Echo-house: exploring a virtual environment by using echolocation

Conference Paper · December 2018

DOI: 10.1145/3292147.3292163

Ronny Andrade

University of Melbourne

Jenny Waycott

University of Melbourne

Echo-House: Exploring a Virtual Environment by Using Echolocation

Ronny Andrade
The University of Melbourne
Australia
andrader@unimelb.edu.au

Steven Baker
The University of Melbourne
Australia
steven.baker@unimelb.edu.au

Jenny Waycott
The University of Melbourne
Australia
jwaycott@unimelb.edu.au

Frank Vetere
The University of Melbourne
Australia
f.vetere@unimelb.edu.au

ABSTRACT

The graphics-intensive nature of most virtual environments (VEs) prevents many people with visual impairment from being able to successfully explore them. A percentage of the population of people with visual impairment are known to use echolocation —sound waves and their reflections— to better explore their surroundings. In this paper, we describe the development of an echolocation-enabled VE (Echo-House) and evaluate the feasibility of using echolocation as a novel technique to explore this environment. Results showed that echolocation gave participants an improved sense of space in the VE. However, the evaluation also identified a range of orientation and mobility issues and found that participants needed additional support to gain confidence in their use of echolocation in the VE. Our findings suggest that with proper support, echolocation has the potential to improve access to VEs for people who are blind or visually impaired by revealing features that would be otherwise inaccessible.

CCS CONCEPTS

• Human-centered computing~Auditory feedback
• Human-centered computing~Accessibility systems and tools

KEYWORDS

Visual impairment, echolocation, virtual environments, exploration; games; accessibility.

ACM Reference format:
Ronny Andrade, Steven Baker, Jenny Waycott and Frank Vetere. 2018. Echo-House: Exploring a Virtual Environment by Using Echolocation. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction (OzCHI'18). ACM, New York, NY, USA, 12 pages.* https://doi.org/1234567890

1 INTRODUCTION

Technology has facilitated access to information and generated new possibilities for entertainment. Virtual experiences, such as virtual environments (VEs) and videogames, and their ever-increasing relevance in popular culture are a prime example of this [9, 18]. However, the high sensory capabilities that most virtual experiences require from their users leads to the exclusion of users with different physical or mental abilities. In the case of people with visual impairment, exclusion from virtual experiences is most evident, as they cannot rely on the visual output that most of these experiences produce.

People with visual impairment need to use nonvisual stimuli to construct a mental map of their surroundings, both in the physical and virtual world [34]. Despite the development of different approaches to navigation of virtual spaces, there are unexplored opportunities to incorporate navigation techniques that people with visual impairment use in the physical world into VEs.

In this paper, we aim to explore the feasibility of building a VE that allows for the simulation of echolocation —the use of sound waves and their echo reverberations — and evaluate the use of echolocation as a

OzCHI'18, 5-7 December 2018, Melbourne, Australia

https://doi.org/1234567890

novel technique to assist people with visual impairment to explore the environment. To do this, we present a prototype echolocation-enabled VE, Echo-House, and discuss findings from a small-scale proof-of-concept laboratory-based evaluation conducted with participants who have a range of visual impairments. While we have focused on the development of a VE, the overlap between VEs and videogames means that our findings could equally be applied to existing videogames or used to create new and more accessible games.

We begin by introducing the concept of sensory substitution, which underpins the way in which people with visual impairment interact with technology. This will be followed by a review of previous work on echolocation and a review of the existing trends in virtual experiences for people with visual impairment, focusing on videogames. We then describe the design decisions made in implementing the prototype of Echo-House and the methodology followed for its evaluation. Finally, results from the prototype evaluation are presented, and we discuss conclusions regarding the feasibility of using echolocation as a novel technique to explore VEs.

1.1 Sensory substitution

In order for people with visual impairment to navigate the world around them, it is necessary for them to use mechanisms that replace visual sensory input with other types of input. Sensory substitution is the translation of information from one sense to another, with auditory and tactile feedback being the most common senses used to replace visual feedback [26]. Research in this field is extensive, with Bach-y-Rita et al. [6] studying this field for decades, and trialing systems that use other sensory organs, such as the tongue [2, 5, 23], or the skin [3, 4] to replace visual information.

Specific sensory-substitution devices, that aim to assist people with visual impairment to navigate the physical world by replacing visual feedback with tactile or auditory feedback, have appeared in the literature: [6, 17, 24, 26]. These devices use cameras, ultrasound or optic emitters paired with a receiver to detect deflections of emitted signals. Once a signal is emitted and detected by the receiver, these devices produce either auditory or tactile feedback to indicate characteristics of the environment, such as distance between the user of the device and potential obstacles. 3D printed tangible interfaces have also been explored as means of sensory substitution to assist people with visual impairment to create a mental map of their surroundings [21, 36, 37].

The sensory-substitution approach has been taken to virtual worlds primarily by using auditory cues as substitutes for visual cues: Ohuchi et al. [29] presented a sound-based virtual maze that participants could explore with a game controller instead of by walking, in order to form a mental map of a location. Sánchez et al. [35] and Connors et al. [15] describe the design, implementation and evaluation of a VE that was modelled after a location in the real world. Players with visual impairment could explore the VE via the use of sound cues. Ohuchi et al. [29] and Connors et al. [15] agree that using sound-based virtual mazes modelled after locations in the real world improves the capabilities of people with visual impairment to create a mental map of an unknown location.

1.2 Echolocation

Echolocation is a skill that has been observed in bats and some aquatic mammals [40]. According to Kolarik et al. [24], and Kupers et al. [25], echolocation in humans is the ability to use self-generated sounds and their echo reverberations to build a mental image of the surrounding spaces. The sound can be generated through a mouth-click, a clap, footsteps, or by hitting a cane on the floor. The use of self-generated sounds and echoes stimulate the visual cortex of the brain of echolocators [24, 39], resulting in a form of sensory substitution by which visual input is replaced by auditory input.

The process of echolocation consists of two steps: transmission, which is the process of producing a sound, usually with the mouth, and reception, which is the process of listening to the sound emitted as it combines with its reverberation over surfaces. Over long distances, there is a perceptible gap between the moment the person listens to the sound they have produced and the upcoming echoes [24, 40].

