Int. J. Human–Computer Studies 177 (2023) 103057

Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Sonification of navigation instructions for people with visual impairment

Dragan Ahmetovic *, Federico Avanzini, Adriano Baratè, Cristian Bernareggi, Marco Ciardullo, Gabriele Galimberti, Luca A. Ludovico, Sergio Mascetti, Giorgio Presti

Università degli Studi di Milano, Dipartimento di Informatica, Via G. Celoria, 18, 20133, Milan, Italy

A R T I C L E I N F O

Keywords:
Visual impairment
Mobility
Navigation assistance
Sonification

A B S T R A C T

Navigation assistance services for people with visual impairment pose the challenge of providing accurate guidance through non-visual navigation instructions. This paper 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 participants having severe visual impairment. 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 people with visual impairment.

1. Introduction

For a traveler visiting an unknown environment, Navigation Assistance (NA) services provide invaluable help in finding the correct path toward a destination. Such services are commonly available as applications for commodity mobile devices and rely on different localization techniques, such as Global Navigation Satellite Systems (GNSS)[1] Katz et al. (2010), beacon networks (Murata et al., 2019), or visual positioning (Coughlan and Manduchi, 2009), coupled with maps and other survey data 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 (Ahmetovic et al., 2016), in which the user is instructed to turn at specific decision points and then proceed along a (usually) straight path.

NA services are particularly important for people with blindness or visual impairment (BVI) because these users have difficulties in accessing visual navigation cues present in the environment. For people with BVI, localization and navigation accuracy in NA services needs to be higher than for users without BVI, 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 users with BVI within an environment (Kunhoth et al., 2019; Murata et al., 2019; Nair et al., 2020).

A second challenge, which we address in this paper, is that the navigation assistance needs to be provided through accurate non-visual instructions that the users with BVI 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 (Ahmetovic et al., 2018b). Thus, prior work has also explored methods to accurately guide users with BVI through verbal instructions (Sato et al., 2019), non-verbal sounds (Mascetti et al., 2016c) or haptic feedback (Azenkot et al., 2011). Prior work has also explored methods to compensate for the inherent user imprecision in following sound instructions, in order to ensure more accurate navigation assistance (Ahmetovic et al., 2019).

1.1. Previous contributions

In our previous work, presented at the International Conference on Pervasive Computing and Communications (Ahmetovic et al., 2019a),

* Corresponding author.
E-mail addresses: dragan.ahmetovic@unimi.it (D. Ahmetovic), federico.avanzini@unimi.it (F. Avanzini), adriano.barate@unimi.it (A. Baratè), cristian.bernareggi@unimi.it (C. Bernareggi), marco.ciardullo@studenti.unimi.it (M. Ciardullo), gabriele.galimberti@unimi.it (G. Galimberti), luca.ludovico@unimi.it (L.A. Ludovico), sergio.mascetti@unimi.it (S. Mascetti), giorgio.presti@unimi.it (G. Presti).

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

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

https://doi.org/10.1016/j.ijhcs.2023.103057
Received 21 October 2022; Received in revised form 12 April 2023; Accepted 16 April 2023
Available online 10 May 2023

we investigated sonification, that is the use of non-speech audio to represent information (Walker and Nees, 2011a), as a means to accurately guide travelers with BVI during NA. Specifically, we explored how people with BVI follow **turn instructions** while supported by continuous sonification feedback. 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[2]:

- **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 to a baseline approach (**Ping** - **P**), used in existing solutions (Ahmetovic et al., 2016), 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 participants with BVI during turns. We discovered that the MS sonification has 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.

1.2. Novel contributions

Our prior work shows that the proposed sonification approaches can be useful to support people with BVI during NA. However, the investigation considered only a portion of the turn-by-turn NA 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 accuracy. To address these issues, this paper 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 NA.
- We design and implement a prototype mobile application that provides turn-by-turn NA 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 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 NA 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 NA solutions for people with BVI (Sato et al., 2019), replacing the baseline approach which is currently used, and resulting in more reliable, accurate, and safe guidance.

2. Related work

People with BVI use auditory, haptic, and vestibular sensing, along with their residual sight (if any) to explore their surroundings and travel to a destination (Golledge et al., 1996). However, non-visual way-finding is less accurate and has a lower sensory range compared to visual navigation, hence it is also longer, more error prone and more cognitively demanding (Ungar, 2000). Thus, people with BVI avoid traversing unfamiliar routes while unassisted (Manduchi and Kurniawan, 2011) and tend to learn new routes first through Orientation and Mobility (O&M) training (Wiener et al., 2010).

Assistive Technologies (ATs) supporting people with BVI during navigation in unfamiliar environments (Csapó et al., 2015), and in particular mobile ATs (Periša et al., 2017) are being researched widely. However, the adoption and usage of such ATs is not homogeneous among people with BVI and largely depends on the user characteristics and usage scenario (Ahmetovic et al., 2019b).

2.1. Orientation and mobility by people with BVI

Orientation and mobility (O&M) by people with BVI has been largely studied in the field of cognitive neuroscience. In these studies, the O&M strategies adopted by blind people (Schinazi et al., 2016) and by people with low vision (Allen, 1977) 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 (Siegel and White, 1975). This representation is based on non-visual sensory information, such as sound, touch, and proprioception. According to the **Route model** (Fletcher, 1980), 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 (Giudice, 2018). 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 (Thinus-Blanc and Gaunet, 1997). Recent work argues that these three different O&M models are not used independently, but they are actually integrated by blind people while navigating (Giudice, 2018).

As far as people with low vision are concerned, the role of residual sight is dominant in O&M strategies (Allen, 1977). Nonetheless, integration of residual sight with other non-visual strategies to acquire information about the environment (*e.g.* echolocation) is essential also for people with low vision (Haymes et al., 1996) and it can be trained in specialized courses (Emerson, 2020).

