Beyond Audio Description: Exploring 360° Video Accessibility with Blind and Low Vision Users Through Collaborative Creation

Lucy Jiang
Cornell University
Ithaca, NY, USA
lucjia@cs.cornell.edu

Mahika Phutane
Cornell University
Ithaca, NY, USA
mahika@cs.cornell.edu

Shiri Azenkot
Cornell Tech
New York, NY, USA
shiri.azenkot@cornell.edu

ABSTRACT

While audio description (AD) is a standard method for making traditional videos more accessible to blind and low vision (BLV) users, we lack an understanding of how to make 360° videos accessible while preserving their immersive nature. Through individual interviews and collaborative design workshops, we explored ways to improve 360° video accessibility with immersion and engagement in mind. Our design workshops presented a unique opportunity for participants with diverse backgrounds to build on each others' personal and professional experiences and collaboratively develop accessible 360° video prototypes. Participants included both AD creators and users, with a focus on BLV AD creators as their perspectives are underrepresented in prior work. We found that immersive video accessibility went beyond an extension of traditional video accessibility techniques. Participants valued accurate vocabulary and different points of view for descriptions, preferred a variety of presentation locations for spatialized AD, appreciated sound effects for setting the mood and subtly guiding, and wished to engage multiple senses to boost engagement. We conclude with implications for immersive media accessibility and future research directions to support disabled people as creators of access technology.

CCS CONCEPTS

• **Human-centered computing → Accessibility**.

KEYWORDS

360° videos, audio description, video accessibility, blind and low vision, design workshop, co-design

ACM Reference Format:
Lucy Jiang, Mahika Phutane, and Shiri Azenkot. 2023. Beyond Audio Description: Exploring 360° Video Accessibility with Blind and Low Vision Users Through Collaborative Creation. In *The 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23), October 22–25, 2023, New York, NY, USA.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3597638.3608381

ASSETS '23, October 22–25, 2023, New York, NY, USA

ACM ISBN 979-8-4007-0220-4/23/10…$15.00
https://doi.org/10.1145/3597638.3608381

1 INTRODUCTION

Despite being touted as "immersive" and "interactive" [35, 81, 82], 360° videos are largely inaccessible to blind and low vision (BLV) users. The standard method for making videos accessible is adding audio description (AD), an additional audio track providing narration of descriptive yet concise visual information [72, 77]. However, there is limited guidance on how to make 360° videos accessible while preserving their immersive and interactive nature.

As the term suggests, 360° videos capture a panoramic field of view, allowing users to engage with different parts of a scene depending on where they orient their bodies or turn their heads, similar to extended reality (XR) technologies. 360° videos can be consumed on head-mounted displays, smartphones, and computers.

Researchers have started to address this accessibility gap [30, 37–39]. However, even with AD applied, BLV people remain excluded from the immersion and engagement afforded by 360° videos. Few studies have explored 360° video accessibility through examining user preferences for 360° descriptions. For example, Fidyka et al. [29–31] utilized focus groups to investigate professional AD creators' and BLV users' preferences for 360° descriptions. They found that users valued descriptions that captured the main action and were interested in spatializing AD for orientation.

Previous work focused greatly on descriptions — **researchers have not yet explored methods of conveying immersion to BLV users in a holistic way**. Prior findings also lacked detail regarding specific design considerations. While researchers have proposed that BLV users are interested in spatial audio, they have not suggested how to spatialize ambient sound effects or earcons. They have also taken a universal design approach, obscuring the breadth of individual preferences for AD. Lastly, many prior works on immersive video accessibility have not included BLV AD creators, professionals in the AD industry who typically write AD with a sighted assistant, framing BLV people as passive consumers rather than creatives with lived experience and writing expertise.

To address these gaps, we conducted a two-part study to explore ideas for enhancing 360° video accessibility. We pose the overarching research question: **How can we make 360° videos accessible?** To scaffold our approach, we investigated sub-questions such as:

- How can audio descriptions best support accessible and immersive 360° video viewing experiences?
- What additional feedback can improve accessibility and immersion for 360° videos?
- How and why should BLV people engage in AD creation?

Our study involved (1) individual interviews and (2) collaborative design workshops with a total of 14 participants. We conducted design workshops to engage participants with different ability levels in co-creating prototypes of accessible 360° videos. To support innovative idea generation, we included a co-design component, drawing on prior work on participatory design and focus groups [64, 70]. Our goals were to highlight individual preferences, discuss ideas and rationale for accessible 360° video designs, and engage both BLV and sighted people in the design process. Of the 14 participants, nine were BLV AD users and five were sighted AD creators. Five BLV participants were professionals in the AD industry and had intersecting experiences as both users and creators of AD.

Through our novel approach with a generative design component, we identify and analyze innovative design ideas brought forth by mixed-ability groups with varying AD experiences. First, we delineate how linguistic and aural components of descriptions, including word choice and earcons, convey immersion in nonvisual ways. Second, we detail how sound design (e.g., sound effects and volume) can establish scenes nonverbally, and present varied preferences on integrating touch, smell, and taste to the 360° video experience. We also share participants' perspectives regarding agency for video exploration and nonvisual attention guidance. Throughout the study, BLV AD creators contributed unique insights grounded in their lived and professional experiences, demonstrating the importance of including disabled people in the process of creating access technology.

Exploring 360° video accessibility establishes a foundation for developing accessible media in any dimension, including images, traditional videos, and the growing space of XR. To our knowledge, our work is one of the first to utilize group design activities that include both BLV and sighted people to elicit insights on (1) preferences for how to make 360° videos accessible and (2) ways to involve BLV people in the process of describing visual content.

2 RELATED WORK

This project is situated in the growing space of 360° videos, which is adjacent to the larger research space of image and video accessibility. We first highlight a subset of image description studies focused on establishing visual description guidelines. Stangl et al. [74, 75] identified that screen reader users universally wanted people, text, and objects to be described in pictures, but desired different details depending on the image's context. Morris et al. [56] found that increasing user agency for understanding and navigating digital images increased satisfaction. Additionally, Quero et al. [20] developed a multimodal prototype for BLV users to interact with art through tactile, haptic, audio, and verbal feedback, finding that additional sensory information was helpful for context and immersion. Generally, researchers advocated for context-aware, flexible, and personalized descriptions [8, 56, 74, 75].

2.1 Video Accessibility

Prior works on video accessibility aimed to understand BLV users' preferences for AD, automate AD creation, and explore innovative AD methods. In practice, AD creators follow practitioner-generated guidelines for content and presentation. For example, guidelines from the AD Coalition and Audio Description Project [1, 5] instruct AD creators to describe what they see rather than infer motivations, be as objective as possible, avoid describing over dialogue and critical portions of music, and more.

Some researchers have studied BLV users' video accessibility preferences for traditional videos. In examining AD preferences across participants with different vision levels, Chmiel and Mazur [23] proposed that middle-ground solutions could cater to most, but customizable options were optimal. Natalie et al. [60, 61] also uncovered BLV users' preferences through the development and analysis of ViScene, a tool to aid sighted novice AD writers. During the evaluation of the tool, BLV viewers' feedback on novice writers' AD centered on the descriptiveness, sufficiency, and clarity of the descriptions [62]. These findings exemplified how including BLV people in the AD creation process could improve the end product. Another study by Jiang and Ladner [42] drew on perspectives of BLV people with expertise as AD users and creators. They created a prototype to facilitate AD authoring through a question-and-answer system, and found that BLV participants wrote descriptions that detailed characters, actions, background settings, and unclear sound effects.

With the advancement of artificial intelligence (AI) technology, some researchers have utilized AI to facilitate AD creation and consumption, especially through human-in-the-loop approaches. Researchers developed tools to automate scene segmentation and script generation [16, 17, 32, 86, 87], identify inaccessible aspects of videos and provide feedback to creators regarding AD word choice and timing [48, 49, 63, 66], and convey position and action through spatialized audio [41]. Others researched how automated systems could give BLV users greater control over what information they could gain from a video [12, 67, 73]. While these works show how automation can simplify AD creation and access, they provide limited insights into BLV users' AD preferences.

Others have designed innovative approaches to presenting AD. Fels et al. [27] proposed a model of AD wherein a character served as a narrator for an AD script and the script utilized first-person pronouns such as "I" and "me." This method supported greater engagement during emotional and exciting scenes, such as a fight scene, but other scenes were not as well-suited to this style due to the narrator's subjectivity. In an investigation of AD for Shakespearean plays, Udo et al. [78] examined how writing descriptions in a style similar to the source material, such as iambic pentameter, could further improve engagement. Additional studies explored how haptics and tactile elements could increase access to videos. McDaniel et al. [53] and Viswanathan et al. [83, 84] utilized vibrotactile belts and gloves to provide spatialized haptic information to users. Though these cues increased user understanding of character positions and movements, they required additional attention to process. Another study found that including touch tours for live theater helped participants better understand the set, props, costumes, and play overall [79]. While these studies give insight to nontraditional methods of making visual content accessible, we build on these findings to further the research conversation about holistic and multisensory video accessibility, with a particular focus on immersive 360° content.

Prior research on video accessibility has largely focused on traditional AD formats, with only a few exceptions [27, 53, 78, 79, 83, 84]. It is generally acknowledged that the primary challenge in AD creation is time — fitting detailed descriptions in between gaps in dialogue [66]. However, with 360° videos, we must consider space as well. More importantly, we build upon this conversation by considering accessibility more broadly: AD is just one component of a larger collage of possible enhancements. In our work, we consider how various components can promote an engaging and immersive video experience.

2.2 360° Video Accessibility

Few researchers have explored 360° video accessibility. As a result, we also build on the larger body of work on 360° video experiences for sighted people. For example, when presented with a variety of 360° videos, sighted users appreciated having access to more information and valued having autonomy over their field of view [46, 47]. Furthermore, Bindman et al. [10] found that wearing a headset helped sighted participants understand their role as an embodied character in a 360° video, leading to higher narrative engagement and empathy. Others identified that spatialized audio improved users' sense of presence while viewing 360° videos [28]. We further explore designs that convey embodiment, and thus immersion, to BLV users in nonvisual ways.

Accessibility for video games and virtual reality (VR) is an active research area. We draw on work in this space as both video games and VR are commonly presented in a first-person perspective, similar to 360° videos. Prior studies explored adding AD to video games [51, 52]. Others found that utilizing audio cues like the Doppler effect to alert players to dangers or implementing spatial sound could help BLV players perceive relative position and orient themselves [33]. Guerrerio et al. [36] examined the tradeoffs between speech and sonification in VR games and proposed that context influenced how to best represent visual objects through audio. However, video game and VR accessibility work has focused primarily on utility rather than engagement, and does not examine how audio or other nonvisual tactics could impact BLV players' perceptions of presence.