Most individuals who echolocate use clicking sounds produced with their tongues [32]. The physical characteristics of the clicking sounds echolocators use have been analyzed. Kolarik et al. [24] and Thaler et al. [40] have found that the mouth-clicks produced by individuals who echolocate last for approximately 3ms to 10ms, with an interval of 250ms between clicks. Sound levels range from 60 to 108 dB SPL, with maximum energy in the frequency range of 8 kHz to 10kHz. Thaler et al. [40] have produced artificial clicks based on the parameters obtained from gathering a large sample of mouth-clicks produced by expert echolocators as a step towards the development of computational models of human echolocation.

It is estimated that 20 per cent to 30 per cent of blind individuals can use echolocation in varying degrees [38], and the functional benefits that echolocation offers include: the ability to determine distance, position, size, shape or material of surfaces or objects by detecting sound reflections over them [24, 38, 39]. These benefits highlight the potential for echolocation to provide a more nuanced form of sensory substitution in VEs when compared to existing auditory-cue-based approaches. Moreover, according to Thaler [38], previous research has found that echolocation improves spatial orientation skills in blind people under experimental conditions.

Some of the obstacles in learning echolocation are the amount of practice it requires, the cognitive load it creates and the fact that training methods are still in the early stages of development [30]. An initial attempt to teach echolocation by using technology is presented by Wu et al. [44], who have developed a mobile application that uses pre-recorded echoes to provide orientation and navigation clues for users exploring a grid-like maze environment. Based on a preliminary evaluation of this system, Wu et al. concluded that using echoes in a virtual maze could be a promising way of increasing people's awareness of echo cues. However, that research is still in development and, to the best of our knowledge, this is the only study to date that explores the potential use of echolocation in a VE. Notably, the aim of Wu et al. 's research is to train people to use echolocation, rather than using echolocation as a way of enabling visually impaired people to explore VEs.

In the following section, we explore the various existing approaches that have been taken to attempt to make virtual experiences, particularly videogames, more accessible to visually impaired or blind users. This is crucial, as people with low or no vision rely on the combination of different mechanisms, such as canes, guide dogs and echolocation, to make sense of their surroundings, it is likely that a range of approaches will be needed to navigate VEs.

1.3 Virtual experiences for visually impaired people

Virtual experiences occur outside the physical world and include virtual environments and videogames. There is an overlap between VEs and videogames; however, a videogame is usually contained within a VE, and a VE may or may not involve game play [9]. There is a growing body of literature that describes virtual experiences designed for people with visual impairment. [15, 20, 26, 29, 34, 41], where the terms "virtual environment" or "videogame" are used interchangeably. These experiences use sensory substitution as their basis for interaction by replacing visual with auditory output and the goal of some of these experiences is to assist people with visual impairment in the creation of a mental map of specific locations [15, 29, 34].

Without sensory substitution, VEs remain barely accessible to blind and visually impaired people [27]. However, a new trend in accessible videogames is emerging: audiogames. These are "computer games that feature complete auditory interfaces, so that they can be played without the use of graphics" [19]. A collection of over 700 of these games can be found on the website http://www.audiogames.net. The site offers games of different genres and hosts a thriving community of game developers and players with visual impairment.

Interaction in audiogames is based on earcons, and voice cues. Earcons are auditory representations of messages, functions, states and labels used in computer systems —and, by extension, in computer games— as alternatives to visual icons [10]. Blattner et al have categorized earcons as representational earcons, also known as auditory icons, and abstract earcons [10]. Representational icons are digitalized versions of naturally-occurring sounds, and abstract earcons are earcons constructed using a single pitch or a group of pitches.

According to White et al. [43], unlike videogames, audiogames lack a consistent set of earcons to convey cues from the game. This lack of consistency means that each new audiogame experience must be learned afresh. Audiogames aim to provide players with visual impairment with mechanisms to focus on exploration as a central part of the gaming experience, making navigation itself a "fun and open-ended part of the experience" [19], even if each new audiogame requires for the player to become familiar with a new set of earcons.

While audiogames can encourage exploration, the current interaction mechanisms limit the degree of freedom that players have for exploring. Ghali et al. [20] have found audiogames rely on the following three interaction patterns: verbal information, where the players are told when to press a key to produce an action; time-based mechanisms, where actions of players are timed by earcons; and constrained navigation mechanisms, where the player usually moves over a grid, with one press of a key corresponding exactly to one step, and movement occurring over a predefined set of positions and in four specific directions: up, down, left and right.

We propose the use of echolocation as a way to assist people with visual impairment in freely exploring VEs using a novel technique that provides a greater degree of freedom than the techniques currently used in audiogames. To achieve this goal, we have designed and evaluated Echo-House, a VE that supports the use of artificial mouth-clicks, claps and echo reverberations, enabling for the simulation of echolocation.

2: DESIGNING ECHO-HOUSE

Echo-House is the echolocation-enabled virtual environment (VE) designed for this study. It was created using Unity 5.6 and the echolocation features were developed with the SteamAudio plug-in, version 2.0 beta 5. Echo-House consisted of three levels with different layouts, as shown in Fig. 1 (not drawn to scale).

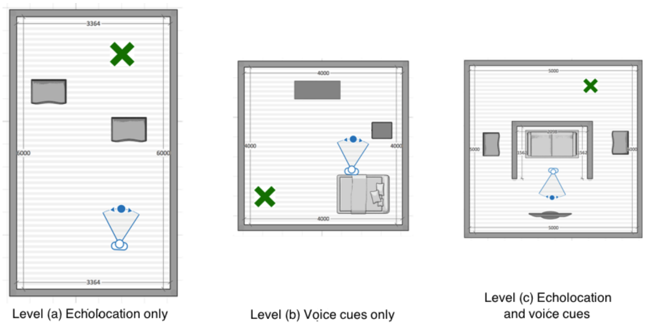

Fig. 1: Layouts of the three levels created as part of Echo-House.

2.1 Layout of Echo-House

The three levels of Echo-House were modelled after rooms in a house. Each level had a goal that participants needed to reach in order to move on to the next level (marked as an X on the layouts presented in Fig. 1). Level (a) represented a long corridor with two tall drawers as obstacles; level (b) was modelled after a bedroom with pieces of furniture around the room, and level (c) was modelled as a living-room with a TV set, a couch, cabinets and a u-shaped inner wall. The use of Unity and the SteamAudio plug-in allowed for the generation of 3D binaural sounds to indicate the position of the goals in each level, creating a sense that each goal had a fixed position in space. Each goal had a unique representational earcon associated with it. The earcon chosen for level (a) was the sound of a cat meowing; in level (b) participants could hear an alarm clock; and in level (c), the sound of candy being unwrapped indicated the position of the goal.