2.2. Navigation systems for people with BVI

Numerous assistive technologies have been researched to support independent mobility for people with BVI. Among these, smartphone-based tools are becoming increasingly popular (Hakobyan et al., 2013) due to their sensing and computational capabilities and the presence of native accessibility features, such as screen readers and magnification. These accessibility features allow people with BVI to interact with most mobile applications, including navigation tools such as maps and Global Positioning System (GPS) navigation (Periša et al., 2017). Specifically designed tools have also been studied. Some approaches rely on smartphone sensors to perceive the surrounding environment and inform the user of the features of interest. Computer vision is used to detect visual cues (Budrionis et al., 2022; See et al., 2022), such as pedestrian crossings (Mascetti et al., 2016b) or traffic lights (Mascetti et al., 2016a), and notify the user. Approaches using Inertial Motion Unit (IMU) sensors have also been proposed to locate users on a traversed path and assist them on the way back (Flores and Manduchi, 2018; Sobnath et al., 2020). Other techniques provide information sourced from online databases. Geographical Information Systems (GIS) are used to provide nearby points of interest (Kacorri et al., 2016, 2018), while street-level imagery is employed to crowd-source accessibility information (Hara et al., 2015). Computer vision analysis of satellite and street-level images is also used to detect mobility cues (Ahmetovic et al., 2017a). The recent introduction of disruptive applications of the AR technology has enabled the development of mobile applications that provide guidance through auditory and haptic messages as well as through visual representations of the pathway (*e.g.*, obstacles, landmarks, points of interest) specifically designed for people with BVI (Medina-Sanchez et al., 2021; Lo Valvo et al., 2021; Froissard et al., 2014).

Turn-by-turn navigation is a guidance paradigm that translates a route into a sequence of straight paths and turning points. It is commonly used by sighted users for outdoor vehicular guidance using GPS. This approach is also useful for people with BVI because no prior knowledge of the environment is needed to follow navigation instructions (Fenech et al., 2010). Due to the low accuracy of GPS systems in indoor environments, alternative methods such as WiFi (Rajamäki et al., 2007), Visual Light Communication (VLC) (Nakajima and Haruyama, 2012), and Bluetooth Low Energy (BLE) beacons (Murata et al., 2018) have been studied to provide meter-level guidance, which is considered to be sufficiently accurate for people with BVI (Murata et al., 2018).

However, even with high accuracy, localization errors in turn-by-turn navigation are still possible (Ahmetovic et al., 2017b). In addition to system errors, user behavior (Ohn-Bar et al., 2018; Guerreiro et al., 2018) can also impact navigation accuracy and cause errors such as veering (Guth and LaDuke, 1994) or imprecision during rotation (Chrastil and Warren, 2017) for both sighted (Sadalla and Montello, 1989) and blind (Marlinsky, 1999) individuals. For sighted people, such errors are commonly compensated through visual inspection (Zacharias and Young, 1981). However, for people with BVI, these errors may impact the navigation outcome or even endanger the user (Ahmetovic et al., 2018b). In particular, turning errors during navigation assistance are related not only to *encoding* (*i.e.*, understanding the correct angle to turn) but also to *execution* (*i.e.*, reproducing the correct turning angle) (Chrastil and Warren, 2017). Specifically, it has been argued that rotation angles are encoded at a 90° rate (Jürgens et al., 1999). Indeed, such angles were shown to be easier to detect and track (Ahmetovic et al., 2018b). Preliminary studies suggest that conveying rotations using more accurate instructions, such as sonification-based interaction, may improve the accuracy of executing the rotations (Ahmetovic et al., 2018a). An alternative approach is to predict and compensate for the inherent user imprecision in following sonification instructions, ensuring more accurate NA (Ahmetovic et al., 2019).

2.3. Sonification techniques supporting people with BVI

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" (Kramer et al., 1999; Walker and Nees, 2011b). The term *sonification* stands for the auditory counterpart of data visualization. The use of non-verbal sound may be advantageous over speech in many applications and in several respects, including robustness to background noise, reduced cognitive load, and linguistic differences (Dingler et al., 2008).

Prior work has proposed taxonomies of sonification functions and techniques (Walker and Nees, 2011b). With specific regard to techniques, some of the most commonly employed ones include audification, model-based, and parameter mapping sonification. The latter in particular (Grond and Berger, 2011) 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, which should not be arbitrary nor excessively complex, in order for the sonification to effectively convey information about the data (Scaletti and Craig, 1991). In this work, 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 (Dubus and Bresin, 2013) 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. The **IS** sonification presented in this work 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 (Coop, 2016). Musification, namely the description of information through music objects, may also enhance engagement in addition to mere data representation. The **MS** sonification presented in this work provides an example of musification.

Several sonification approaches have been developed to improve orientation and mobility performance, especially for people with BVI. Many approaches, in particular, leverage spatialized audio to provide NA (Loomis et al., 2005). For this, some approaches use simple panning of stereo sound (Mascetti et al., 2016c), while others provide immersive 3D sound (Spagnol et al., 2018). The Personal Guidance System (Loomis et al., 1998), in particular, is a seminal work in this field. This system obtains information from a GPS receiver and uses different types of auditory displays, spatial sound delivery methods, and tracker locations (Loomis et al., 2005). Microsoft corporation has also investigated the use 3D sound in the Soundscape project (Anon, 2023) to enrich ambient awareness.

