Pointing Tasks using Spatial Audio on Smartphones for People with Vision Impairments

Abhijeet Singh Raina
J.Mike Walker '66
Department of Mechanical Engineering
Texas A&M University
College Station, Texas 77840.
Email: abhijeetsinghraina@tamu.edu

Ronak R. Mohanty
Meta Reality Labs
Redmond, Washington, 98052.
Email: ronakmohanty@aggienetwork.com

Abhirath Bhuvanesh
J.Mike Walker '66
Department of Mechanical Engineering
Texas A&M University
College Station, Texas 77840.
Email: abhuv@tamu.edu

Divya Prabha J.
University of Arts London
London, U.K.
Email: jdivyaprabha12@gmail.com

Manohar Swaminathan
Microsoft Research
Bangalore, India
Email: manohar.swaminathan@microsoft.com

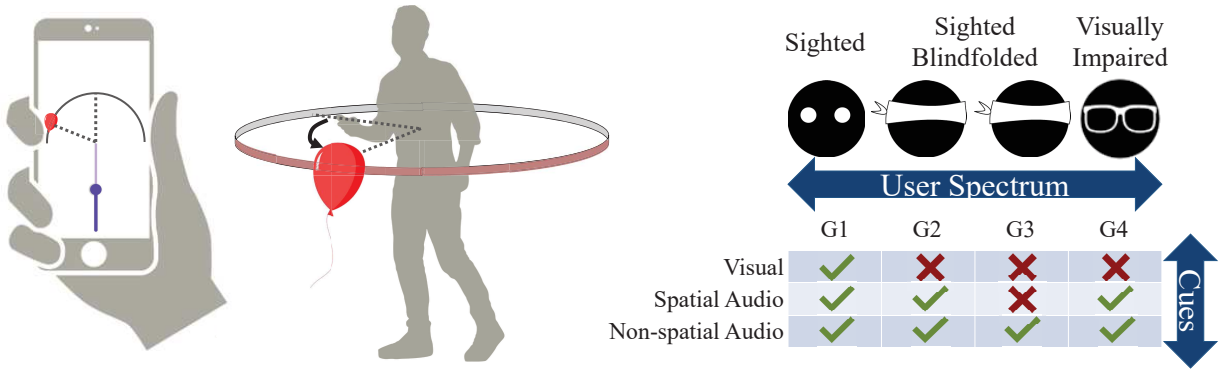
Vinayak R. Krishnamurthy*
J.Mike Walker '66
Department of Mechanical Engineering
Department of Computer Science
& Engineering (by Affiliation)
Texas A&M University
College Station, Texas 77840.
Email: vinayak@tamu.edu

We present an experimental investigation of spatial audio feedback using smartphones to support direction localization in pointing tasks for people with visual impairments (PVIs). We do this using a mobile game based on a bow-and-arrow metaphor. Our game provides a combination of spatial and non-spatial (sound beacon) audio to help the user locate the direction of the target. Our experiments with sighted, sighted-blindfolded, and visually impaired users shows that (a) the efficacy of spatial audio is relatively higher for PVIs than for blindfolded sighted users during the initial reaction time for direction localization, (b) the general behavior between PVIs and blind-folded individuals is statistically similar, and (c) the lack of spatial audio significantly reduces the localization performance even in sighted blind-folded users. Based on our findings, we discuss the system and interaction design implications for making future mobile-based spatial interactions accessible to PVIs.

1 INTRODUCTION & BACKGROUND

The motivation for our research is to enable people with vision impairments (PVIs) to experience and use augmented reality (AR) applications without requir-

*Corresponding author



(a) Interactive context: Mobile shooting game

(b) Experiment Strategy: Sequential cue deprivation

Fig. 1. Our interactive context is defined by (a) a simple shooting game that captures the core interaction task of orientation in a typical AR application on a smartphone (b) sequential cue reduction to understand the effect of removing visual and spatial auditory cues in a sequential manner to comparatively study their effects with respect to PVIs where the visual cue is absent. We define four user groups: G1, G2, G3, and G4 based on the cues provided.

ing expensive investment in hardware. Specifically, our quest is to enable PVIs to perform pointing tasks in 3D in smartphone-based AR applications. Today, many AR applications ranging from games and entertainment to education and training run on commodity smartphones, optionally with very low cost headsets. For instance, *Pokemon Go* [1], one of the fastest mobile AR application to reach US$ 100 million in sales, runs on smartphones with no required add-ons. Further, the capabilities of smartphones are rapidly increasing along multiple dimensions: increased processing power, including built in GPUs, multiple high resolution cameras, a host of sensors, including GPS, IMU, Lidar and others, and 5G connectivity that brings the power of cloud AI/ML within the reach of most smartphone users. Thus powerful AR applications that harness these capabilities are beginning to be available in diverse domains [2–5]. Riding on the same capabilities of smartphones, a newer generation of virtual reality applications, called Lightweight VR [6] that provide semi-immersive VR experiences on smartphones are beginning to appear.

However, most of the efforts previously made to make VR accessible for PVIs proposed systems using expensive, bulky, or custom-built hardware including hapic devices and gloves, head mounted devices, custom earpieces etc. which are highly application oriented [7–10]. Use of such complex hardware challenges the scalability of these proposed techniques for accessible AR/VR. The increasing versatility of modern day smartphones can be utilized to tackle the hardware challenges to some extent [11].

Despite these advances, very little work has been done in creating tools that enable the above experiences to be available to PVIs (see Section 1.2). In this paper, we present the very early steps that we have taken to address this gap. Given the paucity of such work, our goal is to investigate hand-held AR or VR, where the smartphone is held by the user in their hand and moved around to experience the application. Thus we are not considering head-worn AR or VR, which will present more complexities by requiring an additional device to do the pointing or selecting of direction.

We limit our focus to the operation that is fundamental to all such handheld AR/VR applications, that of pointing at an object in 3D space or orienting towards a direction in 3D space, usually by moving the smartphone around. This task is usually accomplished by the sighted users by locating the object of interest visually as they pan the phone around to search for the object. The basic question we seek to address is, how does a person without sight accomplish the same task? More specifically, how can spatial audio be used to accomplish this task in the AR/VR environment?

While there is an extensive volume of work [12–15] on how PVIs perceive through other senses such as sound or touch, how these senses can be meaningfully integrated into a usable interactive AR experience on a smartphone requires deeper inquiry of sensory effects within concrete interactive contexts.

1.1 Problem & Research Questions

In this paper, the question we seek to address is: Can users restricted to using just the auditory sense achieve task performance similar to the one achieved by users with vision? We narrow our focus to the specific task of

orientation i.e., the ability to localize the direction of an object from the egocentric frame of reference of an individual [16].

Auditory feedback for orientation has been extensively studied in perceptual psychology [17–20]. Zahorik et al. [19] report that binaural cues (cues independently transmitted to each ear separately) play a critical role in direction localization. On the other hand, distance perception is affected by a wide range of cues including intensity and spectrum. At the very least, it is understood that combining audio cues effectively for an operational AR environment requires a systematic and controlled series of studies. In this paper, we begin by asking the following questions, in the context of AR with smartphones:

1. What is the difference, if any, between the performance of blindfolded sighted users and PVIs for orientation tasks?
2. What are the similarities and dissimilarities between the actions and motor strategies of blindfolded sighted users and PVIs for execution of the task?
3. How much does spatial audio specifically affect the **accuracy of orientation** (pointing in a desired direction) in the absence of visual feedback (either through blindfolding or because of visual impairment)
4. How much does spatial audio specifically affect the **ability to sustain a given orientation** in the absence of visual feedback (either through blindfolding or because of visual impairment)
5. How much does spatial audio specifically affect **motor strategies** in enabling orientation in the absence of visual feedback (either through blindfolding or because of visual impairment)?

