WatchOut: Obstacle Sonification for People with Visual Impairment or Blindness

Giorgio Presti[1], **Dragan Ahmetovic**[3], **Mattia Ducci**[2], **Cristian Bernareggi**[1], **Luca Ludovico**[1], **Adriano Baratè**[1], **Federico Avanzini**[1], **Sergio Mascetti**[1]

[1,2]Università degli Studi di Milano
Dipartimento di Informatica
[1]{firstname.lastname}@studenti.unimi.it
[2]{firstname.lastname}@unimi.it

[3]Università degli Studi di Torino
Dipartimento di Matematica
{firstname.lastname}@unito.it

ABSTRACT
Independent mobility is one of the main challenges for blind or visually impaired (BVI) people. In particular, BVI people often need to identify and avoid nearby obstacles, for example a bicycle parked on the sidewalk. This is generally achieved with a combination of residual vision, hearing and haptic sensing using the white cane. However, in many cases, BVI people can only perceive obstacles at short distance (typically about 1m, *i.e.*, the white cane detection range), in other situations obstacles are hard to detect (*e.g.*, those elevated from the ground), while others should not be hit by the white cane (*e.g.*, a standing person). Thus, some time and effort are required to identify the object in order to understand how to avoid it.

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 to a BVI user. This contribution presents *WatchOut*, a sonification technique for conveying real-time information about the main characteristics 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-centric approach, involving two iterations of online questionnaires with BVI participants in order to define, improve and evaluate the sonification technique. *WatchOut* was 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.

Author Keywords
Sonification; visual impairment; obstacle avoidance; navigation assistive technologies.

CCS Concepts
•**Human-centered computing** → **Auditory feedback;**
•**Social and professional topics** → **Assistive technologies;**

ASSETS '19, October 28–30, 2019, Pittsburgh, PA, USA

ISBN 978-1-4503-6676-2/19/10. . . $15.00

DOI: https://doi.org/10.1145/3308561.3353779

INTRODUCTION
Orientation and mobility is a challenge for blind or visually impaired (BVI) people who generally need an instructor and a lot of practice to learn how to move independently. Indeed, a number of problems arise during mobility, including: creating a mental map of the area, understanding whether there are incoming vehicles before crossing a road, detecting and avoiding obstacles along the path.

In this contribution, we focus on the last problem. Consider, for example, a BVI person walking on the sidewalk and assume there is a bicycle parked along the path. To be able to avoid the obstacle, the BVI person needs to perceive its main physical characteristics (*e.g.*, its width and height) and estimate the relative position (*e.g.*, its approximate distance and orientation with respect to a front straight direction). Currently, BVI people rely on three main senses to achieve this: residual vision, if any; hearing, in case the object itself produces noise or echo; touch, generally through the white cane.

However, many BVI people cannot rely on vision, at least in some situations (*e.g.*, during the night). Similarly, if the object is not producing any sound (like a parked bicycle), it is only possible to rely on the sound bouncing of the object, and this is possible only under certain conditions (*e.g.*, limited ambient noise). Finally, there are also limitations in the use of touch; for example, the white cane can only detect objects on the ground, it lets the BVI person perceive objects at short distance (about 1m) only, it is preferable not to hit some obstacles with the white cane (*e.g.*, a standing person, or a bicycle that can fall when hit) and some time is required to understand an obstacle size with the white cane.

A number of solutions have been proposed to address these issues, some based on dedicated hardware, others running on mainstream mobile and wearable devices. Most of these solutions focus on the problem of detecting the obstacle. While this is certainly a challenging task, there is an additional and orthogonal problem: to convey salient obstacle characteristics to the BVI user in real time, without requiring long training nor high cognitive workload.

To address this challenge, this contribution presents *WatchOut*, a sonification technique to convey obstacle characteristics in real time. *WatchOut* aims to require little training and low cognitive workload.

The user-centric design process involved two iterations with BVI people and experiments conducted at each iteration. During the first iteration, a sonification technique was defined to convey four main characteristics of an obstacle (distance, position, width, height) as distinct sound properties.

An online survey was administered to 22 BVI people to evaluate the sonification technique. Results exposed limits in the first design, hence giving us the opportunity to fix them. In the second iteration *WatchOut* was improved and this is confirmed by the results of a second questionnaire administered to 9 BVI people. A final evaluation was conducted with a mobile app that detects obstacles using state-of-the art computer vision techniques and sonifies them using *WatchOut*. Results, collected from 13 BVI participants in a real-world environment, show that the application can effectively guide the user to avoid most obstacles (85%). Subjective evaluations by the participants show that the application is considered usable (SUS score of 72.5), and the sonification is considered effective (a score of 4 or more out of 5 for 83% of the participants).

RELATED WORK

Research on obstacle detection and avoidance systems for BVI people has been conducted along two main directions: the investigation of obstacle detection systems and the design of non-visual guidance paradigms for obstacle avoidance. Several works employ sonification techniques to convey relevant information to the user through the auditory channel.

Sonification

The term "sonification" was coined as an auditory counterpart of "visualization", and refers broadly to the use of non-verbal sound to convey information. More specifically, sonification can be defined as "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation" [38]. Using non verbal sound can be advantageous over speech messages in terms of robustness to background noise, reduced cognitive load, and linguistic differences [15].

One key issue in sonification research is the definition of a set of mappings between data dimensions and auditory dimensions. In order to qualify as sonification (*i.e.*, to effectively convey information about the data), the mapping should not be completely arbitrary nor excessively complex. Auditory dimensions can be clustered into five high-level categories: Pitch-related, Timbral, Loudness-related, Spatial, and Temporal [16]. 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 [10]. As an example, the present authors have recently employed a musification approach to sonify angular direction in a wayfinding task [2, 3].

Dubus and Bresin have provided an extensive review of mapping strategies for the sonification of physical quantities [16]. By analyzing the frequency of use of physical and auditory dimensions, they showed that pitch is by far the most used auditory dimension. Moreover, spatial auditory dimensions are almost exclusively used to sonify kinematic quantities such as distance, orientation, velocity, etc.

Obstacle detection

A number of prior works propose obstacle detection based on various types of dedicated hardware, such as ultrasonic sensors to detect obstacles in front and around the user. These sensors can be mounted on the white cane [20], on the eyeglasses [18, 33], on a wearable unit [19], or on a glove [7]. An alternative solution is to use cameras and computer vision algorithms to detect obstacles in front of the user in the range of up to ten meters. They can be mounted on eyeglasses [4, 30] or on a white cane [36]. Finally, a laser range finder mounted on a white cane can be adopted to scan the frontal environment and detect obstacles [41, 13].

