Enable Blind Users' Experience in 3D Virtual Environments: The Scene Weaver Prototype

Harshadha Balasubramanian
Microsoft Research, United Kingdom

Cecily Morrison
Microsoft Research, United Kingdom

Martin Grayson
Microsoft Research, United Kingdom

Zhanat Makhataeva
Microsoft Research, United Kingdom

Rita Marques
Microsoft Research, United Kingdom

Thomas Gable
Microsoft Research, United Kingdom

Dalya Perez
Microsoft Research, United Kingdom

Edward Cutrell
Microsoft Research, United Kingdom

ABSTRACT

Three-dimensional virtual environments are currently inaccessible to people who are blind, as current screen-reading solutions for 2D content are not fully extensible to achieve the needed embodied spatial presence. Forefronting perceptual agency as key to any access approach for users who are blind, we offer Scene Weaving as an interactional metaphor that allows users to choose how and when they perceive the environment and the people in it. We illustrate how this metaphor can be implemented in an example prototype system. In this interactivity, users can control how and when they perceive a virtual museum environment and people within it through a range of interaction mechanisms.

CCS CONCEPTS

• **Human-centered computing**; • **Human computer interaction (HCI)**; • **HCI design and evaluation methods**;

KEYWORDS

Virtual Environments, Virtual Reality, Accessibility, Perceptual Agency, Interaction Metaphor

ACM Reference Format:

Harshadha Balasubramanian, Cecily Morrison, Martin Grayson, Zhanat Makhataeva, Rita Marques, Thomas Gable, Dalya Perez, and Edward Cutrell. 2023. Enable Blind Users' Experience in 3D Virtual Environments: The Scene Weaver Prototype. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3544549.3583909

1 INTRODUCTION

Three-dimensional (3D) virtual environments are currently inaccessible to people who are blind, leaving them excluded from the expansion of 3D digital content into the workplace, education, entertainment, and other contexts [3]. Existing screen-reading solutions for two-dimensional (2D) content are not readily extensible to virtual environments as they lack mechanisms to support the embodied spatial presence critical to experiencing 3D digital content, such as in virtual reality (VR) [11]. While research has explored how increasing the range of perceptual opportunity can provide non-visual mechanisms for embodied experiences, such as spatialized sound (e.g. [1]) and haptics (e.g. [2]), less attention has been paid to the important role of a user's perceptual agency to integrate and make sense of embodied perceptions into a coherent understanding of a scene.

We take the perspective that in a 3D virtual environment a scene is a product of a user's sense-making [9], enacted through a user's perceptual agency in attending to cues in the environment as relevant to their situated context [16]. A user may choose to attend to different cues if they are exploring a space on their own as opposed to socially negotiating that experience with others. They may attend to fewer cues if they are completing a task rather than exploring a space. Sense-making strategies also vary significantly across the blind community, often dependent on past experiences and sensory capabilities [13]. As such, there is no singular representation of a scene shared by all users that assistive systems should aim to describe or reveal.

We argue that access to 3D virtual environments for blind individuals should enable perceptual agency, facilitating a user's agentive role in the dynamic creation of the virtual environment scenes they experience. To this end, we propose scene weaving as an interactional metaphor to sensitize and guide designers and engineers in implementing accessible experiences in the growing number of virtual environments. In weaving, specific threads are chosen and interlaced to form a fabric. In scene weaving, the individual threads are the perceptual cues that are available and are chosen by the user and woven together through the sense-making process. Whether those threads are baked into the virtual environment, like spatialized audio cues, or retrievable through an assistive system, users should be able to select and connect threads together to make sense of their surroundings.

In this interactivity, we share a prototype based on the scene weaving interactional metaphor that puts users and their agency at the center. Taking a virtual museum as the scenario, the system allows blind individuals to choose how and when they perceive the environment and people in it with a range of interaction mechanisms.

