Sonification of Rotation Instructions to Support Navigation of People with Visual Impairment

Dragan Ahmetovic[1], Federico Avanzini[2], Adriano Baratè[2], Cristian Bernareggi[2],
Gabriele Galimberti[3], Luca A. Ludovico[2], Sergio Mascetti[2], Giorgio Presti[2]

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

[2,3]Università degli Studi di Milano
Dipartimento di Informatica
[2]{firstname.lastname}@unimi.it
[3]gabriele.galimberti2@studenti.unimi.it

***Abstract*—Indoor navigation services for people with visual impairment are being researched in academia, and working systems have already been deployed in public places. While previous research mainly focuses on computing the user's position with high accuracy, providing non-visual navigation instructions is also a challenge and naive approaches can fail in helping users reach their target destination or even expose them to hazards.**

In this paper we investigate the problem of guiding users to rotate towards a target direction. We propose three different sonification techniques that provide continuous guidance during rotation, and we compare them to a single-impulse baseline, used in previous works. We also explore three variations that reinforce the proposed techniques by combining them with the baseline. A preliminary study with 10 blind participants highlights two dominant techniques, which we analyze through a follow-up study with 18 participants, from 2 groups with very distant cultural backgrounds. While stark differences emerge in the performance from the two groups, we highlight two clear results common to both: 1) one of the proposed techniques significantly outperforms the baseline, reducing the average rotation error by a factor of 3.5 (from 11° to 3°); 2) the interaction speed of this technique, generally slower than the baseline, significantly improves when combined with the baseline technique.

***Index Terms*—visual impairment, mobility, sonification.**

I. INTRODUCTION

Indoor navigation is an important service when visiting an unknown environment, for example making it faster to reach a destination. For blind or visually impaired (BVI) people, this is even more important: beyond shortening the route and reducing the effort, the solution enables BVI people to reach the destination without assistance and helps them avoiding hazards. The importance of this service for BVI people has been recognized in the literature and a number of existing works propose systems specifically designed for indoor navigation. The main focus of these papers is on how to accurately compute user's position. This is due to the fact that BVI people are less accurate in compensating for an imprecise location, being less skilled in constructing a mental map of the environment. Hence the position should be computed with higher accuracy with that required for sighted people.

In this paper we investigate an orthogonal challenge: how to *guide* a BVI person during navigation in an indoor environment. Existing navigation systems (either indoor or outdoor) share a similar approach to convey navigation instructions to sighted users: the interface is mainly based on graphical information representing the map and the route. However, no consolidated interface has emerged so far for BVI people.

Following a common practice, we model a route as a sequence of nodes (intersections) connected by line segments (straight paths). Consequently, we identify two types of instructions: when the user is at an intersection and needs to rotate towards the next line segment, and when the user is walking along a straight path connecting two nodes. In this contribution we address the former type. Simple verbal instructions are not sufficient for BVI people as they do not convey quantitative information. Indeed, in a previous work, we observed that even coupling verbal instructions with a simple sonification technique is not effective in guiding BVI users during rotations: experimental results show that 5% of rotations either led the user to require assistance or forced the supervisor to stop the experiment in order to avoid danger [1].

We describe two sonification techniques, each one with two variations. Subsequently we show experimental results, conducted with 18 BVI participants, comparing these solutions among themselves and with a benchmark technique adopted in previous work. Results show that all the proposed techniques outperform the benchmark solution in terms of rotation error. In particular, two of them reduce the average rotation error by a factor of 3, compared to the benchmark solution, a difference that also emerges to be statistically significant.

II. RELATED WORK

A. Navigation Systems for BVI people

Numerous assistive technologies have been researched to support independent mobility for BVI people. Among these, smartphone-based tools are becoming increasingly popular [2] due to their sensing and computational capabilities and the presence of native accessibility features, such as screen readers and magnification. These allow BVI people to interact with most applications, including navigation tools such as maps and Global Positioning System (GPS) navigation [3].

Specifically designed tools have also been studied. Some approaches rely on smartphone sensors to perceive the sur-

---

Identify applicable funding agency here. If none, delete this.

rounding environment and inform the user of the features of interest. Computer vision is used to detect visual cues, such as pedestrian crossings [4] or traffic lights [5], and notify the user. Approaches using Inertial Motion Unit (IMU) sensors have also been proposed to detect the traversed path and assist the user on the way back [6].

Other techniques provide information sourced from online databases. Geographical Information Systems (GIS) are used to provide nearby points of interest [7], [8], while street level imagery is employed to crowdsource accessibility information [9]. Computer vision analysis of satellite and street-level images is also used to detect mobility cues [10].

Turn-by-turn navigation is a guidance paradigm which translates a route into a sequence of straight paths and turning points. It is commonly used by sighted users for outdoor vehicular guidance using GPS. This approach is also useful for BVI people because no prior knowledge of the environment is needed to follow navigation instructions [11]. Due to the low accuracy of GPS systems in indoor environments, alternative methods such as WiFi [12], Visual Light Communication (VLC) [13], and Bluetooth Low Energy (BLE) beacons [14] have been studied to provide meter-level guidance, which is considered to be sufficiently accurate for BVI people [14].

However, even with high accuracy, localization errors in turn-by-turn navigation are still possible [15]. In addition to system errors, user behavior [16]–[18] can also impact navigation accuracy and cause errors such as veering [19] or imprecision during rotation [20] for both sighted [21] and blind [22] individuals.

Such errors are commonly compensated through visual inspection [23]. For BVI people, these can impact the navigation outcome or even endanger the user [1].

In particular, turning errors during navigation assistance are related not only to *encoding* (i.e. understanding the correct angle to turn), but also to *execution* (i.e. reproducing the correct turning angle) [20]. In particular, it has been argued that rotation angles are encoded at 90° rate [24]. Indeed, such angles were shown to be easier to detect and track [1]. Preliminary studies suggest that conveying rotations using more accurate instructions, such as sonification-based interaction, may improve the accuracy in executing the rotations [25].

B. Sonification techniques supporting BVI people

The term "sonification" was first used in the 1990s as an auditory counterpart of data visualization, to refer to the use of non-speech sound to convey information. Using non verbal messages is recognized to be advantageous over speech in many applications and in several respect, including robustness to background noise, reduced cognitive load, and linguistic differences [26].

More specifically, sonification is often defined as "the transformation of data relations into perceived relations in an acoustic signal for the purposes of facilitating communication or interpretation" [27]. As a subtype of sonification, musification is the musical representation of data. It goes beyond direct sonification by using higher-level musical features such as tonality and polyphony in order to increase engagement in addition to conveying information [28].

A recent review of the literature on the sonification of physical quantities [29] shows that space-, timbre-, pitch-, and loudness-related sound features can be used to sonify quantities related to kinematics (distance, orientation, velocity, etc.). Accordingly, several sonification approaches have been developed to improve orientation and mobility performance, especially for BVI users. Usoh *et al.* [30] presented a comprehensive study of the application of non-speech sound for navigation and path finding. A more recent review on the use of sound for assisting the mobility of BVI persons is provided by Spagnol *et al.* [31], albeit with a focus on spatial audio.