Klatzky et al. (2006) 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. Usoh et al. (1999) 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 people with BVI is provided by Spagnol et al. (2018), albeit with a focus on spatial audio. In their research, scholars and developers addressed the three main levels of spatial knowledge (Wiener et al., 2009), *i.e.*, knowledge about a point in space (*e.g.*, a landmark, an obstacle, a destination), knowledge about a sequence of points (a path to a destination, or "route knowledge"), and integrated knowledge about the environment (*i.e.*, cognitive-map like knowledge, or "survey knowledge"). Regarding the first level, typical applications deal with non-visual scene representation. These include obstacle detection (Bujacz et al., 2012), identification of relevant or potentially dangerous elements (*e.g.*, stairs or pedestrian crossings) (Mascetti et al., 2016c;

Bujacz et al., 2016), and full-scene representation (Meijer, 1992; Shoval et al., 1998).

Regarding the second level, *i.e.*, knowledge about a sequence of points, which is the most relevant for the purpose of the present study, several approaches have been proposed for the use of sound in wayfinding tasks. 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 (Walker and Lindsay, 2006). NavCog localizes the user through the sensing of the nearby Bluetooth beacons and provides navigation instructions through vocal messages coupled with impulse sonification feedback, which is the same technique used as the baseline in our experiments (Ahmetovic et al., 2016).

Regarding the third level, namely the integrated 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 (Guerreiro et al., 2017). Lahav and coworkers (Lahav and Mioduser, 2008) performed several studies on map exploration by blind subjects, which led to the development of the audio-haptic BlindAid system (Lahav et al., 2012). Katz and coworkers developed a system exploiting a 3D audio virtual environment to investigate the structural properties of spatial representations (Katz et al., 2012). They used a "ears in hand" metaphor (Magnusson et al., 2006), 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. in developing a non-visual system for offline exploration of 2D maps.

3. Sonification design

In the turn-by-turn NA 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 people with BVI convey such distances through verbal instructions at the beginning of each activity (*e.g.*, "proceed 10 meters" (Sato et al., 2019) or "turn to 11 o'clock" (Brady et al., 2015)), and confirm the completion of the activity with a single impulse **Ping** (**P**) sound (Presti et al., 2021). However, the inherent imprecision in following such instructions can lead to navigation errors and potential hazards (Ahmetovic et al., 2018b)

With the aim of providing a more accurate NA for people with BVI, 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 (Ahmetovic et al., 2019a), for the task of conveying turn instructions, we have investigated three sonification techniques (briefly described in Section 1.1): **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 NA systems (Sato et al., 2019). 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 paper, 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.

3.1. 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 (Ludovico and Presti, 2016).

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 (Bass III). Also, the generalizability of this choice is justified by the ubiquity of the major scale in popular music (Johnson, 2009) and a musical predisposition in infancy reported in scientific literature (Trehub, 2001).

Finally, **P** is basically an auditory icon (Gaver, 1989) 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.

3.2. 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 is vaguely resemblant to a Geiger–Müller counter (Geiger et al., 1928). 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.

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.

4. Prototype navigation assistance application

In order to experimentally evaluate the proposed sonification techniques in a real-world environment, we developed a prototype mobile NA application that implements them. The application localizes the user in the environment (see Section 4.1), generates the route towards a target destination (see Section 4.2), and conveys the navigation instructions to the user through an auditory interface (see Section 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 5.

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 (Maidi et al., 2010). 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 (Huang, 2019) and Simultaneous Localization and Mapping (SLAM) (Durrant-Whyte and Bailey, 2006). 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).

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. Fig. 1(a) 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 Fig. 1(b) 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.

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 (Ahmetovic et al., 2021). Another potential application is outdoor visual localization. For this use case, a solution is already implemented in AR libraries.[3]

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 people with BVI (Guth and LaDuke, 1994) 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. Fig. 1(c) 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 Fig. 1(d): 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

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

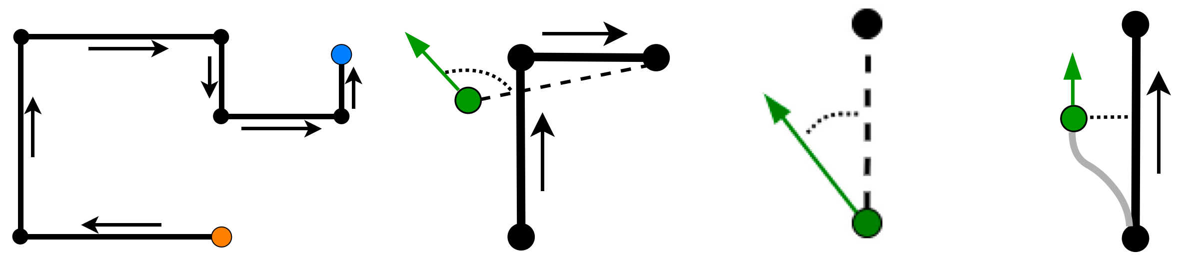


Fig. 1. Navigation strategies.

second compensation strategy, if the *lateral offset* is larger than a given threshold, the user is instructed to move sideways towards the planned route through a verbal message (*e.g.*, "move right"), similarly to prior work (Sato et al., 2019).

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[4]: 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".

5. User study

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

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 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[5] (for each question, we indicate the name we use in the paper to refer to that question):

Pleasantness: "The sound[6] 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.

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 4.2). Specifically, the Mann–Whitney-U test (Mann and Whitney, 1947) was used when two data series were analyzed, and for more data series we used the Kruskal–Wallis test (Kruskal and Wallis, 1952), with the Dunn test for post-hoc comparisons (Dunn, 1961). The target significance levels in multiple comparisons were adjusted using

[4] Henceforth we report the message translation from Italian.

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

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

Table 1
Participants' demographic information.

PID	Age	Gender	Impairment		Expertise		Screen reader		Route	
			Level	Onset	Music	Mobile	Speed	Voice	Familiar	Unfamiliar
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

For impairment: LB=legally blind, TB=totally blind. For route frequency: D=daily, W=weekly, MO=monthly, R=rarely.

the Benjamini–Hochberg False Discovery Rate method (Benjamini and Hochberg, 1995). Instead, subjective Likert-like responses were paired across different sonifications, and therefore Friedman's test (Friedman, 1937) was used. The Nemenyi test, which already accounts for familywise error, was used for post-hoc comparisons (Nemenyi, 1963).