In recent years, obstacle detection has been achieved also through smartphone cameras [35] and depth sensors [8, 22]. Using mainstream devices implies a number of advantages, including limited costs, maintainability, a single interaction paradigm, and higher acceptance by the users [32]. Also, although ultrasonic and laser detection are very quick and accurate, thanks to the evolution of computer vision, obstacle detection based on cameras is more promising for two main reasons [23]: first, cameras are available on mainstream devices such as smartphones and wearable glasses; second, more obstacle features can be extracted from an image (*e.g.*, the color) and also non tangible objects can be detected (*e.g.*, road signs) [26, 25].

Non-visual guidance paradigms for obstacle avoidance

Non-visual paradigms investigated to enable a BVI person to detect obstacles and avoid them before bumping are: haptic and tactile feedback, speech messages and sonification.

Systems using haptic and tactile feedback convey messages on fingers through gloves with piezoelectric actuators [7, 42], on the chest through a 2-d vibration array vest [11], on the arm through an extensible wire [19] or through transcutaneous electrical nerve stimulation [27]. While these techniques provide real time feedback and enable BVI people to instantaneously get a mental map of the position and size of the obstacle, the hardware equipment is perceived as unnatural or invasive [42, 19, 27]. Moreover, gloves prevent hands-free mobility, which is essential for BVI people when holding a white cane.

Other systems adopt speech messages to guide the user around obstacles [40, 12, 4, 14, 22]. Although speech feedback is informative and easy to learn, it is a slow method to inform the user before bumping into an obstacle. Moreover, speech distracts the user from the surrounding environment because much concentration is required to decode a message [40].

Some approaches are based on the sonification of the scene in front of a walking user [31]. Such techniques are generally quicker and possibly less distracting than speech messages, but they also require longer training. For example, "the vOICe" project [28] proposes a scanning method to explore frames captured through a camera and sonified by mapping the vertical position of each pixel to frequency, the horizontal one to time and brightness to loudness. This approach requires a long training; for example, a case study shows that about 8 months training is needed to five blind persons to learn to walk along paths and properly avoid objects using "the vOICe" [34].

In order to ease the learning process, some techniques sonify only some salient features of detected obstacles. Some of them convey the position of an obstacle through spatialized sound sequences from 2 to 8 directions [18, 33] or through 3-d sound [1]. Other techniques convey additional information beyond obstacle position, including its distance from a walking user through frequency [17] or its size through volume [30] or duration [29]. These techniques can be learnt quickly (in the order of minutes or hours), but, differently from the solution presented in this paper, none of them convey synchronously all the obstacle characteristics needed to avoid it: distance, size and direction. Moreover, they were evaluated either with blind-folded people or in a simulated environment with up to five BVI people. Instead, our solution is evaluated in a real-world environment with 13 BVI participants.

FIRST ITERATION: INITIAL SONIFICATION DESIGN

The goal of our research is to develop a system capable of detecting obstacles and conveying their characteristics to BVI users. In the following we describe the initial system design.

System Requirements

During the initial analysis conducted by the team (including a BVI person) two contrasting needs emerged: on the one hand the system should provide detailed obstacle information, possibly regarding multiple obstacles; on the other hand the sonification should be easy to learn, have a low cognitive load, and should not distract the user from ambient noise (*e.g.*, an incoming car). To balance these needs, we defined two requirements. First, the system should be able to identify the most relevant obstacle and convey it to the user. Hence, if two or more obstacles are identified at the same time, the user will be informed about only one of them. Second, as previously observed in literature [43], the system should convey only relevant obstacle characteristics, which in our case are: the *distance* from the user, the horizontal *position* with respect to the user, the obstacle *width* and *height*.

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 binaural stereo signals to the user, thus permitting to exploit additional auditory dimensions in the sonification and to provide more detailed information about obstacles.

Sonification Design

WatchOut should produce a clearly perceivable and intuitively understandable characterisation of the obstacle. As a consequence, the underlying assumption of our design is that obstacles must be rendered as virtual sound sources to the user, and that the virtual sounds emitted by each obstacle carry relevant information about its physical characteristics.

A first issue to solve was to limit, as far as possible, the overlap between real-world sounds coming from the surrounding environment and the sonification added by *WatchOut*. This guideline led us to reject straightforward auditory icons for the detected obstacles, *e.g.*, a bicycle ringing bell to signal the presence of a bike.

On the other side, sonification should be intuitive, thus the mappings should be based on *ecological metaphors*[1] and variations of auditory dimensions should be consistent with those of physical parameters (e.g., obstacle horizontal position may be mapped into left/rigth sound panning, with the amount of panning controlled by the lateralization of the obstacle).

To avoid confusion with the environment, 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 one second[2]. This is clearly perceived as a synthetic sound, since this kind of timbre is very rare in nature (some harmonics are always present in physical phenomena).

A strategic decision involved continuity vs. discretization of values. For example, should we evaluate, and consequently sonify, the distance of the obstacle from the user at a high rate, or simply provide a binary near/far information when the obstacle is within predefined ranges? In order to limit cognitive overload, we chose the former strategy for what concerns *distance*, and the latter strategy for obstacle horizontal *position*, *width* and *height*. The threshold values are partially based on the work by van Erp *et al.* [37]. Consequently, the obstacle characteristics required for the sonification are:

- *distance*: continuous, from 0.1m (close) to 3m (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, right), with an obstacle classified as being on the right if its leftmost part is at least 25cm on the right (analogous for left), in order to create an easily walkable 0.5m corridor;
- *width*: discrete (small, large), where small obstacles easily fits into the walkable corridor (*i.e.*, up to 35cm), and large ones require a bigger trajectory correction;
- *height*: discrete (walkable, to circumvent), where obstacles up to 25cm of height are walkable (the user can step over them, *e.g.*, a sidewalk), and those above are to circumvent.

Consequently, in the design phase we had to select four auditory dimensions to which these four data can be mapped.

Concerning *distance*, we used a temporal auditory dimension. Specifically, we adopted the intermittent sound strategy typical of parking sensors, 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 recognisable, thanks to its widespread use in vehicles. In particular, the base sound repetition rate is mapped linearly between 50 pulses per minute (ppm) for far obstacles, up to 280 ppm for close ones. Even if this technique recalls car parking sonification, the proposed engine as a whole cannot simply be reduced to such a metaphor, since it is designed to convey also other dimensions simultaneously.