While we used 3D binaural sound to signal the goals of each one of the three levels, each level presented a unique experimental condition (described further in the Methodology section below). Level (a) allowed participants to use echolocation but did not give any feedback if they hit an obstacle, level (b) used voice cues indicating the name of the object participants collided with but did not provide echolocation support, and level (c) used both echolocation and voice cues with the name of the objects participants collided with. Table 1 summarizes the layout, auditory icon to indicate the goal, and experimental condition of each level.

Table 1: Summary of the three levels of Echo-House, their layout, auditory icon to indicate the location of the goal, and experimental condition

Level	Layout	Goal cue	Condition
(a)	Corridor	Cat	Echolocation only
(b)	Bedroom	Alarm	Voice cues only
(c)	Living room	Candy	Echolocation and voice cues

2.2 Exploring Echo-House

To explore each level of Echo-House, participants controlled a virtual avatar created from the Unity standard assets. To increase the feeling of presence in the VE, the virtual character could use one of these three auditory icons: footsteps, artificially synthetised mouth-clicks or claps. The three sounds could be produced in all levels; however, the sounds did not produce any echo on level (b). The footsteps were heard automatically as participants walked around Echo-House, whereas participants were able to control when to produce the mouth-click sound or the clapping sound by pressing the space bar or the Control key. Additionally, participants could use the Alt key to hear the 3D binaural cue that indicated the position of the goal. The character used in this study could move like most characters in first-person games: the arrow keys controlled the direction in which the character moved, and the keys a-w-s-d controlled the position of the camera that indicated the avatar's first-person perspective.

2.3 Echo capabilities in Echo-House

By using Unity and the SteamAudio plug-in, we were able to generate real-time physics-based sound reflections based on the geometry of each level. The sound reflections generated by the plug-in used a head-related transfer function (HRTF) to simulate how sound reaches a listener, and the delay produced when the sound reaches each ear. This delay is a key factor for echolocators to estimate distance of surrounding objects or obstacles [44] and through the simulation of the sound reflections and HRTF we gave Echo-House echolocation capabilities.

It must be noted, however, that Echo-House used a generic HRTF provided by the SteamAudio plug-in. While a customized HRTF could have improved participants' sense of presence, person-specific HRTFs are time-consuming to build [33], and the version of the plug-in used in Echo-House does not accept customized HRTFs.

Another feature of the SteamAudio plug-in is the ability to assign materials such as wood, ceramic or carpet to each element —wall, floor or piece of furniture— in Echo-House. Objects modelled with different materials absorb and reflect sound frequencies differently [24, 45]. This difference in sound reflection based on materials produced a more realistic experience, particularly for participants who know how to use echolocation in the physical world.

The artificial mouth-click the virtual avatar emitted was modelled upon the data and model proposed by Thaler et al. [40], which was based on a large number of samples collected from three expert echolocators. With the samples, Thaler et al. were able to produce a model of synthesized mouth-clicks optimized according to the parameters obtained from the three expert echolocators.

2.4 Additional considerations

Apart from the echolocation capabilities, Echo-House incorporated voice cues and representational earcons. At the beginning of each level, participants would hear a brief description of the room they were in and the goal they had to reach. We chose to use pre-recorded voice cues over artificial speech synthesis and representational earcons over abstract earcons in an attempt to create an interface that is more appealing for participants [34].

While Wu et al. [44] proposed a system that enabled the use of echolocation to explore a virtual space, there are significant differences between their implementation and the one presented in this study. In particular, Wu et al. aimed to facilitate the learning of echolocation whereas we are evaluating if it is feasible to explore a virtual space using echolocation as an interaction technique. These different goals resulted in different ways to move around the virtual space. While Wu et al. used a grid-like system, we gave players the ability to move in any direction and for any length they desired. Moreover, our approach to producing echoes is different. Wu et al. relied on pre-recorded echoes, while we produced real-time echoes based on the geometry of the virtual room.

3 METHODOLOGY

To assess whether echolocation can feasibly be used to explore a VE, we conducted a pilot evaluation study of Echo-House. In addition, we aimed to gain in-depth insights about participants' experiences of using echolocation in a VE. The study required participants to use the Echo-House VE for a pre-determined period of time and then participate in a semi-structured interview. Quantitative and qualitative methods were used to analyse data gathered from five participants. This project received approval from the Ethics Committee of the University (Ethics ID 1749810).

3.1 Participants

Participants were recruited via posts on online noticeboards, the University Student Equity and Disability Services, and word-of-mouth. Five participants were recruited for the study. They will be referred to as P1-P5 in this paper. Our decision to not engage with the community of visually impaired people outside of the university for this proof-of-concept evaluation made it difficult for us to obtain a larger sample. All participants self-identified as visually impaired. Table 2 summarizes the characteristics of the participants.

Table 2: Characteristics of participants, including their age range, impairment, whether they know how to echolocate and the assistive technology they used

	Age range	Impairment	Echolocator	Assistive technology
1	50-60	Low vision	No	None
2	20-30	Blind	Yes	Screen reader
3	50-60	Blind	Yes	Screen reader
4	20-30	Low vision	No	None
5	70-80	Low vision, hard of hearing	No	Magnifier

For this study, a decision was made not to use sighted blindfolded control subjects. While this had consequences in relation to the number of participants that could be recruited, previous studies that have compared the echolocation capabilities in blind and sighted participants have found that, on average, blind participants outperform sighted counterparts [25]. Moreover, people with low or no vision have learned a set of orientation and mobility techniques that sighted counterparts have not developed, as the latter rely almost entirely on their sight to navigate new spaces [31].

3.2 Procedure and data collection

The study consisted of a 45-minute playing session and a 15-minute semi-structured interview. While the interview was scheduled to last 15 minutes, P2, P3 and P5 were happy to go beyond the pre-established time. This commitment allowed them to provide additional rich insights on their interaction with Echo-House. During the playing session, participants had to explore the three levels of Echo-House described in Fig. 1 and were encouraged to think aloud to describe what was going on in their heads during the playing session. Think aloud data were included in the qualitative analysis.

