Iterative Design of Sonification Techniques to Support People with Visual Impairments in Obstacle Avoidance

GIORGIO PRESTI, DRAGAN AHMETOVIC, MATTIA DUCCI, CRISTIAN BERNAREGGI,
LUCA A. LUDOVICO, ADRIANO BARATÈ, FEDERICO AVANZINI, and
SERGIO MASCETTI, Department of Computer Science, University of Milan

Obstacle avoidance is a major challenge during independent mobility for blind or visually impaired (BVI) people. Typically, BVI people can only perceive obstacles at a short distance (about 1 m, in case they are using the white cane), and some obstacles are hard to detect (e.g., those elevated from the ground), or should not be hit by the white cane (e.g., a standing person). A solution to these problems can be found in recent computer-vision techniques that can run on mobile and wearable devices to detect obstacles at a distance. However, in addition to detecting obstacles, it is also necessary to convey information about them in real time.

This contribution presents *WatchOut*, a sonification technique for conveying real-time information about the main properties of an obstacle to a BVI person, who can then use this additional feedback to safely navigate in the environment. *WatchOut* was designed with a user-centered approach, involving four iterations of online listening tests with BVI participants in order to define, improve and evaluate the sonification technique, eventually obtaining an almost perfect recognition accuracy. *WatchOut* was also implemented and tested as a module of a mobile app that detects obstacles using state-of-the-art computer vision technology. Results show that the system is considered usable and can guide the users to avoid more than 85% of the obstacles.

CCS Concepts: • **Human-centered computing** → **Empirical studies in accessibility**; • **Social and professional topics** → **Assistive technologies**; • **Information systems** → *Geographic information systems;*

Additional Key Words and Phrases: Turn-by-turn navigation, orientation & mobility, navigation assistance

ACM Reference format:
Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca A. Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. 2021. Iterative Design of Sonification Techniques to Support People with Visual Impairments in Obstacle Avoidance. *ACM Trans. Access. Comput.* 14, 4, Article 19 (October 2021), 27 pages.
https://doi.org/10.1145/3470649

1 INTRODUCTION

There are about 285 million **blind or visually impaired** (**BVI**) people worldwide [59]. For most of them, **Orientation and Mobility** (**O&M**) is a challenge experienced in the everyday life. Issues range from environmental orientation to perception, involving psychosocial, educational

---

Authors' address: G. Presti, D. Ahmetovic, M. Ducci, C. Bernareggi, L. A. Ludovico, A. Baratè, F. Avanzini, and S. Mascetti, Department of Computer Science, University of Milan, Via Celoria 18, Milan 20133, Italy; emails: {giorgio.presti, dragan.ahmetovic}@unimi.it, mattia.ducci@studenti.unimi.it, {cristian.bernareggi, luca.ludovico, adriano.barate, federico.avanzini, sergio.mascetti}@unimi.it.

1936-7228/2021/10-ART19 $15.00
https://doi.org/10.1145/3470649

and administrative dimensions. In particular, one problem is the limited ability to determine objects location and characteristics [81]. Overcoming these issues often requires specific training to learn general O&M strategies and to learn to safely travel few selected routes autonomously.

A number of assistive technologies have been designed to improve O&M performance of BVI people during independent travel. Various comprehensive reviews of assistive devices for BVI people have been published over the last 15 years, addressing not only technical achievements but also current challenges and future perspectives [25, 29, 64, 75]. The noticeable evolution from early electronic travel aids to latest systems based on artificial vision, mixed reality, wearables with built-in cameras, and so on has emphasized the importance of user-centered design, in terms of knowledge of the users' needs, capabilities, and limitations. This is also the approach followed in our work.

When discussing applications for O&M performance, a distinction should be made between orientation and mobility: the former refers the ability to use one's remaining senses to understand one's location in the environment at any given time, while the latter refers to the capacity of facility of movement [28]. Accordingly, applications focusing on orientation address such problems as on-line wayfinding [32, 65], or off-line exploration of maps [27, 45], both in indoor [68] and outdoor [38] contexts. On the other hand, applications focusing on mobility address such problems as detection of hazards and obstacle avoidance along the path. In this contribution, we focus on the latter problem.

As an example, a BVI person moving towards a bicycle parked on a sidewalk would need to detect its presence, assess its size and map its relative directional position and distance, in order to walk around it. This is usually achieved using a combination of senses including: residual vision, if any; hearing, in case the object itself produces or reflects sound; touch, generally through the use of a white cane. However, this is still a challenging task, as many BVI people cannot rely on vision, at least in some situations, for example with low luminosity or strong glare. Furthermore, if an object is not producing any sound, reflects the sound poorly, or there is a strong ambient noise, it is also difficult to rely on the sense of hearing for the task of obstacle avoidance. Similarly, the sense of touch cannot always be reliably used. For example, the white cane only detects objects touching the ground, and within a range of about 1 m. It can also hit or trip bystanders, damage objects (e.g., a bicycle that can fall when hit) or get stuck or tangled. It also requires some time to assess obstacle characteristics by touch to correctly avoid them. In this contribution we focus on white cane users, whose number is much larger than that of individuals who rely on guide dogs.[1] Our solution, as currently designed, would probably be of marginal interest for BVI individuals with a guide dog, because guide dogs can effectively assist in avoiding obstacles. However, our solution is more scalable and affordable.[2]

As discussed in the next section, many assistive technologies have been proposed to support BVI people in detecting obstacles. Some are based on dedicated hardware, others run on mainstream mobile and wearable devices, and most focus on the problem of solely detecting the obstacle. However, while this is certainly a key problem, there is an orthogonal fundamental issue: how to convey salient obstacle properties to the user in real time, without requiring long training and with limited cognitive workload. To tackle this challenge, this contribution presents *WatchOut*, a sonification technique to convey obstacle properties in real time. *WatchOut* aims at being effective also with limited training times. For this investigation we adopted a user-centered design process which involved a series of iterations and online listening experiments with BVI people, for a total of 61

---

[1] While we are not aware of any official statistic about the number of guide dogs, unofficial estimate reports that only 2% of BVI individuals in the USA have a guide dog (https://www.guidingeyes.org/about/faqs).

[2] The overall costs for a guide dog are in the order of 50, 000 USD (https://www.guidingeyes.org/about/faqs).

participants in 4 iterations. Results show that in the first 3 iterations we have been able to progressively increase the effectiveness of the sonification technique with equal conditions, which are mainly affected by the choice of conveying four obstacle properties to the user. During the fourth iteration, we instead decided to experiment with the sonification of three obstacle properties, and this led to an almost perfect recognition rate by the participants.

This contribution extends our previous conference paper [62] along with the following directions:

- We introduce a new sonification mapping design, that improves over the two previously presented designs, while still providing the same information about the obstacle (distance, horizontal position, height and width);
- We present a reduced sonification design that reports only three obstacle properties (distance, horizontal position and height), but provides a more accurate understanding of the obstacle properties;
- We perform a comparative analysis of the accuracy metrics and participants' subjective feedback across the four iterations;
- We discuss the findings resulting from the user studies and we present generalizable knowledge that could be useful for researchers and practitioners when designing similar, sonification-based interactions.

2 RELATED WORK

The topic of obstacle avoidance systems for BVI people has been mainly researched from two perspectives: the investigation of obstacle detection systems, and the design of non-visual guidance paradigms aimed at avoiding detected obstacles.

2.1 Obstacle Detection

Obstacle detection based on various types of dedicated hardware, such as ultrasonic, infrared, or laser range finder sensors to detect obstacles around the user, have been proposed by several prior works. These sensors are very quick and accurate, and they can be handheld [83], wearable [35], or mounted on the white cane [40], eyeglasses [34, 71], or even on gloves [12]. Cameras and computer vision algorithms (mounted on eyeglasses [8, 66] or on a white cane [76]) may represent a valid alternative to recognize obstacles in front of the user in the range of up to 10 meters [48]. Depth sensors have also been used for a more accurate obstacle localization [13, 44, 61].

Smartphone cameras [74] and depth sensors [44] have also been used in recent years for the same goal. Such devices present many advantages, such as limited costs, maintainability, a single interaction paradigm, and higher acceptance by the users [70]. In addition, if compared with range finders, a larger number of obstacle features can be extracted from an image, including color and other non-tangible information, such as symbols on road signs [51, 52].

Among the various solutions based on mobile devices, some use, as in our approach, both depth sensors and video camera feed [36, 43]. The difference is that our solution focuses on how to use sonification techniques to inform the user about the presence of an obstacle. Instead, the work of Jafri et al. [37] focuses on the object recognition only, while Li et al. [43] propose a solution based on speech.

Table 1 summarizes obstacle detection approaches adopted by prior works and the proposed solution.

2.2 Non-Visual Guidance Paradigms for Obstacle Avoidance

Non-visual guidance for obstacle avoidance can be achieved using: haptic and tactile feedback, speech messages, or sonification. Prior works propose haptic and tactile feedback on fingers

Table 1. Obstacle Detection Approaches in the Proposed Approach and Prior Literature

Approach	**Works**
Embedded ultrasonic sensor devices	[22, 24, 34, 40, 71]
Embedded infrared transmitter/sensor coupling devices	[6, 39, 58, 63]
Embedded laser range finder devices	[19, 83]
Embedded devices using video camera feed	[8, 56, 66, 76]
Embedded devices using depth sensors	[13, 44, 60, 61]
Mobile devices using video camera feed	[3, 49, 52, 74]
Mobile devices using depth sensors	[33, 37, 57]
Mobile devices using depth sensors and video camera feed	**This contribution**, [36, 43]

through gloves with piezoelectric actuators [12, 84], on the chest [17] or abdomen [78] through 2-d vibration array vests, on the arm through an extensible wire [35] or through transcutaneous electrical nerve stimulation [53]. While these techniques provide real time feedback and enable BVI people to form a mental map of the position and size of the obstacle, the required hardware and equipment is often perceived as unnatural or invasive [35, 53, 84]. In particular, gloves prevent hands-free mobility, which is essential for BVI people using a white cane. An additional issue concerns the available tactile display parameters. Actuator locations and temporal patterns are reliable parameters, thanks to the high spatial and temporal acuity of the tactile sense, while vibration intensity and frequency are of limited use (as they are affected by actuator differences, attachment method, and so on), and higher order spatio-temporal or intensity-temporal patterns are less explored [26]. Moreover, previous results showed that tactile parameters that are reliable in isolation may not be so in "compound" signals, which may lead to high perceptual and/or cognitive processing demands [78].

Another paradigm relies on speech messages to guide the user around obstacles [8, 18, 20, 44, 82]. Although speech instructions are informative and easy to learn, they require more time to inform the user, who risks bumping into an obstacle before the complete alert message is spoken. Moreover, speech distracts the user from the surrounding environment because decoding a verbal message requires a significant cognitive effort [82].