Here, by performance we mean the accuracy and the time taken for locating an object placed around the individual's body. By actions and motor strategies we believe that a careful investigation of these questions will offer critical insights needed for integrating spatial audio feedback in a practicable manner for PVIs.

1.2 Prior Work

The exploration of multi-modal interfaces in virtual reality environments has a rich history starting way back in the sixties with Heilig's *sensorama* [21]. There are very many works that make fundamental contributions to the perception of spatial audio and its practical use in real and virtual environments. There are numerous past and ongoing research in the use of spatial audio in AR and VR environments (for example, see the special issue on spatial audio in VR, AR and MR [22] and articles therein, including the relatively newer area of 360 degree video [23]. Similarly there are many that examine the use of touch in VR systems [24, 25]. The benefits of combining spatial audio and haptics to enhance presence in such environments has been well recognized and there are many works exploring this combination of sensory inputs [26], but most of them are about enhancing the overall perception of presence, in conjunction with the visual display.

There are few works that discuss VR accessibility for PVIs. Maidenbaum et al. proposed a virtual *EyeCane* for navigation tasks in VR [27]. *PowerUp* laid out some guidelines for making web-based games accessible to PVIs [28]. *NavStick* utilizes a gaming controller to navigate their surroundings with speech as the primary sensory cue [29]. *SeeingVR* provide a set of tools such as text-to-speech, and magnification lens to aid people with low vision to see in VR [30]. The study by Dong et al. showed that in a VR environment using 3D auditory feedback, the experience of PVIs is different when compared with that of sighted people [31]. Some games developed in the past such as *VI-Tennis* and *AudiOdyssey* use a variety of hardware devices to provide a gaming experience for sighted and PVI gamers [32, 33]. Drossos et al. developed a computer based tic-tac-toe game with binaural sound effects for blind children [34]. A few games such as *VBGhost* and *TapBeats* utilize smartphones to play games accessible to PVIs [35, 36]. There are a variety of works which target learning for PVIs using audio-gamification approach [37–39]. This audio-gamification can also be utilized in various applications in AR/VR environments by making use of the capabilities of widely used smartphones.

In parallel, there have been explorations of utilizing spatial audio interfaces for providing directional and distance cues. One of the early works that inspires our approach is by Sanchez et al. [40], who demonstrated a game environment, *AudioDoom* to enable spatial learning for blind children. Frauenberger and Noisternig [41] later proposed a formal software implementation framework dubbed VAR (virtual audio reality) for smooth integration of auditory cues within VR systems. Kolsover et al. [42] integrated the primary manipulative senses, namely, visual, auditory, and haptic senses for providing directional cues in mobile navigation. A more recent work by Brill et al explores a combination of vibrotactile and spatial audio directional cues for pararescue jumpers in the U.S. Air Force [43]. There are a bunch of commercially available *Audio Games* which use a variety of auditory cues such as 3D sound [44–47], stereo sound [48–51] and, verbal cues [52–54]. *Audio Game Hub* is an iOS app available on App Store which also contains a bunch of audio games accessible to PVIs [55–57]. The use of spatial audio for

enabling gamers without sight to play mainstream video games has been proposed in [58], but the work primarily deals with desktop or console gaming scenarios and has not been studied on smartphones.

There are many works that use audio, haptics or a combination of these to help PVIs to navigate in the real world. [42, 59–63]. Microsoft Soundscape [64] is an iOS app that uses the Geowiki Open Street Map to enable PVIs to navigate the real world using spatial audio to speak out points of interest. The use of spatial audio helps in the orientation and localization of points of interest. In addition, in a specific navigation mode, there is an audio beacon (a virtual drumbeat that is located at the coordinate of the destination) played out in spatialized audio that enables the user to determine the relative orientation to the destination. Ross et al. [65] address the broad questions around audio based explorations in virtual environments by extending Microsoft Soundscape. One of the key results from their study is that participants found estimation of the distance to be very challenging and required additional cues of known physical locations or physically being present in known locations. The work by Zhao et al., [66] combines audio with haptics to enable PVIs to navigate in a VR environment to give a better sense of presence by providing the equivalent of a white cane in VR, facilitated by a physical device that is worn by the user.

The visual experiences of a sighted individual play an important role in their development of spatial knowledge [67]. There are many works comparing sighted, sighted-blindfolded, and PVIs across a variety of studies across different age groups. Campus et al. compared sighted-blindfolded and early-blind subjects across spatial bisection, and temporal bisection tasks [68]. Ribadi et al. compared static and dynamic balance among these three groups in adolocents [69]. It was observed that sighted children show consistent improvement in spatial navigation skills with age as compared to blind children [70, 71]. An earlier study by Klatzky et al. found a significant difference across these groups in tasks involving spatial inference but did not find a significant spatial deficit among PVIs [72]. Haptic material perception has also been studied across similar group of participants and no significant advantage has been recorded because of the visual experiences [73]. Similarly, Rovira et al. did not find any significant deficit among blind and sighted adolescents while performing mental rotation of 2D shapes [74]. Cattaneo et al. found differences in spatial bias between sighted individuals and PVIs, where PVIs exhibiting no significant spatial bias in vertical and radial dimensions [75]. Accessible games such as *BlindHero* also compared the performance across these groups [76]. Similar performance in spatial orientation and obstacle avoidance task using spatial audio was observed among sighted-blindfolded and PVI individuals in a study by Bujacz et al. [77].

1.3 Knowledge Gaps & Our work

We observe that a large portion of the body of work available on enabling both sighted as well as blind users deals with navigation tasks wherein the user is in motion. Works that seek to support visually impaired users, while seminal, are largely application-oriented. Our work seeks to complement these works by offering a deeper task-oriented analysis of how and why spatial audio can enable orientation for PVIs in mobile AR. Secondly, our work establishes a crucial connection between the similarities and differences between sighted and visually impaired individuals. This connection serves as an important step in the development of design guidelines for sighted HCI designers who wish to incorporate the experience of PVIs in AR/VR environments while using sound and touch as the primary perceptual cues [10, 31].

Finally, our work looks at the fine-grained process of direction localization in the absence of visual feedback. We specifically show that (a) the efficacy of spatial audio is relatively higher for PVIs than for blindfolded sighted users during the initial reaction time for direction localization, (b) the general behavior between PVIs and blindfolded individuals is statistically similar, and (c) the lack of spatial audio significantly reduces the localization performance even in sighted blind-folded users.

2 APPROACH & RATIONALE

There are three key aspects to our approach: (1) interactive context for a systematic study (Figure 1(a)), (2) perceptual cues for orientation, and (3) experimental strategy (Figure 1(b)).

2.1 Interactive Context

A simple application that captures this problem is the task of shooting a target at any orientation and distance with the user at the origin. Auditory localization depends on interaural loudness difference(IDL) and interaural time difference(ITD) and humans can take advantage of both during horizontal auditory localization tasks [78]. Vertical auditory localization is difficult which depends on the spectrum of sound cues as created by the outer ear, suggesting that vertical localization is difficult when compared with horizontal localization [79]. Hence, to simplify the study task, We restrict the target to be in the horizontal plane in our study. The basic task is for the

users to orient the phone to align with the position of a target in the horizontal plane by moving the hand holding the phone around or by turning their body in place while holding the hand steady or a combination of the above. This is the core operation in most AR or lightweight VR applications and our goal is to make this task accessible to PVIs. The one change we make to a typical AR application is that instead of holding the phone perpendicular to the ground (in landscape or portrait mode) the users hold the phone with its surface parallel to the ground and use their other hand to interact with the touch screen.