CHI EA '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9422-2/23/04.
https://doi.org/10.1145/3544549.3583909

2 RELATED WORK

Screen readers enable blind individuals to access and interact with digital content [6]. They render the textual elements on a screen into spoken audio that are navigable with keyboard shortcuts and/or screen gestures. As they provide information curated into a linear stream, relationships between objects and places are difficult to ascertain quickly or with cognitive ease [10]. While these challenges are significant in common computing experiences such as Web browsing [6], they are amplified in 3D virtual environments because of the sheer volume and organization of information that may be available to users. Despite efforts to extend screen readers to meet the interactive spatiality of 3D virtual environments [15, 19], they remain limited compared with the active use of the body to curate spatial information [8].

Spatial audio has been extensively used to enhance the understanding of spatial relationships and localization of objects. Techniques such as object sonification and audio beacons can provide critical understanding and navigational support of the objects and space in virtual environments [4, 14]. Researchers and designers have also explored the use of haptic feedback and special controllers [5, 7], such as the wearable "Canetroller" that simulates white cane interaction in virtual reality [12, 20]. Combining haptic and audio rendering provides multimodal sensory experiences that enhance immersion, ease cognitive load, and extend spatial understanding.

3 SCENE WEAVER PROTOTYPE

3.1 Scene Weaving Concept

Scene weaving is an interaction metaphor that makes more relatable an epistemological stance of perceptual agency in how individuals who are blind access 3D virtual environments. We have posited that virtual environments are not a singular entity to be described, but that individuals prioritize and integrate perceptual cues in the environment to sense-make a 'scene' that is specific to their individual interests, abilities, and current goals. Similar to the way specific threads are chosen and interlaced to weave a fabric, individuals weave together perceptual cues that are available in the environment to create a scene. The woven fabric, or scene, demands a distinct identification from the virtual environment which contributed to it but does not circumscribe it.

Scene weaving speaks to the processual, creative, and not least embodied nature of sense-making. This proposed interactional metaphor stands in contrast to description of an environment, whether summarized for audio descriptions (e.g., [18]) or approaches inspired by screen readers with pre-defined piecemeal read-outs (e.g., [17]). The loom on which fabric is woven is like the design of the assistive technology, providing a framework for how threads can be woven and perceptual agency that can thereby be achieved. Just as the loom plays a key role in determining the eventual fabric, the design of the assistive technology has a key role in what kind of scene is evoked.

3.2 Scene Weaver Design Elements

In this section, we describe the design elements of a prototype that embodies the scene weaving concept. It has been implemented on a laptop keyboard with mouse or trackpad, and stereo headphones as well as on an Occulus to consider the range of devices that users might have access to. We also implemented controllers using the Xbox adaptive controller to expand possibilities for multimodal and spatial engagement. We selected a virtual museum scenario for the prototype, because this permits us to design for the widest range of exploratory and social interactions that users are likely to undertake.

3.2.1 Environmental Spatial Audio. Environmental spatial audio is provided in a range of ways as a user moves through the environment. Users create audio as they move or turn in the environment through the sonification of their footsteps. The sounds of footsteps and rotations echo in the surroundings, alerting users to how their body is situated with every step, very much like in everyday interactions. Environmental elements are also sounded when the user walks into, or past, them. Spatialized pings, for example, signify objects that users can interact with while thumps indicate the user has walked into a wall and claps indicate the presence of people. Users can also teleport to an object accompanied by a length of sound that represents the distance travelled and spatialized announcements of passing objects, situating the user in relationship to the environment. Such environmental spatial audio is often already in a virtual environment [3], however, the control of such sounds is limited to moving one's body in space. We explore other opportunities to get an overview/preview of a space using audio in the coming sections.