Some have studied BLV people's perspectives on making 360° videos more immersive. Fleet and Herndon [26, 37] interviewed seven blind users to gather ideas for translating immersion in nonvisual ways. The authors live-described three videos using second-person pronouns to reinforce the intent of 360° videos as providing an embodied experience. Despite the linguistic alteration to traditional AD, participants could not distinguish between traditional and 360° video experiences. They also expressed that having more agency felt more immersive, favoring an interactive VR experience over predetermined 360° video flows.

Along a similar vein, Fidyka et al. [29–31] also examined 360° video description preferences. Through focus groups with BLV AD users and sighted AD creators, they found that users prioritized learning about the main action through AD and suggested that head movements could prompt additional descriptions for directional segments of the video. Participants in a subsequent study valued receiving detailed descriptions in a conversational style, and proposed that spatial audio could augment understanding and orientation within a scene [31]. Notably, the authors highlighted three ways to present AD within the soundscape (i.e., the audio mix consisting of AD, background music, etc.) [38, 39, 69]. "Omniscient" AD was audible from all directions, which was most similar to traditional AD. The "friend on sofa" mode mixed the AD into one ear to emulate video co-watching experiences. "Tracked" AD leveraged spatial audio to place the descriptions in the direction of action in 360° scenes.

Parallel to work on AD content and design, Chang et al. [21] developed Omniscribe, an AI-supported tool for 360° video AD creation. The application assisted sighted AD creators in creating *"immersive labels"* such as spatialized AD, scene descriptions, and object descriptions. While these measures helped BLV participants understand the video and granted them greater agency, they also increased the cognitive load for viewing. In our study, we expand on Chang et al.'s ideas for immersive description styles and further investigate BLV peoples' experiences as AD users and creators.

The studies described above provide preliminary insights about 360° video accessibility. However, they lack detail regarding BLV users' perspectives on immersion and engagement and do not explore additional modes of video accessibility beyond traditional AD techniques. We build on prior work to further examine the translation of immersion in 360° videos, understand user preferences for 360° experiences that are both serviceable and engaging, and explore the holistic user experience. Moreover, we expand upon the conversation by drawing on the perspectives of BLV AD creators, who have a unique intersection of skills and lived experience.

3 METHODOLOGY

To explore how to make 360° videos accessible, we drew on participatory design principles and prior research with co-creation components to involve disabled people in the design process [6, 11, 44, 59, 61, 62, 70]. We first conducted individual interviews, then held two collaborative design workshops with five and four participants, respectively. We collected preferences regarding traditional AD, engaged participants in brainstorming how to improve 360° video accessibility, and generated prototypes to make an undescribed 360° video accessible.

3.1 Participants

We recruited 14 participants through social media postings, mailing lists, targeted recruitment, and snowball sampling. We included participants who (1) identified as blind or low vision and regularly watched videos with AD and / or (2) were AD creators. We intentionally recruited participants at the intersection of both groups, BLV AD creators, to ensure that their unique perspectives were represented. Participants were required to be at least 18 years old and comfortable communicating in English. As the design workshop involved using a VR headset, we also screened participants according to the Meta Quest 2 Safety Guidelines [54, 55]. This study and all recruitment materials were approved by the Institutional Review Board at our university.

As shown in Table 1, nine participants identified as blind or low vision, and five participants were sighted. Ten participants were AD creators (including writers, voice talents, and audio engineers); five creators identified as BLV. Participant ages ranged from 18 to 69 (*mean* = 41.86, *SD* = 15.01) and eight identified as women while six identified as men. Participant types are delineated through their pseudonyms. For BLV AD creators, pseudonyms begin with "A." For sighted AD creators, pseudonyms begin with "S." For BLV users, pseudonyms begin with "B."

Table 1: Participant demographics, including participant type, the design workshop (DW) they attended if applicable, self-reported gender and ethnicity, paraphrased vision details, and occupation.

Type	Pseudonym	DW	Gender	Ethnicity	Vision Details	Occupation
BLV AD Creator	Aaron	1	Male	Caucasian	Blind (no light perception or functional vision)	AD advocate and producer turned AD company CEO
	Aleja	2	Female	Hispanic / Latina	Blind (color and light perception)	Quality and inclusion manager / bilingual voice artist / AD advocate
	Annie	2	Female	White	Blind (large objects and high contrast perception)	AD and accessibility consultant / dancer
	Aidan	-	Man	Black	Blind (no vision)	Freelance audio producer / AD advocate / narrator / AD consultant
	Amber	-	Female	White	Blind (small amount of vision in right eye)	Voice talent / university student
Sighted AD Creator	Shane	1	Male	White / Caucasian	Sighted	AD producer / consultant / trainer / speaker
	Scott	1	Male	Caucasian	Sighted	Audio engineer / voice talent / AD writer
	Sarah	2	Female	White	Sighted	Head AD writer / trainer
	Stacy	-	Woman	Caucasian	Sighted	AD writer and producer for live and interactive media
	Steve	-	Male	White	Sighted	Voice talent / AD producer
BLV AD User	Bella	1	Female	Multiracial	Blind (light and object perception)	Digital accessibility advocacy director / accessibility consultant
	Brynn	1	Female	Caucasian	Low vision (more central than peripheral vision)	Accessibility education support associate
	Becky	2	Female	Caucasian	Blind (no vision)	Volunteer for blindness organizations
	Blake	-	Male	Caucasian	Blind (no peripheral vision, central vision varies daily)	Educator / clergyperson / mental health professional

3.2 Procedure

Our study consisted of two components: (1) semi-structured interviews to understand participants' prior experiences and (2) collaborative design workshops to synthesize preferences and create an accessible 360° video prototype. This two-part methodology allowed us to individually probe prior to collaborative idea generation, as recommended by Sanders et al. [70]. The interviews equipped all participants with a shared context that they could build upon during the workshops.

Of the 14 total participants, 13 completed interviews and nine attended one of the two workshops. One workshop participant (Scott) was unable to complete the interview.

We conducted interviews between February and April 2023. The design workshops took place during March and April 2023. The first design workshop (DW1) was conducted in-person at the 38th CSUN Assistive Technology Conference with five participants. The second design workshop (DW2) featured two participants in-person at our university campus in New York City and two participants attending virtually through video conferencing software. Each workshop included mixed-ability groups of AD users and creators, ensuring a greater diversity of perspectives within each discussion.

3.2.1 Interviews. The interviews were structured as follows: general questions about experiences with AD consumption or creation, questions about 360° video experiences in particular, and a brainstorming session using a video probe. Interviews were virtual and lasted between 60 and 90 minutes.

We prepared two sets of questions, one for BLV participants and one for sighted participants. Participants who were BLV AD creators were asked both sets of questions.

BLV participants were asked about their video watching habits and AD preferences. Examples of questions included the following:

- **Video watching habits:** How often do you watch videos with AD? Do you watch videos with others (e.g., social co-watching and co-describing)?
- **Preferences related to AD they watched in the past:** Can you think of any examples of videos that have good AD? What makes an AD experience particularly excellent to you?

AD creators were asked about their roles and AD preferences. Examples of questions included the following:

- **Roles and tasks they had as professionals:** What is your role as an AD creator (writer, narrator, etc.)? Can you describe your everyday tasks as a creator?
- **Preferences related to AD they created in the past:** Can you think of any examples of videos that you described that you are particularly proud of or happy with?

Before the brainstorming session, we presented participants with two clips. Participants first used headphones to listen to a video with spatialized audio [9] to ensure that they understood the concept. Then, they watched the first 95 seconds of the undescribed 360° video "Avatar 2: The Way of Water" [85] using their phones. The video followed the journey of a human who teleported to the Avatar universe, transformed into a blue Avatar character, and explored the environment. To spark meaningful conversations, we selected a 360° video that included (1) spatial audio features, (2) embodied characters, (3) movement represented through walking, running, rolling, or flying, and (4) minimal dialogue to allow participants more space and flexibility for descriptions.

We then engaged participants in brainstorming design ideas based on the video probe. We first invited them to share their thoughts on the video, then transitioned into discussions about exploration and immersion. We asked broad questions such as, "What would make 360° videos more immersive in a nonvisual way?" and "How would you want to explore a 360° video?" We asked participants which point of view (i.e., first, second, or third-person perspective) they preferred for the AD. Based on their selections, we then read out a pre-written description of the video (provided in Supplementary Materials) and asked for their thoughts on the AD's point of view. We also asked which of three options they preferred for AD location: "omniscient," "friend on sofa," or "tracked" [39]. Throughout the interview, we encouraged participants to share the reasoning behind their design choices and asked them to recall specific examples or past experiences to ground their responses.

Participants were compensated with a $25 gift card for their time and contributions.

3.2.2 Design Workshops. The design workshops involved participant introductions, 360° video probe viewing, a design activity to collaboratively design an accessible video experience, and a brief reflection. Both sessions were conducted by two researchers and were approximately 90 minutes long.

We began by asking participants to introduce themselves; during this time, many shared information about their identities and professions. We also encouraged chatter and discussion to build rapport. After introductions, we passed around VR headsets (specifically, the Meta Quest 2) for participants to engage with the 360° video probe. Each participant watched the video one or more times.

Following the same criteria as in our interviews, we presented an undescribed 65-second 360° video titled "The Super Mario Bros Movie 360 / VR Experience" [40], which featured the user embodied as Mario and followed his journey of falling down a drain with Luigi and exploring a grassy plain with mushrooms. We anticipated that the video's whimsical nature would spark creative conversations. Figure 1 shows examples of scenes in the workshop video.

We then began the design activity, where participants were given 75 minutes to work together and create a prototype of an ideal 360° video experience. We encouraged participants to prototype the entire AD experience, including narration, accompanying audio cues, and other sensory cues for improving immersion and engagement with the video. BLV participants were encouraged to ask sighted participants questions throughout the process. One researcher wrote down lines of script dictated by participants and read them out when requested.

Since most participants had only experienced traditional AD, we followed up with questions such as "How would this work in a 360° setting?" or "How can we design descriptions and sound effects to improve this experience?" To combat groupthink, a common concern with research studies involving a group of participants [11], we encouraged participants to share their own preferences and to consider whether the prototype included those. After completing the design activity, we asked participants to reflect on the prototype and the creation process.

All participants were compensated with a $50 gift card and were offered up to $30 in travel reimbursements upon completion of the workshop.

3.3 Data Analysis

We audio recorded and transcribed all interviews and design workshops. Two researchers analyzed the data using inductive coding to identify emerging ideas for 360° video experiences. We individually coded two transcripts, discussed discrepancies, and reached alignment on the codebook. We then split up the remaining interview transcripts for coding.

We also coded the design workshop transcripts using our interview codebook. However, we realized that coding participant quotes alone did not capture group consensus or group interactions. To inform our approach, we referenced prior work on group methodologies, such as focus group research [18, 19, 25, 64]. We developed a schema for capturing and analyzing group dynamics in the workshops, drawing on Duggleby's three layers of focus group analysis: individual, group, and group interactions [25].