And this can be used in other settings like wayfinding [80, 81] or understanding the location of objects around the person in an augmented or mixed reality environment. Utilizing the metaphor of a *virtual crossbow* [50], we designed and implemented a simple Android game app that generates a series of targets at different orientations and distances from the user. The user's task is to shoot the balloons using a slider in the smartphone screen akin to pulling the string of a bow and releasing to shoot an arrow (Figure 1(a)). Effectively, the process of determining the shooting direction maps to the orientation task. Subsequently, the pulling of the slider to shoot the balloon maps to the user's ability to sustain a given orientation while focusing on a non-orientation task (i.e. hitting the balloon). Note that the estimation of distance of the balloon is not considered in our study.

2.2 Perceptual Cues

Given the context of our shooting game, we developed our game to include visual, auditory, and vibrotactile feedback. Specifically, orienting the phone closer to the direction of balloon brings it within the visual range of the phone's screen. Furthermore, as the angular difference between phone and the balloon's direction reaches within a certain threshold, we also display a change in color of the balloon. Given that our primary target audience is PVIs, the reason for adding visual feedback was to establish a reference with respect to which we could understand how the absence of visual feedback would affect the performance and the actions of a user during the orientation task.

Our second and critical perceptual feedback for the orientation task is auditory. Here, we utilize spatial audio, which is a powerful evolutionary human ability that plays a role in drawing the visual sense roughly in the direction of the source of sound and then the visual sense accurately locates the source. Spatial audio as a powerful tool for people with vision impairment has long been recognized and there are many efforts that seek to exploit this sense for orientation [82, 83]. On the other hand, Voss et al. show evidence for superior spatial hearing for blind individuals in the horizontal plane in addition to significant deficits in the vertical plane. [84]. Therefore, we specifically constrained the target balloons to be on the horizontal (azimuth) plane of the user which is further aligned with the display screen of the phone. In addition to spatial audio, we also added a non-spatial audio chirp feedback. These refer to beep tones with varying frequency of beeps to indicate angular deviation from the target. Such cues are quite common in both digital and physical environments (e.g. in accessible pedestrian crossings).

Finally, the association of the extent of the slider draw (pulling the string of the bow) is mapped to the vibrotactile feedback and its mapping to the distance of the target is an interesting area of future research. In this work, we "*cheat*" by providing a distinct "*lock*" sound when the target is in the correct range corresponding to the extent of the draw, prompting the user to release the virtual arrow. We then provide a verbal confirmation of the hit or the miss. The reason for this addition lock sound comes from extensive literature that essentially concludes enabling a person to accurately estimate the distance of an auditory source without visual feedback is currently prohibitively challenging if not completely impossible. Neilsen et al. [85] summarize the results of several auditory experiments in different room conditions and conclude, *"In the anechoic room there is no correspondence between physical and perceived distance"*.

2.3 Experimental Strategy

In order to systematically study the tasks of orientation and ranging, we implement a *sequential cue deprivation* strategy (Figure 1(b)). What we mean by that is we conducted a between-subjects experiment across four groups of users beginning with sighted users who played the game with all feedback mechanisms. This is our reference user group. Following this, we sequentially remove one cue at a time starting from vision, followed by spatial audio. This results in two groups, namely, sighted blind-folded users with spatial audio and beep cues and sighted blind-folded users with beep cues but without spatial audio. Our final group is comprised of the PVI users who are provided with both spatial audio as well as beep cues. These four groups allowed us to explore our research questions.

3 EXPERIMENTAL SETUP

We conducted controlled user studies where participants performed tasks involving spatial orientation of objects which required spatial cognition. We introduced

these tasks to the users in a sequential fashion starting with orientation task, followed by a task requiring both orientation and ranging. These tasks were presented in the form of a 2D shooting game where the goal was to shoot a virtual balloon placed at a distance. Each task corresponds to a specific setting of the game. We *gamified* the study primarily to get better engagement as the participants perform tasks with increased difficulty levels [37].

3.1 Game Design

For this experiment, we developed an Android game in which the goal is to shoot a virtual balloon which is emitting a beep sound in 3D space. Unlike the visual information which can be processed in parallel, non-visual information is described as sequential or serial processing [86]. Hence, we provide one target (balloon) at a time. In this game, a balloon appears at different orientations and distances from the user. The targets are restricted to be on a horizontal plane in front of the user. In order to shoot, the user needs to align the phone in the direction of the balloon by moving their hand/arm with which they are holding the phone in the horizontal plane in front of them. The game provides Auditory, Visual, and Vibrotactile feedback to the users in order to help the user shoot the virtual balloon. Each feedback mode is described below.

Auditory Feedback: In some experimental conditions, the beep sound is spatialized which means that the user is able to identify the beep sound direction using which the user may align the phone in direction of the beep sound. To help with the alignment, another auditory cue is provided in terms of the varying beep frequency of the balloon. The beep frequency is inversely proportional with the angular difference between the phone and the balloon in the horizontal plane. As the user aligns the phone towards the balloon, the beep frequency keeps on increasing and reaches a maximum when the phone is perfectly aligned with the balloon. The game also provides a distinct 'lock' sound whenever the amount of slider pulled correctly corresponds to the distance of the balloon from the user within a margin of ± 5 units, given that the phone is pointing directly towards the balloon within an angular margin of ± 5 degrees.

Vibrotactile Feedback: The user also needs to estimate the distance of the balloon using a vertical slider on the smartphone. The main purpose of this feedback is to add complexity to the tasks rather than providing a means to do distance estimation. The game provides a vibrotactile feedback to the user whenever they pull the slider. This slider represents a virtual bow which the user is pulling by sliding their finger vertically downwards on the screen. The amplitude of the vibration is proportional to the amount by which the slider is pulled. The amplitude keeps on increasing until the slider value correctly corresponds to the distance between the user and the balloon. Upon further pulling of the slider, amplitude will remain constant.

Visual Feedback: The game provides visual feedback in the form of the color of the balloon. The balloon color changes based on how the user is aligned with respect to the balloon and how much the slider is pulled back. Red color represents that the user is neither within the allowable angular threshold (± 5 degrees), nor the slider is pulled back by the correct amount. Yellow color represents that the alignment of the phone is correct i.e., the angular difference between the phone and the balloon is less than ± 5 degrees but the slider value is incorrect. Blue color represents that the slider is pulled back by the correct amount i.e., within ± 5 units of the correct value but the alignment error between the phone and the balloon is more than the threshold. Green color represents that the phone is aligned with the balloon direction and the slider value is also correct.

Once the users think that they are aligned with the balloon and the estimated distance is correct, they can shoot by lifting their finger from the screen. The game then gives a verbal confirmation of a hit or a miss. The game can be played in two modes – Practice mode and Game mode. In Practice mode, if the user misses the target, the target position will remain the same with respect to the user so that the user can try to shoot the same target again. In Practice mode, the target position will change only when the user successfully hits the target. In Game mode, the target position will change after every attempted shot irrespective of a hit or a miss i.e., the user only gets one chance to shoot the target.

4 EXPERIMENT DESIGN

Our experiment design was based on systematic approach to investigate how different sensory inputs help a person to locate an audio source in 3D space. We conducted controlled between-subjects experiments with four different group of participants. Each group performed the same experiment but under different physical conditions and sensory inputs. In the following sections we describe the participation pool, study tasks, experimental procedure, and evaluation metrics followed across

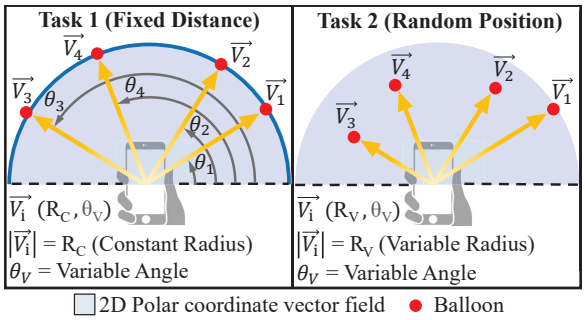


Fig. 2. Users perform two tasks in the study. In Task 1 (left), the balloon is at a fixed distance (R_C) from the user with varying direction (θ_V). In Task 2 (right), both distance and direction changes. $\vec{V_i}$ represents different trials.

all experiments.