3.2.2 Triggered Spatial Audio. Triggered spatial audio gives the user great perceptual agency over what in the environment is sounded. A gaze mode allows users to hear the names of objects that are directly ahead as they move within the space. A left-right sweep on the trackpad allows the user to directly control a spatialized presentation of the non-verbal environmental audio cues (mentioned above) as well as quickly check what is immediately on either side when standing still. A radial sweep, performed by moving a finger on a track pad up from the bottom, reads out the names of objects, people, and architectural features in other rooms in 2-meter increments. These features resemble a cane sweep, encouraging blind users to draw on a familiar practice to confirm or add to their understanding of where they are in relation to specific aspects of the environment as a speed that they control. All the items can be toggled on or controlled through gestures so that the user can reach their chosen overview of the scene before them.

3.2.3 The Navigator. The navigator is a tool that provides efficient access to information that users are seeking: People, Places, and Things. These categories were drawn from audio description practices [18], which have developed structures for providing information that blind individuals are accustomed to. When toggled on, users can select a category and cycle through items in the category. The user maintains their orientation in space, hearing the information spatialized; the camera pans back to just behind the avatar and rotates towards the item in focus, attuning sighted companions to what the blind individual is focusing on to enable a shared experience. For example, the user might select the Things category, and then tap down to get to the arcade machine, hearing it spatialized off to the left.

Users can also find out information that falls into multiple categories with the navigator. With a particular place, person or thing focused, users can select a secondary category. For example, if they choose people, users will hear who is also in this place, who is nearby the object, or how people are interacting with others respectively. Beyond the 'people, places, things' structure, the specific information elements included will differ depending on the virtual environment. The design goal is to allow users to drive the curation of specific information and make fast choices. For example, someone can assess whether it's worth going to a room based on who and what is there without first navigating there. The process is fast because only the details relevant and chosen by the user are heard, in contrast to "scene descriptions."

3.2.4 Interaction Knot. The interaction knot supports users' participation in dynamic, social situations through setting up tailored notifications which alert users to spontaneous cues that other people may generate. Like a knot which surfaces multiple threads simultaneously, this feature prompts users about various cues as they arise, meaning they can save the time and cognitive load that it would take to search for this information. An Interaction Knot permits the rhythm of scene weaving to continue uninterrupted by sudden change. The notifications are preset by users, allowing them to adjust what they want to know according to the context. Keeping users aware of when there is someone close to them and available for interaction gives them agency in initiating interaction or indicating acceptance of interaction (or not). Users can also choose to be told about people who enter the user's interaction space or raise their hand. Similar to knots securing and shaping a woven fabric, the notifications that users preset and respond to while interacting will affect how they make sense of their virtual environment and negotiate building a scene. A tap on the keyboard will orientate the user to the source of the last notification. For maximum flexibility, notifications can be repeated, muted, and turned off using keyboard shortcuts.

Visually orientating towards a person signals the intent to interact, and the reorientation ensures users' spatial representations of the world can be amended in line with where they are positioned in the current interaction. The visual and auditory reorientation creates common ground for blind users and other visitors in the environment to negotiate how they can share the experience. The design of the interaction knot attempts to address the differences in time-sensitivity and cognitive attention between social and environmental interaction.

3.3 Interactive Prototype Walkthrough

Imagine you seek out opportunities to discover vintage tech that takes you back to important memories, like your first phone, the arcade games you used to play with your friends, or the computer in your first job. You have just found out about a virtual exhibition called "The History of Work", and you realize that it may be the best chance to reencounter some iconic pieces that are physically extremely hard to get hold of nowadays. Also, unlike in a physical exhibition, you will not need a person to guide you around who may influence your experience; you can explore the space and objects in your own time, in your own way.

As you enter the virtual museum, you hear a quiet hiss that lets you know you are in the museum. To get an idea of the exhibition space, you tap your foot by hitting X. There are quite a lot of echoes where you are, so this might be a large or long room. You slowly turn left and then right using the arrow keys and there is much less echo. You decide that you are in a long room using these environmentally spatialized audio sounds. Next you decide to find out what is in the room. You move your index finger slowly from left to right on the track pad to trigger the spatial audio of the objects around you and then check what objects are further away by moving your finger up on the track pad: "Old mobile phone; coffee machine; water cooler; IBM PC; laser printer." Each tick tells you that the radius has expanded by two meters.

