VRBubble: Enhancing Peripheral Awareness of Avatars for People with Visual Impairments in Social Virtual Reality

Tiger Ji tfji@wisc.edu University of Wisconsin-Madison Madison, Wisconsin, USA Brianna R. Cochran bcochran2@wisc.edu University of Wisconsin-Madison Madison, Wisconsin, USA Yuhang Zhao yuhang.zhao@cs.wisc.edu University of Wisconsin-Madison Madison, Wisconsin, USA

ABSTRACT

Social Virtual Reality (VR) is growing for remote socialization and collaboration. However, current social VR applications are not accessible to people with visual impairments (PVI) due to their focus on visual experiences. We aim to facilitate social VR accessibility by enhancing PVI's peripheral awareness of surrounding avatar dynamics. We designed VRBubble, an audio-based VR technique that provides surrounding avatar information based on social distances. Based on Hall's proxemic theory, VRBubble divides the social space with three Bubbles-Intimate, Conversation, and Social Bubble-generating spatial audio feedback to distinguish avatars in different bubbles and provide suitable avatar information. We provide three audio alternatives: earcons, verbal notifications, and real-world sound effects. PVI can select and combine their preferred feedback alternatives for different avatars, bubbles, and social contexts. We evaluated VRBubble and an audio beacon baseline with 12 PVI in a navigation and a conversation context. We found that VRBubble significantly enhanced participants' avatar awareness during navigation and enabled avatar identification in both contexts. However, VRBubble was shown to be more distracting in crowded environments.

CCS CONCEPTS

• Human-centered computing \rightarrow Virtual reality; Accessibility technologies.

KEYWORDS

visual impairments, social virtual reality, proxemics, audio feedback

ACM Reference Format:

Tiger Ji, Brianna R. Cochran, and Yuhang Zhao. 2022. VRBubble: Enhancing Peripheral Awareness of Avatars for People with Visual Impairments in Social Virtual Reality. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22), October 23–26, 2022, Athens, Greece.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3517428. 3544821

ASSETS '22, October 23-26, 2022, Athens, Greece

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9258-7/22/10...\$15.00

https://doi.org/10.1145/3517428.3544821

1 INTRODUCTION

Social virtual reality (VR) refers to VR platforms that allow users to socialize with each other in the form of avatars in a virtual space [41]. More and more social VR platforms have been deployed to the market, such as VRChat, Rec Room, and Altspace [22, 43, 78]. Compared to 2D video conferencing systems that cause "Zoom fatigue" [47], social VR offers an immersive and engaging experience that can enhance interpersonal interaction and boost productivity. As a result, social VR has attracted increasing attention in recent years. The VR market was valued at \$7.81 billion as of 2020, and is expected to grow 28.2% annually from 2021 to 2028 [54]. Meta, formerly known as Facebook, also re-branded in 2021 alongside their full commitment to producing a Metaverse, envisioning social VR to be the next generation of Internet that connects everyone.

Unfortunately, similar to most mainstream VR applications, current social VR mainly targets sighted users by providing various visual avatar designs and supporting non-verbal social interactions, which poses barriers to people with visual impairments (PVI). With more than two billion people experiencing visual impairments worldwide [53], it is vital to provide PVI equal access to the emerging social VR as virtual collaboration and gathering increases, especially during the COVID-19 pandemic [50].

Researchers have started tackling the VR accessibility problems for PVI by enabling them to navigate and perceive VR scenes. Some research leveraged or created additional devices (e.g., PHANTOM, game controller thumbsticks) to enable VR navigation by providing haptic feedback and/or audio feedback [7, 24, 30, 46, 49, 68, 85]. Others focused on software solutions, designing accessible interactions based on existing VR setups, such as keyboard-based interactions with spatial audio feedback for a virtual world [52, 59, 72, 74], and more intuitive interactions based on off-the-shelf VR controllers and headsets [86]. However, prior work mainly focused on basic VR tasks such as navigation and object perception. It does not address the unique barriers caused by the dynamic and multiplayer nature of social VR.

Different from a static VR scene with only system-generated objects, social VR is more complex and challenging—human controlled avatars constantly move in the environment and interact with other avatars and objects. All the avatar dynamics are mainly presented visually in social VR and not accessible to PVI. To our knowledge, no existing techniques have focused on the avatar dynamics to support accessible social VR experience for PVI.

We aim to fill this gap by enhancing PVI's awareness of surrounding avatars in social VR. Unlike prior work that required PVI to actively explore and query information from the environment [49, 85], we focus on *peripheral awareness*—the innate ability to unconsciously "maintain and constantly update a sense of our social

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

and physical context" [56]. People can usually maintain peripheral awareness effortlessly without being distracted from their main focus [6]. This ability is especially important in social and collaborative environments since it provides more context for one's activity. For example, when navigating to a specific location, a sighted person can easily notice who has passed by to decide whether to greet that person or start a quick conversation; when in the middle of a conversation, they can stay aware of who just joined the conversation or who is close enough to overhear the conversation.

We seek to facilitate the peripheral awareness of avatars for PVI in social VR. Via an iterative design with six PVI, we designed *VRBubble*, an audio-based VR technique that provides surrounding avatar information based on social distances. Following Hall's proxemic theory [18], we split the virtual space with three "bubbles" centered on the user—the Intimate Bubble, Conversation Bubble, and Social Bubble—to represent different social spaces that are suitable for different social interactions. VRBubble then provides three spatial audio alternatives (i.e., earcons, real-world sound effects, and verbal notifications) to convey the avatar information, such as names, relationships with the PVI (friends or strangers), and their interactions with the bubbles (entering or leaving). PVI can also flexibly select and combine their preferred audio alternatives for different bubbles and avatars to maintain awareness of the avatar dynamics.

We evaluated VRBubble with 12 participants with visual impairments in the context of a navigation task and a conversation task in social VR, with an audio beacon attached to each avatar as the baseline. Our study showed that VRBubble enhanced user avatar awareness while navigating and was effective at providing users with previously inaccessible identifying information about avatars. We also found that PVI generally favor receiving verbal descriptions while navigating, and more brief and intuitive sounds while conversing.

2 RELATED WORK

2.1 Accessibility of Virtual Environments

2.1.1 Audio Techniques for VR accessibility. There has been extensive work that assisted PVI in exploring virtual environments via audio feedback. We summarize different types of audio techniques in prior work below.

Audio beacons. Audio beacons have been used to convey object positions [13, 35, 36, 79]. For example, Walker and Lindsay's [79] study utilized three different audio beacons in navigation guidance. They observed the impacts on PVI's navigation performance as they changed various parameters, such as timbre and distance to a waypoint to trigger the audio. Maidenbaum et al. [36] provided a beeping sound based on the distance between the PVI's avatar and the virtual object in front of them to facilitate navigation in a virtual space. As the avatar got closer to the object, the beeping rose in frequency of beeps. Blind Swordsman [13] was a VR game on mobile devices, where a blind user can hear the spatial audio beacon from the enemies, physically turn to that direction, and tap the touchscreen to swing his sword in the direction he is facing.

Object Sonification. Prior work has also used audio to identify objects within the virtual space [3, 10, 19, 42, 55, 57, 63–65]. For example, de Oliveira et al. [10] recreated a virtual stage and placed

instruments in the environment that generated spacial music. Participants then listened to the music tracks to identify and locate instruments on the stage. Heuten et al. [19] presented a sonification interface to virtual maps for PVI. PVI can listen to the spatialized earcons specific to various geographic objects and landmarks on the map to construct a mental model of a geographical area. AudioDoom [63, 64] was an acoustic virtual environment designed for blind children. Virtual objects and events generated spatial sounds to help users identify objects, navigate the space, and improve cognitive skills.

Echolocation. Virtual echolocation has been emulated through signals and audio reflections [2, 80, 83]. For example, Andrade et al. [2] enabled PVI to use echolocation to navigate a desktop-based virtual world, where the user's avatar can produce mouth-click or clap sounds by pressing a key on a keyboard and hear the sound reflected in the environment. Waters and Abulula [80] presented a sonar system based on the reflectivity of ultrasound used in the echolocation of bats. The audio used was scaled within human hearing ranges, so that PVI can utilize the audio to navigate a VR environment. However, the echolocation method was only used by a small amount of blind people.

User Queried Verbal Descriptions Some works allowed users to select objects in the environment via a list or a grid and hear additional descriptions about the selected object [37, 49, 74, 82]. For instance, Terraformers [82] was a virtual world game designed to be accessible to PVI. It provided a menu for nearby objects. A user can thus navigate the menu and hear audio descriptions of selected objects. Nair et al. [49] also compared similar menu systems with their novel game controller-based navigation technique, allowing PVI to look around a virtual world by scrubbing the thumbstick on a game controller to different directions; the system then announced what was in that direction via spatial verbal descriptions.

2.1.2 Haptic Solutions for VR accessibility. Prior work has also enhanced the accessibility of virtual environments for PVI by generating haptic feedback or creating haptic controllers [25, 33, 66-68, 75, 76, 81, 85]. For example, Jansson et al. [25] enabled PVI to use the stylus on a Phantom Premium device to "touch" a virtual space and receive force feedback to perceive different virtual surfaces and objects. Tzovaras et al. [76] leveraged the CyberGrasp haptic gloves to generate force feedback to a blind user's fingers, providing the illusion that they were navigating the virtual space with a cane. In the same vein, Zhao et al. created Canetroller [85], a wearable haptic VR controller that simulated white cane interaction for blind people in virtual reality. When a user swept the controller and hit a virtual object, a physical resistance and spatial audio sound effect were generated to simulate the feedback of a white cane hitting a real object. A follow-up work by Siu et al. [68] further improved the design of the controller by providing three dimensional force feedback. Recently, Wedoff et al. designed a virtual reality game, called Virtual Showdown [81]. PVI played the game by hitting a virtual ball into the opponent' goal across a virtual table using a bat. In this game, a Kinect was used to track the user's movement. A visually impaired user can then hold a Nintendo Switch controller as the bat and receive both audio and vibration feedback to perceive the relative position of the ball from his bat.

Prior research on VR accessibility focuses on space navigation and object perception. However, social VR introduces additional complications with the dynamic and non-uniform avatars that present social implications. No research has addressed the accessibility of avatars in social VR. Our research aims to fill this gap by facilitating PVI's awareness of avatars in social VR via customizable audio techniques.

2.2 Accessibility of the Real World

As with virtual environments, a myriad of prior work has designed audio techniques to sonify real world environments, including using audio beacons to mark waypoints [20, 84], informing users about nearby objects and landmarks through verbal descriptions [15, 66], generating auditory icons or earcons to identify points of interest [39, 58], and providing echolocation or sonar systems to enable the exploration of surrounding environments [21, 77]. Similar to audio techniques for virtual worlds, these solutions also do not address the dynamic complexity of avatars in a social VR context.

2.2.1 Technologies to Facilitate Real-World Social Activities. More relevant research has been done to assist PVI in real-world social activities by enhancing their awareness of the conversational partners and their non-verbal behaviors. Various wearable or handheld assistive technologies have been developed. These technologies came in a variety of forms, including smartphone applications [87], belts[5, 40], gloves [31, 32], headband [45, 60], glasses [1, 73], and wearable cameras [12, 73]. Some research focused on haptic feedback. For example, Krishna et al. [32] created a haptic glove called VibroGlove that allowed PVI to understand facial expressions of a conversation partner. The glove consisted of several vibration motors mounted on the back of each finger that was used to present six different emotions (i.e., anger, disgust, happiness, fear, surprise, and sadness) through vibration patterns. These patterns were designed based around the shape of the mouth and eye area of each facial expression. Another example was Tactile Ban, a wearable headband prototyped by Qiu et al. [60], which used tactile feedback to provide the feeling of other people's gazes onto the PVI. The band had two vibration patterns based on if a person glanced at the PVI or if their gaze was fixated on them currently.

