Evaluating micro-guidance sonification methods in manual tasks for Blind and Visually Impaired people

RENAN GUARESE, Royal Melbourne Institute of Technology, Australia
RON VAN SCHYNDEL, Royal Melbourne Institute of Technology, Australia
FABIO ZAMBETTA, Royal Melbourne Institute of Technology, Australia

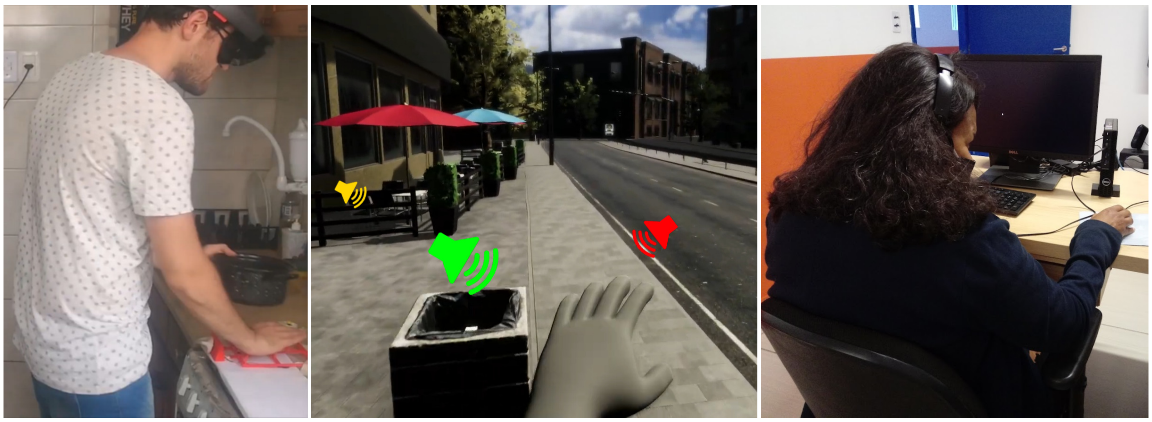

Fig. 1. Left: a blindfolded user searching for a stove and receiving spatialized sound feedback from a MR HMD, during preliminary tests. Center: representation of a sonification coming from multiple locations, under different pitches, timbres, loudness and intermittence, varying according to the user current position and the information being displayed. Right: a remote BVI user evaluating sonification methods for the current experiment in their own office environment.

This paper presents a user evaluation of seven sonification methods in two-dimensional (2D) manual micro-guidance tasks, which can be used as building blocks for spatialized audio in Mixed and Virtual Reality to model next-generation guidance aids for the Blind and Visually Impaired (BVI). The methods were tested in comparable interactive sonifications of 2D positions in a series of hand-navigation assessments with BVI and blindfolded sighted users, to validate the different approaches in environments without any visual feedback. Results highlighted that alternation and spatiality can be useful resources in sonified guidance, and that users accustomed to faster-than-regular audio speed replay tend to have more precise performances, while musical literacy only had a performance effect on methods highly dependent on aural skills. Ultimately, this work corroborates the notion that sonification may help BVI users perform better in day-to-day manual micro-guidance tasks such as retrieving items from a pantry, handling kitchen appliances, and properly discarding trash.

CCS Concepts: • **Human-centered computing** → **Accessibility technologies**; **Empirical studies in accessibility**; **Auditory feedback**; **Mixed / augmented reality**; **User studies**.

Manuscript submitted to ACM

Additional Key Words and Phrases: micro-guidance, assistive technologies, blind and visually impaired guidance, mixed reality accessibility

ACM Reference Format:
Renan Guarese, Ron van Schyndel, and Fabio Zambetta. 2022. Evaluating micro-guidance sonification methods in manual tasks for Blind and Visually Impaired people. In *OzCHI 2022: connected creativity, November 29–December 2, 2022, Canberra, Australia.* ACM, New York, NY, USA, 18 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

In the context of wayfinding, micro-navigation can be defined as the "sensing of the immediate environment for obstacles and hazards" [66]. In a similar fashion, our work uses the broader term micro-guidance to designate both close range navigation and object locating tasks, particularly when hand position is critical. This implies tasks such as putting away and retrieving items from a pantry, locating and handling kitchen appliances, and discarding trash into the proper bins. Despite being a well-studied problem [11, 27], micro-guidance can be a cognitively demanding and error-prone problem for Blind and Visually Impaired (BVI) people [13, 49], highly affecting their day-to-day routines. Especially while navigating inside buildings and public spaces where GPS data is unreliable, BVI people get reduced confidence in independent navigation [21].

In unfamiliar locations, micro-guidance usually requires understanding current position, finding the way to a destination, and maintaining orientation [21]. *In-situ* guidance assistance has thus been used to allow users to rely less on prior knowledge of the environment [18], which could be out-of-date or even non-existent. Adding to the situational aspect of Augmented Reality (AR), we intend on exploring the strengths of audio spatialization, inspired by its use in Virtual Reality (VR) environments, which are commonly filled with a range of spatial audio, from sounds that provide critical notifications to users (e.g., video-game "enemy" footsteps) to those that increase realism (e.g., wind blowing in a nature simulation) [20].

According to Neuhoff [41], although the spatial resolution of the visual system is better than that of the auditory system, we can only see a limited field of vision while we can hear in 360 degrees - despite some areas of directional ambiguity, which can be resolved by encouraging small movements of the user head or sound source [61]. By using current Mixed Reality (MR) equipment such as HoloLens 2, untrained users can consistently localise sound sources with a deviation of less than 10° solely based on spatialization [65]. Such equipment uses generalized Head Related Transfer Functions which, while functional, are often not effective enough for a listener to consistently resolve sound source elevation [36]. Thus, relying solely on spatial audio to represent the position of a target in 3D space is unlikely to be effective, requiring additional sound attributes to be explored.

The methods presented in this work are meant to be part of a larger set of interface cues which we aim to investigate and optimize for different contexts. The scenario set up in this particular experiment is a metaphor for a generic manual task in a flat plane, one of the many different contexts any person goes through in their day-to-day activities. BVI people mostly rely on their hearing and touch senses to orient themselves to their surroundings [39, 62]. Some surfaces, however, can be uncomfortable, unsanitary or even dangerous to touch, such as an already occupied park bench, a dirty garbage can, or a hot stove. By guiding the user's hand via sound, an interactive sonification feedback can precisely inform the location of available seating, the top of a bin or the lit stove burner, as demonstrated in Fig. 1-center. In this manner, the user is able to appropriately place their hand to perform the task, without having to touch any undesired surfaces, as in Fig. 1-left.

The research for this paper was performed during a full lockdown due to the COVID-19 pandemic, so in order to allow the research to proceed, it was decided to have the experiment run fully online, allowing users to participate remotely, from the safety of their houses or places of work. In this paper, we present research on sonification methods *ex-situ*, presenting them in a 2D space based on a flat screen, and using a computer mouse as the interactive control device, serving the purpose of a hand tracker. This enabled the research to proceed online, including BVI participants, as it was unlikely for BVI or sighted users to posses MR or VR devices.

Our work aims to compare the performance of BVI and blindfolded sighted users performing the same real-world tasks while receiving virtual assistance. By simulating a totally blind sonified target-finding manual task, we intend on gathering user data and insights on micro-guidance sonification methods to pave the way for next-generation guidance aids for the Blind and Visually Impaired (BVI) in our subsequent *in-situ* investigation. We hypothesize that BVI users are likely to have developed behaviors facilitating the execution of such tasks. By observing these behaviors, we hope to optimize our methods for BVI users and their needs, developing assistive technologies that can be used in real-life scenarios, besides simulated environments.

The following research questions were the basis of this investigation, which also outline our main contributions:

- RQ1: Is user performance (e.g. in time taken and distance error) affected by different sonification patterns used in micro-guidance tasks?
- RQ2: Is the perceived user workload affected in such tasks?
- RQ3: Do personal traits (e.g. sight condition and knowledge of musical theory) affect the results of particular sonification patterns?
- RQ4: Does the reported user performance correlate to their accuracy in performing sonified micro-guidance tasks?

The remainder of this paper is organised as follows: Section 2 illustrates related works from the literature, Section 3 defines the hypotheses used to base this research, and Section 4 describes the sonification methods used in our experiment. Subsequently, Section 5 details the reasoning behind and the procedures of the user experiment, Section 6 exposes the results found, and Section 7 provides arguments to understand and explain said results. Finally, Section 8 discusses what could be done with the knowledge gained here, and Section 9 concludes the study.