You begin walking through the room and toggle gaze mode on to hear what you are passing: "projector; Jack; vending machine; projector." You turn your gaze mode off and turn your attention to the footsteps, you are not alone. What other rooms are there, you wonder? You toggle the navigator on and tap to the places categories: "main hall; film room; modern office desk." The spatialized audio makes it sound like you are close to the film room – ahh, you must have wandered in. You keep tapping and you hear, "retro desk." That sounds interesting. You tap the 'I' key to see what interactive objects there are: two, it seems. You then tap 'P' and learn that Jack is in this room. You decide to join Jack and teleport to the retro desk room. The sound of the teleport is quite long, so it must be far from the film room.

You have finished with the retro office desk room and are wondering what to do next. You will give a talk later about the arcade, maybe you could have a short play to refresh your memory. You open the navigator, tap to 'Things' and listen for 'Arcade.' You tap 'P' and hear that, "Nobody is interacting with the arcade machine." When you arrive at the arcade, you hear a ping, telling you that you can interact with it and start to play. Two consecutive doorbell sounds signal that two people have arrived in the exhibition. With the navigator activated and focused on 'People,' you learn they are Richard and Jordan. You recognize Richard's voice—do you know him? Focusing on Richard, you hear their self-description: "Richard. Curator of the museum and collector of objects and stories about the history of office work." You tap 'P' to hear who they are with: "Talking to Jordan."

You expect people to interact with you soon, so you activate the interaction knot. You have requested to be notified when people approach and when they raise their hand during your talk. Soon after you activate the interaction knot, you hear a notification sound: You press 'N' to find out more: "Aliana approaching just now." You hear two more notifications—let's check those. "Jack approaching just now. One more notification." You tap to hear the next notification: "Taylor just arrived." You start your presentation with the arcade theme tune. When it ends, there is a notification alert and so you press N to check what this is. "Aliana raised hand just now." You can press enter to accept, and this orientates you to look at Aliana, letting her know you are paying attention. Aliana: says "Hi! Do I know that theme tune?" You reply: "Yes you do—it's Pac Man, of course!"

4 DISCUSSION

This paper presents a prototype that embodies the interactional metaphor of scene weaving to guide blind users' access to the burgeoning world of 3D virtual content and environments. This metaphor aims to make more tangible an epistemological stance of perceptual agency for users that is grounded in the idea that there is no singular representation of a scene shared by all users that assistive systems should aim to describe or reveal. The rich variety of content and experiences offered in 3D virtual environments allows sighted users the ability to pull information from a wide range of affordances and spatial relationships to make sense of the space, objects, and people in the environment. We argue that such perceptual agency must also be granted by the assistive systems offered to blind and low vision users. In contrast to curated information presentation such as audio description or serialized screen-readers, experiences in virtual environments should facilitate a user's direct agency in the dynamic creation and understanding of the scenes that they experience.

REFERENCES

[1] Ronny Andrade, Steven Baker, Jenny Waycott, and Frank Vetere. 2018. Echohouse: Exploring a virtual environment by using echolocation. In OzCHI '18: Proceedings of the 30th Australian Conference on Computer-Human Interaction, ACM, 278–289.

[2] Cecily Morrison, Edward Cutrell, Martin Grayson, Anja Thieme, Alex S. Taylor, Geert Roumen, Camilla Longden, Rita Marques, Abigail Sellen, and Sebastian Tschiatschek. 2021. Social Sensemaking with AI: Designing an Open-ended AI experience with a Blind Child. In The 2021 CHI Proceedings on Human Factors in Computing Systems, 1–12.