This study followed an alternating treatments design [7] to ensure that the order in which participants explored the levels of Echo-House did not affect the final results. Participants were separated in two groups. One group completed the levels in the following order: (a), (b), (c), the other group completed the levels in following order: (c), (a), (b). To prevent carryover effects, or the effects of participants reaching the goal quicker after becoming familiar with the layout, each level had a different layout, as shown in Fig. 1. As already discussed and shown in Table 1 each level had a different experimental condition, with level (a) having echolocation-only, level (b) having voice cues only, and level (c) having echolocation and voice cues. The condition of no echolocation was introduced as a control, while the combined echolocation and voice cue condition allowed us to consider the impact that combining modalities would have on the participants ability to navigate the environment, as compared to echolocation only.

Qualitative and quantitative data were gathered to evaluate the use of echolocation to explore the VE. The participants' playing session was automatically recorded. Video and key logs were generated. Upon completing the playing session, participants took part in a semi-structured interview. If participants knew how to echolocate in the physical world, they were asked a set of slightly different questions that focused on understanding this skill and how they correlated it to the VE. The interviews were audio-recorded after obtaining participants' consent.

Quantitative data obtained from the video recordings of participants' playing session and log files were analyzed using descriptive statistics graphics. The data obtained from the think aloud process and the interviews were analyzed using thematic analysis [1], where the analysis focused on identifying themes related to participants' ability to explore the VE.

3.3 Physical set up

Fig. 2 shows the setup used during the evaluation. Participants sat in front of a monitor, a mouse and a standard English-layout keyboard. The monitor and the mouse were not connected to a computer but were there to act as props and emulate a gaming desktop PC. The keyboard was connected to a laptop that ran Echo-House. The laptop was positioned so that even if participants had low vision they could not look at the screen.

Participants used semi-closed headphones during the gameplay. This type of headphone partially isolates external noises. The pair of headphones used during gameplay provided a flat frequency response, without creating the bass enhancement which is common in most commercial headphones. The echo produced by the elements in the VE was amplified so that it would be easier for participants to detect it.

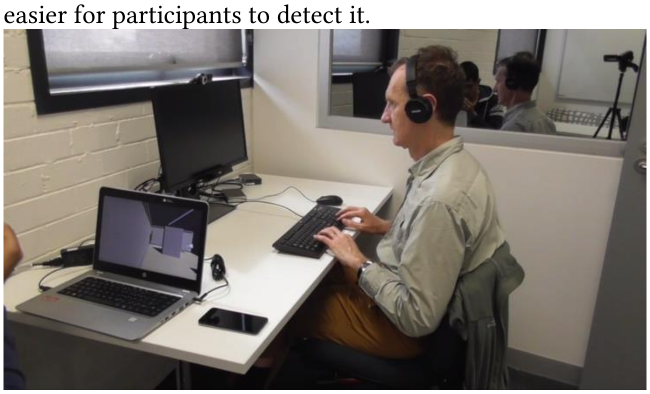

Fig. 2: Participant interacting with Echo-House

All participants had been contacted through digital media before the evaluation, so it was safe to assume that they would be familiar with the use of a standard computer keyboard. Moreover, Sánchez et al. [34] have found the keyboard to be the most common input device used in games for people with visual impairment.

3: RESULTS

In this section, we describe quantitative and qualitative results from the evaluation of Echo-House. Quantitative results include the time participants took to complete each level, and the number of times participants used the different audio cues provided in Echo-House. Qualitative findings center around how echolocation in the VE compares to echolocation in the physical world, participants' opinions on the way they interacted with Echo-House, and participants' need for proper scaffolding in order to take the most advantage of echolocation in Echo-House.

3.1 Quantitative Results

Descriptive statistical analyses of the auto-recorded videos of participant playing session and log data focused on two key aspects: completion time of each level, and number of times participants used the different audio cues available: artificial mouth-clicks, claps, or a 3D binaural sound to indicate the position of the goal. Table 3 shows completion time for each level, and Fig. 3 shows the use of the sound cues.

Table 3 shows that all participants except P2 took less than 30 minutes to complete the three levels of Echo-House. The table also shows P2 and P3 took the longest to complete level (a). This level offered the echolocation-only experimental condition, so P2 and P3 spent several minutes familiarizing themselves with the artificial mouth-click, the clapping sound, and the echo these sounds produced. Finally, the table shows that all participants, except P2 completed level (b) faster than the other levels; level (b) was the voice-cues only control condition.

Table 3: Completion time by level and participant in minutes and seconds

Participant	Level (a)	Level (b)	Level (c)	Total
1	6:25	4:09	17:56	20:30
2	10:24	14:48	22:46	47:58
3	17:30	3:04	9:07	29:41
4	6:40	5:02	8:33	22:15
5	6:59	4:36	9:43	21:18

Fig. 3 shows the total number of times participants used the sound cues available to them: artificial mouth clicks, clapping sounds, or a 3D binaural sound indicating the position of the goal in the virtual space. The Fig. shows that participants who were blind (P2 and P3) generated many more mouth-click and clapping sounds relative to the participants with low vision. This predisposition to use the echolocation cues could be attributed to the fact that P2 and P3 have used echolocation for years. The Fig.

shows that P2, who used the mouth-click and clapping sounds the most, was also the one who used the 3D binaural cue that indicated the position of the goal less often, while P5, who was hard of hearing, relied almost entirely on the 3D binaural cue that indicated the position of the goal. This suggests that participants each had their own preferences for how to navigate through the levels, with the two participants who already used echolocation using the echo cues more often than the other participants. It was to be expected that P5 preferred not to use the echolocation cues, as this participant was hard of hearing; however, P5 could successfully perceive the binaural 3D sound that signaled the position of the goal.