Our analysis framework (presented in Table 2) allowed us to glean insights about both the spoken dialogue and the dynamics of mixed-ability groups with varying AD experiences. At an **individual** level, we coded participant quotes in isolation using our interview codebook. At a **group** level, we considered each workshop group as its own entity using primarily existing codes. At a **group interaction** level, we examined group dynamics, consensus, and rapport. We also took note of participants' identities (i.e., sighted AD creator, BLV AD creator, or BLV user) when generating new codes, noting who led the discussions and who asked or answered questions.

We then conducted thematic analysis [13] by clustering codes, discussing potential groupings, and determining themes that best reflected participant perspectives.

4 FINDINGS

First, we present the prototypes created during the workshops. We then share participants' AD preferences, identified in both the interviews and workshops. Throughout this section, we analyze the collaboratively-created prototypes, as well as the discussions amongst participants, to uncover key design considerations for 360° video descriptions.

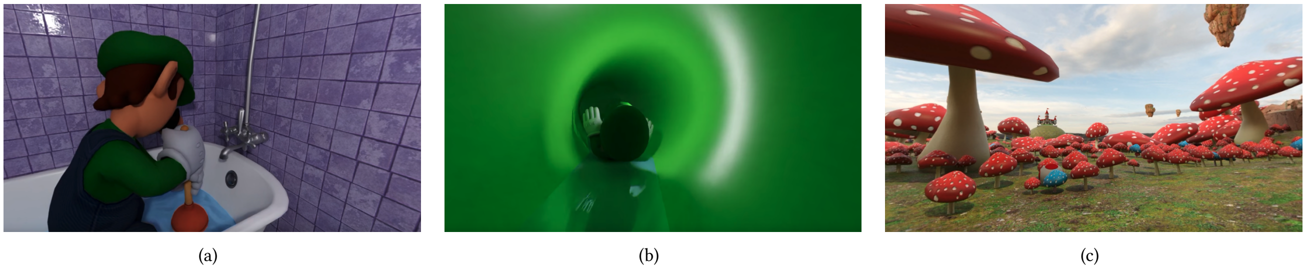


Figure 1: Screenshots from the design workshop video probe. The photos are in chronological order and show (a) Luigi plunging in the bathtub, (b) the user following Luigi down a green pipe, and (c) the grassy plain dotted with mushrooms.

Table 2: Our design workshop analysis framework featuring analysis at the individual, group, and group interaction levels.

	Individual	**Group**	**Group Interaction**
Who	Who said what? (i.e., participant ID, BLV identity)	Which group said what? (i.e., DW1, DW2)	Who was a dominant contributor and who was less vocal?
What	What was the content of what people said?	What did the group agree on (gleaned through the script and excerpts)?	What method did the group use to interact and discuss to come to a consensus?
How	How did the person present the ideas? (i.e., statement, question to the group)	How did the group present the ideas? (i.e., in the script, through discussion)	How was the group dynamic and how did that shape the outcome? (i.e., who was leading the discussion, did participants agree or disagree)
Analysis Process	Using same codebook as interview data	Using same codebook as interview data, treating each group as an individual unit (congruent methodological approach [25])	Creating new codes to capture dynamics

4.1 Prototypes

Workshop prototypes included audio descriptions and augmentative sound effects. Participants engaged in lengthy discussions about AD preferences before committing to script lines. For both groups, the scripts were approximately eight lines long and took between 60-75 minutes to write. Tables 3 and 5 show description scripts generated by DW1, while Table 4 presents the description script created during DW2.

The left columns show lines of AD in the order that they would be inserted as narrations into the video, which correspond to some views shown in Figure 1. The right columns report audio cues that participants wished to present alongside the descriptions. Typically, AD scripts include timestamps, but we did not record timestamps due to time and logistical constraints of the workshop. Table 5 presents an experimental version of description written during DW1, which positioned the AD narrator as a character within the content and resulted in fewer lines of AD.

Both scripts established embodiment by using first or second-person pronouns and including phrases such as *"as Mario."* The scripts were also similar in terms of describing colors, relative directions, and action verbs. These scripts represent styles, conventions, and preferences for AD, which is one component of a much larger space of possibilities for immersive and accessible content. Other components, such as earcons, haptic feedback, and smell or taste, were discussed during interviews and are mentioned below.

4.2 Linguistic Video Description Preferences

Participants felt that the linguistic presentation of AD strongly impacted their engagement and immersion within 360° videos. During the interviews, participants highlighted their individual AD preferences, including level of detail and adherence to the style of the source material, that were grounded in their prior experiences with AD. Many participants' preferences for traditional content were generalizable to 3D contexts, but they acknowledged new challenges such as conveying embodiment.

4.2.1 *Customizable Details, Accurate Vocabulary, and Uncensored Scripts were Engaging.*

BLV participants reported varying preferences for level of detail, and valued having options to customize the AD to their liking. For example, some identified as *"minimalist[s]"* (Aaron), while others thought additional details made the experience *"more immersive... [it's] like you're building a world for me with words"* (Aleja). Annie noted how providing different AD options allowed users greater *"agency over choosing what fits for*

Table 3: Script prototype written by participants in Design Workshop 1.

Line	Description	Audio Cues
1	As Mario, we watch Luigi, a tall thin plumber with a large mustache in a green hat, in a tub plunging a bathtub drain.	Suction cup plunging
2	Luigi gets sucked into the drain. He reaches up and pulls us with him.	Squished vacuum / Mario screaming underwater
3	We slide through a green tunnel.	Hollow echo (curve midrange, high frequency cutoff, small room) / water (mixed back)
4	We reach a fork in the pipe. Luigi goes to the left and we go to the right. We fly out of the pipe.	Water (left and right) / popping out of the pipe (mixed center and back)
5	We land on a grass patch, surrounded by mushrooms, some the size of a house, some the size of a leg.	Dead spots of mushrooms (absence of sound due to an object nearby) / boing from mushroom spring / landing
6	We approach a white toadstool with red polka dots.	Approaching large object / footsteps
7	The toadstool jumps up, wielding a stick.	Unsheathing sound (if it is part of the visuals)
8	We follow the toadstool as he bounces across the mushrooms.	"Boing" from mushroom spring (mixed to account for distance and travel sounds)

Table 4: Script prototype written by participants in Design Workshop 2.

Line	Description	Audio Cues
1	In 3D computer animation, you, as Mario watch your brother Luigi plumbing a tub.	Water, echoes, amplify plunging sounds
2	Your brother falls through the drain, and you fall after him as though on a water slide.	More splashes and more tube / closed sounds
3	You quickly slide down a green pipe and end up at a fork.	
4	Luigi goes left, you go right.	[Description should be read before the dialogue to set it up]
5	You pop out into a landscape covered in red and white mushrooms.	Open space, dream sound (like a musical interlude)
6	A white mushroom with red spots turns into a character and pops out in front of you.	"Boop" sound effect
7	He indicates a blue mushroom.	
8	You continue on and start bouncing from mushroom to mushroom toward a castle in the distance.	"Boing" spring-like sound effect for bouncing, vibrations when bouncing

Table 5: Script prototype of the character-as-narrator style written by participants in Design Workshop 1.

Line	Description
1	Eh Luigi, what's up with the drain? You're getting your mustache wet!
2	Oh no! Which way? You're kidding me! A fork in the —
3	What the! I've never seen so many mushrooms!
4	What's with the stick? Look out with that thing!

you versus people telling you... 'that's what you get.'" However, some warned that overly detailed AD could *"give spoilers"* (Shane) and interfere with suspense. Sarah mentioned that descriptions were typically read before the action occurred, *"unless it's a jump scare, because we want everyone to be scared at the same time"* (Sarah). While prior guidelines acknowledge this principle [5], maintaining equitable audience experiences was also important for 360° videos. So long as there were visual cues guiding sighted viewers, participants wanted description details and spatialized audio cues to convey sudden experiences regardless of where users were facing in the 360° space.

When scripts used fitting vocabulary and did not censor explicit content, participants felt substantially more immersed in the video. AD creators shared how they used linguistic and phonetic means to set the scene and engage users. For example, Sarah chose vocabulary to match the time period or mood of the content she was describing. When describing a historical romance show, she conjured the feeling of period romance novels by using words such as "parlor" instead of more modern phrases such as "living room." Stacy also followed a similar process: *"If an environment generally comes off as harsh or hostile, I will use synonyms that sound a little sharper."* Furthermore, participants thought that censored AD scripts were *"paternalistic, patronizing, [and] discriminatory"* (Aaron) and took users out of the *"raw and vivid"* (Aleja) nature of a piece.

4.2.2 Changing Points of View Conveyed Immersion and Embodiment. For embodied content, almost all participants preferred hearing AD from a first- and second-person perspective (using pronouns such as "I" and "you") over third-person (using pronouns such as "they") as they perceived the former to be more personal and immersive. Participants felt that the point of view of an AD script could convey the feeling of the experience (i.e., embodiment and immersion) to BLV users.

Interestingly, we noticed a shift in participant preferences during their discussions about the interview video probe. A majority of BLV participants (N = 6) initially opted to hear description from a third-person perspective as most existing AD is presented this way. However, they considered alternative perspectives for 360° AD once we clarified that they were visually designed to be the character themself.

During the interviews, participants differed on which perspective was best for 360° content. For example, Stacy had experience describing video games and preferred for descriptions to be from the third-person perspective. Although video games are typically presented from a first-person perspective, they do not allow users full agency over the character's movements. She used verbiage such as *"through the eyes of"* to convey the perspective of the user instead of switching the pronouns in the AD. In contrast, three participants stated that traditional methods of third-person description made them feel like they were *"watching it instead of being in it"* (Aaron). Shane acknowledged that traditional AD felt like being *"outside the fourth wall, but I can see through the fourth wall."* Aaron staunchly advocated for first-person-plural as opposed to second-person perspective, and thought that using "we" instead of "you" felt as though *"the describer is in it with me, rather than sucking me out of the show"* (Aaron). This perspective, which was agreeable to the other participants in DW1, was ultimately reflected in their prototype script.

Despite some differing opinions, a majority of participants (N = 9) preferred second-person description; others noted that their preference depended on the content. Annie stated that *"second-person makes much more sense... you're actually the person experiencing it and not a third party."*

Participants in DW2 expressed that the usage of second-person perspective alone was enough to imply embodiment. Although they concurred on the pronouns to use in the description, they debated whether to include the words "first-person perspective" in the script itself to explicitly convey that the user was intended to be part of the content. Given that first-person entertainment, such as video games, has historically been inaccessible to BLV users, they speculated that audience members may not have prior exposure to the terminology, and its usage could alienate BLV people who were unfamiliar with the concept. As Becky shared, *"it's more direct just to use 'you' in the description, because... the audience could be anyone, and you don't know what they would know."* The group concluded that using the word "you" clearly indicated that the user was experiencing the action and environment firsthand.