Some work focused on audio feedback. For example, Anam et al. [1] used Google glasses to track faces and communicate social signals (i.e., facial features, behavioral expressions, body postures, and emotions) of surrounding people to the PVI verbally. Recently, Morrison et al. [45] designed PeopleLens, a head-mounted device that provided spatial identifying audio to assist blind children with gaze direction and mental mapping of surrounding people. Bump and woodblock sounds were used to guide the user's gaze to center on a face, while names would be read out for identified people in the surroundings as the user's gaze passed over them.

Prior work has focused on enhancing the primary social tasks, conveying information about the conversational partners. Besides the primary task, peripheral awareness (as a secondary task) is also important in social activities to *unconsciously* sense the surrounding dynamics and make *ad hoc* social decisions. However, this ability remains understudied for PVI.

2.2.2 Assistive Peripheral Awareness Technologies. There has been some prior work that facilitated peripheral awareness during collaborative tasks through visual [6, 61] or audio [9, 27, 29, 34] cues. For example, Cadiz et al. [6] created the Sideshow interface to support the user's peripheral awareness of information on their computer. The interface remained on the user's primary display and presented personalized information through visual notifications and summaries, such as number of unread emails or number of friends online. Sakashita et al. [61] enabled remote collaboration involving physical artifacts through the use of video conferencing and motion tracking. They placed video devices to represent each collaborator and automatically oriented devices to emulate the gaze direction of the collaborator. This supported the user's peripheral awareness of what part of the physical artifact a collaborator was focusing on. However, these works focus on sighted people by providing visual feedback.

Some work designed audio feedback to enhance PVI's peripheral awareness in a work collaboration scenarios [9, 27, 34]. For example, Lee et al. [34] designed the CollabAlly browser extension, which provided blind users with audio feedback to support the navigation of content or comment changes. Earcons were utilized to convey which part of the document was being edited and different voices were used for text-to-speech to contextualize which collaborator was editing. Jung [27] used ambient music to convey notifications to individuals within a physical work space. Speakers were placed throughout the space to enable spatial audio, then unique musical notifications were played at an individual's location when events relevant to that individual occurred, such as the reception of an email.

However, this work does not design for the unique challenges posed by a social context. Compared to collaborative tasks which involve a small number of collaborators and allow for asynchronous interactions, a social context can involve a large number of moving avatars, generating more peripheral information and distraction. Thus, a different design is needed to adequately facilitate peripheral awareness for the social context.

3 DESIGN OF VRBUBBLE

We designed *VRBubble*, an audio-based VR interaction technique to enhance the peripheral awareness of avatars in social VR for PVI. Our design followed the method of user-centered design [51]. We first formulated a set of general design considerations based on prior literature to design an initial prototype. With a formative study with six PVI, we further iterated and improved our design [26]. We describe our design process and the final design below.

3.1 General Design Considerations

We formulated the following design considerations (C1-C3) based on insights from prior literature [17, 87].

(C1) Leverage mainstream platforms. We focus on desktop VR, where a user can see the virtual environment on the desktop screen, hear the spatial audio feedback via earphones or speakers, and interact with it via keyboard and mouse. While VR headsets are emerging, most people don't own a headset due to its cost [70], let alone PVI who cannot benefit from the visual feedback from the headset. Instead, desktop computers are more widely used and

many social VR platforms (e.g., VRChat, AltSpace) support desktop access. Moreover, following the VR office concept by Grubert et al. [17], stationary VR in front of a desk that can be controlled by a keyboard can support more comfortable and efficient interaction for longer use and higher productivity in social and collaborative context.

(C2) Convey proper information about avatar dynamics. Prior research has explored PVI's needs for information in real-world social activities and indicated that the top two important information include the identity of surrounding people and their relative location [87]. We thus translate such needs to the virtual world and provide the corresponding avatar information to support PVI in the social VR context.

(C3) Avoid intrusiveness and distraction. Since peripheral awareness is supposed to be effortless [6, 38], we seek to minimize the distraction of our design as well as the user's conscious effort. We thus consider short but intuitive audio feedback at suitable timing to convey avatar information to PVI in an unobtrusive manner. Moreover, instead of having users to actively query information, we focus on proactive notifications to reduce their interaction effort.

3.2 VRBubble based on Hall's Proxemic Theory

To convey surrounding avatars' location information (C2) to PVI without overwhelming them (C3), our design followed Hall's proxemic theory [18] to divide the virtual environment into different social spaces. Hall's proxemic theory correlated physical distances with social interactions that typically happen within that distance, such as the distance for intimate interactions versus the distance for friendly conversation. Three distances were defined: *intimate distance* (1 foot) where people usually feel distress when this space is encroached upon unwillingly, *personal distance* (4 feet) where people interact with familiar people, and *social distance* (12 feet) where conversation with less familiar acquaintances or group happens. The space outside of the social distance is considered to be public space.



Figure 1: Conceptual diagram of bubbles.

We defined three virtual bubbles centered on a user based on the distance thresholds defined by Hall; the *Intimate Bubble, Conversation Bubble*, and *Social Bubbles* (Figure 1). We describe the social indication of each bubble:

Intimate Bubble defines the space within the intimate distance. Avatars in this bubble signify that they are about to collide with the user's avatar. Since PVI cannot visually perceive whether they are too close to others or whether other people are invading their intimate space, it is important to alert them when avatars enter this bubble (C2).

Conversation Bubble represents the space between the intimate and personal distance. Avatars in this space are close enough to start a conversation with. Moreover, the user may need to pay attention to these avatars since they are close to overhear one's ongoing conversation. We thus generate audio feedback to notify the user if any avatar enters or exits this space (C2).

Social Bubble defines the space outside of the personal distance, but still within social distance. Avatars in this space are potential conversational partners, but would not be in the immediate distance to start the conversation. The user could then decide whether she is interested in approaching this person for a conversation. Compared to the Intimate and Conversation Bubble, we expect that avatar information in the Social Bubble would be less important. We thus design more subtle audio feedback to inform the user of the avatars in this bubble (C3).

We do not consider avatars outside of the the social distance (i.e., public space in Hall's theory) since they are much less relevant to the users' current social context (C3).

3.3 Audio Alternative Design via Iterations

Based on the three bubbles, we designed spatial audio feedback to convey the surrounding avatars information, including the avatar identity and motion dynamics between these bubbles (C2), thus enabling PVI to build suitable social interactions upon sufficient avatar awareness. We also sought to make our audio feedback as least distracting and invasive as possible (C3). To achieve these goals, we designed and iterated on different audio feedback alternatives via a formative study.

3.3.1 **Formative Study**. Following the method of user-centered design [51], we conducted a formative study [26] with six PVI (three male, two female, and one who preferred not to say) whose ages ranged from 22 to 58 (*mean* = 44.167, *SD* = 13.348). All participants were legally blind.

Initial Design. In the formative study, we presented our initial design of VRBubble with the three bubbles described before. When an avatar entered or exited a bubble, the user heard an earcon with a verbal description of the avatar's information. Earcons were utilized as they are brief, abstract, and distinctive (C3) sounds that encodes particular information [4]. We used different earcons to represents an avatar's moving dynamics between different bubbles. A twobeat earcon with increasing tone \checkmark (or decreasing tone \checkmark) was used to represent an avatar entering (or leaving) the Social Bubble; similar two-beat earcons with a different timbre were used for the Conversation Bubble; and an "bumping" sound earcon was used to indicate an avatar in the Intimate Bubble. All earcons were accompanied with a more informative verbal description (C2), reporting the avatar's name, relationship with the user (friend or stranger), and the relative position ("nearby" for Conversation Bubble, "In vicinity" for Social Bubble). For example, the user heard "Friend Alice nearby (or no longer nearby)" if Alice's avatar entered (or left) the Conversation Bubble. All audio feedback were spatial audio rendered from the avatar's position. We prototyped VRBubble using web-based VR (C1, details in Section 4.1.2).

Method and Findings. The formative study was hosted virtually through zoom and took roughly 2 hours for each participant to complete. Participants were given access to our VR environment through a url. Participants were introduced to the initial VRBubble design through a short tutorial. We then asked them to experience VRBubble in two tasks: a navigation task, and a conversation task. Both tasks were used to simulate common social contexts they could encounter in VR. During the prototype experiencing process, participants thought aloud, describing whether they like this feature or not and why. By the end, participants discussed how they wanted to improve VRBubble in different social context.

We summarize our major findings from the formative study. (1) While most participants found the earcon design useful, understanding the abstract earcons could create a steep learning curve for PVI. Some participants suggested more intuitive sound effects to present avatar information. (2) Verbal description is clear and easy to understand, but it could be distracting especially in a conversation context. Participants suggested shortening the verbal description to reduce distraction. (3) Participants had different preferences for audio feedback and valued various avatar information differently. Our design should provide the flexibility for users to customize their audio experience for different avatars and social contexts.

3.3.2 **Three Audio Alternatives.** Based on the findings in the formative study, we designed three spatial audio feedback alternatives: earcons, verbal notifications, and real-world sound effects. Each alternative presented similar avatar information, including avatar identity (name and/or relationship with the PVI) and motion dynamics between bubbles (C2). We describe the design of three audio alternatives for each bubble:



Figure 2: Three audio alternatives for different bubbles: (a) earcons, (b) verbal notifications, (c) real-world sound effects.

Earcon. Given that earcon is abstract and has a high learning curve, we used earcons to present simple information, such as distinguishing different bubbles and different avatars (friend avatar vs. stranger avatar). Instead of completely different earcons, we designed associated earcons with distinctions to minimize the learning curve (C3). We used a two-beat earcon with the tone of the last beat

increased $\overset{1}{\mathcal{N}}$ ¹ (or decreased $\overset{1}{\mathcal{N}}$ ²) to indicate an avatar entering (or leaving) the Social Bubble. For the Conversation Bubble, since

avatars in this bubble probably required more immediate attention, we used four-beat earcons with the tone of last beat increased \mathcal{M}^{*} ³ or decreased \mathcal{M}^{*} to represent an avatar entering or leaving this bubble. To distinguish friend and stranger avatars, we adjusted the pitch and speed of the earcons, so that higher pitched, faster earcons indicated friends, while normal pitch and speed indicated strangers ^{5 6 7 8}. We used a game-like bump earcon ⁹ to signify when an avatar was in the Intimate Bubble.

Verbal Notifications. We provided clear and short (C3) verbal notifications to present avatar information. To reduce distraction, we informed avatar identity by only reporting the name. We no longer provided information on if the avatar was a friend or a stranger since in the real world a person would be able to know whether someone was a friend by name. We also used verbal notification to convey the avatar's general position based on their interaction with the bubbles. Specifically, we used "in the area" ¹⁰ and "left the area" ¹¹ to indicate when an avatar entered or left the Social Bubble, and "nearby" ¹² and "no longer nearby" ¹³ for entering or leaving the Conversation Bubble. If an avatar, Alice, entered the Conversation Bubble, a user would hear "Alice nearby." To signify that an avatar was in the Intimate Bubble, we used "collided with" ¹⁴ followed by the avatar name.