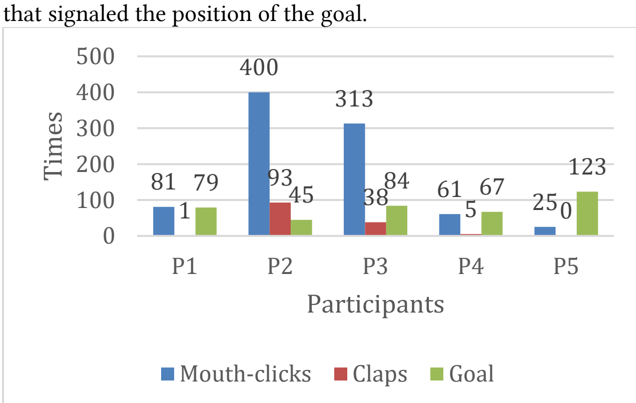

Fig. 3: Total number of times participants used the artificial mouth-clicks, the clapping sounds, or the 3D binaural cues indicating the position of the goal

3.2 Qualitative Results

The following sub-sections describe the common themes found across the interviews with the different participants. These themes include: echolocation in the VE and how it compares to echolocation in the physical world; comments on movement and interaction; and the need for additional support, or scaffolding, to successfully use echolocation in a VE.

3.2.1 Echolocation in the virtual environment. All participants except P5, who is hard of hearing, were able to perceive the echo generated in the VE after emitting the mouth-click or clapping sounds in the levels that supported it. The echo was described by the participants as giving them a sense of space. While exploring level (a), which simulated a long corridor and where echolocation was the only feedback available, P3 stated "The room seems very deep with the [mouth] click", and as soon as P3 reached level (c), which simulated a living-room and provided echolocation and voice cues support, they said "Bigger… much more echo here". Similarly, when reflecting on the experience, P4 said "the echo gave me a sense of space, but I found it hard to work out how far away I was. I think sometimes it was louder than others".

Regarding spatial navigation of the VE, P1 said "by the amount of walking I had to do, [level (a)] felt like a long corridor". Once again, this participant's perception of the environment matches the actual layout. Moreover, the gameplay think aloud comments of P2 and P3 showed them successfully identifying walls and drawers in different levels of Echo-House. For example, P2 said: "it looks like there's something up to the right" when exploring level (a) with the artificial mouth-click and standing next to a wall on their right.

Participants commented on the need to learn to understand the sound emitted and the echo received. P1 said: "I didn't know how to interpret the different sounds to find my way out". This is true for participants who can echolocate as well; after finishing the evaluation of Echo-House, P3 declared: "I was probably learning what that click sounded like and of course that was going to take some time". This shows that participants needed time to become familiar with the echo produced in the VE.

Echolocation works differently for each person because of the uniqueness of the mouth-click and HRTF of each individual [32, 33]. This was evident when P2 and P3, both expert echolocators, interacted with Echo-House. For P3, the artificial mouth-click used in the VE was less sharp than his own: "The click sound isn't as sharp as the one that I usually make myself [even though] I'm not sure if my click is unusually sharp", they said.

Participants also obtained different degrees of spatial information from the sounds available. P3 declared "I get more information from the clap rather than the click, which is unusual. I get more information usually from the click myself", whereas P2 felt they could obtain more information from the mouth-click.

People who use echolocation make sense of their surroundings by listening to the delay of the echo coming back to their left and right ears. P2 and P3 felt this delay, especially on level (c), which simulated a spacious living-room, with P3 saying "[this level] probably feels bigger so I'm hearing more of the sound delay", and P2 saying: "on the left is more delayed that's how you can tell there's more space on the left", while standing on the bottom right corner of level (c). However, regarding the accuracy of the delay, P2 added: "I think maybe the simulator is too quick with the feedback because it's all about the time; the timing of the feedback indicates what the distance is and it's such a small change… it's just happening so quickly. I really have to be listening for it".

Our analysis suggests that while participants successfully perceived the echoes generated and could get some idea of the VE, the echolocation accomplished with Echo-House did not necessarily feel like the echolocation they were used to from their experience in the physical world. P3 said: "So normally as I move closer to a wall I would hear it more clearly than I'm hearing it here". When asked whether using echolocation in the virtual world would affect their echolocation skills in the physical world, P3 said it would not.

3.2.2 Moving and interacting. Participants pointed out the need for more information from the environment and the way in which the virtual character interacted with it. When colliding with obstacles on level (a) (corridor, echolocation only), P1 wondered: “Am I walking past cabinets or getting into the cabinets? I didn’t know if I was walking past [the cabinets] or actually banging into them”. Regarding how the virtual avatar interacted with the environment, this participant declared: “I didn't know how tall things were or if I was just walking past them or if I could climb over them or if I had to go all the way around them”.

P1, P2 and P5 pointed out other physical differences they felt when exploring Echo-House when compared to the physical world. P1 mentioned the ability to hear a sound to indicate where the goal was: “the ability to make the target make some sound, I don't think you could necessarily count on that in real life but it's good that you could do that in the virtual environment”. P2 highlighted the lack of the physical feeling of the room: “if I walk into this [lab] room, I can tell there's concrete under the carpet, the carpet is not a fancy carpet that feels good to walk on, and that [feeling] was not at all on the computer”. P5 explained how they get information from the world via the sense of touch: “I use the balls of my feet to feel the rough things on the edges of the train platform [and] not get lost when I walk from one platform to another. I use the walking stick to prep where steps are, where the edges are. It is a very tactile thing and I was missing this feedback in the game”. This matches previous findings regarding the sense of presence in VEs [33, 42].

Regarding using the arrow keys to move the virtual avatar around the environment, P2 initially expressed concerns about not having a frame of reference for how each key press mapped to the virtual avatar. The virtual avatar required longer presses to walk and P2 was pressing the arrow keys softly for very brief intervals. After a while, the participant figured out for how long to press the keys and declared “It took me a while to get used to them… and then I went into this sort of more explorative mode”. This is visible in the auto-generated video recordings of P2 playing session, where at first, they moved very slowly and after discovering how the key-press affected the movement of the virtual avatar, they started moving more swiftly. While P3 did not comment on the effect of the key-press in the movement of the virtual avatar during the post-interview, video analysis of their playing session revealed a similar behavior, starting with short arrow key presses then using longer key presses as they understood how much it was necessary to press an arrow key for the virtual avatar to move. A similar behavior was not observed on P1, P4 or P5, who have low vision; these participants seemed to move more confidently in the VE.

Participants commented on the controls of the virtual avatar and the separation of the walking and viewing directions. P1 declared: “I wasn't very clear on the difference between moving my head and moving my body”. P2 pointed out the importance of feedback on the position of the head: “ok so I need to remember which way I'm looking… That's tricky, that's a problem in some sense because without any feedback of where it is you can forget”.