4.2.3 Characters Could Serve as Narrators. During DW1, the group experimented with an unconventional form of AD: a character as a narrator. Shane suggested either incorporating a separate character or converting an existing character to be the AD narrator, inspired by independent disability-centric theater companies, other stylistically cast description works, and AD research from Dr. Deborah Fels at Ryerson University [27]. He described AD as *"an aesthetic innovation"* that toed the line *"between telling people, showing people what somebody is looking like... [and] building it into the narration"* (Shane). Although none of the other workshop participants were familiar with this form of AD, they were excited about prototyping the immersive description experience. Aaron, who previously expressed his enjoyment in describing esoteric and nontraditional entertainment, found this to be *"experimental and cool"* and led the group in writing the character-as-narrator script:

> "I would love to see somebody [try this method]... It's figuring out what to say with the intent of letting us know what's happening but not making it just so descriptive. It's smart, very smart. It's harder to do, but I think it could be worth the effort."

The prototype (Table 5) was shorter than the others since the descriptions needed to double as plausible lines of dialogue. Despite this, the character-as-narrator script still conveyed participants' desired details. It described which characters were present, what they were doing (e.g., sliding down the pipe), and defining characteristics (e.g., Luigi's mustache). Participants suggested lines for the script in a playful manner:

> **Bella:** "My mustache is getting stuck on the side of the pipe!"
> **Shane:** Bella, that's perfect! Yes!
> **Bella:** "Because it's so big," or whatever!
> **Shane:** "My mustache is getting wet!" Exactly.
> **Bella:** "Oh no, I'm getting stuck with my mustache!" or whatever.
> **Brynn:** "My mustache is pulling me down this separate pipe" or something like that, you know. Make it funny, make it creative.

As an experienced sound engineer and AD creator, Aaron also expressed that such adaptations to existing AD paradigms *"could be revolutionary in making video games accessible"* in both traditional and immersive formats.

4.3 Aural Video Description Preferences

Participants shared varied thoughts about narration quality, audio mixes, and the spatialization of the AD location. Similar to linguistic presentations of description, audio preferences translated between traditional and 360° content. For example, the audio quality of a description track impacted users' experiences with comprehensibility and enjoyability. Further, many participants found poor audio and other problems in traditional videos to be exacerbated in 360° spaces, furthering inequity between sighted and BLV immersive experiences. When referring to a Disney World ride with a 360° video portion, Amber mentioned, *"I went on... a Pandora ride, and a main portion of this was watching a video... I liked [it], but one thing that was a bit jarring was that the audio description was flat. It wasn't with the rest of the audio, so it felt a bit separated."* Participants wished to have AD in the same number of audio channels (i.e., mono, stereo, spatial) as the source material. For 360° videos and immersive content, participants also explored how to design spatialized descriptions.

4.3.1 Fitting Narrators Indicated Cultural Competency and Quality.

Having high quality narration was crucial. Eight participants mentioned that "bad" narration made them feel disengaged from the content — or worse, stop watching entirely. Participants preferred narrators with nuanced, non-monotonous performances, and some specifically renounced the usage of text-to-speech (TTS) technology. Aidan emphasized *"cultural competency"* in particular when discussing how AD should honor the culture of the characters and their stories. Aleja also described how her excitement to watch a documentary was quashed due to a mismatch between the narrator and the cultural context:

> "The voice has to match the content. I was just made aware of [a show]... all about Black women's experiences with hair. It's a very fraught subject for Black people. And guess who's narrating it? A white woman! And it's like, 'What is that?' ... I really want to see the show, but it makes me sick to my stomach to hear this person doing the narration. And that's not okay... When things don't go quite right, instead of providing me access, you're literally making me not want to tune into something that I otherwise would really enjoy watching."

Overall, participants found poorly cast narrators to be *"distracting [and] ludicrous"* (Blake), breaking any sense of immersion.

4.3.2 Spatialized Audio Description Helped with Immersion and Clarity.

During the interviews, participants had diverse preferences for the spatial location of AD. Two chose "omniscient," two chose "friend on sofa," seven chose "tracked," and two wished to select the presentation location depending on the content (refer to Section 2.2 for definitions).

Despite minimal enthusiasm for "omniscient" AD during the interviews, participants in DW2 were most interested in the "omniscient" option and described it as an "inner monologue." For the Mario video, some participants were uncertain about which voices belonged to which characters. In the following exchange, participants built on each others' comments about how the AD placement could address some of their confusions. They concluded that "omniscient" was easiest to follow, as the consistent location differentiated it from character dialogue in the source material.

> **Sarah:** Right, well to other people's point about the dialogue sort of being in the middle, and the sound effects being around... keeping the audio description and your [character's] own dialogue sort of in the middle would make the most sense. Again, I'm a sighted person, so that would make the most sense to me intuitively, just because that, to me, would serve as sort of your inner monologue.
> **Annie:** I agree with that. Yeah, because for me, it was hard to find. I was expecting the voices to come from where I thought the movements were, because I also didn't know if... the characters speaking were the things that I heard moving, like the sliding or the jumping, or whatever...
> **Aleja:** Yeah, 100%. I envision it being me as Mario at the center, and then Luigi, wherever he happens to be. So that way, we can keep track of where the other person or things are, because you know that you are always at the center, right? It's you. And everything is moving relative to you.

Some participants were interested in the "friend on sofa" option based on their prior AD experiences. Steve recalled using a 360° description syncing application, and described the AD to be like *"a little cute gargoyle that's on your shoulder giving you audio description... placed in a way that's trusted."* During DW1, Brynn also shared that she wanted *"to have options to do either one ear or another. Especially if you're watching a movie with somebody and they don't want to hear your audio description but you still want to have it on, you have one headphone in and one out."* The flexibility of the "friend on sofa" option for choosing which ear to hear the AD in, as well as the privacy affordance in co-watching settings, appealed to participants.

The "tracked" option was generally perceived as the most immersive and interactive. Amber referenced her prior experience with spatial audio and ASMR: *"It feels more directional when they're kind of all around."* Aaron was enthusiastic about "tracked" as a user and creator: *"My personal favorite [is 'tracked']... I want the description to move around. I want it to be part of the thing, and it jumps over here for this, and it jumps over there for that... I would love to design that."* Sarah also expressed that the spatialization of AD could *"draw the viewer to look towards that as well if they have low vision, or to turn their head towards it just so they can hear it the same in both headphones,"* providing subtle guidance to users.

However, three participants recognized the difficulties of spatializing AD for hard of hearing users, and recommended having multiple options for AD location. Others felt that the "tracked" AD could be cognitively intensive, with Steve describing it as *"chaos"* and Annie sharing that *"it could be a lot to take in."* Regardless of the AD presentation location, participants reinforced that it was important to have adequate audio ducking [3] with dialogue and music — in other words, the AD needed to be audible over the source audio.

4.4 Sound Design Considerations for Exploration and Agency

Participants thought that audio cues could support agency in 360° video experiences and allow BLV users to have freedom of exploration and independent information access. They discussed audio cues such as sounds built into the content, augmentative earcons (distinctive sounds that convey an event or information [14]), or prompts to help with orientation. While guidance ensured that BLV users did not miss critical content, participants generally preferred subtle cues that could guide them while not interfering with their exploration.

4.4.1 Sound Effects Alone Could Convey Information. Both workshop groups discussed how sound effects could supplement the source material soundscape. While the original video contained some sound effects, all participants believed that the existing experience was insufficient for communicating information or setting the mood of the scene. Some sighted creators acknowledged the power of sound effects during the interviews — Steve, a voice talent, joked that *"defaulting to sound design [is best]... the less voice the better. I'm smiling because I know I'm putting myself out of a job."* Sarah, who develops AD curricula and guidance, shared how audio could establish the setting and reduce the need for AD during DW2.

> "If something is cutting between two scenes, once you set them both up, if one's at a party, and one is out in a field in the middle of the night, at the party you can hear people talking, and glasses clinking, and music and stuff like that. And in the field, you can hear crickets, or wind, or whatever... you can hear just from the quality of these two things, where you are in space."

A similar idea surfaced during DW1, where participants were dissatisfied with the intricacy of the video's audio experience. Traditional AD typically consists of only one narration track and does not modify a video's original soundscape; however, participants felt that enhancing the video's spatial audio could help convey both information and immersion. Shane suggested that companies hire sound designers to *"create a soundscape that [represents] Mario... this particular adventure, fantasy"* to work towards parity between the audio and visual quality. Aaron felt that the video needed *"a vacuum, watery, rubbery, squishy sound, because once he's in the drain, you need the sound of something being squished into the drain."* He also thought the sound of the characters going through the tube was easily conflated with a waterfall rather than a tunnel due to the lack of echo, and gave detailed suggestions about specific delays, cutoffs, reverbs, and EQs that could augment the spatial audio experience. The final sound effects and audio cues from both workshops are reported in Tables 3 and 4.

4.4.2 Sound and Speech Helped with Orientation. Eleven participants mentioned that spatialized sound effects and speech could seamlessly guide BLV users' attention within a 360° video. During the interview, most participants thought the speech in the Avatar video provided an immersive and lightweight way to reveal what was happening around them. Amber explained, *"Having the person be like, 'Hey!'... That's really helpful, just in terms of knowing which direction to face... It's not audio description, so you're still in the moment."*

The direction and proximity of both sound effects and speech were helpful for users. Aaron felt that speech where *"the voice leads ahead of me, we're running with it, and the voice seems like it maybe is even getting further away"* was helpful for conveying forward movement. Other participants referenced how they utilized the 360° nature of audio in real life to conceptualize their surroundings, and wished for similar cues in immersive videos. People who were not familiar with VR found audio cues to be useful for orientation, but would have felt *"annoyed"* (Sarah) if the guidance was compulsory and *"turned [them] around"* (Sarah) to face the action.

4.4.3 Augmentative Earcons and Prompts Could Help with Orientation. Participants suggested including spatial earcons and prompts to help them understand which direction to face. Acknowledging that 360° videos typically had one focal direction, Aidan proposed having a "north star" or "home base" to use as an absolute reference point to the source content:

> "Is there a way to orient me, whether that be the cardinal, or if it's the clock face? Whatever the case, where is my north? Where's my 12? ... If there was a little tiny beep up at the 12 o'clock position, my north, and then I noticed that that beep is moving a little bit, like, 'Oh wow, there's a little bit of turn happening there.'"

Blake also wished to be informed, but not directed by the AD or earcons: *"If it was directly behind me... let me know that there's more that I'm not experiencing. But I would not want much more hands on [than] that."* Participants felt that guidance would assuage their fears and stresses of missing out on content, thereby improving their experience overall. Annie felt that including guidance in a 360° video experience could make it feel *"more casual... it takes a lot of the stress out of it,"* while Aleja shared: *"I like exploring, but if I fall too far — if I'm going to far the wrong way, or whatever — then I wouldn't mind a little beacon to guide me back to where I need to be."*

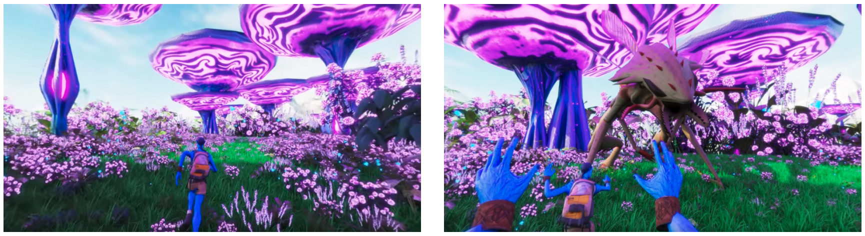
Once again, participants valued customizability and wished to toggle between guiding and exploring modes depending on their mood or the type of content. For example, Aaron preferred having structured guidance towards the action, but *"would want the option to choose on a movie by movie basis."* On the other hand, Aidan noted, *"I would love to be able to explore on my own, but again, knowing where I'm at, right, and being able to get back to where I want to get back to — that type of thing would be cool."* The level of personalization between users and within users' preferences showcased the importance of flexibility for experiencing immersive environments.