Real-world Sound Effect. This design alternative focused on providing realistic real-world sounds to intuitively support peripheral awareness and immersion, thus reducing the amount of effort needed to comprehend the audio notification (C3). We used crowd sound effects as the background sound, with different densities to represent the total number of avatars in all three bubbles within the social distance (12 feet). Two levels of crowd sound effects with increasing volume and densities indicate two levels of stranger avatar amount: 1-5 ¹⁵ and more than 5 ¹⁶ strangers. Two levels of cheering crowd sounds were also used to highlight friend avatars due to their higher importance, indicating two levels of friend avatar

¹Friend entering Social Bubble earcon: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/f_soc_in.mp3?v=1639640993569

²Friend leaving Social Bubble earcon: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/f_soc_out.mp3?v=1639640998113

³Friend entering Conversation Bubble earcon: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/f convo in.mp3?v=1643983723059

⁴ Friend leaving Conversation Bubble earcon: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/f_convo_out.mp3?v=1643983757550

⁵Stranger entering Social Bubble earcon: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/s_soc_in.mp3?v=1649899925021

⁶Stranger leaving Social Bubble earcon: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/s_soc_out.mp3?v=1639641013488

⁷Stranger entering Conversation Bubble earcon: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/s_convo_in.mp3?v=1649900021917

⁸Stranger leaving Conversation Bubble earcon: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/s_convo_out.mp3?v=1643983842245

⁹Intimate Bubble earcon: https://cdn.glitch.me/11db17e0-58dc-4f24-8390bb2e597d1475%2F483602_raclure_game-bump.mp3?v=1636388046675

 $^{^{10}}$ Entering Social Bubble verbal: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/area.mp3?v=1643807674537

 $^{^{11}}Leaving$ Social Bubble verbal: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/noarea.mp3?v=1643807963569

¹²Entering Conversation Bubble verbal: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/nearby.mp3?v=1643807938153

¹³Leaving Conversation Bubble verbal: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/nonearby.mp3?v=1643807989565

¹⁵Low ambient stranger crowd: https://cdn.glitch.me/11db17e0-58dc-4f24-8390bb2e597d1475/stranger_crowd_small.wav?v=1639761138064

¹⁶High ambient stranger crowd: https://cdn.glitch.me/11db17e0-58dc-4f24-8390bb2e597d1475/stranger_crowd_medium.wav?v=1639761063563

amount: 1-5¹⁷ and more than 5¹⁸ friends. Moreover, we put more emphasis on the avatars in the Conversation Bubble since the user was more likely to converse with them. A spatial footstep sound effect was regularly played for every avatar within the Conversation Bubble. We generated different footsteps sounds to distinguish friend avatars from strangers, with the friend avatars presenting a higher pitch heel sound ¹⁹ and the stranger avatars presenting a boot stomp ²⁰. We also generated a realistic bump sound ²¹ to indicate an avatar entering the Intimate Bubble.

3.4 Customization

Customization was key to our design since different PVI have different audio preferences and prioritize different information and goals. Instead of generating all audio feedback together (as we did in the initial design), we allowed users to select and combine different audio alternatives for different bubbles, avatars (e.g., friend vs. stranger), and social context to customize their experiences. For example, a user could choose verbal notifications for friends but earcons for strangers in the Social Bubble, and combine both audio alternatives for both friends and strangers in the Conversation Bubble. The user was also allowed to select none audio feedback for a specific type of avatar or bubble (e.g., stranger avatars in the Social Bubble).

4 EVALUATION

We evaluated VRBubble with 12 PVI in different social contexts. We aimed to understand: (1) How effective is VRBubble in enhancing PVI's peripheral awareness of avatars in different social contexts? (2) How distracting is VRBubble? (3) How do PVI customize their audio experience for different avatars, bubbles, and social contexts?

4.1 Method

4.1.1 **Participants**. We recruited 12 participants with visual impairments (7 male, 5 female) whose ages ranged from 20 to 68 (*mean* = 38.92, *SD* = 13.79, Table 1). No participants were in the formative study. The participants were recruited through the National Federation of the Blind. We used a survey to check the participants' eligibility. Participants were eligible for our study if they were at least 18 years old, legally blind, and capable of independent consent. We also asked about participants' VR experience in the survey. We prioritized participants who had experience with VR or virtual environments. Among the 12 participants, nine had VR experience, such as audio-based VR games, VR for vision therapy, and VR environment exploration. However, no participants had experience with social VR. Participants were compensated at the rate of \$20 per hour.

Table 1: Demographic information of 12 participants.

ID	Age/Sex	Prior VR Experience			
P1	31/M	Vision therapy			
P2	57/M	3D virtual home design game			
P3	37/M	Audio-based VR games; VR studies			
P4	68/M	VR environments; VR studies			
P5	44/F	Simulated street navigation VR app			
P6	34/F	VR headset games			
P7	49/F	Oculus Quest accessibility testing; VR games			
P8	25/F	None			
P9	33/F	None			
P10	46/M	PVI accessible audio games			
P11	23/M	Navigational VR study			
P12	20/M	None			

4.1.2 **Apparatus.** We built a custom VR environment with avatars and implemented both VRBubble and a baseline feature in the VR environment. To evaluate VRBubble in different contexts, we implemented two VR scenarios, a navigation scenario and a conversation scenario. We describe the design and implementation of our study environment.

Baseline. We compared VRBubble's performance to a baseline feature. Our baseline was an audio beacon that consisted of a spatial beeping sound, in intervals, attached to each avatar (except for the participant avatar) in the environment. We chose this baseline since audio beacon was an effective and common way of conveying the presence of people and objects for PVI in the real world [11]. Some current assistive technology for PVI (e.g., Microsoft Soundscape [44]) also uses audio beacon to label target destinations for PVI.

VR environment and Two Scenarios. We built a custom VR environment with two key VR scenarios: a navigation scenario and a conversation scenario. In the navigation scenario, we generated a VR space plan with eight similar navigation routes (Figure 3(a)). All routes were approximately 60 ft with one turn (either turn left or turn right). All routes were considered to be the same according to Ishikawa et al.'s study design [23] since they had same length and same number of turns. Avatars were rendered at along the routes at various distances (Figure 3(a)).

In the conversation scenario, we rendered one avatar in front of the user, conversing with the user. Other avatars were rendered and moving around in the virtual space during the conversation (Figure 3(b)).

Features for Basic Movements To enable participants to navigate in the virtual environment and experience the avatar awareness features, we added two fundamental features to allow PVI to move and turn in a virtual environment: (1) *Movement:* PVI can control the movement of their own avatar via arrow keys, for example, up arrow to move forward and left arrow to move left. Each key press would move the PVI a foot in the virtual space and play a footstep sound to notify the PVI. This footstep sound ²² is centered on the user's avatar and differs from the footsteps used for VRBubble by pitch. (2) *Turning.* A user can turn her avatar to the left (or

 $^{^{17} \}rm Low$ ambient friend crowd: https://cdn.glitch.me/11db17e0-58dc-4f24-8390-bb2e597d1475/friend_crowd_small.wav?v=1639761145228

¹⁸High ambient friend crowd: https://cdn.glitch.me/11db17e0-58dc-4f24-8390bb2e597d1475/friend_crowd_medium.mp3?v=1639761151401

 $^{^{19}}$ Friend footstep: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/fstep.mp3?v=1643754264523

²⁰Stranger footstep: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/ sstep.mp3?v=1643754271035

²¹Collision: https://cdn.glitch.global/11db17e0-58dc-4f24-8390-bb2e597d1475/mixkitquest-game-heavy-stomp-v-3049.wav?v=1644177861935

²²User footstep: https://cdn.glitch.me/11db17e0-58dc-4f24-8390-bb2e597d1475% 2F571698_rainial-co_cotton-thud-7.wav?v=1636411987805



Figure 3: Two tasks: (a) navigation, (b) conversation.

right) by pressing the left (or right) arrow keys while holding the shift key. A ticking sound ²³ would play for each 90 degree turn, providing a sense of how much the PVI had turned.

VR Environment Implementation and Setup. We implemented the VR environment and all the features with A-Frame [71], a web framework that generates 3D spaces via HTML and Javascript. We hosted our prototype on Glitch [16], a browser-based platform that allows users to host and share applications as well as share the source code, to enable a remote setup for the study. Participants and researchers can go to the same web address to join the same VR environment. Once connected, researchers can input unique command sequences via a keyboard to adjust the participant's client, such as adjusting the audio alternatives in VRBubble or moving the participant's avatar to a different VR scenario. Participants' behavioral data (e.g., their positions at different times) was then logged and output into the researchers' browser console.

We asked participants to share their screen with audio via Zoom to ensure the commands from the researchers were properly reflected on the participants' end and to confirm that the audio feedback was properly generated at the participants' end.

4.1.3 **Procedure.** Our study consisted of a single session that lasted 2 hours for each participant. We conducted the study remotely via a Zoom video call. An initial interview was first conducted to ask about participants' prior experience with VR, especially social VR. Then we asked about their social experiences in real life.

We then asked participants to experience our VR environment with the awareness features. Participants were requested to use their earphones for the study, since we generate spatial audio. We sent a url to our VR environment through email before the study. If any participant had difficulty accessing their email, we sent the link again through the Zoom chat or read the link out-loud to them.

After the participant successfully joined the VR environment, we evaluated participants' experience and performance in two context: (1) a navigation context and (2) a conversation context. In each context, participants completed navigation or conversation tasks with two awareness features—VRBubble and the baseline (i.e., audio beacon). We counterbalanced the awareness feature and the context across all participants. We elaborate the details of the remaining study in three phases: tutorial, navigation, an conversation.

Tutorial: Introducing the Features. We conducted a tutorial session to teach participants how to use our VR prototypes. We first introduced the interactions for basic movements (arrows keys to move and shift + left/right arrow key to turn, as explained in

Section 4.1.2). We then rendered an empty VR space and encouraged participants to move around until they were familiar with the controls and the audio feedback produced by the avatar movement. We then introduced VRBubble and the baseline.

In the tutorial for VRBubble, we introduced each audio alternative one by one. For each alternative, we demonstrated the audio feedback of both friend avatar and stranger avatar for each Bubble. We then rendered avatars in the VR space and asked participants to freely explore the environment with this audio alternative until they felt fully familiar with it. During the exploration, participants were asked to think aloud, talking about their experiences with this audio design, including whether they like this design or not and the reason. We counterbalanced the order of the audio alternatives using Latin Square [48]. We used the similar method to introduce the baseline.

Navigation. We evaluated the effectiveness of VRBubble in the navigation context. Each participant conducted 15 trials of navigation tasks along the routes in Figure 3a in different conditions. We randomly selected a route for each trial. To enable participants to follow the route and arrive at the destination, we generated audio-based verbal instructions. Participants heard "turn left (or right)" when they needed to turn, and heard "arrived destination" when they reached the end of the route. For each trial, the participant was first moved to the start point of a route. When the research announced "Start", they would navigate by following the verbal instruction until arriving at the destination. We asked participants to complete each navigation task as quickly as possible.

Participants first conducted three trials of basic navigation tasks without any avatars around. These trials helped participants get familiar with the navigation tasks. We then asked participants to customize their audio feedback in VRBubble based on the navigation task. To guide participants through the customization, we asked them to choose and combine audio feedback for the five following situations: (1) when a friend enters or leaves the Social Bubble; (2) when a stranger enters or leaves the Social Bubble; (3) when a stranger enters or leaves the Conversation Bubble; (4) when a friend enters or leaves the Conversation Bubble; (5) when an avatar collides with the user in the Intimate Bubble. For each situation, participants could select any combination of audio feedback ranging from multiple to none at all (Figure 7). During the customization process, the researcher adjusted the audio alternatives in real time for the participant to experience the selected feedback and make immediate corrections. After customizing VRBubble, we asked about the reason behind their choices and any suggestions they have for customizing. Once the customization phase was completed, the researcher adjusted the audio feedback in VRBubble to the participant's preferred combination.