2 RELATED WORKS

In order to understand the current state-of-the-art in the multiple frontlines this paper faces, a broad literature review was executed. The works researched were grouped together and discussed in the three categories below, according to their similarity in theme. Despite being interesting and proven useful solutions for users with low vision, works dependent on visual cues fall out of the scope of the current investigation and will not be discussed.

2.1 Sonification as an interface

Sonification has been used as an alternative data display by numerous works [7, 33, 41, 42]. Particularly regarding sonification as a tool for guidance methods, [36] compared three situated sonification patterns in VR to inform object locations in a user experiment. Their results indicated that people could benefit from their auditory displays when not relying on vision. This sort of guidance can arguably be useful for any egocentric activity when you cannot rely on vision. For instance, similar works empirically evaluated sonification methods in road crossing experiments with users [2, 10, 34, 63]. In [34], most of the subjects (75%) declared to prefer either of the sonifications tested over speech instructions, while [63] indicated that spatialized sound enhanced the experience of virtual traffic simulations. Besides spatially positioning sounds at the starting and finishing ends of a crosswalk to convey its location and direction, other pieces of information are expected by the user [2, 10], which could be further explored in other audio attributes.

In an audio-only game environment, [64] presented design principles for responsive audio feedback geared towards hand gestures, with results indicating non-speech audio cues lead to measurably higher user immersion. In sports HCI, even one-dimensional sonifications have been effectively used to convey position deviation to athletes [50, 58]. According to Sigrist et al., auditory displays are often not systematically evaluated, being instead selected in an *ad hoc* manner. Since these arbitrarily designed displays may constrain motor learning by reduced motivation, distraction, or misinterpretation even in professional athletes [50], it is imperative to conduct proper evaluations when appointing sonification patterns in assistive technologies for the Visually Impaired.

By comparison, the current work aims to expand on the sonification techniques explored in the aforementioned research studies by evaluating seven different methods, with combined sound attributes dynamically changing according to the mouse position, which acted as a tracker for the user hand position. Other works such as Paré [43] investigate the use of sonification towards spatial navigation as a sensory substitution device (SSD). Although very promising, this research considers SSDs to be a different field of study, which will not be discussed further.

2.2 Non-visual assistive navigation

Based on recent research reviews [13, 25] and by further exploring the area, some related works were found and analyzed. [57] developed a device that generates a comprehensive radar profile of the environment and translates it into a sound representation. Their paper reports the radar and binaural sound engine of the system, which is able to indicate the distances and directions of obstacles. [31] performed a sound perception experiment directing users to adjust the pan and elevation angles of a camera to point it to a target, based solely on audio feedback, with results exposing a correlation between target difficulty and search time.

By exploring *in-situ* navigation, [18, 48] studied the effect of their tool in blind users while they followed a route. Despite exposing valuable results in favor of the use of situated MR, their user experiments focused on building prior knowledge for macro-navigation via virtual navigation. Similarly, with the intention of simulating real outdoor environments in Orientation and Mobility training sessions, [54] developed and tested an accessible VR system providing visual and audio feedback to visually impaired users.

Aimed at outdoor navigation, [3] used spatialized music to compare turn-by-turn and beacon navigation for sighted people, which was similarly investigated by [46], although using a tactile display instead. Another similar work used Augmented Reality to guide BVI users in indoor environments [28]. Their system, however, focused on turn-by-turn guidance via haptic feedback in a vest. Opposed to all of these works, we intend to direct our efforts at micro-guidance in different day-to-day tasks for BVI users, primarily following the design requirements and heuristics recommended by specialized guidance [6, 29, 62] and auditory environment [37, 41, 42] works.

2.3 Sonified Mixed Reality

Concerning Mixed and Augmented reality (MR/AR), several works have explored focusing on sound as a primary medium, i.e. as a substitute for the traditional AR visuals, in several different exploration contexts, such as urban [5, 24, 32], gaming [16, 17, 45], cultural [23, 35, 51], and botanical [59]. Results from user experiments have shown that the use of spatial audio together with sound icons enabled users to explore multiple simultaneous elements with an increased level of usability [5, 23], search performance [32], and immersion [59].

Thus, by subverting the usual expectation that MR is mainly composed of visual elements, one is able to target BVI users with MR assistive tools. Focusing on its non-visual elements, such as audio and haptics, MR heuristics and guidelines [9, 12, 14, 40] can still be followed by contextualizing such elements onto the user's surrounding space, making them pertain to real objects and locations. Katz [24], for instance, exposed a macro-navigation system which is able to guide users to a desired destination through spatialized semantic audio rendering, using a head-centered reference frame. The current investigation, however, pursues a micro-guidance scenario, focusing on hand-centered tasks.

San Marín and Kildal [47] investigated the use of spatialized auditory feedback on AR Head-Mounted Displays (HMDs) to warn users about entering hazard zones while performing an independent navigation task, in a negative feedback loop. Their user study indicated that people preferred a particular auditory feedback method when comparing it to visual alerts. Our study, on the other hand, investigates a positive feedback loop, which guides the user towards a goal, instead of alerting them against it, since it is crucial to differentiate cues and hazards [62].

Other recent works investigated the use of emulated speech as a spatial navigation interface for AR, either spatialized [30] or not [8]. Although having promising results, speech is arguably more limited than sonification, besides being easier to be misinterpreted in-the-wild, since speech messages often prevent the hearing of environmental sounds [34]. Finally, two works [17, 65] tested a particular AR HMD for spatialized sound guidance. Their results showed that people can consistently localize spatial sounds with an acceptable deviation for most tasks, although none of these works compared spatialization to any other attributes of sonification.

3 HYPOTHESES

The following hypotheses were the basis of this research:

H1: Sonification patterns have an effect on user performance (e.g. in time taken, and distance error) in micro-guidance tasks.

This hypothesis attempts to evaluate whether there is a quantitative impact of the proposed sonification methods upon a generic hand displacement task, and if so, what is its magnitude. For each user trial, two attributes are measured during any attempt: the two-dimensional distance from their selected mouse position to the audio target, and the total amount of time they take to finish the trial. By investigating this, it might be possible to ascertain whether there is an optimal combination of sound attributes to inform spatial positions to users non-visually, or perhaps that there is a trade-off between precision and time taken to complete a task.

H2: Sonification patterns have an effect on the perceived workload in micro-guidance tasks.

In order to assess the user-centered impact of sonification, this hypothesis aims to evaluate how the multiple combinations of audio attributes create workload for different users during the tasks, whether there is a quantifiable difference between them, and if so, what is it. By applying the NASA Task Load Index (TLX) [19] questionnaire to users after a set of trials, it is possible to quantitatively ascertain how they were impacted by a specific method in six different attributes: mental demand, physical demand, temporal demand, perceived performance, effort, and frustration. This way, it may be possible to gauge whether any of the sonifications are too cumbersome to be used in real life scenarios.

H3: Personal traits (e.g. sight condition, previous knowledge of musical theory, and speech audio speed preference) have an effect on the results of particular sonification patterns.

Although blindness generally leads to the emergence of superior auditory localization abilities, particular types of vision loss (such as in the frontal region of space) may have a detrimental effect on the development of accurate spatial hearing [26]. Additionally, age-related hearing loss is the third most prevalent chronic medical condition among older adults [56]. Thus, visually impaired or not, users should not be expected to have similar levels of auditory acuity. According to Neuhoff, empirical sonification researchers should evaluate their designs with a focus on the poorest rather than the best analytical listeners in their target user population [41]. Thus, this hypothesis attempts to evaluate which sound patterns present a correlation between high performance results and high aural skills among users, if any. Beyond trained ears, it is important to verify how crucial musical literacy is in the understanding of sonification metaphors [42].

H4: There is a correlation between the perceived user performance and their accuracy results in performing sonified micro-guidance tasks.

Although audio instructions are provided for each technique, it is expected that some users might misinterpret their appropriate mapping and scaling of information to each of the acoustic attributes used [42]. This hypothesis aims to investigate whether users are misled by any of the patterns, e.g., due to different sonification metaphors, inability to distinguish tempo, stereo or pitch variation. We posit that methods with high perceived performance and low accuracy results may lead to poor decision-making.