4.5 Multisensory Interactions and Feedback

Participants were interested in novel ways of interacting with videos in multisensory (i.e., tactile, olfactory, or gustatory) ways, but some were concerned about including smell and taste in their experience.

4.5.1 Haptic and Tactile Feedback Increased Engagement and Understanding. All interview participants (N = 13) were interested in haptic feedback and tactile augmentations. They referenced existing examples of tactile analog content, such as theater pre-shows and museum exhibits, and haptic digital content, such as vibrations from video game controllers.

Referring to his experience describing museum exhibits, Shane explained how touch tours allowed BLV patrons to gain a hands-on and interactive understanding of historical artifacts and works of art. He shared that some live theaters had *"set displays in the lobby,*



(a) (b)

Figure 2: Screenshots from the interview video probe. The photos show (a) the brightly-colored environment in the Avatar world and (b) a large bug that participants encountered.

[with] all kinds of tactile elements of the props, models, spots, swatches of fabric that represent the costumes" (Shane). As a blind dancer and AD creator, Annie explained how touch tours made decorative and symbolic items, such as a nutcracker, more understandable and concrete to those who were unfamiliar with them.

Both BLV and sighted participants were intrigued by the prospect of receiving *"vibro-tactile feedback"* (Bella) from mobile devices and VR controllers. As a Braille reader, Aleja thought that tactile feedback for 360° video experiences was reminiscent of using her Braille display. She suggested that tactile elements such as textures could be presented on a *"a touch pad, like on our computers, but on the [controller]"* (Aleja). Four participants acknowledged that additional sensory elements could benefit d/Deaf and hard of hearing or deafblind users.

> "Haptics are fun... highlighting things might be a different way of access for people who are deafblind, getting more tactile feedback. I think we often really underutilize... things like different textures or feelings that could emphasize what's happening on screen or in the environment. [Making] it more immersive is also more accessible." (Annie)

Steve drew on his experiences with VR and video games to imagine potential haptic feedback signals. He suggested *"to have a little thump or some sort of indication: 'Hey, you're getting close to something that you want to explore.'"* Stacy, an avid gamer, mentioned technologies such as haptic chairs and vests that provided more feedback than controllers alone. She brainstormed novel ways for haptics to compensate for sound that was obscured by AD:

> "Haptics are a great way to imply very loud sounds. While narration is going on, the audio mix will duck very loud sounds if you are narrating over them... [if] there's a huge explosion, a lot of rocks are crumbling – you [have] to describe that, even if you want to let parts of the explosion come through. So haptics could imply some of the background audio... that is being ducked or slightly sacrificed for the sake of narration."

4.5.2 Smell and Taste were Interesting, but Less Critical. Most were familiar with the concept of engaging olfactory senses through past experiences with 4D interactions, and six participants were open to incorporating it into their 360° video experiences. When commenting on the Avatar video from the interview, Blake shared how scent could represent the environment: *"So I'm assuming that there's grass or flowers... [Having] the smell of grass, versus the smell of fresh rain, versus the smell of flowers, would give a very different sensation of what's going on there."*

Others were interested in trying this feature, but had reservations about unpleasant smells making them nauseous. In contrast to most participants, Aaron mentioned, *"I've been to theaters plenty of times with the air and the water, and the vibrations... I'm not a big fan."* The diversity of participant preferences reiterated the importance of customizable settings to allow people to engage in ways that best-suited them. Overall, fewer participants were interested in taste, with many citing the bug in the Avatar video probe as a grotesque taste they would not want to experience. Figure 2 shows screenshots from the Avatar video of the environment and the bug, as mentioned by participants.

4.6 Both Quality and Quantity of Described Content Were Important

BLV users felt that AD granted them *"autonomy"* (Aleja) in their viewing choices. However, they were frustrated by insufficient access, due to descriptions being unavailable, inconsistently offered across platforms, or poorly executed. In particular, Aaron expressed his frustrations about the lack of availability of AD despite the passage of digital accessibility legislation, and mentioned that there was no monetary excuse for inaccessibility.

> "Blind people tend to be just happy to get whatever they get, instead of saying, 'Hey, we're entitled to more.' ... [As an audio describer,] I'm also intimately familiar with the price that they pay for audio description services, and I know that we are the lowest price item of anything they do... it costs less money than it does to feed the crew."

However, many observed that when companies attempted to meet legislative compliance, AD quantity increased while quality decreased. New technologies were a substandard replacement for human-written, narrated, and mixed AD:

> "As more and more streaming platforms are required to have audio description, what's really disappointing is that a lot of providers don't care, just because it's accessibility, and they go, 'Oh, it's a box we need to check, like closed captions.'" (Sarah)

Examples of declining quality due to cost-cutting included using TTS technology or hiring less experienced narrators. As a BLV narrator, Aidan shared his view on declining AD quality:

> "From a consumer perspective, I worry about the TTS, I'm worried about the quality. I see a lot of companies who are taking [shortcuts] to cheapen the product... [As] someone who works within the industry, I can tell you that the cost cutting is awful, because that... impacts the number of opportunities that you can get... it impacts blind narrators more than non-blind narrators."

Although manually creating AD requires more time than using AI, participants agreed that the difference was noticeable. Steve, an experienced voice talent, acknowledged the gap between high- and low-quality narration:

> "I think there's a professionalism involved. The best comparison I can think of is going to a local community theater sing-along karaoke versus going to Broadway... there's a difference even though it might be the same song, the same lyrics, the same notes, the same piano."

4.7 Including BLV AD Creators

4.7.1 Becoming a BLV AD Creator Was Viewed as Controversial. All participants advocated for the inclusion of BLV consultants and quality control experts. However, creators such as Aleja shared that the path to working in the AD industry was not easy, and that their involvement in AD creation was met with scrutiny:

> "I, along with a sighted consultant, wrote the audio description for [a documentary]. That was super controversial. It still is. There's a lot of people who don't believe that blind people have any space in writing anything visual. I don't agree with that. I think that we can perform all roles of the AD workflow... blind writers are out there."

Others argued that involving BLV people in AD creation was analogous to involving disabled people in other occupations with appropriate accommodations.

> "The industry needs to not just be all about creating **for** folks, but folks want to create for themselves and for others... Folks have a real hard time accepting that the accommodation for a blind person to write audio description is someone or something... I just see it as an accommodation, like everything else." (Aidan)

The sighted creators also emphasized the value of involving BLV people as experts in AD creation. When brainstorming how to best describe 360° videos, sighted creators often wished to consult BLV users for their preferences instead of making assumptions. For example, Stacy actively solicited feedback from BLV users on a livestream when describing videos. During DW2, Sarah acknowledged her positionality as a sighted person and mentioned that she didn't want to *"dominate the discussion."* She also shared that incorporating disability justice in AD creator training provided a *"critical piece of context for getting into AD"* (Sarah).

4.7.2 BLV Experts Contributed Unique Insights During the Design Workshops. During the workshops, BLV participants often asked detailed questions about characters' appearances, actions, and settings, and identified segments that did not make sense based on the audio. For example, Aleja recognized a body of water from the audio, but was confused about what it actually was: *"It sounds like he's just kind of tromp, tromp, tromping through some type of body of water — a swamp, a beach, a puddle. I mean, it could be anything."*

Though the Mario franchise was familiar and well-established to the sighted participants, BLV participants identified gaps in prior understanding that needed to be clarified in the descriptions. For example, Annie inquired about the stylistic and artistic presentation of the visuals to better understand the video's overall atmosphere during DW2.

> "What is the coloring, or what is the style like, visually? ... Is it 3D? Is it claymation? Is it cartoons? Is it live action? Is it very bright colored? Is it neon video gamey surreal? Is it very cartoon-like? ... Even if we're not seeing it visually, that can give some context."

These questions sparked a larger discussion about the inaccessibility of existing media and how describing content now could not remediate decades of inaccessibility to *"cultural touchpoints"* (Bella). Annie also related her memories of access barriers with the Mario games to her current experiences:

> "It's this nostalgic thing for some people. But then again, if you've never played this... it's a little bit awkward, because you're making it accessible to people where the games may not be accessible... you're kind of saying, 'This is a nostalgic thing.' But oh, you never knew that anyway."

BLV participants also discussed how much detail to include in the final prototype. They considered all BLV participants' preferences when collaboratively generating concise yet vivid AD. Notably, Aaron emphasized his initial confusion about whether the characters were humanoid, something that sighted creators in both workshops did not mention until asked. The following excerpt showcases the group dynamic between BLV participants in DW1:

> **Aaron:** Let's just say we're Mario, right. And then what about the description of other characters? Is that good?
> **Bella:** What if you say we're Mario, and then describe a little bit about him? Or no, is that too much?
> **Aaron:** Right, but really short. I think it would have to be short. Because you don't want a 20-minute —
> **Bella:** Yeah, no, no, I agree. What are the defining features of him? The color of his hat.
> **Brynn:** Yep.
> **Aaron:** And the fact that he's human! Because my thing about Toad was I didn't know if he was human or something else. So you at least gotta tell people they're a short pudgy human with an Italian mustache. That's good, right?

BLV participants' clarifying questions helped refine the AD script to be as precise as possible. For example, BLV creators asked for more detail after hearing snippets of ad-hoc descriptions from the sighted creators. Aleja asked, *"Are they tumbling down this thing? Are they rushing? ... What exactly is the movement here, as they're going down?"* Aaron also purposely paused the conversation during the workshop to garner feedback from other BLV attendees. He asked, *"So when someone says they see a red polka dot toadstool and you imagine it in your mind, do you think that it's a red mushroom with white polka dots, or a white mushroom with red polka dots?"*

Lastly, BLV and sighted participants acknowledged each others' contributions to the collaborative AD creation process. Sighted participants emphasized their respect for their BLV peers, with Shane commenting on Aaron's experience as an *"expert sound editor."* Sighted participants also joked about being the *"token sighted people"* (Shane) in the group. At the outset of DW1, Aaron drew on his experiences of writing description with sighted assistants, sharing that *"we're going to have to rely on our sighted companions here to be really clear and eager describers about what was there."*

5 DISCUSSION

In this paper, we detailed design considerations for making 360° videos more accessible, guided by input from BLV and sighted AD experts. To our knowledge, our work is one of the first to consider BLV people as creators of AD and video accessibility more broadly, rather than just users. We now discuss immersive media accessibility, consider areas for future work and research in the field of video accessibility, and reflect upon our methodology.

