbalanced to offset possible learning effects. Specifically, the order of routes and tasks, and the association between routes and tasks were defined following a Latin square design. During the navigation, two supervisors followed the participant, without interfering with the navigation but being ready to take action in case of potential hazards.

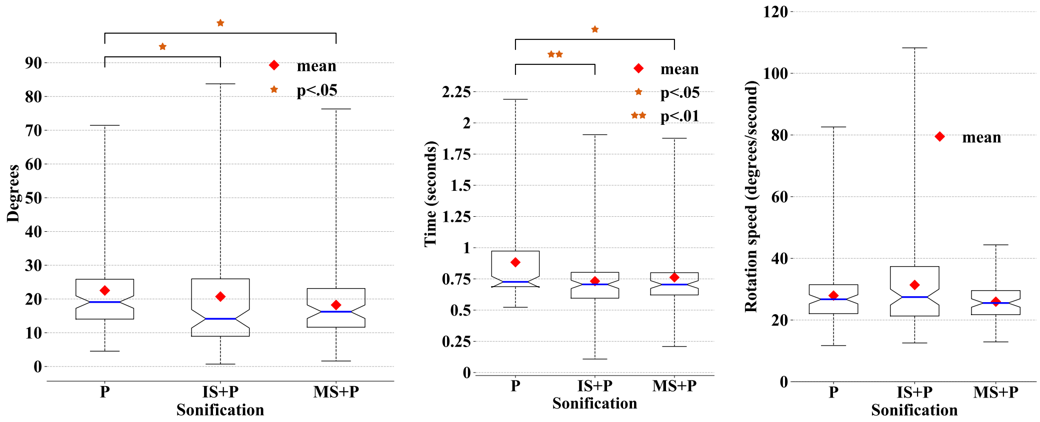
The fourth and last phase of the study was the final questionnaire that collected subjective feedback from the participants regarding the sonifications and the navigation tasks. Specific questions are reported in Section 3.5.1.

3.6 Experimental Results

In total, we collected data from 313 pairs of turns and straight paths (34.78 ± 6.73 for each participant). Note that the minimum number of turns was 18 for each participant (six turns in each route) but additional turns were required to correct the participants' orientation. No significant difference was found in the number of rotations required with different sonifications (12.67 ± 4.50, 12.44 ± 2.75, and 9.67 ± 3.20 for **P**, **IS+P**, and **MS+P**, respectively).

3.6.1 Supervisors' Interventions

Overall, most of the experiment was conducted with the participants navigating independently in the unknown environment, with only sporadic supervisors' intervention (*i.e.*, 7 interventions in 313 pairs of turns and straight paths), to support the participant or to prevent possible hazards. In three cases, despite the training task, the participant did not understand how to interpret the sonification instructions and therefore an additional explanation was required during the navigation task. In one case, support was needed due to a positioning error (a visual marker was not framed) that resulted in the participant being guided out of the planned route. In one case, the participant touched an obstacle on the side of the route with the white cane and stopped to ask the supervisors what to do. In another case, the participant could not hear the



(a) Over-rotation error (b) Over-rotation time (c) Rotation speed

Figure 3.2: Rotation Metrics

vocal instruction and asked the supervisor to repeat it. Finally, while using **IS+P** a participant made an over-rotation but the orientation offset was below the threshold of 30°, so the application did not instruct to re-rotate. However, in the following straight path, the participant got close to a bush and thus was stopped by a supervisor.

3.6.2 Sonification Impact on Turns

In the analysis of over-rotation errors, over-rotation times, and rotation velocities, we excluded 37 turns due to anomalous situations (*e.g.*, the participants interrupted the rotation movement to share a comment with the supervisor), so we take into account 276 turns. Full results are reported in Table 3.2.

Sonification	**Over-Rotation**		**Rotation**
	Error	Time	Velocity
P	$22.52 \pm 13.12°$	$0.88 \pm 0.33\,s$	$27.93 \pm 9.54°/s$
IS+P	$20.71 \pm 17.73°$	$0.73 \pm 0.30\,s$	$31.38 \pm 14.61°/s$
MS+P	$18.22 \pm 11.27°$	$0.76 \pm 0.27\,s$	$25.95 \pm 6.80°/s$

Table 3.2: Impact of sonifications on turns

Considering the over-rotation errors (see Figure 3.2a), the differences across the sonifications were found to be significant ($H_{(2)} = 8.53$, $p < .05$). Pairwise comparisons show that rotation errors for both **MS+P** and **IS+P** were significantly ($p < .05$) lower than for **P**. The sonifications also significantly ($H_{(2)} = 11.62$, $p < .01$) impacted the over-rotation times (see Figure 3.2b). Also in this case the rotation times for **MS+P** and **IS+P** were lower than for **P**. No significant differences were instead found between the rotation speeds achieved with the proposed sonifications and with the baseline (see Figure 3.2c).

One finding reported in prior literature highlights that BLV people tend to incur larger rotations errors when the requested rotation angle is small [13].

Our results confirm this finding (see Figure 3.3). Indeed, when the target rotation angle is smaller than 60°, participants incur significantly ($p < .01$) larger rotation errors for all sonifications (see Table 3.3).

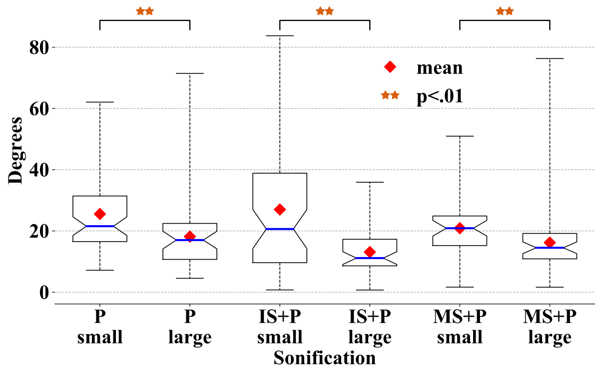


Figure 3.3: Comparing over-rotation error for *small* (*i.e.*, less than 60°) and *large* (*i.e.*, more than 60°) turns.

Sonification	**Over-Rotation**	
	Small turns	Large turns
P	$25.55 \pm 13.53°$	$18.16 \pm 11.15°$
IS+P	$27.01 \pm 21.02°$	$13.10 \pm 7.35°$
MS+P	$20.90 \pm 10.79°$	$16.19 \pm 11.19°$

Table 3.3: Impact of sonifications on over-rotations for small and large turns.

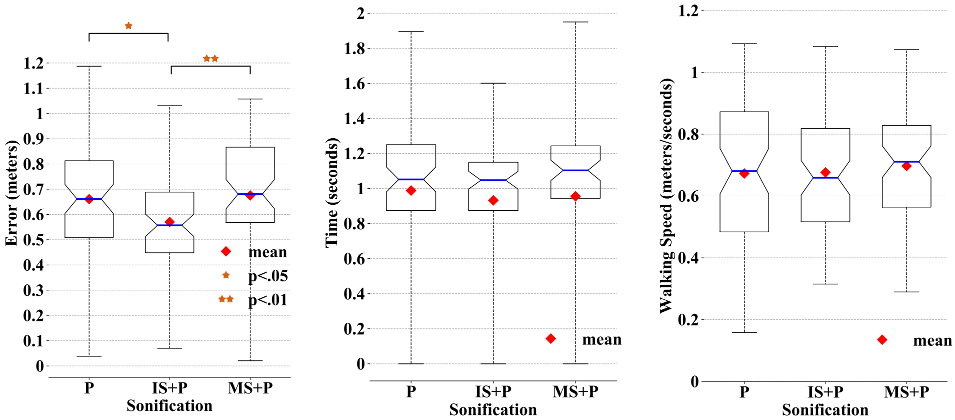
3.6.3 Sonification Impact on Straight Paths

In the analysis of over-walking errors, over-walking times, and walking speeds, we excluded straight paths that were shorter than one meter. These paths were often generated due to compensation instructions and were too short to properly convey the sonifications. Thus, we consider the 208 remaining straight paths. The full results of the analysis are reported in Table 3.4.

Sonification	**Over-Walking**		**Walking**
	Error	Time	Speed
P	$0.66 \pm 0.24\,m$	$0.99 \pm 0.46\,s$	$0.67 \pm 0.21\,m/s$
IS+P	$0.57 \pm 0.18\,m$	$0.93 \pm 0.46\,s$	$0.68 \pm 0.19\,m/s$
MS+P	$0.68 \pm 0.23\,m$	$0.96 \pm 0.47\,s$	$0.70 \pm 0.19\,m/s$

Table 3.4: Impact of sonifications on straight paths

For what concerns over-walking errors (see Figure 3.4a), significant differences ($H_{(2)} = 12.98$, $p < .01$) were found among the sonifications. In this case **IS+P** produced significantly lower over-walking errors with respect to both **P**



(a) Over-walking error (b) Over-walking time (c) Walking speed

Figure 3.4: Metrics on straight paths.

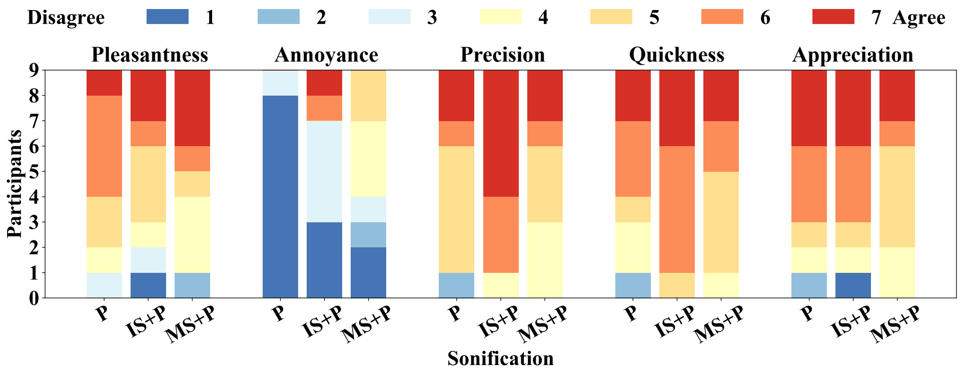


Figure 3.5: Answers to the closed questions

and **MS+P**. Instead, no significant differences emerge for over-walking times and average walking speeds considering different sonifications (see Figure 3.4b and Figure 3.4c).

3.6.4 Participants' Subjective Feedback

Figure 3.5 shows the distribution of the answers to the closed questions presented in Section 3.5.1. Mean values and standard deviations[8] are reported in Table 3.5.

Sonification	Pleasant	Annoying	Precise	Quick	Appreciated
P	5.33 ± 1.15	1.22 ± 0.63	5.22 ± 1.40	5.55 ± 0.96	5.56 ± 1.57
IS+P	4.77 ± 1.81	3.11 ± 2.02	6.33 ± 0.94	6.22 ± 0.63	5.44 ± 1.83
MS+P	5.11 ± 1.66	3.22 ± 1.47	5.22 ± 1.13	5.55 ± 0.96	5.33 ± 1.05

Table 3.5: Subjective feedback about sonifications

[8] We report means for multipoint items as they were found to better indicate central tendency than medians [126]

Considering the **Pleasantness** metric, no statistically significant differences emerged among the sonifications. However, **P7** commented:

> **P7**:*I am ok with both [P] and [IS+P]. Perhaps I would use the first one for familiar routes since I already know the distances, while on a new route I would use the second one to have more information on the distances.*

Also **P5** and **P8** mentioned that **MS+P** was mentally demanding (*e.g.*, **P5**: "I need to concentrate"). **P5** appreciated **P** the most because this sonification does not mask the environmental sound, but also recognized that **IS+P** clearly conveys the distance to the target:

> **P5**:*[With **P**] I could hear the background environment ... [**IS+P**] gave me a good understanding of how close the target was.*

Similarly to **P5**, **P6** also observed that **P** does not mask the background noise, but still gave a lower score to this sonification because it does not provide precise navigation instructions:

> **P6**: *With ping [i.e., **P**] I felt unsure and it gave me fewer references.*

For **Annoyance**, significant differences were found (($H_{(2)} = 8.59$, $p < .05$) and **P** emerged to be less annoying than **MS+P**. **P1** commented that the sound produced by **IS+P** made her anxious, while **P4** noted that he prefers sounds unrelated to music and hence he reported **MS+P** to be annoying. **P3** noted that the continuous and rapid ticking of **IS+P** is annoying, while **P2**, **P7** and **P8** observed that the sound of **IS+P** can mask the surroundings noise. Interestingly, although **MS+P** produces a less continuous sound, which does not mask the surrounding noise as much as **IS+P**, it was even less appreciated, on average, than **IS+P**. On this, **P3** provided an interesting insight:

> **P3**:*Despite the pauses [between notes in **MS+P**] I had to focus on the sound rather than the surrounding.*

Considering **Precision** and **Quickness**, although no statistically significant difference emerged, **IS+P** obtained a higher average score and was viewed positively by almost all participants: **P1**, **P2**, **P6** and **P9** commented that **IS+P** provides a more precise perception of the distance to the target (either rotating or moving on a straight path), **P5** and **P7** observed that this sonification is simpler and safer, and **P8** perceived that **IS+P** uses fewer instructions. The only exception is **P3**, who experienced problems while using **IS+P** and, in contrast with the other participants, commented that the sonification is not very precise.

Considering the **Appreciation**, five participants indicated **IS+P** as the most appreciated sonification (**P2**, **P4**, **P5**, **P6**, **P7**)[9]. In contrast with them,

[9]**P4** and **P5** gave the maximum score to **P** and **IS+P** but commented that they would prefer using **IS+P**.

P1 gave the lowest score to **IS+P**, perceiving it to be annoying, while she gave the maximum score to **MS+P** motivating that the sound is pleasant, precise, and fast.

One main remark emerged from the semi-structured interviews: while many participants recognized that the instruction to move sideways was indeed useful (**P3**, **P4**, **P5**, **P8**, **P9**), four participants (**P2**, **P3**, **P4**, **P7**) suggested to add a sonification for this instruction as well, in order to inform how much the user should move sideways. There are two additional insights that we believe to be particularly interesting: **P7** suggests using various sonifications for different actions while **P9** and **P5** question the idea of turn-by-turn navigation:

> **P9**: *I am not used to walking then stopping and rotating, I usually do it all continuously.*

3.7 Discussion

3.7.1 Study results

Our previous work [5] shows that, for rotations, **IS+P** and **MS+P** both yield significantly better performance than **P**, with no significant differences between **IS+P** and **MS+P**. The experimental evaluation presented in this study confirms our previous results regarding rotations, also in a realistic experimental scenario. Indeed **IS+P** and **MS+P** yielded significantly better results than **P** considering over-rotation error and over-rotation time. Instead, considering straight paths, **IS+P** achieves a lower over-walking error than both **P** and **MS+P** and no significant differences emerge in over-walking time. One possible reason that motivates the better performance by **IS+P** on straight paths is that it provides more accurate feedback in the proximity of the target, due to the exponential mapping of the distance to sound frequency. The sound is also continuous, providing distance information even when the user slows down or stops, an approach that some participants reported being more informative. Instead, due to the linear distance mapping, **MS+P** plays notes at regular distances only, which on longer straight paths may be quite far between. Thus, the user is less induced to slow down when approaching the target, and therefore the technique results are less accurate.

A second interesting result concerns the perceived annoyance caused by the different sonification techniques. In our previous study, conducted in a laboratory setting, no significant differences emerged. We believe that, since the users did not have to move and orient themselves in an unknown environment, they did not perceive the proposed sonifications as annoying. Conversely, here we show that in a real-world environment the participants did perceive **MS+P** as being more annoying than **P**. Furthermore, the participants commented that both proposed sonifications could be annoying in some situations. Specifically, **IS+P** was reported to be distracting from the surrounding soundscape. We suspect that this could be related to the repetition frequency of the sound (with higher frequencies being able to guide the user more accurately but at the same time being more annoying), or to the sound itself. As a future work, we will explore alternative sound designs, evaluating their impact on the perceived annoyance. Another interesting result is that **MS+P** was reported to be cognitively demanding, as the participants needed to pay attention while waiting for the notes to be played. This was not expected, as we believed that a more discrete sound would have been less invasive for the participants.

A third interesting result regards the design of the navigation assistance system, which was based on the idea that using one sonification for both turns and straight paths would result in navigation assistance instructions being easier to learn and follow. The above results and participants' comments suggest that the opposite approach, in which different instructions are sonified in a different way, might be preferable. **MS+P** could be used for rotations, while **IS+P** could be used for straight paths. Another possible approach could be to let the user select the preferred sonification, a choice that can be influenced by a number of factors, including the extent to which the user knows the route. Indeed, as observed by some of the participants, **IS+P** is more informative when the route is unknown but could be too distracting along a (partially) known route.

3.7.2 Experimental Environment

Differently from our prior investigation, in this study we evaluated the impact of the proposed sonifications in a real-world scenario. Specifically, the experimental evaluation was conducted on a university campus for three main reasons: to limit the hazards for the participants, to avoid unpredictable obstacles along the route (*e.g.*, a standing person, a parked bicycle) that could impact the study results, and to test the system in an environment where participants would not be guided by environmental features (*e.g.*, a corridor) that can be used as references, haptically (*e.g.*, by following a corridor wall with the white cane) or acoustically (thanks to echolocation, BLV people can, for example, feel the presence of an open door in a corridor).

These characteristics, however, might not be present in other scenarios, and hence could influence the navigation process and its outcome. For example, in our study, the participants knew that, in case of a hazard during the experiment, the supervisor would intervene. This might have impacted the way the participants navigated or their walking speed. Similarly, in an environment with obstacles, the participants would have most likely had to avoid them, which would have influenced their walking routes. Also, while there was some noise in the area due to the presence of numerous students and a few passing cars, the overall background noise was different than in a typical outdoor environment with many passing vehicles. The difference in the background noise could have potentially influenced the ability of the users to perceive the instructions and the sonifications.

There are also two factors that made the experimental setting harder to navigate than routes usually traversed by the participants. First, the participants had no previous knowledge of the experimental routes. This is particularly challenging because BLV people most commonly travel on familiar routes or study the route beforehand, as shown in Section 3.5.2. Second, the participants had no environment reference points, and in particular no haptic references. One common practice among BLV people that use the white cane (as all our participants did) is to follow haptic references, like walls or sidewalk edges. In the experimental setting, we intentionally decided to avoid these references, so that they would not affect the results, but clearly, this makes the walking task much more challenging.

3.7.3 Experimental Apparatus

Study results may have also been influenced by the experimental apparatus used. In particular, the effect of the navigation assistance system prototype, which supported the experimental evaluation of the sonification techniques, was not explored. In our experiment, we did not evaluate the accuracy of the navigation assistance system itself since this was out of the scope of the investigation. However, since the over-rotation and over-walking errors were computed using the navigation assistance system localization, we acknowledge that the accuracy of these measurements depends on the accuracy of the navigation assistance system. To maximize the localization accuracy, and therefore the reliability of the over-rotation and over-walking measurements, the localization markers were positioned in the proximity of the turning points, where these measurements were taken. Considering the navigation strategy and the guidance instructions used by the navigation assistance system, our design adopted a turn-by-turn navigation schema. While the proposed design is not the only possible one for a navigation assistance system, we highlight that our design choices are consistent with prior solutions in the literature [167].

Another peculiarity of the study is the use of a chest mount. This study design choice helped to simplify the experimental protocol because, if the participants held the device with their hands, it would have been harder to distinguish body rotations from hand rotations. Participants would have also needed to concentrate on pointing the device at the right angle (*e.g.*, not towards the ground), which is known to be hard for BLV people [138]. However, a chest mount may not be a practical solution for a number of reasons, including that its acceptance by the users needs to be evaluated and, as observed by one participant (P2), the smartphone can be easily stolen. Alternative solutions could be wearable devices, like smart glasses.

3.7.4 Participants

In our previous investigation, the experimental setting was minimal and easily replicable, which made it possible to recruit many participants, in multiple places, thus minimizing geographical bias and allowing us to consider also the impact of cultural factors on the results of the rotation instructions. In this study, since the experimental setting was constructed in a specific real-world environment, we could only recruit local participants that were available to come to our university campus. Furthermore, in order to limit confounding factors that could impact the results, we only involved participants who are blind. As a result, the study involved a small number of participants, a limitation that is common in the field of research on assistive technologies [172].

Additionally, unlike the participants from the previous study, most of the participants in this study had similar characteristics. In particular, most had similar self-reported expertise with music and smartphones and similar screen reader settings. Therefore, differently from our prior work, we were not able to analyze these characteristics in our study. Similarly, we could not assess for possible cultural biases in our population sample (for example, using a major scale in **MS** may be less effective for non-Western musical cultures).

3.7.5 Interaction Design Guidelines

Based on the results of the user study and on the experience gained during the definition of our interaction approach, we provide preliminary guidelines for the design of navigation assistance verbal and sonification instructions.

- The combination of verbal instructions and sonification, where an action is requested to the user verbally (*e.g.*, "rotate right"), and its quantity is conveyed through sonification, represents a practical solution for navigation assistance, as it is easy to learn and provides accurate guidance.
- The navigation interaction paradigm should account for distractions and possible memory overload of the user. Thus, the system should provide the functionality to repeat the most recent navigation instruction upon user request.
- On straight paths it is more effective to have an exponential distance mapping that provides more accurate information when the user is approaching the target and less distraction when the user is far.
- There is a trade-off in selecting the frequency at which the continuous navigation sound is repeated. A frequent repetition is generally more informative and can provide more accurate guidance, but it can also negatively impact the perception of the environmental sound.
- Personalization of the guidance instructions and sonification parameters is needed to account for the users' subjective preferences and different usage contexts (*e.g.*, known vs. unknown routes).
- Some users feel reassured when a sound is played periodically, also when the user is not moving or moving slowly as a reminder that the system is still providing guidance. This is particularly relevant on long straight paths.