4 SONIFICATION METHODS

The methods proposed here are adapted to convey a position in a two-dimensional grid. Given a particular target in this grid, each method will sound differently according to the current user's mouse distance to said target, either by communicating a single 2D distance or the distance in each Cartesian axis (X and Y) via different sound components. In this implementation, for every method that uses varying pitch as an attribute to convey distance, each pixel of distance adds 1 Hz to its frequency, up to a maximum of 1 KHz (generally at an out-of-screen position). For our tests, the exploration space is not constrained by the computer screen size, nor its resolution, as the mouse movement is allowed to continue off-screen.

The spatialized methods - which arguably work as sound beacons [6] - are controlled by the Unity engine using its default stereo pan and a linear volume rolloff, while all non-spatialized methods maintain a constant user-defined volume. Similar to [44], a sine wave was used as a basis in most methods, in order to avoid confusion with any environmental sounds. In an effort to take into consideration Neuhoff's suggestions [41], a wide variety of acoustic dimensions were used to map the spatial parameters among the sonifications in order to compare them. The Audacity digital audio editor software was used to generate the wave files, while the Unity engine controls their modulation during the test.

Pitch Only (PO). The PO method uses a single sine wave conveying the 2D Euclidean distance from the mouse cursor to the target for both of their ears (headphone channels), with its pitch starting at 40 Hz (exactly atop the target) and increasing in pitch the further it moves towards any direction. Out of the seven sonifications tested, this is the only method that encodes a single 2D distance value instead of two single-dimension distances. This was done deliberately in order to have a simple and easy-to-understand sonification as a basis, specially since pitch is one of the most widely used acoustic dimensions in auditory display [42].

Spatialized Pitch (SP). Similar to the previous method, the Y axis in SP is represented by a pitch variation in a single sine wave, starting at 40 Hz (exactly atop the target) and increasing in pitch the further it moves towards either direction in its Cartesian Y position. The X axis, on the other hand, is demonstrated in a stereo spatialization fashion, similar to several works [4, 17, 43], e.g., if the target is positioned to the right of the mouse cursor, that same sine wave will only be heard in the right ear. Likewise, if the sound comes from the left ear, it means that the target is to the left of the cursor, with the sound flipping from side to side as the mouse position crosses over the target. Additionally, the amplitude (volume) decreases in a linear rolloff as the cursor distance to the target increases, and vice-versa.

Channel-based Pitch (CbP). In CbP, each ear is fed with a different sine wave concomitantly, each being independent of the other. The left channel represents the X axis, and the right channel, the Y axis. Each channel independently varies its pitch exactly as the earlier ones: with its pitch starting at 40 Hz (exactly atop the target) and increasing in pitch the further it moves towards either direction. This method is a channel-independent version of PO, serving the purpose of testing whether users can correctly interpret audio feedback independently on each ear, and if so, how cognitively demanding is that.

Alternated Pitch (AP). Like SP, the Y axis in AP is represented by a pitch variation in a single sine wave, starting at 40 Hz (exactly atop the target) and increasing in pitch the further it moves towards either direction in its Cartesian Y position. Meanwhile, the X axis has its distance from the target conveyed by an on-off alternation (tempo) of that same wave, always maintaining a constant "off" interval (0.03 s) and varying the length of its "on" signal to a sixth of the current pixel distance in seconds. The closer the mouse is to the target, the lesser will be the wave's mark-to-space ratio, i.e., this alternation will sound faster and more frequent. In the same sense, the farthest the cursor is to the target, the wave's mark-to-space ratio will be greater, i.e., this alternation will sound slower and less frequent. This method is similar to the beeping parking sensors of modern cars, and is also frequently used in sonifications to convey urgency [36].

Spatialized Alternated Pitch (SAP). SAP uses all of the same signal patterns as the AP method. However, on top of it, it also uses stereo spatialization for its representation of the X axis distance. This means that beyond the wave on-off alternation, the signal will only be sounded in the channel regarding the side of the target the mouse cursor is currently found in. Exactly like the SP method, SAP also uses stereo spatialization for its representation of the X axis distance, similar to several works[4, 17, 43]

Spatialized One-dimensional Harmonic-based Pitch (H1D). In the H1D method, two audio sources are used. One is a fixed 100 Hz sine wave, never changing its pitch. The other is a sawtooth wave with a varying pitch, changing according to the Y axis distance from the mouse cursor to the target, starting at 40 Hz (exactly atop the target) and increasing in pitch the further it moves towards either direction in its Cartesian Y position. Every time the pitch of the second wave nears an octave of the first, the two audio sources will harmonize with each other, conveying the user's progress toward the target. Additionally, it also uses stereo spatialization for its representation of the X axis distance, exactly like the SP and SAP methods. This method is a harmonic-based version of SP, serving the purpose of testing whether users can correctly identify harmonizing sounds and distinguish octaves in audio feedback, and if so, how cognitively demanding is that.

Spatialized Two-dimensional Harmonic-based Pitch (H2D). The H2D method uses all of the same signal patterns H1D does. However, it is also composed of a third audio source: a second sawtooth wave with a varying pitch, changing according to the X axis distance from the cursor to the target, also starting at 40 Hz (exactly atop the target) and increasing in pitch the further it moves towards either direction in its Cartesian X position. This means that three audio sources will always be heard, one being fixed and the other two harmonizing with it at various positions according to their respective axes. This method is a multisignal version of H1D, being introduced to the experiment in order to understand the human capacity of distinguishing multiple sound signals at the same time, and how cognitively demanding it is.

5 USER EXPERIMENT

The global COVID-19 pandemic forced us to pursue a remote user experiment. For the test to be run online, from the safety of users' homes, we assumed the use of common personal devices such as a mouse and headphones, as seen in Fig. 1-right. The decision to not use current MR HMDs was also based on a recent trend in interaction research that exposes the benefits of evaluating idealized user interfaces that are unencumbered by hardware limitations [15], aiming instead to set precedents for future devices. Although some BVI participants expressed concerns about not being familiar with using a mouse, we assured them they would not have to rely on a visual cursor, as the only purpose of the mouse was as a tracker of their hand position.

Thus, an application was developed and made available online to users with a self-guided version of the experiment that took on average 30 minutes to be completed. The data was anonymously collected in real time into a remote server. Extensive and clear instructions were given before the tests (e.g. the equipment required as set up, and how to download, run and perform the experiment) and prior to each method (what each parameter of the current sonification being tested meant). For the interface to be accessible by both sighted and BVI users alike, all instructions were narrated, they had to be heard at least one time, with the option to repeat at will, and no visual elements were displayed: only a black screen was rendered. Users were guided by the seven different sound patterns to perform mouse positioning tasks, following an audio target. The mouse cursor position was reset after every trial, in order to lessen its influence on the completion time of the following task.

In each task, participants were asked to move the mouse with their hand according to the best position they understood the audio was guiding them to. Fourteen repetitions were done for each of the sound patterns: the first four serving as a tutorial and the remaining ten contributing to our data collection. Each user had three out of the seven methods chosen for them by the application, with their orders being distributed among users via Latin square, as a means to minimize learning bias. The same was done with the order of the target positions among trials.

6 RESULTS

In total, 47 volunteers participated in the experiment, out of which 9 were Visually Impaired (5 totally blind and 4 had low vision). Although the subsample of BVI people is smaller, this is in line with other high-quality work in the field [8, 30, 31, 43, 54], who also note the difficulty of recruiting this demographic group, oftentimes including blindfolded sighted users as well, in order to make comparisons between the groups and assess a greater population in the experiments. Users were in an age group between 18 and 67 (Avg = 31.69, Std = 11.02). 10 of the subjects identified themselves as female (21.2%), 31 as male (65.9%) and 6 decided not to declare a gender (12.7%). 25 people claimed they were from Brazil (59.6%), 7 from Australia (14.9%) and 8 from different countries (Colombia, India, Netherlands, New Zealand, Peru, Philippines, Sweden, and USA).