4.1 Participants

We recruited a mix of 48 participants. Out of these, 36 were sighted participants and 12 participants had complete loss of vision. 3 sighted participants had prior experience with spatial audio through gaming consoles and prior user studies. We asked all the participants to wear headphones/earphones while playing the game in order to experience the spatial auditory cues. Other than the visual impairment, all the participants could comfortably use both of their upper limbs and had normal hearing capability.

We divided the 36 sighted participants into 3 user groups with 12 participants in each group. Group 4 comprised of PVI participants formed the fourth group. The groups were divided based on physical condition of the participants and the sensory inputs provided while performing the experiment. The details about the different groups are as follows.

Group 1 **(G1):** All the participants in this group played the game with their eyes open. Also, the beep sound of the balloon was spatialized i.e., the beep volume in each ear is determined based on the position of the target with respect to the sagittal plane of the user. If the target lies on the left side of the sagittal plane, the beep volume is higher in the left ear as compared to the right, and vice versa.

Group 2 **(G2):** All the participants in this group played the game with their eyes closed. The beep sound of the balloon was spatialized similar to that of G1.

Group 3 **(G3):** Similar to G2, all the participants in group-3 played the game with their eyes closed with the difference that the beep sound of the balloon was not spatialized. Hence, the beep volume in both the ears was the same.

Group 4 **(G4):** All the PVI participants were grouped together. Similar to G1 and G2, the beep sound of the balloon was spatialized. G2 closely matches this group in terms of the physical conditions and sensory inputs.

We would like to note here that our participant pool was recruited in US from groups 1—3 and group 4 was recruited from India. This was done due to our specific focus on accessibility of mobile interactions in underrepresented groups with low fidelity smartphones. Having said this, the PVIs recruited for our study were all well acquainted with their own specific device models with audio-enabled capabilities.

4.2 Tasks

In this experiment, each user performed two tasks with a common goal of shooting the balloon. The tasks were separated based on how the balloon position changes with respect to the user across different trials. Below, we define the two tasks.

Task 1 **– Fixed Distance:** In this task, the distance of the balloon with respect to the user is fixed and only the direction of the balloon changes between trials (Fig. 2 (left)). This means that the user will need to pull the slider the exact same amount in each trial to successfully hit the balloon. Between two consecutive trials, only the alignment of the balloon will change in the horizontal plane in front of the user.

Task 2 **– Random Position:** In this task, both the distance and direction of the balloon changes between two consecutive trials (Fig. 2 (right)). The amount of slider pull required between two consecutive trials is different.

4.3 Procedure

Each study took approximately 45 minutes for the sighted individuals and around 1.5 hours with the PVIs. The study with sighted participants was done in person using a Samsung Note 10 smartphone with the mobile app pre-loaded. However, we made special considerations for the study protocol with PVIs (group 4) due to severe restrictions imposed due to COVID-19. For group 4, the study was conducted via a video conference call wherein participants first downloaded, installed, and performed initial trials. We specifically took the following steps to ensure proper data collection:

We had two study administrators, one in the US and

one in India who were constantly guiding the participants through video conferencing, monitoring their actions and giving verbal feedback wherever necessary to closely imitate the study protocol followed during the in-person testing.
During the online studies, the data was streamed to an online server in real-time and was checked for any anomalies by the study administrator in the US.
The PVIs recruited for the research study were all well-acquainted with using a touch screen mobile. This helped the study administrators guide the participants via Zoom in an efficient manner.

Each session started with a general introduction of the game to familiarize the user with the interface and different interaction techniques used in the game. Finally, they were asked to fill a demographic questionnaire before beginning with the study tasks.

Each of the sighted users shot the target 5 times in the Practice Mode of the game at the beginning of each task (Fixed Distance and Random Position) followed by the 5 study shots for each task in the Game Mode. In total, each user performed 10 practice shots and 10 study shots. The PVIs required upto 8 trial shots to understand the interface and to learn its use. They did the same 5 shots in the study phase for each task as the sighted participants. The practice shots helped the users to acquire adequate practice of the tasks before beginning with the study shots. All the users performed the tasks in a sequential manner. The users started the game by first performing the Fixed Distance task followed by the Random Position task. Participants in G2 and G3 were asked to close their eyes at all times while playing the game. Each trial took $15-30$ seconds to complete across all the user groups. In total, 960 trials were recorded across all the four user groups.

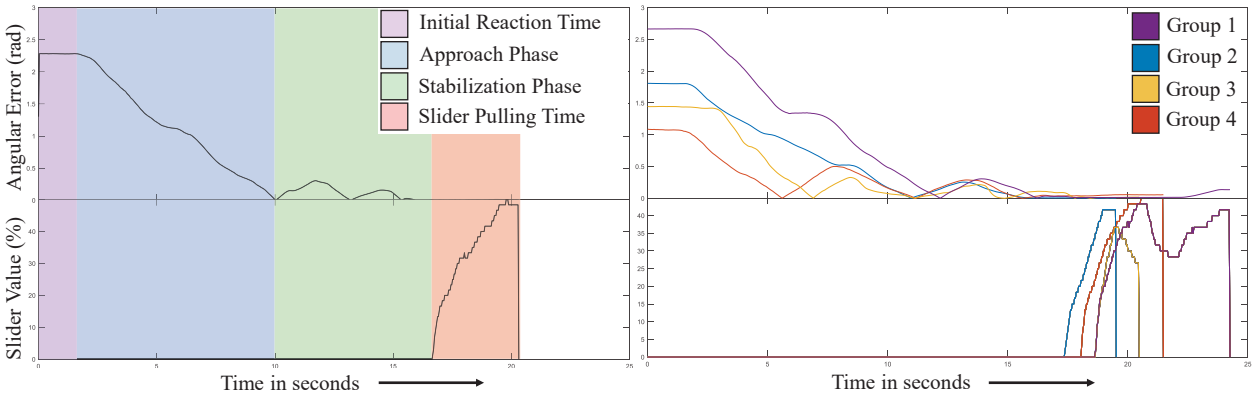
4.4 Data and Evaluation Metrics

For each trial performed by the participant, we recorded the raw event log containing (a) time taken for each trial, (b) phone's accelerometer data, (c) phone's gyroscope data, (d) balloon (target) position in 3D space, (e) no. of hits and misses, (f) position of the user in 3D space, (g) position of touch on the screen, (h) amount of slider pull, and (i) angular difference between the phone and the target. Since the studies for Group-4 were conducted online, the raw event log was live-streamed to an online server and was checked for any anomalies by the study administrator. In order to analyze the actions performed by the user to achieve the goal of shooting the target, we take a closer look at the IMU data (rotation of the phone) and touch inputs (slider pulling) by the user. We further elaborate on this action analysis in the following sub-section.

4.4.1 Action Analysis

The users performed two actions in order to shoot the target i.e., rotate the phone and pull the slider. To analyze these actions, we plotted angular error between the target and the user v/s time, along with the slider value v/s time (Fig. 3 (a)) for every trial and observed a common trend in the motor strategy of the users across all the groups (Fig. 3 (b)). Based on our observations, we segmented these actions into four phases – Initial reaction time, Approach phase, Stabilization phase, and Slider pulling time. Each of these phases are defined below.