[1] In this case *ecological metaphor* means that the sonification is coherent with users's real-world sensory and cognitive experience [9].
[2] Examples for all sounds presented in this contribution are available https://watchoutobstacles.netlify.com/

In order to represent the horizontal *position* in the field of view, we adopted a spatial auditory dimension, *i.e.*, panning: in case of left or right obstacle positions the base sound is played only on the corresponding left/right channel, while for the center position it is played in both channels. Once again, the mapping is intuitive, as it is an exaggeration of the physical behaviour of the obstacle under the hypothesis that it is the sound source.

Width was associated with pitch, as larger objects typically produce lower sounds (think of violin vs. double bass timbre, or small vs. large dog barking); consequently, a high pitch (a C6 note) corresponds to a small obstacle, whereas a low one (a C4) to a large object. A two octave interval was chosen to make the pitches 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 may introduce unpredictable affective reactions.

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

Evaluation Methodology

We developed an online questionnaire to evaluate the designed sonification approach and distributed it through mailing lists and user groups for BVI people. The questionnaire was filled anonymously. The demographic data collected included sex, age (in ten-year ranges from $18-27$ to $58+$), visual impairment ("low vision" or "blindness"), duration of the condition (since birth, $<$ 5years, $5-10$years, $>$ 10years), prior musical and mobile technology experience (low, medium, high).

After an initial training with examples of sound parameters used, the participants were presented a number of audio samples produced with the proposed technique, that simulated obstacles with different characteristics. Even though the sound parameter used to map the distance characteristic is continuous, for simplicity, we consider only two distance settings: near and far. The participants were asked to listen to the audio samples, and select the correct obstacle characteristics. A final set of questions assessed the participants' opinion on the unpleasantness and comprehensibility of the presented sounds.

Results of the First Questionnaire

The questionnaire was completed by 22 BVI participants. Other 28 did not complete it, and one participant did not have visual impairments. Thus, they were discarded from the analysis. Among the participants 14 were male and 8 female. 18 participants reported being blind while 4 had low vision. The reported age was $18-27$ for 8 participants, $28-37$ for 4 of them, $38-47$ for 3, $48-57$ for 5, and over 58 for 2. 11 participants were visually impaired since birth, 3 for less than 5 years, 4 between 5 and 10 years, and 4 for more than 10. "High" musical expertise was reported by 4 participants, 8 reported "Medium", and 10 reported "Low" expertise. In terms of mobile technology expertise, 1 participant selected "Low", 6 selected "Medium" and 15 selected "High".

The accuracy for a given characteristic is measured as the ratio of correctly categorized sounds by the participants for that characteristic. Instead, the global accuracy is the ratio of sounds correctly categorized for all four characteristics. The global accuracy for the first questionnaire was 0.34 ± 0.17[3]. To interpret the low global accuracy score, we consider the accuracy scores for single characteristics, shown in Figure 1a. The analysis reveals that height and width characteristics had a much lower accuracy (0.65 ± 0.12 and 0.64 ± 0.09 respectively) compared to distance and position characteristics (0.92 ± 0.05 and 0.92 ± 0.04 respectively).

This was also reflected in the final observations which measured the sound comprehensibility and unpleasantness on a 1-5 Likert-like scale, with open ended questions to motivate the score. Prior work motivates the usage of the mean for Likert-like scores [21]. Thus we will continue using the same notation[3] for these scores. Participants on average selected 3.37 ± 1.20 for the comprehensibility and 1.56 ± 0.89 for the sound unpleasantness. While the scores were overall positive, 7 participants reported problems in correctly understanding width and height characteristics, 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 characteristics. Specifically, participants under 38 years of age had significantly lower (0.56 ± 0.14) accuracy scores for width characteristic compared to others (0.71 ± 0.20), measured using Mann-Whitney U test with Bonferroni correction[4] [39] ($U = 57.0$, $p < 0.05/4$). Conversely, participants who were 38 and above had significantly lower (0.51 ± 0.18) height characteristic accuracy scores ($U = 19.0$, $p < 0.01/4$) compared to others (0.76 ± 0.13). Low vision participants had lower accuracy scores for all 4 characteristics (width: 0.56 ± 0.14, height: 0.58 ± 0.19, distance: 0.80 ± 0.20, position: 0.84 ± 0.15) compared to 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 had consistently higher accuracy scores for width (0.84 ± 0.26), distance (1.0 ± 0.0) and position (0.93 ± 0.12) characteristics, compared to 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, $U = 80.0$, $p < 0.05/4$).

SECOND ITERATION: IMPROVED SONIFICATION

Design of the improved sonification

The evaluation of the first iteration showed that *height* and *width* characteristics are harder to grasp, and present a high variability across demographic groups (age and disability). While this seems to be mitigated by musical expertise, we cannot expect the general populace to have such skills.

[3] We use *mean*$\pm$*standard deviation* notation.

[4] We report the significance level as α/m where α is the desired overall significance level and m is the number of multiple comparisons.

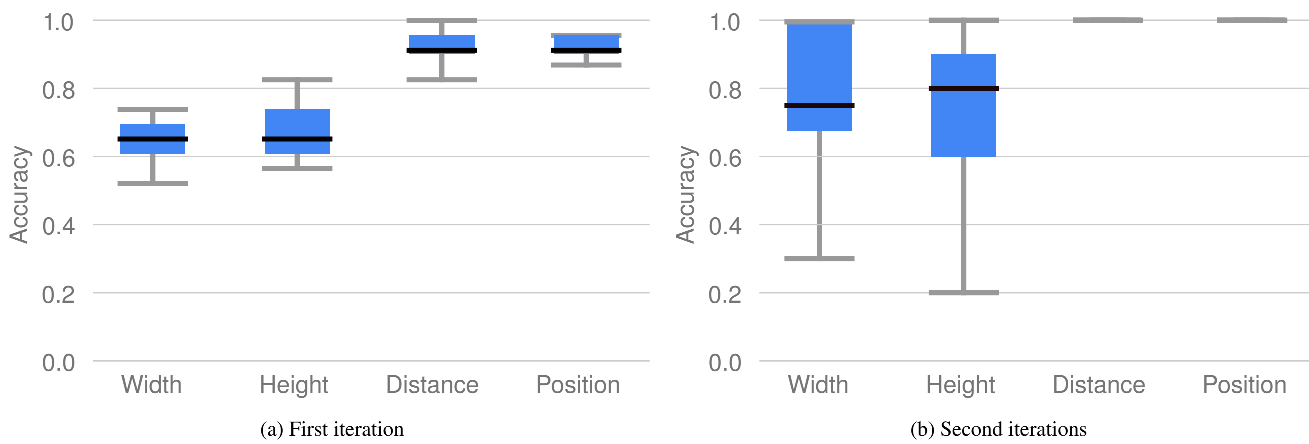


Figure 1: Accuracy of the four obstacle characteristics in the first and second iteration.