20 people (42.5%) declared to have at least some knowledge of musical theory, while 23 (48.9%) claimed to have very little to none. 27 people (57.4%) said that they usually prefer listening to speech audio at regular speed (1x), while 14 (29.7%) claimed to prefer a faster-than-regular speed (1.25x or more). Only 3 subjects (6.3%) claimed not to make regular use of a mouse. No statistically significant difference was found in comparisons between age, gender, country,

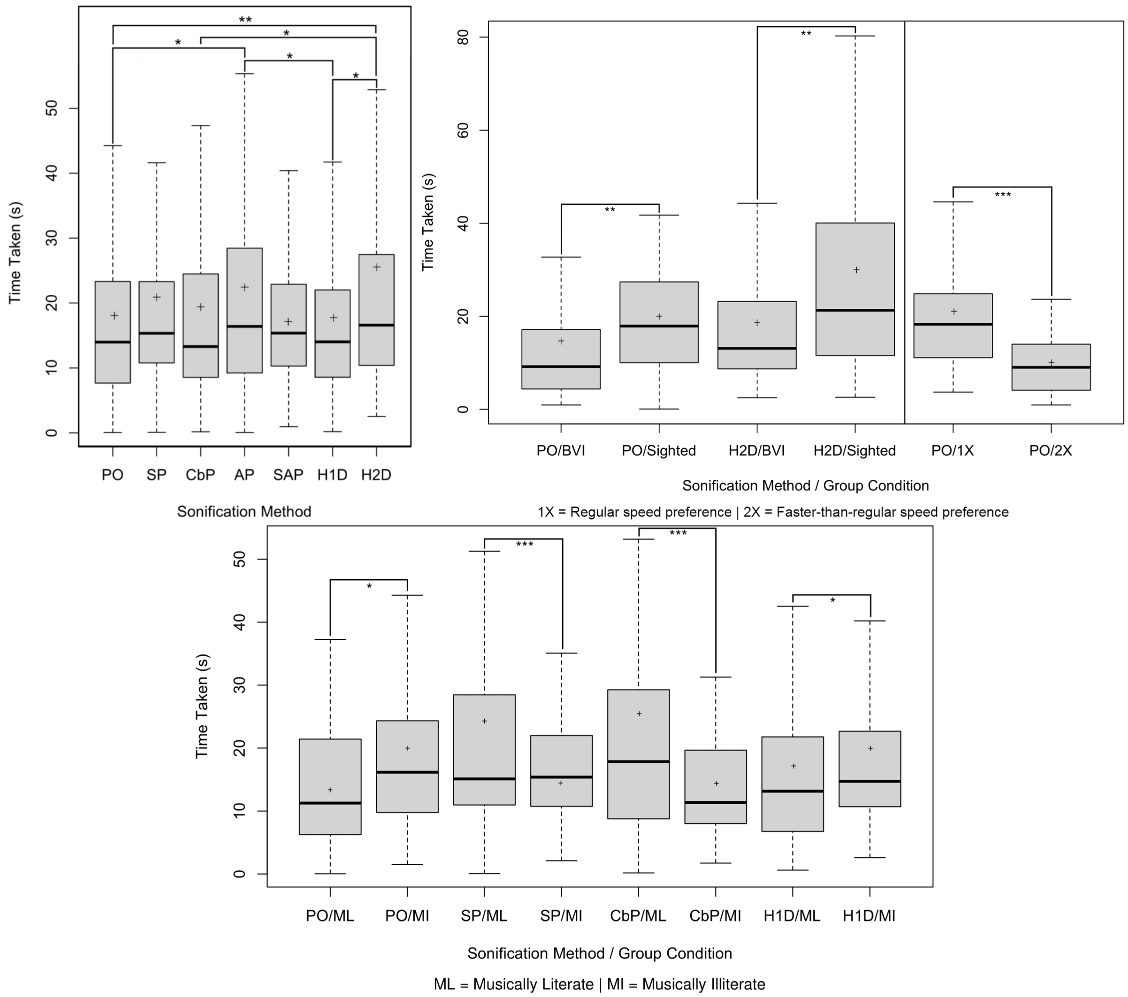

Fig. 2. Top Left: Time taken results for each sonification method. Top Right: Time taken results for each sonification method with participants being grouped by sight condition and audio speed preference. Bottom: Time taken results for each sonification method with participants being grouped by musical literacy. The p-value is denoted with asterisks, namely * for $p \leq 0.05$, ** for $p \leq 0.01$, and *** for $p \leq 0.001$.

and familiarity with a mouse device. The experiment results do not include any volunteers who reported any hearing disabilities. For all of the results of paired comparisons (sight condition, audio speed preference, and musical literacy), whenever the sizes of the two populations were not equal, the data used was based on the smaller of the two, whilst the remaining subjects for the other population were discarded in that particular comparison. The trials discarded were the ones that happened last.

6.1 Time taken

Considering the time taken results per trial for each method, and due to the data not being normally distributed based on Shapiro–Wilk tests, a Friedman test shows an effect of the sonification method used on the time taken to find the target [$X^2(6) = 14.368, p = 0.03$]. Wilcoxon rank sum tests with continuity correction were run for each pair of methods, the significant results can be seen in Fig. 2 (top left). Despite not having achieved statistically significant differences from any other method, SAP recorded the smallest average time, with H1D being a close second, being significantly less time-demanding than AP, and H2D. Similarly, CbP presented the smallest median, being significantly less time-demanding than H2D. Beyond that, the multiple significant differences in time taken results between the sonifications help supporting hypothesis **H1**.

Sight Condition: When dividing the users into two groups, a Wilcoxon test showed an effect of the user sight condition on the median time taken for PO, and H2D. This outlined that BVI users were significantly faster than their sighted peers in these two methods, as seen in Fig. 2 (top right).

Audio Speed preference: Regarding preference for speech audio speed, a Wilcoxon test demonstrated an effect of this preference on the median time taken only for the PO method, which presented significantly smaller time results for people used to faster-than-regular speech audio speeds, which can be seen in Fig. 2 (top right).

Musical Literacy: A Wilcoxon test also demonstrated an effect of the user knowledge of musical theory on the average time taken for PO, SP, CbP, H1D. However, this effect had slightly conflicting results between methods, since musical literates presented significantly smaller time medians for PO, SP, and H1D than their non-musical peers, while also having significantly greater time medians for the CbP method, as shown in Fig. 2 (bottom).

6.2 Distance from target

Considering the two-dimensional distance results per trial for each method, due to the data not being normally distributed based on Shapiro–Wilk tests, a Friedman test shows an effect of the sonification method used on the average distance to the target [$X^2(6) = 80.39, p < 0.001$]. Wilcoxon rank sum tests with continuity correction were run for each pair of methods, the significant results can be seen in Fig. 3 (top left). Notably, the Spatialized Pitch and the Spatialized Alternated Pitch methods had the most consistent results. SP also had the smallest average results, presenting significantly smaller distances than every other method except for PO. Additionally, the statistically significant differences in distance error results between the methods support hypothesis **H1**.

Sight Condition: When dividing the users into two groups based on sight condition, no method presented a significant difference in the comparison between the results of visually impaired and sighted people.

Audio Speed preference: Regarding preference for speech audio speed, a Wilcoxon test demonstrated an effect of this preference on the average distance from target for SP, AP, SAP, H1D, and H2D. Except for H1D, all of the aforementioned methods exposed significantly smaller average distance errors for the users who claimed to prefer faster-than-regular speech audio speeds, as seen in Fig. 3 (bottom). When looking at their median values, however, only the AP and SAP methods presented smaller median distance errors for the same group.