After the customization, participants continued to perform another 12 trials of navigation tasks with avatars around. Participants were asked to complete two tasks at the same time: a primary navigation task (i.e., navigating to the destination following the verbal instructions) and a secondary avatar awareness task (i.e., memorizing how many avatars they've passed by and who they are). We asked participants to reach the destination as quickly as possible, but also try their best to perceive and memorize the surrounding avatars. Before the navigation task, we informed the participants ahead that we would asked them to report the avatar number and

²³Turn audio: https://cdn.glitch.me/11db17e0-58dc-4f24-8390-bb2e597d1475% 2F268108_nenadsimic_button-tick.wav?v=1636008594574

names they remembered after each trial. However, if the task was too difficult, we asked participants to prioritize the primary navigation task.

Participants completed above tasks in two awareness conditions: using VRBubble with the customized audio alternatives, and using the baseline. Each condition included six trials of navigation. To further understand the effectiveness of VRBubble in situations with different avatar complexity, we generated avatars in two amount levels: Low Amount (1-5 avatars) and High Amount (6-10 avatars). We decided whether a trial had low or high amount of avatars randomly but ensured that participants experienced each avatar amount level for three trials per awareness condition. For each trial, we randomly generated avatars within the corresponding avatar amount range, and the avatars were randomly distributed along the navigation route. We guaranteed that all avatars were within the social distance (12 feet) to the route, so that all avatars were detectable by VRBubble. After each navigation trial, we asked participants to report their perceived avatar information, including the total avatar number, the number of friend avatars, the number of stranger avatars, and the avatar names they heard. We timed participants' navigation time for all trials.

By the end, we asked about participants' general experience with VRBubble and the baseline. Participants rated the effectiveness, distraction, and immersion of both features via 5-point likert scale scores. Participants also discussed their feedback and suggestions for improving VRBubble in the navigation context.

Conversation. The procedure of the conversation task was similar to the navigation task. This phase of the study was conducted in the conversation scenario (Figure 3b). Participants conducted three trials of conversation task in different conditions. For each trial, the participant faced an avatar who asked 10 yes/no questions. Participants were asked to answer all questions as quickly and accurately as possible. A researcher on the team acted as the conversation avatar to ask the questions. The next question was asked immediately after the participant answered the prior question.

Participants first conducted one trial of "dry-run" task, answering 10 questions without any distraction to get familiar with the task. We then asked them to customize the VRBubble feedback again for the conversation context. Participants then continued to perform two trials of conversation tasks with avatars around. One trial used the customized VRBubble and the other used the baseline. For each trial, participants conducted two tasks simultaneously: a primary conversation task (i.e., answering 10 yes/no questions) and a secondary awareness task (i.e., perceiving and memorizing surrounding avatars). The primary task was prioritized if the task was too difficult. After each trial, we asked participants to recall the avatar information, including total avatar number, the number of friends and strangers, and the avatar names.

To balance the avatar complexity in each trial, we pre-defined two similar groups of avatar moving dynamics. Both groups had eight avatars spawn in and move around the user over the duration of 70 seconds during the conversation process. Each group had four friend avatars and four stranger avatars. However, the order of avatars and the timing they spawn in were different between the two groups. We counterbalanced the avatar groups with the awareness feature used across participants. We pre-planned 30 yes/no questions. All questions were easy-toanswer questions that asked about a specific fact about the participants to minimize the variances caused by participants' literacy and cognitive ability. Some example questions included "Have you ever been to Italy?", "Are you a coffee drinker?" We randomly pooled 10 questions for each trial and also made sure not to repeat any questions across the study for each participant. We measured participants' response time for each question.

After all trials, we asked all 30 yes/no questions again without any distractions and collected participants' answers as the correct answers. Participants then assessed the effectiveness, distraction, and immersion of VRBubble and the baseline for the conversation context via 5-point likert scale scores. Finally, participants talked about their qualitative feedback and suggestions for VRBubble in the conversation context.

We ended the study by discussing with the participants about their general experiences with VRBubble compared to the baseline, as well as their desired awareness technology for different contexts in social VR in the future.

4.1.4 **Analysis**. We detail our analysis methods for both the quantitative and qualitative data in our study.

Navigation Performance We first analyzed the impact of VR-Bubble on participants' performance in the navigation task, including navigation time and the error rate when recalling the number of avatars they passed by. We had three within-subject factors: *Feature* (two levels: *VRBubble*, *Baseline*) that defined the avatar awareness feature participants used, *AvatarAmount* (*Low: 1-5, High: 6-10*) that specified the amount of avatars in the space, and *Trial* (1-3) that defined each navigation task in a specific condition. We had two measures: *NavigationTime*—the time taken in seconds by the participant to reach the end of the navigation path, and *Avatar*-*ErrorRate*—the ratio of the difference between the reported avatar number and the actual number to the actual number of avatars.

We also added a between-subject factor, *Order* (*VRBubble-Baseline*, *Baseline-VRBubble*) to our model for counterbalance validation. The Shapiro-Wilk test showed that both NavigationTime (W = .867, p < .001) and AvatarErrorRate (W = .898, p < .001) deviated significantly from a normal distribution, so we used Aligned Rank Transform for nonparametric factorial ANOVAs (ART) [28] ²⁴ to model our data. We found no significant effect of Order ($F_{1,10} = 2.296, p = .161$) on the navigation time and no significant effect of the interaction between Order and other factors on navigation time. Similarly, we also found no significant effects of Order ($F_{1,10} = .0005, p = .982$) on avatar error rate and no significant effect of the interaction between Order and other factors on error rate.

Conversation Performance We then analyzed the impact of VRBubble on participants' performance in the conversation task, including the response time per question, accuracy of answering questions, and the error rate when recalling the number of passedby avatars during the conversation. We had two within-subject factors: *Feature (VRBubble, Baseline)*, and *AvatarGroup (Group1, Group2)* that indicated which pre-defined set of avatar dynamics was used during each conversation trial. We had three measures: *AvatarErrorRate, ResponseTime*—the mean response time for the

²⁴ART is a non-parametric approach to factorial ANOVA or linear mixed effect model

participant to answer a question, and *IncorrectAnswers*—the number of questions that participants answered inconsistently compared to the standard answers or questions that needed to be repeated.

A between-subject factor, *Order* (*VRBubble-Baseline, Baseline*, *VRBubble*), was added to valid our counterbalancing. The Shapiro-Wilk test showed that ResponseTime (W = .886, p = .011) and IncorrectAnswers (W = .654, p < .001) were non-normally distributed, while AvatarErrorRate (W = .972, p = .708) followed a normal distribution. We thus used ART to model ResponseTime and IncorrectAnswers, and used ANOVA [14] for AvatarErrorRate. We found no significant effect of Order on the response time ($F_{1,8} = .293, p = .603$) or questions answered incorrectly ($F_{1,8} = .802, p = .397$) and no significant effect of the interactions between Order and other factors on response time or questions answered incorrectly. Similarly, we found no significant effects of Order ($F_{1,8} = .702, p = .426$) or its interactions on the error rate of recalling avatars.

Moreover, we found no significant effect of AvatarGroup or its interactions on ResponseTime ($F_{1,8} = .026$, p = .876), IncorrectAnswers ($F_{1,8} = .82$, p = .392), and AvatarErrorRate ($F_{1,8} = .157$, p = .097). These results confirmed that the two pre-defined sets of avatars had similar complexity.

Post hoc Analysis. Based on the results of the ART and ANOVA models, we conducted post hoc tests to further investigate the relationship between the different levels of significant factors. Specifically, we used paired t-test for normally distributed data, and used the pairwise Wilcoxon signed-rank test for non-normally distributed data. We corrected the p-value threshold with Bonferroni Correction.

Qualitative Analysis. We video recorded the whole study and transcribed participants' verbal feedback and responses to the interview questions with an automatic transcription service, Otter.ai. Researchers on the team also read through the transcripts and corrected the auto transcription errors in the transcripts. We analyzed the transcripts using the the qualitative analysis method described by Saldaña [62]. Two researchers first independently coded the first four participants' transcripts. They then discussed and generated a codebook upon agreement. One researcher then followed the codebook and coded the rest of the data. When a new code emerged, the two researchers discussed again and updated the codebook. The codes were then organized and categorized into themes using affinity diagram.

4.2 Results

We report the impact of VRBubble on participants' performance in different context, including both a navigation and a conversation task. We also report participants' experiences with different audio alternatives and their preferences when customizing VRBubble for different context.

4.2.1 VRBubble in Navigation. For the navigation task, we found that VRBubble provided the user with more detailed and accurate information about surrounding avatars, compared to the baseline. However, we also noticed that VRBubble could be more distracting than the baseline. We report the detailed results below.

Avatar Amount Estimation. We analyzed the effectiveness of VRBubble on enhancing PVI's awareness of the amount of avatars nearby in a navigation context. An ART analysis showed a significant effect of the awareness feature on participants' error rate when estimating the amount of avatars they passed by ($F_{1,110} = 33.54$, p < .001). We then conducted a *post hoc* paired Wilcoxon signed-rank test and found that participants had a significantly lower error rate when using VRBubble than the baseline (V = 1566.5, p < .001, error rate with VRBubble: *mean* = .239, SD = .26, baseline: *mean* = .465, SD = .323). This result demonstrated that VRBubble significantly improved PVI's awareness of the surrounding avatar amount.

We examined the error rate more closely for environments with different avatar complexity (AvatarAmount): low amount of avatars vs high amount of avatars. A *post hoc* Wilcoxon signed-rank test showed that VRBubble significantly decreased participants' error rate of avatar number estimation regardless the avatar amount was high (V = 585, p < .001) or low (V = 259, p = .01 < .025 with Bonferroni Correction). This indicated that the effectiveness of VRBubble was consistent across environments with different avatar complexity.

Participants' qualitative feedback also explained the effectiveness of VRBubble. Ten participants expressed positive sentiment towards VRBubble, emphasizing its ability to provide distinguishable audio feedback for each avatar, with seven noting VRBubble created a better sense of distance for avatars via the different bubbles. In contrast, participants found the audio beacon baseline hard to use since all avatars shared the same beacon sound and it was difficult to distinguish different avatars, especially when avatars were close together. P6 "had a hard time differentiating if it was the beacon for the additional avatar or the same avatar. I had a hard time distinguishing how many people or avatars there were."

Avatar Identification. Compared to the baseline, VRBubble enabled participants to collect more detailed information about surrounding avatars, including avatar names and the relationship to the user.

When customizing VRBubble, nine participants selected verbal notification that reported avatar names. With this feature, participants in general recalled the names of passed-by avatars accurately with a mean accuracy of 0.715 (SD = .305). Six participants expressly liked VRBubble's capacity to provide information about each individual avatar. For example, P3 emphasized the importance of name identification in engaging in social interactions: "[The name] would potentially determine whether or not I would want to go and speak to them ... or even somebody I want to avoid ... I can see that having been named there is a useful feature." However, we noticed that the avatar identification accuracy of some participants dropped to below 0.5 in more crowded avatar environments. This was because the verbal notifications from different avatars would overlap and become confusing when the avatar reached high amounts. As P5 noted, "If it's not too many people happening at once then I can feel it out, but then when so many people are happening, then I can't really figure it out."