In particular, experimentation showed that base sound pitch was intuitively linked to the obstacle's vertical position (this kind of association seems innate, at least in Western culture [5]), so it was a more suitable dimension for sonifying *height*. We thus decided to use the base sound pitch to convey this kind of information, by linking the low pitch to a walkable obstacle, and the high one to the obstacles to circumvent.

Therefore, the *width* characteristic needs to be mapped to an auditory dimension that is clearer and less prone to misunderstanding. Leaving loudness aside, which is hard to evaluate in a noisy context and difficult to calibrate on different devices, the choice fell on polyphony (a higher-level musical feature). In particular, a single-pitch was associated with a small obstacle, while a three-note chord (*i.e.*, a cluster) with larger obstacles.

The chord chosen 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 characteristics, 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.

Results of the Second Questionnaire

We repeated the online survey study with the improved sonification approach. The evaluation methodology replicated the one from the previous questionnaire. However we added as an additional prerequisite that the participants should have not previously filed the first questionnaire, in order to avoid possible learning effects. For the second questionnaire, we collected 9 answers from BVI people. This time, there were no incomplete questionnaires, however one participant reported he doesn't have any visual impairment and was discarded from the analysis.

Of the BVI participants, 4 were male, and 5 female. Participants in the $18-27$ age group were 4, 2 were in the $28-37$ range, 1 was $38-47$, 2 were $48-57$, there were no over 58. 5 participants reported being blind, while 4 defined their sight condition as Low vision. Participants who were visually impaired since birth were 6, none of them were visually impaired for less than 5 years, 1 between 5 and 10 years, and 2 for more than 10 years. The self assessed musical expertise reported by the participants was "High" for 3, "Medium" for 5, and "Low" for 1. 1 participant considered his mobile technology expertise to be "Low", 4 selected "Medium" and 4 "High".

The perceived comprehensibility of sounds remained substantially the same (3.22 ± 1.09) in the second questionnaire. However, considering the actual results obtained by the participants in terms of accuracy for all four characteristics, the score for the improved sonification was 0.53 ± 0.13. The improvement was significant for each characteristic (see Figure 1b): for height it improved to 0.70 ± 0.30 ($U = 82.0$, $p < 0.5$), for distance to 0.99 ± 0.03 ($U = 27.0$, $p < 0.01$), for width to 0.74 ± 0.25 ($U = 71.0$, $p < 0.5$), and for position to 0.99 ± 0.04 ($U = 24.0$, $p < 0.01$). We believe that, with the improved understanding for height and width characteristics, also other characteristics were easier to disambiguate, and therefore yielded better accuracy scores. However, the improved understanding of the new sound properties comes at the cost of a higher unpleasantness score reported by the participants (2.88 ± 1.05).

We also notice that the variability between different demographic groups, highlighted in the results of the first questionnaire, tend to disappear with improved overall accuracy. Indeed, with respect to the first questionnaire, 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, and also in the height score accuracy for participants who were 38 and over (0.77 ± 0.35, $U = 59.0$, $p < 0.01$). However, no significant differences were found between those groups in the second questionnaire. Similarly, no differences were found between low vision and blind participants, as well as those with different musical expertise levels.

EVALUATION IN THE REAL WORLD

System Prototype

In order to test the effectiveness of *WatchOut* in the real world, we developed an iOS application that detects obstacles and sonifies them to the user. The app is logically organised in two main modules: one to detect the obstacle, the other to sonify their characteristics. The sonification is implemented with the AudioKit framework[5] according to the design described in the previous sections, with the following adaptations:

- to signal the user that the sonification is active but no obstacles are in sight, an idle sound was added: a quiet pink noise sound[6] modulated by a 1s fade in and out, interleaved by 5 seconds of silence. This sound is turned off and 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 12dB to compensate the frequency response of the device.

Figure 2 provides a high-level block diagram of the implementation of the sonification engine. Hexagons in the rightmost column contain the main characteristics extracted from the visual scene. These act as controllers of audio modules (rectangles), which generate (*sinusoid*, *impulse*, and *pink noise*) or modify (*filter* and *volume envelope*) sound. Finally, thanks to adders and switches (circles and diamonds), signals are routed to the audio output.

Even if the focus of the study is the sonification technique, it is worth providing a description of the obstacle detection system. This module relies on Apple's ARKit 2[7], a framework to support augmented reality that exposes a number of high level computer vision primitives to the developer. A particularly relevant aspect is that these functions ease the process of converting the position of a 2D point framed by device camera into a 3D position in a given reference frame. Since the user's position and orientation in the reference frame are known, it is easy to obtain the 3D position of a point in the reference frame centred on the user. For this reason, in the following, when we refer to any geometrical object (point, plane, etc.), we assume that we know its 3D position with respect to the user.

The obstacle detection module takes in input two main sources of information, both provided by ARKit: planes and point cloud detection. Considering the former, ARKit detects vertical and horizontal planes; the lower horizontal plane is the ground plane where the user is walking on, so it should not be detected as an obstacle. The other planes (both horizontal and vertical) should instead be considered as obstacles. Figure 3a shows the ground plane, and three other planes: two vertical and a horizontal one. In addition to planes, there are other objects that should be detected as obstacles. Figure 3b shows an example: the bicycle is not identified as a plane (which is intuitively correct), but it is still an obstacle that the detection module should detect.

[5]https://audiokit.io/

[6]A broadband noise whose power is inversely proportional to the frequency, reminding the sound of a distant waterfall