Chapter 4

Modular navigation system architecture

Prior research on navigation assistance for BLV people mainly focuses on two aspects: the technology used to localize the user, and the interaction approach for guiding the user during navigation. Other important aspects that bridge these two parts, such as the routing algorithm, the data model of the environment, and the navigation logic, are seldom addressed. In this chapter, we propose a modular architecture design for navigation assistants, and we propose a detailed design of two key modules, implementing aspects related to routing and guidance, which are rarely addressed in prior literature.

4.1 Contribution

This chapter proposes three key contributions:

- First, a decomposition of the navigation assistance problem is conceptualized, reflecting a design approach that focuses on the separation of concerns of each part (see Section 4.2). The proposed approach breaks down the navigation task into four distinct layers: Localization, Routing, Guidance, and Interface. This conceptual framework allows for modularity in design and implementation.

- As a second contribution, our work focuses on the design of the Routing and Guidance layers, which have been overlooked in the literature but are critical to generating clear instructions for BLV users (see Section 4.3). A solution implementing a state of the art approach for the Localization and Interaction layers, and the proposed design for *Routing* and *Guidance* layers, is also proposed.

- Finally, as the last key contribution, a user study is conducted with eight BLV participants (see Section 4.4). Results show that *NavGraph*, compared to a state-of-the-art solution, provides clearer instructions and safer navigation.

4.2 Problem decomposition

As previously mentioned in Section 1.3, we propose to structure the problem of BLV navigation assistance into four separate layers (see Figure 4.1):

- *Localization*, that is the problem of, computing the user's pose, usually in terms of a 2D position and 1D orientation. Different technologies can be adopted to address this problem, either outdoor or indoor, including GNSS [106], Bluetooth beacons [9], or visual-inertial odometry [93].

- *Routing*, that is the problem of deciding the *next target* toward which the user should be directed. Solving this problem generally involves computing a multi-step route from the current position to the destination and then selecting, given the context in the user's proximity, which is the next target.

- *Guidance*, that is the problem of selecting the instructions to guide the user toward the next target. The problem is particularly relevant for the navigation of BLV people because there is a need to provide salient instructions without overwhelming users with too much information (*e.g.*, audio feedback) [10, 144].

- *Interface*, that is the problem of conveying the guidance instructions to the users, using a combination of audio, haptic, and visual information. In particular, for audio information, it is possible to use combinations of synthesized speech, sonification [90], and earcons [30].

This structure reflects a *separation of concerns* design approach, aimed at making it simpler to address the overall problem of navigating a BLV person. Specifically, this division allows each layer to be managed independently, ensuring that changes in one layer do not impact the others. In the following we design an overall architecture of a navigation assistance system. This framework, called *NavGraph*, details the functioning of each identified layer, and proposes a novel design for the Guidance and the Navigation layers, which are not sufficiently discussed in prior works.

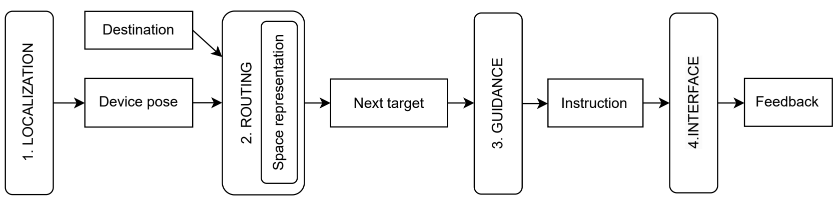


Figure 4.1: Decomposition of the navigation problem.

4.3 *NavGraph* design

The previously introduced problem decomposition highlights a division among the layers of a navigation system. This division allows each layer to be managed independently, ensuring that changes in one layer do not impact the others. In particular, we present a design for *Routing* and *Guidance* layers, which have been overlooked in the literature. Indeed, many papers propose solutions oriented towards improving *Localization* and *Interface* without analyzing the influence of changes in the *Routing* and *Guidance* layers. Section 4.3.1 describe the system architecture, while Section 4.3.2 provides formal definitions that are used to model the problem and that are used to compute the navigation (see Section 4.3.3). Finally, Section 4.3.4 describe the proposed solution for selecting the instructions to convey addressing the **Guidance** problem.

4.3.1 Solution architecture

We design a system architecture (see Fig. 4.1) based on four modules, each implementing a solution to the problems identified in Section 4.1.

The *Localization* module computes the device's pose (2D position and orientation). We assume that this module runs periodically, providing location updates to the following modules.

The *Routing* module computes the route from the current device position to the destination. In this thesis, we do not address the problem of how the user selects the destination, because it depends on the specific domain. For example, in a road-navigation app, the user might type the address, while a museum application can present a set of artworks as possible destinations. To compute the route, the *Routing* module uses a set of data structures defining the relevant characteristics of the environment that we call *space representation* and that are formalized in Section 4.3.2. The route computed by the *Routing* module is represented as a list of intermediate targets that the user should reach. What is particularly relevant is the *next target* that is the first element of this list and that represents the point where the user should be directed.

The *Guidance* module select the messages to guide the user toward the next target. This module is not aware of the actual device's position in the entire environment but only takes in input the local information that includes the *current edge* along which the user is moving and the relative position of the user with respect to the next target. The output of the *Guidance* module is an instruction, composed of an *instruction type*, which is a label from a pre-defined set (e.g., "rotate") possibly associated with a quantity (e.g., "30 degrees").

Finally, the *Interface* module conveys the instructions to the user. This includes defining how and when to convey the instructions. Indeed, while the *Guidance* can potentially generate instructions at a high frequency, it is the *Interface* module that defines when to convey them. For example, if messages are conveyed in the form of verbal instructions, the *Interface* module can decide to avoid repetition of the same instruction. Alternatively, if the *Interface* module outputs the instructions on a user interface (e.g., an arrow indicating the direction), the interface can be updated every time a new instruction is received from the *Guidance* module. In the following, we do not focus on the *Interface* module, as existing solutions can be adopted [3, 5, 9, 95].

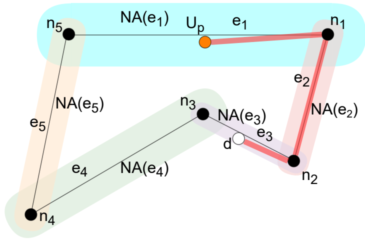


Figure 4.2: Example of a *Navigation Graph* and navigation area.

4.3.2 Space representation

In this thesis, we model the user position over a two-dimensional space S. Even though the user's and device's pose do not necessarily match (in particular for the orientation) in the following, for the sake of simplicity, we refer to the *user*'s position $u_p \in S$ and orientation $u_o \in [0, 2\pi)$.

We define a *Navigation Graph* where edges and nodes represent straight paths and waypoints, respectively. The *Navigation Graph* is used to compute the routes and hence a weighting function among edges is required. Different weighting functions can be defined, for example taking into account the expected effort to traverse an edge. Formally, the *Navigation Graph* $\mathcal{G}$ is an undirected and weighted graph $\langle N, E, W\rangle$ where N is a set of nodes, $E \subseteq N^2$ a set of edges, and W a weighting function $W : E \rightarrow \mathbb{R}^+$. We denote the position of each node n in S as $n.p$. An example of a weighting function that only takes into account the length of an edge in S can be defined as follows: for each $e = \langle n_1, n_2\rangle \in E$, $w(e) = dist(n_1.p, n_2.p)$, where $dist$ is the euclidean distance.

Each edge intuitively defines a straight path that the user can follow. In some cases, the users are constrained to walk along an edge (*e.g.*, in a corridor), while in open spaces users can walk freely, possibly moving away from the edge they should be following. To represent this, we introduce the concept of *Navigation Area* of an edge, which is the region of space around that edge. The union of all Navigation Areas for all edges in the graph is the *Graph Navigation Area*. Formally, for each edge e, we defined a linear distance δ_e and we define the navigation area of e as the set of points closer than δ_e to e:

$$NA(e) = \{p \in S | dist(p, e) < \delta_e\}$$

where, given $e = \langle n_1, n_2\rangle$, $dist(p, e)$ is the euclidean distance between p and the line segment $\overline{n_1 n_2}$. Fig. 4.2 shows an example of a *Navigation Graph* with three edges, each with its navigation area.

We model the destination d as a point in S. We assume that the user has reached the destination if the user is in the *destination area*: a circle centered in the destination with radius δ_d. Since the Graph Navigation Area defines where the user is expected to walk, we assume that the system should prevent the user from moving outside it. For this reason, we assume that the destination area is contained in the Graph Navigation Area.

4.3.3 The *Routing* module

The objective of the *Routing* module is to process the space representation, taking into account the user's position and the destination to compute the local information that is then provided to the *Guidance* module. There are the following cases.

1. The user has arrived at the destination *i.e.*, the user position is in the destination area.

2. The user is outside the graph navigation area and should be guided to re-enter in the navigation area.

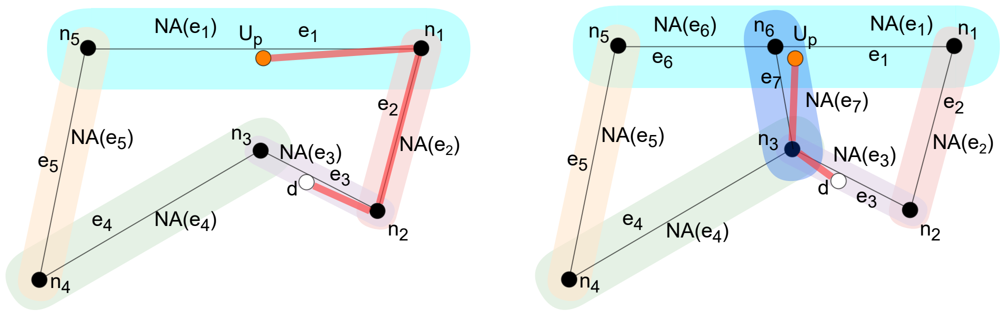
3. The user and the destination are in the navigation area of the same edge ($i.e.$, $\exists e \in E | u_p \in NA(e) \wedge d \in NA(e)$). In this case the user can be directly guided toward the destination.

4. In all other cases, the *Routing* module computes the route from the user's position to the destination along the graph navigation area. The remaining of this section describes this procedure by first providing an intuitive idea of how the user can navigate to the destination and then by showing how to compute it.

Intuition. Let's first consider the case in which the user's position and the destination are each contained in the navigation area of a single edge. An example is shown in Fig. 4.3a. In this case, the user is in the navigation area of edge e_1 and can be guided toward n_1 or n_5 along the navigation area. Then, from either n_1 or n_5 the user can follow the *Navigation Graph* toward n_2 or n_3 that are the vertexes of e_3. From there, the user can be guided toward the destination along the navigation area of e_3. In a more general case, the user position and/or the destination can be in the navigation area of more than one edge. The example in Fig. 4.3b shows the case in which the user's position is contained in the navigation areas of two edges (i.e., e_1 and e_6) and hence the user can be guided toward any node among n_1, n_5, n_6 and n_3.

Computation Algorithm 1 shows the procedure to compute the route. It takes in input the user's position, the destination and the *Navigation Graph*. The procedure first creates two new nodes, n_u and n_d, representing the user's position and the destination, respectively (Lines 1 and 2). The procedure then computes the set NA_u of edges whose navigation area contains the user position and, similarly, the set NA_d of edges whose navigation area contains the destination (Lines 3 and 4). In Line 5 the procedure computes the set of nodes N_u that are the vertexes of edges in E_u. Similarly (Line 6) it computes the set E_d for the destination. Next, the procedure creates the set of edges E_u that contains, for each node n in N_u one edge between n_u and n. Similarly for the destination (Lines 7 and 8). Then, the procedure extends $\mathcal{G}$ by adding n_u and n_d to the set of nodes and E_u and E_d to the set or edges. Finally, the procedure returns the shortest route on $\mathcal{G}'$ between n_u and n_d.

Note that the routing procedure seamlessly solves the **Target update** problem (see Section 2.3) because it automatically computes a new route when



(a) User's position in the navigation area of a single edge

(b) User's position in the navigation area of two edges

Figure 4.3: Example of Routing from the user's position (orange dot) to the destination position (white dot).

Algorithm 1 Routing procedure

***Routing* procedure**:
Input: u_p, d, $\mathcal{G} = \langle N, E, W \rangle$
Output: shortest route from u_p to d along the navigation area of $\mathcal{G}$.
n_u is a node representing the user's position
n_d is a node representing the destination position
$NA_u = \{e \in E | u_p \in NA(e)\}$
$NA_d = \{e \in E | d \in NA(e)\}$
$N_u = \bigcup_{e=\langle n_1,n_2 \rangle \in NA_u} \{n_1, n_2\}$
$N_d = \bigcup_{e=\langle n_1,n_2 \rangle \in NA_d} \{n_1, n_2\}$
$E_u = \bigcup_{n \in N_u} \langle n_u, n \rangle$
$E_d = \bigcup_{n \in N_d} \langle n_d, n \rangle$
$\mathcal{G}' = \langle N \cup n_u \cup n_d, E \cup E_u, \cup E_d, W \rangle$
return the shortest route on $\mathcal{G}'$ between n_u and n_d

the user enters in the navigation area of a new edge. Consider for example the situation in Fig. 4.3a: the user is initially guided towards n_1 (Section 4.3.4 describes how this is achieved). Once the user enters into the navigation area of e_2, this edge will be part of NA_u and thus n_2 will be part of N_u (see Algorithm 1). Consequently the shortest route algorithm will compute a path that starts from the user's position, passes through n_2 and then arrives at the destination hence updating the next target, that becomes n_2.

4.3.4 *Guidance* module

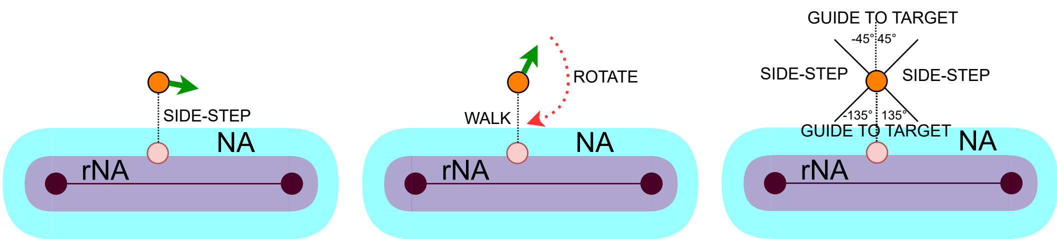
The module is designed according to two design principles: **safety**, achieved by minimizing the time spent outside the navigation area, which can be potentially hazardous for the user, and **parsimonious correction** to reduce the number of instructions to correct the user's pose by avoiding minor corrections.

In line with the previous literature [3, 167] the *Guidance* module generates four types of instructions: **arrived**, **walk**, **rotate**, and **side-step**. Considering the four cases described in Section 4.3.3, the *Guidance* module returns **arrived**, if the user arrived at destination (case 1). Instead, if the user is outside the graph navigation area (case 2), the aim is to guide them to return to the navigation area with **walk**, **rotate**, and **side-step**, as discussed in Section 4.3.4. In cases 3 and 4, the *Guidance* module guides the user towards a target (the destination or the first node on the route, respectively) by using **walk** and **rotate** messages. This is discussed in Section 4.3.4.

Guidance to the navigation area

A common situation, during navigation along a straight path, is that the user veers thus walking outside the navigation area. When this happens, it is common that the user is directed approximately in the correct direction (*i.e.*, parallel to the navigation area). Fig. 4.4a shows an example. In this case, it is possible to guide the user to re-enter into the navigation area with a **side-step** instruction (e.g., a side-step toward the right, as in the example of Fig. 4.4a). The *Guidance* module applies the *parsimonious correction* principle: instead of guiding the user to just re-enter into the navigation area, it guides the user deeper into the navigation area. The reason is that, if the user is guided to just re-enter the navigation area, there is the risk that the user could exit again from it after a short while. This problem is mitigated by guiding the user closer to the actual edge. Technically, we define a *restricted Navigation Area* (*rNA*), which has the same definition as the navigation area, but with a smaller δ_e (in our experiments we used a value for δ_e of the rNA that is half that of the navigation area). Once the user re-enters the *rNA*, guidance resumes as described in Section 4.3.4.

While the situation depicted in Fig. 4.4a is common, there are situations in which the *side-step* instruction would not guide the user back to the navigation area. This happens when the user orientation is almost perpendicular to the navigation area, as exemplified by Fig. 4.4b. In this case, there is the need to guide the user to **rotate** and then to **walk** to re-enter into the *rNA*. In this



(a) Side-step to return inside the *rNA*

(b) Turn and walk toward the *rNA*

(c) Possible instructions

Figure 4.4: Instructions provided when the user is outside the navigation area. The orange circle is the user's position; the green arrow is the user's orientation; the side-step or walk direction is the dotted line from the orange circle to the rNA.

case, the *rNA* is defined as a target area and the user is guided towards it as described in Section 4.3.4.

In order to decide whether to use a **side-step** instruction or define a next target, we define an angular reference system oriented perpendicularly to the navigation area as depicted in Fig. 4.4c. If the user orientation is between 45° and 135°, the user is instructed to do a **side-step** to the right, otherwise, if the user orientation is between −135° and −45°, the user is instructed to do a **side-step** to the left. If, at any point while re-entering into the navigation area, the user orientation is not in one of these two ranges, the *rNA* is defined as a target area and the user is guided with a sequence or **rotate** and **walk** messages as described in Section 4.3.4.

Guidance to target

In the following we describe how the *Guidance* module creates instructions to guide a user towards a target area. There are three situations in which the user can be guided towards a target area:

- When the user is re-entering into the *rNA*. In this case the target area is the *rNA* of the current edge as explained in Section 4.3.4.
- When the user and the destination are in the navigation area of the same edge (case 3 in Section 4.3.3). In this case, the target area is the destination area.
- When the *Routing* module computes a route towards the destination (case 4 in Section 4.3.3). In this case, the target area is the intersection between the navigation area of the current edge and the navigation area of the next edge. We explain the intuition with the example of Fig. 4.5: the user is navigating along the current edge e_1 toward the intersection of the navigation areas of e_1 and e_2 (the area with purple outline). As soon as the user reaches this area, since the user is now in the navigation area of e_2, the routing procedure (Algorithm 1) computes a new shortest path having n_2 as the first node and hence the user will be guided towards the following target area.

Guidance towards a target area works as follows: if the user is heading towards the target area, then a **walk** instruction is returned. Otherwise, the

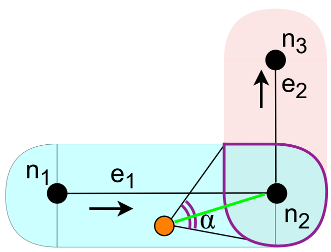


Figure 4.5: Example of target area computed as the intersection of two navigation areas.

user is instructed to rotate to correct the orientation. Also in this case, the *Guidance* module adopts a parsimonious correction approach: instead of correcting the orientation to the barely minimum to re-align with the target area, the module requires the user completely re-align to the center of the target area.

The intuitive procedure described above is formalized as follows: consider the angle α centered on the user's position (see Fig. 4.5) subtending the target area. As long as the user orientation is in α, then **walk** instruction is returned. Otherwise, the *Guidance* module starts correcting the user's orientation with a **rotate** instruction until the user is aligned with the bisector of α. Then, a **walk** instruction is returned again.

Note that the above approach is subject to a problem: when the user is far away from the target (e.g., on a very long edge), it could happen that α is a small angle, making it hard for the user to keep the orientation within this angle, hence resulting in frequent orientation corrections. For this reason, the size of α is computed as the maximum between the angle subtending the target area and a fixed angular threshold (30° in our experiments).

4.4 User study

To understand how the proposed architecture and specifically the design of the *Routing* and *Guidance* modules affect the navigation experience of people who are BLV we conducted a user study with eight blind participants with no residual vision (see Section 4.4.6).

4.4.1 Experimental design

The experimental evaluation followed a within-group design in which the independent variable (condition) was the version of the navigation app used: one representing the state-of-the-art solution (**SOTA**), and the other our proposed solution (**NEW**). The experimental apparatus is described in detail in Section 4.4.3. For a given condition, each participant did two navigation tasks (in total four different paths) with the condition order alternated and counterbalanced across participants to avoid effects of order (see Section 4.4.5 for the detailed experimental procedure).

Various metrics were analyzed as dependent variables (see Section 4.4.2). Since we adopted a within-group design, differences between conditions for a given metric were tested for statistical significance using paired tests. Specifically, when data series were normally distributed we used Paired T-test [181] and Wilcoxon signed-rank test [199] otherwise. Normality of the distributions was checked using Shapiro-Wilk's test [161]. In some cases, an additional between-group factor was considered. In such cases, we used a Two-Way Anova [179] for normally distributed data, and Scheirer–Ray–Hare test [169], otherwise Post hoc analyses were conducted using the Wilcoxon signed-rank test and Paired T-test, respectively. When multiple comparisons were con-

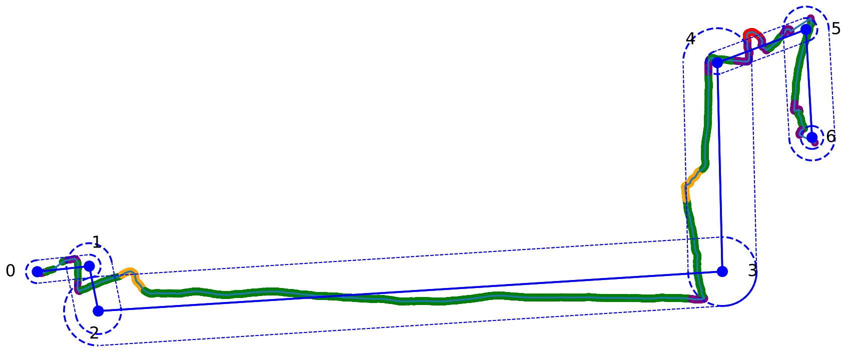


Figure 4.6: Example of a user's trajectory. The figure shows a navigation graph with 7 nodes and 6 edges together with its navigation areas (blue). The trajectory of the user is shown as a trace with segments of various colors corresponding to segment in which the user walks into the *NA* (green), rotates into the *NA* (purple), side-steps to return into the *rNA* (orange), rotates to return into the *rNA* (red).

ducted, we used the Kruskal-Wallis test [116], adjusted using the Benjamini-Hochberg False Discovery Rate method [25].