3.3.3 Need for additional scaffolding. Participants agreed that spending more time interacting with Echo-House or having a training phase would have improved their understanding of the mouth-click and the echolocation capabilities of Echo-House. According to P3: “it took a while to get used to [the click]. This probably means that for someone to really use this it's just a little bit of training because you're incredibly attuned to the dynamics of your own head for the echolocation”. For P4 “it’s more a matter of getting used to what the difference in sound means”.

P2, P3 and P4 agreed that it is necessary to incorporate additional orientation cues. P4 commented: “if I started turning my head I would lose track of where I was”. Similarly, P3 said: “The main thing that I missed was some feedback on the head position”, and P2 said: “there isn't enough information to tell which one is the direction I'm facing, and I would just like to have some more clue”.

P2 and P3 proposed different solutions to the problem of needing additional orientation cues. P2 suggested: “just hold down Control and left or right to take you to the nearest compass point. This could give me a certain orientation so if I do go too far to the left it'll bring me back on track”. P3 suggestion was: “it could be quite easy to add a key that just re centers the head so that you now have a baseline”. Additional general suggestions from participants included: giving more detailed instructions and voice descriptions of each level (P1, P3, P5) and mapping one press of the arrow keys to one footstep (P2, P3, P5).

4 DISCUSSION

Thaler and Goodale [39] argue that echolocation can help blind individuals represent “spatial relationships between objects, or the spatial structure of a scene”. Virtual environments (VEs) and videogames contain complex scenes that are almost impossible for people with visual impairment to freely explore, and audiogames provide limited mechanisms for exploration. Addressing these limitations, echolocation could help people with visual impairment make sense of scenes and objects in VEs or videogames by helping them understand the relationships between the spatial objects in these scenes.

The aim of this study was to analyze if it is possible to build a VE that enables participants to simulate echolocation, and to evaluate the feasibility of using echolocation to explore this environment. To achieve this aim, we designed an echolocation-enabled prototype of a VE called Echo-House and tested it with a group of five participants, including two blind participants who are expert echolocators (P2, P3) and three participants with low vision (P1, P4, P5).

One notable limitation of our evaluation study is that we were only able to recruit a small number of participants. This is because we aimed to recruit individuals who were located on campus of the university where the research took place, and who were blind or had visual impairments, and we decided not to use sighted blindfolded participants to guarantee the relevance of our results. These criteria made it difficult to find suitable participants. Furthermore, this was a feasibility study and we wanted to find out if using echolocation in a VE was possible at all before engaging a larger number of participants.

Sample size is a topic of heated debate in qualitative research, with small samples having limited acceptability in the research community [13, 16]. The key to determining the ideal sample size in qualitative studies is data saturation, which is the point at which no new information or themes appear from conducting more interviews [11, 22]. Guest et al. [22] Boddy [11], and Morse [28] argue that, in order to determine the sample size of a qualitative study, the scope of the study, the quality of the data, the time spent with each participant, and the homogeneity of the population being studied should be considered. Even though the sample size of this study is small, we consider that the scope of this study — being a proof-of concept laboratory-based evaluation—, the time spent with the participants —no less than an hour, with P2, P3, and P5 staying for longer—, the rich quality of the data provided by expert echolocators, and the relative homogeneity of studying a population who share a common trait and similar beliefs, justify the small sample size of this study.

While it could be argued that a participant who is hard of hearing should not be taken into account in a study investigating echolocation, we have decided to keep the insights provided by P5, as they illustrated the limitations of the technique, and the fact that this participant managed to complete all the levels highlights the importance of providing additional means of feedback that are not always present in current VEs.

All participants completed the three levels available in Echo-House, with level (b), the level without echolocation, being the level completed the fastest by most participants, as shown in Table 3. While this result may at first seem counter intuitive, qualitative data reveals this result can be explained by participants taking more time to familiarize themselves with the echolocation features of the program and subsequently spending additional time interrogating spaces before navigating to the goal. The extent to which participants used echolocation varied significantly, with the two expert echolocators using the artificial mouth-click far more often than the participants with low vision, as seen in Fig. 3. The fact that some participants completed the levels using the artificial mouth-click sparingly and that the level that did not offer echolocation capabilities was completed faster seems to indicate that 3D binaural sound is sufficient to locate goals in a VE. However, echolocation was nonetheless useful, especially for participants who can echolocate in the physical world, as it enabled them to identify features such as the drawers in level (a) (echolocation only) and proximity to walls. Without the ability to use echolocation in the VE, these features would have been inaccessible for them. Therefore, we consider echolocation in VEs has the potential to improve the ability of a blind person to explore a VE, albeit research with a larger sample will be necessary to make more definitive claims.

Echolocation helps people with visual impairment create a mental map of their surroundings by letting them estimate distance to obstacles [38]. While our evaluation of Echo-House did not assess participants' ability to create a mental map of the VE, incidental findings, such as participants describing level (a) as being long, or their ability to perceive the drawers or their proximity to walls on level (a), or their description of level (c) as feeling "bigger", suggest that echolocation could help participants create a mental image of a VE. Future research could focus on assessing the ability of participants to create mental maps of a VE through echolocation.

In education, scaffolding refers to providing aids to assist students in learning a new skill [8]. We believe that with sufficient scaffolding, the usefulness of echolocation to explore VEs could be increased. Moreover, participants identified possible scaffolding techniques that could be introduced in future iterations of Echo-House, such as providing a training phase or providing additional location cues. As echolocation varies from person to person due to the unique characteristics of each individual's mouth-click sound and HRTF [32, 33], providing customizable click sounds, where players can change the frequency and pitch of the sound, could improve their usefulness.