Ten participants' VRBubble customization supported friend identification (i.e., earcon or real-world sound effect for Conversation Bubble). Most participants distinguished friends and strangers accurately with a mean accuracy of 0.658 for friends (SD = .367) and 0.656 for strangers (SD = .376). Ten participants agreed that discerning between friends and strangers was important for them to decide what social actions to take. As P2 said, "I might want to ask a stranger what time it is. But if I hear friendly footsteps coming towards me, I want to know who's approaching and I can strike up conversation." However, P4 was an outlier who consistently reported many more avatars than the actual friend and stranger numbers in the space. This was because P4 used the abstract earcons for the Conversation Bubble and found it difficult to discern the pitch changes between friends and strangers.

The two participants (P5, P10) chose verbal notification for the Conversation Bubble and thus could not distinguish friends and strangers. They customized VRBubble based on their real-life experiences, as in real-life they would recognize if someone was a friend or stranger by name. We envisioned that users would be able to distinguish friends and strangers via verbally reported names in a real social VR setting after long-term use.

Impact on Navigation Time. We compared the effects of awareness feature on participants' navigation time. With an ART model, we found a significant effect of awareness feature on navigation time ($F_1 = 36.708, p < .001$). A post hoc paired Wilcoxon signed-rank test with correction showed that participants navigated significantly slower when using VRBubble than using the beacon baseline (V = 482, p < .001, navigation time with VRBubble: *mean* = 36.507, SD = 14.897, baseline: *mean* = 31.293, SD = 11.033). The result indicated that VRBubble distracted participants more than the baseline in the navigation task.

We further investigated the effect of VRBubble under different avatar complexity (low amount of avatars vs. high amount). We found a significant effect of the interaction between Feature and AvatarAmount on participants' navigation time ($F_1 = 13.676$, p <.001). Using a *post hoc* Wilcoxon signed-rank test with Bonferroni Correction, we found that, there's no significant difference in time between VRBubble and baseline when the avatar number was low (V = 209, p = .052), but participants walked significantly slower with VRBubble than the baseline when the avatar number was high (V = 66, p < .001). This indicated that VRBubble was only more distracting when the environments became crowded.

The distraction of VRBubble could be caused by the more diverse audio feedback provided by VRBubble, where participants spent longer time to associate the sounds with different events. Six participants mentioned being distracted and slowed down since they tried to remember the meaning of the different sounds. As P5 explained, "If I go too fast, I might miss it ... I have to pay attention on top of all the other avatars leaving, coming in, and coming out."

Another reason that may cause a slower navigation was participants' interest. Since VRBubble enabled participants to receive more information about the surrounding avatars, it aroused participants' interest to explore the environment during the navigation. In our study, while we emphasized that navigation was the primary task, we still observed that some participants (e.g., P2, P5, P12) stopped, looked around, and even moved their avatar left and right to explore nearby avatars with VRBubble. In contrast, since the baseline did not provide much information, participants mostly focused on the navigation.

4.2.2 VRBubble in Conversation. Unlike in the navigation task, VRBubble generally did not perform significantly different from

the baseline in the conversation context. We report the detailed measures below.

Avatar Amount Estimation. We analyzed the effectiveness of VRBubble on enhancing PVI's awareness of avatar amount nearby in a conversation context. An ANOVA analysis showed no significant effect of the awareness feature on participants' error rate when estimating the amount of avatar they passed by ($F_{1,8} = 1.431$, p =.266). Compared to the navigation task, we found that participants made more errors when estimating the avatar number with VRBubble in the conversation task (mean = .305, SD = .191). The error rates went especially high when there are more avatars around. Interestingly, we noticed that the reported avatar number by participants never exceeded five across all trials, even though there were usually more avatars present. Specifically, eight participants encountered more than five avatars in the conversation task, and the highest avatar number was eight. This might indicate a limit in how much audio feedback can be peripherally processed (or a upper bound of avatar number that can be perceived) while focusing on an audio-centered task such as conversation. As P4 noted, "Because you had to answer the question, so you couldn't put all your brainpower into [the surrounding avatars]. I just don't think the human brain is segmented enough to handle both those tasks at the same time."

Avatar Identification. Similar to the navigation task, participants appreciated VRBubble's ability to identify surrounding avatars in real-time since it enabled them to dynamically adapt to the social environments. Ten participants indicated that knowing who was in the immediate surroundings had an impact on their social behaviours and what they would say during a conversation.

Seven participants customized VRBubble to allow for name identification. However, participants' avatar recognition accuracy were lower (*mean* = .655, *SD* = .282) compared to the navigation task (*mean* = .715, *SD* = .305). Moreover, ten participants customized VRBubble to distinguish friends from strangers. Similarly, they were less accurate compared to those in the navigation task, both when identifying friends (*mean* = .567, *SD* = .323) and when identifying strangers (*mean* = .467, *SD* = .302). The lower accuracy could be attributed to the more attention-demanding nature of the conversation task, and the direct conflict between verbal notifications and the conversation.

Impact on Conversation. We investigated the impact of VR-Bubble on participants' ability to answer questions in a conversation. With an ART model, we found no significant effect of the awareness feature on the number of questions answered incorrectly $(F_{1,8} = .647, p = .444)$. Most participants answered all questions correctly with both features. Four participants (P2, P4, P5, P12) presented incorrect answers while using VRBubble, and four participants (P4, P7, P10, P11) did so with the baseline. P2 was the only participant who answered two questions incorrectly, while everyone else answered one question incorrectly.

Moreover, we compared the effects of VRBubble and the baseline on participants' response time to questions in the conversation. With an ART model, we found no significant effect of Feature on response time ($F_{1,8} = .206, p = .662$). These results showed that VRBubble and baseline were generally at the same distraction level in the conversation task.



Figure 4: Likert scale score averages for each task: navigation Task (nav) and conversation task (conv).

4.2.3 Perceived Experience with VRBubble. We compared participants subjective experience with VRBubble and the baseline from three dimensions, including effectiveness, distraction, and immersion. Figure 4 shows the mean scores given by participants for each dimension. Participants perceived VRBubble (*mean* = 3.917, *SD* = .9) to be more effective than the baseline (*mean* = 2.667, *SD* = 1.231) at enhancing avatar awareness in the navigation task, while perceiving VRBubble (*mean* = 3.167, *SD* = 1.403) to be slightly less effective than the baseline (*mean* = 3.25, *SD* = 1.485) for the conversation task. These scores matched our expectations given that the conversation task was more attention-demanding and participants had less cognitive capacity to use VRBubble.

Participants perceived more immersion in the virtual space while using VRBubble (*mean* = 4.333, SD = .888) than the baseline (*mean* = 2.667, SD = 1.557) during the navigation task. They also felt more immersed during the conversation task when using VR-Bubble (*mean* = 3.583, SD = 1.443) than when using the baseline (*mean* = 3.083, SD = 1.505). Six participants found the audio beacon baseline to be annoying or distracting, and thus diminishing their immersive experience. In contrast, VRBubble provided a better sense of presence by dividing the social space and associating the avatar dynamics with social indications. As P3 mentioned, "*The bubble system was much more immersive, just because it gives you a couple of clear lines between when you're entering and leaving someone's space. The beacon is just super impersonal.*"

Interestingly, participants perceived VRBubble (*mean* = 1.75, SD = .754) to be less distracting than the baseline (*mean* = 2.583, SD = 1.379) for the navigation task, despite taking significantly more time to complete the task with VRBubble. This further confirmed that the slower navigation could be caused by participant's increased interest in exploring the surrounding with VRBubble instead of distraction. For the conversation task, participants perceived VR-Bubble (*mean* = 3.167, SD = 1.467) to be slightly more distracting than the baseline (*mean* = 3.083, SD = 1.443). Five participants expressed that the amount of information from VRBubble caused distractions during conversation, suggesting simpler or reduced audio was preferred for audio-intensive tasks.

4.2.4 Experience with VRBubble and its Audio Alternatives. All participants agreed that the Bubble concept based on social distances was novel and effective, and expressed favor that VRBubble provided them with more information than the baseline did. Seven participants liked that VRBubble separated the space and conveyed a sense of distance to the user. P3 noted, "I like the three different distances. If I walk into a social situation, I wish I had a bubble radar just to know when people are in my immediate vicinity versus people a little ways away. It is just nice to have a concept of relative [distance]." Participants also liked the sense of avatar presence and movement that was conveyed by the feedback from the bubbles. We then elaborate participants' experiences with different audio alternatives below.

Earcon. Four participants liked the brevity and non-intrusiveness of the earcons. Unlike the continuous audio feedback in the baseline, participants liked that the earcons discretely notified them every time an avatar entered a bubble. For example, P8 had difficulty with discerning spatial audio. She reported, *"I like how concrete [the earcon] is. I have a really hard time with spatial awareness. So, without necessarily having to keep track of the space because there are those different [earcons], it's very helpful."* Three participants noted that earcons for each avatar event could be easily noticed and distinguished, which could be used to reduce audio overloading in crowded environments. As P4 indicated, *"To signify the person is leaving, I'd want the [earcon] sound that shows the person's going out. Because then you don't have too many verbal cues, you know, to not listen too much."*

The main deterrent of earcons was the deep learning curve. Six participants found it hard to recognize different earcon sounds. P4 pointed out that the meanings of earcons would have to become second nature in order to be effective in a dynamic social environment: "You don't want to have to think about it, you want to immediately react." Five participants had issues with discerning the pitch difference between friend and stranger earcons. As P1 said, "there were more and more unexpected [earcons] that threw me off and, being not as familiar with the tones, second guessing if [the earcon] was higher [pitched] or lower, entering or exiting." P4 suggested that instead of using different pitches, completely different earcons should be used for different types of avatars.

Verbal Notification. Nine participants liked that the verbal notifications were informative and easy to understand. They expressed a desire for sufficient information about their surroundings to adjust their behaviors, especially in the context of conversation. P3 emphasized that distinguishing between friends and strangers can help protect his privacy, *"there may be something that you're willing to share with a friend, but not necessarily with a complete stranger."*

The drawback of verbal notifications was its verbosity, which caused more audio overlapping between avatars. Nine participants mentioned this problem in the study. Four participants also noted that hearing a verbal notification immediately drew their attention, which could be distracting in certain scenarios, such as conversation. Due to these issues, three participants cared only about verbal notifications for important people, such as friends, rather than being verbally informed about every avatar.

Real-world Sound Effect. Six participants liked the footstep and crowd noises since they were ambient and non-distracting. The audio had no sudden alerts and was quiet enough to tune out when focusing on the primary task. The crowd sounds were also immersive and helped convey a general sense of the social environment, such as if the user was at an informal party or a business meeting. For example, P4 preferred using real-world sound effects in the conversation task for an immersive experience. As he expressed, "I was not gearing [my customization] toward the best [avatar recognition] results but [was] gearing it toward getting a real feel for it." P3 even suggested providing "different types of ambient background noises that meet the theme of whatever rooms." Two participants (P3, P6) also felt that the two levels of ambient crowd sound with different loudness was sufficient, and any further levels may be too loud or unable to contribute any additional useful information. Moreover, two participants (P1, P2) liked the spatial footstep audio in the Conversation Bubble, which helped them constantly track important avatars in the conversational distance.

However, six participants disliked the constant real-world sound effects since it was hard to for them to distinguish each avatar via the general crowd sounds. P6 thus suggested combining the crowd sound with an overall description of the avatar number. Two participants (P7, P10) had difficulty distinguishing friend and stranger avatars via the different footstep sounds, especially with other louder audio in the environment.