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

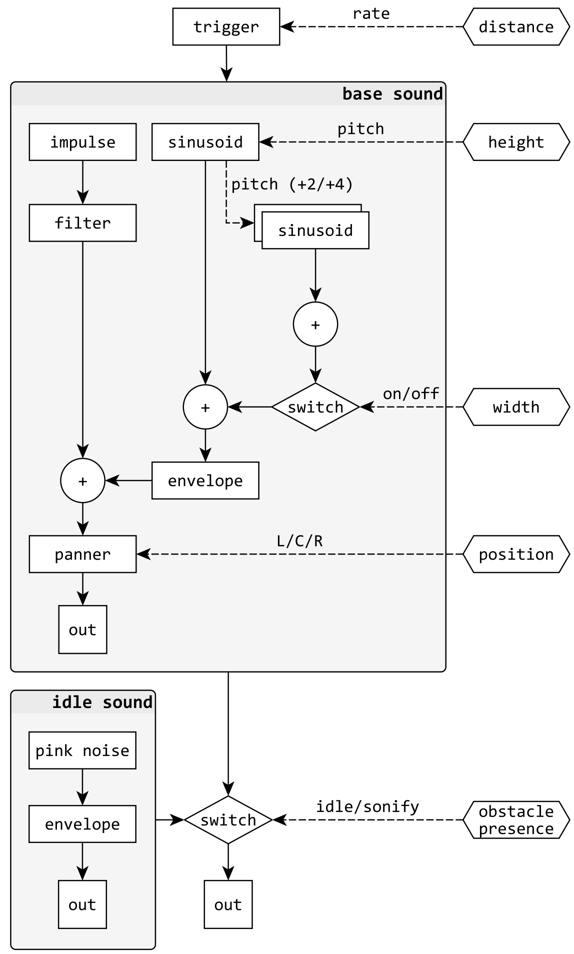
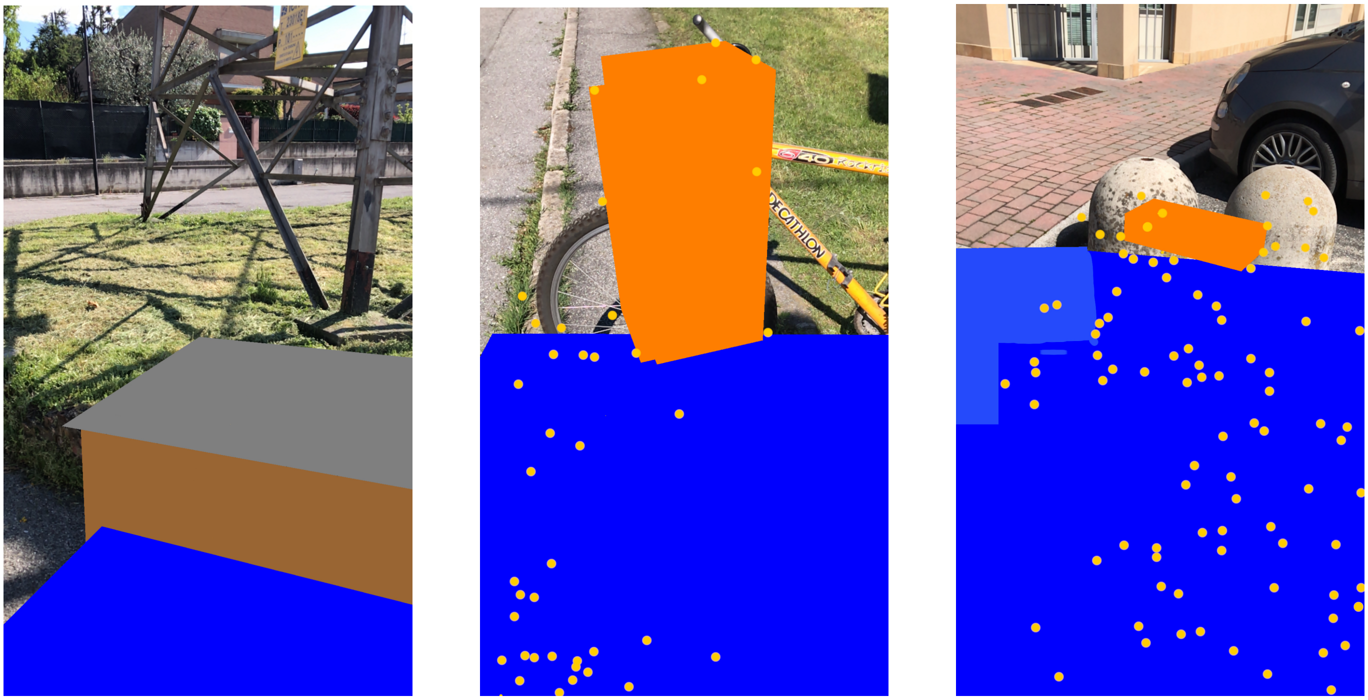


Figure 2: A block diagram of the sonification engine. Rectangular boxes are audio modules generating or modifying sound; hexagons represent controllers; solid lines represent audio data; dashed lines represent control signals.

For these cases the detection module relies on point cloud information from ARKit. Point cloud is a sparse set of *feature points*, each representing a notable feature in the image[8] and generally positioned on the objects contours. ARKit provides point clouds at a tunable framerate. In our experiments we used a frequency of 30Hz. Thanks to a function (called *hit test*) exposed by ARKit[9], we can detect feature points belonging to a plane (*e.g.*, the yellow points on the blue plane in Figure 3b); these points are ignored in the following computation as the obstacle they represent has already been considered. In principle, all remaining feature points represent obstacles. However, the computation of these points is subject to noise, so classifying each of them as an obstacle would result in a high number of false positives. For this reason, our detection module first aggregates feature points in each frame to form *candidate objects* and then validates them across multiple frames.

[8]https://developer.apple.com/documentation/arkit/arframe/2887449-rawfeaturepoints

[9]https://developer.apple.com/documentation/arkit/arscnview/2875544-hittest



(a) Three planes detected by ARKit: the ground plane (blue), one higher horizontal plane (gray) and a vertical plane (brown).

(b) A bicycle is detected as a candidate obstacle (also note that features points in the ground plane are pruned).

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

Figure 3: Examples for the obstacle detection procedure.

All feature points provided by ARKit at a given frame (except those belonging to a plane) are grouped to form candidate objects: two feature points are associated with the same candidate object if their distance is bounded by a threshold (50cm in our experiments). The intuition for this choice is that even if the two points actually belong to two distinct objects (like in Figure 3c) these objects are close by and the user cannot walk between them, so it is convenient to recognise them as a single obstacle.

To combine candidate objects among different frames, the temporal domain is divided into non-overlapping consecutive windows (of 0.75s in our experimental setup). The detection module counts how many times each candidate object overlaps with other candidate objects from other frames in the same window. If this count is larger than a threshold (2 in our setup), the object is *validated*.

The last step of the recogniser module identifies the obstacle that is more relevant for the user. To do so, at the end of each window, planes and validated objects are sorted according to a total order relation that takes into account which obstacle is more urgent to sonify. Briefly, obstacles in front of the user have higher priority than those on the sides and, among those in front of the user (or on the sides), the closer obstacle is the most relevant. The obstacle with higher priority is given to the sonification module.