Several approaches and applications have been developed for each of the three main levels of spatial knowledge [32], i.e. knowledge about a point in space (e.g., a landmark, an obstacle, a destination), knowledge about a sequence of points (a path to a destination, or "route knowledge"), and integrated knowledge about the environment (i.e. cognitive-map like knowledge, or "survey knowledge").

Regarding the first level, typical applications deal with non-visual scene representation. These include obstacle detection [33], identification of relevant or potentially dangerous elements (e.g., stairs or pedestrian crossings) [34], [35], and full-scene representation [36], [37].

Regarding the second level, which is the most relevant for the purpose of the present study, several approaches have been proposed for the use of sound in wayfinding tasks. The Personal Guidance System [38] obtains information from a GPS receiver and uses different types of auditory displays, spatial sound delivery methods, and tracker locations. The SWAN system employs non-speech auditory beacons that change in timbre and position with increasing levels of subject's practice: the learning experience of the system improves both speed and accuracy in navigation performances [39]. NavCog localizes the user through the sensing of the nearby bluetooth beacons and provides navigation instructions through vocal messages coupled with impulse sonification feedback, which is the same technique used as the baseline in our experiments [40].

Regarding the third level, various applications use sound to help a user form a cognitive map of an unknown environment, either "on-line" (i.e. while exploring the environment) or "off-line" (i.e. through the learning of maps of the environment in advance). Guerreiro *et al.* developed a gesture based virtual navigation system used to explore a turn-by-turn representation of the environment [41]. Lahav and coworkers [42] performed several studies on map exploration by blind subjects, which led to the development of the audio-haptic BlindAid system [43]. Katz and coworkers developed a system exploiting a 3D audio virtual environment to investigate structural properties of spatial representations [44]. They used a "ears in hand" metaphor [45], based on an egocentric view of the virtual map in which the ears of the user are virtually placed at the position of the hand or handheld device used to explore the map. The same metaphor was used by Geronazzo *et al.* [46] in developing a non-visual system for off-line exploration of

2D maps.

III. RESEARCH GOALS

The goal of our research is to find the most effective sonification approaches to guide BVI users towards their target using an indoor navigation system (INS) deployed on a mobile device (e.g., a smartphone). An INS tracks a user's location (e.g., using BLE beacons [14]) and monitors user's orientation through the mobile device inertial sensors. It then computes a route from the navigation starting point to the destination as a set of waypoints (intersections) connected by straight line segments. Therefore the INS must instruct the person on (i) how to walk on a straight line between the intersections, (ii) the actual reaching of an intersection, and (iii) how to rotate at the intersection in order to face the right turning direction.

In this work we address solely the problem (iii), and investigate solutions to guide the user while rotating at intersections. We aim to design a solution that can be used on existing mobile devices, without requiring additional hardware. This choice has the additional advantage of minimizing invasiveness, which is a concern for BVI people [47]. Beyond the screen (which is clearly not accessible by blind users), mobile devices have typically two output interfaces: audio and vibration. We focus on monaural audio, delivered through mobile device speakers. The investigation of vibration, stereo audio (using less invasive bone-conducting headphones), and multi-modal feedback is left as future work.

Within this frame, our research questions are:

Q1 How different sonification approaches impact the rotation accuracy and time?
Q2 How users perceive the proposed sonifications?
Q3 How learning effects, musical education and diverse cultural background impact user performance and how they perceive different sonification techniques?

We here assume that the user can interact with an INS (e.g., to insert the target destination) using native assistive technologies for mobile devices such as screen readers, which are available on both iOS and Android devices. We also assume that the device orientation is consistent with the user orientation. Therefore the application should instruct the users not to rotate the device with respect to themselves while walking.

IV. THE SONIFICATION TECHNIQUES

As shown in Sec. II-B, the problem of providing an audio description of distances from a target is not completely new. The novel aspect of this work is that it compares approaches that employ different sonification strategies: some use temporal (rhythmic) parameters, others use melodic and musical parameters; some employ a discrete sonification mapping, others employ a continuous mapping. The approaches are compared in terms of both user performance (angular error, completion time) and user acceptance (pleasantness, invasiveness, intuitiveness). To address **Q1**, we have implemented and compared the following sonification approaches[1]:

- *Ping* – The only feedback is an impulsive sound emitted when the target angle is reached. This approach was used in previous studies [40], and is retained here as a baseline.
- *Intermittent sound (IS)* – This sonification triggers impulsive "beeping" sounds at a variable rate, which is inversely proportional to the angular distance. It is vaguely resemblant of a Geiger-Müller counter. Our implementation employs a noisy unpitched pulse, which is repeated at rates from 1 to 15 Hz.
- *Amplitude modulation (AM)* – This approach employs a sinusoidal sound, modulated in amplitude by a low-frequency (sub-audio) sinusoidal signal. The frequency of the modulating signal is inversely proportional to the angular distance, producing a slowly pulsing sound at large angular distances, which becomes stationary when the target is reached. Our implementation employs a 440 Hz sinusoid (A_4 in scientific pitch notation – SPN), modulated by a sub-audio sinusoid from 1 to 15 Hz.
- *Musical scale (MS)* – The initial distance from the target angle is subdivided into eight circular sectors, corresponding to as many grades of an ascending major scale. While approaching the target angle, the user enters new areas, thus triggering new notes; when the target has been passed, a descending note is produced. Our implementation employs samples of piano notes and covers a C major scale starting at C_4 in SPN.

We implemented *IS*, *AM*, and *MS* approaches both alone ("base sonifications") and coupled with the ping when on target ("compound sonifications"). The former set was useful to compare sonification techniques, the latter to evaluate possible improvements by the reinforcement of an additional impulse.

An important aspect that can affect sonification interaction is the functional mapping between the kinematic quantity (angular distance) and the sonification parameters (e.g., rate of the discrete intermittent sound in *IS*, or frequency of the sub-audio modulating frequency in *AM*). For example, linear or an exponential mapping can be employed; a behavior that linearly responds to the distance from the target is expected to be easier to predict: e.g., for*MS*, this would imply eight circular sectors of the same amplitude and consequently moving at a constant angular speed would trigger notes at a constant pace. An exponential behavior, conversely, provides a finer tracking of user's motion when the target is approaching. Recalling the previous example, the amplitude of circular sectors would become narrower in proximity of the target angle.

Based on a preliminary evaluation with a BVI user, we selected the exponential mapping for *IS* and *AM*, as it was judged to provide a more precise tracking and encouraged the user to slow down when approaching the target. Conversely, the exponential mapping was considered to be confusing for *MS*, as several notes would be triggered for small movements

---

[1] Examples of the proposed sonifications are provided at https://earcons.netlify.com/

in proximity of the target: consequently, a linear mapping was chosen. This difference is coherent with the *MS* sonification mapping being inherently discrete (triggering notes at a discrete set of angles), while those of *IS* and *AM* are continuous.