Finally, sonification approaches can be used. The term "sonification", as defined by Walker et al., is "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation" [79]. Representing an auditory counterpart of the term "visualization", it broadly refers to the use of non-verbal sound to convey information.

Sonification can be advantageous over speech messages in terms of robustness to background noise, reduced cognitive load, and linguistic differences [21], possibly at the expense of longer training. Moreover, the availability of a large number of reliable auditory dimensions permits in principle to simultaneously convey multiple features of the visual scene to the listener, thus overcoming the limitations of haptic and tactile feedback. Finally, sonification approaches can rely on general-purpose and non-invasive hardware for the display, such as the mobile phone speaker, or bone-conducting headphones.

2.3 Sonification Approaches

Various taxonomic descriptions of sonification have been proposed, in terms of both functions and techniques [79]. Specifically, commonly employed techniques include audification, auditory icons, earcons, model-based, and parameter mapping sonification. The latter [31] is used in this work, and amounts to representing changes in some data dimension with changes in an auditory dimension. Thus, a distinguishing feature of this type of sonification is the *mapping function* between data

Table 2. Systems for Obstacle Avoidance Based on Sonification (in Order of Publication Date)

Work	**Obstacle properties**	**Auditory dimensions**
Ifukube et al. [34]	Position, Size	Spatial (binaural), Loudness-related
Meijer [54]	Position, Other (brightness)	Timbral, Temporal, Loudness-related
Shoval et al. [71]	Distance, Position	Spatial (binaural), Loudness-related
González-Mora et al. [30]	Distance, Position	Spatial (binaural), Loudness-related
Aguerrevere et al. [1]	Distance, Position	Spatial (binaural), Loudness-related
Sainarayanan et al. [66]	Position, Other (brightness)	Timbral, Temporal, Loudness-related
Striem-Amit et al. [73]	Position, Other (brightness)	Timbral, Temporal, Loudness-related
Metsiritrakul et al. [55]	Distance, Position, Size	Pitch-related, Temporal (duration), Spatial (binaural)
This contribution	Distance, Position, Size (height, width)	Pitch-related, Temporal (intermittence), Spatial (stereo panning, reverberation), Timbral, Musical (polyphony)

Sonified object properties include Distance, (angular) Position, Size (height and/or width), Other. Auditory dimensions include the five high-level categories discussed in Section 2.3.

and auditory dimensions. In particular, in order to qualify as sonification (i.e., to effectively convey information about the data), such mapping should not be arbitrary nor excessively complex.

Auditory dimensions can be clustered into five high-level categories [23]: Pitch-related, Timbral, Loudness-related, Spatial, and Temporal. Higher-level musical features may also be used, such as tonality and polyphony: in this case the term “musification” is often used, to refer to a musical representation of data [16]. As an example, the present authors have recently employed a musification approach to sonify angular movements in the task of accurately rotating towards a direction [2], which is a critical issue during assisted navigation [4]. In their extensive review of mapping strategies for the sonification of physical quantities, Dubus and Bresin [23] analyzed the frequency of use of physical and auditory dimensions and showed that pitch is by far the most used auditory dimension. Moreover, they revealed that spatial auditory dimensions are almost exclusively used to sonify kinematic quantities such as distance, orientation, velocity, and so on.

Several works have explored the sonification of features extracted from a visual scene [67], with the goal of aiding obstacle avoidance by a BVI person. A historical example is “The vOICe” vision technology [54], which sonifies successive frames captured by a camera by interpreting the corresponding pictures as spectrograms (i.e., mapping the vertical position of each pixel to frequency, the horizontal one to time, and brightness to loudness). As a result, the amount of sonified data are enormous (the complete visual image facing the user, regardless of salience evaluations), and a long training is consequently required: about 8 months were needed to five blind persons to learn to walk along paths and properly avoid objects [73].

In order to reduce cognitive load and ease the learning process, more recent works sonify only some features of detected obstacles, through specific auditory dimensions. Spatial dimensions are often used, ranging from simple lateralization (i.e., panning of stereo sounds) to immersive 3D sound (see two extensive reviews on the uses of spatial audio for electronic travel aids at large [46, 72]).

Table 2 gives a schematic view of various systems. Even though it does not provide an exhaustive review of previous research, it does show some common trends and approaches. In particular, it can be noticed that all of these works convey the angular position of an obstacle, mostly through spatial or temporal dimensions, where the displayed positions are quantized over a discrete set of possibilities [34, 71], or vary continuously according to the visual scene [1]. Distance is the second most sonified obstacle property, typically through loudness-related auditory dimensions [1, 30, 53, 71]. Size-related obstacle properties are more rarely sonified, through loudness-related [34] or even pitch-related and temporal [55] dimensions.

Table 2 also shows that only one of the reviewed systems [55] conveys synchronously all the obstacle properties needed to avoid it: distance, size, and angular position. Out *WatchOut* system also sonifies all these properties, albeit using a different approach for mapping them into auditory dimensions.

Moreover, the studies cited above evaluated either with blind-folded people or in a simulated environment with up to five BVI people. Instead, our sonification technique is evaluated with a total of 61 participants, and its application in a navigation app is also experimented in the real-world with 13 BVI participants [62].

3 THE *WATCHOUT* SYSTEM

Giudice [28] argues that the greatest challenges in O&M for BVI people are about insufficient information access rather than vision loss *per se*, and that applications that make the right spatial information available through other senses can ultimately lead to equivalent performance between sighted and BVI individuals. The goal of our research is to overcome insufficient access to navigation-critical information by developing a mobile system capable of detecting obstacles and conveying their properties to BVI users through an effective sonification strategy. The choice of sonification over alternative approaches (tactile feedback or speech messages) is based on the discussion presented in Section 2.2.

While the main contribution of this article is the iterative design of the sonification strategy presented in Section 4, here we describe the main development phases of the mobile app, and specifically its analysis, design, implementation, and evaluation as an iOS prototype (a more detailed description of the mobile application and its modules is available in our previous article [62]). Specifically, we describe two main components: the obstacle detection module, which visually identifies obstacles framed by the mobile device camera in augmented reality, and the sonification engine, which was first used during the design of sonification strategies, and then ported to the mobile system for real-time sonification generation.

3.1 Analysis and Design

The initial analysis conducted by the team (including a BVI person) revealed two contrasting needs:

- **completeness:** the system should provide detailed obstacle information, possibly regarding multiple obstacles.
- **understandability:** the obstacle information should be conveyed with a technique that is robust to ambient noise, easy to learn, with low cognitive load, and that does not distract the user from ambient noise (e.g., an incoming car).

To balance between these needs, we defined two requirements. First, the system should identify the most relevant obstacle (see Section 3.2 for details) and convey it to the user. Hence, if two or more obstacles are identified at the same time, the user will be informed about one of them only. Second, as previously observed in literature [85], the system should convey only a limited set of relevant obstacle properties. Based on the existing literature, we chose the following obstacle properties to be identified and sonified: the frontal *distance* from the user [30, 78], the horizontal *position* with respect to the user (i.e., in front of the user, on the left/right) [1, 34, 71], and the obstacle *size* [55, 66]. Regarding the latter property (i.e., size), we first decided to favor the need of completeness by providing both the obstacle *width* and *height*. Instead, during the fourth iteration of the sonification design (see Section 4) we decided to focus on the need of understandability by dropping the *width* dimension. The choice of dropping the *width* property is due to the fact that, while it can be useful to figure out what the obstacle is, this information is less useful than the others to guide the user to avoid the obstacle.

One strategic decision involved the granularity of the information provided to the subject. The choice of discrete or continuous data dimensions, and the implications for the design process, have been discussed in the sonification literature [31]. As an example, for the *distance* dimension, one may evaluate and sonify the distance of the obstacle from the user as a continuous time-dependent signal or may simply provide a binary near/far information when the obstacle is within predefined ranges. Similar considerations apply to the remaining data dimensions.

In order to limit cognitive overload, we chose the former strategy for *distance*, and the latter strategy for *position* and *size*. Treating distance as a continuous dimension, in particular, permits to convey information related to the relative velocity of the obstacle with respect to the subject: although obstacle velocity is not explicitly sonified, the rate of change of the sonification for distance provides this information implicitly. The threshold values are partially based on the work by van Erp et al. [78]. The obstacle properties initially selected for the sonification are:

- *Distance* – Continuous, from 0.1 m (close) to 3 m (far). The lower threshold is needed to avoid detecting the user, and the upper one doubles the range of the white cane, while preventing far objects from causing cognitive overload;
- *Position* – Discrete (left, center, and right), with an obstacle classified as being on the right if its leftmost part is at least 25 cm to the right of the median plane of the camera reference system (analogous for left classification), in order to create an easily walkable 0.5 m corridor;
- *Width* – Discrete (narrow, wide), where narrow obstacles easily fit into the walkable corridor (i.e., up to 35 cm), and large ones require a bigger trajectory correction;
- *Height* – Discrete (walkable, to circumvent), where obstacles up to 25 cm of height are walkable (the user can step over them, e.g., a sidewalk), and those above are to circumvent.

As a consequence of the two contrasting needs discussed above, we require our solution to be used in the real world with bone-conducting headphones. These have the advantage of minimizing invasiveness, since the ear canal is not occluded and thus the environment can be heard. At the same time, they deliver stereo signals to the user, thus permitting to exploit additional spatial auditory dimensions in the sonification (particularly, panning of the stereo signal), and to provide more detailed information about obstacles.

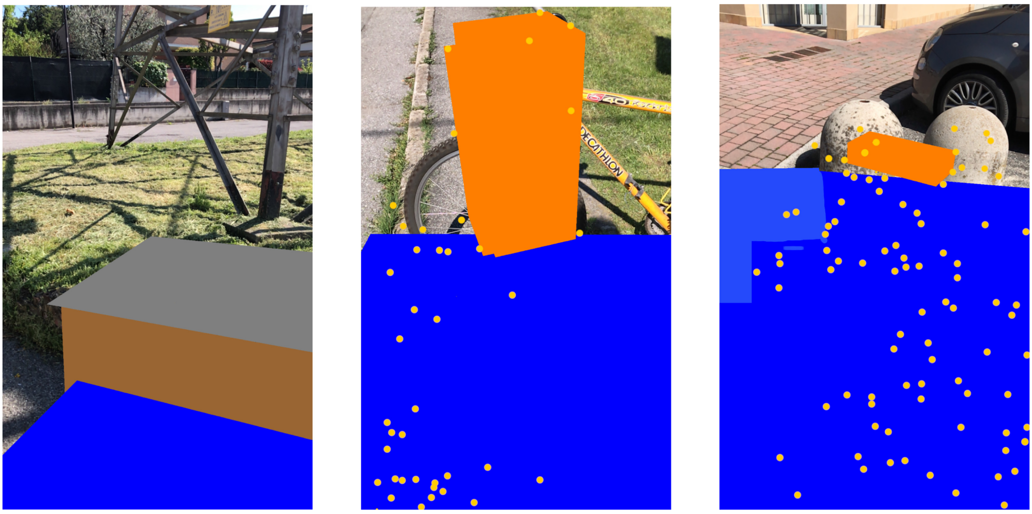
3.2 Obstacle Detection

The obstacle detection module uses Apple's ARKit framework[3] to identify and localize the obstacles in 3D with respect to the user. This framework exposes a number of high level computer vision primitives that can be used to reconstruct the 3D structure of the surrounding environment from the 2D images sourced from the device camera. In particular, it identifies a set of salient 3D feature points, called point cloud, belonging to the objects and the environment, and it is able to detect planes (walls, ground) from the co-planar points identified in the point cloud. The lower horizontal plane is considered the ground plane on which the user is walking, while the other planes are considered as obstacles, as shown in Figure 1(a).