5.2. Participants

The study involved 9 participants, whose demographic data are reported in Table 1. The inclusion criteria for this study required the participants to be people with BVI, unable to orient themselves with residual sight (namely persons 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).[7] 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 (Strait et al., 2010) and thus the ability to comprehend sonification techniques. We also collected the participant'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 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.[8]

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.

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

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

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 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 (people with BVI 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 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 people with BVI to avoid hazards.

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 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 pre-defined 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.

Table 2
Impact of sonifications on turns.

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

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 5.1.

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).

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 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.

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 2.

Considering the over-rotation errors (see Fig. 2(a)), 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 Fig. 2(b)). 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 Fig. 2(c)).

One finding reported in prior literature highlights that people with BVI tend to incur larger rotations errors when the requested rotation angle is small (Ahmetovic et al., 2018b). Our results confirm this finding (see Fig. 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).

Table 3
Impact of sonifications on over-rotations for small and large 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 4
Impact of sonifications on straight paths.

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

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 4.

For what concerns over-walking errors (see Fig. 4(a)), 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** and **MS+P**. Instead, no significant differences emerge for over-walking times and average walking speeds considering different sonifications (see Figs. 4(b) and 4(c)).

6.4. Participants' subjective feedback

Fig. 5 shows the distribution of the answers to the closed questions presented in Section 5.1. Mean values and standard deviations[9] are reported in Table 5.

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:

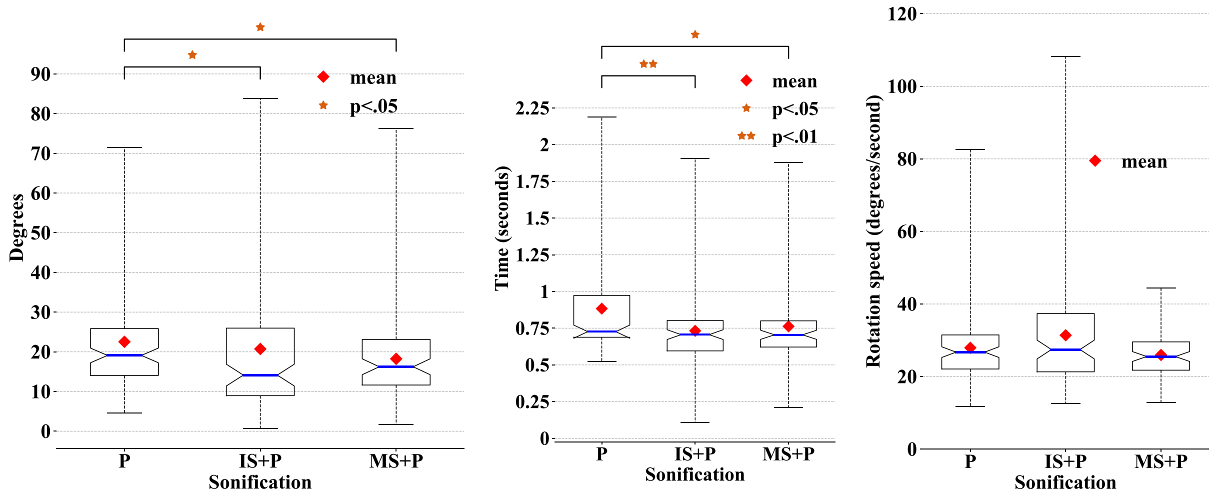
> **P5**:*[With **P**] I could hear the background environment ... [**IS+P**] gave me a good understanding of how close the target was.*

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**

[9] We report means for multipoint items as they were found to better indicate central tendency than medians (Lewis, 1993).



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

Fig. 2. Rotation Metrics.

Table 5
Subjective feedback about sonifications.

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

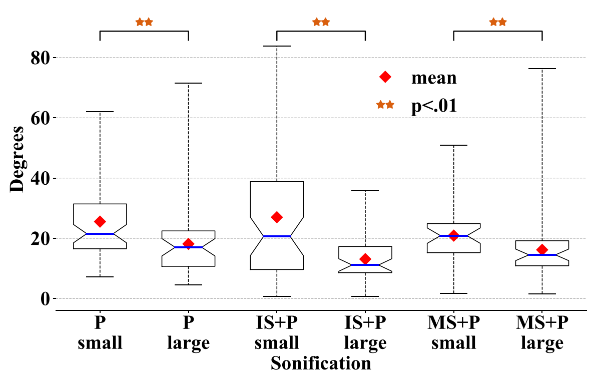


Fig. 3. Comparing over-rotation error for *small* (*i.e.*, less than 60°) and *large* (*i.e.*, more than 60°) turns.

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**).[10] In contrast with them, **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:

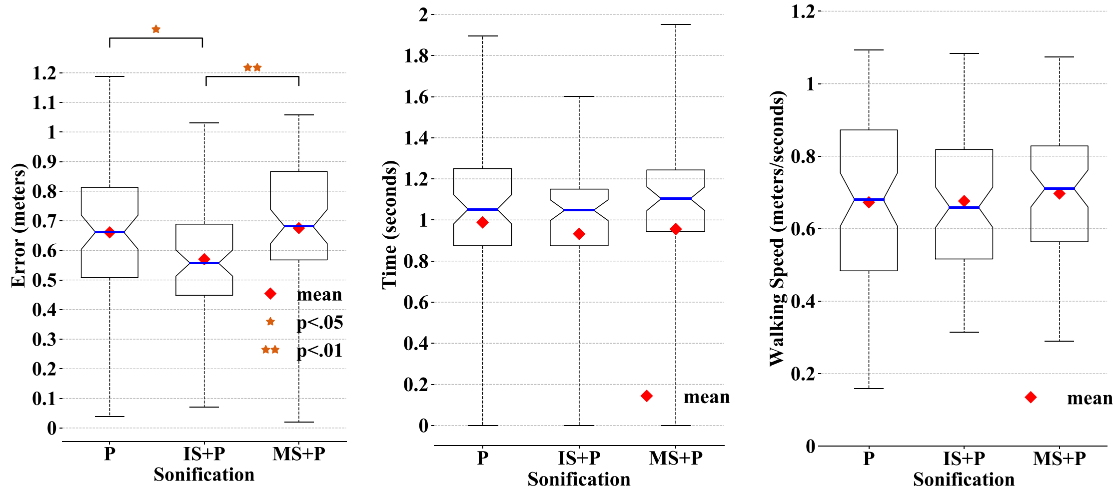
> **P9**: *I am not used to walking then stopping and rotating, I usually do it all continuously.*

7. Discussion

7.1. Study results

Our previous work (Ahmetovic et al., 2019a) 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 paper confirms our previous results regarding rotations, also in a realistic experimental scenario. Indeed **IS+P** and **MS+P** resulted significantly better 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

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



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

Fig. 4. Metrics on straight paths.

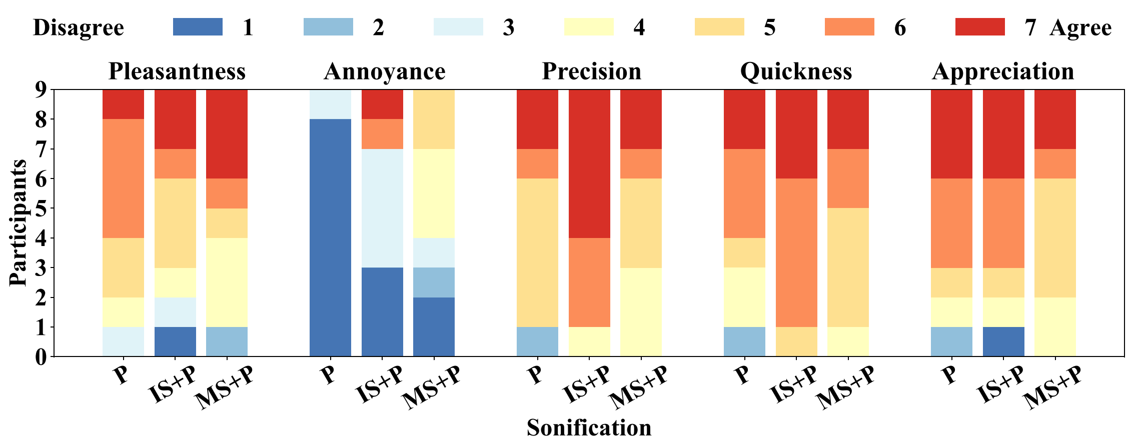


Fig. 5. Answers to the closed questions.

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 NA system, which was based on the idea that using one sonification for both turns and straight paths would result in NA 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.

7.2. Experimental environment

Differently from our prior investigation, in this paper 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, people with BVI 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 people with BVI most commonly travel on familiar routes or study the route beforehand, as shown in Section 5.2. Second, the participants had no environment reference points, and in particular no haptic references. One common practice among people with BVI 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.

7.3. Experimental apparatus

Study results may have also been influenced by the experimental apparatus used. In particular, the effect of the NA 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 NA 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 NA system localization, we acknowledge that the accuracy of these measurements depends on the accuracy of the NA 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 NA system, our design adopted a turn-by-turn navigation schema. While the proposed design is not the only possible one for a NA system, we highlight that our design choices are consistent with prior solutions in the literature (Sato et al., 2019).

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 people with BVI (Mascetti et al., 2016a). 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.

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 work, 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 (Sears and Hanson, 2011).

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).

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 NA 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 NA, 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 provided. A frequent sound 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.

8. Conclusion and future work

This paper proposes two sonification techniques designed to augment navigation instructions in NA systems with the goal of providing more accurate guidance for people with BVI. Experimental results show that both techniques outperform a baseline solution adopted in existing systems (*i.e.*, **P**) during turn instructions, and that one (*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 (Ahmetovic et al., 2016) that adopt the baseline solution.

As future work, we intend to investigate the use of different sonifications for different tasks, for example using **MS+P** during rotations and **IS+P** for straight paths. Also, 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 paper 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 paper 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.

CRediT authorship contribution statement

Dragan Ahmetovic: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision, Project administration. **Federico Avanzini:** Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition. **Adriano Baratè:** Conceptualization. **Cristian Bernareggi:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Marco Ciardullo:** Software, Investigation. **Gabriele Galimberti:** Software, Validation, Formal analysis, Investigation, Visualization, Writing – original draft. **Luca A. Ludovico:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing. **Sergio Mascetti:** Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Supervision, Funding acquisition, Project administration. **Giorgio Presti:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Federico Avanzini reports equipment, drugs, or supplies was provided by MAS Acoustics.

Data availability

Data will be made available on request.

Acknowledgments

This research was partially supported by the Italian Ministry of Foreign Affairs under grant PGR10051. The acoustic-treatment materials employed in the silent room have been kindly provided by MAS Acoustics.

References

Ahmetovic, D., Avanzini, F., Baratè, A., Bernareggi, C., Galimberti, G., Ludovico, L.A., Mascetti, S., Presti, G., 2018a. Sonification of pathways for people with visual impairments. In: International Conference on Computers and Accessibility.

Ahmetovic, D., Avanzini, F., Baratè, A., Bernareggi, C., Galimberti, G., Ludovico, L.A., Mascetti, S., Presti, G., 2019a. Sonification of rotation instructions to support navigation of people with visual impairment. In: International Conference on Pervasive Computing and Communications.