V. EXPERIMENTAL SETTINGS

To assess the efficacy of our approach, we conducted two studies having the same goal but different characteristics. The preliminary study acquired initial formative results, and was used to test, refine and consolidate the experimental protocol. The main study applied the adjusted experimental protocol to a different, more numerous test group.

A. Participants

For the preliminary study, we had 7 completely blind participants, 5 male and 2 female. They were between 18 and 62 years old and all had prior experience with smartphones.

Table I reports the demographics data of the individuals participating in the main study. To address **Q3**, we had 2 groups of participants; 13 from Uganda ($U1$ to $U13$) and 5 from Italy ($I1$ to $I5$). Participants from Uganda were 7 male and 6 female, between 11 and 19 years of age (15.5 ± 2.3[2]). All were completely blind, with blindness onset age between birth and 6 years of age (2.5 ± 2.8); 4 of them had prior music training and none had prior experience with smartphones.

Participants from Italy were 3 female and 2 male, between 25 and 66 years of age (50.2 ± 19.2). Also in this case all participants were blind and their blindness onset age varied between birth and 44 years of age (19.3 ± 18.8); 4 of them had prior music training, and all of them had prior experience with smartphones, including mobility assistance apps and sonification interaction techniques.

B. Apparatus

The experiments were administered through a mobile app running on an Android Pixel 2 device. The same device was employed across all experiments to minimize the biases due to differences in technological components. In order to reduce involuntary actions, some buttons (e.g., Home, Back, etc.) were disabled by covering them with tape. The application used the default Android text-to-speech synthesizer and it was designed to be used by blind people without a screen reader software (e.g., TalkBack) so that participants did not have to get acquainted with this technology.

The application used device inertial sensors to record the participants' angular rotation speed at a sampling frequency of 100 Hz. Since the rotation error introduced by inertial sensors can be considered negligible [4], these data were used to compute the ground truth of participants' rotations.

Since the goal of the investigation was to measure exclusively the user accuracy during rotation, the experiment took place in a silent room and no headphones or other assistive tools (like a white cane) were used. The participants sat on a swivel chair and the smartphone was anchored to a wooden armrest (left or right, depending on the participant's handedness) through a holding mechanism. Inside the room, besides the participant, there were up to two supervisors with the task of checking that the experiment followed the planned protocol.

TABLE I: Main study participants' demographic data

ID	Sex	Age	Origin	VI onset	Experience with	
					Music	Smartphone
Preliminary Study Participants						
P1	M	20	Italy	Birth	yes	yes
P2	F	39	Italy	Birth	yes	yes
P3	M	45	Italy	21	yes	yes
P4	M	62	Italy	Birth	yes	yes
P5	F	47	Italy	25	no	yes
P6	M	25	Italy	Birth	yes	yes
P7	M	18	Italy	Birth	yes	yes
Main Study Participants						
U1	M	16	Uganda	Birth	no	no
U2	M	18	Uganda	Birth	no	no
U3	M	14	Uganda	Birth	no	no
U4	F	13	Uganda	5	yes	no
U5	M	16	Uganda	6	yes	no
U6	F	16	Uganda	Birth	no	no
U7	F	15	Uganda	5	no	no
U8	F	19	Uganda	Birth	no	no
U9	M	11	Uganda	6	no	no
U10	M	13	Uganda	5	yes	no
U11	F	16	Uganda	Birth	no	no
U12	F	17	Uganda	Birth	no	no
U13	M	18	Uganda	6	yes	no
I1	M	37	Italy	Birth	yes	yes
I2	F	25	Italy	11	yes	yes
I3	F	64	Italy	22	yes	yes
I4	F	66	Italy	Birth	yes	yes
I5	M	59	Italy	44	no	yes

C. Preliminary Study

The preliminary study was divided into two sessions. The first session evaluated the base sonification techniques with respect to the baseline approach (*Ping*). The second session, conducted with a different set of participants, evaluated the compound sonifications with respect to the baseline approach. The average completion time was about 30 minutes per participant.

The experiment included a training and a test phase. During the training, participants were given an overview of the experiments and a description of the proposed sonification techniques. Afterwards, participants could perform training trials with the provided techniques. The experimenters would observe the participants' behavior and, if any issues would arise at this point, explain how to correctly perform the experiments.

Afterwards, participants would perform the test trials in a random order. For each trial, participants were instructed to turn left or right with a verbal message and then had to rotate guided by one of the sonification techniques. Participants could end the trial with a tap on the screen when confident to have reached the desired angle. For each technique, we specified three predefined angles: narrow (50°), medium (80°), and large (130°). In total, 48 trials were tested, resulting from 4

---

[2] As a convention *Mean* ± *Standard Deviation* will be used

conditions (i.e. the sonifications) × 3 angles × 2 directions (clockwise and counterclockwise) × 2 repetitions.

D. Main Study

The main study focused on a reduced set of sonifications, namely *IS* and *MS*, with and without the ping reinforcement, plus the *Ping* sonification on its own[3]. This allowed us to focus on the most promising techniques emerging from the preliminary study, and to simultaneously test all the sonifications with each participant through additional task repetitions, without prolonging the duration of the study.

Indeed, thanks to an improved organization of the training and the use of earcons, despite an increased number of trials the average time to complete the experiment was cut to about 25 minutes. In total, the experiment involved 18 BVI participants ($U1 - U13, I1 - I5$).

Concerning the activity structure, a tutorial phase was added to accustom participants with the app interface. Participants were guided through an example of interaction with the app using the baseline sonification. Experimenters would make sure that participants comprehend the task, otherwise they would explain how to interact with the app correctly. The following training phase explored the sonifications, presented in a fixed order. Two trials were allowed for each technique.

During the preliminary study, participants would not know which technique they were about to test; thus, in addition to performing the task, they also had to identify the sonification at the same time. To address this issue, in the main study we introduced *earcons* [48] (i.e. brief, distinctive sounds) before each trial, so as to intuitively inform participants about the kind of experience they had to face. These earcons mimicked a brief interaction with the corresponding sonification. This decoupled the time and the movements needed to perform the actual task from the time needed to recognize the sonification.

The test phase was split into two parts, spaced by a 3 minutes break. Again, for each technique, three angles were proposed: narrow (50°), medium (80°), and large (130°). This time, the trials were organized in groups of 5, 1 trial for each condition (i.e. sonification technique), and the conditions within a group were randomized.

This organization into groups was not perceivable by participants and was designed to minimize the learning and repetitions by uniformly distributing the conditions across the trials, hence avoiding the learning and repetition biases that were possible in the preliminary study. In total, 90 trials were tested: 5 conditions × 3 angles × 2 directions (clockwise and counterclockwise) × 3 repetitions.

After the tests a final questionnaire was administered to obtain participants' subjective evaluations of the experience (**Q3**). We asked them to evaluate 5 metrics for each sonification, based on their personal opinion: pleasantness of the sonification, annoyance with the interaction, guidance precision, speed of the interaction, and overall appreciation. A Likert-like scale ranging between 1 and 7 was used to map these values.