4.4.2 Metrics

We collected both objective and subjective data. Most of the objective data are computed from the system logs, which include the user's trajectory and other variables computed by the navigation application. For example, Fig. 4.6 was computed from the collected data and shows the trajectory of **P2** on path 2. The blue lines represent the space representation, in terms of the *Navigation Graph* (solid line) and the navigation area (dashed lines). The user trajectory is represented in various colors, depending on the status of the system: green and purple represent the situation in which the user is within the navigation area and is instructed to walk or rotate, respectively. Orange and red are used, instead, to represent the situation in which the user is outside the navigation area and is instructed to do a side-step and to rotate, respectively.

From the system logs we computed the following data, which we then analyze in Section 4.5:

- **Number of instructions**: the number of turn, walk, and side-step instructions and their total;
- **Lateral distance from the current edge**: the average lateral distance from the current edge;
- **Time duration**: the time duration of the navigation task;
- **Time outside the navigation area**: the total amount of time spent outside the navigation area;

- **Instructions cluster**. Given a temporal parameter t, an instruction cluster is a set of consecutive instructions such that the temporal distance between each of them and the following is less then t.

In addition to the system logs, we also collected data manually annotated by the supervisor, including data about the supervisors' interventions, *i.e.*, the situations in which the supervisors intervened during the test, either because the participant requested support or to prevent potential hazards to the participant.

Participants' subjective feedback was collected through two types of questionnaires, one administered after each task, and one administered at the end of the experiment[1]. The questionnaire administered after each task consists of three Likert-like scale items and one open question. The three Likert-like scale items had possible answers ranging from 1 (disagree) to 5 (agree) and for each item the supervisor also invited the participant to motivate or discuss their answers. In the following, we refer to each topic using the term reported in bold (*i.e.*, **Timely**, **Clarity**, **Conciseness**, **Completeness**).

- (Likert-like scale item.) The instructions were provided timely. (**Timely**)
- (Likert-like scale item.) The instructions seem clear to me. (**Clarity**)
- (Likert-like scale item.) The instructions are concise. (**Conciseness**)
- (Open question) Are the instructions complete? (**Completeness**)

At the end of all tasks, we collected additional comments based on the following questions. Again, we refer to each topic using the term reported in bold (*i.e.*, **General differences** and **Quantity of instructions**).

- Did you perceive any differences between the two versions of the navigation app? (**General differences**)
- Did you notice any differences in the quantity and frequency of instructions provided by the two versions of the app? (**Quantity of instructions**)

4.4.3 Experimental apparatus

The study was conducted using an iPhone 14 Pro, running the two versions (**NEW** and **SOTA**) of the navigation app. Specifically, for **SOTA** we adopted a solution used in previous work [3]. Although this solution was not explicitly organized into the four modules we identify in this thesis, we note that the it addresses the *Routing* and *Guidance* sub-problems in a similar way as other solutions present in the recent literature [167]. Instead, **NEW** uses the same *Localization* and *Interface* modules as **SOTA** and the *Routing* and *Guidance* modules presented in Section 4.3.

[1]The experiments were conducted in Italian; we report the English translation.

The *Localization* module in both versions of the app is implemented with a combination of visual-inertial odometry and image tracking enabled by native Augmented Reality libraries[2]. This solution was used as it is easy to implement and can achieve accurate and stable localization (and direction) if visual markers are placed at short intervals. Indeed, in our experiments visual markers are placed in known locations (approximately every 5m) along the path, so that the drift due to inertial odometry is corrected during navigation.

Differently from **NEW** (see Section 4.3), the *Routing* module in **SOTA** adopts a state-of-the-art solution [3] using a *Navigation Graph* to represent the set of possible routes. When the navigation begins, the *Routing* module of **SOTA** computes the next target and then guides the user towards it. When the user's distance from the next target is less than the δ_e value of the next edge, the next target is updated and the procedure iterates until the destination is reached.

In contrast to **NEW**, the *Guidance* module of **SOTA** has no concept of a *rNA*. Thus, it simply produces a side-step instruction when the user exits the graph navigation area but does not guide the user back to the *rNA*. Also, **SOTA** uses an angle α fixed to 30°.

The *Interface* module, used for both **NEW** and **SOTA**, is based on prior work [3], and it implements a combination of verbal messages (rotate, walk, side-step), earcons and sonification, used to convey quantities (*e.g.*, the amount of rotation). A difference between the *Interface* module of **NEW** and the one in **SOTA** is that the former sonifies the side-step instructions while the latter does not, since in **SOTA** the user is only instructed to make a single lateral step (*e.g.*, "Step right").

To ensure that the device camera can correctly frame the visual markers, maintain a clear view of the surrounding environment, and have a consistent orientation in the direction of the participant's body, the device was positioned on the participant's chest, placed on a chest harness. The auditory information was provided through the device speaker, without using headphones, in order not to mask the surrounding soundscape, which is important for BLV travelers to avoid hazards.

4.4.4 Experimental Setting

The study was conducted on the campus of the University of Milan. We defined four routes with the same number of planned turns (*i.e.*, six turns each) and approximately the same length (103m, 99m, 97m, and 101m, respectively). For each route, the planned turns were equally divided into left and right turns. The routes were chosen to avoid hazards and objects that the participant can use to support orientation and navigation (like walls, sidewalks or trails) because we wanted the participants to rely on the provided instructions only, without being guided by other objects in the environment. For example, we wanted to avoid having participants walk along a straight path by following a wall (BLV individuals are usually trained to walk along walls using the

[2] `https://developer.apple.com/augmented-reality/arkit/`

white cane or echolocation). Furthermore, we used visual cues on the ground to clearly define the paths and determine where a marker should be placed, making it easier for supervisors to carry out the test.

4.4.5 Experimental protocol

The study was organized into four main phases: a preliminary questionnaire, a study introduction, the navigation phase, and a final questionnaire.

The preliminary questionnaire collected the participants' demographic data, expertise, and habits. The main results of the preliminary questionnaire are reported in Section 4.4.6.

In the second phase we briefly introduced our study, provided an overview of the navigation tasks, and explained to the participant how to follow the instructions provided by the application.

During the navigation phase, each participant was asked to complete four tasks. In each task, the participant would traverse one of the routes with one of the two versions of the app. The app versions and routes were counterbalanced across the participants using a Latin Square Design to account for the effects of order. Before each task, participants were introduced to the app version used in the task through a short tutorial. In the tutorial the supervisor explained which instructions and sonifications are provided by the app and how to react to them. The participant was then asked to try the navigation system on a short test route consisting of two short straight paths and one rotation. Once the tutorial was completed, participants were accompanied by the supervisor at the starting point of the route and were asked to follow the app instructions until reaching the destination. During the navigation, two supervisors followed the participant, without interfering with the navigation but being ready to take action in case of potential hazards. After each task, we collected additional subjective feedback as described in Section 4.4.2.

The fourth and last phase of the study collected additional subjective feedback and comments from the participants regarding the whole study and the perceived differences between the app versions as described in Section 4.4.2.

4.4.6 Participants

The study involved 8 participants, whose demographic data are reported in Table 4.1. The inclusion criteria for this study required the participants to be people with blindness, unable to orient themselves with residual sight (namely persons with severe visual impairment), and without multiple disabilities.

Three participants self-identified as female, the others as male. The age of participants was[3] between 24 and 66 (47 ± 14). We collected information about the participants' use of mobile devices in mobility. In particular Table 4.1 reports the participant's self-reported expertise with the use of mobile devices (Likert-like scale items ranging from 1 to 5), frequency of use of mobile devices and details about their usual screen reader settings. We also investigated how

[3] We use the notation $x \pm y$ to denote average x and standard deviation y.

Table 4.1: Participants' demographic information. G.=gender. For impairment: FB=fully blind. For route frequency: Fam.=familiar, Unfam.=unfamiliar, D=daily, W=weekly, MO=monthly, R=rarely.

			Impairment		Mobile		Screen reader		Route	
PID	**Age**	**G.**	**Level**	**Onset**	**Expertice**	**Frequency**	**speed**	**voice**	**Fam.**	**Unfam.**
P1	55	M	FB	birth	4	D	70	M	W	MO
P2	42	F	FB	birth	4	D	80	F	D	D
P3	53	F	FB	birth	5	W	75	M	D	R
P4	46	M	FB	birth	5	D	>=75	F	D	W
P5	24	M	FB	birth	5	D	65	F	D	R
P6	28	F	FB	birth	5	W	70	M	W	MO
P7	66	M	FB	age 6	3	D	55	M	D	R
P8	62	M	FB	age 10	5	W	65-85	F	D	MO

participants listen to the screen reader in mobility. Four participants reported that they listen to the screen reader while walking using the device speaker (**P1**, **P6**, **P7**, **P8**). **P6** added that she uses stereo earphones depending on the situation. Other 4 participants use single earphone while walking (**P2**, **P3**, **P4**, **P5**). All the participants use app during mobility such as *Ariadne GPS*[4], *BlindSquare*[5], *FourSquare*[6], *Google Maps*[7], *iMove*[8], *Apple Maps*[9], *Moovit*[10], *Siri*[11], *Soundscape*[12] and *VoiceVista*[13].

Finally, we asked the participants how often they travel independently along familiar and unfamiliar routes. Six participants reported traveling familiar routes daily, the others weekly. Instead, only one participant travels unfamiliar routes daily, the others weekly (one participant), monthly (3 participant) or even more rarely (3 participants).

4.5 Experimental results

4.5.1 Analysis of movement data

All participants completed the tasks and navigated independently during the large part of the experiment. On 24 occasions the supervisor had to temporarily stop the experiment to prevent possible hazards (*e.g.*, although the area was closed to traffic, on a few occasions vehicles passed by) or to recover from major localization problems (*e.g.*, the position markers, which were placed on

[4] `https://www.ariadnegps.eu/en/`
[5] `https://blindsquare.com`
[6] `https://apps.apple.com/it/app/foursquare-city-guide/id306934924`
[7] `https://www.google.com/maps/`
[8] `https://apps.apple.com/us/app/imove-around/id593874954`
[9] `https://www.apple.com/maps`
[10] `https://www.moovit.com/`
[11] `https://www.apple.com/siri`
[12] `https://www.scottishtecharmy.org/soundscape`
[13] `https://apps.apple.com/it/app/voicevista/id6450388413`

supports, were moved or knocked over by the wind). In these cases, the supervisor took note of the intervention and later edited the collected data, to exclude the data automatically collected by the system during the supervisor intervention.

In total, we collected data from 32 paths (4 paths for each participant), equally divided between the two conditions. Participants walked for $720.77s \pm 138.21s$ seconds in total and covered a distance of $414.01m \pm 15.64m$. The total number of instructions for each participant was 123.5 ± 26.54, with 56.75 ± 9.86 rotate instruction, 60.38 ± 14.9 walk instructions and 6.38 ± 3.64 side-step instructions.

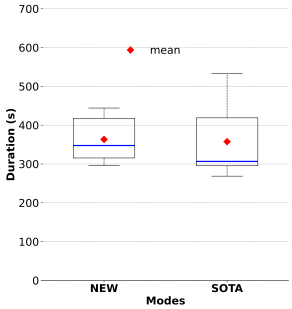
Duration and number of instructions

Fig. 4.7 shows a comparison of **NEW** and **SOTA** in terms of duration of the navigation and number of instructions. Considering the navigation time, no statistically significant difference emerges between **NEW** ($363.33s \pm 53.69s$) and **SOTA** ($356.98s \pm 88.77s$). Instead, considering the total number of instructions, **NEW** (51.62 ± 8.99) had significantly less ($p < .01$, $t(7) = -5.27$) instructions than **SOTA** (71.88 ± 17.98). We also conducted additional analyses, considering only turn, walk, or side-step instructions between the two conditions. In this case, both factors (app version and instruction type) yielded significant differences ($p < .001$) as well as their interaction ($p < 0.05$). Pairwise comparisons show that **NEW** had significantly less turn (**NEW**: 22.75 ± 4.02 vs **SOTA**: 34 ± 6.18, $p < .001$, $t(7) = -8.71$) and walk (**NEW**: 25.88 ± 4.86 vs **SOTA**: 34.5 ± 10.27, $p < .01$, $t(7) = -3.79$) instructions, while no significant difference was found in the number of side-step instructions (**NEW**: 3 ± 1.41 vs **SOTA**: 3.38 ± 2.91).

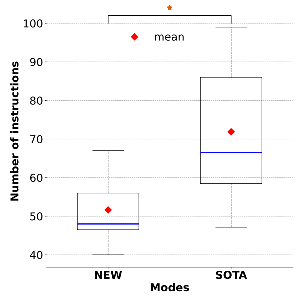
Time outside the navigation area

Fig. 4.8 shows how much time the participants spent outside the navigation area (**NEW**: $17.81s \pm 13.73s$, **SOTA**: $44.28s \pm 18.81s$). The difference is statically significant ($p < .05$, $W = 0$). A related metric is the average distance from the user's position to the current edge (see Fig. 4.8b). Also in this case (**NEW**: $0.85m \pm 0.06$, **SOTA**: $1.13m \pm 0.21m$), the difference is statistically significant ($p < .05$, $t(7) = -4.19$).

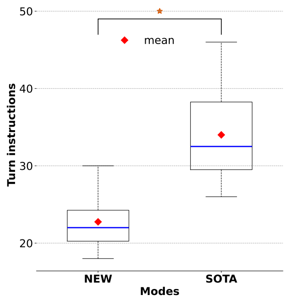
To better understand the phenomenon, we also analyzed the time spent outside the navigation when the current edge had a δ_e (the "radius" of the navigation area, see Section 4.3.2) equal to 1, 2, and 3 meters (see Figs. 4.8c, 4.8d and 4.8e, respectively). Both factors (app version and navigation area radius) yielded significant differences ($p < .001$) but not their interaction. The difference is more accentuated (**NEW**: $6.52s \pm 2.86s$ vs **SOTA**: $29.83s \pm 9.27s$) when $\delta_e = 1m$ ($p < .001$, $t(7) = -5.84$). Instead, when $\delta_e = 2m$ or $\delta_e = 3m$ the difference is not statistically significant.



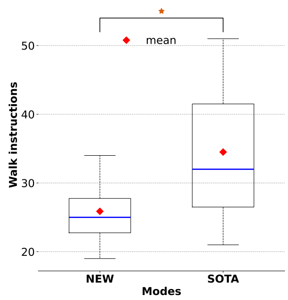
(a) Overall time duration



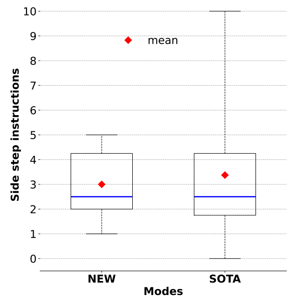
(b) All instructions



(c) Turn instructions



(d) Walk instructions



(e) Side-step instructions

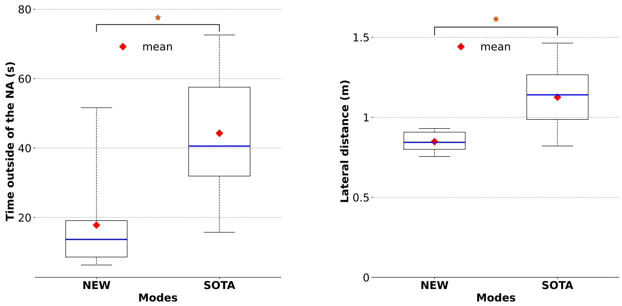
Figure 4.7: Total navigation time and number of instructions ($\star$ significant).

Instructions burst

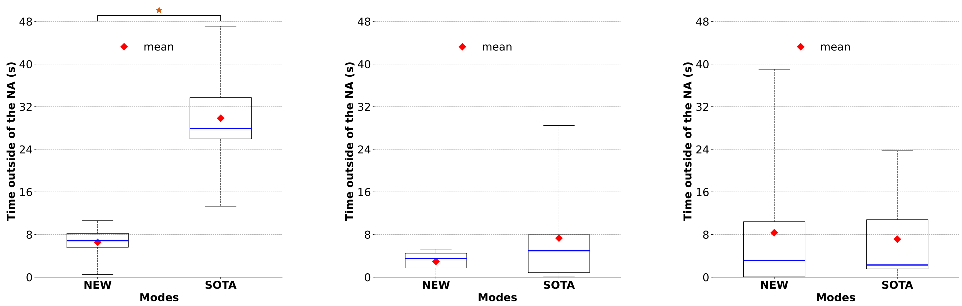
A possible undesirable behavior of a navigation system is to provide a "instructions burst": a sequence of requests one after another. To capture this situation, we define the concept of instructions cluster (Section 4.4.2) considering a maximum time distance between two consecutive requests, and we compute the average size of the instructions clusters for each participant, comparing **NEW** and **SOTA** (see Fig. 4.9). For all considered maximum time distances, the differences are statistically significant. Specifically, both app version factor ($p < .001$) and maximum time distance ($p < .05$) yielded significant differences ($p < .001$) but not their interaction. For $t = 0.5s$ the average size of instruction clusters is $1.01s \pm 0.02s$ for **NEW** and $1.10s \pm 0.06s$ for **SOTA** ($p < .01$, $W = 0$). For $t = 1s$, it is $1.02s \pm 0.02s$ for **NEW** and $1.17s \pm 0.06s$ for **SOTA** ($p < .001$, $t(7) = -7.70$), and for $t = 2s$ it is $1.04s \pm 0.02s$ for **NEW** and $1.29s \pm 0.08s$ for **SOTA** ($p < .001$, $t(7) = -8.14$).

4.5.2 Participants' Subjective Feedback

Fig. 4.10 shows the distribution of the scores for the Likert-like scale items (see Section 4.4.2). Considering the **Timely** (**NEW**: 4.13 ± 0.86, **SOTA**: $3.88 \pm$

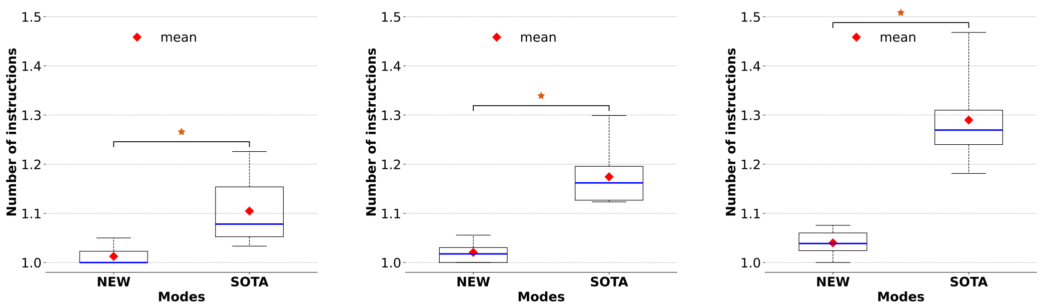


(a) Time outside the NA (b) Lateral distance from current edge



(c) Time outside the NA ($\delta_e = 1m$) (d) Time outside the NA ($\delta_e = 2m$) (e) Time outside the NA ($\delta_e = 3m$)

Figure 4.8: Time outside the navigation area (NA) and average distance from edge ($\star$ significant).



(a) Instruction cluster size $t = 0.5s$ (b) Instruction cluster size $t = 1s$ (c) Instruction cluster size $t = 2s$

Figure 4.9: Size of instruction clusters ($\star$ significant).

0.86)[14] and **Conciseness** (**NEW**: 4.75 ± 0.43, **SOTA**: 4.75 ± 0.43) metrics, no statistically significant differences were found. Instead, for **Clarity**, **NEW**

[14]We report means for multipoint items as they were found to better indicate central tendency than medians [126].

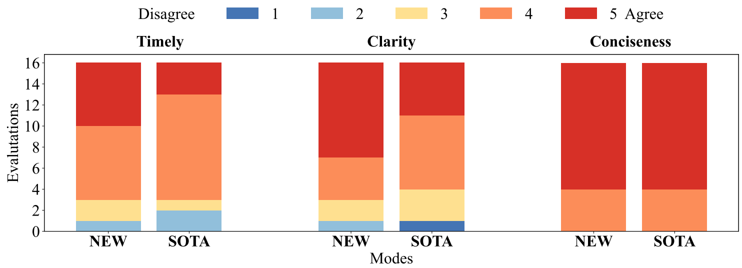


Figure 4.10: Participant' subjective feedback for likert-like items.

(4.31 ± 0.923) was perceived to be significantly clearer ($W = 4.5$, $p < .05$) than **SOTA** (3.94 ± 1.03).

Considering the **Timely** topic, some participants (**P2**, **P5** and **P8**) clarified the answer explaining that it is not desirable to have a system that is too quick in providing instructions.

> **P2**: *[With **SOTA**] you risk being too prompt if you correct me immediately.*

> **P5**: *[With **SOTA**], the instructions kept changing frequently, it was too real-time.*

Instead, considering **Clarity**, four participants (**P2**, **P3**, **P4** and **P5**) reported encountering the instructions cluster problem when using **SOTA**.