Experimental Setting

We conducted a user study with 13 participants navigating on predefined outdoor paths which contained obstacles of various kinds. The aim of the investigation was not to assess the accuracy of the obstacle detection system, but rather to evaluate the participants' opinion on the functionality and in particular on the proposed sonification technique. Thus, we implemented the study as a think-aloud protocol and we collected the participants' subjective feedback through Likert-like scale questions including System Usability Scale (SUS) questionnaire (see Table 4a) and a number of questions related to the proposed sonification technique.

Participants

We recruited 13 participants, 7 male and 6 female, by word-of-mouth and through mailing lists. None of the participants was included in the preliminary surveys, however $P10$ participated in the first online questionnaire. The participants' demographic information, reported in Table 1, included the same information as the online questionnaires.

All low vision participants walked without assistive tools, while all blind participants used a white cane, excluding $P11$ who requested to walk with the experimenter without using the cane. $P10$ and $P12$ similarly requested to walk with the experimenter for safety concerns, but also used the cane. In these cases the experimenter always stood at the side of the participant, holding the participant's arm and letting the participant act as a guide. The experimenter would only halt the participant in case of danger.

ID	Sex	Age	Visual Impairment		Experience with		
			Condition	Since	Music	Tech.	Area
P1	M	28-37	Low vision	birth	High	Low	Some
P2	M	48-57	Low vision	birth	Med.	Low	NA
P3	M	28-37	Blind	birth	High	Med.	Some
P4	M	48-57	Blind	10+y	Med.	Low	Yes
P5	M	58+	Blind	5-10y	Med.	High	Some
P6	M	18-27	Low vision	10+y	High	Med.	Yes
P7	F	38-47	Blind	10+y	Low	Low	Yes
P8	F	58+	Blind	10+y	Low	Low	Yes
P9	F	38-47	Blind	10+y	Med.	Med.	Yes
P10	F	38-47	Blind	birth	Med.	Med.	No
P11	F	18-27	Blind	birth	Low	High	Some
P12	M	18-27	Blind	10+y	Med.	Med.	Yes
P13	F	18-27	Low vision	birth	High	High	Yes

Table 1: Participants' demographic information.

All participants excluding $P10$ had prior experience of at least a portion of the testing environment, however they did not know of the presence and positions of specific obstacles. $P2$ did not complete the post questionnaire and therefore we do not know the area expertise for this participant.

Testing Environments

Due to different geographic locations of the participants and experimenters, a number of outdoor testing environments were selected to be convenient to reach. To present consistent visual conditions for the obstacle detection system, all tests were conducted during sunny days. All environments were selected with the following constraints:

- Sidewalk or a similar pedestrian-safe area of overall length around 500m.
- No changes in altitude because they may be detected as obstacles by the obstacle detection approach.
- Obstacles unknown to the participant (parked cars, traffic bollards, curb, walls on the side and architectural barriers).

Apparatus

For the experiments we use an iPhone 7+ device with the obstacle detection prototype installed and "Aftershockz Bluez" bone conducting earphones. The smartphone is positioned in portrait mode and attached on the chest of the participant using a chest strap. The chosen device has a vertical field of view of 65.5° (46.4° horizontal). Therefore, the device is oriented with an inclination of 60° with respect to the ground plane so that the phone has a good view range of all the possible obstacles from the participants' feet onward. The app measures the inclination of the device during the setup phase and notifies the experimenter if the inclination is correct using visual and verbal messages. Once the direction of the device is correct, the experimenter asks the participant to slowly rotate the torso left and right in order to calibrate the ground plane detection. Once the ground plane is detected and appears on the device screen, a verbal message notifies that a test can start.

Protocol

After meeting at a mutually convenient place, the experimenter collects participant's demographic data and explains the goal and the structure of the experiment. This information includes how the test app works, the description of the sonification approach and its semantics. The participant is reminded to voice thoughts and feelings during the tests.

The participant is taken to the training location to acquire confidence with the app and the sonification approach. The training consists in exploring 4 predefined obstacles (a wall, a traffic bollard, a person, and stairs of at least 3 steps), one at the time, with the testing app. The experimenter notifies the participant of the obstacle type and instructs the participant to explore the obstacle with the cane or by hands. Then, the app is started and the participant is advised to explore the obstacle again, this time with the app, from different distances and angles for at least a minute. From a safe position in order not to influence the detection but to be able to intervene if necessary, the experimenter makes sure that the participant has explored different sonification configurations, or otherwise directs the participant to explore them. This training phase lasted up to 5 minutes for each participant.

Afterwards, the experimenter leads the participants to the starting point of a route, opportunely chosen before the experiment to satisfy testing environment criteria. The experimenter starts the test and provides verbal instructions for the predefined path. The experimenter follows the participant (or walks by the participant for those who requested it), annotating which obstacles were correctly avoided (True Positives – *TP*), the situations in which the participants attempted to avoid an obstacle which was not present (False Positives – *FP*), and which obstacles were not avoided (False Negatives – *FN*), in which case the experimenter stops the user before bumping into the obstacle. The participants were informed that due to changes in light and shadows sporadic *FP* detections may happen, but that those should not persist more than an instant and therefore to ignore detections that do not appear consistent through time. After each obstacle, the experimenter reports what it was to the participant. The test stage lasts until the end of the path or for 15 minutes.

After the test stage, the experimenter presents the final questionnaire to the participant. This questionnaire includes SUS evaluation of the obstacle detection system and Likert-like scale questions measuring the participants' opinion on the sound pleasantness, effectiveness and the capability to distinguish the four sound characteristics and the corresponding obstacle properties. Open ended questions investigate participants' opinion on better/worse detected obstacles, and suggestions for improvements of the detection system and the sonification approach.

Experimental Results

We report the quantitative analysis of the obstacles detected on the path, the analysis of SUS scores and sonification questionnaires, and participants' observations on the system and on the sonification approach.

Analysis of the Obstacle Detection

On average, the participants encountered 12.77 ± 4.66 obstacles. Among those, 11 ± 3.76 were correctly avoided (*TP*), while 1.77 ± 1.69 were not (*FN*). Additionally, on average 2.46 ± 1.85 were the occurrences in which participants attempted to avoid obstacles when none was present (*FP*). This mostly happened due to sporadic false detections of shadows as objects, which did not persist in the following frames, but participants avoided anyway due to safety concerns.

We analyze the precision $p = TP/(TP+FP)$ and recall $r = TP/(TP+FN)$ metrics, commonly used to measure the impact respectively of *FP* or *FN* on the detection accuracy. The closer these scores are to 1.0, the less is the effect of *FP* or *FN* with respect to *TP*. On average, the precision of the obstacle avoidance is $p = 0.82 \pm 0.11$, while the recall score is $r = 0.87 \pm 0.11$. We do not detect any statistically significant difference among these scores with respect to different demographic characteristics of the participants.