Initial Reaction Phase: This phase starts when the trial starts i.e., the participant starts to hear the beeping sound. The phase ends when the user starts to move the phone either towards or away from the target, thus capturing the time taken by the user to react to the cues.
Approach Phase: During this phase, the user moves the phone to align with the target. This phase follows immediately after the Initial reaction time ends and the user starts to show some movement of the phone. This phase ends when the angular error between the phone and the target first reaches zero or it reaches the minimum value recorded during that trial. This phase can be interpreted as the initial gross movement made by the user towards the target.
Stabilization Phase: This phase represents the time spent by the user to do fine adjustments to the orientation of the phone in order to properly align with the target. In this phase we generally see a wavy pattern in the angular error v/s time plot and with every wave the angular error decreases, showing that the users generally took an iterative approach to do fine adjustments to the alignment. The phase starts immediately after the end of approach phase and ends as soon as the phone movement stops or becomes relatively small.
Slider Pulling Phase: In this phase the user pulls the slider back to estimate the distance of the target. The phase starts when the user starts to pull the slider and ends when they stop pulling and release the slider by lifting their finger off the screen of the phone. This phase is independent of the other three phases, and it sometimes overlaps with the stabilization phase.



(a) Action segmentation for sample trial

(b) Representative trials for four user groups

Fig. 3. (a) We segment the actions of the users into four phases: Initial reaction time, Approach phase, Stabilization phase, and Slider pulling time. (b) A representative sample trial for each group, showing similarities in the actions and motor strategies across all groups.

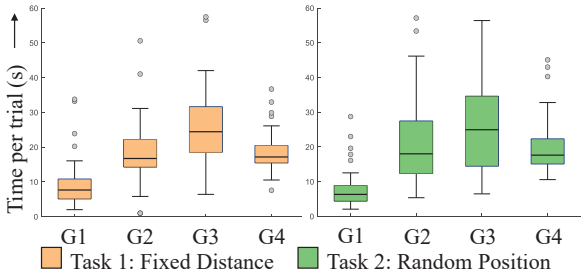


Fig. 4. Time per trial for the four groups across the two tasks

5 RESULTS

In the following sections, we report the statistical analysis of the user performance metrics. Furthermore we discuss the key findings and insights gained from our data collection, observation, and user-feedback from all trials performed by the participant. We present the analysis of the total time taken per trial for each user group across the two tasks. Subsequently, we segment each trial into different phases based on the general strategy used by the participants to shoot the target.

5.1 Time Taken Per Trial

In this sub-section, we compare the total time taken by each participant per trial. We do this comparison across the two tasks for all user groups (See Fig. 4). We first tested the data for normality using Shapiro-Wilk test and found out that the data is not from a normal distribution. We further conducted hypothesis testing using Kruskal-Wallis test ($\alpha = 0.05$) which is the non-parametric statistical equivalent of one-way ANOVA test.

5.1.1 Comparison across tasks

We performed pair-wise comparison of the time taken by the participants in each group across the tasks. For this comparison, we considered group 1 as the gold standard because it had access to all the sensory cues.

For Task 1: Significant differences were observed between G2 ($p < 0.001$), G3 ($p < 0.001$), and G4 ($p < 0.001$) when compared to G1. This is along expected lines since G1 has the visual input while the others do not. Significant differences were observed between the mean time taken to complete each trial between G2 and G3 ($p = 0.01$), and G3 and G4 ($p = 0.03$). This brings out the contribution of spatial audio to the task at hand since G3 is deprived of spatial audio cues compared to G2 and hence takes longer to complete the task. What is significant is that sighted blindfolded users without the spatial audio cues took a longer time (M = 22.32 sec) compared to PVIs clearly establishing the value of spatial audio cues for orientation. It is important to note that no significant differences were found between G2 and G4 indicating that blindfolded sighted users and PVIs had similar task completion times for this task, though the higher standard deviation for G2 could possibly indicate that PVIs were possibly better at utilizing spatial audio cues for orientation than the blindfolded participants.

For Task 2: Significant differences were observed between G2 ($p < 0.001$), G3 ($p < 0.001$), and G4 ($p < 0.001$) when compared to G1. This is expected since G1 had all the sensory cues including vision. No significant differences were observed between the completion times for task 3 between G2 and G3 ($p = 0.12$), G2 and G4 ($p = 0.23$), and G3 and G4 ($p = 0.51$). It appears that the effect of distance estimation overpowers that of direc-

tion localization (task 1). Given that distance estimation is known to be difficult through auditory feedback, the lack of differences is expected. Having said that, we do observe a higher mean, median and inter-quartile range for G3 ($\mu = 26.38$ sec, M = 20.99 sec) as compared to G2 ($\mu = 24.28$ sec, M = 17.28 sec) and G4 ($\mu = 21.38$ sec, M = 17.93 sec). This suggests that even with the difficulty of distance estimation, the lack of spatial audio for G3 is what likely resulted in wider spread of the distribution.

5.2 Analysis of Action Phases

We compare the time taken in each of the four phases by each participant (See Fig. 5). We do this comparison across the two tasks for each user group. Similar to the previous analysis, we first tested the data for normality using Shapiro-Wilk test and found out that the data is not from a normal distribution. We further conducted hypothesis testing using Kruskal-Wallis non-parametric test ($\alpha = 0.05$). We performed pair-wise comparison of the time taken in each phase by the participants in each group across the two tasks. Similar to the previous analysis, we considered group 1 as the gold standard.

For Task 1: It is important to note that no significant differences were found between G2 and G4 for mean times in initial reaction time ($p = 0.1$), approach phase ($p = 0.23$), stabilization phase ($p = 0.94$), and slider pulling phase ($p = 0.22$) indicating that the perceptual experiences of the sighted blindfolded and the PVIs were roughly similar in the orientation task. Significant differences were observed in the mean initial reaction time for each trial between G2 and G3 ($p < 0.001$), and G3 and G4 ($p < 0.001$). The shorter reaction times for G3 compared to G2 and G4 could be explained by the fact that without the spatial audio cues G3 participants had to move to even find out the relative position of the target to their current orientation, whereas the other two groups had to spend some time extracting this information from the spatial audio cue before starting their movement. Mean time for stabilization phase was significantly more for G3 and G4 ($p = 0.04$) with much higher standard deviation indicating the utility of spatial audio cues.

For Task 2: The relative differences between various groups and their statistical significance were roughly similar to that of Task 1. In general, all the time durations became lesser than for Task 1 indicating the learning effect on the participants since Task 2 was the last Task to be performed.

6 DESIGN IMPLICATIONS

6.1 Hardware Challenges

A key methodological issue we faced in our study was the difference between the protocols for the sighted users and PVIs. While the study with sighted individuals was conducted with a single smartphone in person, the study with PVIs was conducted remotely with each PVI user using their personal smartphone. The lack of consistency in the smartphone models resulted in two critical hardware-specific issues because of which two of the enrolled participants were unable to get a single hit in 10 trials. As a result, we terminated the study in order to prevent discomfort and feeling of inadequacy by conveying to them that there must be hardware issues.

Upon looking closely at the smartphone (Inertial Measurement Unit) IMU data, we found that these users primarily faced issues due to abnormally high noise resulting in complete loss of control over the game. Further analysis showed that the vibrotactile feedback generation was very compute intensive. Given that the study was being conducted while on call with the researcher, the Talkback and vibrotactile feedback being a load on the CPU, we speculate that there is a likelihood of latencies that depend on the phone in the audio lock feedback. Finally, we also faced difficulties with different aspects of the smartphone such as battery drainage, audio-lock cue being delayed causing them to miss, and difficulty with Talkback gestures being recognized. An earlier work on Talkback with PVI users also indicated the issues with touch screen sensitivity and gesture recognition being non-uniform across phones [87]. Thus hardware and OS versions should be carefully considered by researchers working with audio and tactile interfaces on smartphones for PVIs.