> **P3**: *I received two opposing instructions very close to each other, and I didn't have time to follow the first one, meaning I couldn't perform the action of the first; instead, I carried out the action of the second.*

Similarly, one participant reported frequent unnecessary instructions with **SOTA**:

> **P5**: *[With **SOTA**], the instructions kept changing frequently. It was too real-time. I couldn't orient myself properly right away. [With **NEW**], I was oriented and then I reached the end of a path. I never had to change direction mid-course.*

Also **P4** observed that in two cases he had to immediately correct himself after the first instruction, such as turning left and then right.Finally, considering **Completeness**, three participants (**P3**, **P6**, **P7**) observed that both version of the navigation app lack of a functionality to identify and avoid objects.

Considering the **General differences** question, **P5** perceived a difference in the "guidance mode". Instead, considering the **Quantity of instructions** question, one participant was unsure (**P1**), three did not perceive any difference in the number of instructions (**P3**, **P4**, **P6**), two perceived that **NEW** provided fewer instructions (**P5**, **P7**), and two perceived that **SOTA** provided fewer instructions (**P2**, **P8**).

4.6 Discussion

We discuss the key results of our investigation, namely the benefits of the design principles that we adopt in our solution (Section 4.6.1), and the benefits of the problem decomposition approach (Section 4.6.2). We also address the key limitations of our methodological approach (Section 4.6.3).

4.6.1 Benefits of the Adopted Design Principles

The experimental results give evidence that the *Routing* and *Guidance* modules have a significant impact on the user experience. In particular, our experiments show that the adoption of the **safety** and **parsimonious correction** design principles can reduce a) the number of instructions, b) instructions bursts, and c) the time spent outside the navigation area. Reducing the total number of instructions makes the system less invasive (in the sense that it reduces the number of audio instructions) and more secure since it allows the user to focus on surrounding audio feedback (*e.g.*, approaching vehicles). Reducing the number of instructions also impacts on the instructions bursts, that, as reported by the participants, have a negative impact on the clarity of the navigation system. Indeed, the proposed solution was also perceived to be significantly clearer than the benchmark. Finally, reducing the time spent outside the navigation area is also relevant: the navigation area represents a location where the user is expected to move (*e.g.*, the sidewalk), while the region outside the navigation area (*e.g.*, the roadway) can be potentially hazardous.

4.6.2 Benefits of the Problem Decomposition Approach

In this work we suggest to decompose the overall problem of navigation assistance for people who are BLV into four sub-problems. We argue that this approach has three main advantages. First, it is a conceptual framework to better analyze and classify the contributions in the state of the art, most of which focus on the *Localization* or the *Interface* sub-problems. Second, the identification of the *Routing* and *Guidance* sub-problems suggests that specific solutions should be designed as well. Instead, most solutions in the state of the art disregarded these two problems which, as we show in this study, can have a significant impact on the navigation experience. In this perspective we believe that our work poses the basis for a stream of research aimed at studying (and comparing) other solutions for these specific sub-problems. The third advantage is related to software modularization: the identification of four independent software components facilitates software re-use. For example, the app prototype that we developed for the experiments can be easily modified to adopt a different *Localization* module (*e.g.*, based on Bluetooth beacons as the *Navcog* system [167]). It should also be possible to experiment with different *Interface* modules as well.

4.6.3 Limitations

We discuss three key limitations of our work: due to experimental setting, due to the apparatus used, and due to the limited demographic and cultural variety among the participants.

Experimental Setting Limitations

In this work we evaluated the impact of the proposed navigation approach in a real-world scenario. Specifically, the experimental evaluation was conducted on a university campus for three main reasons: to limit the hazards for the participants, to limit unpredictable obstacles along the route (*e.g.*, a parked bicycle) that could impact the study results, and to test the system in an environment where participants would not be guided by environmental features (*e.g.*, a corridor) that can be used as references, haptically (*e.g.*, by following a corridor wall with the white cane) or acoustically (thanks to echolocation, BLV travelers can, for example, sense the presence of an open door in a corridor).

These characteristics, however, might not be present in other scenarios, and hence could influence the navigation process and its outcome. For example, in our study, the participants knew that, in case of a hazard during the experiment, the supervisor would intervene. This might have impacted the way the participants navigated or affected their walking speed. Similarly, in an environment with unpredictable obstacles, the participants would have most likely had to avoid them, which would have influenced their walking routes. Also, while there was some noise in the area due to the presence of numerous students and a few passing cars, the overall background noise was different than in a typical outdoor environment with many passing vehicles. The difference in the background noise could have potentially influenced the ability of the users to perceive the instructions and the sonifications.

There are also two factors that made the experimental setting harder to navigate than routes usually traversed by the participants. First, the participants had no previous knowledge of the experimental routes. This is particularly challenging because BLV individuals most commonly travel on familiar routes or study the route beforehand, as shown in Section 4.4.6. Second, the participants had no environment reference points, and in particular no haptic references. One common practice among BLV individuals who use the white cane (as all our participants did) is to follow haptic references, like walls or sidewalk edges. In the experimental setting, we intentionally decided to avoid these references, so that they would not affect the results, but clearly, this makes the walking task much more challenging.

Apparatus Limitations

Study results may have also been influenced by the experimental apparatus used. Specifically, we did not measure the impact of the localization accuracy since this was out of the scope of the investigation. However, since the navigation instructions depend on the system localization accuracy, we acknowledge that this factor could be relevant as well. To account for this issue, we positioned the localization markers approximately every 5m, thus mitigating the localization drift. We empirically observed that the drift effect was limited with the only exception (excluded in the analysis of the results) in which the location markers were accidentally moved or knocked down (e.g., by the wind).

Another peculiarity of the study is the use of a chest harness. This study design choice helped to simplify the experimental protocol because, if the participants held the device with their hands, it would have been harder to distinguish body rotations from hand rotations. Participants would have also needed to concentrate on pointing the device at the right angle (*e.g.*, not towards the ground), which is known to be hard for BLV individuals [138]. However, a chest harness may not be a practical solution for a number of reasons, including that its acceptance by the users needs to be evaluated and the smartphone can be easily stolen. Alternative solutions could be wearable devices, like smart glasses, and other types of bags or ways of securing the smartphone.

Demographic and Cultural Bias

In this work, since the experimental setting was constructed in a specific real-world environment, we could only recruit participants from the same geographical area. Furthermore, in order to limit confounding factors that could impact the results, we only involved participants who are blind with no residual vision. As a result, the study involved a small number of participants, a limitation that is common in the field of research on assistive technologies [172].

Chapter 5

Noise adaptation techniques

BLV People access mobile devices using screen readers. However, noisy environments can impair screen reader intelligibility. During mobility, this could disorient or even endanger the user. To address this issue, we propose three screen reader speech compensation techniques based on environmental noise: speech rate slowing, adaptive volume increase, and adaptive equalization. Through a study with twelve participants in three simulated noise scenarios, we evaluate screen reader intelligibility and the perceived distraction from the soundscape, with and without compensations. Four of the proposed compensations, in particular those that pair speech rate reduction with volume or equalization adaptation, significantly improve screen reader's speech intelligibility in all the considered scenarios, and the compensations do not have a significant impact on the distraction from the soundscape.

5.1 Background

BLV people commonly use screen readers [103, 195, 103], to access the native functionalities of mobile device [104] as well as assistive capabilities provided by third-party developers. However, screen reader usage during mobility presents three main issues. First, many BLV people feel that the speech feedback is intrusive to people around them, or even perceived negatively by others [175]. This effect is even more apparent in people with invisible disabilities, like mild low-vision [61]. Second, the environmental noise may mask screen reader speech [33, 117], preventing people from hearing it. This problem is especially present in very loud scenarios, such as the subway [1], crowded places, and while walking in the traffic [1]. Because of this, BLV people frequently use headphones while listening to screen reader during mobility [193]. Third, the soundscape may also be partially covered by the screen reader speech, in particular when using headphones [50]. Thus, the speech could draw away the user's attention from the environmental sound cues which may be useful for orientation (*e.g.*, acoustic traffic signals) or for avoiding dangerous situations (*e.g.*, an approaching car). Most users partially offset such issues by using a single headphone, while only a minority adopt pass-through or bone conduction headphones due to high prices, unfamiliarity with these solutions, and concerns with their output sound quality level [193]. Manual adaptation of speech rate [33] or volume [1] is also reported as possible coping mechanisms. However, the users may not have the time to manually change the setting in case of sudden noise and the action can distract them from their current activity.

Approaches for the automatic adaptation of sound based on ambient noise, especially in mobile contexts, have been proposed using several different techniques. These techniques can be grouped into two broad clusters: active noise cancellation and noise-based processing of the information carrying audio signal. In the first cluster, active noise cancellation headphones or headsets use adaptive audio processing for reducing the background noise [142]. However, noise-canceling solutions are not suited for the purposes of this work since important sound cues from the environment may be rendered inaudible, making mobility more difficult or even endangering the user [205].

The second cluster comprises methods that automatically adjust volume, dynamic range, or other sound features based on ambient noise, which is in turn monitored and analyzed through built-in microphones on mobile devices. One example is the adaptive control of the smartphone volume based on user activity and ambient noise, with the purpose of improving the perception and recognition of alert and notification sounds in noisy environments [17]. Another example is the adaptive control of the dynamic range (compression) of the audio being played, depending on the level of the environmental noise around the listener, which is measured using the microphone on the mobile device [140]. None of the previous works address the challenge of improving screen reader intelligibility.

5.2 Contribution

To address that the noisy environments impair screen reader intelligibility masking the screen reader output, our work [8] proposes a solution to dynamically adapt the screen reader output based on noise-based processing. The goal is to improve the screen reader intelligibility, without increasing the distraction from the surrounding soundscape and without permanently reducing the screen reader speech rate. To this end, we designed three compensation techniques:

- **Rate** – Applies a flat speech rate reduction in presence of environmental noise;
- **Vol** – Adaptively increases or decreases speech volume based on the intensity of environmental noise;
- **Eq** – Adaptively increases or decreases speech volume only for frequencies impacted by environmental noise.

A preliminary evaluation, conducted with 4 BLV participants, assessed the effect of the proposed techniques on the screen reader speech intelligibility in presence of background noise. The participants commented that making the speech both louder and slower would make it easier to comprehend. Thus, in a second preliminary evaluation, conducted with other 6 representative participants, we also included conditions obtained as combinations of speech slowing with the other two compensations: **Rate+Vol** and **Rate+Eq**. A final evaluation, conducted with 12 participants, also evaluated the distraction caused by the screen reader speech with respect to the background soundscape.

Specifically, in the main study, we evaluated the compensations in three typical noise scenarios [1]: **Crowd**, **Traffic**, and **Subway**. The scenarios were realistically simulated in a silent chamber with a quadraphonic audio setup, playing 4-channel real-world recordings at a sound pressure level consistent with the real situation. For the experiment, we used a corpus of text sentences [158], read by the screen reader and conveyed through bone conduction headphones, with and without compensation. We measured speech intelligibility as the percentage of the correctly understood words, and the distraction caused by the screen reader speech with respect to the soundscape by asking the participants to pinpoint the direction of a contextual sound played concurrently with the speech feedback.

The results confirm that noisy environments severely impact screen reader speech intelligibility but compensation approaches can mitigate this effect significantly. In particular, the combined compensations (**Rate+Vol** and **Rate+Eq**) provide consistent improvement in all soundscape scenarios. The perceived distraction from the surrounding soundscape, caused by the screen reader speech, does not significantly change when compensations are used. This indicates that the proposed compensations can improve speech intelligibility without impacting the user's ability to pay attention to the surrounding environment. Therefore, our approach is a practical solution to improve the understanding of screen reader speech feedback in noisy environments and during mobility.

5.3 Compensation Techniques Design

We designed three base compensation techniques (**Rate**, **Vol** and **Eq**) and two techniques combining **Rate** with the remaining compensations (**Rate+Vol** and **Rate+Eq**). In the following, we indicate that no compensation technique was used with the term None. The compensation techniques have been designed to be deployable on mobile devices, without requiring proprietary hardware: they only need an audio output device (*e.g.*, headphones), a microphone (possibly integrated in the headphones), and limited computational power.

5.3.1 Adaptive Volume

The key idea of dynamic volume compensation (**Vol**) is to adjust the speech volume in an adaptive way, depending on the environmental noise that surrounds the device. The goal is to keep the *Signal to Noise Ratio* (SNR) between

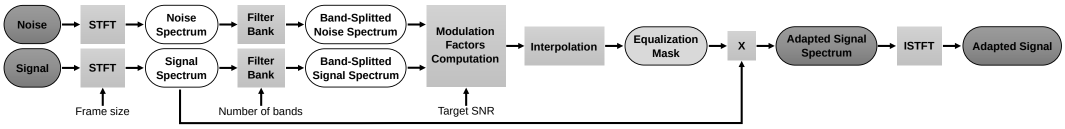


Figure 5.1: Adaptive sonification based on spectral equalization.

the speech level and the soundscape level constant but, at the same time, to limit the intervention to an acceptable range (in order to avoid lowering the volume too much in quiet scenarios or saturating the device in loud scenarios).

To achieve this, a gain factor $k(t)$ is computed as the ratio between the target SNR, which can be tuned based on user's hearing and preferences, and a running measure of the actual SNR, denoted in Eq. (5.1) by SNR_{target} and $SNR_{running}(t)$ respectively.

$$k(t) = \frac{SNR_{target}}{SNR_{running}(t)} = SNR_{target} \cdot \frac{RMS_{\text{noise}}(t)}{RMS_{\text{signal}}(t)}, \tag{5.1}$$

In Eq. (5.1) RMS stands for *Root Mean Square*, more specifically, the *running* RMS, *i.e.* the RMS evaluated on a moving temporal window. The larger the window, the smoother the intervention, at the cost of slower responsiveness. This processing completely occurs in the time domain. **Vol** compensation can be fine-tuned by setting two main parameters: SNR_{target}, and the running RMS window size. Moreover, in order to avoid excessive volume compensation (low k values would render the sound almost inaudible in quiet environments, while high values would result in audible signal distortion and may cause pain or even hearing damage), we added the possibility of limiting $k(t)$ to a range $R_k = [k_{min}, k_{max}]$.

We chose the value $SNR_{target} = +1.2$ as it was found to be the preferred value for speech signals in a previous study on acceptable noise levels and preferred SNRs for speech and music [124]. Values for $R_k = [-3, +20]$ dB were chosen in order to avoid signal clipping as well as to minimize attenuation. For the estimation of the running RMS, we used a temporal window with duration 46ms and an overlap of 23ms between successive windows, in order to minimize artifacts (glitches) resulting from an exceedingly fast adaptation.

5.3.2 Adaptive Equalization

The goal of this technique is to use spectral modifications, rather than volume adjustment, to reduce the masking of the verbal message by the background noise. In fact, the concept of masking does not merely depend on volume; rather, it consists in the alteration of the perception of the single frequency components of the message if the noise occupies the same spectrum region with sufficiently high energy. Thus, adaptive Equalization compensation (**Eq**) accounts for the proper *auditory masking* effect by realizing a multi-band version of **Vol**. First, the signal is split into a number of frequency bands, then a different modulation factor is computed for each band, so as to reduce masking effects only where needed.

Fig. 5.1 shows the process to compute the adapted signal. The first step is to take the *Short Time Fourier Transform* (STFT), to extract the spectra of both the signal and the noise. Then, these spectra are processed through an ad-hoc filter bank to divide them into a given number of bands. Ideally, the band subdivision should approximate what in psychoacoustics is called a *critical band*, that is the band of audio frequencies within which a second tone will interfere with the perception of the first tone by auditory masking, since they activate the same area of the basilar membrane [63]. Equivalently, this filter bank should serve the purpose of simulating the human auditory system by modeling those band-pass filters, whose width follows the Equivalent Rectangular Bandwidth (ERB) scale [143]. Nevertheless, when dealing with speech signals, such a fine subdivision in frequency bands followed by a per-band gain modulation may result in a complete cancellation of vocal formants, which are the spectral characteristics of vocal signal conveying information. To avoid this loss of information, we opted for a reduced number of bands. Finally, a different gain factor is calculated for each band (whereas for **Vol** a single modulation factor is applied to the whole signal). The per-band gain factors are computed with the same approach summarized in Eq. (5.1), the only difference being that values are now computed for each band. To apply the computed gain changes, instead of multiplying each band for the corresponding factor and summing the bands together (which would introduce some phase issues), we interpolate the factors so to obtain an equalization curve with the same resolution of the STFT frames. In particular, we use a Piecewise Cubic Hermite Interpolating Polynomial, which has the property of preserving the maximum and minimum points of the initial function, thus avoiding equalization overshooting or undershooting. Once the desired equalization curve is computed, multiplying it by the spectrum of the signal returns the spectrum of the adaptive signal. The latter is brought back into the time domain through the Inverse Short Time Fourier Transform (ISTFT), thus obtaining the adaptive audio signal.

For all the parameters in common with the **Vol** compensation, we chose the same values. In addition to those, **Eq** compensation requires one additional parameter, which is the number of bands. We used 5 bands, which resulted in good preservation of the vocal formants, as discussed above.

5.3.3 Adaptive Speech Rate

BLV people have higher listening rates than sighted people [33] and usually keep the screen reader at the highest possible speed allowing them to understand the speech in a quiet environment, as also confirmed by our participants. However, it is yet unclear how noise impacts screen reader intelligibility by BLV people [33]. Additionally, as reported in prior works, expert screen reader users often change their screen reader speech rate based on context [1], and this approach is also considered a possible adaptation to improve screen reader intelligibility in presence of noise [33].

Speech rate compensation (**Rate**) applies a flat reduction of the screen reader speech rate when noise is detected. We implemented this as a 15% speed reduction with respect to the *base speech rate* (*i.e.*, participant's preferred speech rate). This parameter was empirically selected after interacting with the preliminary study participants: when we asked for their preferred speech rates, six out of ten answered with a range of rates and in four cases the range size was 15% (in the other two cases the range was 5%). To avoid the speech rate to change within a sentence or a single word, the solution does not adapt continuously based on the changes in the ambient noise. Investigating the design space of the continuous screen reader speech rate adaptation based on the changes in environmental noise is a possible future extension of our work.

5.3.4 Combined Compensations

Based on the comments from the participants in the preliminary studies, we have also explored the combinations of the proposed compensation techniques. The three base compensation techniques change the way speech is reproduced (**Vol**, **Eq**), or the way it is generated (**Rate**). Combining **Vol** and **Eq** is not meaningful, as it would merely reduce to an **Eq** technique with a different SNR. Instead, we combine **Rate** with either **Vol** or **Eq**, obtaining two additional techniquee, **Rate**+**Vol** and **Rate**+**Eq**, in which the speech is generated at the speed defined by **Rate** and reproduced after applying **Vol** or **Eq**.

5.4 Evaluation

We conducted a set of user studies to assess the effect of the proposed compensation techniques on speech intelligibility in the presence of noise, and on the distraction caused by the screen reader speech with respect to the environment soundscape. The research was approved by the Ethics Committee of our University. As part of the iterative design process, we first evaluated our techniques with sighted participants [71], then we conducted three studies with representative BLV participants. The first two of these are preliminary studies, which are described in Section 5.4.1. The last one is the main study, which is described in detail in the rest of this section.

5.4.1 Preliminary Studies

Two preliminary studies were conducted with 4 and 6 BLV participants, respectively, as a part of our design and parameter tuning process. In the first study, we assessed the three base compensation techniques (**Rate**, **Vol**, **Eq**). For this, we simulated a noisy soundscape and reproduced speech feedback (possibly compensated) that the participants were asked to listen and repeat (see Section 5.4.3). Motivated by the participants' comments, in the second study we introduced the combined compensations: **Rate**+**Vol** and **Rate**+**Eq**. We also corrected the compensation parameters, in particular, to avoid the crackling noise that would appear with specific sound frequencies on bone conduction headphones, as mentioned in Section 5.4.3.

Considering the evaluation methodology, we introduced two main changes between the preliminary studies and the main study. First, in the preliminary studies, we assessed distraction as a subjective measure reported by the participants. Instead, in the main study, we introduced an objective measure of the distraction caused by the speech feedback, by assessing the participants' ability to pinpoint the direction of an environmental sound, reproduced concurrently with the speech. The second difference regards the speech rate. In the preliminary studies, all participants used the same speech rates. Instead, in the main study, we configured the screen reader speech rates based on the preferred settings for each participant.

5.4.2 Audio Stimuli

In order to quantitatively assess both the speech intelligibility and distraction, we prepared a set of **audio stimuli**, each combining a **soundscape**, a (possibly compensated) **speech signal** that the participants are asked to listen and repeat despite the background soundscape, and a **contextual sound**, the direction of which the participants are asked to pinpoint while the speech feedback is played. More specifically, audio stimuli are audio tracks with six channels. Four channels reproduce the soundscape with its contextual sound from one of three directions (*i.e.*, left, right, front) and two channels reproduce the speech signal. The speech signal is always reproduced at the same time offset with respect to the soundscape in order to have consistent listening conditions. Instead, each contextual sound is played in a random instant while the speech signal is played.

To generate the **speech signals**, we followed the approach proposed in [158]. Each sentence (in Italian) is constructed from a word table with 5 columns and 10 rows, by randomly picking one word from each column. The combination of the extracted words forms a sentence that is grammatically correct but semantically unpredictable. The resulting combinations have the advantage of focusing the participant's attention on the actual comprehensibility of the sentences. The sentences were reproduced using the *VoiceOver*[1] screen reader (female voice). The screen reader allows the user to specify a preferred speech rate, indicated as a percentage of the maximum speed. Thus, for each participant, we generated personalized **speech signals** by specifying the speech rate that the participant is used to.