VI. EXPERIMENTAL RESULTS

To measure the performance of the techniques described in Section IV, we defined the following metrics:

- **Rotation error** is the angular error between the target angle and the participant's direction at the end of a task;
- **Rotation time** is the time between the moment a participant starts a task and the moment he/she ends the task;
- **Rotation velocity** is the average angular speed achieved during the rotation;
- **Zero crossings count** measures the number of times the participant crossed the target angle before ending a task. A high value means that the participant moved back and forward multiple times over the target angle before being confident of the direction.

A. Preliminary study

Henceforth, we define three levels of significance ; $p < 0.05$ is denoted with $\star$, $p < 0.01$ with $\star\star$, and $p < 0.001$ with $\star\star\star$.

Our results reveal that *MS* ($4.43° \pm 5.63°$) was less error prone than *Ping* ($6.49° \pm 4.91°$), and the difference was found to be statistically significant* using a Wilcoxon rank-sum test (see Figure 1a). However no significant difference was found between the baseline and *IS* ($6.11° \pm 4.99°$) or *AM* ($8.20° \pm 5.33°$). We think that the high rate pulsations in the proximity of the target direction, which characterize these techniques, make it harder to precisely pinpoint it.

Considering the rotation time, *AM* ($6.39 \pm 1.18\,s$) was significantly* faster than *Ping* ($7.65 \pm 2.21\,s$). Similarly, *IS* ($6.53 \pm 1.09\,s$) was also significantly* faster than *Ping* (see Figure 1c). Conversely, *MS* was the slowest on average ($9.63 \pm 4.69\,s$), but the difference with *Ping* was not statistically significant.

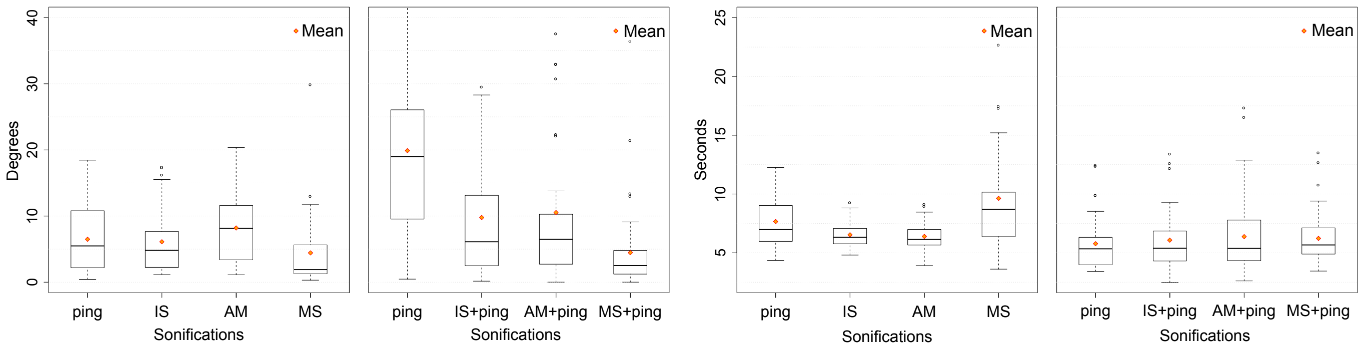
Coupling the proposed techniques with *Ping* was shown to improve their accuracy (see Figure 1b). Indeed, in the second session, all three proposed approaches were more precise than *Ping****. Instead, the addition of the *Ping* reinforcement did not provide improvements over the baseline in terms of rotation time (see Figure 1d). The results from the preliminary study motivate our choice of discarding *AM* from the main study. Indeed, *AM* and *AM+Ping* are conceptually similar to *IS* and *IS+Ping*, respectively. They exhibit similar behaviors in terms of performance, but a lower average accuracy.

B. Main Study

1) General Findings: The overall findings of the main study are shown in Table II. The study partially confirmed the previous results, but also shed light on new findings. Since in this study we performed paired tests with all sonifications, we were able to conduct comparisons also between base and compound sonifications, as well as investigate possible interactions between the sonification type and the presence of the reinforcement. In other words we investigated whether these two main effects (type of sonification and presence of

---

[3] Henceforth, when referring to the main study, we use "base sonifications" to refer to *IS* and *MS* while we use "compound sonifications" to refer to *IS+Ping* and *MS+Ping*



(a) Session 1 Rotation Error (b) Session 2 Rotation Error (c) Session 1 Rotation Time (d) Session 2 Rotation Time

Fig. 1: Results of the preliminary study

TABLE II: Main study general findings

Technique	**Error**		**Time**		**Velocity**		**0 cross**		**Pleasant**		**Annoying**		**Precise**		**Fast**		**Appreciated**	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Ping	11.11°	12.78°	$7.15\,s$	$3.5\,s$	$17.76^\circ/s$	$7.23^\circ/s$	2.03	1.41	3.56	1.29	4.50	1.50	5.39	1.04	5.50	0.92	4.39	1.65
IS	9.89°	12.40°	$9.19\,s$	$4.81\,s$	$14.52^\circ/s$	$5.76^\circ/s$	1.83	1.44	5.00	1.03	4.67	1.08	5.00	1.08	4.83	1.34	5.11	0.90
MS	3.25°	4.58°	$10.7\,s$	$4.45\,s$	$12.65^\circ/s$	$5.43^\circ/s$	2.49	1.65	5.83	1.04	3.56	1.42	5.72	0.96	4.61	1.82	6.06	1.06
IS+Ping	6.24°	10.04°	$8.28\,s$	$3.24\,s$	$12.83^\circ/s$	$5.66^\circ/s$	1.94	1.52	4.39	0.85	4.39	1.46	5.33	1.24	4.22	1.56	4.89	0.96
MS+Ping	3.17°	5.21°	$9.23\,s$	$3.69\,s$	$13.55^\circ/s$	$5.34^\circ/s$	2.17	1.47	4.61	1.42	4.06	1.70	5.72	1.07	4.78	1.35	5.22	1.06

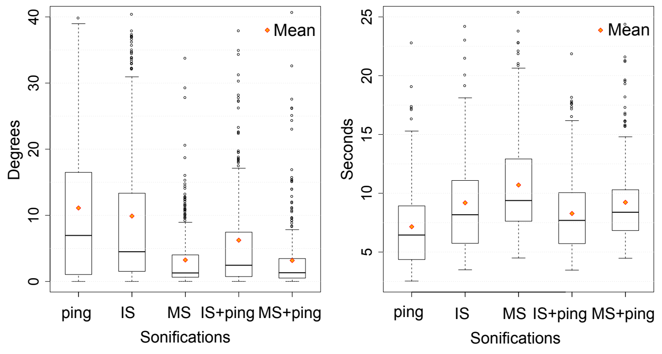
reinforcement) synergize to produce an interaction effect that influences the results (e.g., a sonification and ping reinforcement may not cause improvement independently, but they might jointly). For this purpose, we used the Scheirer-Ray-Hare Test [49].