Ahmetovic, D., Bernareggi, C., Keller, K., Mascetti, S., 2021. MusA: artwork accessibility through augmented reality for people with low vision. In: International Web for All Conference.

Ahmetovic, D., Gleason, C., Ruan, C., Kitani, K., Takagi, H., Asakawa, C., 2016. NavCog: A navigational cognitive assistant for the blind. In: International Conference on Human Computer Interaction with Mobile Devices and Services.

Ahmetovic, D., Guerreiro, J., Ohn-Bar, E., Kitani, K.M., Asakawa, C., 2019b. Impact of expertise on interaction preferences for navigation assistance of visually impaired individuals. In: International Cross-Disciplinary Conference on Web Accessibility.

Ahmetovic, D., Manduchi, R., Coughlan, J.M., Mascetti, S., 2017a. Mind your crossings: Mining GIS imagery for crosswalk localization. Trans. Access. Comput..

Ahmetovic, D., Mascetti, S., Bernareggi, C., Guerreiro, J., Oh, U., Asakawa, C., 2019. Deep learning compensation of rotation errors during navigation assistance for people with visual impairments or blindness. Trans. Access. Comput..

Ahmetovic, D., Murata, M., Gleason, C., Brady, E., Takagi, H., Kitani, K., Asakawa, C., 2017b. Achieving practical and accurate indoor navigation for people with visual impairments. In: International Web for All Conference.

Ahmetovic, D., Oh, U., Mascetti, S., Asakawa, C., 2018b. Turn right: Analysis of rotation errors in turn-by-turn navigation for individuals with visual impairments. In: International Conference on Computers and Accessibility.

Allen, D., 1977. Orientation & mobility for persons with low vision. J. Vis. Impair. Blind. 71 (1), 13–15.

Anon, 2023. Microsoft soundscape. URL https://www.microsoft.com/en-us/research/product/soundscape/.

Azenkot, S., Ladner, R.E., Wobbrock, J.O., 2011. Smartphone haptic feedback for nonvisual wayfinding. In: International Conference on Computers and Accessibility.

Bass III, J.B., White supremacy and the major scale: Reframing narratives to teach the blues, improvise, and check in with students. Res. Issues Music Educ. 17 (1), 2.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol..

Brady, E.L., Sato, D., Ruan, C., Takagi, H., Asakawa, C., 2015. Exploring interface design for independent navigation by people with visual impairments. In: International Conference on Computers and Accessibility.

Budrionis, A., Plikynas, D., Daniušis, P., Indrulionis, A., 2022. Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review. Assist. Technol..

Bujacz, M., Kropidlowski, K., Ivanica, G., Moldoveanu, A., Saitis, C., Csapo, A., Wersenyi, G., Spagnol, S., Johannesson, O.I., Unnthorsson, R., et al., 2016. Sound of vision-spatial audio output and sonification approaches. In: International Conference on Computers Helping People with Special Needs.

Bujacz, M., Skulimowski, P., Strumillo, P., 2012. Naviton – A prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired. J. Audio Eng. Soc..

Chrastil, E., Warren, W., 2017. Rotational error in path integration: Encoding and execution errors in angle reproduction. Exp. Brain Res..

Coop, A.D., 2016. Sonification, musification, and synthesis of absolute program music. In: Int. Conf. Auditory Display.

Coughlan, J., Manduchi, R., 2009. Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users. Int. J. Artif. Intell. Tools.

Csapó, Á., Wersényi, G., Nagy, H., Stockman, T., 2015. A survey of assistive technologies and applications for blind users on mobile platforms: A review and foundation for research. J. Multimodal User Interfaces.

Dingler, T., Lindsay, J., Walker, B.N., 2008. Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In: Int. Conf. Auditory Display.

Dubus, G., Bresin, R., 2013. A systematic review of mapping strategies for the sonification of physical quantities. PLoS One.

Dunn, O.J., 1961. Multiple comparisons among means. J. Amer. Statist. Assoc..

Durrant-Whyte, H., Bailey, T., 2006. Simultaneous localization and mapping: Part I. Robot. Autom. Mag..

Emerson, R.W., 2020. Orientation and mobility. In: Movement and Visual Impairment. Routledge, pp. 32–46.

Fenech, E.P., Drews, F.A., Bakdash, J.Z., 2010. The effects of acoustic turn-by-turn navigation on wayfinding. In: Human Factors and Ergonomics Society.

Fletcher, J.F., 1980. Spatial representation in blind children. 1: Development compared to sighted children. J. Vis. Impair. Blind. 74 (10), 381–385.

Flores, G.H., Manduchi, R., 2018. WeAllWalk: An annotated dataset of inertial sensor time series from blind walkers. Trans. Access. Comput.

Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Amer. Statist. Assoc..