SUS and Sound Characteristics

We analyze the SUS scores registered by the participants (see Figure 4). We recall that *P*2 did not complete the post questionnaire and therefore was excluded from this and the following analyses. The average SUS score for all the participants was 72.5 ± 13.23, indicating a "Good" usability score [6]. There were no significant differences in the overall SUS or single score with respect to different demographic characteristics. Instead, regarding single scores, a significant difference was found between low vision and blind participants for questions 1 and 7. Specifically, for question 1, low vision participants had a score of 2.67 ± 0.58, whereas blind participants averaged 3.78 ± 0.83 ($U = 5.0$, $p < 0.05/4$). Instead, for question 7, low vision participants had a score of 5.0 ± 0.0, whereas blind participants averaged 3.67 ± 1.22 ($U = 3.0$, $p < 0.05/4$).

On a Likert-like scale from 1 to 5, the system was found to be overall useful, scoring 4.25 ± 0.72. Sound unpleasantness was 2.33 ± 1.18 on average, while the perceived effectiveness of the sounds was 4.25 ± 0.92. None of the scores resulted significantly different among diverse demographic groups.

In terms of the perceived easiness to distinguish diverse sound properties and corresponding obstacle characteristics, position scored an average value of 4.33 ± 0.98, distance 4.08 ± 0.9, height 3.5 ± 1.45 and size 3.5 ± 1.38. No significant differences between the different characteristics were registered using a Kruskal-Wallis test.

Participants' Comments and Observations

Think aloud notes and open ended questions were analyzed, revealing that in general participants appreciated the system, but they also highlighted limitations and proposed improvements for the approach. Some of the obstacles were easier to deal with than others. Specifically walls were easier to avoid for 5, flower pots for 3, cars and poles for 2 participants. On the other hand stairs and traffic bollards were harder to address for 3 and 2 participants respectively. However those opinions were not shared by all participants. For example *P*12 found cars harder to avoid, while *P*9 had more problems with walls. *P*4 reported that obstacles which generated inconsistent height sonification were harder. This could happen for tall objects such as poles, if the upper portion is not detected in some frames. Also *P*9 and *P*11 requested better management of concurrent obstacles at different heights. For *P*5 very tall obstacles such as branches, were hard to understand and avoid. On this matter *P*1, *P*4 and *P*7 also voiced the desire for specific head-level obstacle detection, given the inherent danger such obstacles represent [24]. Instead, *P*11 reported difficulties for objects in movement (*e.g.*, passing cars).

Seven participants expressed the desire for vibration as an additional feedback modality, using the phone (*P*3, *P*7, *P*11, *P*13), bracelets (*P*1), cane (*P*10) or directly on bone conducting headphones (*P*1, *P*12). Five participants wanted more verbal cues such as names of specific obstacles (*P*4, *P*7, *P*11, *P*12), or quantities such as distance (*P*10). Three participants requested configurable sounds, such as musical instruments (*P*5, *P*9) or tunes (*P*6), to associate to obstacles or their characteristics, and *P*1 suggested to have the frequency of notifications configurable related to obstacle type. *P*5 and *P*7 would like more accurate distance measurements, while *P*3 and *P*13 wanted higher responsiveness, and *P*13 was interested in the possibility of adding a night vision modality, useful for some low vision users.

DISCUSSION

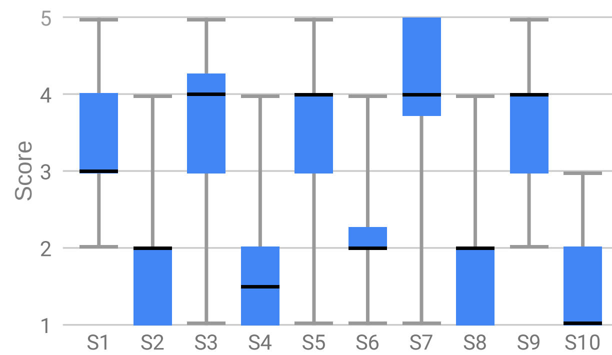
The data we collected from the three experiments (two questionnaires and the real world evaluation) can help us better understand *WatchOut* and its application in the system prototype we developed. We now discuss four main points.

1) During the second iteration we attempted to improve how *WatchOut* sonifies obstacle width and height. Results show that we achieved our objective, proving that the new mappings are based on metaphors that are cognitively more robust. Moreover, the effectiveness of the sonification for the two remaining obstacle characteristics (distance and position) was also improved. This was unexpected, as we did not change the mapping of these characteristics into sound. We believe that this result is due to the fact that the sonification experimented in the second iteration is globally more intuitive and less cognitive demanding; consequently all the characteristics are more easily understood.

Despite the improved results in the second iteration, global accuracy is 0.53 ± 0.13, meaning that it is still hard for the users to correctly identify all four characteristics at the same time. In particular, considering that distance and position are almost always correctly detected, there are still problems with height and width. We believe that this is mainly due to the very limited learning time. However, while this is a limitation of our current solution, we should also observe that it does not affect the effectiveness of *WatchOut* in the real world as we showed that most of the obstacles can be avoided thanks to our prototype. A possible interpretation for this result is that, given precise information about obstacle distance and position, the two remaining characteristics can be deduced by combining the audio feedback with the context and other senses (e.g., residual sight).

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.

Figure 4: SUS questions and scores for the system.

2) Another difference between the results in the two iterations is that, in the second one, participants considered the sound to be more unpleasant ($1,56$ on average in the first, $2,89$ in the second). The values in the second iteration are consistent with those obtained in the real world evaluation (average value of $2,33$). This suggests that, while we succeeded in making the sound more understandable, we also made it less pleasant for the participants. 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, the employed clusters are highly dissonant, being composed of three contiguous tones.

3) In the real world evaluation we observed a large majority of true positives, but also some false negatives and false positives. Considering the experimental design, in principle we are unable to know whether a false positive or negative is due to a wrong detection or to a wrong interpretation of the sonified instruction by the user. However in most of the cases the experimenter was able to see the device screen showing debug information (similar to those depicted in Figure 3). In these cases the experimenter observed that the errors were most often due to a wrong detection. This was expected, as the obstacle detection is an early prototype and these results show that it actually needs to be improved. However, at the same time, these results suggest that *WatchOut* can effectively convey obstacle characteristics to the user when the detection module provides correct results.