Other obstacles, which are not detected as planes, are identified using the point cloud information (see Figure 1(b)). The detection module first clusters feature points in each frame, based on

---

[3] https://developer.apple.com/arkit/.



(a) Detected planes: ground (blue), a higher horizontal plane (gray) and a vertical plane (brown)

(b) A bicycle detected as a candidate obstacle (note that ground plane features points are pruned)

(c) Two adjacent objects are recognized as a single one

Fig. 1. Examples for the obstacle detection procedure.

their distance (a threshold of 0.5 m is used), to form *candidate objects*. Even if aggregated points actually belong to distinct objects (like in Figure 1(c)), such objects have to be very close to each other. Therefore, the user would not be able to walk between them, which makes it convenient to recognize them as a single obstacle. Candidate objects that overlap across different frames (2 in our experimental setup), in a given temporal window (0.75 s in our setup), are considered valid obstacles. The recognizer module identifies the most relevant obstacle among those detected in a single temporal window. The system prioritizes frontal obstacles, and among these, the ones closest to the user. Similarly, if only lateral obstacles are present, the closest one to the user is returned. The most relevant obstacle is then sonified using the sonification engine.

3.3 Sonification Engine

Besides temporal granularity (see Section 3.1), sonification approaches can also be classified in terms of their "level of abstraction" [47]: real-world sounds are at a symbolic level of abstraction, whereas synthesized sounds are typically far from our experience and cannot be immediately associated with a symbolical meaning. In our case, the system should limit the overlap between real-world sounds coming from the environment and the sonification added by *WatchOut*. This led us to reject straightforward auditory icons for the detected obstacles, e.g., a bicycle ringing bell to signal the presence of a bike. Instead, the base sound for the sonification was chosen to be a sine wave coupled with a percussive initial transient (i.e., a filtered impulse sound), with the sine wave decaying exponentially in about 1 second.[4] This is perceived as a synthetic sound, since this kind of timbre is rare in nature (some harmonics are always present in physical phenomena).

In order to improve the usability of the application in real-world conditions, two additional sound processing steps were included:

---

[4] Examples for all sounds presented in this contribution are available https://watchoutobstacles.netlify.com/.

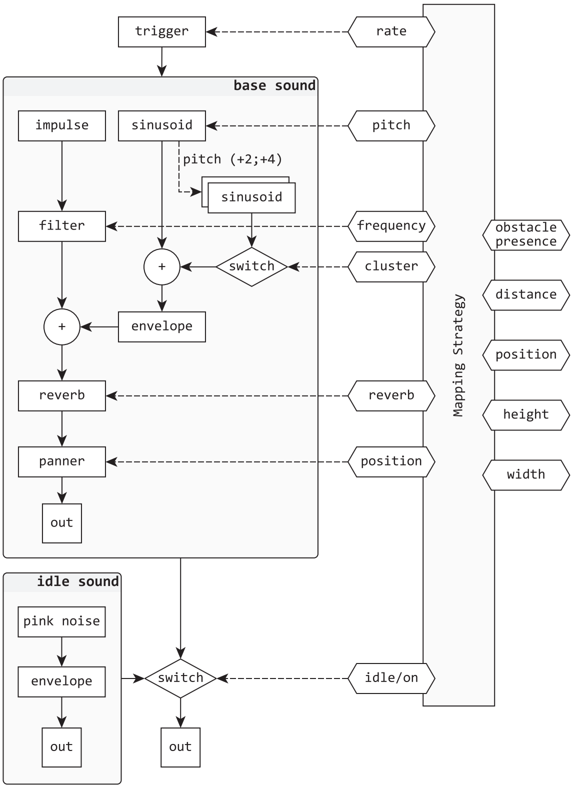


Fig. 2. Block diagram of the sonification engine. Rectangular boxes are audio modules generating/modifying sound; hexagons are controllers; solid lines represent audio data; dashed lines represent control signals.

- to signal the user that the sonification is active but no obstacles are in sight, an idle sound was added: a quiet pink noise[5] modulated by a 1second fade in and out, interleaved by 5 seconds of silence. This sound is replaced by the proper sonification when an obstacle is in sight;
- since the sonification is going to be heard via bone-conducting headphones, we boost the volume of the low pitched sound by 12 dB to compensate the frequency response of the device.

Figure 2 shows a high-level block diagram of the sonification engine. Hexagons are both the main properties extracted from the visual scene and synthesizer engine controllers. These are connected by a matrix according to the desired mapping strategy. Audio modules (rectangles) can generate (*sinusoid*, *impulse*, and *pink noise*) or modify (*filter*, *panner*, *reverb*, and *volume envelope*) sound. Finally, signals are routed to the audio output thanks to adders and switches (circles and diamonds).

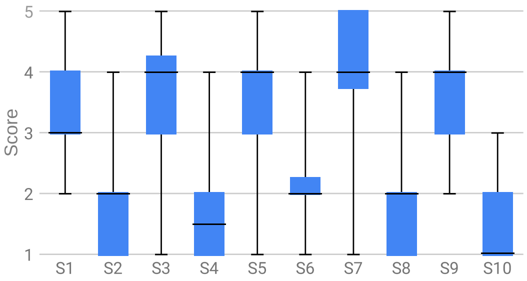
The sonification engine implementation, used to create samples for online listening tests, was created in Max.[6] In the mobile obstacle detection app prototype, described in [62], the port of the sonification engine was created with the AudioKit framework.[7]

---

[5] A broadband noise whose power is inversely proportional to the frequency, reminding the sound of a distant waterfall.
[6] https://cycling74.com/products/max.
[7] https://audiokit.io/.

S1	I think that I would like to use this system frequently
S2	I found the system unnecessarily complex
S3	I thought the system was easy to use
S4	I think I would need support of a technical person to use the system
S5	I found the various functions in this system were well integrated
S6	I thought there was too much inconsistency in this system
S7	I imagine that most people would learn to use this system very quickly
S8	I found the system very cumbersome to use
S9	I felt very confident using the system
S10	I needed to learn a lot of things before I could get going with the system

(a) SUS Questions.



(b) SUS Scores.

Fig. 3. SUS questions and scores for the system.

3.4 Usability Evaluation

The usability of the *WatchOut* mobile system was assessed in the real world through a user study. Full details about the study are presented in our previous conference paper [62], while this section summarizes it for the sake of completeness.

The study consisted of a series of navigation tasks on predefined outdoor paths. All paths shared an overall length of around 500 m, on plain terrain, and included obstacles unknown to the participant, such as curbs, traffic bollards, parked cars, walls on the side and architectural barriers). The study was exclusively aimed at evaluating the participants' opinion on the functionality and usability of the system. Participants' subjective feedback was collected through Likert-like scale questions including **System Usability Scale** (**SUS**) questionnaire [14] (see Table 3(a)).

A total of 13 BVI participants (7 male and 6 female) was recruited. In a preliminary training phase (lasting 5 minutes approximately), participants explored 4 predefined obstacles (a wall, a traffic bollard, a person, and stairs of at least 3 steps), first being notified by the experimenter of the obstacle type, and then exploring obstacles autonomously through the system. Afterwards, the actual test was started on the predefined path. The experimenter followed participants, annotating their performance and possibly aiding them in case of need. The test lasted until the end of the path or for 15 minutes. After the test, participants filled in the final questionnaire, which included SUS evaluation of the obstacle detection system, Likert-like scale questions measuring the participants' opinion on the sonification, and open ended questions addressing participants' opinions and suggestions.

Figure 3 reports the SUS scores collected from the participants. The average SUS score for all the participants was 72.5 ± 13.23, indicating a “Good” usability score [11]. Details on the statistical analysis can be found in [62]. The participants' observations unveiled important obstacle types, possible reinforcement cues such as vibration feedback, and the user requirements with respect to the interaction preferences with the system. While such results are not directly applicable to our sonification feedback design, they point to the need of improving especially the hardware setup for the recognition of specific obstacle types. For example, external cameras or phone holders could be used to be able to view head level obstacles. Additionally, to support identification of obstacle types, the detection algorithm will need to be improved, introducing object recognition capability.

4 USER-CENTERED SONIFICATION MAPPING DESIGN

We conducted a series of design iterations to define an effective sonification mapping between the identified obstacle properties and auditory dimensions. In each iteration we defined an initial sonification mapping proposal, designed using the sonification engine described in Section 3.3. The proposed sonification was evaluated in a follow-up user study, conducted as a series of listening tests through an online survey. The identified limitations of the approach were then addressed in the next iteration. In all the design iterations we followed the principle of basing mappings on ecological metaphors, with the goal of maximizing intuitiveness and learnability. In this context, the locution "ecological metaphor" is used to mean that the sonification is coherent with users' real-world sensory and cognitive experience [15], so that variations of auditory dimensions are consistent with those of physical parameters (e.g., obstacle horizontal position may be mapped into left/right sound panning, with the amount of panning controlled by the lateralization of the obstacle). In this section, we first outline the experimental setup common to all the design iterations. Then, for each iteration, we report the mapping strategy adopted and the corresponding experimental results.

4.1 Online Listening Tests Experimental Design

Online listening tests, created as a survey on LimeSurvey platform,[8] were distributed through Italian BVI user groups and mailing lists. Any BVI person could participate, but only for one iteration, to avoid possible learning effects. The tests were filled anonymously and included five parts.

(1) Introduction. The participants are provided information on our research and in particular, on the study. If they are willing to participate, they are invited to use headphones in order to better hear the proposed audio samples. When ready, the participants can initiate the survey.

(2) Demographic Questions. The second part collected demographic data including sex, age (in ten year ranges from $18 - 27$ to 58+), visual impairment category ("low vision" or "blindness"), and duration of the condition (since birth, <5years, $5 - 10$years, >10years). We also collected self-reported prior musical and mobile technology experience (low, medium, and high).