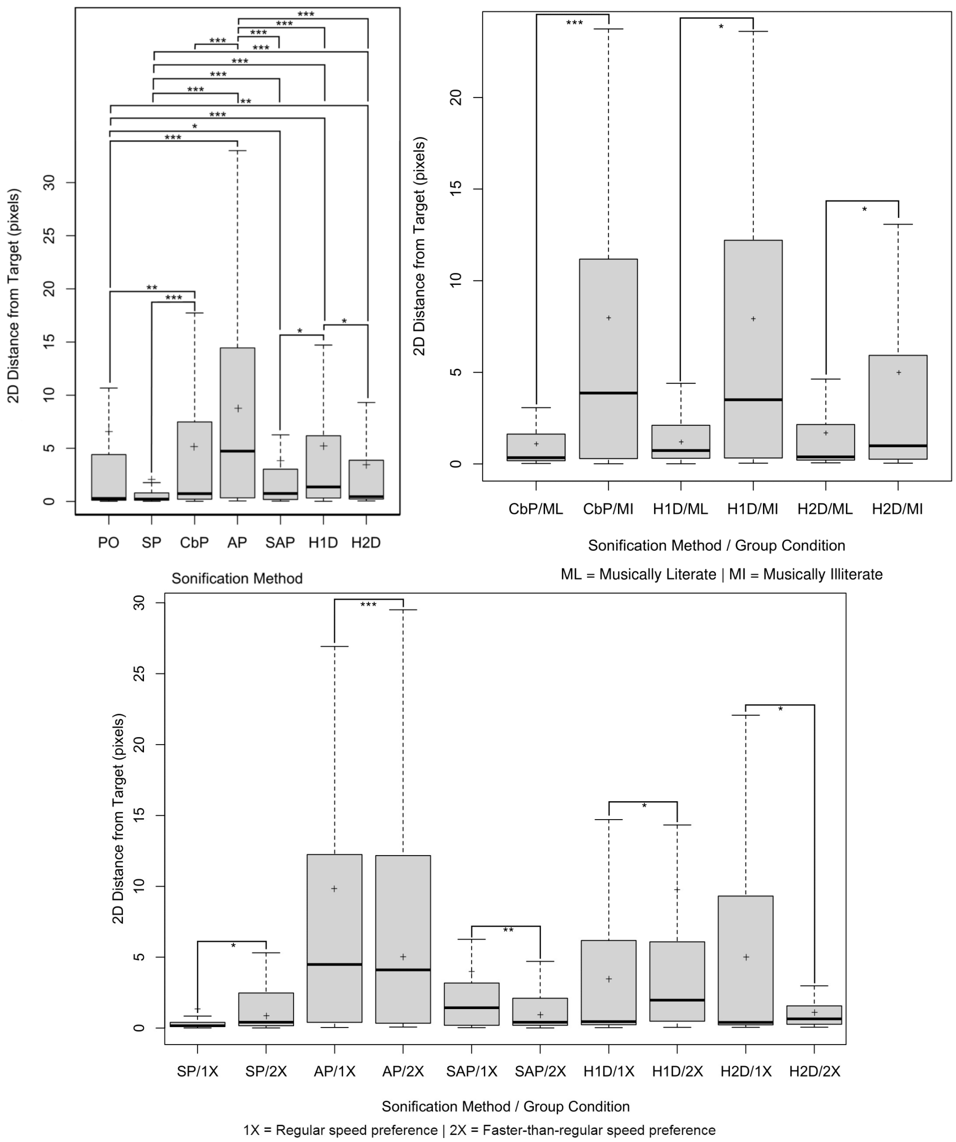

Fig. 3. Top Left: Distance error results for each sonification method. Top Right: Distance error results for each sonification method with participants being grouped by musical literacy. Bottom: Distance error results for each sonification method with participants being grouped by audio speed preference.

Musical Literacy: A Wilcoxon test demonstrated an effect of the user knowledge of musical theory on the median distance from target for CbP, H1D, and H2D. This highlighted that musically literate participants made significantly smaller distance errors than their non-musical peers in these three methods, as shown in Fig. 3 (top right).

6.3 NASA TLX

As for the NASA Task Load Index test [19], which measures the workload of generic procedures in 6 different attributes, the results can be seen in Fig. 4. The simplest methods – PO and SP, the ones we hypothesized did not require having high aural skills *per se* - achieved the smallest mental demand scores. In the overall TLX raw results, PO, SP and SAP were the only methods to score below 40. Both the highly divergent set of scores across the methods, and their multiple correlations between performance results and workload, support hypothesis **H2**.

Cognitive load: Taking into account the perceived Mental Demand attribute of the TLX test, we are able to make inferences about the cognitive load generated by each method [52]. In this particular score, as per Fig. 4, PO presented the least mental demand, while only PO, SP and SAP scored below 50. This arguably makes them the three sonifications with which users had the least cognitively loaded experiences during the tests.

User confidence: In decision-making, this work will consider the user confidence during a particular guess to be a function of the TLX perceived performance and the measured time taken for said guess [53]. In that aspect, based on Figures 4 and 2, it is possible to notice how PO and SAP had the two lowest results in perceived performance (since low scores mean better perceived performance) and in time taken. This arguably makes them the two methods with which users were most confident in estimating a mouse position during the trials.

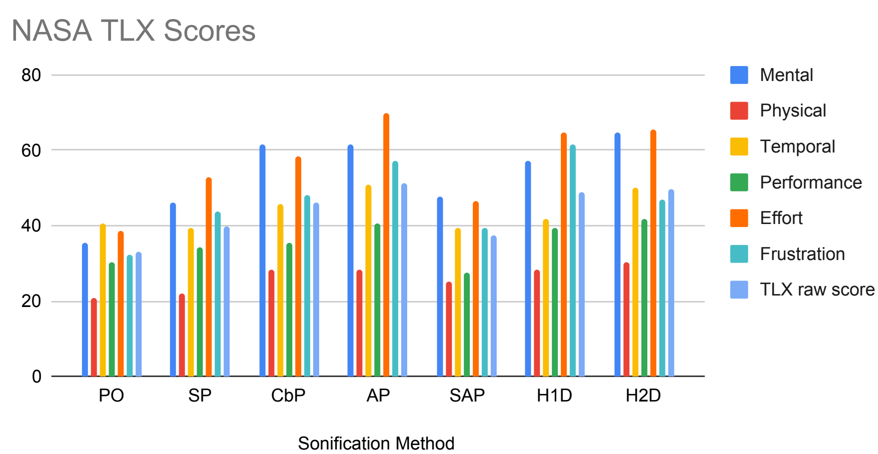

Fig. 4. NASA TLX scores for each sonification method.

7 DISCUSSION

Concerning the results disclosed in the previous section, some insights should be pointed out. Overall, alternation and spatiality proved to be very useful resources in sonified guidance for most subjects, with the successful exploration of harmonics highly relying on user musical literacy. In addition, users accustomed to faster-than-regular audio speed replay tended to have more precise performance with most methods. Besides these key findings, additional remarks were grouped by theme in the remainder of the section.

Pitch: Despite presenting the overall best results in workload, the Pitch Only method seems rather unreliable. Even though it was evaluated as one the two methods with which people were the most confident, it also presented the second highest average distance error, i.e., this self-reported performance notion was highly overestimated by users, which might have affected their decision making. Given this disparity, the Pitch Only method fails to prove hypothesis **H4**. Since this is the only sonification that encodes a single 2D distance value instead of two single-dimension distances, we can arguably assume this to be the reason of this lack of accuracy.

Originally meant as a spatialized version of PO to understand the impact of stereo on sonification, the Spatialized Pitch technique proved its value. Making each axis encoding independent from one another, SP exposed the least error-prone distance results, the second least mental-demanding and third smallest TLX scores out of every method. Its simplicity and the ease-of-understanding of its spatial features likely explain the performance results. This is consistent with the work of [6], which claimed that sensory navigation through audio beacons outperforms turn-by-turn navigation in the creation of mental maps.

While the Channel-based Pitch method may seem straight-forward since each axis is represented by a separate channel, it did not perform as expected. It was in fact the fourth most error-prone and had the fourth highest workload score. We can hypothesize that the under-performance of the method was caused by the constant confusion of which channel meant which axis, since multiple users mentioned they had to move the mouse in a straight line to be reminded of the proper order. Since there is no common-sense correlation between the pairs "left-right" and "X-Y", this confusion is reasonable. Despite that, with average results, it still shows that users can correctly interpret audio feedback independently on each ear, and with a reasonable cognitive demand.

Alternation: Based on a highly adopted sonification method, the Alternated Pitch technique was developed using the same metaphor as the parking sensor beeper of modern cars. Being the most error-prone method by far and having the highest workload, we hypothesize that some users assumed they were meant to get away from the faster beeping sounds, which could have made them confused about the entire AP sonification.

Openly designed as a stereo-based version of AP, the Spatialized Alternated Pitch added the stereo component onto the X-axis encoding of the former method. SAP reported the least error-prone results for a single axis, also achieving the third smallest 2D distance average and being the least time-demanding on average. Even though spatial audio lacks precision for conveying sound source elevation, which is crucial for locating objects in 3D space [36], this strongly shows the advantages of adding a spatiality component to a sonification method, instead of using spatiality as its only attribute.

Harmonics: To understand the human capacity of distinguishing multiple sound signals at the same time, two harmonic-based methods were introduced to the experiment: H1D and H2D. With varying results which were highly dependent on the user previous knowledge of musical theory, H1D presented the second highest median distance errors and the third highest workload score. Similarly, H2D was the most time-demanding method and had the second highest workload. This is all expected since the majority of the participants (58%) claimed to have little to no previous knowledge of musical theory, which allows us to assume their ears were not accustomed to identifying octaves.