5.1 Implications for Accessible, Engaging, and Immersive Media

Our work uncovered ways to improve nonvisual immersion in 360° videos and highlighted how BLV people contribute to AD creation. While our findings centered on 360° videos, our proposed design considerations are generalizable across many types of media. For example, these findings can be applied to making traditional videos or even images more immersive via soundscapes and haptics, elaborating on ideas presented by Morris et al. [56]. Beyond digital interactions, designers can draw on these design ideas to improve accessibility for board and card games [43] or formative learning experiences for BLV children [15].

These insights can also be applied to the video game space. Prior work briefly investigated the intersection of AD and video games [33, 36, 51, 52]. Only a few participants drew connections between AD and video games; most BLV participants were unfamiliar with the video game space due to its overall inaccessibility. However, after engaging with the videos, they expressed that utilizing creative techniques to design AD for video games and other immersive content would be *"revolutionary"* (Aaron). Work from Microsoft on making racing games more accessible bolsters the importance of continuing research on visual description and sound design to guide users without removing their agency in immersive environments [22]. Through our study, we share additional methods for conveying embodiment in first-person content, discuss how earcons can help with orientation and guidance, and consider how integrating additional senses can improve accessibility and immersion.

Findings regarding the novel task of describing across an omnidirectional space are also applicable to making extended reality more accessible. Many BLV participants in our study expressed interest in XR, but chose not to explore the technology due to their uncertainty about its accessibility. While efforts to add descriptions to VR utilized TTS technology for description [34, 88], our work shows that BLV people found human-created descriptions to be more immersive due to the usage of contextual vocabulary and fitting narrators. We encourage future work on accessible XR to further investigate nonvisual representations of visual information and immersive ways to present this information while including BLV people in the design process.

BLV participants discussed how they were excluded from discourse on popular culture, as many *"cultural touchpoints"* (Bella) were expressed only visually in the media. They felt that current efforts to make content accessible were inadequate remediation for decades of prior inaccessibility and cultural exclusion. However, they were excited about potential new ways for presenting AD, such as the character-as-narrator format mentioned by Shane and first introduced by Fels et al. [27]. Participants' eagerness to try new methods of AD motivates further research and innovation in this space.

Our findings on best practices for 360° video accessibility can serve as a foundation for designers and AD creators to establish accessibility standards for immersive content, such as first-person video games and XR.

5.2 Artificial Intelligence Has Potential to Increase AD Quality and Quantity

Researchers have studied how AI can make human processes for creating AD more efficient by automating parts of the workflow [12, 16, 17, 32, 49, 66, 86, 87]. Future research should consider how to leverage AI to support AD in traditional or 360° videos; however, as we highlighted, descriptions comprise just one of many components of video accessibility. For example, AI could help in determining what directional segments of a 360° video to describe. Using computer vision on key frames or natural language processing on dialogue, researchers can identify an optimal viewing path that captures key points, guiding creators to describe salient video segments while optimizing for minimal head turning.

The rapidly advancing space of AI, which includes multimodal large language models such as GPT-4 [65] and improved video captioning systems [4, 24], has potential to further automate components of AD creation and increase the amount of AD available to users. However, it is essential to consider how AI-assisted AD creation is executed. AD experts observed that the increasing quantity of AD on the market is often coupled with lower quality to meet regulatory mandates. Others referred to the proliferation of low quality AD with TTS as *"cheapening"* (Aidan) the product. We recognize that AI could enable companies to release the cheapest minimum viable product to BLV users for the sake of compliance, rather than work towards full accessibility and equity.

Certain applications of AI could increase user agency and video accessibility. For example, more advanced conversational visual question answering agents could support BLV users in writing AD, similar to BLV AD creators' current workflows with sighted assistants [42]. AI could also modify AD verbosity or content based on user preferences to better suit users' video watching goals. Udo et al. [80] examined the merits of integrating AD into the production process of a piece; AI could even be used to assess scripts prior to filming to proactively confirm if they provide enough space for AD.

Prior works on image and VR accessibility have largely taken a utilitarian approach to access, and do not place emphasis on BLV users' enjoyment when interacting with content. Tensions between higher quantity and quality of AD often focus on what is serviceable, rather than what is enjoyable. During the interviews and workshops, participants often gave examples of content they watched for leisure and emphasized that AD had the potential to be *"an aesthetic innovation"* (Shane). Our work considers *why* BLV people engage with content — for entertainment and enjoyment — a prominent component of the content consumption experience that is under-addressed in most prior research.

Future work can explore open questions about nonvisual access to content. For example, what are BLV users' thoughts on using advanced AI to create access for subjective content such as art or videos? How can we harness emerging multimodal LLMs to create greater quantities of AD while maintaining quality? We recommend further exploration at this intersection of artificial intelligence and video accessibility.

5.3 BLV Involvement Can Improve Video Accessibility Outcomes

Studies on accessible visual media typically position BLV people as passive consumers, assuming that their only interactions with access technology are as users [12, 16, 17, 23, 29–31, 33, 51–53, 87]. However, the invention of AD is typically credited to blind visionaries [76].

Few studies have considered how BLV and sighted people can work together to create access, such as Natalie et al.'s [61, 62] works on BLV feedback for novice-written AD and Muehlbradt and Kane's [59] study on collaborative image captioning. While prior research provides valuable insights into BLV description cocreation, we synthesized insights from larger groups — 4 to 5 people instead of dyads [59, 61, 62] — and showcased tensions between BLV participants' individual preferences. Our mixed-ability approach also differs from group design sessions featuring only BLV users, as studied by Morrison et al. [57] and Siu et al [71].

Our findings highlight how BLV participants' input can enrich and improve the AD industry. Prior to this study, the vast majority of BLV participants (N = 8) had minimal exposure to XR. As a result, their perspectives reflected a broad unfamiliarity with this technology and first-person entertainment formats, which led to novel design contributions and questions that extended beyond existing paradigms. For example, during both workshops, BLV participants consistently asked questions about parts left undescribed by the sighted participants. They also discussed their nuanced feelings of *"nostalgia"* (Annie) towards the workshop probe, given the long history yet prior inaccessibility of the Mario universe. While prior work has reported that sighted participants miss information desired by BLV viewers [49, 58, 61, 62], the consideration of BLV users' prior cultural context introduced additional complexity to the level of detail the AD needed to provide. For example, participants asked about the visual style of the piece, whether the characters were humanoid, and what the sound effects meant.

The diverse contributions of BLV and sighted participants showcase the value of including both perspectives when crafting descriptions. As we analyzed the design workshops for group dynamics, we found that BLV AD creators often drove discourse on key points due to their rich expertise as narrators, audio engineers, writers, and AD users. Understanding BLV people's AD preferences is an ongoing discussion, and including BLV participants in AD creation ensures that this communication channel between users and creators remains open [42].

Lastly, Bennett et al. [7] propose the interdependence framework as one that "emphasizes how myriad people and devices come together to build access, with special attention to acknowledging the work of people with disabilities." We apply this framework to AD creation and recognize how the unique contributions of BLV and sighted people can be integrated. Broadly, we encourage future work to include disabled people as both creators and users of access technologies.

5.4 Benefits and Limitations to Our Method

According to prior research on interviews and focus groups, generalizing individual preferences to broader populations can be difficult [50]. As one of the first studies to conduct design workshops with mixed groups of BLV and sighted users [2, 6], we critically evaluate our method to determine whether and how our design workshops generated novel results. We find that the discussions, disagreements, and conclusions generated through this method gave us fundamental insights into both the AD process and the needs of BLV people, which would not have been possible to obtain with interviews alone.

5.4.1 Individual Interviews Prepared Participants for the Design Workshop.

The participatory framework proposed by Sanders et al. [70] stated that individual probing followed by group idea generating led to the strongest results. We engaged participants in individual interviews before the group activity to provide them with a common vocabulary and experience to discuss 360° video preferences more broadly. During the workshops, multiple participants referenced the interview video probe [85] to compare and contrast design features such as spatial audio. Citing prior materials helped them ground their suggestions, which moved groups towards consensus more quickly. For future studies that employ this two-part research method, we recommend including complementary activities in both components to elicit richer insights.

5.4.2 Design Workshop Activities Drove Discussion.

Our design workshops centered on the 360° video probe, a whimsical tale of Mario and Luigi falling down a drain and to a magical land [40]. The fantastical nature of this video sparked lively debate as it was *"difficult to describe"* (Aaron) and encouraged participants to ask clarifying questions. A key element of this video, the embodied character, provoked further discussion surrounding embodiment through sound and AD. Moving forward, we suggest that researchers select group activities to bring forth individual participants' differences and spark conversation and debate.

We chose animated videos as they are prevalent in 360° content, video games, and VR, but acknowledge that non-animated videos are also common. Although the minimal amount of dialogue in the video opened more space for discussion of relevant descriptions, many videos have more dialogue, which introduces known timing challenges. Future work can examine differences between immersion in animated versus live action 360° video content and explore videos with a greater variety of dialogue levels.

5.4.3 Establishing Rapport Led to Observed Power Imbalances. We reflect on group dynamics and corroborate that establishing rapport is important [45]. Interestingly, yet unsurprisingly, many of the workshop participants were acquaintances or friends with each other, primarily through work connections as AD creators. Participants who did not know others (all of whom happened to be BLV AD users) appeared more hesitant to share their opinions within a group of people with more expertise. Although we found it necessary for participants to share their occupations and relations to AD with the group to contextualize their comments, we observed a power dynamic between creators and users. We acknowledge this as a potential limitation to the design workshop method. In line with limitations of other qualitative user studies [68], we also recognize that our work captures a limited set of perspectives.

Future research could study interactions between participants with different ability levels and varying amounts of topical expertise. This could guide best practices for building rapport and credibility among participants while reducing the detrimental effects of power imbalances.

5.4.4 Group Interactions Gave Insight Into the AD Creation Process. During the workshops, we identified that sighted creators wished to defer to BLV users and did not want to *"dominate the discussion"* (Sarah), but their verbal descriptions were necessary to provide BLV participants with context about the visuals of the video. This tension was ultimately resolved while creating AD, as (1) sighted creators provided preliminary descriptions, (2) BLV participants asked clarifying questions, (3) BLV and sighted participants worked together to refine the descriptions, and (4) all participants came to consensus on the final lines of AD.

As with prior work on BLV involvement in creating visual access [42, 59, 73], question and answer exchanges supported BLV people in engaging with the writing process and contributing their unique perspectives. These conversations gave insight into the inner workings of AD creation workflows that incorporated BLV users as well as sighted creators.