(3) Training. The training included one page for each sonified property (thus, four pages in the first three iterations and three pages in the last one, as explained in the following). Each page introduces a single dimension and allows the participant to listen to the possible variations of the sound for that dimension (e.g., near/far for the frontal distance). A written description explained how the considered audio dimension varies and specified the setting associated with each audio sample. Each sample could be played through a link and replayed as much as desired. We highlight that, even though the auditory dimension used to map the distance property is continuous, for simplicity, during the training and the tests we considered only two distance settings: near and far.

(4) Listening Tests. Then, the participants were presented with a number of testing samples that simulated obstacles with different properties. The number of samples corresponded to all the possible combinations of obstacle properties, and each sample was presented once. Each test was presented on a separate page and included a link to the audio sample. Afterwards, a page with four multiple choice questions was presented, one for each obstacle property, and the participants were asked to select the correct ones. For each online listening test, we measured the **accuracy for a given property** as the ratio of correctly categorized sounds by the participants for that property. We also measured the **global accuracy**, that is the ratio of sounds correctly categorized for all four properties. Since the goal was to assess the accuracy rate in identifying the obstacle properties, the effect of order was deemed negligible, and the samples were presented in a fixed order.

(5) User Preferences and Opinions. A final set of questions was included to collect the participants' subjective feedback on the considered sonification mapping. In particular, we asked the

---

[8] https://www.limesurvey.org.

Table 3. Mapping between Obstacle Properties and Auditory Dimensions in the Four Iterations

Iteration	Distance	Horiz. position	Width	Height
1	intermittence	panning	pitch	timbre
2	intermittence	panning	polyphony	pitch
3	intermittence	panning	reverberation	pitch
4	intermittence	panning	-	pitch

participants to evaluate the **comprehensibility** and the **unpleasantness** of the presented sonification mappings, on a 1–5 Likert-like scale. Additional open-ended questions were also recorded to allow the participants to provide a motivation for the submitted scores, as well as to report further comments and suggestions.

4.2 First Iteration

4.2.1 Design. The sonification mappings between obstacle properties and different auditory dimensions are summarized in Table 3 for all iterations. For the first iteration, we used a temporal auditory dimension to sonify *distance*. Specifically, we employed the rate of intermittence of the sound, associating high/low-repetition rate to near/far obstacles, respectively.

This kind of sonification, that intuitively recalls the sound pulses emitted by sonar-like sensors, is easily recognizable, thanks to its widespread use in vehicles. However, we note that, while this technique recalls car parking sonification, the proposed sonification mapping as a whole cannot simply be reduced to such a metaphor, since it is designed to convey also other dimensions concurrently.

In the proposed configuration, the base sound intermittence rate is set to vary linearly, between 50 **pulses per minute** (**ppm**) for far obstacles, up to 280 ppm for close ones.

In order to represent the horizontal *position* of the obstacle in the field of view, we adopted a spatial auditory dimension, i.e., panning. In case the obstacle is on the left or on the right side with respect to the user's orientation, the base sound is played only on the corresponding channel. Instead, for the obstacle in the center position (in front of the user), the sound is played in both channels. Once again, the mapping is intuitive, as it is an exaggeration of the physical behavior of the obstacle under the hypothesis that it is the sound source.

The *width* of the obstacle is associated with sound pitch. A low pitch (a C4 note) corresponds to a large obstacle, whereas a high pitch (a C6 note) corresponds to a narrow object. The underlying metaphor is that larger objects typically produce lower sounds (think of a violin vs. double bass timbre, or small vs. large dog barking). An interval of two octaves was selected to make the difference in pitch clearly distinguishable also to non-musicians (i.e., more than the average pitch difference between male and female voices), and to avoid peculiar musical intervals, which could introduce unpredictable affective reactions.

Finally, the *height* of the obstacle was mapped to a timbral auditory dimension, i.e., the cut-off frequency of a band-pass resonant filter applied to the percussive layer of the base sound. Specifically, a cut-off frequency of 130Hz was used for walkable obstacles, while those to be circumvented were mapped to a cut-off frequency of 6 kHz. This choice was made because high frequencies are generally considered more alarming than low ones.

4.2.2 Evaluation. For the listening test we recruited 22 participants. The aggregated demographic information on the participants for all four iterations is reported in Table 4. Detailed list of participants is reported in Appendix A.

Table 4. Aggregated Demographic Information of Participants

Iteration			1	2	3	4	Tot.
# of participants			22	9	12	18	61
Demographics	Sex	F	8	5	5	10	28
		M	14	4	7	8	33
	Age	18–27	8	4	6	8	26
		28–37	4	2	4	2	12
		38–47	3	1	1	7	12
		48–57	5	2	1	1	9
		58+	2	0	0	0	2
Impairment	Type	Blind	14	5	10	12	41
		Low vision	8	4	2	6	20
	Onset	Birth	11	6	3	8	28
		>10y	4	2	2	1	9
		5–10y	4	1	2	3	10
		<5y	3	0	5	6	14
Expertise	Music	Low	10	1	5	7	23
		Medium	8	5	5	6	24
		High	4	3	2	5	14
	Mobile	Low	1	1	3	5	10
		Medium	6	4	5	6	21
		High	15	4	4	7	30

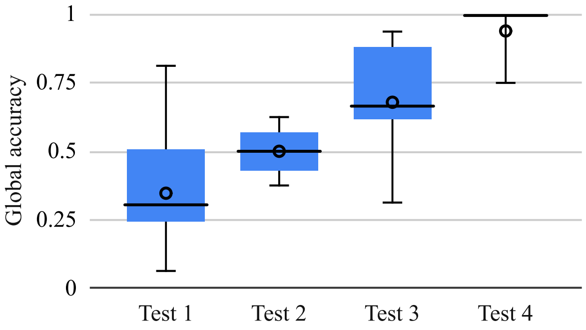


Fig. 4. Global accuracy across the four iterations.

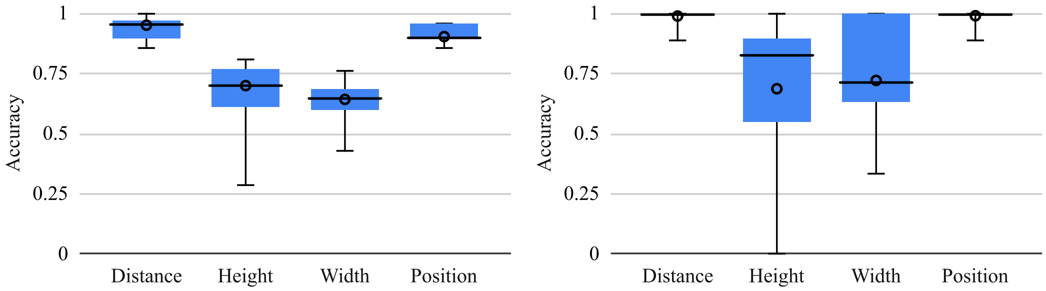
The global accuracy for the first listening test (see Figure 4) was 0.35 ± 0.19.[9] In order to interpret this low global accuracy score, we consider the accuracy scores for single properties, shown in Figure 5(a). The analysis reveals that height and width properties scored much lower ,accuracies (0.65 ± 0.12 and 0.64 ± 0.09, respectively) compared with distance and position (0.92 ± 0.05 and 0.92 ± 0.04, respectively).

This was also reflected in responses to the final set of questions (see Figure 6). Participants on average selected 3.00 ± 1.04[10] for the comprehensibility and 1.50 ± 1.00 for the sound unpleasantness. While the scores were overall positive, 7 participants reported problems in correctly

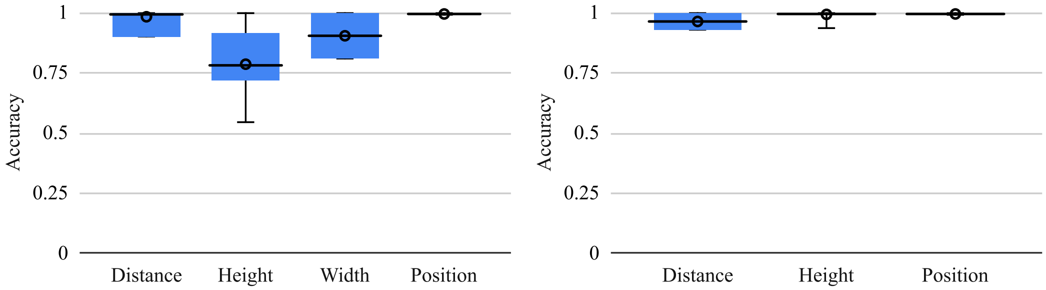
---

[9] We use *mean±standard deviation* notation.

[10] For Likert-like scores, reporting mean is suggested in prior work [42]. Thus, we continue using the same notation[9].

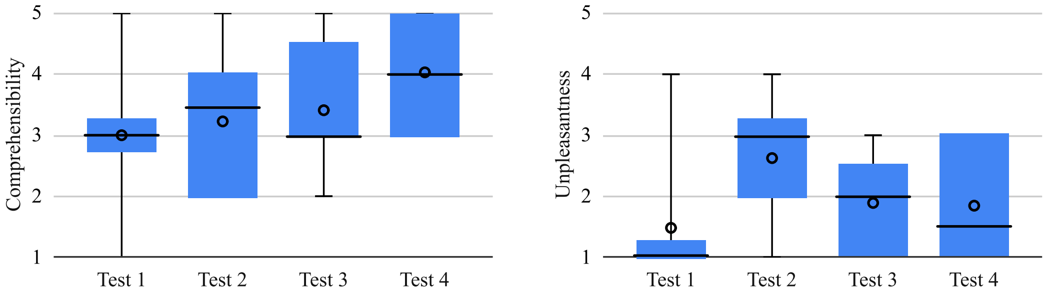


(a) First iteration (b) Second iteration



(c) Third iteration (d) Fourth iteration

Fig. 5. Accuracy of the four obstacle properties in the four iterations.



(a) Comprehensibility (b) Unpleasantness

Fig. 6. Subjective feedback on comprehensibility and unpleasantness across iterations.

understanding width and height properties, and, in particular, one participant reported that the sound pitch feels more natural for representing the obstacle height property.

Considering the participants' demographics, we tested for the effect of age, visual condition, onset and music expertise on the accuracy score. We notice that age influences the accuracy score for the width and height properties. Specifically, participants under 38 years of age ($N = 12$[11]) had significantly lower (0.56 ± 0.14) accuracy scores for width property compared with others ($N = 10$, 0.71 ± 0.20), measured using Mann–Whitney U test with Bonferroni correction[12] [80] ($U = 57.0$, $p < 0.05/4$). Conversely, participants who were 38 and above ($N = 10$) had significantly lower

---

[11] N denotes the number of participants in the group, not datapoints (which is defined as N^* number of trials).

[12] We report significance level as α/m where α is the goal significance level and m is the number of multiple comparisons.