Sight Condition: Mirzaei and Kaufmann [38] compared the results from deaf people and persons without hearing problems in similar spatial object localization experiments, in order to collect insights from their differences. Inspired by them, we briefly demonstrated in Fig. 2 (top right) how two sonification methods presented significant differences regarding time taken for the trials when grouping BVI and sighted individuals, while no significant differences were found for the distance error results comparing these groups. This might indicate that BVI individuals are more likely to complete a micro-guidance task in a smaller amount of time than sighted people, while neither of the two groups are likely to outperform the other in the precision for said task. These results are also consistent with Paré [43], which showed how blind participants were faster than their sighted peers while navigating through an obstacle course. Overall, it is important to note that the small BVI population is a weakness for this particular grouping of the results, which is likely why only a couple of methods presented statistically significant differences.

Musical Literacy: Concerning musicality, it is expected that musically literate people are more capable to outperform their non-musical peers when listening to sonified feedback [42]. Beyond knowing that prior musical experience overall helps to follow this sort of instructions more accurately, without being distracted by the combination of sonification feedbacks [1], it is also important to understand which of these combinations are most impacted by prior knowledge. When training is not available, we believe it is mandatory to offer users the option to be guided by methods that minimize this impact, following Neuhoff's suggestion of targeting the poorest analytical listeners [41]. Thus, it is necessary to avoid methods highly dependent of aural skills, such as CbP, H1D and H2D.

Speech audio speed preference: Perhaps one the most interesting set of results in this work is how the user preference on speech audio speed significantly affected the distance user performance on most of the sonification methods. People used to listening to speech audio on faster-than-normal speeds already have trained ears, hence it is expected this would affect their performance. As popular video streaming, podcast and chat platforms enable options for speeding-up videos and audio files, it is conceivable that more regular users get accustomed to it, enhancing their aural skills. All of the results based on condition groups reported here support **H3**.

8 FUTURE WORK

By considering the results of the user tests, we aim to adapt audio-based micro-guidance methods into real-life scenarios, aligned with the BVI community needs. As previously stated, one of the main weaknesses of this papers is the small BVI population, which we intend on increasing for future efforts. Given that the experiments were run remotely, and thus accommodations had to be made to include a larger population, the results here presented somewhat rely on the quality of the devices used by the subjects, especially their headphones and mouse. We aim to run the next experiments in person, to ascertain a standardized use of hardware.

Besides that, while expanding the assistive concept onto different day-to-day tasks, more dimensions will have to be considered and translated into sound format. For instance, while crossing a road, besides spatially positioning sounds at the starting and finishing ends of a crosswalk to convey its location and direction, other pieces of information are expected by the user [2, 10, 63]. To make sure it is safe for a BVI person to start crossing the road, the current traffic light status needs to be shown, as well as the user's updated position along the crossing, the incoming flow of people and even the sudden elevation of a curb.

By shifting the sound attributes, it is possible to create adaptive sound icons [59] that alert the listener to environmental characteristics as they walk along, performing their daily tasks. In a cooking scenario, on the other hand, a BVI person will need to interact with a hot oven and stove. When required to insert ingredients into boiling water, instead of having to position themselves along a one-dimensional line (as in the crosswalk example), their hands will have to be three-dimensionally positioned on top of a pot, far enough not to touch it and close enough not to splash boiling water [17].

By tracking the user hand in space, allied to spatial reasoning of the surrounding environment from a device such as the HoloLens 2, it is possible to inform users of the optimal hand position with an adaptation of the sonification methods presented in this work. For instance, a combination of audio sources aligning in pitch from the three cartesian coordinates (XYZ). By developing these methods as situated AR, we will be able to test different combinations in performance and usability via user experiments with BVI subjects, advancing the current micro-guidance state-of-the-art both in assistive technology and sonified augmented reality.

9 CONCLUSION

In the current *ex-situ* investigation of a two-part study, we presented the proposal, development, user testing, and analysis of sonification methods aimed at aiding decision-making in a micro-guidance context. The intended use case application is to help BVI users to locate 2D positions situated around them during manual tasks. Using a standard computer, mouse (acting as hand tracker), and headphones, users were able to interpret spatialized sound targets into proper locations in front of them without the need for speech instructions, having their performances assessed in multiple ways.

The approaches presented in this paper demonstrated to have different effects on manual position-finding tasks. The least error-prone and effort demanding methods were exposed, suggesting that specific techniques may be used depending on the task priority (e.g. time or accuracy). Suggestions have also been made as to proceed with further design by combining different methods and testing the application in AR-guided sonification experiments, with Orientation and Mobility experts, and BVI people as users. Furthermore, this was a relevant step towards sonified AR research, as there still is a gap in assistive and accessible applications for this demographic group that needs to be filled, and the proper interaction and sonification techniques need to be in place as to support its reach into the general public as to promote inclusion, which will be addressed in the upcoming *in-situ* investigation of this study.

We also aim to compare the performances of BVI and blindfolded sighted users within the same contexts of performing real tasks while being virtually assisted, considering what we learned thus far in the effects of sonification. We hypothesize that BVI users are likely to have behaviors in place facilitating performance in such tasks. By observing these behaviors, we aim to improve our micro-guidance methods for BVI users and their needs, developing assistive technologies that can be used in real-life scenarios, besides simulated environments.

Although some experiences of blindness in VR already exist [22, 55, 60], the use of MR head mounted displays has not been widely studied in the same context. For example, digital empathy towards BVI people has yet to be explored in MR contexts which, by raising awareness, could assist in promoting developers in designing experiences that are more inclusive and accessible.

ACKNOWLEDGMENTS

We would like to thank all experiment participants, the Associação de Cegos do Rio Grande do Sul for assisting in participant recruiting and the Australian Technology Network of Universities for providing funding through the ATN-Latin American PhD Scholarship.

REFERENCES

[1] 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 *2019 IEEE International Conference on Pervasive Computing and Communications (PerCom.* 1–10. https://doi.org/10.1109/PERCOM.2019.8767407

[2] Dragan Ahmetovic, Roberto Manduchi, James M. Coughlan, and Sergio Mascetti. 2017. Mind Your Crossings: Mining GIS Imagery for Crosswalk Localization. *ACM Trans. Access. Comput.* 9, 4, Article 11 (April 2017), 25 pages. https://doi.org/10.1145/3046790

[3] Robert Albrecht, R. Väänänen, and T. Lokki. 2016. Guided by music: pedestrian and cyclist navigation with route and beacon guidance. *Personal and Ubiquitous Computing* 20 (2016), 121–145.

[4] amit barde, matt ward, robert lindeman, and mark billinghurst. 2020. the use of spatialised auditory and visual cues for target acqusition in a search task. *journal of the audio engineering society* (august 2020).

[5] Costas Boletsis and Dimitra Chasanidou. 2018. Smart tourism in cities: Exploring urban destinations with audio augmented reality. *ACM International Conference Proceeding Series*, 515–521. https://doi.org/10.1145/3197768.3201549

[6] Gregory D. Clemenson, Antonella Maselli, Alexander J. Fiannaca, Amos Miller, and Mar Gonzalez-Franco. 2021. Rethinking GPS navigation: creating cognitive maps through auditory clues. *Scientific Reports* 11 (apr 2021). https://doi.org/10.1038/s41598-021-87148-4

[7] Robert F. Cohen, Valerie Haven, Jessica A. Lanzoni, Arthur Meacham, Joelle Skaff, and Michael Wissell. 2006. Using an Audio Interface to Assist Users Who Are Visually Impaired with Steering Tasks. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility* (Portland, Oregon, USA) *(Assets '06)*. Association for Computing Machinery, New York, NY, USA, 119–124. https://doi.org/10.1145/1168987.1169008

[8] James M. Coughlan, Brandon Biggs, Marc-Aurèle Rivière, and Huiying Shen. 2020. An Audio-Based 3D Spatial Guidance AR System for Blind Users. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Roberto Manduchi, Mario Covarrubias Rodriguez, and Petr Peňáz (Eds.). Springer International Publishing, Cham, 475–484.