Four **soundscapes** were selected, three with a high and one with a low noise level. Suburban soundscape (low noise) is a baseline condition in which the speech compensation is not needed. It consists of a recording of the noise in a suburban residential area with low traffic. The noisy soundscapes are a subway station during train arrival (**Subway**), a crowded market (**Crowd**), and a trafficked city street (**Traffic**) as they were reported to be particularly challenging scenarios [1]. Each soundscape has a duration of 15 seconds.

For each soundscape, we also prepared a **contextual sound**: the sound of a closing gate in **Suburban**, the sound of a cash receipt being printed

[1] `https://www.apple.com/accessibility/vision/`

in **Crowd**, the sound of opening bus doors in **Traffic**, and the intermittent signal of opened doors of a subway in **Subway**. All the contextual sounds were chosen without introducing vocal sounds in the soundscapes, to avoid confusion with the speech signal. Each contextual sound exists in three variations: one is played from the participants' left, one from the right, and one in front of the participant. The recordings of the soundscapes and contextual sounds are available online[2].

5.4.3 Apparatus

The experimental set-up mimics a real-world scenario in which ambient noise reaches the listener's ears through a purely acoustical path, while the speech signal is delivered through headphones. Consequently, the audio stimuli described in the previous section were reproduced through two distinct audio streams: one for the soundscapes and the contextual sound, the other for the speech signal.

It was first necessary to recreate a realistic simulation, in order to provide participants with the impression of being immersed in an everyday acoustic environment. Such a goal was addressed through a quadraphonic reproduction system: four audio channels were routed to the corresponding speakers arranged into a square, with the listener in the center. The speakers are positioned on the front-left, front-right, back-left, and back-right with respect to the listener. With respect to stereophony, quadraphony allows a better spatialization of the sound and, consequently, the possibility for the listeners to better locate the sound events presented to them. The speakers were installed in a silent chamber (where the tests took place), a soundproof room acoustically treated to dampen sound reflections, thus enabling to simulate wide sound scenarios, despite its limited size. Contextual sounds are played by two speakers, depending on the direction. For example, when the contextual sound is played on the left of the participant, the two left speakers (*i.e.*, front-left, back-left) are used.

Speech signals were conveyed to the participant using Z8 Docooler bone conduction headphones. This choice had a twofold motivation: first, they ensure the maximum possible transparency with respect to the external soundscape, as they do not occlude the ear canal; second, they guarantee a good response in the characteristic speech frequency range (*i.e.* 200 Hz to 4 kHz). Nevertheless, we experimentally verified that these specific headphones introduce an unpleasant vibration around 230 Hz, therefore we processed all the headphone signals with a bell equalizer centered on the aforesaid frequency, with a filter gain of -4.5 dB, and an overall gain of $+2.4$ dB in order to linearize the device response. We calibrated the system to produce an uncompensated speech signal with average $SNR = 0$ with respect to a baseline outdoor soundscape without strong sources of noise (*i.e.*, **Suburban**). This means that the uncompensated speech signal is tuned to be intelligible in such a scenario, without being intrusive to others. Experiments show that this is indeed the

[2] `https://noise-soundscapes.netlify.app/`

case (see Section 5.5).

5.4.4 Evaluation Protocol

The evaluation was organized into five phases: initial questionnaire, instructions, calibration, listening tasks, and a final open-ended questionnaire. For each participant, the experiment lasted for about one hour. The initial questionnaire collects the participants' information (see Section 5.4.6). In the instruction phase, the participant is invited to seat on a chair in the silent chamber and to wear the bone conduction headphones. Then, the supervisor explains how the experiment works and in particular that the participant has to repeat the sentence in the speech signal as they understand it and indicate with their finger the direction from which the contextual sound comes (left, right, front). While explaining the test, the supervisor reproduces each soundscape, an example of a speech signal, and the contextual sounds.

During calibration, the base speech rate to be used during the test is assessed. For this, the **Suburban** soundscape is used, as it represents a situation with little environmental noise. The supervisor plays an audio stimulus with the preferred speech rate reported by the participant. If the participant can correctly understand the sentence, this speech rate is selected. Otherwise, the process iterates with a speech rate slower by 5%, until the participant can correctly understand the sentence.

The listening task phase consists of a set of 61 tasks. During each task, the supervisor plays an audio stimulus and the participant repeats the sentence and indicates the direction of the contextual sound. The supervisor takes note of the answers provided by the participant. Four initial tasks are used for training, to ensure that the participant correctly understood what to do. The results of these tasks are not recorded. The following 57 tasks include three audio stimuli with the **Suburban** soundscape and 54 with the noisy soundscapes (*i.e.*, **Crowd**, **Traffic**, **Subway**). The three audio stimuli with the **Suburban** soundscape serve as a baseline and use an uncompensated speech signal. Each of them has a contextual sound played from a different direction. For each noisy soundscape there are 18 audio stimuli, 6 for each direction of the contextual sound: one without compensated speech signal, others with a different compensation each. The 57 tasks were organized into 3 sessions, separated by a short break of about 2 minutes. To minimize effects of order, compensations and directions of contextual sound were counter-balanced with a Latin-square design.

Finally, we asked a series of open-ended questions (see Table 5.2), investigating participants' opinions on the presented compensations and the experiment in general. Additionally, we collected their comments and suggestion for improvements. This part of the experiment was organized as a semi-structured interview and, starting from the questions, the supervisor invited the participant to discuss and report comments.

5.4.5 Metrics and Data Analysis

We define the soundscape and compensation technique (if any) as independent variables. The dependent variables are:

- **Speech Intelligibility**, defined as the percentage of the correctly understood words for each listening task;
- **Perceived Direction**, defined as the percentage of the correctly identified contextual sound directions.

To account for the small number of participants, common in accessibility research [172], a repeated measure design is adopted [36, 172]. Score differences were analyzed for statistical significance using Friedman Chi Square test [68], with Dunn post-hoc test for pairwise comparisons [54]. We also verified, for each scenario and compensation condition, that there was no significant learning effect across the tasks, using the Mann Kendall Trend Test [108, 133]. Benjamini–Hochberg FDR method [25] was used for multiple test corrections.

During the evaluation, for each participant, there are exactly three tasks with the same pair of soundscape and compensation. We compute the average for speech intelligibility and perceived direction metrics among these three tasks. Each average value is then used as a data point in the analysis.

5.4.6 Participants

We recruited 12 BLV participants (3 female) through local associations, social networks, and word-of-mouth. The recruiting criteria required that (a) the participant was not involved in the preliminary studies, (b) the participant is blind, according to the World Health Organization Classification [185], and (c) the participant does not have a hearing impairment (normal hearing). As reported in Table 5.1, participants had an average age of 47 ($SD = 14.63$) and they were all blind since birth with the exception of **P3**, **P4**, and **P11**.

We collected participants' self-reported expertise with mobile devices and music, as Likert-like scale items ranging from 1 (no expertise) to 5 (high expertise). Average mobile device expertise was 3.73 ($SD = 1.14$), while musical expertise scores averaged 3 ($SD = 1.41$). All the participants use *VoiceOver* screen reader on iOS devices, except **P1** who uses *TalkBack* on an Android device. All except for **P10** use the screen reader with a female voice. Preferential screen reader speech rates reported by the participants ranged between 55% and 90% with an average of 66% ($SD = 10.57\%$). Following the calibration phase, seven participants (**P1**, **P3**, **P4**, **P6**, **P8**, **P9**, **P11**) used their reported speech rate. Instead, five participants (**P2**, **P5**, **P7**, **P10**, **P12**) used a speech rate that was lower than what they initially reported (70%, 80%, 70%, 70%, 55%, respectively).

All participants travel known routes frequently but only two participant travel unfamiliar routes daily (**P2**, **P12**). Others traverse unknown routes only when needed (**P6**, **P10**, **P11**), and possibly avoid them (**P3**). Except for **P1**, **P3** and **P10**, participants interact with their mobile device through screen

Table 5.1: Participants' demographic information. For route frequency: Fam.=familiar, Unfam.=unfamiliar

				Expertise		**Screen reader**		**Route frequency**	
ID	**Age**	**Sex**	**Onset**	**music**	**mobile**	**speed**	**voice**	**Fam.**	**Unfam.**
P1	21	M	birth	3	2	60	F	daily	rarely
P2	40	F	birth	1	4	80	F	daily	daily
P3	59	M	18	1	2	66	F	daily	rarely
P4	64	M	<18	3	4	55	F	daily	rarely
P5	23	M	birth	3	5	90	F	daily	monthly
P6	45	M	birth	3	5	67	F	daily	rarely
P7	40	F	birth	5	4	75	F	every 48h	rarely
P8	35	M	birth	5	4	60	F	daily	weekly
P9	60	M	>18	4	4	60	F	daily	rarely
P10	54	M	birth	4	4	80	M	daily	monthly
P11	61	M	birth	1	5	55	F	daily	rarely
P12	62	F	birth	3	2	60	F	daily	daily

reader during mobility on daily basis. However, some (**P1**, **P3**, **P9**, **P10**) avoid listening the screen reader while they need to focus on the environmental sound (*e.g.*, while walking), unless the screen reader feedback is highly relevant (*e.g.*, navigation instructions in an unknown environment).

We also asked the participants how they usually listen to the screen reader while not moving (*e.g.*, home, office) and while in mobility (*e.g.*, walking, waiting for the train at the station). In general, participants use different audio output devices in different situations. While not in mobility, all participants except for **P5** and **P11** listen to the screen reader through the mobile device speaker but some (**P5**, **P7**, **P9**) alternate this with the use of earphones, headphones, and mono earphone. One participant reported using a Bluetooth speaker (**P11**). During mobility, the preferred listening hardware also changes depending on the situation. The mobile device speaker is used by all participants except **P2** and **P8**, but some (**P5**, **P6**) report that it is sometimes necessary to approach the speaker to the ear in noisy environments. **P9** prefers the mobile device speaker as it allows him to listen to the environmental sound. Many participants (**P2**, **P5**, **P7**, **P8**, **P11**) reported using single earphone (either wired or wireless) as it allows them to listen to environmental sound. Many participants (**P3**, **P6**, **P7**, **P9**, **P11**) reported using stereo earphones when they do not need to focus on the environmental sound (*e.g.*, while on a train). One participant (**P9**) uses *Apple AirPods* as they allow to listen to the environmental sound (we assume the participant uses the "transparency mode" available for this device), and one participant (**P11**) reported using bone conduction headphones.

Half of the participants (**P1**, **P5**, **P6**, **P7**, **P10**, **P12**) cope with noisy environments by approaching the mobile device to their ear to listen to the screen reader. Also, half of the participants (**P2**, **P3**, **P4**, **P6**, **P8**, **P10**) reported that they increase the screen reader volume, and one participant (**P9**) adjusts the volume to balance between the volume of the screen reader and the volume of the environmental sound. Three participants (**P4**, **P7**, **P5**) use stereo earphones when they do not need to concentrate on the environmental sound. Three participants report slowing the *VoiceOver* speech rate in noisy environments (**P3**, **P4**, **P7**). In particular **P3** and **P4** use a *VoiceOver* option to read the text word-by-word or character-by-character. Other coping mechanisms are to find a quieter place to listen to the screen reader (**P10**) or to listen during quieter moments (**P11**), for example during subway train stops.

Finally, we asked which assistive technologies participants use in mobility. Most reported using general-purpose navigation tools (*e.g.*, *Google maps*) combined with assistive technologies for BLV people. Only **P9** did not use any navigation application. Popular applications are those for checking public transport timetables and for providing accessible directions, like *BlindSquare*[3], *Lazarillo*[4], *NearBy Explorer*[5], *Ariadne GPS*[6], *ViaOpta Nav*[7], *TomTom navigator*[8], *Microsoft Soundscape*[9], and *SeeingAI*[10].

5.5 Experimental Results

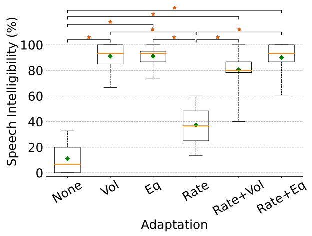
5.5.1 Quantitative Analysis

Speech Intelligibility

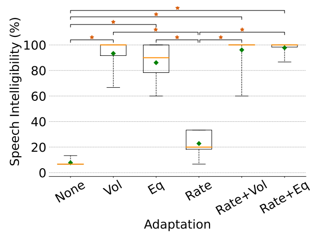
As expected, in the control condition (*i.e.*, **suburban** soundscape) the participants were able to recognize nearly all the words ($M = 99.44\%$, $SD = 1.84\%$). Instead, in presence of noise (see Fig. 5.2), without any compensation, the percentage of correctly recognized words drops to $11.11 \pm 12.57\%$, $7.78 \pm 2.48\%$ and $7.22 \pm 9.98\%$, for **Crowd**, **Traffic** and **Subway** soundscapes, respectively. This result confirms that, without any compensation, synthetic speech intelligibility is severely impaired in presence of typical environment noise.

The proposed compensation techniques increase the speech intelligibility average score in all soundscapes. The only exception is the **Rate** compensation in the **Subway** soundscape, for which the intelligibility is lower than without compensation. Specifically, for all the soundscapes, the effect of the sonification technique was found to be significant (**Crowd** $H_{(5)} = 47.17$, $p < .001$; **Traffic** $H_{(5)} = 52.98$, $p < .001$; **Subway** $H_{(5)} = 50.49$, $p < .001$). Pairwise comparisons reveal that **Vol**, **Eq**, **Rate+Vol** and **Rate+Eq** significantly improve the speech intelligibility with respect to both **None** and **Rate**, considering all soundscapes. In particular, **Rate+Eq** achieves the highest intelligibility score for the **Traffic** ($97.78 \pm 4.16\%$) and **Subway** ($96.11 \pm 4.27\%$) soundscapes. In-

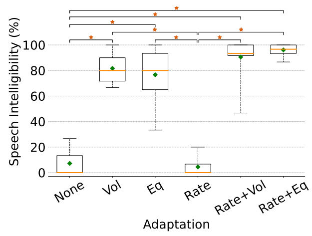
[3]https://blindsquare.com
[4]https://lazarillo.app
[5]https://tech.aph.org/ne_info.htm
[6]https://www.ariadnegps.eu/en/
[7]https://www.itcares.it/portfolio/viaoptanav/
[8]https://www.tomtom.com/it_it/navigation/mobile-apps/go-navigation-app
[9]https://www.microsoft.com/en-us/research/product/soundscape/
[10]https://www.seeingai.com/



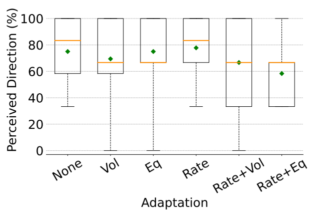
(a) Speech Intelligibility in **crowd**



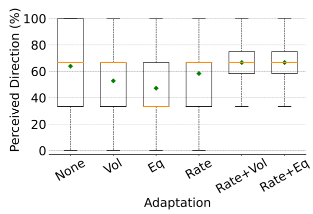
(b) Speech Intelligibility in **traffic**



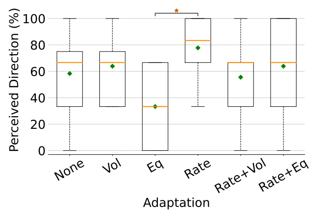
(c) Speech Intelligibility in **subway**



(d) Perceived Direction in **crowd**



(e) Perceived Direction in **traffic**



(f) Perceived Direction in **subway**

Figure 5.2: Speech Intelligibility and Perceived Direction results (◆ mean; ⋆ significant).

stead, in the **Crowd** soundscape, **Vol** (91.11±10.30%) and **Eq** (91.11±7.37%) show the highest improvement over the uncompensated speech.

Perceived Direction

In the **Suburban** scenario participants can recognize the correct direction of the contextual sound in almost all cases ($M = 92\%$, $SD = 14\%$). In the other soundscapes, recognizing the correct direction of the contextual sound is much harder. Indeed, the perceived direction score falls to $75 \pm 28\%$ in the **Crowd** soundscape, $64 \pm 32\%$ for **Traffic** and $58 \pm 31\%$ for the **Subway** scenario (see Fig. 5.2f). Contrary to our expectations, the screen reader speech compensations do not seem to worsen the perceived direction score over the uncompensated condition. Indeed, no significant differences emerge among compensations. The only significant group difference ($H_{(5)} = 15.27$, $p < .001$) is detected in the **Subway** soundscape, where pairwise comparisons reveal that **Rate** is significantly better than **Eq** compensation.

5.5.2 Answers to the Final Open-Ended Questions

We summarize the participants' answers to the final questions, which are reported in Table 5.2. All participants agreed that the soundscapes realistically reproduce real-world situations in which it could be hard to listen to the screen reader (Q1). Only **P5** stated that he would not have problems in noisy scenarios, as he would set the volume to the maximum. However, he also admitted that he would have problems in the **Subway**, scenario, as it is particularly noisy.

Table 5.2: Final open-ended questions

Q1	Do the soundscapes realistically reproduce situations in which it can be hard to listen to the screen reader?
Q2	In which soundscape was it harder to understand the screen reader? Which one was easiest?
Q3	Were contextual sounds [the term was introduced to the participant before] realistic? Which contextual sound was harder to perceive?
Q4	Some sentences were read aloud with a slower rate. Did you perceive this?
Q5	Some sentences were read aloud with a higher volume. Did you perceive this?
Q6	Some sentences were read aloud with a higher volume on some frequencies only. Did you perceive this?
Q7	Some sentences were read aloud with a combination of slower rate and higher volume. Did you perceive this?
Q8	What do you think of the bone-conduction headphones that you used during the experiments?

Besides **P3**, who had more difficulties in distinguishing sounds in the **Traffic** scenario, all other participants considered **Subway** as the soundscape in which it was harder to understand the screen reader (Q2). They noted that this is the noisiest soundscape, it masks the speech output and hence it is more cognitively demanding. Nine participants (**P1**, **P2**, **P3**, **P5**, **P6**, **P7**, **P8**, **P10**, **P12**) considered **Suburban** as the simplest soundscape, as it is less noisy and the volume of the screen reader is higher and separable from the soundscape. Instead, two participants (**P4**, **P11**) considered **Crowd** as the simplest soundscape. In particular, **P4** noted that sounds and voices are more distinct. Finally, **P9** reported that **Traffic** is the simplest soundscape because environmental noise and speech were clearly distinguished in this soundscape.

Considering the contextual sounds (Q3), six participants (**P3**, **P5**, **P6**, **P8**, **P9**, **P10**) regarded them as realistic. However, **P2** considered the **Subway** contextual sound too short and **P11** believed it was not loud enough. **P7** reported that the **Suburban** contextual sound was hard to hear and **P12** stated that it was different than expected. **P2** perceived the sound of cash receipt in **Crowd** as too loud, while **P7** and **P12** considered it unrealistic. Finally, the contextual sound in **Traffic** was perceived to be different than expected by three participants (**P3**, **P5**, **P7**) and hard to separate from the soundscape (**P11**).

All participants were able to perceive the **Vol** compensation and noted that it improves intelligibility (Q5). Eight participants (**P1**, **P3**, **P4**, **P7**, **P9**, **P10**, **P11**, **P12**) were able to perceive the presence of the **Rate** compensation (Q4), and all of them regarded it as useful. Only five participants (**P3**, **P4**, **P7**, **P9**, **P11**) perceived **Eq** (Q6). Since **Vol** and **Eq** are similar, we suspect that most participants were not able to distinguish between the two. Similarly, only five participants (**P4**, **P5**, **P7**, **P9**, **P10**) perceived the combined compensations (Q7).

Considering the use of the bone conduction headphones (Q8), 9 participants (**P3**, **P4**, **P6**, **P7**, **P8**, **P9**, **P10**, **P11**, **P12**) agreed that this is an acceptable solution, mainly because they do not prevent the hearing of environmental sound. However, some participants (**P1**, **P3**) reported that they are uncomfortable and heavy. In particular, **P2** noted that the sound is "external" and hence hard to understand.

5.5.3 Participants' Comments

Participants also provided spontaneous comments. We report the key topics that emerged from their analysis.

Approach Usefulness

Participants found the proposed approach useful for screen reader usage in noisy environments (**P1**, **P4**, **P7**, **P8**). In particular, **P7**, who already used manual adjustment of volume and speech rate, requested for it to be implemented in the same way as the automated adaptation of screen illumination to the ambient light on mobile devices[11]:

> *"It would be great if they [compensations] were made similar to the automatic screen illumination."*

Combined Compensations

Participants considered the combined compensations useful (**P5**, **P7**, **P9**, **P12**), confirming the findings from the preliminary studies. **P12** noted that **Rate** alone does not improve intelligibility, but it can be useful when paired with **Vol**, which is consistent with quantitative results. Similarly, **P7** noted that raising the volume and lowering the speech rate helps in noisy contexts. In particular, **P5** noted:

> *"I sometimes wondered how I was able to understand sentences that I did not previously understand. These [combined compensations] make life easier."*

Trade-off between intelligibility and distraction

Participants also commented on the inherent trade-off between the ability to correctly understand the screen reader speech and to pay attention to the environment (**P1**, **P2**, **P3**, **P4**, **P6**, **P9**, **P10**, **P11**). While most of the comments related to the soundscape and speech volume, some referred to the speech rate (**P10**, **P11**). **P11** in particular felt that slowing the speech rate lowers the volume of some words, while **P10** observed:

[11] All quotes have been translated from Italian.

"It is not always useful because slowing too much dilates the time needed [to listen]."

Modifications to the Approach

Participants also suggested possible improvements. **P10** mentioned that some screen reader voices mask specific consonant sounds, which makes them harder to understand, and therefore we should consider also the screen reader voice used. **P5** also suggests another possible compensation, which is to delay the screen reader speech when strong background noise is detected:

"I would've liked to hear the message when the [subway] sound fades out. Instead, the message was played when the sound was on the rise."