In the case of rotation error (see Figure 2a), the interaction effect as well as the main effects were found to be significant***. As in our prior study, *MS* was the most accurate among the base sonification techniques, significantly*** better than *Ping*. Again, no significant difference was found between *IS* and *Ping*. However, *IS+Ping* performed significantly*** better than both *Ping* and *IS*. Instead, while *MS+Ping* was significantly*** better than *Ping*, it provided no significant improvement compared to *MS*.

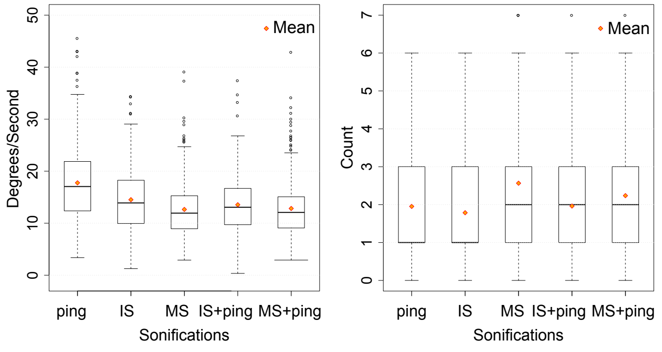
Considering the rotation time (see Figure 2b), the main effects were found to be significant*** both in terms of sonification type and the presence of reinforcement, but no interaction effect was detected. While *Ping* had significantly*** shorter rotation times than all the remaining techniques, *MS+Ping* also improved significantly over *MS*, whereas *IS+Ping* was not statistically different from *IS*.

The results about rotation time are generally in accordance with the analysis of the rotation velocity (see Figure 2c). Indeed, also in this case *Ping* is significantly*** faster than all remaining sonifications. However, we note that *MS* and *MS+Ping* have no significant differences.

An apparently contradictory result is that *MS+Ping* takes significantly** less time than *MS* without having a significantly higher rotation velocity. This can be explained considering zero crossings count (see Figure 2d): *MS+Ping* has a lower zero crossings count than *MS*. This means that participants did not rotate faster with *MS+Ping*, but they adjusted their orientation less frequently, resulting in lower rotation time.



(a) Rotation Error (b) Rotation Time



(c) Rotation Velocity (d) Zero Crossing Count

Fig. 2: Results of the main study

2) *Qualitative data analysis:* We analyzed the responses provided by participants (see Table II). In general, *Ping* was found to be significantly*** less *pleasant* than the remaining base sonifications, as well as the compound sonifications,

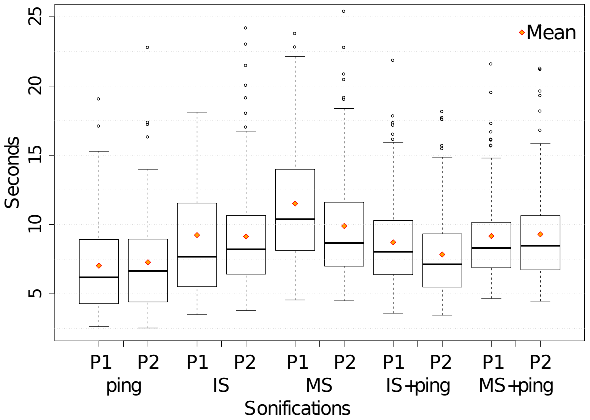


Fig. 3: Learning effect between Part 1 and 2: rotation time

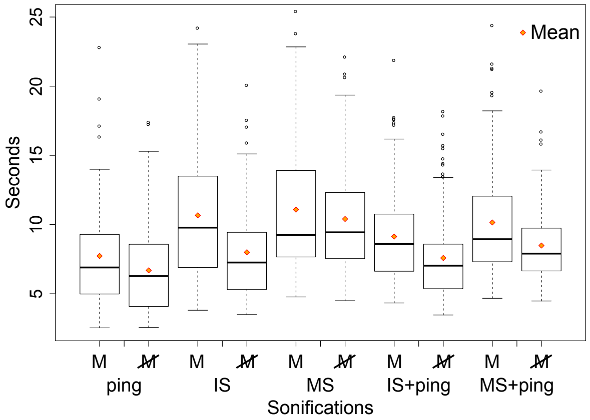


Fig. 4: Musical education: M - with, $\not M$ - without

however at a much lower significance* level. The addition of *Ping* reduced the pleasantness of the base sonifications, which in particular was significant for *MS* vs *MS+Ping***.

No significant differences among sonifications were found for *annoyance* and *precision*. However, for the perceived interaction *speed*, *Ping* was considered faster than *IS***.

In terms of general *appreciation*, musical scale ranked highest and was the only sonification significantly better than all the remaining ones. In particular, *MS* was significantly*** better than *Ping*. While *MS+Ping* was the second best along this metric, it was still significantly* less appreciated than *MS*.

3) Learning Effects: We also investigated the presence and the relevance of the learning effects on the analyzed metrics. For this purpose, we compared the results of the 45 trials of part 1 against the 45 trials of part 2, divided by sonification type (see Figure 3). While no significant differences were found in terms of rotation error, *MS* was found to significantly*** improve the rotation time between part 1 ($11.52 \pm 4.7\,s$) and part 2 ($9.89 \pm 4.04\,s$). Similarly, *IS+Ping* rotation time also improved** between part 1 ($8.72 \pm 3.31\,s$) and part 2 ($7.84 \pm 3.12\,s$).

4) Musical Education: Participants with prior music experience ($9.75 \pm 4.66\,s$) had significantly*** longer rotation times than others ($8.23 \pm 3.53\,s$) for all sonifications but *MS* ($9.89 \pm 4.04\,s$). See Figure 4. Those who had prior musical experience ($5.88° \pm 8.50°$) also scored lower rotation errors than the others ($7.39° \pm 11.30°$). While this trend is present for all sonifications, no significant difference is evident, due to the high variability of the data. The only exception is *MS+Ping*: in this case participants with prior musical experience ($2.37° \pm 3.79°$) scored significantly** lower rotation errors than those without ($3.83° \pm 6.06°$).

In terms of perceived qualitative differences, *MS* (6.38 ± 0.74) was found to be more precise** for participants with prior musical experience than for those without (5.20 ± 0.79). Compound sonifications were also found to be less annoying* for participants with musical experience than for those without. Specifically, *IS+Ping* scored 3.50 ± 1.69 for those with musical experience and 5.10 ± 0.74 for those without. Similarly, *MS+Ping* scored 3.00 ± 1.85 for those with musical experience and 4.90 ± 0.99 for those without. For *MS+Ping*, the overall appreciation was also higher* for participants with prior musical experience (5.88 ± 1.13) than for those without (4.70 ± 0.67).

5) Cultural Differences: Cultural differences are summarized in Figure 5. Participants from Uganda scored significantly lower*** rotation times ($7.89 \pm 3.11\,s$) than those form Italy ($11.56 \pm 4.51\,s$). At the same time, participants from Uganda also had significantly higher*** rotation error ($8.39° \pm 9.95°$) than participants from Italy ($2.44° \pm 2.77°$).