[9] Marcelo De Paiva Guimarães and Valeria Farinazzo Martins. 2014. A Checklist to Evaluate Augmented Reality Applications. In *2014 XVI Symposium on Virtual and Augmented Reality*. 45–52. https://doi.org/10.1109/SVR.2014.17

[10] Manfred Diaz, Roger Girgis, Thomas Fevens, and Jeremy Cooperstock. 2017. To Veer or Not to Veer: Learning from Experts How to Stay Within the Crosswalk. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. 1470–1479. https://doi.org/10.1109/ICCVW.2017.174

[11] Jens Einsiedler, Oliver Sawade, Bernd Schäufele, Marcus Witzke, and Ilja Radusch. 2012. Indoor micro navigation utilizing local infrastructure-based positioning. In *2012 IEEE Intelligent Vehicles Symposium*. 993–998. https://doi.org/10.1109/IVS.2012.6232262

[12] Tristan C. Endsley, Kelly A. Sprehn, Ryan M. Brill, Kimberly J. Ryan, Emily C. Vincent, and James M. Martin. 2017. Augmented Reality Design Heuristics: Designing for Dynamic Interactions. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, 1 (2017), 2100–2104. https://doi.org/10.1177/1541931213602007

[13] Agebson Rocha Façanha, Ticianne Darin, Windson Viana, and Jaime Sánchez. 2020. O&M Indoor Virtual Environments for People Who Are Blind: A Systematic Literature Review. *ACM Trans. Access. Comput.* 13, 2, Article 9a (Aug. 2020), 42 pages. https://doi.org/10.1145/3395769

[14] Nathan Gale, Pejman Mirza-Babaei, and Isabel Pedersen. 2015. Heuristic guidelines for wearable augmented reality applications. *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, 529–534. https://doi.org/10.1145/2793107.2810309

[15] Jerônimo G Grandi, Zekun Cao, Mark Ogren, and Regis Kopper. 2021. Design and Simulation of Next-Generation Augmented Reality User Interfaces in Virtual Reality. In *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 23–29. https://doi.org/10.1109/VRW52623.2021.00011

[16] Renan Guarese, Franklin Bastidas, João Becker, Mariane Giambastiani, Yhonatan Iquiapaza, Lennon Macedo, Luciana Nedel, Anderson Maciel, Fabio Zambetta, and Ron Schyndel. 2021. Cooking in the dark: a mixed reality empathy experience for the embodiment of blindness. In *Proceedings of the 1st XR in Games Workshop* (New York). SBC, Porto Alegre, RS, Brasil. https://doi.org/10.5753/xr_in_games.2021.15680

[17] Renan Guarese, Franklin Bastidas, João Becker, Mariane Giambastiani, Yhonatan Iquiapaza, Lennon Macedo, Luciana Nedel, Anderson Maciel, Fabio Zambetta, and Ron van Schyndel. 2021. Cooking in the Dark: Exploring Spatial Audio as MR Assistive Technology for the Visually Impaired. In *Human-Computer-Interaction – INTERACT 2021*, Carmelo Ardito, Rosa Lanzilotti, Alessio Malizia, Helen Petrie, Antonio Piccinno, Giuseppe Desolda, and Kori Inkpen (Eds.). Springer International Publishing, Cham, 318–322.

[18] J Guerreiro, Daisuke Sato, Dragan Ahmetovic, Eshed Ohn-Bar, Kris Kitani, and Chieko Asakawa. 2019. Virtual Navigation for Blind People: Transferring Route Knowledge to the Real-World. *International Journal of Human-Computer Studies* 135 (10 2019), 102369. https://doi.org/10.1016/j.ijhcs.2019.102369

[19] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139 – 183. https://doi.org/10.1016/S0166-4115(08)62386-9

[20] Dhruv Jain, Sasa Junuzovic, Eyal Ofek, Mike Sinclair, John Porter, Chris Yoon, Swetha Machanavajhala, and Meredith Ringel Morris. 2021. *A Taxonomy of Sounds in Virtual Reality*. Association for Computing Machinery, New York, NY, USA, 160–170. https://doi.org/10.1145/3461778.3462106

[21] Watthanasak Jeamwatthanachai, Mike Wald, and Gary Wills. 2019. Indoor navigation by blind people: Behaviors and challenges in unfamiliar spaces and buildings. *British Journal of Visual Impairment* 37, 2 (2019), 140–153. https://doi.org/10.1177/0264619619833723 arXiv:https://doi.org/10.1177/0264619619833723

[22] Redouane Kachach, Pablo Perez, Alvaro Villegas, and Ester Gonzalez-Sosa. 2020. Blindness Visualizer: A Simulated Navigation Experience. *Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VRW 2020*, 504–506. https://doi.org/10.1109/VRW50115.2020.00107

[23] Fatima Zahra Kaghat, Ahmed Azough, Mohammed Fakhour, and Mohammed Meknassi. 2020. A new audio augmented reality interaction and adaptation model for museum visits. *Computers and Electrical Engineering* 84 (6 2020). https://doi.org/10.1016/j.compeleceng.2020.106606

[24] Brian Katz, Florian Dramas, Gaëtan Parseihian, Olivier Gutierrez, Slim Kammoun, Adrien Brilhault, Lucie Brunet, Mathieu Gallay, Bernard Oriola, Malika Auvray, Philippe Truillet, Michel Denis, Simon Thorpe, and Christophe Jouffrais. 2012. NAVIG: augmented reality guidance system for the

visually impaired. *Technology and Disability* 24 (11 2012), 163–178. https://doi.org/10.1007/s10055-012-0213-6

[25] Sulaiman Khan, Shah Nazir, and Habib Ullah Khan. 2021. Analysis of Navigation Assistants for Blind and Visually Impaired People: A Systematic Review. *IEEE Access* 9 (2021), 26712–26734. https://doi.org/10.1109/ACCESS.2021.3052415

[26] Andrew J. King. 2014. What happens to your hearing if you are born blind? *Brain* 137, 1 (01 2014), 6–8. https://doi.org/10.1093/brain/awt346 arXiv:https://academic.oup.com/brain/article-pdf/137/1/6/11141097/awt346.pdf

[27] Hicham Lahdili, Zine El, and Abidine Ismaili. 2019. Micro navigation system for smart parking area. *Int. J. Intelligent Enterprise* 6 (2019), 25–27. Issue 4.

[28] Young Hoon Lee and Gérard Medioni. 2015. Wearable RGBD Indoor Navigation System for the Blind. In *Computer Vision - ECCV 2014 Workshops*, Lourdes Agapito, Michael M. Bronstein, and Carsten Rother (Eds.). Springer International Publishing, Cham, 493–508.

[29] Laura Lewis, Sarah Sharples, Ed Chandler, and John Worsfold. 2015. Hearing the way: Requirements and preferences for technology-supported navigation aids. *Applied Ergonomics* 48 (2015), 56–69. https://doi.org/10.1016/j.apergo.2014.11.004

[30] Yang Liu, Noelle RB Stiles, and Markus Meister. 2018. Augmented reality powers a cognitive assistant for the blind. *eLife* 7 (nov 2018), e37841. https://doi.org/10.7554/eLife.37841

[31] Jacobus C. Lock, Iain D. Gilchrist, Iain D. Gilchrist, Grzegorz Cielniak, and Nicola Bellotto. 2020. Experimental Analysis of a Spatialised Audio Interface for People with Visual Impairments. *ACM Trans. Access. Comput.* 13, 4, Article 17 (oct 2020), 21 pages. https://doi.org/10.1145/3412325

[32] Alexander Marquardt, Christina Trepkowski, Tom David Eibich, Jens Maiero, Ernst Kruijff, and Johannes Schoning. 2020. Comparing Non-Visual and Visual Guidance Methods for Narrow Field of View Augmented Reality Displays. *IEEE Transactions on Visualization and Computer Graphics* 26 (12 2020), 3389–3401. Issue 12. https://doi.org/10.1109/TVCG.2020.3023605

[33] Sergio Mascetti, Andrea Gerino, Cristian Bernareggi, and Lorenzo Picinali. 2017. On the Evaluation of Novel Sonification Techniques for Non-Visual Shape Exploration. *ACM Trans. Access. Comput.* 9, 4, Article 13 (April 2017), 28 pages. https://doi.org/10.1145/3046789

[34] Sergio Mascetti, Lorenzo Picinali, Andrea Gerino, Dragan Ahmetovic, and Cristian Bernareggi. 2016. Sonification of guidance data during road crossing for people with visual impairments or blindness. *International Journal of Human-Computer Studies* 85 (2016), 16–26. https://doi.org/10.1016/j.ijhcs.2015.08.003 Data Sonification and Sound Design in Interactive Systems.