(0.51 ± 0.18) height property accuracy scores ($U = 19.0$, $p < 0.01/4$) compared with others ($N = 13$, 0.76 ± 0.13). Low vision participants ($N = 4$) had lower accuracy scores for all 4 properties (width: 0.56 ± 0.14, height: 0.58 ± 0.19, distance: 0.80 ± 0.20, position: 0.84 ± 0.15) compared with blind participants (width: 0.64 ± 0.10, height: 0.66 ± 0.12, distance: 0.95 ± 0.05, position: 0.93 ± 0.05), but none resulted statistically significant. Considering the musical expertise, participants with "High" level of expertise ($N = 15$) had consistently higher accuracy scores for width (0.84 ± 0.26), distance (1.0 ± 0.0) and position (0.93 ± 0.12) properties, compared with the other participants (width: 0.59 ± 0.06, $U = 30.5$, $p < 0.01/4$; distance: 0.91 ± 0.06, $U = 16.0$, $p < 0.01/4$; position: 0.92 ± 0.03, $U = 80.0$, $p < 0.05/4$, $N = 7$).

4.3 Second Iteration

4.3.1 Design. In the first iteration, the recognition of the *height* and *width* properties was subject to lower accuracy than the two remaining dimensions, and it was variable across demographic groups (age and disability). We argue that the relatively low recognition rates for *height* and *width* may be attributed to the employed metaphors, which generate some confusion. In particular, pitch is known to be intuitively associated with height, and this kind of association seems innate at least in Western culture [9], whereas in the first iteration pitch was instead associated with obstacle *width*. In addition, results showed that recognition is better, and variability is reduced for subjects with musical expertise, which suggests that musical skills improve recognition. In particular, the employed auditory dimension for sonifying *height* (cut-off frequency of a resonant band-pass filter) may be too subtle for the general populace which cannot be expected to possess the timbral awareness exhibited by musically trained subjects.

In light of the above considerations, we conducted a second iteration of the sonification design using pitch to sonify *height*, with low pitch (C4) associated with a walkable obstacle and the high one (C6) to the obstacles that must be circumvented. As for *width*, the choice fell on polyphony (a higher-level musical feature). In particular, a single-pitch was associated with a narrow obstacle, while a three-note chord (i.e., a *cluster*) with wider obstacles. The chosen chord is realized by adding two notes that are 2 and 4 semitones higher than the fundamental (i.e., C-D-E). The compactness of the chord retains the clear distinction between high and low pitch, and the strong difference between a single note and a cluster should be perceivable also to non-musicians.

The filter of the percussive layer is then set to a first order low-pass filter with a cut-off frequency of 130Hz. Even if it is no longer modulated by obstacle properties, we decided to keep the percussive layer, since it helps in the perception of panning, and overlapping repetitions of the base sound for close obstacles.

4.3.2 Evaluation. For the second test, we collected 9 answers from BVI people (see Table 4). In this case the global accuracy (see Figure 4) was 0.5 ± 0.09, and a significant improvement was noticed for all four individual properties (see Figure 5(b)): for height it improved to 0.70 ± 0.30 ($U = 82.0$, $p < 0.05$), for distance to 0.99 ± 0.03 ($U = 27.0$, $p < 0.01$), for width to 0.74 ± 0.25 ($U = 71.0$, $p < 0.05$), and for position to 0.99 ± 0.04 ($U = 24.0$, $p < 0.01$).

We believe that, with the improved understanding of the height and width obstacle properties, other properties were also easier to disambiguate, and therefore yielded better accuracy scores. However, in the case of the second iteration, this improvement came at the cost of an increase in the unpleasantness score (see Figure 6) reported by the participants (2.88 ± 1.05). Indeed, the unpleasantness score in the case of second iteration was significantly higher than in iteration 1 ($p < 0.01$) and iteration 4 ($p < 0.05$), as computed using Kruskal–Wallis test (χ^2: 8.58, $p < 0.05$) with Dunn post-hoc test. Instead, the perceived comprehensibility of sounds remained substantially the same (3.22 ± 1.09).

We also notice that the variability between different demographic groups, highlighted in the results of the first listening test, tend to disappear with improved overall accuracy. Indeed, with respect to the first test, considering participants' age, we notice improved results in the width score accuracy (0.73 ± 0.26, $U = 79.0$, $p < 0.05$) for participants under 38 ($N = 6$), and also in the height score accuracy for participants who were 38 and over (0.77 ± 0.35, $U = 59.0$, $p < 0.01$, $N = 3$). However, no significant differences were found between those groups in the second test. Similarly, no differences were found between low-vision and blind participants, as well as those with different musical expertise levels.

4.4 Third Iteration

4.4.1 Design. Despite the improvements, the experimental results in the second iteration still shows a relatively low accuracy in height and width recognition. This can be due to the use of polyphony (chords) as an auditory dimension for width. As a matter of fact, this may interfere with the perception of pitch (associated with height), especially for non-musicians, even though we specifically designed a compact cluster to avoid that. Additionally, since from an auditory standpoint polyphony was the only new dimension introduced in the second iteration with respect to the first one, it can be deduced that this dimension is the main responsible for the lowering in the pleasantness scores.

Based on these considerations, we changed the sonification strategy for width, and we adopted an ecological rather than a musical metaphor. Specifically, we adopted a spatial auditory dimension: reverberation. Wide obstacles are associated with a reverberated sound, while narrow ones are associated with a dry, non reverberated sound. The working hypothesis was that it should be intuitive to associate size with reverberation, since in an everyday listening context this is a phenomenon strictly connected to both environment size and object size (i.e., larger objects generate more acoustic reflections than smaller objects with the same surface properties).

The reverb block in the sonification engine (see Figure 2) was implemented using a so-called "perceptual" synthetic reverberation algorithm [77], which synthesizes early reflections through four delay lines, and the late reverberation tail through a feedback delay network with 16 channels [86, Channel 5]. This results in a perceptually convincing reverberation effect at low computational costs.

The reverberation time was set to 1 second, which roughly corresponds to that of a medium-sized conference room, or a small theater. Such a value should make the reverberation clearly audible, but at the same time not too invasive.

4.4.2 Evaluation. The test for the third iteration involved 12 BVI participants (see Table 4). The accuracy scores for single properties are shown in Figure 5(c). Even if the average accuracy for *height* is not significantly different from the previous iteration, its standard deviation dropped considerably (0.78 ± 0.13). Moreover, the average accuracy for *width* increased, together with a minor reduction of its standard deviation: 0.91 ± 0.08 ($U = 84.0$, $p < 0.5$). Taken together, these results produced the best overall accuracy yet (0.71 ± 0.19), as shown in Figure 4.

Despite an increase in perceived comprehensibility (3.25 ± 1.29) and a reduction in the unpleasantness score (2.00±0.85), these results were not significantly different with respect to the previous iterations (see Figure 6). The reverberation effect was also reported as hard to understand by one participant, while another still reported that height and width were easily confused.

The new sonification strategy reintroduced some small but significant variability across different categories of people that we observed in the first iteration but not in the second. In particular, *height* is better recognized by people over 28 (0.81 ± 0.40, $N = 6$ *v.s.* 0.78 ± 0.13, $N = 6$; $U = 54.50$, $p < .01$), people with late insurgence of visual impairment (0.75 ± 0.14, $N = 7$ *v.s.* 0.85 ± 0.21, $N = 5$;

$U = 71.00$, $p < .05$), and those that are musically skilled (0.68 ± 0.20, $N = 7$ *v.s.* 0.86 ± 0.12, $N = 5$; $U = 37.00$, $p < .001$). Similarly, *width* is better recognized by technologically skilled people (0.89 ± 0.11, $N = 4$ *v.s.* 0.96 ± 0.11, $N = 8$; $U = 81.00$, $p < .05$).

Interestingly, when looking at this inter-group variability, it can be noticed that the recognition of *height* was the most affected, even if the changes with respect to the previous iteration regarded the sonification of *width* only. This result supports our conjecture that the previous sonification for width might interfere with the recognition of height and suggests the existence of some form of perceptual or cognitive relationship between these two parameters, which deserves further attention.

4.5 Fourth Iteration

4.5.1 Design. Although the third iteration provided significant improvements over the initial sonification design, we wanted to explore a drastically different direction aimed at evaluating the effect of cognitive load on the overall recognition accuracy and on the sonification pleasantness. Specifically, we decided to reduce the number of sonified dimensions from four to three.

After discussing the topic among the designers and with blind individuals, we decided to drop the *width* dimension, which is considered less relevant than the others. Indeed, *distance* and *position* provide crucial information to avoid obstacles and hence cannot be dropped. Also, *height* can provide information on the most unexpected (and dangerous) obstacles for BVI individuals i.e., hanging obstacles, which cannot be perceived with the white cane [10]. While *width* is still considered important, it can be considered less important than the others also because it actually helps avoiding the obstacle when it is not-walkable and in front of the user; however, in this case a wide object (e.g., a wall) can often be perceived with echolocation. Note that similar considerations led van Erp et al. [78] to consider these very same dimensions (position, distance, and height) for vibrotactile obstacle detection display. As a consequence, the employed sounds were the same as in the third iteration, with the exclusion of reverberated ones.

4.5.2 Evaluation. For the evaluation we recruited 18 BVI participants (see Table 4). The reduction of the sonified dimensions and of the cognitive load resulted in a very high overall accuracy (0.96 ± 0.09). With regard to individual dimensions (see Figure 5(d)), the *height* score improved considerably with respect to the previous iteration, reaching its maximum across all iterations (0.78 ± 0.13 in iteration 3, 0.99 ± 0.02 in iteration 4; $U = 4.50$, $p < .001$). Statistically significant differences were not found for other dimensions.

The perceived comprehensibility was the highest (4.06 ± 1.11) among all the iterations ($p < 0.01$), as computed using Kruskal-Wallis test ($\chi^2 : 41.79$, $p < 0.05$) with Dunn post-hoc test. The unpleasantness score in iteration 4 dropped to 1.72 ± 0.89, placing this iteration close to the first (and most pleasant) one.

Finally, inter-group variability dropped. The only significant inter-group difference was found in *distance* recognition, in favor of people with low vision (1.00 ± 0.00, $N = 6$ *v.s.* 0.95 ± 0.05, $N = 12$; $U = 16.00$, $p < .05$).

5 DISCUSSION

The data from the sonification design iteration studies and the real-world experiment help us to better understand how sonification can be used to convey obstacle properties in the context of an obstacle avoidance system. In particular, here we discuss how the amount of the considered obstacle properties and their mapping onto different auditory dimensions influence the participants' ability to concurrently discern different obstacle properties, and their subjective perception of the sonification approach.

5.1 Discernibility of Obstacle Properties

Our initial design of the sonification interaction considered four properties that were deemed most useful for the obstacle avoidance task: distance, position, width, and height. We conducted multiple design iterations aimed towards selecting ecologically valid, intuitive, and easy to distinguish auditory dimensions that could be used to convey these properties. This task proved to be quite complex. Indeed, during the first iteration, the width and height obstacle properties, associated with sound pitch and timbre, respectively, were easily confused by the participants. We hypothesized that there were two causes for this issue: (1) the association of sound pitch to the width obstacle property because this dimension is commonly linked to height, and (2) the documented interaction between the pitch and the timbre property [41], which resulted in them being harder to understand for those with lower musical expertise.