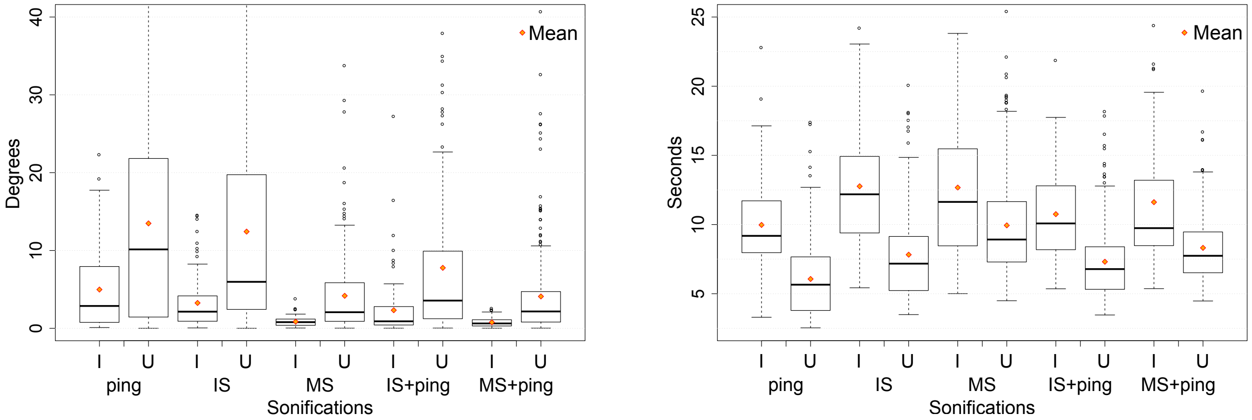
Ugandan participants considered *Ping* less pleasant*** (2.92 ± 0.76) and generally less appreciated*** (3.62 ± 1.12) than Italian participants (5.20 ± 0.84 and 6.40 ± 0.89 respectively). *IS+Ping* was considered less pleasant* (4.08 ± 0.64), less precise** (4.85 ± 0.99) and slower*** (3.38 ± 0.65) by Ugandan participants than by those in Italy (5.20 ± 0.84; 6.60 ± 0.89; and 6.40 ± 0.89 respectively). Similarly, *MS+Ping* was less pleasant*** (3.92 ± 0.86), less precise* (5.38 ± 0.96), slower** (4.23 ± 1.01), and less appreciated*** (4.69 ± 0.63) for participants from Uganda, than for those in Italy (6.40 ± 0.89; 6.60 ± 0.89; 6.20 ± 1.10; and 6.60 ± 0.55 respectively). *MS* was also considered much more pleasant*** for Italian participants, with a perfect score of 7, than for participants from Uganda (5.38 ± 0.87).

VII. DISCUSSION

A. General Findings

Regarding the baseline sonification (*Ping*), our results report that the rotation errors are in the same range ($11.11° \pm 12.78°$) as prior studies conducted in the wild [1] ($14.9° \pm 9.9°$). This level of error appears to be caused by a delay in the user reaction to the notification stimuli, which is provided when reaching the target angle. Instead, the proposed sonifications are capable of a more precise guidance than the baseline sonification because they provide additional feedback continuously, in order to help the user pinpoint the target angle more accurately.

Specifically, *MS* plays ascending notes at fixed angular distances while approaching the target angle, and descending notes when moving away from it. This approach has 2 key effects: 1) as the user rotates, feedback is provided frequently,



(a) Rotation Error

(b) Rotation Time

Fig. 5: Cultural differences

which causes the user to pay attention to the feedback and rotate slowly; 2) as the highest note is played, it preempts the user to stop rotating because the target direction is imminent.

The notification anticipates the target direction of about 6° for narrow angles (50°) up to about 16° for large angles (130°). Our intuition is that providing feedback before the target direction offsets the inaccuracy inherent to the baseline sonification, thus resulting in a more accurate guidance. Indeed the results of our study show that *MS* achieves a rotation error of $3.25° \pm 4.58°$, much lower than all other approaches.

While *MS* outperforms other base sonifications in terms of rotation accuracy, the continuous feedback results in lower rotation speeds and, therefore, higher rotation times for this sonification. Furthermore, no specific feedback is provided when crossing the target angle, and therefore participants would adjust their orientation multiple times before being satisfied by their orientation. Conversely, *IS* accelerates the frequency of the feedback impulses in proximity of the target angle, prompting the user to slow the rotation only at that point. While this approach does not manage to reach the same accuracy as *MS*, it achieves faster rotation times and less adjustments in proximity of the target angle.

By comparing the base and the compound sonifications, we are also able to corroborate the intuition, built in the preliminary study, that adding the *Ping* reinforcement to *MS* significantly decreases its rotation time, because it provides a clear feedback upon reaching the target angle. This improvement is advantageous also because it does not worsen the accuracy of the sonification. The use of the reinforcement *Ping* is also confirmed to increase the accuracy of *IS*, due to a more precise feedback when reaching the target angle. Also in this case, there is no worsening of the rotation time.

B. Differences Between the Preliminary and the Main Studies

While the main study generally confirms the results hinted in the preliminary study, it also disproves some of the initial intuitions. In the preliminary study, both *IS* and *AM* were significantly faster than the baseline, while there were no significant differences between the compound sonifications and the baseline. We recall that in the preliminary study the tests for base and compound sonifications were conducted by two distinct set of users. Conversely, the main study shows that the *Ping* baseline remains faster than all the other sonifications.

An explanation for this difference may be found in the improved experimental protocol. With the addition of the earcons, participants do not spend time to identify the sonification when performing a task. In particular, this may affect the time needed to identify *Ping*, which is the only sonification that does not immediately provide a feedback, possibly causing a delay in the participant's interaction.

C. Qualitative Findings

Analyzing the qualitative data, we noticed that participants found all sonifications to be more pleasant than the baseline. This is a confirmation of our intuition that participants prefer a continuous guidance over a single impulse notification. Furthermore, considering the overall appreciation, *MS* stands out as the participants' favorite sonification, significantly better than the baseline. At the same time, no significant differences emerge considering the perceived annoyance among the diverse sonifications, hinting that none of the sonifications is considered by participants to cause significant cognitive overload. We believe that annoyance should be evaluated in a real context, in which the user is navigating through an indoor space, with the need to listen and focus on environmental sounds.

In contrast with the experimental results, participants did not perceive any of the presented sonifications to be more accurate than others. Not being able to assess whether the guidance is accurate may lead the user away from the path without noticing it, which was shown to impact the navigation outcome in previous works [1]. As for the rotation time, while participants correctly determined that the baseline sonification was the fastest, they were not able to correctly differentiate between the remaining ones. For example, *IS+Ping*, which was the fastest during the trials, was regarded as the slowest in the questionnaires, and the only one considered to be significantly slower than the baseline. We believe that it was harder for participants to evaluate the speed of the sonifications using continuous guidance, since they were focused on following instructions rather than evaluating the interaction speed.