[35] Dario Di Mauro, Davide Maria Calandra, Francesco Cutugno, and Daniela D'auria. 2015. A 3D Audio Augmented Reality System for a Cultural Heritage Management and Fruition. *Journal of Digital Information Management* 13 (2015). https://doi.org/10.13140/RG.2.1.2171.9520

[36] Keenan R. May, Briana Sobel, Jeff Wilson, and Bruce N. Walker. 2019. Auditory Displays to Facilitate Object Targeting in 3D Space. International Community for Auditory Display, 155–162. https://doi.org/10.21785/icad2019.008

[37] Keenan R. May, Brianna J. Tomlinson, Xiaomeng Ma, Phillip Roberts, and Bruce N. Walker. 2020. Spotlights and Soundscapes: On the Design of Mixed Reality Auditory Environments for Persons with Visual Impairment. *ACM Trans. Access. Comput.* 13, 2, Article 8 (April 2020), 47 pages. https://doi.org/10.1145/3378576

[38] Mohammadreza Mirzaei, Peter Kán, and Hannes Kaufmann. 2021. Multi-modal Spatial Object Localization in Virtual Reality for Deaf and Hard-of-Hearing People. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 588–596. https://doi.org/10.1109/VR50410.2021.00084

[39] Sandy Murillo. 2015. How Do Blind and Visually Impaired People Get Around? In *Chicago Lighthouse Website*. https://chicagolighthouse.org/sandys-view/getting-around/ Accessed: 2021-10-29.

[40] Rabia Murtza, Stephen Monroe, and Robert J. Youmans. 2017. Heuristic Evaluation for Virtual Reality Systems. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 61, 1 (2017), 2067–2071. https://doi.org/10.1177/1541931213602000 arXiv:https://doi.org/10.1177/1541931213602000

[41] John G. Neuhoff. 2019. Is Sonification Doomed to Fail? International Community for Auditory Display, 327–330. https://doi.org/10.21785/icad2019.069

[42] John G. Neuhoff, Rebecca Knight, and Joseph Wayand. 2002. Pitch change, sonification and musical expertise: which way is up. In *In Proceedings of the 2002 International Conference on Auditory display*.

[43] Samuel Paré, Maxime Bleau, Ismaël Djerourou, Vincent Malotaux, Ron Kupers, and Maurice Ptito. 2021. Spatial navigation with horizontally spatialized sounds in early and late blind individuals. *PLoS ONE* 16 (02 2021). https://doi.org/10.1371/journal.pone.0247448

[44] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. 2019. WatchOut: Obstacle Sonification for People with Visual Impairment or Blindness. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 402–413. https://doi.org/10.1145/3308561.3353779

[45] Emmanouel Rovithis, Nikolaos Moustakas, Andreas Floros, and Kostas Vogklis. 2019. Audio legends: Investigating sonic interaction in an augmented reality audio game. *Multimodal Technologies and Interaction* 3 (12 2019). Issue 4. https://doi.org/10.3390/mti3040073

[46] Sonja Rümelin, Enrico Rukzio, and Robert Hardy. 2011. NaviRadar: A Novel Tactile Information Display for Pedestrian Navigation. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 293–302. https://doi.org/10.1145/2047196.2047234

[47] Ane San Martín and Johan Kildal. 2019. Audio-Visual AR to Improve Awareness of Hazard Zones Around Robots. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3290607.3312996

[48] Daisuke Sato, Uran Oh, João Guerreiro, Dragan Ahmetovic, Kakuya Naito, Hironobu Takagi, Kris M. Kitani, and Chieko Asakawa. 2019. NavCog3 in the Wild: Large-Scale Blind Indoor Navigation Assistant with Semantic Features. *ACM Trans. Access. Comput.* 12, 3, Article 14 (Aug. 2019), 30 pages. https://doi.org/10.1145/3340319

[49] Quan Shi. 2018. *Crowdsourcing based micro navigation system for visually impaired.* Master's thesis.

[50] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review. *Psychonomic Bulletin & Review* 20 (2013), 21–53.

[51] Marjan Sikora, Mladen Russo, Jurica Derek, and Ante Jurčević. 2018. Soundscape of an archaeological site recreated with audio augmented reality. *ACM Transactions on Multimedia Computing, Communications and Applications* 14 (8 2018). Issue 3. https://doi.org/10.1145/3230652

[52] Alexander Skulmowski and Günter Daniel Rey. 2017. Measuring Cognitive Load in Embodied Learning Settings. *Frontiers in Psychology* 8 (2017), 1191. https://doi.org/10.3389/fpsyg.2017.01191

[53] Romin W. Tafarodi, Alan B. Milne, and Alyson J. Smith. 1999. The Confidence of Choice: Evidence for an Augmentation Effect on Self-Perceived Performance. *Personality and Social Psychology Bulletin* 25, 11 (1999), 1405–1416. https://doi.org/10.1177/0146167299259006 arXiv:https://doi.org/10.1177/0146167299259006

[54] Lauren Thevin, Carine Briant, and Anke M. Brock. 2020. X-Road: Virtual Reality Glasses for Orientation and Mobility Training of People with Visual Impairments. *ACM Trans. Access. Comput.* 13, 2, Article 7 (April 2020), 47 pages. https://doi.org/10.1145/3377879

[55] Lauren Thevin and Tonja Machulla. 2020. Three Common Misconceptions about Visual Impairments. *Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VRW 2020*, 517–518. https://doi.org/10.1109/VRW50115.2020.00110

[56] Patricia A Tun, Sandra McCoy, and Arthur Wingfield. 2009. Aging, Hearing Acuity, and the Attentional Costs of Effortful Listening. *Psychology and Aging* (09 2009), 761–766. https://doi.org/10.1037/a0014802

[57] Christoph Urbanietz, Gerald Enzner, Alexander Orth, Patrick Kwiatkowski, and Nils Pohl. 2019. A Radar-based Navigation Assistance Device With Binaural Sound Interface for Vision-impaired People. 236–243. https://doi.org/10.21785/icad2019.023

[58] Vincent van Rheden, Thomas Grah, and Alexander Meschtscherjakov. 2020. Sonification Approaches in Sports in the Past Decade: A Literature Review. In *Proceedings of the 15th International Conference on Audio Mostly* (Graz, Austria) *(AM '20)*. Association for Computing Machinery, New York, NY, USA, 199–205. https://doi.org/10.1145/3411109.3411126

[59] Yolanda Vazquez-Alvarez, Ian Oakley, and Stephen A. Brewster. 2012. Auditory display design for exploration in mobile audio-augmented reality. *Personal and Ubiquitous Computing* 16 (12 2012), 987–999. Issue 8. https://doi.org/10.1007/s00779-011-0459-0

[60] Zhiquan Wang, Huimin Liu, Yucong Pan, and Christos Mousas. 2020. Experiencing and Navigating Virtual Reality without Sight. *Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VRW 2020*, 519–520. https://doi.org/10.1109/VRW50115.2020.00114

[61] Frederic L Wightman and Doris J Kistler. 1999. Resolution of front-back ambiguity in spatial hearing by listener and source movement. *The Journal of the Acoustical Society of America* 105 (apr 1999). https://doi.org/10.1121/1.426899

[62] Michele A. Williams, Caroline Galbraith, Shaun K. Kane, and Amy Hurst. 2014. "just Let the Cane Hit It": How the Blind and Sighted See Navigation Differently. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility* (Rochester, New York, USA) *(ASSETS '14)*. Association for Computing Machinery, New York, NY, USA, 217–224. https://doi.org/10.1145/2661334.2661380

[63] Haojie Wu, Daniel H. Ashmead, Haley Adams, and Bobby Bodenheimer. 2018. 3D Sound Rendering in a Virtual Environment to Evaluate Pedestrian Street Crossing Decisions at a Roundabout. In *2018 IEEE 4th VR Workshop on Sonic Interactions for Virtual Environments (SIVE)*. 1–6. https://doi.org/10.1109/SIVE.2018.8577195

[64] Wenjie Wu and Stefan Rank. 2015. Audio Feedback Design Principles for Hand Gestures in Audio-Only Games. In *Proceedings of the Audio Mostly 2015 on Interaction With Sound* (Thessaloniki, Greece) *(AM '15)*. Association for Computing Machinery, New York, NY, USA, Article 38, 6 pages. https://doi.org/10.1145/2814895.2814925

[65] A. T. Zijlstra. 2017. *Using the HoloLens' Spatial Sound System to aid the Visually Impaired when Navigating Indoors.* Master's thesis.

[66] Michael Zöllner, Stephan Huber, Hans-Christian Jetter, and Harald Reiterer. 2011. NAVI – A Proof-of-Concept of a Mobile Navigational Aid for Visually Impaired Based on the Microsoft Kinect. In *Human-Computer Interaction – INTERACT 2011*, Pedro Campos, Nicholas Graham, Joaquim Jorge, Nuno Nunes, Philippe Palanque, and Marco Winckler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 584–587.