6 CONCLUSION

In this study, we used a novel approach to elicit innovative design ideas for accessible 360° video experiences and identify the benefits of including disabled people as creators of access technology. Through individual interviews and collaborative design workshops with BLV and sighted AD experts, we learned that 360° video accessibility and immersion can be conveyed in multiple ways: through the linguistic and aural presentation of AD, through sound design that leverages both sound effects and earcons, and through multisensory augmentations such as tactile or haptic feedback. Participants also preferred non-intrusive guidance while navigating the videos. We identified how BLV participants' contributions about details and sound effects improved the AD creation process and video accessibility overall. Understanding BLV people's perspectives on nonvisual immersion and engagement opens up possibilities for future accessible experiences across many types of media, including photos, videos, video games, and extended reality.

ACKNOWLEDGMENTS

We thank our anonymous reviewers for their valuable feedback and experience. We also thank Kelly Jiang, Dr. Joel Snyder, Daniel Zhu, Sharon Heung, and Emma McDonnell for their support and input. We are grateful to our study participants for their involvement and insights through both stages of the study. This work was partially supported by a gift from Meta (Meta Platforms, Inc.).

REFERENCES

[1] ADI AD Guidelines Committee. 2003. *Guidelines for Audio Describers.* https://adp.acb.org/guidelines.html

[2] Jérémy Albouys-Perrois, Jérémy Laviole, Carine Briant, and Anke M Brock. 2018. Towards a multisensory augmented reality map for blind and low vision people: A participatory design approach. In *Proceedings of the 2018 CHI conference on human factors in computing systems.* 1–14. https://doi.org/10.1145/3173574.3174203

[3] Apple Inc. 2023. *Change your VoiceOver settings on iPhone.* https://support.apple.com/guide/iphone/change-your-voiceover-settings-iphfa3d32c50/ios

[4] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. 2021. Vivit: A video vision transformer. In *Proceedings of the IEEE/CVF international conference on computer vision.* 6836–6846. https://doi.org/10.1109/ICCV48922.2021.00676

[5] Audio Description Coalition. 2009. *Standards for Audio Description and Code of Professional Conduct for Describers.* https://www.perkins.org/wp-content/uploads/elearning-media/adc_standards.pdf

[6] Shiri Azenkot, Catherine Feng, and Maya Cakmak. 2016. Enabling building service robots to guide blind people a participatory design approach. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI).* IEEE, 3–10. https://doi.org/10.1109/HRI.2016.7451727

[7] Cynthia L Bennett, Erin Brady, and Stacy M Branham. 2018. Interdependence as a frame for assistive technology research and design. In *Proceedings of the 20th international acm sigaccess conference on computers and accessibility.* 161–173. https://doi.org/10.1145/3234695.3236348

[8] Cynthia L Bennett, Cole Gleason, Morgan Klaus Scheuerman, Jeffrey P Bigham, Anhong Guo, and Alexandra To. 2021. "It's Complicated": Negotiating Accessibility and (Mis) Representation in Image Descriptions of Race, Gender, and Disability. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–19. https://doi.org/10.1145/3411764.3445498

[9] Binaulab Audio 3D. 2016. *Queen - Bohemian Rhapsody - 3D AUDIO (TOTAL IMMERSION).* https://www.youtube.com/watch?v=VnzIIhLNHqg

[10] Samantha W Bindman, Lisa M Castaneda, Mike Scanlon, and Anna Cechony. 2018. Am I a bunny? The impact of high and low immersion platforms and viewers' perceptions of role on presence, narrative engagement, and empathy during an animated 360 video. In *Proceedings of the 2018 CHI conference on human factors in computing systems.* 1–11. https://doi.org/10.1145/3173574.3174031

[11] William Boateng. 2012. Evaluating the efficacy of focus group discussion (FGD) in qualitative social research. *International Journal of Business and Social Science* 3, 7 (2012). https://s3.wp.wsu.edu/uploads/sites/2154/2015/09/Evaluating-the-Efficacy-of-Focus-Group-Discussion-in-Qualitative-Social-Research.pdf

[12] Aditya Bodi, Pooyan Fazli, Shasta Ihorn, Yue-Ting Siu, Andrew T Scott, Lothar Narins, Yash Kant, Abhishek Das, and Ilmi Yoon. 2021. Automated Video Description for Blind and Low Vision Users. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–7. https://doi.org/10.1145/3411763.3451810

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. https://doi.org/10.1191/1478088706qp063oa

[14] Diane Brauner. 2023. *iCons and Earcons: Critical but often overlooked tech skills.* https://www.perkins.org/resource/icons-and-earcons-critical-often-overlooked-tech-skills/

[15] Emeline Brulé and Gilles Bailly. 2018. Taking into Account Sensory Knowledge: The case of geo-techologies for children with visual impairments. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–14. https://doi.org/10.1145/3173574.3173810

[16] Virginia P Campos, Tiago MU de Araújo, Guido L de Souza Filho, and Luiz MG Gonçalves. 2020. CineAD: a system for automated audio description script generation for the visually impaired. *Universal Access in the Information Society* 19 (2020), 99–111. https://doi.org/10.1007/s10209-018-0634-4

[17] Virgínia P Campos, Luiz MG Gonçalves, Wesnydy L Ribeiro, Tiago MU Araújo, Thaís G Do Rego, Pedro HV Figueiredo, Suanny FS Vieira, Thiago FS Costa,

Caio C Moraes, Alexandre CS Cruz, et al. 2023. Machine Generation of Audio Description for Blind and Visually Impaired People. *ACM Transactions on Accessible Computing* 16, 2 (2023), 1–28. https://doi.org/10.1145/3590955

[18] Martha Ann Carey. 1995. Comment: Concerns in the analysis of focus group data. *Qualitative health research* 5, 4 (1995), 487–495. https://doi.org/10.1177/104973239500500409

[19] Martha Ann Carey and Mickey W Smith. 1994. Capturing the group effect in focus groups: A special concern in analysis. *Qualitative health research* 4, 1 (1994), 123–127. https://doi.org/10.1177/104973239400400108

[20] Luis Cavazos Quero, Jorge Iranzo Bartolomé, Seonggu Lee, En Han, Sunhee Kim, and Jundong Cho. 2018. An interactive multimodal guide to improve art accessibility for blind people. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 346–348. https://doi.org/10.1145/3234695.3241033

[21] Ruei-Che Chang, Chao-Hsien Ting, Chia-Sheng Hung, Wan-Chen Lee, Liang-Jin Chen, Yu-Tzu Chao, Bing-Yu Chen, and Anhong Guo. 2022. OmniScribe: Authoring Immersive Audio Descriptions for 360 Videos. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14. https://doi.org/10.1145/3526113.3545613

[22] Neha Chintala. 2023. *From Blind Driving Assists to One Touch Driving, Meet The Most Accessible Forza Motorsport Ever.* https://news.xbox.com/en-us/2023/04/27/forza-motorsport-accessibility-features-blind-driving/

[23] Agnieszka Chmiel and Iwona Mazur. 2022. A homogenous or heterogeneous audience? Audio description preferences of persons with congenital blindness, non-congenital blindness and low vision. *Perspectives* 30, 3 (2022), 552–567. https://doi.org/10.1080/0907676X.2021.1913198

[24] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020). https://doi.org/10.48550/arXiv.2010.11929

[25] Wendy Duggleby. 2005. What about focus group interaction data? *Qualitative health research* 15, 6 (2005), 832–840. https://doi.org/10.1177/1049732304273916

[26] Equal Entry. 2020. *Audio Descriptions for 360 Degree Video: Best Practices.* https://www.youtube.com/watch?v=jOX6gxUZq8w

[27] Deborah I Fels, John Patrick Udo, Peter Ting, Jonas E Diamond, and Jeremy I Diamond. 2006. Odd Job Jack described: a universal design approach to described video. *Universal Access in the Information society* 5 (2006), 73–81. https://doi.org/10.1007/s10209-006-0025-0

[28] Richard E Ferdig, Karl W Kosko, and Enrico Gandolfi. 2020. The use of ambisonic audio to improve presence, focus, and noticing while viewing 360 video. *Journal For Virtual Worlds Research* 13, 2-3 (2020). https://jvwr-ojs-utexas.tdl.org/jvwr/article/view/7422

[29] Anita Fidyka and Anna Matamala. 2018. Audio description in 360° videos: Results from focus groups in Barcelona and Kraków. *Translation Spaces* 7, 2 (2018), 285–303. https://doi.org/10.1075/ts.18018.fid

[30] Anita Fidyka and Anna Matamala. 2021. Retelling narrative in 360° videos: Implications for audio description. *Translation Studies* 14, 3 (2021), 298–312. https://doi.org/10.1080/14781700.2021.1888783

[31] Anita Fidyka, Anna Matamala, Olga Soler Vilageliu, and Blanca Arias-Badia. 2021. Audio description in 360° content: results from a reception study. *Skase Journal of Translation and Interpretation* 14, 1 (2021), 14–32. http://www.skase.sk/Volumes/JTI20/pdf_doc/02.pdf

[32] Langis Gagnon, Claude Chapdelaine, David Byrns, Samuel Foucher, Maguelonne Heritier, and Vishwa Gupta. 2010. A computer-vision-assisted system for videodescription scripting. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 41–48. https://doi.org/10.1109/CVPRW.2010.5543575

[33] David Gonçalves, Manuel Piçarra, Pedro Pais, João Guerreiro, and André Rodrigues. 2023. "My Zelda Cane": Strategies Used by Blind Players to Play Visual-Centric Digital Games. In *Proceedings of the 2023 CHI conference on human factors in computing systems*. 1–15. https://doi.org/10.1145/3544548.3580702

[34] Ricardo E Gonzalez Penuela, Wren Poremba, Christina Trice, and Shiri Azenkot. 2022. Hands-On: Using Gestures to Control Descriptions of a Virtual Environment for People with Visual Impairments. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–4. https://doi.org/10.1145/3526114.3558669

[35] Nick Grudin. 2015. *Introducing 360 Video on Facebook.* https://www.facebook.com/formedia/blog/introducing-360-video-on-facebook

[36] João Guerreiro, Yujin Kim, Rodrigo Nogueira, SeungA Chung, André Rodrigues, and Uran Oh. 2023. The Design Space of the Auditory Representation of Objects and Their Behaviours in Virtual Reality for Blind People. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2763–2773. https://doi.org/10.1109/TVCG.2023.3247094

[37] James Herndon and Chancey Fleet. 2020. *Audio Descriptions for 360-Degree Video: Recap of Webinar.* https://equalentry.com/audio-descriptions-for-360-degree-video-recap/

[38] ImAc Project. 2019. *Audio description for 360° content.* https://www.imacproject.eu/2019/09/02/audio-description-for-360-content/

[39] ImAc Project. 2019. *Audio description in 3D audio.* https://www.imacproject.eu/2019/01/14/audio-description-in-3d-audio/

[40] Imperial Potato. 2022. *The Super Mario Bros Movie 360/VR Experience.* https://www.youtube.com/watch?v=ykLs1GohDFE