4) Considering the SUS scores in the real world evaluation, we observed statistically significant differences between blind and low vision participants. In particular blind participants more often reported they would use the system frequently (SUS question 1). Our interpretation is that the system is more relevant for blind users because low vision users can use residual vision and therefore mobility is generally easier for them. Low vision participants also find it easier to learn to use the system compared to blind participants (SUS question 7). One possible interpretation for this result is that, by using residual vision, these participants can easily associate what they can see with the feedback they receive from *WatchOut*, hence the system results easier to learn for them.

CONCLUSIONS AND FUTURE WORK

This paper presents *WatchOut*, a sonification technique aimed at conveying obstacles information to BVI users in real time. The experimental results show that *WatchOut* can convey obstacle distance and position (center/left/right) with almost perfect accuracy (above 0.99), while it is less effective to convey obstacles height and width (accuracy of 0.71 ± 0.29 and 0.76 ± 0.23, respectively). Despite this, when applied in the real word, *WatchOut* is effective in guiding users to avoid obstacles.

A number of ideas for future work emerged from the comments by the participants to the experiments. In particular, participants frequently suggested adding vibration feedback, possibly pairing sonification with verbal cues (including names of specific obstacles), allowing the users to personalize the interaction and the sounds used, and tackling the problem of head level obstacles, which are also recognized as particularly dangerous in prior literature [23].

Considering the sonification technique, we intend to investigate how to further improve the accuracy for the height and width parameters. This includes evaluating how learning can affect this metric. Indeed, in this paper users trained for few minutes only, and we suspect that a slightly longer training could substantially improve the accuracy.

We also intend to better investigate the sonification effectiveness in the real world, for example to estimate whether ambient noise can affect it. This can be achieved in two different ways. First, by conducting experiments in a controlled environment where a synthetic sound is produced to emulate urban noise. A complementary solution is to conduct the evaluation in a real urban environment, being able to distinguish errors caused by the detection technique from those resulting from the sonification technique.

In the future we also intend to further investigate the recognition module, improving the accuracy and robustness of the prototype we used in this contribution to test *WatchOut* in the real world.

REFERENCES

[1] Daniel Aguerrevere, Maroof Choudhury, and Armando Barreto. 2004. Portable 3D sound/sonar navigation system for blind individuals. In *2nd LACCEI Int. Latin Amer. Caribbean Conf. Eng. Technol. Miami, FL.*

[2] Dragan Ahmetovic, Federico Avanzini, Adriano Baratè, Cristian Bernareggi, Gabriele Galimberti, Luca A. Ludovico, Sergio Mascetti, and Giorgio Presti. 2018. Sonification of Pathways for People with Visual Impairments. In *Proc. Int. ACM SIGACCESS Conf. on Computers and Accessibility (ASSETS'18)*. Galway, 379–381.

[3] 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 *Proc. IEEE Int. Conf. on Pervasive Computing and Communications (PerCom 2019)*. Kyoto, 332–341.

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

[5] 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, USA, 64–71.

[6] 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, 114–123.

[7] M Bouzit, A Chaibi, KJ De Laurentis, and C Mavroidis. 2004. Tactile feedback navigation handle for the visually impaired. In *ASME 2004 International Mechanical Engineering Congress and Exposition*. American Society of Mechanical Engineers, 1171–1177.

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

[9] Egon Brunswik and Joe Kamiya. 1953. Ecological cue-validity of 'proximity' and of other Gestalt factors. *Am. J. Psych.* 66, 1, 20–32.

[10] Allan D. Coop. 2016. Sonification, Musification, and Synthesis of Absolute Program Music. In *Proc. Int. Conf. Auditory Display (ICAD2016)*. Canberra.

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

[12] 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, 25–35.

[13] 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, 95.

[14] 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, 1575–1589.

[15] Tilman Dingler, Jeffrey Lindsay, and Bruce N. Walker. 2008. Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech. In *Proc. Int. Conf. Auditory Display (ICAD2008)*. Paris.

[16] Gaël Dubus and Roberto Bresin. 2013. A Systematic Review of Mapping Strategies for the Sonification of Physical Quantities. *PLoS ONE* 8, e82491. Issue 12.

[17] José L González-Mora, A Rodriguez-Hernandez, LF 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 *International Work-Conference on Artificial Neural Networks*. Springer, 321–330.

[18] Tohru Ifukube, Tadayuki Sasaki, and Chen Peng. 1991. A blind mobility aid modeled after echolocation of bats. *IEEE Transactions on biomedical engineering* 38, 5, 461–465.

[19] 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 *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. ACM, 1483–1488.

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

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

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

[23] Roberto Manduchi. 2012. Mobile vision as assistive technology for the blind: An experimental study. In *International Conference on Computers for Handicapped Persons*. Springer, 9–16.

[24] 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, 44–54.

[25] Sergio Mascetti, Dragan Ahmetovic, Andrea Gerino, and Cristian Bernareggi. 2016a. Zebrarecognizer: Pedestrian crossing recognition for people with visual impairment or blindness. *Pattern Recognition*.

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

[27] Simon Meers and Koren Ward. 2005. A substitute vision system for providing 3D perception and GPS navigation via electro-tactile stimulation.

[28] Peter BL Meijer. 1992. An experimental system for auditory image representations. *IEEE transactions on biomedical engineering* 39, 2, 112–121.

[29] Kawin Metsiritrakul, Atiwong Suchato, and Proadpran Punyabukkana. 2017. Obstacle Avoidance Feedback System for the Blind using Stereo Sound. In *Proceedings of the 11th International Convention on Rehabilitation Engineering and Assistive Technology*. Singapore Therapeutic, Assistive & Rehabilitative Technologies (START) Centre, 21.

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

[31] Rajib Sarkar, Sambit Bakshi, and Pankaj K Sa. 2012. Review on image sonification: a non-visual scene representation. In *Recent Advances in Information Technology (RAIT), 2012 1st International Conference on*. IEEE, 86–90.

[32] Kristen Shinohara and Jacob O Wobbrock. 2011. In the shadow of misperception: assistive technology use and social interactions. In *C. on Human Factors in Computing Systems*. ACM.

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

[34] 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, e33136.

[35] 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.

[36] 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, 131–136.

[37] 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, 23.

[38] 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.

[39] Eric W Weisstein. 2004. Bonferroni correction.

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

[41] Dan Yuan and Roberto Manduchi. 2005. Dynamic environment exploration using a virtual white cane. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 1. IEEE, 243–249.

[42] John Zelek, Richard Audette, Jocelyn Balthazaar, and Craig Dunk. 1999. A stereo-vision system for the visually impaired. *University of Guelph* 1999.

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