5.6 Discussion

5.6.1 Main Results

The proposed compensation techniques were found to be effective in improving synthetic speech intelligibility across different noisy soundscapes. Specifically, four of the five proposed compensation techniques significantly improved the speech intelligibility in all three noisy soundscapes. The only compensation that did not significantly improve intelligibility is **Rate**. Indeed, some of the participants initially felt that reducing speech rate would not improve intelligibility in presence of noise. For example, one participant in the instruction phase of the preliminary evaluation argued:

"I don't think that diminishing the speed would impact the speech understanding."

However, after the listening tasks, most participants confirmed that combining rate with the other compensations was useful for louder soundscapes. For example, the same participant retracted the prior opinion:

"Both raising volume and reducing speed are useful. In general, I think lower speed is more important"

In agreement with this opinion, results show that in two scenarios (**Traffic** and **Subway**), the combined compensations improve on average over the other compensations. While this difference is not significant, possibly due to the limited sample size, it confirms the intuition of the participants to the preliminary study that combining the compensation techniques can further improve speech intelligibility in noisy scenarios.

We initially hypothesized that compensations would distract the participants from the environment soundscape. However, no significant difference was observed in the **Perceived Direction** score across different conditions. Such a result can be explained by a comment provided during preliminary experiments: when the speech is hard to understand (either too quiet or too fast) it requires additional concentration, hence distracting from the environmental sound. This means that in some cases compensations might actually *reduce* distraction.

5.6.2 Experimental Design Limitations

In order to ensure both participants' safety and experimental repeatability, we conducted the evaluation in a controlled environment. This has some implications, including the fact that the soundscapes had to be simulated. To ensure highly realistic simulations, in our experimental settings we used real-world recordings of the target soundscape scenarios, replayed in a silent chamber with a quadraphonic speaker setup and settings replicating real-world listening conditions. Indeed, all study participants agreed that the experimental setting was realistic.

Another consequence is that the participants were in a safe environment, rather than in the real-world, so they did not actually need to focus on the soundscape to prevent dangers. This may have impacted their level of attention with respect to the background sounds. Additionally, distraction was measured as the ability to pinpoint the direction of a single sound, while in general continuous attention should be devoted to the ambient soundscape. To account for this, the contextual sound was played at a random time, and therefore the participants were stimulated to pay attention to the soundscape continuously.

To ensure uniformity among the tests we always used the same three soundscapes, reproducing the speech signal always at the same time. While this guarantees that all sentences are read with consistent background noise, participants reported that they got used to the soundscape and were less distracted by it after a while. The experimental design could be improved by using longer soundscapes and by changing the starting point of the soundscape playback across repetitions. However, this would require collecting a larger set of datapoints, to compensate for the differences in the listening conditions.

Another consequence of placing sentences in the middle of the soundscapes is that, after the sentence reading ends, there are still a few seconds before the soundscape ends. The participants would wait until the end of the soundscape before repeating what they heard. This required an additional memorization effort and in some cases resulted in participants forgetting what they heard. For this reason, the participants who reported this issue were asked to repeat the sentence immediately, without waiting for the soundscape to end.

To measure speech intelligibility without an influence of an external semantic context we used a random sentence generation approach [158]. One limitation of this approach is that, while the exact words are unpredictable, the phrase structure is always the same (*e.g.*, the first word is always a name). Thus the participants could try to infer a word even if it is not clearly heard. However, this is not unrealistic: in many real-world applications, the screen reader often provides structurally similar messages with a small set of possible words (*e.g.*, "Turn right/left" in a navigation system).

Finally, our experiments involved solely Italian-speaking participants. While we expect the results to generalize to other languages, such an assumption would need to be evaluated with participants speaking different languages. Similarly, while the technique should be applicable to different listening hardware such as the speaker or in-ear headphones, we conducted the experiments solely with bone-conduction headphones. Thus, additional experiments would be needed to verify that the approach generalizes to other listening hardware.

5.6.3 Technique Limitations

In designing the **Rate** compensation technique, we decided to apply a flat speed reduction. A different solution could consist in reducing the speech rate proportionally with the noise level. However, for this, we would need to study the correlation between speech intelligibility at different speech rates and environmental noise. This would require a process of parameter tuning that is out of the scope of this study.

Correlating the environmental noise level with the amount of compensation is also a problem for **Vol** and **Eq**. However, for these compensations, it was possible to tune the settings based on prior works studying the impact of the difference between speech and noise volume on speech intelligibility [141, 196]. Still, fine-tuning these parameters to the specific application domain could help to achieve even better results.

While the speech signals were pre-recorded for convenience, the proposed compensations can be computed in real-time. Indeed, assuming that the compensation techniques are implemented as a part of the speech synthesizer, they would require monitoring the environmental noise and changing the speech generation parameters. In order to rapidly adapt to the environmental noise, the proposed technique employs temporal windows of 46 ms. Since changing the speech generation parameters does not produce any additional delay, the overall delay remains within 50 ms, which can be considered real-time. In order to compute the compensations in real-time, the device microphone could be used to collect samples of the environmental noise. If the microphone is expected to be covered (*e.g.*, if the device is in the user's pocket), an external microphone could be used (*e.g.*, the headset microphone).

Chapter 6

Conclusion

This thesis proposes three main contributions to address three problems: guiding BLV individuals during turns while navigating a route, improving screen reader intelligibility in noisy environments, and designing and evaluating a navigation system to enhance the overall navigation experience. As the first contribution, the thesis proposes two sonification techniques designed to augment navigation instructions in navigation assistance systems with the goal of providing more accurate guidance for BLV people. Experimental results show that both techniques outperform a baseline solution adopted in existing systems (*i.e.*, **P**) during turn instructions, and that one of them (*i.e.*, **IS+P**) also outperforms the baseline on straight path guidance. This result is particularly relevant because our solution can be used to easily improve the guidance accuracy of existing systems [9] that adopt the baseline solution.

As second contribution, the thesis proposes a new approach for defining the system architecture for turn-by-turn navigation systems, based on a separation of concerns design that decomposes the overall navigation system into a set of four layers. This thesis also addresses the design of two of the layers: *Routing* and *Guidance*, by adopting a design principle aimed at reducing the number of navigation instructions issued to the BLV traveler. Experimental results with eight BLV participants show that the developed system outperforms a baseline solution adopted in existing navigation systems, providing fewer and clearer instructions, and causing the user to spend less time in potentially unsafe areas. These results are significant for two reasons. Firstly, they can be used to improve the navigation experience of existing systems [3, 167]. Secondly, this thesis's investigation paves the way for a new stream of research aimed at designing and engineering solutions for the *Routing* and *Guidance* layers that, as we show in this thesis, have a significant impact on the navigation experience.

As third contribution, this thesis proposes audio compensation techniques to improve intelligibility of screen reader speech when background noise is present. The thesis provides experimental evidence that the proposed techniques can effectively improve screen reader intelligibility in noisy environments, without distracting the user from the soundscape. The proposed compensations are a practical solution that can be easily implemented in existing mobile screen readers.

All these results contribute to enhancing the navigation experience for BLV people, making it more accessible. Through the implementation of innovative sonification techniques, a new design for a navigation system architecture, and compensation methods for screen reader intelligibility in the presence of noise, this thesis not only addresses key challenges, but also sets the stage for future advancements in assistive technology. By demonstrating the effectiveness of these solutions, the research underscores the potential for significant improvements in mobility and independence for BLV people. The implications of this work extend beyond academic research offering practical insights for developers of assistive technologies, potentially influencing industry design choices. Ultimately, this thesis advocates for a more inclusive approach to technology design, emphasizing the importance of developing systems that accommodate the unique needs of BLV people, thereby fostering greater integration into society and everyday life.

6.1 Future work on sonification of navigation instructions

As future work, as suggested by some participants, we want to experiment with how users personalize the sonification depending on the context, for example, when traversing known or unknown routes. We also intend to compare the sonifications in different environments, both indoor (*e.g.*, museums or airports) and outdoor (*e.g.*, roads with loud traffic noise), possibly using even more accurate positioning techniques (for example using the LIDAR sensor). In such different environments, the user would need to pay attention to the background soundscape, while surrounded by various environmental noises. Thus, one of the issues may be that the instructions and the sonifications are difficult to hear and follow. To address this issue, we will explore the dynamic adaptation of the instructions and sonifications, for example by automatically adapting their volume or other sound properties based on the soundscape loudness. Another issue is related to how the users carry the device: using the chest mount is impractical in many contexts and future experiments should evaluate how other solutions (*e.g.*, holding the device in hand) impact the navigation.

The experiments reported in this thesis were conducted with a relatively small set of participants, with similar needs (*i.e.*, all participants were blind and had similar characteristics). We intend to evaluate whether the results reported in this thesis generalize to users with low vision, who are able to rely on residual sight to integrate the auditory instructions with the visual ones. Finally, we intend to systematically study the navigation strategy. For example, as suggested by some participants, it may be preferable to provide sonifications also to quantify the lateral shift instructions. Also, we will explore the effect of different tuning parameters on the navigation outcome, including, for example, the *orientation offset* threshold above which the user is instructed to rotate.

6.2 Future work on the navigation problem

In future work, we plan to evaluate the navigation system in multiple environments, both indoor (*e.g.*, museums or airports) and outdoor (*e.g.*, roads with loud traffic noise), possibly using more accurate positioning techniques (*e.g.*, the LIDAR sensor). We are also interested in assessing the impact of adding obstacle detection and avoidance features on the navigation experience and in determining how to integrate these features into the existing *Routing* and *Guidance* modules.

The experiments reported in this thesis were conducted with a relatively small set of participants, with similar needs (*e.g.*, all participants were totally blind). We intend to evaluate whether the results reported in this thesis generalize to users with low vision, who are able to rely on residual sight to integrate auditory instructions with visual instructions.

6.3 Future work on Audio adaptation feedback in presence of noise

This work paves the way for various research directions. First, it is possible to investigate other compensation techniques, by altering different speech properties (*e.g.*, pitch). Second, we intend to explore how different environmental noise characteristics affect the compensation techniques, analyzing correlations between environmental noise characteristics and speech intelligibility at different levels of compensation (*e.g.*, speech rates). Additional factors should also be taken into account, including the language and the listening hardware (headphones, speakers). A third research direction is to investigate compensation techniques for sonification instructions, which have been proposed for navigation assistance [139].

Bibliography

[1] Ali Abdolrahmani, Ravi Kuber, and Amy Hurst. An empirical investigation of the situationally-induced impairments experienced by blind mobile device users. In *Proceedings of the 13th International Web for All Conference*, pages 1–8, 2016.

[2] Iyad Abu Doush, Sawsan Alshatnawi, Abdel-Karim Al-Tamimi, Bushra Alhasan, and Safaa Hamasha. Isab: integrated indoor navigation system for the blind. *Interacting with Computers*, 29(2):181–202, 2017.

[3] Dragan Ahmetovic, Federico Avanzini, Adriano Baratè, Cristian Bernareggi, Marco Ciardullo, Gabriele Galimberti, Luca A Ludovico, Sergio Mascetti, and Giorgio Presti. Sonification of navigation instructions for people with visual impairment. *International Journal of Human-Computer Studies*, 177:103057, 2023.

[4] Dragan Ahmetovic, Federico Avanzini, Adriano Baratè, Cristian Bernareggi, Gabriele Galimberti, Luca A Ludovico, Sergio Mascetti, and Giorgio Presti. Sonification of pathways for people with visual impairments. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 379–381, 2018.

[5] Dragan Ahmetovic, Federico Avanzini, Adriano Baratè, Cristian Bernareggi, Gabriele Galimberti, Luca A Ludovico, Sergio Mascetti, and Giorgio Presti. Sonification of rotation instructions to support navigation of people with visual impairment. In *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom*, pages 1–10. IEEE, 2019.

[6] Dragan Ahmetovic, Cristian Bernareggi, Andrea Gerino, and Sergio Mascetti. Zebrarecognizer: Efficient and precise localization of pedestrian crossings. In *2014 22nd International Conference on Pattern Recognition*, pages 2566–2571. IEEE, 2014.

[7] Dragan Ahmetovic, Cristian Bernareggi, Kristian Keller, and Sergio Mascetti. Musa: artwork accessibility through augmented reality for people with low vision. In *Proceedings of the 18th International Web for All Conference*, pages 1–9, 2021.

[8] Dragan Ahmetovic, Gabriele Galimberti, Federico Avanzini, Cristian Bernareggi, Luca Andrea Ludovico, Giorgio Presti, Gianluca Vasco, and

Sergio Mascetti. Enhancing screen reader intelligibility in noisy environments. *IEEE Transactions on Human-Machine Systems*, 53(4):771–780, 2023.

[9] Dragan Ahmetovic, Cole Gleason, Chengxiong Ruan, Kris Kitani, Hironobu Takagi, and Chieko Asakawa. Navcog: a navigational cognitive assistant for the blind. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 90–99, 2016.

[10] Dragan Ahmetovic, João Guerreiro, Eshed Ohn-Bar, Kris M Kitani, and Chieko Asakawa. Impact of expertise on interaction preferences for navigation assistance of visually impaired individuals. In *Proceedings of the 16th International Web for All Conference*, pages 1–9, 2019.

[11] Dragan Ahmetovic, Roberto Manduchi, James M Coughlan, and Sergio Mascetti. Mind your crossings: Mining gis imagery for crosswalk localization. *ACM Transactions on Accessible Computing (TACCESS)*, 9(4):1–25, 2017.

[12] Dragan Ahmetovic, Sergio Mascetti, Cristian Bernareggi, João Guerreiro, Uran Oh, and Chieko Asakawa. Deep learning compensation of rotation errors during navigation assistance for people with visual impairments or blindness. *ACM Transactions on Accessible Computing (TACCESS)*, 12(4):1–19, 2019.

[13] Dragan Ahmetovic, Uran Oh, Sergio Mascetti, and Chieko Asakawa. Turn right: Analysis of rotation errors in turn-by-turn navigation for individuals with visual impairments. In *International Conference on Computers and Accessibility*, 2018.

[14] Dragan Ahmetovic, Daisuke Sato, Uran Oh, Tatsuya Ishihara, Kris Kitani, and Chieko Asakawa. Recog: Supporting blind people in recognizing personal objects. In *Conference on Human Factors in Computing Systems*. ACM, 2020.

[15] Shurug Al-Khalifa and Muna Al-Razgan. Ebsar: Indoor guidance for the visually impaired. *Computers & Electrical Engineering*, 54:26–39, 2016.

[16] Dennis Allen. Orientation & mobility for persons with low vision. *Journal of Visual Impairment & Blindness*, 71(1):13–15, 1977.

[17] VD Ambeth Kumar, S Malathi, Abhishek Kumar, Kalyana C Veluvolu, et al. Active volume control in smart phones based on user activity and ambient noise. *Sensors*, 20(15):4117, 2020.

[18] Aries Arditi and YingLi Tian. User interface preferences in the design of a camera-based navigation and wayfinding aid. *Journal of Visual Impairment & Blindness*, 107(2):118–129, 2013.

[19] Shiri Azenkot, Richard E Ladner, and Jacob O Wobbrock. Smartphone haptic feedback for nonvisual wayfinding. In *international conference on Computers and accessibility*, 2011.

[20] Shiri Azenkot and Nicole B Lee. Exploring the use of speech input by blind people on mobile devices. In *International conference on computers and accessibility*. ACM, 2013.

[21] Shiri Azenkot, Sanjana Prasain, Alan Borning, Emily Fortuna, Richard E Ladner, and Jacob O Wobbrock. Enhancing independence and safety for blind and deaf-blind public transit riders. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 3247–3256, 2011.

[22] Maryam Bandukda, Catherine Holloway, Aneesha Singh, Giulia Barbareschi, and Nadia Berthouze. Opportunities for supporting self-efficacy through orientation & mobility training technologies for blind and partially sighted people. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–13, 2021.

[23] John B Bass III. White supremacy and the major scale: Reframing narratives to teach the blues, improvise, and check in with students. *Research & Issues in Music Education*, 17(1):2, 2017.

[24] J MALVERIN Benjamin. The laser cane. *Bulletin of prosthetics research*, 11(2):443–450, 1974.

[25] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 1995.

[26] Shripad Bhatlawande, Manjunatha Mahadevappa, and Jayanta Mukhopadhyay. Way-finding electronic bracelet for visually impaired people. In *2013 IEEE Point-of-Care Healthcare Technologies (PHT)*, pages 260–263. IEEE, 2013.

[27] James Biagioni, Tomas Gerlich, Timothy Merrifield, and Jakob Eriksson. Easytracker: automatic transit tracking, mapping, and arrival time prediction using smartphones. In *proceedings of the 9th ACM conference on embedded networked sensor systems*, pages 68–81, 2011.

[28] Jeffrey P Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, et al. Vizwiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 333–342, 2010.

[29] Werner Bischof, Elmar Krajnc, Markus Dornhofer, and Michael Ulm. Navcom–wlan communication between public transport vehicles and

smart phones to support visually impaired and blind people. In *International Conference on Computers for Handicapped Persons*, pages 91–98. Springer, 2012.

[30] Meera M Blattner, Denise A Sumikawa, and Robert M Greenberg. Earcons and icons: Their structure and common design principles. *Human–Computer Interaction*, 4(1):11–44, 1989.

[31] Mounir Bousbia-Salah, Maamar Bettayeb, and Allal Larbi. A navigation aid for blind people. *Journal of Intelligent & Robotic Systems*, 64:387–400, 2011.

[32] Erin L Brady, Daisuke Sato, Chengxiong Ruan, Hironobu Takagi, and Chieko Asakawa. Exploring interface design for independent navigation by people with visual impairments. In *International Conference on Computers and Accessibility*, 2015.

[33] Danielle Bragg, Cynthia Bennett, Katharina Reinecke, and Richard Ladner. A large inclusive study of human listening rates. In *Conference on Human Factors in Computing Systems*. ACM, 2018.

[34] Robin N Brewer and Vaishnav Kameswaran. Understanding the power of control in autonomous vehicles for people with vision impairment. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 185–197, 2018.

[35] Michael Brock and Per Ola Kristensson. Supporting blind navigation using depth sensing and sonification. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 255–258, 2013.

[36] Marc Brysbaert. How many participants do we have to include in properly powered experiments? a tutorial of power analysis with reference tables. *Journal of cognition*, 2019.

[37] Michal Bujacz, Karol Kropidlowski, Gabriel Ivanica, Alin Moldoveanu, Charalampos Saitis, Adam Csapo, György Wersenyi, Simone Spagnol, Omar I Johannesson, Runar Unnthorsson, et al. Sound of vision-spatial audio output and sonification approaches. In *International Conference on Computers Helping People with Special Needs*, 2016.

[38] Michal Bujacz, Piotr Skulimowski, and Pawel Strumillo. Naviton – a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired. *Journal of the Audio Engineering Society*, 2012.

[39] Hsuan-Eng Chen, Yi-Ying Lin, Chien-Hsing Chen, and I-Fang Wang. Blindnavi: A navigation app for the visually impaired smartphone user. In *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems*, pages 19–24, 2015.

[40] Ruiqi Cheng, Kaiwei Wang, Longqing Lin, and Kailun Yang. Visual localization of key positions for visually impaired people. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 2893–2898. IEEE, 2018.

[41] Seyed Ali Cheraghi, Anup Sharma, Vinod Namboodiri, and Güler Arsal. Safeexit4aii: an inclusive indoor emergency evacuation system for people with disabilities. In *Proceedings of the 16th International Web for All Conference*, pages 1–10, 2019.

[42] Sakmongkon Chumkamon, Peranitti Tuvaphanthaphiphat, and Phongsak Keeratiwintakorn. A blind navigation system using rfid for indoor environments. In *2008 5th international conference on electrical engineering/electronics, computer, telecommunications and information technology*, volume 2, pages 765–768. IEEE, 2008.

[43] Daeun Joyce Chung, Muya Guoji, Nina Mindel, Alexis Malkin, Fernando Alberotrio, Shane Lowe, Chris McNally, Casandra Xavier, and Paul Ruvolo. Large-scale, longitudinal, hybrid participatory design program to create navigation technology for the blind. *arXiv preprint arXiv:2410.00192*, 2024.

[44] Achituv Cohen and Sagi Dalyot. Route planning for blind pedestrians using openstreetmap. *Environment and Planning B: Urban Analytics and City Science*, 48(6):1511–1526, 2021.

[45] Angela Constantinescu, Karin Müller, Monica Haurilet, Vanessa Petrausch, and Rainer Stiefelhagen. Bring the environment to life: A sonification module for people with visual impairments to improve situation awareness. In *Proceedings of the 2020 International Conference on Multimodal Interaction*, pages 50–59, 2020.

[46] Allan D. Coop. Sonification, musification, and synthesis of absolute program music. In *Proceedings of the International Conference on Auditory Display*, 2016.

[47] James Coughlan and Roberto Manduchi. Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users. *International Journal on Artificial Intelligence Tools*, 2009.

[48] James M Coughlan and Joshua Miele. Ar4vi: Ar as an accessibility tool for people with visual impairments. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pages 288–292. IEEE, 2017.

[49] Ryan Crabb, Seyed Ali Cheraghi, and James M Coughlan. A lightweight approach to localization for blind and visually impaired travelers. *Sensors*, 23(5):2701, 2023.

[50] Rafael Jeferson Pezzuto Damaceno, Juliana Cristina Braga, and Jesús Pascual Mena-Chalco. Mobile device accessibility for the visually impaired: problems mapping and recommendations. *Universal Access in the Information Society*, 2018.

[51] Uddipan Das and Brian Hong. Findmyway: toward developing an accessible wayfinding application for visually impaired in an indoor space. In *International Conference on Computers Helping People with Special Needs*, pages 428–435. Springer, 2024.