Froissard, B., Konik, H., Trémeau, A., Dinet, É., 2014. Contribution of augmented reality solutions to assist visually impaired people in their mobility. In: International Conference on Universal Access in Human-Computer Interaction.
Gaver, W.W., 1989. The SonicFinder: An interface that uses auditory icons. Hum.–Comput. Interact. 4 (1), 67–94.
Geiger, H., Müller, W., et al., 1928. Das Elektronenzählrohr. Physikalische Zeitschrift.
Geronazzo, M., Bedin, A., Brayda, L., Campus, C., Avanzini, F., Interactive spatial sonification for non-visual exploration of virtual maps. Int. J. Hum.-Comput. Stud..
Giudice, N.A., 2018. Navigating without vision: Principles of blind spatial cognition. In: Handbook of Behavioral and Cognitive Geography. Edward Elgar Publishing, pp. 260–288.
Golledge, R.G., Klatzky, R.L., Loomis, J.M., 1996. Cognitive mapping and wayfinding by adults without vision. In: The Construction of Cognitive Maps.
Grond, F., Berger, J., 2011. Parameter mapping sonification. In: The Sonification Handbook. Berlin.
Guerreiro, J., Ahmetovic, D., Kitani, K.M., Asakawa, C., 2017. Virtual navigation for blind people: Building sequential representations of the real-world. In: International Conference on Computers and Accessibility.
Guerreiro, J., Ohn-Bar, E., Ahmetovic, D., Kitani, K., Asakawa, C., 2018. How context and user behavior affect indoor navigation assistance for blind people. In: International Web for All Conference.
Guth, D., LaDuke, R., 1994. The veering tendency of blind pedestrians: An analysis of the problem and literature review. J. Vis. Impair. Blind..
Hakobyan, L., Lumsden, J., O'Sullivan, D., Bartlett, H., 2013. Mobile assistive technologies for the visually impaired. Surv. Ophthalmol.
Hara, K., Azenkot, S., Campbell, M., Bennett, C.L., Le, V., Pannella, S., Moore, R., Minckler, K., Ng, R.H., Froehlich, J.E., 2015. Improving public transit accessibility for blind riders by crowdsourcing bus stop landmark locations with google street view: An extended analysis. Trans. Access. Comput..
Haymes, S., Guest, D., Heyes, A., Johnston, A., 1996. The relationship of vision and psychological variables to the orientation and mobility of visually impaired persons. J. Vis. Impair. Blind. 90 (4), 314–324.
Huang, G., 2019. Visual-inertial navigation: A concise review. In: International Conference on Robotics and Automation.
Johnson, M., 2009. Pop Music Theory. Harmony, Form, and Composition. Lulu Press.
Jürgens, R., Boss, T., Becker, W., 1999. Estimation of self-turning in the dark: Comparison between active and passive rotation. Brain Res..
Kacorri, H., Mascetti, S., Gerino, A., Ahmetovic, D., Alampi, V., Takagi, H., Asakawa, C., 2018. Insights on assistive orientation and mobility of people with visual impairment based on large-scale longitudinal data. Trans. Access. Comput..
Kacorri, H., Mascetti, S., Gerino, A., Ahmetovic, D., Takagi, H., Asakawa, C., 2016. Supporting orientation of people with visual impairment: Analysis of large scale usage data. In: International Conference on Computers and Accessibility.
Katz, B.F.G., Kammoun, S., Parseihian, G., et al., 2012. NAVIG: Augmented reality guidance system for the visually impaired. Virtual Reality.
Katz, B.F., Truillet, P., Thorpe, S., Jouffrais, C., 2010. NAVIG: Navigation assisted by artificial vision and GNSS. In: Workshop on Multimodal Location Based Techniques for Extreme Navigation, Finland.
Klatzky, R.L., Marston, J.R., Giudice, N.A., Golledge, R.G., Loomis, J.M., 2006. Cognitive load of navigating without vision when guided by virtual sound versus spatial language. J. Exp. Psychol.: Appl. 12 (4), 223.
Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J.H., Miner, N., Neuhoff, J., 1999. The Sonification Report: Status of the Field and Research Agenda. Tech. Rep., National Science Foundation.
Kruskal, W.H., Wallis, W.A., 1952. Use of ranks in one-criterion variance analysis. J. Amer. Statist. Assoc..
Kunhoth, J., Karkar, A., Al-Maadeed, S., Al-Attiyah, A., 2019. Comparative analysis of computer-vision and BLE technology based indoor navigation systems for people with visual impairments. Int. J. Health Geogr..
Lahav, O., Mioduser, D., 2008. Construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind. Comp. Hum. Behav..
Lahav, O., Schloerb, D.W., Kumar, S., et al., 2012. A virtual environment for people who are blind - A usability study. J. Assist. Technol..
Lewis, J.R., 1993. Multipoint scales: Mean and median differences and observed significance levels. Int. J. Hum.-Comput. Interact..
Lo Valvo, A., Croce, D., Garlisi, D., Giuliano, F., Giarré, L., Tinnirello, I., 2021. A navigation and augmented reality system for visually impaired people. Sensors.
Loomis, J.M., Golledge, R.G., Klatzky, R.L., 1998. Navigation system for the blind: Auditory display modes and guidance. Presence 7 (2), 193–203.
Loomis, J.M., Marston, J.R., Golledge, R.G., et al., 2005. Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance. J. Vis. Impair. Blind..
Ludovico, L.A., Presti, G., 2016. The sonification space: A reference system for sonification tasks. Int. J. Hum.-Comput. Stud. 85, 72–77. http://dx.doi.org/10.1016/j.ijhcs.2015.08.008.
Magnusson, C., Danielsson, H., Rassmus-Gröhn, K., 2006. Non visual haptic audio tools for virtual environments. In: Haptic and Audio Int. Design.
Maidi, M., Didier, J.-Y., Ababsa, F., Mallem, M., 2010. A performance study for camera pose estimation using visual marker based tracking. Mach. Vis. Appl..
Manduchi, R., Kurniawan, S., 2011. Mobility-related accidents experienced by people with visual impairment. Res. Pract. Vis. Impairment Blindness.
Mann, H.B., Whitney, D.R., 1947. On a test of whether one of two random variables is stochastically larger than the other. Ann. Math. Stat..