D. Presence and Entity of the Learning Effects

Thanks to the improved ordering of test trials in the main study, we were able to analyze how the rotation time and error change between the first and the second part of the activity. This was not possible in the preliminary study, as the rotations were completely randomized and, therefore, there was no clear cut between the initial rotations and the following ones. While there were no significant differences on the rotation error metric, we uncovered the presence of a significant learning effect on the rotation time for the *MS* sonification.

This sonification is regarded as more complex to understand without prior experience. It plays higher notes when the user is approaching the target angle, but no clear feedback is given when the user reaches it (i.e. that the last note played is the highest one). Only after moving away from it a lower note is played to signal that the distance from the target is increasing. We believe that, with some experience, users can learn to identify the highest note, and, consequently, be able to complete the task without adjusting their orientation as much.

E. Musical Education

With the improved study design, we were also able to assess the impact of the user characteristics on the results. We first focused on the level of the musical expertise, discovering that those who have prior musical expertise tend to pay more attention to the sonifications, thus scoring overall higher rotation times but lower rotation errors. In particular, in the case of *MS+Ping* the rotation error is significantly lower for those with prior musical experience. One possible explanation is that, with this sonification approach, prior musical experience helps to follow the provided instructions more accurately, without being distracted by the combination of sonification feedbacks.

This is reflected in the qualitative scores. Indeed the *MS+Ping* sonification is overall more appreciated by those with prior musical experience that by those without. Those with prior musical experience were also more tolerant towards both *MS+Ping* and *IS+Ping*. We speculate that the compound sonifications, characterized by a dual feedback, are more cognitively demanding than base sonifications; indeed these sonifications are considered more annoying by participants without prior musical experience. Those with prior musical experience also considered *MS* to be more precise, further supporting their appreciation for musical scale-based interactions.

F. Cultural Differences

The analysis of the results, divided by the origin of participants (Uganda and Italy), highlighted stark differences between the two groups. While participants from Uganda had much higher rotation errors, they also achieved much lower rotation times for all the sonifications. The surprising finding is that both metrics followed the same trend for the two groups for every sonification. For example, in both cases *MS* and *MS+Ping* were the most accurate; conversely, *Ping* had the lowest rotation time, yet it caused most errors.

Considering the qualitative results, participants from Uganda generally had a lower appreciation of all sonifications, in particular of the compound ones. They found such sonifications less pleasant, but also less accurate and fast. *MS+Ping* and *Ping* were also overall less appreciated compared to Italian participants, while *MS* and *Ping* were found to be less pleasant.

We suspect that these differences are due to the fact that participants from Uganda were less familiar with the presented interactions, assistive technologies and smartphones in general, and therefore they found interacting with the system less pleasant and more annoying. For the same reason they were not aware on what accuracy and interaction time is expected when guided with sonification-based interactions, and, therefore, were much faster yet less accurate than Italian participants. This shows that cultural differences have to be considered when designing interaction paradigms for navigation assistance.

VIII. CONCLUSION AND FUTURE WORK

Assistive technologies for navigation enable BVI people to acquire independence during everyday mobility, thus improving their quality of life. Existing systems are becoming more accurate in localizing and routing BVI users. However, the guidance techniques used in these systems have limited accuracy and may cause navigation errors or even endanger the user. In particular errors occurring while following rotation instructions were shown to influence the outcome of the navigation in existing systems.

Based on the results of our experiments, we propose two approaches (*IS* and *MS*) which use sonification to provide continuous guidance during rotation. Both are appreciated by the users, and they achieve a much higher accuracy than the baseline solution commonly used in existing navigation assistance tools. However, the proposed guidance techniques also result in longer rotation times compared to the baseline. This issue is solved by combining the proposed approaches with the baseline sonification, by adding an impulse feedback in the proximity of the target rotation angle. These compound sonifications reduce rotation time without penalizing rotation accuracy and hence are promising solutions to be used in actual navigation systems.

As future work we intend to investigate sonification guidance for other navigation tasks, such as frontal movements. We will also investigate spatialized and stereophonic audio cues, as well as combined auditory and vibro-haptic feedback. We will validate the proposed guidance interaction paradigms in more naturalistic scenarios, using bone conducting headphones that enable each user to hear their personalized instructions but do not prevent them from listening to the surrounding sounds.

The designed interaction paradigms will be ultimately integrated in turn by turn indoor navigation systems such as Navcog [14], and outdoor navigation tools [8]. The integrated systems will be evaluated through supervised experiments and remote data collection and analysis [7], [8].

REFERENCES

[1] D. Ahmetovic, U. Oh, S. Mascetti *et al.*, "Turn right: Analysis of rotation errors in turn-by-turn navigation for individuals with visual impairments," in *Int. Conf. on Computer and Accessibility*, 2018.

[2] L. Hakobyan, J. Lumsden, D. O'Sullivan *et al.*, "Mobile assistive technologies for the visually impaired," *Ophthalmology*, 2013.

[3] M. Periša, I. Cvitić, and R. E. Sente, "Comparative analysis of mobile phone application solutions accessibility for informing visually impaired persons in traffic environment," in *Services for mobility and mobility as a service*, 2017.

[4] S. Mascetti, D. Ahmetovic, A. Gerino *et al.*, "Zebrarecognizer: Pedestrian crossing recognition for people with visual impairment or blindness," *Pattern Recognition*, vol. 60, 2016.

[5] S. Mascetti, D. Ahmetovic, A. Gerino, and C. Bernareggi, "Robust traffic lights detection on mobile devices for pedestrians with visual impairment," *Comp. Vision and Image Understanding*, vol. 148, 2016.

[6] G. H. Flores and R. Manduchi, "Weallwalk: An annotated dataset of inertial sensor time series from blind walkers," *Trans. on Accessible Computing*, vol. 11, no. 1, 2018.

[7] H. Kacorri, S. Mascetti, A. Gerino *et al.*, "Supporting orientation of people with visual impairment: Analysis of large scale usage data," in *Int. Conf. on Computer and Accessibility*. ACM, 2016.

[8] K. Hernisa, M. Sergio, G. Andrea *et al.*, "Insights on assistive orientation and mobility of people with visual impairment based on large-scale longitudinal data," *Trans. on Accessible Computing*, vol. 11, no. 1, 2018.

[9] K. Hara, S. Azenkot, M. Campbell *et al.*, "Improving public transit accessibility for blind riders by crowdsourcing bus stop landmark locations with google street view: An extended analysis," *Trans. on Accessible Computing*, vol. 6, no. 2, 2015.

[10] D. Ahmetovic, R. Manduchi, J. M. Coughlan *et al.*, "Mind your crossings: Mining gis imagery for crosswalk localization," *Trans. on Accessible Computing*, vol. 9, no. 4, 2017.

[11] E. P. Fenech, F. A. Drews, and J. Z. Bakdash, "The effects of acoustic turn-by-turn navigation on wayfinding," in *Human factors and ergonomics society*, vol. 54, no. 23. SAGE Publications, 2010.

[12] J. Rajamäki, P. Viinikainen, J. Tuomisto *et al.*, "Laureapop indoor navigation service for the visually impaired in a wlan environment," in *Conf. on Electronics, Hardware, Wireless and Optical Comm.*, 2007.