[52] Tilman Dingler, Jeffrey Lindsay, Bruce N Walker, et al. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the 14th International Conference on Auditory Display, Paris, France*, pages 1–6, 2008.

[53] Gaël Dubus and Roberto Bresin. A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE*, 2013.

[54] Olive Jean Dunn. Multiple comparisons among means. *Journal of the American statistical association*, 1961.

[55] Hugh Durrant-Whyte and Tim Bailey. Simultaneous localization and mapping: part i. *Robotics & automation magazine*, 2006.

[56] Fatma El-Zahraa El-Taher, Ayman Taha, Jane Courtney, and Susan Mckeever. A systematic review of urban navigation systems for visually impaired people. *Sensors*, 21(9):3103, 2021.

[57] Wafa Elmannai and Khaled Elleithy. Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions. *Sensors*, 17(3):565, 2017.

[58] Fatmaelzahraa Eltaher, Luis Miralles-Pechuán, Jane Courtney, and Susan Mckeever. Saferoute: A safer outdoor navigation algorithm with smart routing for people with visual impairment. In *Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments*, pages 1–8, 2024.

[59] Robert Wall Emerson. Orientation and mobility. In *Movement and Visual Impairment*, pages 32–46. Routledge, 2020.

[60] René Farcy, Roger Leroux, Alain Jucha, Roland Damaschini, Colette Grégoire, and Aziz Zogaghi. Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view. In *Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Technology for Inclusion*, volume 12, page 12. Citeseer, 2006.

[61] Heather A Faucett, Kate E Ringland, Amanda LL Cullen, and Gillian R Hayes. (in) visibility in disability and assistive technology. *Transactions on Accessible Computing*, 2017.

[62] Alexander Fiannaca, Ilias Apostolopoulous, and Eelke Folmer. Headlock: a wearable navigation aid that helps blind cane users traverse large open spaces. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, pages 19–26, 2014.

[63] Harvey Fletcher and Wilden A Munson. Loudness, its definition, measurement and calculation. *Bell System Technical Journal*, 1933.

[64] Janet F Fletcher. Spatial representation in blind children. 1: Development compared to sighted children. *Journal of Visual Impairment & Blindness*, 74(10):381–385, 1980.

[65] German Flores, Sri Kurniawan, Roberto Manduchi, Eric Martinson, Lourdes M Morales, and Emrah Akin Sisbot. Vibrotactile guidance for wayfinding of blind walkers. *IEEE transactions on haptics*, 8(3):306–317, 2015.

[66] German Flores and Roberto Manduchi. Easy return: an app for indoor backtracking assistance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2018.

[67] German H Flores and Roberto Manduchi. A public transit assistant for blind bus passengers. *IEEE Pervasive Computing*, 17(1):49–59, 2018.

[68] Milton Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 1937.

[69] Giovanni Fusco and James M Coughlan. Indoor localization using computer vision and visual-inertial odometry. In *Computers Helping People with Special Needs: 16th International Conference, ICCHP 2018, Linz, Austria, July 11-13, 2018, Proceedings, Part II 16*, pages 86–93. Springer, 2018.

[70] Giovanni Fusco and James M Coughlan. Indoor localization for visually impaired travelers using computer vision on a smartphone. In *Proceedings of the 17th international web for all conference*, pages 1–11, 2020.

[71] Gabriele Galimberti. Auditory feedback to compensate audible instructions to support people with visual impairment. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–3, 2021.

[72] Giovanni Galioto, Ilenia Tinnirello, Daniele Croce, Federica Inderst, Federica Pascucci, and Laura Giarré. Sensor fusion localization and navigation for visually impaired people. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 471–473, 2017.

[73] Thomas Gallagher, Elyse Wise, Binghao Li, Andrew G Dempster, Chris Rizos, and Euan Ramsey-Stewart. Indoor positioning system based on sensor fusion for the blind and visually impaired. In *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–9. IEEE, 2012.

[74] Simon Gallo, Dominique Chapuis, Laura Santos-Carreras, Yeongmi Kim, P Retornaz, H Bleuler, and R Gassert. Augmented white cane with multimodal haptic feedback. In *2010 3rd IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics*, pages 149–155. IEEE, 2010.

[75] William W Gaver. The sonicfinder: An interface that uses auditory icons. *Human–Computer Interaction*, 4(1):67–94, 1989.

[76] Hans Geiger, Walther Müller, et al. Das elektronenzählrohr. *Physikalische Zeitschrift*, 1928.

[77] Michele Geronazzo, Alberto Bedin, Luca Brayda, Claudio Campus, and Federico Avanzini. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies*, 2016.

[78] 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.

[79] Cole Gleason, Dragan Ahmetovic, Saiph Savage, Carlos Toxtli, Carl Posthuma, Chieko Asakawa, Kris M Kitani, and Jeffrey P Bigham. Crowdsourcing the installation and maintenance of indoor localization infrastructure to support blind navigation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2(1):1–25, 2018.

[80] Reginald G Golledge, Jack M Loomis, Roberta L Klatzky, Andreas Flury, and Xiao Li Yang. Designing a personal guidance system to aid navigation without sight: Progress on the gis component. *International Journal of Geographical Information System*, 5(4):373–395, 1991.

[81] Nora Griffin-Shirley, Laura Bozeman, The Nguyen, Vitalis Othuon, Anita Page, Juhyun Hahm, Juyoung Hahm, and Jaehoon Lee. A survey of blind orientation and mobility specialists about the accommodations and teaching strategies they use when providing orientation and mobility services. *Journal of Visual Impairment & Blindness*, 115(3):190–203, 2021.

[82] Florian Grond and Jonathan Berger. Parameter mapping sonification. In *The sonification handbook*. Logos Verlag Berlin GmbH, Berlin, 2011.

[83] João Guerreiro, Dragan Ahmetovic, Kris M Kitani, and Chieko Asakawa. Virtual navigation for blind people: Building sequential representations of the real-world. In *International Conference on computers and accessibility*, 2017.

[84] João Guerreiro, Dragan Ahmetovic, Daisuke Sato, Kris Kitani, and Chieko Asakawa. Airport accessibility and navigation assistance for people with visual impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019.

[85] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M Kitani, and Chieko Asakawa. Cabot: Designing and evaluating an autonomous navigation robot for blind people. In *Proceedings of the 21st international ACM SIGACCESS conference on computers and accessibility*, pages 68–82, 2019.

[86] Danna Gurari, Qing Li, Abigale J Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P Bigham. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3608–3617, 2018.

[87] Fredrik Gustafsson, Fredrik Gunnarsson, Niclas Bergman, Urban Forssell, Jonas Jansson, Rickard Karlsson, and P-J Nordlund. Particle filters for positioning, navigation, and tracking. *IEEE Transactions on signal processing*, 50(2):425–437, 2002.

[88] David Guth and Robert LaDuke. The veering tendency of blind pedestrians: An analysis of the problem and literature review. *Journal of Visual Impairment and Blindness*, 1994.

[89] SA Haymes, DJ Guest, AD Heyes, and AW Johnston. The relationship of vision and psychological variables to the orientation and mobility of visually impaired persons. *Journal of Visual Impairment & Blindness*, 90(4):314–324, 1996.

[90] Thomas Hermann, Andy Hunt, John G Neuhoff, et al. *The sonification handbook*, volume 1. Logos Verlag Berlin, 2011.

[91] Sebastian Hilsenbeck, Dmytro Bobkov, Georg Schroth, Robert Huitl, and Eckehard Steinbach. Graph-based data fusion of pedometer and wifi measurements for mobile indoor positioning. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pages 147–158, 2014.

[92] Brian Hoyle and Dean Waters. Mobility at: The batcane (ultracane). In *Assistive technology for visually impaired and blind people*, pages 209–229. Springer, 2008.

[93] Guoquan Huang. Visual-inertial navigation: A concise review. In *International conference on robotics and automation*, 2019.

[94] Tatsuya Ishihara, Kris M Kitani, Chieko Asakawa, and Michitaka Hirose. Deep radio-visual localization. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 596–605. IEEE, 2018.

[95] Gaurav Jain, Basel Hindi, Zihao Zhang, Koushik Srinivasula, Mingyu Xie, Mahshid Ghasemi, Daniel Weiner, Sophie Ana Paris, Xin Yi Therese Xu, Michael Malcolm, et al. Streetnav: Leveraging street cameras to support precise outdoor navigation for blind pedestrians. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pages 1–21, 2024.

[96] Antonio R Jimenez, Fernando Seco, Carlos Prieto, and Jorge Guevara. A comparison of pedestrian dead-reckoning algorithms using a low-cost mems imu. In *2009 IEEE International Symposium on Intelligent Signal Processing*, pages 37–42. IEEE, 2009.

[97] Michael Johnson. *Pop Music Theory. Harmony, Form, and Composition.* Lulu Press, 2009.

[98] Hernisa Kacorri, Sergio Mascetti, Andrea Gerino, Dragan Ahmetovic, Valeria Alampi, Hironobu Takagi, and Chieko Asakawa. Insights on assistive orientation and mobility of people with visual impairment based on large-scale longitudinal data. *Transactions on Accessible Computing*, 2018.

[99] Hernisa Kacorri, Sergio Mascetti, Andrea Gerino, Dragan Ahmetovic, Hironobu Takagi, and Chieko Asakawa. Supporting orientation of people with visual impairment: Analysis of large scale usage data. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 151–159, 2016.

[100] Vaishnav Kameswaran, Alexander J. Fiannaca, Melanie Kneisel, Amy Karlson, Edward Cutrell, and Meredith Ringel Morris. Understanding in-situ use of commonly available navigation technologies by people with visual impairments. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–12, 2020.

[101] Slim Kammoun, Christophe Jouffrais, Tiago Guerreiro, Hugo Nicolau, and Joaquim Jorge. Guiding blind people with haptic feedback. *Frontiers in Accessibility for Pervasive Computing (Pervasive 2012)*, 3, 2012.

[102] Slim Kammoun, Marc J M Macé, Bernard Oriola, and Christophe Jouffrais. Towards a geographic information system facilitating navigation of visually impaired users. In *Computers Helping People with Special Needs: 13th International Conference, ICCHP 2012, Linz, Austria, July 11-13, 2012, Proceedings, Part II 13*, pages 521–528. Springer, 2012.

[103] Shaun K Kane, Jeffrey P Bigham, and Jacob O Wobbrock. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Conf. on comp. and access.* ACM, 2008.

[104] Shaun K Kane, Chandrika Jayant, Jacob O Wobbrock, and Richard E Ladner. Freedom to roam: a study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Conference on Computers and accessibility.* ACM, 2009.

[105] Brian F. G. Katz, Slim Kammoun, Gaëtan Parseihian, et al. Navig: augmented reality guidance system for the visually impaired. *Virtual Reality*, 2012.

[106] Brian FG Katz, Slim Kammoun, Gaëtan Parseihian, Olivier Gutierrez, Adrien Brilhault, Malika Auvray, Philippe Truillet, Michel Denis, Simon Thorpe, and Christophe Jouffrais. Navig: Augmented reality guidance system for the visually impaired: Combining object localization, gnss, and spatial audio. *Virtual Reality*, 16:253–269, 2012.

[107] Robert K Katzschmann, Brandon Araki, and Daniela Rus. Safe local navigation for visually impaired users with a time-of-flight and haptic feedback device. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(3):583–593, 2018.

[108] Maurice George Kendall. *Rank correlation methods.* Griffin, 1948.

[109] Shah Khusro, Babar Shah, Inayat Khan, and Sumayya Rahman. Haptic feedback to assist blind people in indoor environment using vibration patterns. *Sensors*, 22(1):361, 2022.

[110] Cherry Kim, Seongjin Wang, Dongjun Kim, Shijia Pan, and Jun Han. Impactile: Tile-level localization for the visually impaired using structural uniqueness. In *Proceedings of the 25th International Workshop on Mobile Computing Systems and Applications*, pages 86–92, 2024.

[111] Jee-Eun Kim, Masahiro Bessho, Shinsuke Kobayashi, Noboru Koshizuka, and Ken Sakamura. Navigating visually impaired travelers in a large train station using smartphone and bluetooth low energy. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*, pages 604–611, 2016.

[112] Laehyun Kim, Sehyung Park, Sooyong Lee, and Sungdo Ha. An electronic traveler aid for the blind using multiple range sensors. *IEICE Electronics Express*, 6(11):794–799, 2009.

[113] Roberta L Klatzky, James R Marston, Nicholas A Giudice, Reginald G Golledge, and Jack M Loomis. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. *Journal of experimental psychology: Applied*, 12(4):223, 2006.

[114] Reinhard Koutny, Peter Heumader, and Klaus Miesenberger. A mobile guidance platform for public transportation. In *International Conference on Computers for Handicapped Persons*, pages 58–64. Springer, 2014.

[115] Gregory Kramer, Bruce Walker, Terri Bonebright, Perry Cook, John H Flowers, Nadine Miner, and John Neuhoff. The sonification report: Status of the field and research agenda. Technical report, National Science Foundation, 1999.

[116] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 1952.

[117] Ravi Kuber, Amanda Hastings, Matthew Tretter, and Dónal Fitzpatrick. Determining the accessibility of mobile screen readers for blind users. In *IASTED Conference on Human-Computer Interaction*. ACTA, 2012.

[118] Masaya Kubota, Masaki Kuribayashi, Seita Kayukawa, Hironobu Takagi, Chieko Asakawa, and Shigeo Morishima. Snap&nav: Smartphone-based indoor navigation system for blind people via floor map analysis and intersection detection. *Proceedings of the ACM on Human-Computer Interaction*, 8(MHCI):1–22, 2024.

[119] Harish Kumar, S Harish, SJ Lakshan, et al. Decision making algorithm for blind navigation assistance using deep learning. In *2022 1st International Conference on Computational Science and Technology (ICCST)*, pages 268–272. IEEE, 2022.

[120] Jayakanth Kunhoth, AbdelGhani Karkar, Somaya Al-Maadeed, and Asma Al-Attiyah. Comparative analysis of computer-vision and ble technology based indoor navigation systems for people with visual impairments. *International Journal of Health Geographics*, 2019.

[121] Orly Lahav and David Mioduser. Construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind. *Comp. in Human Behavior*, 2008.

[122] Orly Lahav, David W. Schloerb, Siddarth Kumar, et al. A virtual environment for people who are blind - a usability study. *Journal of Assistive Technologies*, 2012.

[123] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*. Morgan Kaufmann, 2017.

[124] Donguk Lee, James D Lewis, Patti M Johnstone, and Patrick N Plyler. Acceptable noise levels and preferred signal-to-noise ratios for speech and music. *Ear and Hearing*, 43(3):1013–1022, 2022.

[125] Young Hoon Lee and Gerard Medioni. Wearable rgbd indoor navigation system for the blind. In *European Conference on Computer Vision*, pages 493–508. Springer, 2014.

[126] James R Lewis. Multipoint scales: Mean and median differences and observed significance levels. *International Journal of Human-Computer Interaction*, 1993.

[127] Bing Li, Juan Pablo Munoz, Xuejian Rong, Qingtian Chen, Jizhong Xiao, Yingli Tian, Aries Arditi, and Mohammed Yousuf. Vision-based mobile indoor assistive navigation aid for blind people. *IEEE transactions on mobile computing*, 18(3):702–714, 2018.

[128] Jack M Loomis, Reginald G Golledge, and Roberta L Klatzky. Navigation system for the blind: Auditory display modes and guidance. *Presence*, 7(2):193–203, 1998.

[129] Jack M Loomis, James R Marston, Reginald G Golledge, and Roberta L Klatzky. Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. *Journal of visual impairment & blindness*, 99(4):219–232, 2005.

[130] Luca Andrea Ludovico and Giorgio Presti. The sonification space: A reference system for sonification tasks. *International Journal of Human-Computer Studies*, 85:72–77, 2016.

[131] Charlotte Magnusson, Henrik Danielsson, and Kirsten Rassmus-Gröhn. Non visual haptic audio tools for virtual environments. In *Haptic and Audio Int. Design*. Springer-Verlag Berlin Heidelberg, 2006.

[132] Madjid Maidi, Jean-Yves Didier, Fakhreddine Ababsa, and Malik Mallem. A performance study for camera pose estimation using visual marker based tracking. *Machine Vision and Applications*, 2010.

[133] Henry B Mann. Nonparametric tests against trend. *Econometrica: Journal of the econometric society*, pages 245–259, 1945.

[134] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, 1947.

[135] Steve Mann, Jason Huang, Ryan Janzen, Raymond Lo, Valmiki Rampersad, Alexander Chen, and Taqveer Doha. Blind navigation with a wearable range camera and vibrotactile helmet. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1325–1328, 2011.

[136] Michał Markiewicz and Marek Skomorowski. Public transport information system for visually impaired and blind people. In *International Conference on Transport Systems Telematics*, pages 271–277. Springer, 2010.

[137] Natalina Martiniello, Werner Eisenbarth, Christine Lehane, Aaron Johnson, and Walter Wittich. Exploring the use of smartphones and tablets among people with visual impairments: Are mainstream devices replacing the use of traditional visual aids? *Assistive Technology*, 2019.

[138] Sergio Mascetti, Dragan Ahmetovic, Andrea Gerino, Cristian Bernareggi, Mario Busso, and Alessandro Rizzi. Robust traffic lights detection on mobile devices for pedestrians with visual impairment. *Computer Vision and Image Understanding*, 2016.

[139] Sergio Mascetti, Lorenzo Picinali, Andrea Gerino, Dragan Ahmetovic, and Cristian Bernareggi. Sonification of guidance data during road crossing for people with visual impairments or blindness. *International Journal of Human-Computer Studies*, 2016.

[140] Andrew Mason, Nick Jillings, Zheng Ma, Joshua D Reiss, and Frank Melchior. Adaptive audio reproduction using personalized compression. In *Proc. 57th Audio Engineering Society Conference: 57th International Conference*, 2015.

[141] David McShefferty, William M Whitmer, and Michael A Akeroyd. The just-noticeable difference in speech-to-noise ratio. *Trends in hearing*, 19, 2015.

[142] Brett RC Molesworth, Marion Burgess, and Daniel Kwon. The use of noise cancelling headphones to improve concurrent task performance in a noisy environment. *Applied Acoustics*, 74(1), 2013.

[143] Brian CJ Moore and Brian R Glasberg. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The journal of the acoustical society of America*, 74(3):750–753, 1983.

[144] Karin Müller, Christin Engel, Claudia Loitsch, Rainer Stiefelhagen, and Gerhard Weber. Traveling more independently: a study on the diverse needs and challenges of people with visual or mobility impairments in unfamiliar indoor environments. *ACM Transactions on Accessible Computing (TACCESS)*, 15(2):1–44, 2022.

[145] Vidya N Murali and James M Coughlan. Smartphone-based crosswalk detection and localization for visually impaired pedestrians. In *2013 IEEE international conference on multimedia and expo workshops (ICMEW)*, pages 1–7. IEEE, 2013.

[146] Masayuki Murata, Dragan Ahmetovic, Daisuke Sato, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. Smartphone-based indoor localization for blind navigation across building complexes. In *International Conference on Pervasive Computing and Communications*, 2018.

[147] Masayuki Murata, Dragan Ahmetovic, Daisuke Sato, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. Smartphone-based localization for blind navigation in building-scale indoor environments. *Pervasive and Mobile Computing*, 57:14–32, 2019.

[148] Qingquan Na, Hui Zhou, Hailei Yuan, Mengfan Gui, and Hongjing Teng. Improving walking path generation through biped constraint in indoor navigation system for visually impaired individuals. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[149] Vishnu Nair, Greg Olmschenk, William H Seiple, and Zhigang Zhu. Assist: Evaluating the usability and performance of an indoor navigation assistant for blind and visually impaired people. *Assistive Technology*, 2020.

[150] Peter Bjorn Nemenyi. *Distribution-free multiple comparisons.* Princeton University, 1963.

[151] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999, 2017.

[152] Qun Niu, Ning Liu, Jianjun Huang, Yangze Luo, Suining He, Tao He, S-H Gary Chan, and Xiaonan Luo. Deepnavi: A deep signal-fusion framework for accurate and applicable indoor navigation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(3):1–24, 2019.

[153] Eshed Ohn-Bar, João Guerreiro, Kris Kitani, and Chieko Asakawa. Variability in reactions to instructional guidance during smartphone-based assisted navigation of blind users. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, 2(3):1–25, 2018.

[154] Wenyan Ou, Jiaming Zhang, Kunyu Peng, Kailun Yang, Gerhard Jaworek, Karin Müller, and Rainer Stiefelhagen. Indoor navigation assistance for visually impaired people via dynamic slam and panoptic segmentation with an rgb-d sensor. In *International Conference on Computers Helping People with Special Needs*, pages 160–168. Springer, 2022.

[155] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. Watchout: Obstacle sonification for people with visual impairment or blindness. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 402–413, 2019.

[156] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca A Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. Iterative design of sonification techniques to support people with visual impairments in obstacle avoidance. *ACM Transactions on Accessible Computing (TACCESS)*, 14(4):1–27, 2021.

[157] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Andrea Ludovico, Adriano Baratè, Federico Avanzini, and Sergio

Mascetti. Iterative design of sonification techniques to support people with visual impairments in obstacle avoidance. *ACM Transactions on Accessible Computing*, 14(4):1–27, 2021.

[158] Giuseppina Emma Puglisi, Anna Warzybok, Sabine Hochmuth, Chiara Visentin, Arianna Astolfi, Nicola Prodi, and Birger Kollmeier. An italian matrix sentence test for the evaluation of speech intelligibility in noise. *Int. J. of Audiology*, 2015.