Our participants noted an impaired sense of presence in the VE due to the lack of tactile feedback and the different feeling of echolocation in the virtual world. This is analogous to the need for visually impaired people to make use of different mechanisms, such as canes, guide dogs and echolocation, to create a richer sense of presence in real world scenarios. Presence refers to an individual's sense of physically being in a VE [12, 33]. Similar results have been found before in other evaluations of VEs [33, 42]; therefore, this detachment from the VE was to be expected. Moreover, using the arrow keys to control the direction in which the virtual avatar walked and the keys a-w-s-d to control the direction it was looking caused confusion in participants. Participants suggested simple alternatives to reduce confusion; they recommended mapping the pressing of the arrow keys with individual footsteps and adding more environmental cues to have a baseline they could go back to if they got disoriented. Additionally, it is possible that the use of different input devices such as joysticks, head-tracking devices, or allowing participants to walk-in-place could make interaction clearer, whilst increasing the sense of presence [14, 42],

While there are many studies on echolocation in the physical world and its perceived benefits [24, 25, 32, 38-40] and some studies exploring the use of sound to navigate VEs [15, 26, 29, 34], the only study we have found that uses echolocation to navigate VEs is the preliminary research by Wu et al. [44]. Our study provides an in-depth analysis of blind and visually impaired participants using echolocation in a VE, thereby contributing to this emerging and important research area. Moreover, the fact that the software plug-in used for the development of Echo-House is available for two main game-developing platforms means that the results of this research can easily and quickly be added to existing or future VEs and videogames, making the findings of this research applicable in contexts other than laboratory evaluations.

As designers continue to develop technology to address the needs of people with different capabilities, they will have to explore novel interaction techniques to improve accessibility. These techniques are important not only for providing access to digital information and services, but also for ensuring people with different abilities are able to take part in activities that have become a central part of our digital culture, such as playing videogames and exploring virtual environments.

5 CONCLUSION

In this paper, we have outlined the design of an echolocation-enabled virtual environment (VE). and presented the results of a proof-of-concept laboratory-based evaluation of this environment. Our initial evaluation revealed that while 3D binaural sound allows participants to find specific goals in a VE, echolocation enables them to explore the environment and identify certain features that would be otherwise imperceptible to them. Additionally, our evaluation revealed the need to provide proper scaffolding for participants to be able to make the most out of the echolocation capabilities of the VE. Other issues such as the need to establish a strong sense of presence and defining good controls and feedback mechanisms were also uncovered.

Given the importance of virtual experiences in our modern culture, it is necessary to design technologies that support access for everyone regardless of their sensory capabilities. If the proper support is provided, echolocation could enable people with visual impairment to access virtual environments.

Acknowledgements

Ronny Andrade's work was supported by a Melbourne Research Scholarship and the Interaction Design Lab – The University of Melbourne. We would also like to thank our participants for their time and rich insights.

REFERENCES

[1] Jodi Aronson. 1995. A pragmatic view of thematic analysis. *The qualitative report*, 2, 1. 1-3.

[2] Paul Bach-y-Rita, K Kaczmarek and K Meier. 1998. The tongue as a man-machine interface: A wireless communication system. in *Proceeding of the 1998 international symposium on information theory and its applications*, 79-81.

[3] Paul Bach-y-Rita. 1996. Conservation of space and energy in the brain. *Restorative neurology and neuroscience*, 10, 1. 1-3.

[4] Paul Bach-y-Rita. 1999. Theoretical aspects of sensory substitution and of neurotransmission-related reorganization in spinal cord injury. *Spinal Cord*, 37, 7. 465.

[5] Paul Bach-y-Rita, Kurt A Kaczmarek, Mitchell E Tyler and Jorge Garcia-Lara. 1998. Form perception with a 49-point electrotactile stimulus array on the tongue: a technical note. *Journal of rehabilitation research and development*, 35, 4. 427.

[6] Paul Bach-y-Rita, Mitchell E Tyler and Kurt A Kaczmarek. 2003. Seeing with the brain. *International journal of human-computer interaction*, 15, 2. 285-295.

[7] David H Barlow and Steven C Hayes. 1979. Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of applied behavior analysis*, 12, 2. 199-210.

[8] Brian R. Belland. 2014. Scaffolding: Definition, Current Debates, and Future Directions. in Spector, J.M., Merrill, M.D., Elen, J. and Bishop, M.J. eds. *Handbook of Research on Educational Communications and Technology*, Springer New York, New York, NY, 505-518.

[9] Rafaello Bergonse. 2017. Fifty Years on, What Exactly is a Videogame? An Essentialistic Definitional Approach. *The Computer Games Journal*, 6, 4. 239-255. DOI: http://dx.doi.org/10.1007/s40869-017-0045-4

[10] Meera M Blattner, Denise A Sumikawa and Robert M Greenberg. 1989. Earcons and icons: Their structure and common design principles. *Human–Computer Interaction*, 4, 1. 11-44.

[11] Clive Roland Boddy. 2016. Sample size for qualitative research. *Qualitative Market Research: An International Journal*, 19, 4. 426-432.

[12] Saniye Tugba Bulu. 2012. Place presence, social presence, co-presence, and satisfaction in virtual worlds. *Computers & Education*, 58, 1. 154-161.

[13] Kelly Caine. 2016. Local standards for sample size at CHI. in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ACM, 981-992.

[14] Weiya Chen, Anthony Plancoulaine, Nicolas Férey, Damien Touraine, Julien Nelson and Patrick Bourdot. 2013. 6DoF navigation in virtual worlds: comparison of joystick-based and head-controlled paradigms. in *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, ACM, 111-114.

[15] Erin C Connors, Elizabeth R Chrastil, Jaime Sánchez and Lotfi B Merabet. 2014. Action video game play and transfer of navigation and spatial cognition skills in adolescents who are blind. *Frontiers in human neuroscience*, 8, Article 133. DOI: https://doi.org/10.3389/fnhum.2014.00133.