[3] Eelke Folmer, Bei Yuan, Dave Carr, and Manjari Sapre. 2009. TextSL: A command-based virtual world interface for the visually impaired. In ASSETS'09 - Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility, ACM, 59–66.

[4] José Luis González-Mora, A. Rodriguez-Hernandez, Enrique Burunat, Martin F., and Miguel A. Castellano. 2006. Seeing the world by hearing: Virtual Acoustic Space (VAS) a new space perception system for blind people. In 2006 2nd International Conference on Information & Communication Technologies, IEEE, 837–842.

[5] Hanne de Jaegher and Ezequiel di Paolo. 2007. Participatory sense-making: An inactive approach to social cognition. Phenomenol Cogn Sci 6, 4 (December 2007), 485–507.

[6] Orly Lahav and David Mioduser. 2001. Multisensory virtual environment for supporting blind persons' acquisition of spatial cognitive mapping–a case study. In EdMedia+ Innovate Learning, Association for the Advancement of Computing in Education (AACE), Norfolk, VA, USA, 1046–1051.

[7] Martez Mott, Edward Cutrell, Mar Gonzalez Franco, Christian Holz, Eyal Ofek, Richard Stoakley, and Meredith Ringel Morris. 2019. Accessible by design: An opportunity for virtual reality. In Adjunct Proceedings of the 2019 IEEE International Symposium on Mixed and Augmented Reality, ISMAR-Adjunct 2019, Institute of Electrical and Electronics Engineers Inc., 451–454.

[8] Darren J. Reed. 2019. Dancing with Data: Introducing a Creative Interactional Metaphor. Sociol Res Online 25, 4 (2019), 533–548.

[9] Yvonne Rogers, Helen Sharp, and Jennifer Preece. 2011. Interaction design: Beyond human-computer interaction (3rd edition ed.). John Wiley & Sons.

[10] Jaime Sánchez and Mauricio Lumbreras. 2009. Virtual Environment Interaction Through 3D Audio by Blind Children. Cyber Psychology & Behavior 2, 2 (January 2009), 101–111.

[11] Anastasia Schaadhardt, Alexis Hiniker, and Jacob O. Wobbrock. 2021. Understanding Blind Screen-Reader Users' Experiences of Digital Artboards. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, 1–19.

[12] S. K. Semwal. 2001. MoVE: Mobiltiy training in haptic virtual environment. piglet.uccs.edu/~ semwal/NSF2001PS.pdf.

[13] 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 (CHI '20), Association for Computing Machinery, New York, NY, USA, 1–13.

[14] Anja Thieme, Cynthia L Bennett, Cecily Morrison, Edward Cutrell, and Alex S Taylor. 2018. "I Can Do Everything but See!" – How People with Vision Impairments Negotiate Their Abilities in Social Contexts. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18), Association for Computing Machinery, New York, NY, USA, 1–14.

[15] M.A Torres-Gil, O. Casanova-Gonzalez, and José Luis González-Mora. 2010. Applications of virtual reality for visually impaired people. WSEAS transactions on computers 9, 2 (2010), 184–193.

[16] Dimitrios Tzovaras, Georgios Nikolakis, George Fergadis, Stratos Malasiotis, and Modestos Stavrakis. 2002. Design and implementation of virtual environments training of the visually impaired. Proceedings of the fifth international ACM conference on Assistive technologies - Assets '02 (2002), 48.

[17] Rosa van der Veen, Jeroen Peeters, and Ambra Trotto. 2018. Charged utopia VR: exploring embodied sense-making in the virtual space. In TEI '18: Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction, Association for Computing Machinery, Inc, 292–298.

[18] Herman Wahidin, Jenny Waycott, and Steven Baker. 2018. The challenges in adopting assistive technologies in the workplace for people with visual impairments. In Proceedings of the 30th Australian Conference on Computer-Human Interaction, ACM Press, 432–442.

[19] 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 CHI '19: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery.

[20] World Health Organization. 2021. Global strategy on digital health 2020-2025.