[159] Pablo-Alejandro Quinones, Tammy Greene, Rayoung Yang, and Mark Newman. Supporting visually impaired navigation: a needs-finding study. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1645–1650. ACM, 2011.

[160] Jyri Rajamäki, Petri Viinikainen, Julius Tuomisto, Thomas Sederholm, and Miika Säämänen. Laureapop indoor navigation service for the visually impaired in a wlan environment. In *Proceedings of the 6th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications*, pages 96–101. World Scientific and Engineering Academy and Society (WSEAS), 2007.

[161] Nornadiah Mohd Razali, Yap Bee Wah, et al. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*, 2(1):21–33, 2011.

[162] Peng Ren, Jonathan Lam, Roberto Manduchi, and Fatemeh Mirzaei. Experiments with routenav, a wayfinding app for blind travelers in a transit hub. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–15, 2023.

[163] Maria Rosiak, Mateusz Kawulok, and Michał Maćkowski. The effectiveness of uwb-based indoor positioning systems for the navigation of visually impaired individuals. *Applied Sciences*, 14(13):5646, 2024.

[164] Philipp Ruchti, Bastian Steder, Michael Ruhnke, and Wolfram Burgard. Localization on openstreetmap data using a 3d laser scanner. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 5260–5265. IEEE, 2015.

[165] Yessica Sáez, José Muñoz, Francisco Canto, Antony García, and Héctor Montes. Assisting visually impaired people in the public transport system through rf-communication and embedded systems. *Sensors*, 19(6):1282, 2019.

[166] Georgina Santos and Nikolay Nikolaev. Mobility as a service and public transport: a rapid literature review and the case of moovit. *Sustainability*, 13(7):3666, 2021.

[167] Daisuke Sato, Uran Oh, João Guerreiro, Dragan Ahmetovic, Kakuya Naito, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. Navcog3 in the wild: Large-scale blind indoor navigation assistant with semantic features. *Transactions on Accessible Computing*, 2019.

[168] Stefano Scheggi, Agostino Talarico, and Domenico Prattichizzo. A remote guidance system for blind and visually impaired people via vibrotactile haptic feedback. In *22nd Mediterranean conference on control and automation*, pages 20–23. IEEE, 2014.

[169] C James Scheirer, William S Ray, and Nathan Hare. The analysis of ranked data derived from completely randomized factorial designs. *Biometrics*, pages 429–434, 1976.

[170] Victor R Schinazi, Tyler Thrash, and Daniel-Robert Chebat. Spatial navigation by congenitally blind individuals. *Wiley Interdisciplinary Reviews: Cognitive Science*, 7(1):37–58, 2016.

[171] Stephen Se. Zebra-crossing detection for the partially sighted. In *Proceedings IEEE conference on computer vision and pattern recognition. CVPR 2000 (Cat. No. PR00662)*, volume 2, pages 211–217. IEEE, 2000.

[172] Andrew Sears and Vicki Hanson. Representing users in accessibility research. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 2235–2238, 2011.

[173] Apurv Shaha, Shubham Rewari, and Sankaradithyan Gunasekharan. Swsvip-smart walking stick for the visually impaired people using low latency communication. In *2018 international conference on smart city and emerging technology (ICSCET)*, pages 1–5. IEEE, 2018.

[174] Swapnil Shingte and Rajesh Patil. A passenger bus alert and accident system for blind person navigational. *Int. J. Sci. Res. Sci. Technol*, 4(8):282–288, 2018.

[175] Kristen Shinohara and Jacob O Wobbrock. In the shadow of misperception: assistive technology use and social interactions. In *SIGCHI conference on human factors in computing systems*. ACM, 2011.

[176] A Siegel. The development of spatial representations of large-scale environments. *Advances in Child Development and Behavior/Academic Press*, 1975.

[177] Alexa F Siu, Mike Sinclair, Robert Kovacs, Eyal Ofek, Christian Holz, and Edward Cutrell. Virtual reality without vision: A haptic and auditory white cane to navigate complex virtual worlds. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020.

[178] Simone Spagnol, György Wersényi, Michał Bujacz, et al. Current use and future perspectives of spatial audio technologies in electronic travel aids. *Wireless Comm. and Mobile Comp.*, 2018.

[179] Lars St, Svante Wold, et al. Analysis of variance (anova). *Chemometrics and intelligent laboratory systems*, 6(4):259–272, 1989.

[180] Dana L. Strait, Nina Kraus, Alexandra Parbery-Clark, and Richard Ashley. Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. *Hearing Research*, 2010.

[181] Student. The probable error of a mean. *Biometrika*, pages 1–25, 1908.

[182] Martin Swobodzinski and Amy T Parker. A comprehensive examination of electronic wayfinding technology for visually impaired travelers in an urban environment. 2019.

[183] Arvind Thiagarajan, James Biagioni, Tomas Gerlich, and Jakob Eriksson. Cooperative transit tracking using smart-phones. In *Proceedings of the 8th ACM conference on embedded networked sensor systems*, pages 85–98, 2010.

[184] Catherine Thinus-Blanc and Florence Gaunet. Representation of space in blind persons: vision as a spatial sense? *Psychological bulletin*, 121(1):20, 1997.

[185] Bjora Thylefors, AD Negrel, Ramachandra Pararajasegaram, and KY Dadzie. Global data on blindness. *Bull of the WHO*, 73(1), 1995.

[186] Sandra E Trehub. Musical predispositions in infancy. *Annals of the New York academy of sciences*, 930(1):1–16, 2001.

[187] Sylvie Treuillet and Eric Royer. Outdoor/indoor vision-based localization for blind pedestrian navigation assistance. *International journal of Image and Graphics*, 10(04):481–496, 2010.

[188] Chia Hsuan Tsai, Fatemeh Elyasi, Peng Ren, and Roberto Manduchi. All the way there and back: Inertial-based, phone-in-pocket indoor wayfinding and backtracking apps for blind travelers. *ACM Transactions on Accessible Computing*, 2024.

[189] Martin Usoh, Kevin Arthur, Mary C. Whitton, et al. Walking > walking-in-place > flying, in virtual environments. In *Proceedings of SIGGRAPH*, 1999.

[190] Bruce N. Walker and Jeffrey Lindsay. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. *Human Factors*, 2006.

[191] Bruce N. Walker and Michael A. Nees. Theory of sonification. In *The sonification handbook*. Logos Verlag Berlin GmbH, 2011.

[192] Bruce N Walker and Michael A Nees. Theory of sonification. *The sonification handbook*, 2011.

[193] Ruolin Wang, Chun Yu, Xing-Dong Yang, Weijie He, and Yuanchun Shi. Eartouch: facilitating smartphone use for visually impaired people in mobile and public scenarios. In *Conf. on human factors in comp. systems*. ACM, 2019.

[194] Yunqing Wang and Katherine J Kuchenbecker. Halo: Haptic alerts for low-hanging obstacles in white cane navigation. In *2012 IEEE Haptics Symposium (HAPTICS)*, pages 527–532. IEEE, 2012.

[195] WebAIM. Screen reader user survey, 2021.

[196] Adam Weisser and Jörg M Buchholz. Conversational speech levels and signal-to-noise ratios in realistic acoustic conditions. *The Journal of the Acoustical Society of America*, 145(1):349–360, 2019.

[197] Jan M. Wiener, Simon J. Büchner, and Christoph Hölscher. Taxonomy of human wayfinding tasks: A knowledge-based approach. *Spatial Cognition & Comp.*, 2009.

[198] William R Wiener, Richard L Welsh, and Bruce B Blasch. *Foundations of orientation and mobility*, volume 1. American Foundation for the Blind, 2010.

[199] Frank Wilcoxon. Individual comparisons by ranking methods. In *Breakthroughs in statistics: Methodology and distribution*, pages 196–202. Springer, 1992.

[200] Jeff Wilson, Bruce N Walker, Jeffrey Lindsay, Craig Cambias, and Frank Dellaert. Swan: System for wearable audio navigation. In *2007 11th IEEE international symposium on wearable computers*, pages 91–98. IEEE, 2007.

[201] Sandy Wong. Traveling with blindness: A qualitative space-time approach to understanding visual impairment and urban mobility. *Health & Place*, 49:85–92, 2018.

[202] Yuan Wu, Hai-Bing Zhu, Qing-Xiu Du, and Shu-Ming Tang. A survey of the research status of pedestrian dead reckoning systems based on inertial sensors. *International Journal of Automation and Computing*, 16:65–83, 2019.

[203] Shuchang Xu, Ciyuan Yang, Wenhao Ge, Chun Yu, and Yuanchun Shi. Virtual paving: Rendering a smooth path for people with visual impairment through vibrotactile and audio feedback. *Proceedings of the ACM*

on Interactive, Mobile, Wearable and Ubiquitous Technologies, 4(3):1–25, 2020.

[204] Ciyuan Yang, Shuchang Xu, Tianyu Yu, Guanhong Liu, Chun Yu, and Yuanchun Shi. Lightguide: Directing visually impaired people along a path using light cues. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(2):1–27, 2021.

[205] Hanlu Ye, Meethu Malu, Uran Oh, and Leah Findlater. Current and future mobile and wearable device use by people with visual impairments. In *Conference on Human Factors in Computing Systems*. ACM, 2014.

[206] Chris Yoon, Ryan Louie, Jeremy Ryan, MinhKhang Vu, Hyegi Bang, William Derksen, and Paul Ruvolo. Leveraging augmented reality to create apps for people with visual disabilities: A case study in indoor navigation. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pages 210–221, 2019.

[207] Carolyn Yu, Yee Li, Tsai-Yuan Huang, Wei-An Hsieh, Shao-Yu Lee, I-Hui Yeh, Gang Ku Lin, Neng-Hao Yu, Hsien-Hui Tang, and Yung-Ju Chang. Busmyfriend: Designing a bus reservation service for people with visual impairments in taipei. In *Companion Publication of the 2020 ACM Designing Interactive Systems Conference*, pages 91–96, 2020.

[208] Muhammad Yusro, Kun Mean Hou, E Pissaloux, HL Shi, Kalamullah Ramli, and Dodi Sudiana. Sees: Concept and design of a smart environment explorer stick. In *2013 6th International Conference on Human System Interactions (HSI)*, pages 70–77. IEEE, 2013.

[209] Yuhang Zhao, Elizabeth Kupferstein, Hathaitorn Rojnirun, Leah Findlater, and Shiri Azenkot. The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–14, 2020.

[210] Pengfei Zhou, Yuanqing Zheng, and Mo Li. How long to wait? predicting bus arrival time with mobile phone based participatory sensing. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 379–392, 2012.

[211] Yancong Zhu, Cong Wang, Wei Liu, and Yi Lv. Running guide: Design of a marathon navigation system for visually impaired people. In *Proceedings of the Seventh International Symposium of Chinese CHI*, pages 7–15, 2019.

[212] Susanne Zimmermann-Janschitz, Bettina Mandl, and Antonia Dückelmann. Clustering the mobility needs of persons with visual impairment or legal blindness. *Transportation Research Record*, 2650(1):66–73, 2017.

Appendix A

State of the art contributions

The table shows the main contributions of each relevant paper in the literature according to the four layers presented in Section 1.3.

Table A.1: Related papers with layers and services implemented.
L = *Localization*, R = *Routing*, G = *Guidance*, I = *Interface*

Paper	**L**	**R**	**G**	**I**	**Main contributions**
[2]	X	X			Localization: Combination of WiFi positioning, beacon positioning and Radio Frequency Identification positioning. Routing: Grid representation, the grid cells indicate where it is possible to go, graph representation derived from the grid representation.
[3]		X	X	X	Routing: Graph representation, turn-by-turn navigation, veering correction, update the target at a given threshold to the waypoint. Guidance: Navigation Instructions. Interface: Audio feedback.
[4]			X	X	Guidance: Navigation Instructions. Interface: Audio feedback.
[7]	X			X	Localization: Visual-inertial odometry using Augmented reality, Visual landmarks. Interface: Audio feedback. Services: Descriptor of the Environmental features.

Paper	L	R	G	I	Main contributions
[9]	X	X		X	Localization: Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons. Routing: Graph representation, turn-by-turn navigation, veering correction. Guidance: Navigation Instructions, Point of Interest information. Interface: Audio feedback. Services: Point of interest setting and search.
[10]			X	X	Guidance: Navigation Instructions, updates about mobility. Interface: Multimodal feedback.
[12]			X		Guidance: Navigation Instructions, security announcements.
[15]	X				Localization: Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons, Inertial sensor positioning.
[19] [65] [101] [109] [168]				X	Interface: Haptic feedback.
[29]	X			X	Localization: WiFi fingerprinting. Interface: Audio feedback. Services: Bus Transport Assistance.
[31]			X	X	Guidance: Navigation Instructions, updates about mobility. Interface: Multimodal feedback. Services: Obstacle detection and avoidance.
[35] [155] [156] [38]				X	Interface: Audio feedback. Services: Obstacle detection and avoidance.
[40]	X				Localization: GNSS positioning combined with visual information.
[41]	X	X	X		Localization: Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons. Routing: graph representation, turn-by-turn navigation.
[42]	X	X			Localization: Radio Frequency Identification positioning. Routing: grid representation, graph representation derived from the grid representation, turn-by-turn navigation, veering correction.

Paper	L	R	G	I	Main contributions
[43]	X	X		X	Localization: Visual-inertial odometry. Routing: graph representation, turn-by-turn navigation, the connection between anchors represents a straight path, anchors as nodes. Interface: Multimodal feedback. Services: Record and repeat the route.
[45][107]				X	Interface: Haptic feedback. Services: Obstacle detection and avoidance.
[48]				X	Interface: Audio feedback. Services: Point of interest setting and search.
[49]	X				Localization: Visual inertial-odometry, Visual landmarks, Particle filter
[51]	X			X	Localization: Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons. Interface: Audio feedback.
[62]			X	X	Guidance: Navigation Instructions, security announcements. Interface: Audio feedback
[66]	X		X		Localization: Inertial sensor positioning. Guidance: Navigation Instructions, updates about mobility. Services: Record and repeat the route.
[67]	X			X	Localization: WiFi fingerprinting. Interface: Audio feedback. Services: Bus Transport Assistance.
[69]	X			X	Localization: Visual-inertial odometry using Augmented reality, Visual landmarks. Interface: Audio feedback.
[72]	X				Localization: Combination of Visual inertial-odometry with Inertial sensor positioning.
[73]	X				Localization: Combination of WiFi Fingerprinting positioning and Inertial sensor positioning.
[79]	X	X			Localization: Crowdsourcing-based Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons. Routing: Graph representation, turn-by-turn navigation, veering correction, paths organized as a network of routes between the beacons.
[80]	X			X	Localization: GNSS positioning. Interface: Audio feedback. Services: Geographic Information system.

Paper	**L**	**R**	**G**	**I**	**Main contributions**
[84]	X		X		Localization: Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons. Guidance: Navigation Instructions, point of interest information, security announcements and updates about mobility.
[85]	X			X	Localization: LiDAR scanning, Scene understanding using CNN. Interface: Haptic feedback. Services: Descriptor of the Environmental features. Services: Obstacle detection and avoidance.
[87]	X				Localization: Particle filter.
[91]	X	X			Localization: Combination of WiFi Fingerprinting positioning and Inertial sensor positioning. Routing: Graph representation, turn-by-turn navigation, veering correction.
[94]	X				Localization: Received Signal Strength Indication (RSSI) of BLE beacons combined with images in Deep Learning using a CNN.
[95]	X	X	X	X	Localization: Visual-inertial odometry using existing street cameras. Routing: Graph representation, turn-by-turn navigation, veering correction. Guidance: Navigation Instructions, security announcements and updates about mobility. Interface: Multimodal feedback.
[106]	X			X	Localization: GNSS positioning combined with visual landmarks. Interface: Audio feedback.
[110]	X			X	Localization: Acoustic-based localization. Interface: Audio feedback.
[111]	X	X		X	Localization: Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons. Routing: Multiple representation, hierarchy of areas and graph representation, update the target at a given threshold to the point of reference. Interface: Multimodal feedback.
[113]				X	Interface: Audio feedback.

Paper	L	R	G	I	Main contributions
[118]	X	X			Localization: LiDAR scanning, floormap detection and understanding. Routing: Graph representation, turn-by-turn navigation.
[125]	X	X	X	X	Localization: Combination of Visual inertial-odometry with Inertial sensor positioning. Routing: Grid, the grid cells indicate where it is possible to go or unknown areas, recompute route based on obstacle on the trajectory. Guidance: Navigation Instructions. Interface: Haptic feedback. Services: Obstacle detection and avoidance.
[127]		X		X	Routing: grid representation, the grid cells indicate where it is possible to go or unknown areas, graph representation derived from the grid representation, recompute route based on obstacle on the trajectory. Interface: Multimodal feedback. Services: Obstacle detection and avoidance.
[128] [129]	X			X	Localization: GNSS positioning. Interface: Audio feedback.
[135]				X	Interface: Haptic feedback. Services: Obstacle detection and avoidance.
[136]	X			X	Localization: GNSS positioning. Interface: Audio feedback. Services: Public Transportation Mobility Support.
[146] [147]	X				Localization: Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons, Inertial sensors positioning, Particle filter
[148]	X	X	X	X	Localization: LiDAR scanning, Inertial sensor positioning. Routing: grid representation, the target is the next gait point, recompute route based on obstacle on the trajectory. Guidance: Navigation Instructions. Interface: Haptic feedback.
[152]	X				Localization: Scene understanding using CNN, Recurrent Neural Networks for inertial sensors.
[160]	X				Localization: WiFi fingerprinting.

Paper	L	R	G	I	Main contributions
[162]	X	X	X	X	Localization: Inertial sensors combined with a Particle filter. Routing: Tile representation, update the target at a given threshold to the next tile, veering correction. Guidance: Navigation Instructions, Point of Interest information. Interface: Multimodal feedback.
[163]	X				Localization: Received Signal Strength Indication (RSSI) of BLE beacons combined with Ultra Wide Band beacons using Two-Way Ranging.
[167]	X		X	X	Localization: Fingerprinting of Received Signal Strength Indication (RSSI) of BLE beacons, Semantic features. Guidance: Navigation Instructions, Point of Interest information. Interface: Audio feedback. Services: Point of interest setting and search.
[174]	X			X	Localization: GNSS positioning combined with Inertial sensor positioning. Interface: Audio feedback. Services: Bus Transport Assistance.
[188]	X	X	X		Localization: Inertial sensors combined with a Particle filter. Routing: graph representation, turn-by-turn navigation, veering correction. Guidance: Navigation Instructions, updates about mobility. Services: Record and repeat the route.
[187]	X			X	Localization: Visual landmarks. Interface: Audio feedback.
[200]	X	X		X	Localization: GNSS positioning combined with Inertial sensor positioning. Routing: Graph representation, turn-by-turn navigation, veering correction. Interface: Audio feedback. Services: Record and repeat the route.
[203]			X	X	Guidance: Navigation Instructions, security announcements and updates about mobility. Interface: Haptic feedback.
[204]			X	X	Guidance: Navigation Instructions. Interface: Haptic feedback.

Paper	**L**	**R**	**G**	**I**	**Main contributions**
[206]		X	X	X	Routing: Multiple representation, checkoff zones and graph representation, update the target when user enter in the check-off zone. Guidance: Navigation Instructions. Services: Record and repeat the route.
[207]			X	X	Guidance: Navigation Instructions, updates about mobility. Interface: Haptic feedback. Services: Bus Transport Assistance.
[209]			X	X	Guidance: Navigation Instructions, security announcements and updates about mobility. Interface: Multimodal feedback.
[210]	X				Localization: Crowdsourcing-based inertial sensors segments, Acoustic-based localization. Services: Bus Transport Assistance.
[211]				X	Interface: Multimodal feedback.

Acknowledgments

The PhD is a long path of training and study, but above all, of work, requiring great effort, fatigue, and sacrifices.

First and foremost, I would like to express my deepest gratitude to Professor Sergio Mascetti and Professor Dragan Ahmetovic for giving me the opportunity to work with them, for the patience, support, and trust they have always shown me. Thank you for helping me and for giving me the chance to prove myself, even when it seemed unlikely that I would reach this milestone. I am infinitely grateful.

A heartfelt thank you to Dr. James Coughlan and the Smith-Kettlewell Eye Research Institute for offering me the opportunity to visit San Francisco and for funding part of my PhD scholarship. I would also like to thank Assia Andrao and Retina Italia odv for giving me the opportunity to work on research projects outside of my main field and for the financial support during my journey.

I would then like to thank my parents, who shared this journey with me and emotionally supported me during the toughest moments of the PhD.

A special thanks to my collaborators, who contributed significantly to the research activities. In particular, I would like to express my gratitude to M. Ciardullo, A. Mariucci, L. Morali, M. Galimberti (Uncle), and A. Galimberti (Dad): they have truly understand the immense effort that scientific research entails and what it means to dedicate oneself to this profession.

I also thank the people with whom I shared, in full or in part, the PhD journey, working and engaging with each other every day: Marco , Priyankar, Aiman, Luca A., Riccardo, Annalisa, Michele, Matteo M., Matteo C, Federico A., Giorgio P., Luca A. L, and Cristian.

Finally, a special thought goes to those who offered me friendship, emotional support along this journey: Silvia < 3, Daniele M., Daniele L., Luca P., Edoardo, Lorenzo, Andrea P., Andrea M., Diego, Jacopo, Francesca, India, Giulia, Chiara M., Chiara P., Ash, Robin, Matteo R., Mattia, Roxana, Jillian, Alex, Wahid, Valentina, Marina, Kevin, and Laura. Thank you all for making this experience lighter.

Lastly, I would like to thank all the individuals with visual impairments who contributed to the completion of this thesis and who invested their time in achieving this result and those that will follow.

Thank you!