[13] M. Nakajima and S. Haruyama, "Indoor navigation system for visually impaired people using visible light communication and compensated geomagnetic sensing," in *Communications in China*. IEEE, 2012.

[14] M. Murata, D. Ahmetovic, D. Sato *et al.*, "Smartphone-based indoor localization for blind navigation across building complexes," in *Int. Conf. on Pervasive Computing and Communications*. IEEE, 2018.

[15] D. Ahmetovic, M. Murata, C. Gleason *et al.*, "Achieving practical and accurate indoor navigation for people with visual impairments," in *Web for All Conf.* ACM, 2017.

[16] E. Ohn-Bar, J. Guerreiro, D. Ahmetovic *et al.*, "Modeling expertise in assistive navigation interfaces for blind people," in *Conf. on Intelligent User Interfaces*. ACM, 2018.

[17] J. Guerreiro, E. Ohn-Bar, D. Ahmetovic *et al.*, "How context and user behavior affect indoor navigation assistance for blind people," in *Web for All Conf.* ACM, 2018.

[18] H. Kacorri, E. Ohn-Bar, K. M. Kitani *et al.*, "Environmental factors in indoor navigation based on real-world trajectories of blind users," in *Conf. on Human Factors in Computing Systems*. ACM, 2018.

[19] D. Guth and R. LaDuke, "The veering tendency of blind pedestrians: An analysis of the problem and literature review," *J. of Visual Impairment and Blindness*, 1994.

[20] E. Chrastil and W. Warren, "Rotational error in path integration: encoding and execution errors in angle reproduction." *Experimental brain research*, vol. 235, no. 6, 2017.

[21] E. K. Sadalla and D. R. Montello, "Remembering changes in direction," *Environment and Behavior*, vol. 21, no. 3, 1989.

[22] V. Marlinsky, "Vestibular and vestibulo-proprioceptive perception of motion in the horizontal plane in blindfolded man-ii. estimations of rotations about the earth-vertical axis," *Neuroscience*, vol. 90, no. 2, 1999.

[23] G. Zacharias and L. Young, "Influence of combined visual and vestibular cues on human perception and control of horizontal rotation," *Experimental brain research*, vol. 41, no. 2, 1981.

[24] R. Jürgens, T. Boss, and W. Becker, "Estimation of self-turning in the dark: comparison between active and passive rotation," *Brain Res.*, 1999.

[25] D. Ahmetovic, F. Avanzini, A. Baratè *et al.*, "Sonification of pathways for people with visual impairments," in *Int. Conf. on Computer and Accessibility*. ACM, 2018.

[26] T. Dingler, J. Lindsay, and B. N. Walker, "Learnabiltiy of sound cues for environmental features: Auditory icons, earcons, spearcons, and speech," in *Int. Conf. Auditory Display*, 2008.

[27] B. N. Walker and M. A. Nees, "Theory of sonification," in *The sonification handbook*, 2011.

[28] A. D. Coop, "Sonification, musification, and synthesis of absolute program music," in *Int. Conf. Auditory Display*, 2016.

[29] G. Dubus and R. Bresin, "A systematic review of mapping strategies for the sonification of physical quantities," *PLoS ONE*, vol. 8, 2013.

[30] M. Usoh, K. Arthur, M. C. Whitton *et al.*, "Walking $>$ walking-in-place $>$ flying, in virtual environments," in *ACM SIGGRAPH*, 1999.

[31] S. Spagnol, G. Wersényi, M. Bujacz *et al.*, "Current use and future perspectives of spatial audio technologies in electronic travel aids," *Wireless Comm. and Mobile Comp.*, 2018.

[32] J. M. Wiener, S. J. Büchner, and C. Hölscher, "Taxonomy of human wayfinding tasks: A knowledge-based approach," *Spatial Cognition & Comp.*, 2009.

[33] M. Bujacz, P. Skulimowski, and P. Strumillo, "Naviton – a prototype mobility aid for auditory presentation of three-dimensional scenes to the visually impaired," *J. Audio Eng. Soc.*, vol. 60, 2012.

[34] M. Bujacz, K. Kropidlowski, G. Ivanica *et al.*, "Sound of vision - spatial audio output and sonification approaches," in *Int. Conf. on Computer Helping People with Special Needs*. Lecture Notes in Computer Science 9759, Springer Verlag, 2016.

[35] S. Mascetti, L. Picinali, A. Gerino *et al.*, "Sonification of guidance data during road crossing for people with visual impairments or blindness," *Int. J. of Human-Computer Studies*, 2016.

[36] P. B. Meijer, "An experimental system for auditory image representations," *Trans. on Biomedical Engineering*, 1992.

[37] S. Shoval, J. Borenstein, and Y. Koren, "Auditory guidance with the navbelt-a computerized travel aid for the blind," *Trans. on Systems, Man, and Cybernetics*, 1998.

[38] J. M. Loomis, J. R. Marston, R. G. Golledge *et al.*, "Personal guidance system for people with visual impairment: A comparison of spatial displays for route guidance," *J. of Visual impairment and Blindness*, 2005.

[39] B. N. Walker and J. Lindsay, "Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice," *Human Factors*, 2006.

[40] D. Ahmetovic, C. Gleason, C. Ruan *et al.*, "Navcog: A navigational cognitive assistant for the blind," in *I. C. on Human Comp. Int. with Mobile Devices and Ser.*, 2016.

[41] J. Guerreiro, D. Ahmetovic, K. M. Kitani *et al.*, "Virtual navigation for blind people: Building sequential representations of the real-world," in *Int. Conf. on Computer and Accessibility*. ACM, 2017.

[42] O. Lahav and D. Mioduser, "Construction of cognitive maps of unknown spaces using a multi-sensory virtual environment for people who are blind," *Comp. in Human Behavior*, vol. 6, no. 1, 2008.

[43] O. Lahav, D. W. Schloerb, S. Kumar *et al.*, "A virtual environment for people who are blind - a usability study," *J. of Assistive Technologies*, 2012.

[44] B. F. G. Katz, S. Kammoun, G. Parseihian *et al.*, "Navig: augmented reality guidance system for the visually impaired," *Virtual Reality*, 2012.

[45] C. Magnusson, H. Danielsson, and K. Rassmus-Gröhn, "Non visual haptic audio tools for virtual environments," in *Haptic and Audio Int. Design*, 2006.

[46] M. Geronazzo, A. Bedin, L. Brayda *et al.*, "Interactive spatial sonification for non-visual exploration of virtual maps," *Int. J. of Human-Computer Studies*, vol. 85, Jan. 2016.

[47] R. G. Golledge, J. R. Marston, J. M. Loomis *et al.*, "Stated preferences for components of a personal guidance system for nonvisual navigation." *J. of Visual Impairment and Blindness*, vol. 98, 2004.

[48] D. A. Sumikawa, "Guidelines for the integration of audio cues into computer user interfaces," Tech. Rep., 1985.

[49] C. Dytham, "Scheirer-Ray-Hare test," *Choosing and Using Statistics: A Biologist's Guide*, 2003.