In the second iteration we remapped sound pitch to the height dimension and adopted polyphony as the auditory dimension to define width. With this change, an overall improvement in accuracy was noticed for all four considered dimensions. This was unexpected, as we did not change the mapping of distance and position dimensions. We believe that an improvement in the understandability for width and height also influenced the global understanding of the sonification. Nonetheless width and height were still performing worse than distance and position, and a decrease in the perceived pleasantness of the interaction technique was also noticed. We speculated that the usage of polyphony resulted in a higher mental workload, in particular for non-musicians, and that it could still lead to confusion with respect to the pitch dimension.

In the third iteration we therefore substituted polyphony with reverb for the width property since this auditory dimension is commonly associated with larger objects. Again, an overall improvement was detected, in particular for the width dimension, but the comprehension of all the four obstacle properties was still lower than we expected. Based on the results of the three iterations, we concluded that the considered dimensions might be too numerous for a clear and concurrent understanding.

Thus, in the fourth iteration we reduced the sonified dimensions to three: distance, position, and height. Indeed, this led to an almost perfect recognition of the remaining properties during the follow-up study. This finding may indicate that for sonification tasks that require reliable understanding of multiple concurrent dimensions, there is a limit to the number of auditory dimensions that can be discerned with a short training. A possible explanation for this fact, supported by prior literature [7], is that changes in some of the auditory dimensions are harder to distinguish, and therefore selecting multiple properties that are easy to discern becomes harder when the dimensions to sonify are many. This finding will be useful for researchers and practitioners designing sonification interaction techniques requiring multiple properties to be conveyed at the same time.

5.2 Subjective Perception of the Sonification Designs

Throughout the iterative sonification design process, we noticed that the participants' overall perception of the sound comprehensibility appeared to vary in agreement with the measured global accuracy of the sonification. We were not able to test this correlation for statistical significance, due to the small sample size. Nonetheless, we highlight that both metrics display an increasing trend across the iterations, as shown in Figures 4 and 6(a). This may indicate that the users are able to perceive how well they are capable of following the proposed sonification mapping. Thus, even without an objective measurement of the accuracy, participants' subjective perception could be used to estimate the effectiveness of the sonification. A result of this finding could be a quicker turnaround of the sonification design iterations, without the need to compute the accuracy metrics at every iteration.

Instead, the changes in the unpleasantness score varied based on the considered auditory dimensions and did not seem to impact the accuracy considering global results. However, it seemed to correlate to the high variance in the accuracy of specific metrics, in this case height and width. In particular, the sound mapping presented in the second iteration was considered to be more unpleasant with respect to all others, and this difference was statistically significant with respect to the first and the last iterations. Indeed, we argue that the increased unpleasantness in the second iteration is mainly due to the sonification of *width*, and specifically to the use of musical clusters to sonify large obstacles. While this approach proved to be effective and informative for most of the participants, the employed clusters are highly dissonant, being composed of three contiguous tones, which was found to be unpleasant and impacted the accuracy for others.

5.3 Results Generalizability

While the driving motivation for our research is to find an effective sonification technique for a specific application (i.e., obstacle detection), we argue that our results can be generalized along with different directions.

First, our results show that, with a short training, the sonification of four dimensions (i.e., obstacle properties) has a global accuracy of at most 0.71 (in the third iteration), while reducing the dimensions to 3, the global accuracy is almost perfect (0.96). This is partially in line with exiting results in the literature. As an example, Schuett et al. [69] show that the recognition accuracy decreases when the number of dimensions changes from 3 to 5. We cannot directly compare the results presented in our article from those obtained by Schuett et al., because the experimental setup is different. However, we can observe that with three dimensions we obtained an almost perfect accuracy (this is not the case in the article by Schuett et al.) and that the drop in accuracy from adding a single dimension is drastic, while Schuett et al. observed a moderate decrease in accuracy when adding two dimensions. So, our results confirm those presented by Schuett et al and further show that even adding a single dimension could have a strong negative impact on the understanding of the sonified quantities. This suggests that there might be a threshold limit in the number of concurrent dimensions that can be accurately conveyed to the users, in absence of a longer training.

A second possible generalization of our results concerns the application of the designed sonifications within different assistive technologies. A number of assistive technologies based on augmented reality interaction have been proposed in the last years, including some aimed at people with visual impairments [5, 50]. These applications might require conveying multiple spatial relations concurrently (e.g., the position and the distance of an object with respect to the user) through sonification. In such cases, our contribution could provide a solution to sonify three or four dimensions at a time. If needed, by adding a suitable training stage to the procedure, possibly more dimensions could be considered.

Another aspect of our contribution that can be generalized, beyond the specific application, is the employed research methodology. We adopted an instance of the user-centric design approach in which we can easily perform fast-paced design iterations of sonification mappings and test them through online surveys. This approach is practical and convenient, especially when involving participants with disabilities, who may have difficulties to attend to user studies in distant and unfamiliar places. With this approach, our results show that we have been able to incrementally improve the recognition accuracy of obstacle characteristics over four iterations, hence showing that the methodology is indeed effective.

5.4 Limitations

5.4.1 Online Survey Usage. While the usage of a web-based survey has numerous benefits, as outlined above, it also exposes the study to possible limitations due to the lack of control over the participant pool. Due to this we were not able to guarantee the demographic consistency between the participants for each survey, which might introduce some additional variance in the obtained results. Additionally, through an online form, it is difficult to obtain very detailed information on the participants, and it is not possible to ask further clarifications. Thus, our knowledge of the participants is not exhaustive. For the same reason we also were limited in the capability to collect user feedback besides the one provided spontaneously in the open comments. Indeed, for the second and the fourth iterations we had no comments provided by the participants.

Finally, it is also impossible to observe the participants during the study. This impacts the study for two reasons. First, direct observation is a valuable source of information (e.g., non verbal feedback) for researchers and it was not accessible with our methodology. Second, we cannot be sure that the participants performed the study seriously or read everything with attention. For example, Participant 1.15 (see Appendix A) reported that he did not listen to all sounds during the training. This participant had scores consistent with others. Thus, we wonder how many others did not follow the procedure perfectly, and how that impacted the results of our investigation. Similarly we cannot exclude the presence of errors with the web survey or participants' computers.

5.4.2 Validity of the Proposed Sonifications. The proposed sonification mappings explore only some of the possible combinations in this very ample design space. We were driven by a sound scientific approach, using ecologically valid associations between obstacle properties and auditory dimensions, based on prior literature [9, 15] and our experience as researchers in the field. However, we cannot exclude that other mappings could have yielded better results. In particular, we are not able to factor out that what we consider as ecologically valid sonification mappings could be perceived in a different way by the participants. While the improvement in the results across the iterations suggests that our approach should be valid, the persistently low scores for height and width obstacle properties, in the iterations with 4 sonified dimensions, may hint otherwise. Thus, in future we will explore the feasibility of radically different sonification mappings, possibly using auditory dimensions selected in a different way, for example based on a prior investigation with end users.

Another possible limitation is that our user pool was recruited from one single source. Indeed, all of our participants are Italian. In particular, many of them are members of Italian associations of BVI people. This suggest that they might share demographic characteristics that could be influential to their perception of the auditory dimensions and therefore might influence the obtained scores. We believe that the scores would be similar in most European countries, but we cannot exclude that the generalizability of our claims might be limited in other parts of the world, for example low income settings.

6 CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

This article presents *WatchOut*, a sonification technique aimed at conveying obstacles information to BVI users in real time. After an initial usability study of the approach in the real world, which highlighted that mobile-driven obstacle detection is indeed a feasible and appreciated functionality, we polished and refined the sonification mapping used, throughout four design iterations conducted using an online survey. The resulting technique was shown to be capable of conveying multiple obstacle properties concurrently.

From the start we were able to convey distance and position obstacle properties with almost perfect accuracy, due to an effective mapping of these properties to ecologically corresponding auditory dimensions. These mappings were therefore preserved during all the four iterations. For width and height obstacle properties, however, we had much harder time in identifying suitable sonification mappings to appropriate sound dimensions. For height, we believe that pitch, which is commonly associated with this property, is an adequate and ecologically valid representation. For width, the reverberation dimension yielded the best results, reaching a global accuracy of about 0.71.

However, for safety-critical applications, such as obstacle detection, the obtained results can still be considered insufficient. Thus, we ultimately proposed a sonification design that sacrifices one obstacle property but achieves near-perfect global accuracy (0.96). This may indicate that an overall reduction in the number of the sonified dimensions makes it easier to distinguish different obstacle characteristics. An alternative hypothesis is that the interaction between the height and width sonification mappings was responsible for the lower score and therefore an improvement is possible if a better mapping of these dimensions is found.

6.2 Future Work

As future work we will further explore these two possible directions. On one hand, we will progress in our investigation of the possible sonification mappings, striving to identify better ecologically associations. On the other hand, we will analyze how the accuracy in concurrently discerning multiple obstacle properties varies by modulating the numerosity of the sonified auditory dimensions. Furthermore, we will study how the defined approach and the resulting sonifications can generalize to other applications, both in the field of assistive technologies for BVI people, and more in general for non-visual human–computer interaction.

In the future we also intend to further investigate the mobile system as well as the recognition module, improving the accuracy and robustness of the prototype we used in this contribution to test *WatchOut* in the real world. These tests will be conducted with a larger number of participants and in different scenarios in order to thoroughly validate the implemented system. For this test, in addition to the familiarity with the area, the musical and mobile technology expertise, we will also evaluate the impact of the independent travel expertise of the participants.

For the obstacle detection mobile application, several ideas for future work also emerged from the comments provided by the participants. In particular, they suggested to render our approach multi-modal by adding vibration feedback and possibly verbal cues, such as names of specific obstacles. Another proposal is to allow the users to personalize the interaction and the sounds used. Finally, we were suggested to tackle the problem of head level obstacles, which are recognized as a common danger for people with BVI during mobility [48].