[41] Gaurav Jain, Basel Hindi, Connor Courtien, Conrad Wyrick, Xin Yi Therese Xu, Michael C Malcolm, and Brian A Smith. 2023. Towards Accessible Sports Broadcasts for Blind and Low-Vision Viewers. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7. https://doi.org/10.1145/3544549.3585610

[42] Lucy Jiang and Richard Ladner. 2022. Co-Designing Systems to Support Blind and Low Vision Audio Description Writers. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–3. https://doi.org/10.1145/3517428.3550394

[43] Gabriella M Johnson and Shaun K Kane. 2020. Game changer: accessible audio and tactile guidance for board and card games. In *Proceedings of the 17th International Web for All Conference*. 1–12. https://doi.org/10.1145/3371300.3383347

[44] Finn Kensing and Jeanette Blomberg. 1998. Participatory design: Issues and concerns. *Computer supported cooperative work (CSCW)* 7, 3 (1998), 167–185. https://doi.org/10.1023/A:1008689307411

[45] Christopher A Le Dantec and Sarah Fox. 2015. Strangers at the gate: Gaining access, building rapport, and co-constructing community-based research. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 1348–1358. https://doi.org/10.1145/2675133.2675147

[46] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. Tell me where to look: Investigating ways for assisting focus in 360 video. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 2535–2545. https://doi.org/10.1145/3025453.3025757

[47] Yung-Ta Lin, Yi-Chi Liao, Shan-Yuan Teng, Yi-Ju Chung, Liwei Chan, and Bing-Yu Chen. 2017. Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 255–265. https://doi.org/10.1145/3126594.3126656

[48] Xingyu Liu, Patrick Carrington, Xiang'Anthony' Chen, and Amy Pavel. 2021. What Makes Videos Accessible to Blind and Visually Impaired People?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14. https://doi.org/10.1145/3411764.3445233

[49] Xingyu "Bruce" Liu, Ruolin Wang, Dingzeyu Li, Xiang Anthony Chen, and Amy Pavel. 2022. CrossA11y: Identifying Video Accessibility Issues via Cross-modal Grounding. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–14. https://doi.org/10.1145/3526113.3545703

[50] Rachael Luck. 2003. Dialogue in participatory design. *Design studies* 24, 6 (2003), 523–535. https://doi.org/10.1016/S0142-694X(03)00040-1

[51] Carme Mangiron and Xiaochun Zhang. 2016. Game accessibility for the blind: Current overview and the potential application of audio description as the way forward. *Researching audio description: New approaches* (2016), 75–95. https://doi.org/10.1057/978-1-137-56917-2_5

[52] Carme Mangiron and Xiaochun Zhang. 2022. Video games and audio description. In *The Routledge Handbook of Audio Description*. Routledge, 377–390. https://doi.org/10.4324/9781003003052-29

[53] Troy McDaniel, Lakshmie Narayan Viswanathan, and Sethuraman Panchanathan. 2013. An evaluation of haptic descriptions for audio described films for individuals who are blind. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6. https://doi.org/10.1109/ICME.2013.6607554

[54] Meta. 2023. *Health & Safety Warnings.* https://www.meta.com/legal/quest/health-and-safety-warnings/

[55] Meta. 2023. *Meta Quest 2 Health & Safety Manual.* https://www.oculus.com/safety-center/quest-2/

[56] Meredith Ringel Morris, Jazette Johnson, Cynthia L Bennett, and Edward Cutrell. 2018. Rich representations of visual content for screen reader users. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–11. https://doi.org/10.1145/3173574.3173633

[57] Cecily Morrison, Edward Cutrell, Anupama Dhareshwar, Kevin Doherty, Anja Thieme, and Alex Taylor. 2017. Imagining artificial intelligence applications with people with visual disabilities using tactile ideation. In *Proceedings of the 19th international acm sigaccess conference on computers and accessibility*. 81–90. https://doi.org/10.1145/3132525.3132530

[58] Martez E Mott, John Tang, and Edward Cutrell. 2023. Accessibility of Profile Pictures: Alt Text and Beyond to Express Identity Online. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13. https://doi.org/10.1145/3544548.3580710

[59] Annika Muehlbradt and Shaun K Kane. 2022. What's in an ALT Tag? Exploring Caption Content Priorities through Collaborative Captioning. *ACM Transactions on Accessible Computing (TACCESS)* 15, 1 (2022), 1–32. https://doi.org/10.1145/3507659

[60] Rosiana Natalie, Ebrima Jarjue, Hernisa Kacorri, and Kotaro Hara. 2020. Viscene: A collaborative authoring tool for scene descriptions in videos. In *Proceedings of*

the 22nd International ACM SIGACCESS Conference on Computers and Accessibility. 1–4. https://doi.org/10.1145/3373625.3418030

[61] Rosiana Natalie, Jolene Loh, Huei Suen Tan, Joshua Tseng, Ian Luke Yi-Ren Chan, Ebrima H Jarjue, Hernisa Kacorri, and Kotaro Hara. 2021. The efficacy of collaborative authoring of video scene descriptions. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15. https://doi.org/10.1145/3441852.3471201

[62] Rosiana Natalie, Jolene Loh, Huei Suen Tan, Joshua Tseng, Hernisa Kacorri, and Kotaro Hara. 2021. Uncovering patterns in reviewers' feedback to scene description authors. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4. https://doi.org/10.1145/3441852.3476550

[63] Rosiana Natalie, Joshua Tseng, Hernisa Kacorri, and Kotaro Hara. 2023. Supporting Novices Author Audio Descriptions via Automatic Feedback. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18. https://doi.org/10.1145/3544548.3581023

[64] Anthony J Onwuegbuzie, Wendy B Dickinson, Nancy L Leech, and Annmarie G Zoran. 2009. A qualitative framework for collecting and analyzing data in focus group research. *International journal of qualitative methods* 8, 3 (2009), 1–21. https://doi.org/10.1177/160940690900800301

[65] OpenAI. 2023. *GPT-4*. https://openai.com/product/gpt-4

[66] Amy Pavel, Gabriel Reyes, and Jeffrey P Bigham. 2020. Rescribe: Authoring and automatically editing audio descriptions. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 747–759. https://doi.org/10.1145/3379337.3415864

[67] Yi-Hao Peng, Jeffrey P Bigham, and Amy Pavel. 2021. Slidecho: Flexible non-visual exploration of presentation videos. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12. https://doi.org/10.1145/3441852.3471234

[68] André Queirós, Daniel Faria, and Fernando Almeida. 2017. Strengths and limitations of qualitative and quantitative research methods. *European journal of education studies* (2017). https://oapub.org/edu/index.php/ejes/article/view/1017/2934

[69] RNIB. 2018. *Audio description for 360-degree content*. https://www.rnib.org.uk/news/audio-description-for-360-degree-content/

[70] Elizabeth B-N Sanders, Eva Brandt, and Thomas Binder. 2010. A framework for organizing the tools and techniques of participatory design. In *Proceedings of the 11th biennial participatory design conference*. 195–198. https://doi.org/10.1145/1900441.1900476

[71] Alexa Siu, Gene SH Kim, Sile O'Modhrain, and Sean Follmer. 2022. Supporting accessible data visualization through audio data narratives. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–19. https://doi.org/10.1145/3491102.3517678

[72] Joel Snyder. 2005. Audio description: The visual made verbal. In *International Congress Series*, Vol. 1282. Elsevier, 935–939. https://doi.org/10.1016/j.ics.2005.05.215

[73] Abigale Stangl, Shasta Ihorn, Yue-Ting Siu, Aditya Bodi, Mar Castanon, Lothar Narins, and Ilmi Yoon. 2023. The Potential of a Visual Dialogue Agent In a Tandem Automated Audio Description System for Videos. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–16.

[74] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. Person, Shoes, Tree. Is the Person Naked? What People with Vision Impairments Want in Image Descriptions. In *Proceedings of the 2020 chi conference on human factors in computing systems*. 1–13. https://doi.org/10.1145/3313831.3376404

[75] Abigale Stangl, Nitin Verma, Kenneth R Fleischmann, Meredith Ringel Morris, and Danna Gurari. 2021. Going Beyond One-Size-Fits-All Image Descriptions to Satisfy the Information Wants of People Who are Blind or Have Low Vision. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–15. https://doi.org/10.1145/3441852.3471233

[76] The American Council of the Blind. 2023. *All About Audio Description*. https://adp.acb.org/ad.html

[77] The American Council of the Blind. 2023. *The Audio Description Project*. https://adp.acb.org/

[78] John Patrick Udo, Bertha Acevedo, and Deborah I Fels. 2010. Horatio audio-describes Shakespeare's Hamlet: Blind and low-vision theatre-goers evaluate an unconventional audio description strategy. *British Journal of Visual Impairment* 28, 2 (2010), 139–156. https://doi.org/10.1177/0264619609359753

[79] John-Patrick Udo and Deborah I Fels. 2010. Enhancing the entertainment experience of blind and low-vision theatregoers through touch tours. *Disability & Society* 25, 2 (2010), 231–240. https://doi.org/10.1080/09687590903537497

[80] John-Patrick Udo and Deborah I Fels. 2010. The rogue poster-children of universal design: Closed captioning and audio description. *Journal of Engineering Design* 21, 2-3 (2010), 207–221. https://doi.org/10.1080/09544820903310691

[81] Sanjeev Verma. 2015. *360-degree videos now on Google Cardboard and iOS*. https://blog.youtube/news-and-events/360-degree-videos-now-on-google/

[82] Sanjeev Verma. 2015. *A new way to see and share your world with 360-degree video*. https://blog.youtube/news-and-events/a-new-way-to-see-and-share-your-world/

[83] Lakshmie Narayan Viswanathan, Troy McDaniel, Sreekar Krishna, and Sethuraman Panchanathan. 2010. Haptics in audio described movies. In *2010 IEEE International Symposium on Haptic Audio Visual Environments and Games*. IEEE, 1–2. https://doi.org/10.1109/HAVE.2010.5623958

[84] Lakshmie Narayan Viswanathan, Troy McDaniel, and Sethuraman Panchanathan. 2011. Audio-haptic description in movies. In *International Conference on Human-Computer Interaction*. Springer, 414–418. https://doi.org/10.1007/978-3-642-22098-2_83

[85] VR Planet. 2022. *360° VR || Avatar 2: The Way of Water*. https://www.youtube.com/watch?v=V8Abp4MEg80

[86] Yujia Wang, Wei Liang, Haikun Huang, Yongqi Zhang, Dingzeyu Li, and Lap-Fai Yu. 2021. Toward automatic audio description generation for accessible videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12. https://doi.org/10.1145/3411764.3445347

[87] Beste F Yuksel, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Soo Jung Kim, Joshua Junhee Lee, Seung Jung Jin, Yue-Ting Siu, Joshua A Miele, and Ilmi Yoon. 2020. Human-in-the-Loop Machine Learning to Increase Video Accessibility for Visually Impaired and Blind Users. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 47–60. http://doi.org/10.1145/3357236.3395433

[88] 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. https://doi.org/10.1145/3290605.3300341