4.2.5 Customization of Audio Alternatives. All participants appreciated the flexibility of customizing the audio feedback in VRBubble. For example, P8 had an auditory processing disorder that caused large quantities of sound to be difficult to process. As she mentioned, "[I like] the idea that you can customize how you're going to handle an event or have part of an event made accessible to you via sound. It is not something most people think about and I really appreciate it."

Participants customized the audio alternatives for different social tasks, avatar roles, and bubbles (Figure 7). Figure 5 shows the distribution of audio alternatives chosen for the navigation task, and Figure 6 shows the distribution for the conversation task. While no participants selected the exactly same combinations, we saw some patterns across participants that we report below.



Figure 5: VRBubble customization for navigation task: Sideby-side comparison between amount of times that each audio alternative was selected for friends and strangers, at each bubble. Intimate bubble is shared between friends and strangers.

Navigation vs. Conversation Task. In general, we found that verbal notification was most popular for the navigation task, while earcons and real-world sound effects were more popular in the



Figure 6: VRBubble customization for the conversation task: side-by-side comparison between amount of times audio alternative is selected for friends and strangers, at each bubble. Intimate bubble is shared between friends and strangers.

conversation task (except for Intimate Bubble), as shown between Figure 5 and Figure 6. While seven participants used the same combination for both navigation and conversation tasks, five participants changed their customization. We found that participants tended to simplify their audio combination in the conversation task, and all changes happened in Social Bubble. The changes included removing an audio alternative (P4, P11 removed verbal notification, P5 removed real-world sound effect for strangers, and P7 removed earcons for friends), and changing verbal notifications to more brief earcons (P3 for strangers, P11 for friends). This pattern suggested that users preferred shorter and non-verbal feedback during audiofocused tasks like conversation, whereas more informative feedback was preferred for exploration-based tasks like navigation.

Friends vs Strangers. Participants had divergent preferences when distinguishing friends and strangers. Some participants (P2, P5, P7, P8) chose more detailed information for friends and felt that stranger information was less important especially in the Social Bubble. P2 and P7 even did not want any audio feedback for strangers in the Social Bubble. On the other hand, some participants (P3, P11) wanted to know more about strangers: P3 selected earcons for friends and more informative verbal notification for strangers in the Social Bubble in the navigation task, while P11 added earcons for the strangers to alert himself. They explained that in a real use case, users were likely to already know enough information about strangers. As P3 described, *"The verbal would be useful because that might indicate to me it's a stranger I don't know. And this is their name.*"

Patterns across Bubbles. Most participants preferred the realworld sound effects for the Intimate Bubble, while some participants selected the abstract earcons. Participants mostly made the decision based on which audio feedback they found more pleasant. Interesting, P5 was the only participant who expressed interest in knowing who they bumped into, thus choosing verbal notification. This may suggest that in social VR, avatar collision was an event that PVI wanted to avoid but not important enough for PVI to care about the details.

	Navigation Task					Conversation Task				
ID	Social	Social Conversational Conversational		Intimato	Social	Social	Conversational	Conversational	Intimato	
	Friends	Strangers	Friends	Strangers	munate	Friends	Strangers	Friends	Strangers	minhate
1			0	Q	% 9	-4/1-		0	Q	9
2	\bigcirc	N/A	\bigcirc	\$ 9		\bigcirc	N/A	\bigcirc	%	-₩
3		Q	\bigcirc	Q				Q	Q	
4	\bigcirc					٩,	\$ 9	29	29	29
5			\bigcirc	\mathbf{Q}	\bigcirc		\bigcirc	Q	\bigcirc	Q
6			% <i>p</i>	29				٩,	29	٩.
7		N/A			29	N/A	N/A			.
8		٠.	\$ 9	\$ <i>9</i>	29		\$ 9	\$ 9	٩,	% 9
9	\bigcirc	\bigcirc			1 9	\bigcirc	\bigcirc			٩,
10	\bigcirc	\bigcirc	\bigcirc	\bigcirc	1 9	\bigcirc	\bigcirc	Q	\bigcirc	9
11	\Diamond		\$ 9		29			% 9	29	29
12		-41-			-44-					

🐠 = Earcons 💭 = Verbal 🍤 = Real-world 📃 = Changed Customization Between Tasks

Figure 7: Chart of participants' VRBubble customization in navigation and conversation tasks.

For the Social and Conversation Bubbles, we found that participants had different preferences for different bubbles without clear patterns, and only P10 used verbal notification for both bubbles.

Audio Combinations. The most popular audio combination was between verbal notifications and real-world sound effects, which was adopted by four participants (P4, P5, and P9 used the combination for the Social Bubble, P2 used it at the Conversation Bubble). Two participants (P7, P11) also combined earcons with verbal notifications. Participants used such combinations to add immersion to the detailed information provided by verbal notifications. As P5 described, "Because that will add more reality to what I'm doing, instead of just a boring computer thing." Only P12 used a combination of earcons and real-world footsteps. As he explained, "The abstract [earcon] is my favorite and that's why I chose most of the options here. But I thought it would be helpful to have the footsteps as well just so it's a little bit different [from just the earcon]."

5 DISCUSSION

With VRBubble, we contributed the first VR technique that leveraged the social distances (i.e., Bubbles) in Hall's proxemic theory [18] to enhance PVI's peripheral awareness of avatars in a complex and dynamic social VR context. VRBubble enabled user customization by providing three spatial audio alternatives—earcons, verbal notifications, and real-world sound effects—and allowing users to select and combine their preferred audio feedback for different bubbles, avatars, and social contexts.

Our evaluation with 12 PVI demonstrated the effectiveness of VRBubble: It significantly enhanced participants' awareness of the amount of avatars they passed by in a navigation task and enabled participants to identify most avatars, including their names and relationship to the users in both navigation and conversation tasks. However, we found that VRBubble could be distracting especially in crowded environments with high amount of avatars. Our research was the first step towards the general accessibility of the dynamic social VR. Based on our exploration, we discuss the takeaways, design implications, technology generalizability, and the lessons we learnt from our remote study.

5.1 Audio Preferences for Peripheral Awareness

Besides seeing VRBubble as a whole VR technique, our research also explored PVI's experiences and preferences of perceiving peripheral information via different audio modalities. Unlike prior work that studied the use of different audio feedback in various primary tasks (e.g., [1, 45, 87]), we focused on using audio to convey peripheral dynamics and investigated the feasibility of different audio modalities in two representative social contexts: navigation and conversation. Our study showed that compared to a fixed design (e.g., nonadjustable audio beacons), flexible audio customization allows users to effectively control distraction from a given primary task. In our study, five participants changed their audio selection after experiencing a different social task (Table 1), highlighting the need for customization depending on the social context.

Our research identified PVI's customization patterns. First, participants' audio preferences were driven by the cognitive load of the primary task. They preferred more detailed and descriptive audio modality (e.g., verbal notification) in less-attention-demanding tasks, but could only handle brief and distinct audio feedback (e.g., earcons, sound effects) in more attention-demanding and audiofocused tasks. Second, the importance of peripheral information was another factor that influenced PVI's audio selection. Specifically, the more detailed verbal descriptions or multiple audio alternative combinations were selected for more important avatars.

Our results also suggested (but need further investigation) that people may have a upper limit of capacity to perceive peripheral information in attention-demanding tasks (e.g., perceiving five avatars at most in a conversation task), so that they had to give up on less important peripheral information, such as information on strangers in the Social Bubble. While our research focused on the social VR context, the preferences of audio modalities by PVI to receive peripheral information could be applied to broader use cases, such as situational awareness systems for drivers. ASSETS '22, October 23-26, 2022, Athens, Greece

5.2 Design Implications

Based on participants' experiences with VRBubble, we distill design implications to inspire more accessible and immersive social VR experience for PVI.

Context-Adaptive Feedback. Our study indicated that participants preferred different audio modalities for different social contexts. However, social activity is a highly dynamic process where users' context and primary task can change quickly. For example, a user could be navigating, but stop to chat with a passing friend in the next moment. Thus, it could be difficult for PVI to manually adjust their audio combinations every time to match the fast-changing social contexts. Some participants suggested pre-defining multiple sets of audio combinations for different contexts and designing easy ways to toggle between them. As P9 suggested, "[You can have] a walking around mode, and a talking conversation mode. If you exit a conversation and you're walking around, switch back to names." To reduce the effort of active user control, researchers could consider context-adaptive feedback mechanism, which recognizes users' social context with AI technology and automatically switches to users' preferred feedback based on the current context.

Balance Automatic and Manually-queried Feedback. VR-Bubble focused on conveying peripheral information, thus providing proactive feedback that was automatically triggered by environmental changes to reduce users' interaction efforts. However, in our study, several participants indicated the desire for actively querying avatar information since they may have missed some prior information or need more detailed information about particular avatars of interest. We acknowledge that both types of feedback have their values. For example, our study showed that while proactive feedback were suitable for important and urgent information, too much feedback can distract and overwhelm the users. Meanwhile, manually-queried feedback can be a complement to provide detailed information on demand and reduce feedback overload. It is thus important to consider how to balance proactive and manually-queried feedback to optimize users' experience in social VR, especially in crowded virtual environments or cognitively heavy tasks. For example, proactive notifications can be used for important and nearby avatars, and manually-queried feedback can serve to provide additional information based on users' interest. Additionally, users should also have the option to manually repeat any information provided by proactive feedback.

Enable Third-Party Audio Design. While providing different audio alternatives, VRBubble provided only pre-defined audio effects, especially for earcons and real-world sound effects. Interestingly, our study indicated that participants had distinct hearing abilities, audio preferences, and audio aesthetic. For example, some participants (P2, P8) found the real-world footstep more pleasant to hear, while some (P4, P9) preferred the more abstract earcons. Even for the same audio modality, participants had different preferences. Some (P3, P6, P11, P12) liked the current earcon design, while some (P1, P4, P5, P9, P10) found them difficult to distinguish. To better fulfill the users' needs, one solution could be allowing users to design and upload their own audio source to represent particular avatars. However, designing and generating audio source could be technically challenging, especially for people with visual impairments. It is thus important to consider the trade-off between

the users' effort and satisfaction. Online communities for avatar sonification could also be encouraged to provide PVI a place to search and find their preferred audio source, similar to Thingiverse [69] which serves as an online community that provides 3D model resources for makers.

5.3 Generalizability

We discuss the generalizability of our technique from both the platform and the users' perspectives. While implemented and evaluated on a desktop-based VR platform, VRBubble can be easily transferred to HMD-based VR as it focuses on proactive audio design that is not device or platform specific. It can be adapted to other hardware through combination with suitable input techniques (e.g., input via controllers or body/head movements).

Beyond PVI, VRBubble could also be applied for sighted users in social VR. Various real-world presence cues, such as breathing or wind movement are not conveyed through visual information. Thus providing sighted users with additional peripheral information via audio feedback could enhance immersion in social VR, as well as ensuring users' privacy (i.e., being aware of potential out of sight eavesdroppers).

5.4 Evaluation via a Remote Study

We conducted our study remotely due to the restriction of the pandemic. We summarize our experience and the lessons we learnt from the remote study. First, we found it difficult for participants to setup all the software and hardware required by our study independently. To guide participants through the whole setup process, we connected with the participants 30 minutes ahead of the study via Zoom, sent them the url to our VR environment, and asked them to share screen with us during the setup process. Participants generally appreciated our assistance via Zoom and felt the use of Zoom sharing function was effective for the remote study. As P7 noted, "You've just proven a feature not in virtual reality right now—feature sharing. You just showed how important that is for someone with a disability because accessibility [functionality] doesn't always work."