6.2 Future Challenges with Distance Estimation

Even though the study mainly focused on orientation, we also collected some preliminary data of the ranging aspect of the task which was facilitated by the vibrotactile feedback and the lock sound. We recorded the number of times the participants were able to successfully hit the balloon. Since a successful hit requires both orientation and distance estimate to be correct, the number of hits gave us a few insights on the ranging aspect of the task. We observed lower number of hits for task 2 which combined both orientation and ranging as compared to task 1 across all four user groups. This may be due to the additional ranging task that users had to do for task 2. We observed that the vibrotactile feedback for distance estimation was not effective and it is not a trivial problem. In retrospect this is expected because touch is a near

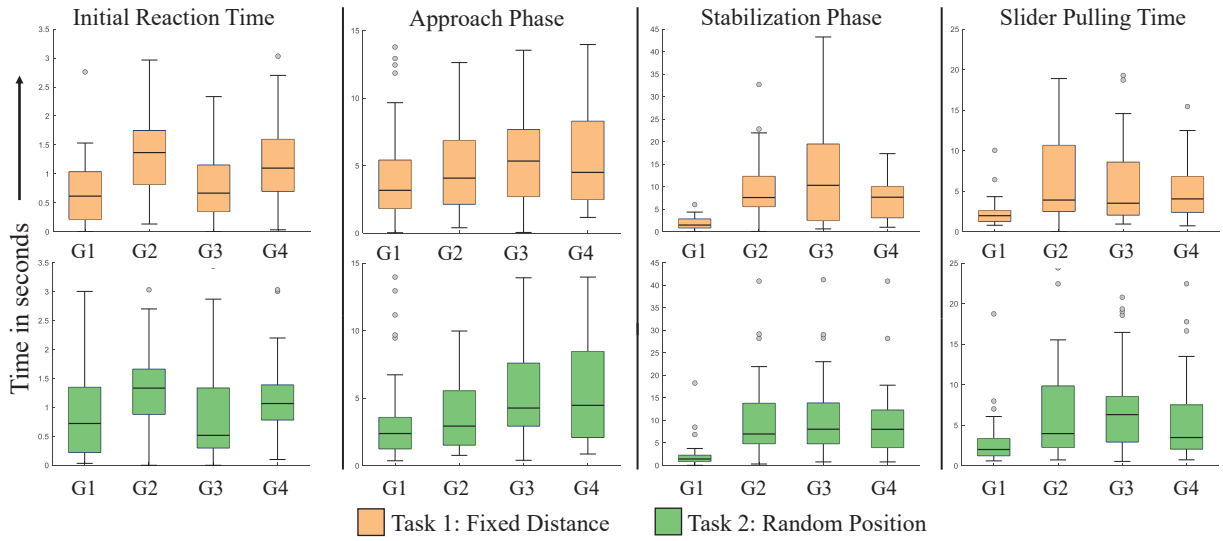


Fig. 5. Time taken for each phase for the three tasks across the four groups

sense and has no bearing on far distance estimation from an evolution standpoint. Hence what we are attempting to do is sensory substitution [88] where with repeated usage the amplitude of the vibration could be mapped to the distance of the target. Our blindfolded subjects and PVIs relied on the lock sound for determining the range rather than the amplitude of the vibrations. Our sighted non-blindfolded participants used the color cue to determine when the slider has been pulled to the right amount. We also observed a monotonous decrease in the number of hits when going from G1 to G4. Participants of G2 performed better than that of G3 in terms of no. of hits which further highlights the effectiveness of spatial audio in orientation tasks. PVIs had much worse hit percentage than the other groups, even though based on the analysis of their phase-wise performance, it is clear that they were able to orient the phone with similar processes as the other groups. We believe that the hardware challenges as discussed in the previous section were the main reason for lower hits by PVIs. We believe that there is yet a richer set of questions regarding tactile cues for PVIs that warrants an isolated study to determine the effectiveness of the tactile feedback for range estimation.

6.3 Potential interaction guidelines for PVI

Another avenue for future research is the study of the ergonomics and affordances of a smartphone in such accessible AR applications. For instance, some of our participants held the phone in one hand and used the other hand to pull the slider while there were a few who used one hand to do both. Some participants used the device while standing while some others performed the trials while sitting down. Since the studies with PVIs were done remotely on a single device, we did not have the opportunity to study these aspects in detail. We asked a few questions about their method of use, but a controlled study of cue perceptions in different configurations is important to arrive at design and interaction guidelines for diverse end applications.

7 CONCLUSION

In this paper, we sought to develop a fundamental understanding of two specific spatial tasks, namely, orientation and ranging for persons with visual impairment. The motivation to do so stemmed from the fact that affordable mobile AR needs a wide variety of guidelines for system, interaction, and feedback mechanism design. While we experimented with both spatial audio and vibrotactile feedback, we observed that spatial audio cues provide sufficient help in orientation tasks. The next stage of the work is to extend the interaction to 360 degrees around the user instead of just on the azimuthal plane while continuing our effort to provide the means for precise distance estimation. Systematic research is required to explore new sensory cues which will make the distance estimation easier. There is a need to explore and study the role of vibrotactile feedback in tasks which require manipulating objects close to the body. Having said that, this research is still only a glimpse of the rich research that is yet to be done in the domain of accessibility in AR systems.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant #2008800. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Pokemon go: Elevating the conversation around accessible gaming https://www.afb.org/aw/17/9/15295 Accessed: 2021-07-15.

[2] Bacca Acosta, J. L., Baldiris Navarro, S. M., Fabregat Gesa, R., Graf, S., et al., 2014, "Augmented reality trends in education: a systematic review of research and applications," *Journal of Educational Technology and Society, 2014, vol. 17, núm. 4, p. 133-149*.

[3] Zhuang, Y., Qu, T., Wong, J., Wan, W., Claudia tom Dieck, M., and Jung, T., 2020, "Using 5g mobile to enable the growing slate of vr and ar applications," In *Augmented Reality and Virtual Reality*. Springer, pp. 185–194.

[4] Li, X., Yi, W., Chi, H.-L., Wang, X., and Chan, A. P., 2018, "A critical review of virtual and augmented reality (vr/ar) applications in construction safety," *Automation in Construction,* **86**, pp. 150–162.

[5] Sahu, C. K., Young, C., and Rai, R., 2021, "Artificial intelligence (ai) in augmented reality (ar)-assisted manufacturing applications: a review," *International Journal of Production Research,* **59**(16), pp. 4903–4959.

[6] Virtual reality without the geeky headsets: Lightweight vr for smartphones https://venturebeat.com/2015/03/25/virtual-reality-without-the-geeky-headsets-lightweight-vr-for-smartphones/ Accessed: 2021-07-15.

[7] Kreimeier, J., Karg, P., and Götzelmann, T., 2020, "Tabletop virtual haptics: Feasibility study for the exploration of 2.5 d virtual objects by blind and visually impaired with consumer data gloves," In Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments, pp. 1–10.

[8] Bujacz, M., Kropidlowski, K., Ivanica, G., Moldoveanu, A., Saitis, C., Csapo, A., Wersenyi, G., Spagnol, S., Johannesson, O. I., Unnthorsson, R., et al., 2016, "Sound of vision-spatial audio output and sonification approaches," In International Conference on Computers Helping People with Special Needs, Springer, pp. 202–209.

[9] Rodriguez, A., Boada, I., and Sbert, M., 2018, "An arduino-based device for visually impaired people to play videogames," *Multimedia Tools and Applications,* **77**(15), pp. 19591–19613.

[10] Kreimeier, J., and Götzelmann, T., 2020, "Two decades of touchable and walkable virtual reality for blind and visually impaired people: A high-level taxonomy," *Multimodal Technologies and Interaction,* **4**(4), p. 79.