Marlinsky, V., 1999. Vestibular and vestibulo-proprioceptive perception of motion in the horizontal plane in blindfolded man. Neuroscience.
Mascetti, S., Ahmetovic, D., Gerino, A., Bernareggi, C., 2016a. Robust traffic lights detection on mobile devices for pedestrians with visual impairment. Comput. Vis. Image Underst..
Mascetti, S., Ahmetovic, D., Gerino, A., Bernareggi, C., 2016b. Zebrarecognizer: Pedestrian crossing recognition for people with visual impairment or blindness. Pattern Recognit..
Mascetti, S., Picinali, L., Gerino, A., Ahmetovic, D., Bernareggi, C., 2016c. Sonification of guidance data during road crossing for people with visual impairments or blindness. Int. J. Hum.-Comput. Stud..
Medina-Sanchez, E.H., Mikusova, M., Callejas-Cuervo, M., 2021. An interactive model based on a mobile application and augmented reality as a tool to support safe and efficient mobility of people with visual limitations in sustainable urban environments. Sustainability.
Meijer, P.B., 1992. An experimental system for auditory image representations. Trans. Biomed. Eng..
Murata, M., Ahmetovic, D., Sato, D., Takagi, H., Kitani, K.M., Asakawa, C., 2018. Smartphone-based indoor localization for blind navigation across building complexes. In: International Conference on Pervasive Computing and Communications.
Murata, M., Ahmetovic, D., Sato, D., Takagi, H., Kitani, K.M., Asakawa, C., 2019. Smartphone-based localization for blind navigation in building-scale indoor environments. Pervasive Mob. Comput..
Nair, V., Olmschenk, G., Seiple, W.H., Zhu, Z., 2020. ASSIST: Evaluating the usability and performance of an indoor navigation assistant for blind and visually impaired people. Assist. Technol..
Nakajima, M., Haruyama, S., 2012. Indoor navigation system for visually impaired people using visible light communication and compensated geomagnetic sensing. In: Communications in China.
Nemenyi, P.B., 1963. Distribution-Free Multiple Comparisons.
Ohn-Bar, E., Guerreiro, J., Ahmetovic, D., Kitani, K.M., Asakawa, C., 2018. Modeling expertise in assistive navigation interfaces for blind people. In: International Conference on Intelligent User Interfaces.
Periša, M., Cvitić, I., Sente, R.E., 2017. Comparative analysis of mobile phone application solutions accessibility for informing visually impaired persons in traffic environment. In: Services for Mobility and Mobility As a Service.
Presti, G., Ahmetovic, D., Ducci, M., Bernareggi, C., Ludovico, L.A., Baratè, A., Avanzini, F., Mascetti, S., 2021. Iterative design of sonification techniques to support people with visual impairments in obstacle avoidance. ACM Trans. Access. Comput. 14 (4), 1–27. http://dx.doi.org/10.1145/3470649.
Rajamäki, J., Viinikainen, P., Tuomisto, J., Sederholm, T., Säämänen, M., 2007. Laureapop indoor navigation service for the visually impaired in a WLAN environment. In: International Conference on Electronics, Hardware, Wireless and Optical Communications.
Sadalla, E.K., Montello, D.R., 1989. Remembering changes in direction. Environ. Behav..
Sato, D., Oh, U., Guerreiro, J., Ahmetovic, D., Naito, K., Takagi, H., Kitani, K.M., Asakawa, C., 2019. NavCog3 in the wild: Large-scale blind indoor navigation assistant with semantic features. Trans. Access. Comput..
Scaletti, C., Craig, A.B., 1991. Using sound to extract meaning from complex data. In: Extracting Meaning from Complex Data: Processing, Display, Interaction II, Vol. 1459. SPIE, pp. 207–219.
Schinazi, V.R., Thrash, T., Chebat, D.-R., 2016. Spatial navigation by congenitally blind individuals. WIREs Cogn. Sci..
Sears, A., Hanson, V., 2011. Representing users in accessibility research. In: Conference on Human Factors in Computing Systems.
See, A.R., Sasing, B.G., Advincula, W.D., 2022. A smartphone-based mobility assistant using depth imaging for visually impaired and blind. Appl. Sci..
Shoval, S., Borenstein, J., Koren, Y., 1998. Auditory guidance with the navbelt-A computerized travel aid for the blind. Trans. Syst. Man Cybern..
Siegel, A.W., White, S.H., 1975. The development of spatial representations of large-scale environments. Adv. Child Dev. Behav. 10, 9–55.
Sobnath, D., Rehman, I.U., Nasralla, M.M., 2020. Smart cities to improve mobility and quality of life of the visually impaired. Technol. Trends Improved Mobil. Visual. Impair..
Spagnol, S., Wersényi, G., Bujacz, M., et al., 2018. Current use and future perspectives of spatial audio technologies in electronic travel aids. Wirel. Commun. Mob. Comput..
Strait, D.L., Kraus, N., Parbery-Clark, A., Ashley, R., 2010. Musical experience shapes top-down auditory mechanisms: Evidence from masking and auditory attention performance. Hear. Res..
Thinus-Blanc, C., Gaunet, F., 1997. Representation of space in blind persons: Vision as a spatial sense? Psychol. Bull. 121 (1), 20.
Trehub, S.E., 2001. Musical predispositions in infancy. Ann. New York Acad. Sci. 930 (1), 1–16.
Ungar, S., 2000. Cognitive mapping without visual experience. In: Cognitive Mapping: Past, Present, and Future.

Usoh, M., Arthur, K., Whitton, M.C., et al., 1999. Walking > walking-in-place > flying, in virtual environments. In: ACM SIGGRAPH.
Walker, B.N., Lindsay, J., 2006. Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice. Hum. Factors.
Walker, B.N., Nees, M.A., 2011a. Theory of sonification. In: The Sonification Handbook.
Walker, B.N., Nees, M.A., 2011b. Theory of sonification. In: The Sonification Handbook.
Wiener, J.M., Büchner, S.J., Hölscher, C., 2009. Taxonomy of human wayfinding tasks: A knowledge-based approach. Spatial Cogn. Comput..
Wiener, W.R., Welsh, R.L., Blasch, B.B., 2010. Foundations of Orientation and Mobility.
Zacharias, G., Young, L., 1981. Influence of combined visual and vestibular cues on human perception and control of horizontal rotation. Exp. Brain Res..