APPENDIX

A PARTICIPANTS DEMOGRAPHIC INFORMATION

Table 5. Participants' Demographic Information for the First Iteration

ID	Sex	Age	Visual Impairment		Experience	
			Condition	Since	Music	Mobile
1,1	F	38–47	Blind	<5	High	High
1,2	M	28–37	Blind	Birth	Med.	High
1,3	F	28–37	Blind	<5	Low	Med.
1,4	M	18–27	Blind	Birth	Med.	High
1,5	M	28–37	Blind	5–10	Low	High
1,6	F	18–27	Blind	Birth	Low	Med.
1,7	F	48–57	Blind	5–10	Low	Low
1,8	M	18–27	Blind	Birth	Low	High
1,9	M	18–27	Blind	Birth	Low	High
1,10	F	28–37	Blind	Birth	High	High
1,11	M	48–57	Blind	>10	Med.	High
1,12	F	18–27	Blind	Birth	Med.	High
1,13	M	48–57	Blind	<5	Med.	Med.
1,14	M	18–27	Blind	Birth	High	High
1,15	M	>58	Blind	>10	Med.	Med.
1,16	M	48–57	Blind	5–10	Med.	High
1,17	M	>58	Blind	Birth	Low	High
1,18	F	38–47	Low vision	>10	Low	High
1,19	M	48–57	Blind	Birth	High	High
1,20	M	18–27	Low vision	>10	Low	Med.
1,21	F	18–27	Low vision	5–10	Med.	High
1,22	M	38–47	Low vision	Birth	Low	Med.

Table 6. Participants' Demographic Information for the Second Iteration

ID	Sex	Age	Visual Impairment		Experience	
			Condition	Since	Music	Mobile
2,1	M	18–27	Blind	Birth	Med.	Med.
2,2	F	38–47	Low vision	>10	High	Med.
2,3	M	48–57	Blind	Birth	High	High
2,4	F	18–27	Blind	Birth	High	High
2,5	F	18–27	Low vision	Birth	Med.	High
2,6	F	28–37	Low vision	5–10	Med.	Med.
2,7	F	18–27	Low vision	>10	Low	Med.
2,8	M	28–37	Blind	Birth	Med.	High
2,9	M	48–57	Blind	Birth	Med.	Low

Table 7. Participants' Demographic Information for the Third Iteration

ID	Sex	Age	Visual Impairment		Experience	
			Condition	Since	Music	Mobile
3,1	M	28–37	Blind	Birth	Low	Med.
3,2	F	18–27	Blind	<5	High	Med.
3,3	F	28–37	Blind	>10	Med.	High
3,4	M	18–27	Blind	<5	Low	Low
3,5	M	18–27	Low vision	5–10	Low	Med.
3,6	F	48–57	Blind	Birth	Med.	Med.
3,7	M	28–37	Blind	>10	Low	High
3,8	M	18–27	Blind	<5	Med.	Med.
3,9	F	38–47	Blind	<5	Med.	High
3,10	M	18–27	Blind	5–10	Med.	Low
3,11	F	18–27	Low vision	Birth	High	High
3,12	M	28–37	Blind	<5	Low	Low

Table 8. Participants' Demographic Information for the Fourth Iteration

ID	Sex	Age	Visual Impairment		Experience	
			Condition	Since	Music	Mobile
4,1	F	18–27	Blind	Birth	Low	Med.
4,2	F	38–47	Blind	<5	Med.	Med.
4,3	M	28–37	Blind	Birth	High	High
4,4	F	38–47	Low vision	Birth	Med.	Low
4,5	F	18–27	Low vision	Birth	High	High
4,6	M	48–57	Blind	Birth	Med.	Med.
4,7	M	18–27	Blind	5–10	Low	Med.
4,8	M	38–47	Blind	<5	Low	Low
4,9	M	38–47	Blind	5–10	Med.	Low
4,10	M	18–27	Low vision	5–10	Low	Low
4,11	F	38–47	Blind	<5	Med.	High
4,12	M	18–27	Low vision	<5	High	High
4,13	F	38–47	Blind	>10	Low	Low
4,14	F	18–27	Low vision	<5	High	High
4,15	F	38–47	Blind	<5	Med.	Med.
4,16	F	28–37	Blind	Birth	Low	Med.
4,17	F	18–27	Blind	Birth	High	High
4,18	M	18–27	Low vision	Birth	Low	High

REFERENCES

[1] Daniel Aguerrevere, Maroof Choudhury, and Armando Barreto. 2004. Portable 3D sound/sonar navigation system for blind individuals. In *Proceedings of the 2nd LACCEI International Latin American Caribbean Conference. Engineering Technology.*

[2] Dragan Ahmetovic, Federico Avanzini, Adriano Baratè, Cristian Bernareggi, Gabriele Galimberti, Luca A. Ludovico, Sergio Mascetti, and Giorgio Presti. 2019. Sonification of rotation instructions to support navigation of people with visual impairment. In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications.* 332–341.

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

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

[5] Dragan Ahmetovic, Daisuke Sato, Uran Oh, Tatsuya Ishihara, Kris Kitani, and Chieko Asakawa. 2020. ReCog: Supporting blind people in recognizing personal objects. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* ACM, 1–12. https://doi.org/10.1145/3313831.3376143

[6] Amjed S. Al-Fahoum, Heba B. Al-Hmoud, and Ausaila A. Al-Fraihat. 2013. A smart infrared microcontroller-based blind guidance system. *Active and Passive Electronic Components* 2013, 16 (2013). https://doi.org/10.1155/2013/726480

[7] Janet E. Anderson and Penelope Sanderson. 2009. Sonification design for complex work domains: Dimensions and distractors. *Journal of Experimental Psychology: Applied* 15, 3 (2009), 183–198.

[8] Andreas, Joachim Diepstraten, and Thomas Ertl. 2004. Design and development of an indoor navigation and object identification system for the blind. In *Proceedings of the International Conference ACM Sigaccess Accessibility and Computing.* ACM, 147–152.

[9] Richard Ashley. 2004. Musical pitch space across modalities: Spatial and other mappings through language and culture. In *Proceedings of the 8th International Conference on Music Perception and Cognition*, S. D. Lipscomb, R. Ashley, R. O. Gjerdingen, and P. Webster (Eds.). Causal Productions, Evanston, Illinois, 64–71.

[10] Jinqiang Bai, Zhaoxiang Liu, Yimin Lin, Ye Li, Shiguo Lian, and Dijun Liu. 2019. Wearable travel aid for environment perception and navigation of visually impaired people. *Electronics* 8, 6 (2019), 697.

[11] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies* 4, 3 (2009), 114–123.

[12] M. Bouzit, A. Chaibi, K. J. De Laurentis, and C. Mavroidis. 2004. Tactile feedback navigation handle for the visually impaired. In *Proceedings of the ASME International Mechanical Engineering Congress and Exposition.* 1171–1177.

[13] Michael Brock and Per Ola Kristensson. 2013. Supporting blind navigation using depth sensing and sonification. In *Proceedings of the ACM Conference Pervasive and Ubiquitous Computing Adjunct Publication.* ACM, 255–258.

[14] John Brooke. 1996. SUS: A quick and dirty usability scale. *Usability Evaluation in Industry.* P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland (Eds.), Taylor & Francis. 189–194.

[15] Egon Brunswik and Joe Kamiya. 1953. Ecological cue-validity of 'proximity' and of other Gestalt factors. *American Journal of Psychology* 66, 1 (1953), 20–32.

[16] Allan D. Coop. 2016. Sonification, musification, and synthesis of absolute program music. In *Proceedings of the 22nd Annual International Conference on Auditory Display.* 177–183.

[17] Dimitrios Dakopoulos, Sanjay K. Boddhu, and Nikolaos Bourbakis. 2007. A 2D vibration array as an assistive device for visually impaired. In *Proceedings of the IEEE 7th International Symposium on BioInformatics and BioEngineering.* IEEE, 930–937.

[18] D. Dakopoulos and N. G. Bourbakis. 2010. Wearable obstacle avoidance electronic travel aids for blind: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40, 1 (2010), 25–35.

[19] Quoc Dang, Youngjoon Chee, Duy Pham, and Young Suh. 2016. A virtual blind cane using a line laser-based vision system and an inertial measurement unit. *Sensors* 16, 1 (2016), 95.

[20] Rahul Dhod, Gurmohan Singh, Gagandeep Singh, and Manjit Kaur. 2017. Low cost GPS and GSM based navigational aid for visually impaired people. *Wireless Personal Communications* 92, 4 (2017), 1575–1589.

[21] Tilman Dingler, Jeffrey Lindsay, and Bruce N. Walker. 2008. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proceedings of the 14th International Conference on Auditory Display.*

[22] A. G. Dodds, J. D. Armstrong, and C. A. Shingledecker. 1981. The nottingham obstacle detector: Development and evaluation. *Journal of Visual Impairment & Blindness* 75, 5 (1981), 203–209.

[23] Gaël Dubus and Roberto Bresin. 2013. A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE* 8, 12 (2013), e82491.

[24] Larisa Dunai, Guillermo Peris Fajarnes, Victor Santiago Praderas, Beatriz Defez Garcia, and Ismael Lengua Lengua. 2010. Real-time assistance prototype–a new navigation aid for blind people. In *Proceedings of the IECON 2010 36th Annual Conference on IEEE Industrial Electronics Society.* IEEE, 1173–1178.

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

[26] ETSI. 2002. *Human Factors: Guidelines on the Multimodality of Icons, Symbols and Pictograms.* Technical Report EG 202 048 (V1.1.1).

[27] Michele Geronazzo, Alberto Bedin, Luca Brayda, Claudio Campus, and Federico Avanzini. 2016. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal on Human Computer Studies* 85 (2016), 4–15.

[28] Nicholas A. Giudice. 2018. Navigating without vision: Principles of blind spatial cognition. In *Handbook of Behavioral and Cognitive Geography*, Daniel R. Montello (Ed.). Edward Elgar Publishing, 260–288.

[29] Nicholas A. Giudice and Gordon E. Legge. 2008. Blind navigation and the role of technology. In *The Engineering Handbook of Smart Technology for Aging, Disability, and Independence*, A. Helal, M. Mokhtari, and B. Abdulrazak (Eds.). Wiley, 479–500.

[30] José L. González-Mora, A. Rodriguez-Hernandez, L. F. Rodriguez-Ramos, L. Díaz-Saco, and N. Sosa. 1999. Development of a new space perception system for blind people, based on the creation of a virtual acoustic space. In *Proceedings of the International Work-Conference on Artificial Neural Networks*. Springer, 321–330.

[31] Florian Grond and Jonathan Berger. 2011. Parameter mapping sonification. In *The Sonification Handbook*, T. Hermann, A. Hunt, and J. G. Neuhoff (Eds.). Logos Verlag, Berlin, 363–397.

[32] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M. Kitani, and Chieko Asakawa. 2019. Cabot: Designing and evaluating an autonomous navigation robot for blind people. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 68–82.

[33] Feng Hu, Norbu Tsering, Hao Tang, and Zhigang Zhu. 2016. Indoor localization for the visually impaired using a 3d sensor. In *Proceedings of the 31st Annual International Technology and Persons with Disabilities Conference*.

[34] Tohru Ifukube, Tadayuki Sasaki, and Chen Peng. 1991. A blind mobility aid modeled after echolocation of bats. *IEEE Transactions on Biomedical Engineering* 38, 5 (1991), 461–465.