[11] Blum, J. R., Bouchard, M., and Cooperstock, J. R., 2011, "What's around me? spatialized audio augmented reality for blind users with a smartphone," In International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services, Springer, pp. 49–62.

[12] Cappagli, G., Finocchietti, S., Cocchi, E., Giammari, G., Zumiani, R., Cuppone, A. V., Baud-Bovy, G., and Gori, M., 2019, "Audio motor training improves mobility and spatial cognition in visually impaired children," *Scientific reports,* **9**(1), pp. 1–9.

[13] Brock, A., 2013, "Interactive maps for visually impaired people: design, usability and spatial cognition," PhD thesis, Université Toulouse 3 Paul Sabatier.

[14] Heller, M. A., and Ballesteros, S., 2006, "Picture perception and spatial cognition in visually impaired people," In *Touch and Blindness*. Psychology Press, pp. 57–80.

[15] Aprile, G., Cappagli, G., Morelli, F., Gori, M., and Signorini, S., 2020, "Standardized and experimental tools to assess spatial cognition in visually impaired children: a mini-review," *Frontiers in Neuroscience,* **14**, p. 988.

[16] Klatzky, R. L., 1998, "Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections," In Spatial cognition, Springer, pp. 1–17.

[17] Loomis, J. M., Klatzky, R. L., Philbeck, J. W., and Golledge, R. G., 1998, "Assessing auditory distance perception using perceptually directed action," *Perception & Psychophysics,* **60**(6), pp. 966–980.

[18] Klatzky, R. L., Lippa, Y., Loomis, J. M., and Golledge, R. G., 2002, "Learning directions of objects specified by vision, spatial audition, or auditory spatial language," *Learning & Memory,* **9**(6), pp. 364–367.

[19] Zahorik, P., Brungart, D. S., and Bronkhorst, A. W., 2005, "Auditory distance perception in humans: A summary of past and present research," *ACTA Acus-*

tica united with Acustica, **91**(3), pp. 409–420.

[20] Katz, B. F., and Picinali, L., 2011, "Spatial audio applied to research with the blind," *Advances in sound localization*, pp. 225–250.

[21] Heilig, M., 2002, "The cinema of the future," *Translated by Uri Feldman. In Multimedia: From Wagner to Virtual Reality. Edited by Randall Packer and Ken Jordan. Expanded ed. New York: WW Norton*, pp. 239–251.

[22] Cohen, M., Villegas, J., and Barfield, W., 2015, "Special issue on spatial sound in virtual, augmented, and mixed-reality environments," *Virtual Reality,* **19**(3), Nov, pp. 147–148.

[23] Bala, P., Masu, R., Nisi, V., and Nunes, N., 2019, *"When the Elephant Trumps": A Comparative Study on Spatial Audio for Orientation in 360º Videos* Association for Computing Machinery, New York, NY, USA, p. 1–13.

[24] Burdea, G. C., 1999, "Haptic feedback for virtual reality," In Virtual reality and prototyping workshop, Vol. 2, Citeseer, pp. 17–29.

[25] Magnenat-Thalmann, N., and Bonanni, U., 2006, "Haptics in virtual reality and multimedia," *IEEE MultiMedia,* **13**(3), pp. 6–11.

[26] Burke, J., Prewett, M. S., Gray, A. A., Yang, L., Stilson, F. R. B., Coovert, M., Elliott, L., and Redden, E., 2006, "Comparing the effects of visual-auditory and visual-tactile feedback on user performance: a meta-analysis," In ICMI '06.

[27] Maidenbaum, S., Levy-Tzedek, S., Chebat, D.-R., and Amedi, A., 2013, "Increasing accessibility to the blind of virtual environments, using a virtual mobility aid based on the" eyecane": Feasibility study," *PloS one,* **8**(8), p. e72555.

[28] Trewin, S., Hanson, V. L., Laff, M. R., and Cavender, A., 2008, "Powerup: An accessible virtual world," In Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility, Assets '08, Association for Computing Machinery, p. 177–184.

[29] Nair, V., Karp, J. L., Silverman, S., Kalra, M., Lehv, H., Jamil, F., and Smith, B. A., 2021, *NavStick: Making Video Games Blind-Accessible via the Ability to Look Around* Association for Computing Machinery, New York, NY, USA, p. 538–551.

[30] Zhao, Y., Cutrell, E., Holz, C., Morris, M. R., Ofek, E., and Wilson, A. D., 2019, *SeeingVR: A Set of Tools to Make Virtual Reality More Accessible to People with Low Vision* Association for Computing Machinery, New York, NY, USA, p. 1–14.

[31] Dong, M., Wang, H., and Guo, R., 2017, "Towards understanding the differences of using 3d auditory feedback in virtual environments between people with and without visual impairments," In 2017 IEEE 3rd VR Workshop on Sonic Interactions for Virtual Environments (SIVE), pp. 1–5.

[32] Morelli, T., Foley, J., Columna, L., Lieberman, L., and Folmer, E., 2010, "Vi-tennis: a vibrotactile/audio exergame for players who are visually impaired," In Proceedings of the Fifth International Conference on the Foundations of Digital Games, pp. 147–154.

[33] Glinert, E., and Wyse, L., 2007, "Audiodyssey: an accessible video game for both sighted and non-sighted gamers," In Proceedings of the 2007 conference on Future Play, pp. 251–252.

[34] Drossos, K., Zormpas, N., Giannakopoulos, G., and Floros, A., 2015, "Accessible games for blind children, empowered by binaural sound," In Proceedings of the 8th ACM international conference on pervasive technologies related to assistive environments, pp. 1–8.

[35] Milne, L. R., Bennett, C. L., and Ladner, R. E., 2013, "Vbghost: a braille-based educational smartphone game for children," In Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility, pp. 1–2.

[36] Kim, J., and Ricaurte, J., 2011, "Tapbeats: accessible and mobile casual gaming," In The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility, pp. 285–286.

[37] Balan, O., Moldoveanu, A., Moldoveanu, F., and Dascalu, M.-I., 2014, "Audio games—a novel approach towards effective learning in the case of visually-impaired people," In Proceedings of seventh international conference of education, research and innovation, Seville.

[38] Leporini, B., and Palmucci, E., 2017, "A mobile educational game accessible to all, including screen reading users on a touch-screen device," In Proceedings of the 16th World Conference on Mobile and Contextual Learning, pp. 1–4.

[39] Rovithis, E., and Floros, A., 2018, "Astrosonic: an educational audio gamification approach," In Proceedings of the DCAC Conference, Interdisciplinary Creativity in Arts and Technology, Corfu, Greece, pp. 1–2.

[40] Sánchez, J., and Lumbreras, M., 1999, "Virtual environment interaction through 3d audio by blind children," *CyberPsychology & Behavior,* **2**(2), pp. 101–111.

[41] Frauenberger, C., and Noistering, M., 2003, "3d audio interfaces for the blind," Georgia Institute of Technology.

[42] Koslover, R. L., Gleeson, B. T., De Bever, J. T., and Provancher, W. R., 2011, "Mobile navigation using haptic, audio, and visual direction cues with a handheld test platform," *IEEE Transactions on Haptics,* **5**(1), pp. 33–38.

[43] Brill, J. C., and Scerra, V. E., 2014, "Effectiveness of vibrotactile and spatial audio directional cues for usaf pararescue jumpers (pjs)," *Advances in Cognitive Engineering and Neuroergonomics,* **11**, p. 33.

[44] A blind legend https://store.steampowered.com/app/437530/A_Blind_Legend/ Accessed: 2021-07-15.

[45] Ear monsters https://www.audiogames.net/db.php?id=Ear+monsters Accessed: 2021-07-15.

[46] Blindside http://www.blindsidegame.com/ Accessed: 2021-07-15.

[47] Three d velocity https://github.com/munawarb/Three-D-Velocity Accessed: 2021-07-15.