[16] Mira Crouch and Heather McKenzie. 2006. The logic of small samples in interview-based qualitative research. *Social science information,* 45, 4. 483-499.
[17] Giulia V Elli, Stefania Benetti and Olivier Collignon. 2014. Is there a future for sensory substitution outside academic laboratories? *Multisensory research,* 27, 5-6. 271-291.
[18] ESA. 2017. Essential Facts About the Computer and Video Game Industry, *Entertainment Software Association. Industry report.*
[19] Johnny Friberg and Dan Gärdenfors. 2004. Audio games: new perspectives on game audio. in *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, ACM, 148-154.
[20] Neveen I Ghali, Omar Soluiman, Nashwa El-Bendary, Tamer M Nassef, Sara A Ahmed, Yomna M Elbarawy and Aboul Ella Hassanien. 2012. Virtual reality technology for blind and visual impaired people: reviews and recent advances. *Advances in Robotics and Virtual Reality,* 26. 363-385.
[21] Timo Götzelmann. 2016. LucentMaps: 3D printed audiovisual tactile maps for blind and visually impaired people. in *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 81-90.
[22] Greg Guest, Arwen Bunce and Laura Johnson. 2006. How many interviews are enough? An experiment with data saturation and variability. *Field methods,* 18, 1. 59-82.
[23] KA Kaczmarek. 2000. Effect of electrode geometry and intensity control method on comfort of electrotactile stimulation on the tongue. *Proc. of the ASME, Dynamic Systems and Contorol Division, 2000* 2. 1239-1243.
[24] Andrew J Kolarik, Silvia Cirstea, Shahina Pardhan and Brian CJ Moore. 2014. A summary of research investigating echolocation abilities of blind and sighted humans. *Hearing research,* 310. 60-68.
[25] Ron Kupers and Maurice Ptito. 2014. Compensatory plasticity and cross-modal reorganization following early visual deprivation. *Neuroscience & Biobehavioral Reviews,* 41. 36-52.
[26] Shachar Maidenbaum, Galit Buchs, Sami Abboud, Ori Lavi-Rotbain and Amir Amedi. 2016. Perception of graphical virtual environments by blind users via sensory substitution. *PloS one,* 11, 2. DOI: https://doi.org/10.1371/journal.pone.0147501.
[27] Shachar Maidenbaum, Shelly Levy-Tzedek, Daniel-Robert Chebat and Amir Amedi. 2013. Increasing accessibility to the blind of virtual environments, using a virtual mobility aid based on the" EyeCane": Feasibility study. *PloS one,* 8, 8. DOI: https://doi.org/10.1371/journal.pone.0072555.
[28] Janice M. Morse. 2000. Determining Sample Size. *Qualitative Health Research,* 10, 1. 3-5.
[29] Makoto Ohuchi, Yukio Iwaya and Yôiti Suzuki. 2006. Cognitive-map forming of the blind in virtual sound environment. In *Proceedings of the 12th International Conference on Auditory Display (ICAD2006)*, London, UK,
[30] David Pelegrin Garcia, Monika Rychtáriková, Christ Glorieux and Brian FG Katz. 2014. Interactive auralization of self-generated oral sounds in virtual acoustic environments for research in human echolocation. in *Proceedings of Forum Acusticum 2014,* 7, 12. 43-55.
[31] Lorenzo Picinali, Amandine Afonso, Michel Denis and Brian FG Katz. 2014. Exploration of architectural spaces by blind people using auditory virtual reality for the construction of spatial knowledge. *International Journal of Human-Computer Studies,* 72, 4. 393-407.
[32] Juan Antonio Martínez Rojas, Jesús Alpuente Hermosilla, Rocío Sánchez Montero and Pablo Luis López Espí. 2009. Physical analysis of several organic signals for human echolocation: oral vacuum pulses. *Acta acustica united with Acustica,* 95, 2. 325-330.
[33] Maria V Sanchez-Vives and Mel Slater. 2005. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience,* 6, 4. 332.
[34] Jaime Sánchez, Ticianne Darin and Rossana Andrade. 2015. Multimodal Videogames for the Cognition of People Who Are Blind: Trends and Issues. in *International Conference on Universal Access in Human-Computer Interaction*, Springer, 535-546.
[35] Jaime Sánchez, Mauricio Sáenz, Alvaro Pascual-Leone and Lotfi Merabet. 2010. Enhancing navigation skills through audio gaming. in *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, ACM, 3991-3996.
[36] Lei Shi, Yuhang Zhao and Shiri Azenkot. 2017. Designing Interactions for 3D Printed Models with Blind People. in *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 200-209.
[37] Brandon Taylor, Anind Dey, Dan Siewiorek and Asim Smailagic. 2016. Customizable 3D Printed Tactile Maps as Interactive Overlays in *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 71-79.
[38] Lore Thaler. 2013. Echolocation may have real-life advantages for blind people: an analysis of survey data. *Frontiers in physiology,* 4, 98. DOI: https://doi.org/10.3389/fphys.2013.00098.
[39] Lore Thaler and Melvyn A Goodale. 2016. Echolocation in humans: an overview. *Wiley Interdisciplinary Reviews: Cognitive Science,* 7, 6. 382-393.
[40] Lore Thaler, Galen M Reich, Xinyu Zhang, Dinghe Wang, Graeme E Smith, Zeng Tao, Raja Syamsul Azmir Bin Raja Abdullah, Mikhail Cherniakov, Christopher J Baker and Daniel Kish. 2017. Mouth-clicks used by blind expert human echolocators–signal description and model based signal synthesis. *PLOS Computational Biology,* 13, 8. DOI: https://doi.org/10.1371/journal.pcbi.1005670.
[41] Javier Torrente, Manuel Freire, Pablo Moreno-Ger and Baltasar Fernández-Manjón. 2015. Evaluation of semi-automatically generated accessible interfaces for educational games. *Computers & Education,* 83. 103-117.
[42] Martin Usoh, Kevin Arthur, Mary C Whitton, Rui Bastos, Anthony Steed, Mel Slater and Frederick P Brooks Jr. 1999. Walking, walking-in-place, flying, in virtual environments. in *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, ACM Press/Addison-Wesley Publishing Co., 359-364.
[43] Gareth R White, Geraldine Fitzpatrick and Graham McAllister. 2008. Toward accessible 3D virtual environments for the blind and visually impaired. in *Proceedings of the 3rd international conference on Digital Interactive Media in Entertainment and Arts*, ACM, 134-141.
[44] Wenyu Wu, Rudina Morina, Arley Schenker, Alexander Gotsis, Harsha Chivukula, Madeline Gardner, Felix Liu, Spencer Barton, Steven Woyach and Bruno Sinopoli. 2017. EchoExplorer: A Game App for Understanding Echolocation and Learning to Navigate Using Echo Cues. in *The 23rd International Conference on Auditory Display (ICAD 2017)*, Georgia Institute of Technology.
[45] Bajram Zeqiri, Werner Scholl and Stephen P Robinson. 2010. Measurement and testing of the acoustic properties of materials: a review. *Metrologia,* 47, 2. S156.