Second, it is hard to guarantee the consistency of participants' experience in a remote study, depending on each individual's devices and environments. For example, they may receive different audio quality due to the different headphones, and some of them may experience feedback delay due to unstable Internet access. For this study, we tried to ensure a certain level of consistency by checking that participants had the pre-requisite equipment, including a keyboard, a pair of headphones that supports spatial audio, and a Microsoft Edge browser installed to access our VR environment. Participants shared their screen and audio via Zoom, so that we can check whether our feature was functioning properly in real time. However, the audio share feature in Zoom did not support spatial audio, which made it impossible to fully confirm users' experience. In the future, how to setup a remote study platform and enable researchers to easily control and confirm users' experience is a vital research direction to explore.

5.5 Limitations and Future Work

Our research has limitations. Since our study focused on the design and evaluation of the audio feedback (i.e., the output) in VRBubble,

we used a Wizard-of-Oz [8] setup to enable participants to customize the audio alternatives. Thus, how to enable PVI to efficiently and easily control and customize VRBubble's audio modalities (i.e., the input) still mains unaddressed. A future work of interest would be to explore effective input modalities to support fast interaction for PVI in social VR.

Additionally, our current evaluation used mock-up social VR environments with system-generated avatars, which may not fully reflect the avatar dynamics in the real social VR environments. For example, avatars in social VR may be conversing rather than standing silently, so that they might be located or identified by their voice. This can potentially change users' experience with VRBubble. Future research should evaluate VRBubble in more realistic social VR scenarios and adapt the design to different situations; for example, a context-aware technique that adjusts the audio feedback based on avatars' voice activity (e.g., muting or reducing the volume of audio notifications for a speaking avatar).

6 CONCLUSION

We presented VRBubble, a social VR technique to augment peripheral awareness by utilizing Hall's proxemic theory. VRBubble provided three audio alternatives—earcons, verbal notification, and real-world sound effects—for PVI to select and combine to maintain aware of different avatars at different social distances. We compared VRBubble to a standard audio beacon baseline via a user study with 12 participants with visual impairments. We assessed the effect of VRBubble on participants' ability to identify avatars as well as its distraction in different primary social tasks. Participants found VRBubble effective at providing previously inaccessible avatar information, and expressed the need for customizing audio feedback in different social VR contexts. Our study contributed a novel accessibility technique to inform the social VR dynamics and provided implications for the future design of accessible VR for people with visual impairments.

ACKNOWLEDGMENTS

We thank the National Federation of the Blind for helping us recruit for our study, as well as the anonymous participants that provided their perspective.

REFERENCES

- ASM Iftekhar Anam, Shahinur Alam, and Mohammed Yeasin. 2014. Expression: A dyadic conversation aid using Google Glass for people who are blind or visually impaired. In 6th International Conference on Mobile Computing, Applications and Services. IEEE, 57–64.
- [2] Ronny Andrade, Steven Baker, Jenny Waycott, and Frank Vetere. 2018. Echohouse: Exploring a virtual environment by using echolocation. In Proceedings of the 30th Australian Conference on Computer-Human Interaction. 278–289.
- [3] Matthew T Atkinson, Sabahattin Gucukoglu, Colin HC Machin, and Adrian E Lawrence. 2006. Making the mainstream accessible: redefining the game. In Proceedings of the 2006 ACM SIGGRAPH Symposium on Videogames. 21–28.
- [4] Stephen A Brewster, Peter C Wright, and Alistair DN Edwards. 1993. An evaluation of earcons for use in auditory human-computer interfaces. In Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems. 222-227.
- [5] Hendrik P Buimer, Marian Bittner, Tjerk Kostelijk, Thea M Van Der Geest, Abdellatif Nemri, Richard JA Van Wezel, and Yan Zhao. 2018. Conveying facial expressions to blind and visually impaired persons through a wearable vibrotactile device. *PloS one* 13, 3 (2018), e0194737.
- [6] J.J. Cadiz, Gina Venolia, Gavin Jancke, and Anoop Gupta. 2002. Designing and deploying an information awareness interface. CSCW 02 (November 2002). https://doi.org/10.1145/587078.587122

- [7] Chetz Colwell, Helen Petrie, Diana Kornbrot, Andrew Hardwick, and Stephen Furner. 1998. Haptic virtual reality for blind computer users. ASSETS 98 (January 1998). https://doi.org/10.1145/274497.274515
- [8] Nils Dahlback, Arne Jonsson, and Lars Ahrenberg. 1993. WIZARD OF OZ STUD-IES – WHY AND HOW. Intelligent User Interfaces (1993).
- [9] Maitraye Das, Thomas Barlow McHugh, Anne Marie Piper, and Darren Gergle. 2022. Co11ab: Augmenting Accessibility in Synchronous Collaborative Writing for People with Vision Impairments. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 196, 18 pages. https://doi.org/10.1145/3491102.3501918
- [10] Patricia A. de Oliveira, Erich P. Lotto, Ana Grasielle D. Correa, Luis G. G. Taboada, Laisa C. P. Costa, and Roseli D. Lopes. 2015. Virtual Stage: An Immersive Musical Game for People with Visual Impairment. In 2015 14th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames). 135–141. https://doi.org/ 10.1109/SBGames.2015.26
- [11] Euan Freeman, Graham Wilson, Stephen Brewster, Gabriel Baud-Bovy, Charlotte Magnusson, and Hector Caltenco. 2017. Audible Beacons and Wearables in Schools: Helping Young Visually Impaired Children Play and Move Independently. *CHI* 17 (January 2017). https://doi.org/10.1145/3025453.3025518
- [12] Lakshmi Gade, Sreekar Krishna, and Sethuraman Panchanathan. 2009. Person localization using a wearable camera towards enhancing social interactions for individuals with visual impairment. In Proceedings of the 1st ACM SIGMM international workshop on Media studies and implementations that help improving access to disabled users. 53–62.
- [13] Evil Dog Games. 2014. Blind Swordsman. https://devpost.com/software/blindswordsman last accessed 18 Jan 2014.
- [14] Ellen R Girden. 1992. ANOVA: Repeated measures. Number 84. Sage.
- [15] Cole Gleason, Alexander J. Fiannaca, Melanie Kneisel, Edward Cutrell, and Meredith Ringel Morris. 2018. FootNotes: Geo-Referenced Audio Annotations for Nonvisual Exploration. 2, 3, Article 109 (sep 2018), 24 pages. https: //doi.org/10.1145/3264919
- [16] Glitch. 2000. Glitch. https://glitch.com last accessed 12 January 2022.
- [17] Jens Grubert, Eyal Ofek, Michel Pahud, and Per Ola Kristensson. 2018. The office of the future: Virtual, portable, and global. *IEEE computer graphics and applications* 38, 6 (2018), 125–133.
- [18] Edward T Hall. 1966. The Hidden Dimension. Doubleday.
- [19] Wilko Heuten, Daniel Wichmann, and Susanne Boll. 2006. Interactive 3D sonification for the exploration of city maps. Nordic conference on Human-computer interaction: changing roles (2006), 155–164.
- [20] Simon Holland, David Morse, and Henrik Gedenryd. 2002. AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. *Personal and Ubiquitous Computing* 6 (09 2002). https://doi.org/10.1007/s007790200025
- [21] Seyedeh Maryam Fakhr Hosseini, Andreas Riener, Rahul Bose, and Myounghoon Jeon. 2014. "Listen2dRoom": Helping visually impaired people navigate indoor environments using an ultrasonic sensor-based orientation aid.
- [22] Rec Room Inc. 2016. Rec Room. Oculus Quest, Microsoft Windows, PlayStation, iOS, Android, Xbox.
- [23] Toru Ishikawa, Hiromichi Fujiwara, Osamu Imai, and Atsuyuki Okabe. 2008. Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *Journal of environmental psychology* 28, 1 (2008), 74–82.
- [24] Gunnar Jansson. 1999. Can a Haptic Display Rendering of Virtual Three-Dimensional Objects be useful for People with Visual Impairments? 93, 7 (July 1999). https://doi.org/10.1177/0145482X9909300707
- [25] Gunnar Jansson, Helen Petrie, Chetz Colwell, Diana Kornbrot, J Fänger, H König, Katarina Billberger, Andrew Hardwick, and Stephen Furner. 1999. Haptic virtual environments for blind people: Exploratory experiments with two devices. *International journal of virtual reality* 4, 1 (1999), 8–17.
- [26] Tiger F. Ji, Brianna R Cochran, and Yuhang Zhao. 2022. Demonstration of VRBubble: Enhancing Peripheral Avatar Awareness for People with Visual Impairments in Social Virtual Reality (CHIEA '22). Association for Computing Machinery, New York, NY, USA, Article 401, 6 pages. https://doi.org/10.1145/3491101.3519657
- [27] Ralf Jung. 2008. Smart sound environments: merging intentional soundscapes, nonspeech audio cues and ambient intelligence. *The Journal of the Acoustical Society of America* 123, 5 (2008), 3935–3935. https://doi.org/10.1121/1.2936002 arXiv:https://doi.org/10.1121/1.2936002
- [28] Matthew Kay, Lisa A. Elkin, James J. Higgins, and Jacob O. Wobbrock. 2021. ARTool: Aligned Rank Transform for Nonparametric Factorial ANOVAs. https: //doi.org/10.5281/zenodo.594511 R package version 0.11.1.
- [29] Fredrik Kilander and Pekka Loennqvist. 2002. A whisper in the woods an ambient soundscape for peripheral awareness of remote processes.
- [30] Julian Kreimeier and Timo Gotzelmann. 2019. First Steps Towards Walk-In-Place Locomotion and Haptic Feedback in Virtual Reality for Visually Impaired. CHI EA 19 (May 2019). https://doi.org/10.1145/3290607.3312944
- [31] Sreekar Krishna, Shantanu Bala, Troy McDaniel, Stephen McGuire, and Sethuraman Panchanathan. 2010. VibroGlove: an assistive technology aid for conveying facial expressions. In CHI'10 Extended Abstracts on Human Factors in Computing

ASSETS '22, October 23-26, 2022, Athens, Greece

Systems. 3637-3642.