[48] Nebula https://graycompany.com.br/.

[49] Sixthsense : All new 3d sound horror shooting game https://apps.apple.com/us/app/sixthsense-all-new-3d-sound/id580417319?ign-mpt=uo%3D2 Accessed: 2021-07-15.

[50] Audio archery - archery for your ears https://audio-archery-archery-for-your-ears-ios.soft112.com/ Accessed: 2021-07-15.

[51] Snow race https://www.agarchive.net/pages/devs/realitySoftware.html Accessed: 2021-07-15.

[52] Memory https://www.audiogames.net/db.php?id=tpbmemory Accessed: 2021-07-15.

[53] Shades of doom http://www.gmagames.com/sod.shtml Accessed: 2021-07-15.

[54] Sarah and the castle of witchcraft and wizardry for windows http://dracoent.com/Windows/SarahAndTheCastleOfWitchCraftAndWizardry Accessed: 2021-07-15.

[55] Audio games hub - hunt https://apps.apple.com/us/app/audio-game-hub/id1101972684 Accessed: 2021-07-15.

[56] Audio game hub - archery https://apps.apple.com/us/app/audio-game-hub/id1101972684 Accessed: 2021-07-15.

[57] Audio games hub - samurai tournament https://apps.apple.com/us/app/audio-game-hub/id1101972684 Accessed: 2021-07-15.

[58] Swaminathan, M., Pareddy, S., Sawant, T. S., and Agarwal, S., 2018, "Video gaming for the vision impaired," In Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '18, Association for Computing Machinery, p. 465–467.

[59] Xiao, J., Joseph, S. L., Zhang, X., Li, B., Li, X., and Zhang, J., 2015, "An assistive navigation framework for the visually impaired," *IEEE transactions on human-machine systems,* **45**(5), pp. 635–640.

[60] Ungar, S., Espinosa, A., Blades, M., Ochaíta, E., and Spencer, C., 1998, "Blind and visually impaired people using tactile maps," *Cartographic Perspectives,* **28**, pp. 4–12.

[61] Coughlan, J., and Manduchi, R., 2009, "Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users," *International Journal on Artificial Intelligence Tools,* **18**(03), pp. 379–397.

[62] Shadi, S., Hadi, S., Nazari, M. A., and Hardt, W., 2019, "Outdoor navigation for visually impaired based on deep learning," In Proc. CEUR Workshop Proc, Vol. 2514, pp. 97–406.

[63] Nie, M., Ren, J., Li, Z., Niu, J., Qiu, Y., Zhu, Y., and Tong, S., 2009, "Soundview: an auditory guidance system based on environment understanding for the visually impaired people," In 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, pp. 7240–7243.

[64] Microsoft soundscape https://www.microsoft.com/en-us/research/product/soundscape/ Accessed: 2021-07-15.

[65] Ross, A. S., Cutrell, E., Fiannaca, A., Kneisel, M., and Morris, M. R. "Use cases and impact of audio-based virtual exploration," In CHI 2019 Workshop on Hacking Blind Navigation.

[66] Zhao, Y., Bennett, C. L., Benko, H., Cutrell, E., Holz, C., Morris, M. R., and Sinclair, M., 2018, "Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation," In Proceedings of the 2018 CHI conference on human factors in computing systems, pp. 1–14.

[67] Proulx, M. J., Brown, D. J., and Pasqualotto, A., 2018, "The processing of what, where, and how: Insights from spatial navigation via sensory substitution," In PROCEEDINGS-BRITISH ACADEMY, Vol. 219, Oxford University Press, pp. 150–166.

[68] Campus, C., Sandini, G., Amadeo, M. B., and Gori, M., 2019, "Stronger responses in the visual cortex of sighted compared to blind individuals during auditory space representation," *Scientific reports,* **9**(1), pp. 1–12.

[69] Ribadi, H., Rider, R. A., and Toole, T., 1987, "A comparison of static and dynamic balance in congenitally blind, sighted, and sighted blindfolded adolescents," *Adapted Physical Activity Quarterly,* **4**(3), pp. 220–225.

[70] Morrongiello, B. A., Timney, B., Humphrey, G., Anderson, S., and Skory, C., 1995, "Spatial knowledge in blind and sighted children," *Journal of Experimental Child Psychology,* **59**(2), pp. 211–233.

[71] Fletcher, J. F., 1980, "Spatial representation in blind children. 1: Development compared to sighted children," *Journal of Visual Impairment & Blindness,* **74**(10), pp. 381–385.

[72] Klatzky, R. L., Golledge, R. G., Loomis, J. M., Cicinelli, J. G., and Pellegrino, J. W., 1995, "Performance of blind and sighted persons on spatial tasks," *Journal of Visual Impairment & Blindness,* **89**(1), pp. 70–82.

[73] Baumgartner, E., Wiebel, C. B., and Gegenfurtner, K. R., 2015, "A comparison of haptic material perception in blind and sighted individuals," *Vision research,* **115**, pp. 238–245.

[74] Rovira, K., Deschamps, L., and Baena-Gomez, D., 2011, "Mental rotation in blind and sighted adolescents: The effects of haptic strategies," *European Review of Applied Psychology,* **61**(3), pp. 153–160.

[75] Cattaneo, Z., Fantino, M., Tinti, C., Pascual-Leone, A., Silvanto, J., and Vecchi, T., 2011, "Spatial biases in peripersonal space in sighted and blind individuals revealed by a haptic line bisection paradigm.," *Journal of Experimental Psychology: Human Perception and Performance,* **37**(4), p. 1110.

[76] Yuan, B., and Folmer, E., 2008, "Blind hero: enabling guitar hero for the visually impaired," In Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility, pp. 169–176.

[77] Bujacz, M., Skulimowski, P., and Strumillo, P., 2011, "Sonification of 3d scenes using personalized spatial audio to aid visually impaired persons," International Community for Auditory Display.

[78] Moore, B. C., 2012, *An introduction to the psychology of hearing* Brill.

[79] Middlebrooks, J. C., and Green, D. M., 1991, "Sound localization by human listeners," *Annual review of psychology,* **42**(1), pp. 135–159.

[80] Giudice, N. A., 2018, "Navigating without vision: Principles of blind spatial cognition," In *Handbook of behavioral and cognitive geography*. Edward Elgar Publishing.

[81] Rodriguez-Sanchez, M., Moreno-Alvarez, M., Martín, E., Borromeo, S., and Hernandez-Tamames, J., 2014, "Accessible smartphones for blind users: A case study for a wayfinding system," *Expert Systems with Applications,* **41**(16), pp. 7210–7222.

[82] Yamamoto, N., and Shelton, A. L., 2009, "Orientation dependence of spatial memory acquired from auditory experience," *Psychonomic Bulletin & Review,* **16**(2), pp. 301–305.

[83] Viaud-Delmon, I., and Warusfel, O., 2014, "From ear to body: the auditory-motor loop in spatial cognition," *Frontiers in neuroscience,* **8**, p. 283.

[84] Voss, P., 2016, "Auditory spatial perception without vision," *Frontiers in psychology,* **7**, p. 1960.

[85] nielsen, s. h., 1992, "Auditory distance perception in different rooms," *journal of the audio engineering society*, march.

[86] Henriques, D. Y., and Soechting, J. F., 2005, "Approaches to the study of haptic sensing," *Journal of Neurophysiology,* **93**(6), pp. 3036–3043.

[87] Jain, M., Diwakar, N., and Swaminathan, M., 2021, *Smartphone Usage by Expert Blind Users* Association for Computing Machinery, New York, NY, USA.

[88] y Rita, P. B., and W. Kercel, S., 2003, "Sensory substitution and the human–machine interface," *Trends in Cognitive Sciences,* **7**(12), pp. 541–546.