[35] Kiyohide Ito, Makoto Okamoto, Junichi Akita, Tetsuo Ono, Ikuko Gyobu, Tomohito Takagi, Takahiro Hoshi, and Yu Mishima. 2005. CyARM: An alternative aid device for blind persons. In *Proceedings of the CHI'05 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1483–1488.

[36] Rabia Jafri, Rodrigo Louzada Campos, Syed Abid Ali, and Hamid R. Arabnia. 2017. Visual and infrared sensor data-based obstacle detection for the visually impaired using the Google project tango tablet development kit and the unity engine. *IEEE Access* 6 (2017), 443–454.

[37] Rabia Jafri and Marwa Mahmoud Khan. 2016. Obstacle detection and avoidance for the visually impaired in indoors environments using Googles Project Tango device. In *Proceedings of the International Conference on Computers Helping People with Special Needs*. Springer, 179–185.

[38] Hernisa Kacorri, Sergio Mascetti, Andrea Gerino, Dragan Ahmetovic, Valeria Alampi, Hironobu Takagi, and Chieko Asakawa. 2018. Insights on assistive orientation and mobility of people with visual impairment based on large-scale longitudinal data. *ACM Transactions on Accessible Computing* 11, 1 (2018), 1–28.

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

[40] Leslie Kay. 1964. An ultrasonic sensing probe as a mobility aid for the blind. *Ultrasonics* 2, 2 (1964), 53–59.

[41] Carol L. Krumhansl and Paul Iverson. 1992. Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance* 18, 3 (1992), 739.

[42] James R. Lewis. 1993. Multipoint scales: Mean and median differences and observed significance levels. *International Journal of Human-Computer Interaction* 5, 4 (1993), 383–392.

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

[44] Hong Liu, Jun Wang, Xiangdong Wang, and Yueliang Qian. 2015. iSee: Obstacle detection and feedback system for the blind. In *Adjunct Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the ACM International Symposium on Wearable Computers*. ACM, 197–200.

[45] Esther Loeliger and Tony Stockman. 2014. Wayfinding without visual cues: Evaluation of an interactive audio map system. *Interacting with Computers* 26, 5 (2014), 403–416.

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

[47] Luca Andrea Ludovico and Giorgio Presti. 2016. The sonification space: A reference system for sonification tasks. *International Journal of Human-Computer Studies* 85 (2016), 72–77. https://doi.org/10.1016/j.ijhcs.2015.08.008

[48] Roberto Manduchi. 2012. Mobile vision as assistive technology for the blind: An experimental study. In *Proceedings of the International Conference on Computers Helping People with Special Needs*. Springer, 9–16.

[49] Roberto Manduchi and Sri Kurniawan. 2011. Mobility-related accidents experienced by people with visual impairment. *AER Journal: Research and Practice in Visual Impairment and Blindness* 4, 2 (2011), 44–54.

[50] Nii Mante and James D. Weiland. 2018. Visually impaired users can locate and grasp objects under the guidance of computer vision and non-visual feedback. In *Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 1–4.

[51] Sergio Mascetti, Dragan Ahmetovic, Andrea Gerino, and Cristian Bernareggi. 2016. Zebrarecognizer: Pedestrian crossing recognition for people with visual impairment or blindness. *Pattern Recognition* 60 (2016), 405–419.

[52] Sergio Mascetti, Dragan Ahmetovic, Andrea Gerino, Cristian Bernareggi, Mario Busso, and Alessandro Rizzi. 2016. Robust traffic lights detection on mobile devices for pedestrians with visual impairment. *Computer Vision and Image Understanding* 148, C (2016), 123–135.

[53] Simon Meers and Koren Ward. 2005. A substitute vision system for providing 3D perception and GPS navigation via electro-tactile stimulation. In *Proceedings of International Conference on Sensing Technology*.

[54] Peter B. L. Meijer. 1992. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering* 39, 2 (1992), 112–121.

[55] Kawin Metsiritrakul, Atiwong Suchato, and Proadpran Punyabukkana. 2017. Obstacle avoidance feedback system for the blind using stereo sound. In *Proceedings of the International Convention on Rehabilitation Engineering and Assistive Technology*. Therapeutic, Assistive & Rehabilitative Technologies (START) Centre, 1–21.

[56] Davide Mulfari. 2018. A TensorFlow-based assistive technology system for users with visual impairments. In *Proceedings of the Internet of Accessible Things*. 1–2.

[57] Rai Munoz, Xuejian Rong, and Yingli Tian. 2016. Depth-aware indoor staircase detection and recognition for the visually impaired. In *Proceedings of the 2016 IEEE International Conference on Multimedia & Expo Workshops*. IEEE, 1–6.

[58] Ayat A. Nada, Mahmoud A. Fakhr, and Ahmed F. Seddik. 2015. Assistive infrared sensor based smart stick for blind people. In *Proceedings of the 2015 Science and Information Conference*. IEEE, 1149–1154.

[59] World Health Organization et al. 2011. *World Report on Disability 2011*. World Health Organization.

[60] Huy-Hieu Pham, Thi-Lan Le, and Nicolas Vuillerme. 2016. Real-time obstacle detection system in indoor environment for the visually impaired using microsoft kinect sensor. *Journal of Sensors* 2016, 11 (2016), 1–13.

[61] Matteo Poggi and Stefano Mattoccia. 2016. A wearable mobility aid for the visually impaired based on embedded 3D vision and deep learning. In *Proceedings of the 2016 IEEE Symposium on Computers and Communication*. IEEE, 208–213.

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

[63] Rosali Pyun, Yeongmi Kim, Pascal Wespe, Roger Gassert, and Stefan Schneller. 2013. Advanced augmented white cane with obstacle height and distance feedback. In *Proceedings of the 2013 IEEE 13th International Conference on Rehabilitation Robotics*. IEEE, 1–6.

[64] Santiago Real and Alvaro Araujo. 2019. Navigation systems for the blind and visually impaired: Past work, challenges, and open problems. *Sensors* 19, 15 (2019), 3404.

[65] Manaswi Saha, Alexander J. Fiannaca, Melanie Kneisel, Edward Cutrell, and Meredith Ringel Morris. 2019. Closing the gap: Designing for the last-few-meters wayfinding problem for people with visual impairments. In *Proceedings of the 21st International acm Sigaccess Conference on Computers and Accessibility*. 222–235.

[66] Gopala Sainarayanan, R. Nagarajan, and Sazali Yaacob. 2007. Fuzzy image processing scheme for autonomous navigation of human blind. *Applied Soft Computing* 7, 1 (2007), 257–264.

[67] Rajib Sarkar, Sambit Bakshi, and Pankaj K. Sa. 2012. Review on image sonification: A non-visual scene representation. In *Proceedings of the Internationaol Conference on Recent Advances in Information Technology*. IEEE, 86–90.

[68] Daisuke Sato, Uran Oh, João Guerreiro, Dragan Ahmetovic, Kakuya Naito, Hironobu Takagi, Kris M. Kitani, and Chieko Asakawa. 2019. NavCog3 in the wild: Large-scale blind indoor navigation assistant with semantic features. *ACM Transactions on Accessible Computing* 12, 3 (2019), 1–30.

[69] Jonathan H. Schuett, Riley J. Winton, Jared M. Batterman, and Bruce N. Walker. 2014. Auditory weather reports: Demonstrating listener comprehension of five concurrent variables. In *Proceedings of the 9th Audio Mostly: A Conference on Interaction With Sound*. 1–7.

[70] Kristen Shinohara and Jacob O. Wobbrock. 2011. In the shadow of misperception: Assistive technology use and social interactions. In *Proceedings of the International Conference on Human Factors in Computing Systems*. ACM, 705–714.

[71] Shraga Shoval, Johann Borenstein, and Yoram Koren. 1994. Mobile robot obstacle avoidance in a computerized travel aid for the blind. In *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2023–2028.

[72] Simone Spagnol, György Wersényi, Michał Bujacz, Oana Bălan, Marcelo Herrera Martínez, Alin Moldoveanu, and Runar Unnthorsson. 2018. Current use and future perspectives of spatial audio technologies in electronic travel aids. *Wireless Communications and Mobile Computing* 2018, 2 (2018), 1–17.

[73] Ella Striem-Amit, Miriam Guendelman, and Amir Amedi. 2012. 'Visual'acuity of the congenitally blind using visual-to-auditory sensory substitution. *PloS One* 7, 3 (2012), e33136.

[74] Ruxandra Tapu, Bogdan Mocanu, Andrei Bursuc, and Titus Zaharia. 2013. A smartphone-based obstacle detection and classification system for assisting visually impaired people. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 444–451.

[75] Ruxandra Tapu, Bogdan Mocanu, and Titus Zaharia. 2018. Wearable assistive devices for visually impaired: A state of the art survey. *Pattern Recognition Letters* 137, (2018), 37–52.

[76] Iwan Ulrich and Johann Borenstein. 2001. The GuideCane-applying mobile robot technologies to assist the visually impaired. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans* 31, 2 (2001), 131–136.

[77] Vesa Valimaki, Julian D. Parker, Lauri Savioja, Julius O. Smith, and Jonathan S. Abel. 2012. Fifty years of artificial reverberation. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 5 (2012), 1421–1448.

[78] Jan B. F. van Erp, Liselotte C. M. Kroon, Tina Mioch, and Katja I. Paul. 2017. Obstacle detection display for visually impaired: Coding of direction, distance, and height on a vibrotactile waist band. *Frontiers in ICT* 4 (2017), 23.

[79] Bruce N. Walker and Michael A. Nees. 2011. Theory of sonification. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John G. Neuhoff (Eds.). Logos Verlag, Berlin, 9–39.

[80] Eric W. Weisstein. 2004. Bonferroni correction. Wolfram Research, Inc. https://mathworld.wolfram.com/BonferroniCorrection.html.

[81] William R. Wiener, Richard L. Welsh, and Bruce B. Blasch (Eds.). 2010. *Foundations of Orientation and Mobility* (3rd. ed.). Vol. 1. American Foundation for the Blind.

[82] Jeff Wilson, Bruce N. Walker, Jeffrey Lindsay, Craig Cambias, and Frank Dellaert. 2007. Swan: System for wearable audio navigation. In *Proceedings of the IEEE International Symposium on Wearable Computers*. IEEE, 91–98.

[83] Dan Yuan and Roberto Manduchi. 2005. Dynamic environment exploration using a virtual white cane. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 1. IEEE, 243–249.

[84] John Zelek, Richard Audette, Jocelyn Balthazaar, and Craig Dunk. 1999. A stereo-vision system for the visually impaired. *University of Guelph*. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.26.1446.

[85] Limin Zeng. 2012. Non-visual 2D representation of obstacles. *ACM SIGACCESS Accessibility and Computing* 102 (2012), 49–54.

[86] Franz Zotter and Matthias Frank. 2019. *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer.

Received July 2020; revised February 2021; accepted June 2021