- [32] Sreekar Krishna, Dirk Colbry, John Black, Vineeth Balasubramanian, and Sethuraman Panchanathan. 2008. A systematic requirements analysis and development of an assistive device to enhance the social interaction of people who are blind or visually impaired. In Workshop on Computer Vision Applications for the Visually Impaired.
- [33] Orly Lahav and David Mioduser. 2004. Exploration of unknown spaces by people who are blind using a multi-sensory virtual environment. *Journal of Special Education Technology* 19, 3 (2004), 15–23.
- [34] Cheuk Yin Phipson Lee, Zhuohao Zhang, Jaylin Herskovitz, JooYoung Seo, and Anhong Guo. 2022. CollabAlly: Accessible Collaboration Awareness in Document Editing. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, Article 596, 17 pages. https://doi.org/10.1145/3491102. 3517635
- [35] Tapio Lokki and Matti Grohn. 2005. Navigation with auditory cues in a virtual environment. *IEEE MultiMedia* 12, 2 (2005), 80–86.
- [36] 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 (2013), e72555.
- [37] Masaki Matsuo, Takahiro Miura, Masatsugu Sakajiri, Junji Onishi, and Tsukasa Ono. 2016. Audible Mapper & ShadowRine: Development of Map Editor Using only Sound in Accessible Game for Blind Users, and Accessible Action RPG for Visually Impaired Gamers. In International Conference on Computers Helping People with Special Needs. Springer, 537–544.
- [38] Tara Matthews and Jennifer Mankoff. 2005. A toolkit for evaluating peripheral awareness displays. In From the Awareness Systems Workshop at CHI.
- [39] Keenan R May, Brianna J Tomlinson, Xiaomeng Ma, Phillip Roberts, and Bruce N Walker. 2020. Spotlights and soundscapes: On the design of mixed reality auditory environments for persons with visual impairment. ACM Transactions on Accessible Computing (TACCESS) 13, 2 (2020), 1–47.
- [40] Troy McDaniel, Sreekar Krishna, Vineeth Balasubramanian, Dirk Colbry, and Sethuraman Panchanathan. 2008. Using a haptic belt to convey non-verbal communication cues during social interactions to individuals who are blind. In 2008 IEEE international workshop on haptic audio visual environments and games. IEEE, 13–18.
- [41] Joshua McVeigh-Schultz, Anya Kolesnichenko, and Katherine Isbister. 2019. Shaping pro-social interaction in VR: an emerging design framework. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 1–12.
- [42] Lotfi Merabet and Jaime Sanchez. 2009. Audio-based navigation using virtual environments: combining technology and neuroscience. AER Journal: Research and Practice in Visual Impairment and Blindness 2, 3 (2009), 128–137.
- [43] Microsoft. 2013. Altspace. Oculus Quest, Microsoft Windows.
- [44] Microsoft. 2022. Microsoft Soundscape. Microsoft Windows.
- [45] Cecily Morrison, Ed Cutrell, Martin Grayson, Geert Roumen, Rita Faia Marques, Anja Thieme, Alex Taylor, and Abigail Sellen. 2021. PeopleLens. *Interactions* 28, 3 (2021), 10–13.
- [46] Konstantinos Moustakas, Georgios Nikolakis, Konstantinos Kostopoulos, Dimitrious Tzovaras, and Michael G. Strintzis. 2007. Haptic Rendering of Visual Data for the Visually Impaired. *IEEE* (January 2007). https://doi.org/10.1145/274497. 274515
- [47] Robby Nadler. 2020. Understanding "Zoom fatigue": Theorizing spatial dynamics as third skins in computer-mediated communication. *Computers and Composition* 58 (October 2020). https://doi.org/10.1016/j.compcom.2020.102613
- [48] Gerhard Nahler. 2009. Latin square. Springer Vienna, Vienna, 105–105. https: //doi.org/10.1007/978-3-211-89836-9_776
- [49] Vishnu Nair, Jay L Karp, Samuel Silverman, Mohar Kalra, Hollis Lehv, Faizan Jamil, and Brian A Smith. 2021. NavStick: Making Video Games Blind-Accessible via the Ability to Look Around. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 538–551.
- [50] Research Nester. 2021. Virtual Collaboration Market Segmentation by Tools (Instant Communication, Project Management Tools, Cloud-Based Tools, and Others); and by End-User (IT & Telecom, BFSI, Retail, Healthcare, Logistics & Transportation, Education, Manufacturing, and Others) – Global Demand Analysis and Opportunity Outlook 2021-2029. https://www.researchnester.com/ reports/virtual-collaboration-market/2994 last accessed 10 January 2022.
- [51] Donald A. Norman. 1986. User Centered System Design; New Perspectives on Human-Computer Interaction. L. Erlbaum Associates Inc.
- [52] Bugra Oktay and Eelke Folmer. 2010. TextSL: a screen reader accessible interface for second life. W4A 10 (April 2010). https://doi.org/10.1145/1805986.1806017
- [53] World Health Organization. 2021. Blindness and vision impairment. https://www.who.int/news-room/fact-sheets/detail/blindness-and-visualimpairment#:~:text=Key%20facts,has%20yet%20to%20be%20addressed. last accessed 26 Jan 2022.

- [54] Andrew Osterland. 2021. Virtual Reality Headset Market Size, Share & Trends Analysis Report By End-device (Low-end, High-end), By Product Type (Standalone, Smartphone-enabled), By Application (Gaming, Education), And Segments Forecasts, 2021 - 2028. https://www.grandviewresearch.com/industryanalysis/virtual-reality-vr-headset-market last accessed 26 Jan 2022.
- [55] Jamie Pauls. 2020. Vintage Games Series, Part 4: Immerse Yourself in the World of Shades of Doom. https://www.afb.org/aw/21/12/17336 last accessed 18 Jan 2022.
- [56] Elin R. Pedersen and Tomas Sokoler. 1997. AROMA: abstract representation of presence supporting mutual awareness. CHI 97 (March 1997). https://doi.org/10. 1145/258549.258584
- [57] Aaron Preece. 2018. A Review of A Hero's Call, an Accessible Role Playing Game from Out of Sight Games. https://www.afb.org/aw/19/3/15113 last accessed 18 Jan 2022.
- [58] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. 2019. WatchOut: Obstacle Sonification for People with Visual Impairment or Blindness. In The 21st International ACM SIGACCESS Conference on Computers and Accessibility (Pittsburgh, PA, USA) (ASSETS '19). Association for Computing Machinery, New York, NY, USA, 402–413. https://doi.org/10.1145/3308561.3353779
- [59] Radegast Project. 2009. Radegast. Microsoft Windows, Mac, Linux.
- [60] Shi Qiu, Matthias Rauterberg, and Jun Hu. 2016. Designing and evaluating a wearable device for accessing gaze signals from the sighted. In *International Conference on Universal Access in Human-Computer Interaction*. Springer, 454– 464.
- [61] Mose Sakashita, E. Andy Ricci, Jatin Arora, and François Guimbretière. 2022. RemoteCoDe: Robotic Embodiment for Enhancing Peripheral Awareness in Remote Collaboration Tasks. Proc. ACM Hum.-Comput. Interact. 6, CSCW1, Article 63 (apr 2022), 22 pages. https://doi.org/10.1145/3512910
- [62] Johnny Saldaña. 2021. The coding manual for qualitative researchers. sage.
- [63] Jaime Sanchez and Mauricio Lumbreras. 1997. Hyperstories: Interactive narrative in virtual worlds. Hypertextes et hypermédias (1997), 329–338.
- [64] Jaime Sánchez and Mauricio Lumbreras. 1999. Virtual environment interaction through 3D audio by blind children. CyberPsychology & Behavior 2, 2 (1999), 101–111.
- [65] JH Sánchez and MA Sáenz. 2006. Assisting the mobilization through subway networks by users with visual disabilities. Virtual Reality & Assoc. Tech (2006).
- [66] Daisuke Sato, Uran Oh, João Guerreiro, Dragan Ahmetovic, Kakuya Naito, Hironobu Takagi, Kris M. Kitani, and Chieko Asakawa. 2019. NavCog3 in the Wild: Large-Scale Blind Indoor Navigation Assistant with Semantic Features. ACM Trans. Access. Comput. 12, 3, Article 14 (aug 2019), 30 pages. https://doi.org/10.1145/3340319
- [67] David W Schloerb, Orly Lahav, Joseph G Desloge, and Mandayam A Srinivasan. 2010. BlindAid: Virtual environment system for self-reliant trip planning and orientation and mobility training. In 2010 IEEE Haptics Symposium. IEEE, 363– 370.
- [68] Alexa F Siu, Mike Sinclair, Robert Kovacs, Eyal Ofek, Christian Holz, and Edward Cutrell. 2020. Virtual reality without vision: A haptic and auditory white cane to navigate complex virtual worlds. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
- [69] Zach Smith and Bre Pettis. 2008. Thingiverse. https://www.thingiverse.com/ last accessed 12 April 2022.
- [70] Statista. 2019. Leading barriers to mass adoption of VR according to XR professionals worldwide as of the 3rd quarter of 2019. https://www.statista.com/ statistics/1099109/barriers-to-mass-consumer-adoption-of-vr/ last accessed 7 July 2015.
- [71] Supermedium. 2015. A-Frame. Cross-platform.
- [72] Manohar Swaminathan, Sujeath Pareddy, Tanuja S. Sawant, and Shubi Agarwal. 2018. Video Gaming for the Vision Impaired. ASSETS 18 (October 2018). https: //doi.org/10.1145/3234695.3241025
- [73] Juan R Terven, Joaquin Salas, and Bogdan Raducanu. 2014. Robust head gestures recognition for assistive technology. In *Mexican Conference on Pattern Recognition*. Springer, 152–161.
- [74] Shari Trewin, Vicki L Hanson, Mark R Laff, and Anna Cavender. 2008. PowerUp: an accessible virtual world. In Proceedings of the 10th international ACM SIGAC-CESS conference on Computers and accessibility. 177–184.
- [75] Dimitrios Tzovaras, Konstantinos Moustakas, Georgios Nikolakis, and Michael G Strintzis. 2009. Interactive mixed reality white cane simulation for the training of the blind and the visually impaired. *Personal and Ubiquitous Computing* 13, 1 (2009), 51–58.
- [76] Dimitrios Tzovaras, Georgios Nikolakis, George Fergadis, Stratos Malasiotis, and Modestos Stavrakis. 2002. Design and implementation of virtual environments training of the visually impaired. In Proceedings of the fifth international ACM conference on Assistive technologies. 41–48.
- [77] Pablo Vera, Daniel Zenteno, and Joaquín Salas. 2014. A Smartphone-Based Virtual White Cane. Pattern Anal. Appl. 17, 3 (aug 2014), 623–632. https://doi.org/10. 1007/s10044-013-0328-8
- [78] Inc VRChat. 2017. VRChat. Oculus Quest, Microsoft Windows.

- [79] Bruce N. Walker and Jeffrey Lindsay. 2006. Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice. *Human Factors* 48, 2 (2006), 265–278.
- [80] Dean A. Waters and Husam H. Abulula. 2001. The virtual bat: Echolocation in virtual reality.
- [81] Ryan Wedoff, Lindsay Ball, Amelia Wang, Yi Xuan Khoo, Lauren Lieberman, and Kyle Rector. 2019. Virtual showdown: An accessible virtual reality game with scaffolds for youth with visual impairments. In Proceedings of the 2019 CHI conference on human factors in computing systems. 1–15.
- [82] Thomas Westin. 2004. Game accessibility case study: Terraformers-a real-time 3D graphic game. In Proceedings of the 5th International Conference on Disability, Virtual Reality and Associated Technologies, ICDVRAT.
- [83] Matt Wilkerson, Amanda Koenig, and James Daniel. 2010. Does a Sonar System Make a Blind Maze Navigation Computer Game More "Fun"?. In Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility (Orlando, Florida, USA) (ASSETS '10). Association for Computing Machinery, New York, NY, USA, 309–310. https://doi.org/10.1145/1878803.1878886
- [84] Jeff Wilson, Bruce N. Walker, Jeffrey Lindsay, Craig Cambias, and Frank Dellaert. 2007. SWAN: System for Wearable Audio Navigation. In 2007 11th IEEE International Symposium on Wearable Computers. 91–98. https://doi.org/10.1109/ISWC. 2007.4373786
- [85] Yuhang Zhao, Cynthia L Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. 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. 1–14.
- [86] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D Wilson. 2019. Seeingvr: A set of tools to make virtual reality more accessible to people with low vision. In Proceedings of the 2019 CHI conference on human factors in computing systems. 1–14.
- [87] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A face recognition application for people with visual impairments: Understanding use beyond the lab. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 1–14.