



Cognitive Load as a Hidden Cost of Empathy: Tradeoffs in Using VR for Inclusive Design

Dawei Xiong, Zhijun Ma, Hao Ling, Xun Wu, Mingming Fan & Xu Zhang

To cite this article: Dawei Xiong, Zhijun Ma, Hao Ling, Xun Wu, Mingming Fan & Xu Zhang (24 Sep 2026): Cognitive Load as a Hidden Cost of Empathy: Tradeoffs in Using VR for Inclusive Design, International Journal of Human-Computer Interaction, DOI: [10.1080/10447318.2026.2730903](https://doi.org/10.1080/10447318.2026.2730903)

To link to this article: <https://doi.org/10.1080/10447318.2026.2730903>



Published online: 24 Sep 2026.



Submit your article to this journal [↗](#)



Article views: 70



View related articles [↗](#)



View Crossmark data [↗](#)



Cognitive Load as a Hidden Cost of Empathy: Tradeoffs in Using VR for Inclusive Design

Dawei Xiong^a , Zhijun Ma^b , Hao Ling^a , Xun Wu^a , Mingming Fan^{b,c}  and Xu Zhang^a 

^aInnovation, Policy and Entrepreneurship Thrust, Society Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China; ^bComputational Media and Arts Thrust, Information Hub, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China; ^cDivision of Integrative Systems and Design, The Hong Kong University of Science and Technology, Hong Kong, China

ABSTRACT

Virtual reality (VR) is used to bridge the “empathy gap” in inclusive design, but whether emotionally evocative simulations improve design outcomes remains unclear. We conducted a between-subjects experiment comparing passive video with non-emotional and emotional interactive VR simulations of low-vision mobility challenges, informed by interviews and co-design with low-vision participants. Emotional intensity was manipulated via non-player-character feedback while task structure and task-relevant information were held constant. We measured empathy change and assessed design ideas by alignment with user-prioritized needs. Both VR conditions increased empathy relative to video, but emotional VR produced no additional measurable gain, whereas non-emotional VR produced stronger user-need-aligned ideation than both other conditions. Drawing on cognitive load theory, we propose that emotionally charged cues may introduce extraneous demands during ideation, offering a working hypothesis for this divergence. Our work provides guidelines for calibrating VR emotional intensity to balance empathic engagement and design ideation.

KEYWORDS

Virtual reality; empathy; inclusive design; cognitive load; accessibility

1. Introduction

Inclusive design aims to create products and services that are accessible and usable by the widest possible range of people, yet its success depends on bridging the “empathy gap” between designers and users with experiences vastly different from their own (Motahar et al., 2023). This gap is particularly pronounced when designing for individuals with accessibility needs (Mack et al., 2021; Righi et al., 2017). With 2.2 billion people globally living with vision impairment (World Health Organization, 2023), developing solutions that address their needs remains an important design challenge. Designers who typically rely on sight may find it difficult to understand the lived realities of navigating the world with low vision. Traditional empathy-building methods, such as interviews or personas, may not fully convey the embodied and often non-verbalized challenges users face, potentially leading to solutions based on incomplete assumptions (Bennett & Rosner, 2019; Motahar et al., 2023).

To address the limitations of traditional methods, the human-computer interaction (HCI) community has increasingly turned to virtual reality (VR) as a potential solution. Often dubbed an “empathy machine,” VR offers a unique capacity for first-person embodiment, allowing individuals to virtually step into another person’s shoes and experience a simulation of their world (Bevan et al., 2019). This capability is grounded in the principles of embodied cognition, which posit that understanding emerges not from abstract thought alone, but through the body’s direct interaction with an environment (Clark, 1998). By simulating the sensory and physical conditions of others, VR can facilitate a

deeper, more intuitive form of perspective-taking than is possible with passive media. Research has demonstrated VR's effectiveness in fostering empathy toward marginalized groups, such as homeless populations (Herrera et al., 2018) and children with auditory accessibility needs (Embøl et al., 2021), suggesting its immense potential for design.

Despite VR's promise, a common assumption in academic and popular discourse is that more empathy, particularly the emotional resonance evoked by immersive media, leads to better design outcomes. This assumption has encouraged the deliberate use of emotional cues to heighten affective empathy. VR studies frequently incorporate expressive facial and bodily displays, evocative soundscapes, or first-person testimonial narratives to elicit emotional resonance (Barbot & Kaufman, 2020; Bujic et al., 2020; Van Loon et al., 2018). Such cues are intended to increase immersion and evoke compassion or distress, thereby strengthening empathic concern.

However, empathy comprises both cognitive (perspective-taking) and affective (emotional sharing) components (Davis, 1983). While interrelated, these dimensions can have distinct and sometimes opposing effects on problem-solving. Psychological research suggests that heightened emotional engagement can, in some cases, impair social-cognitive reasoning and contribute to biased judgments (Bloom, 2017; Kanske et al., 2016). This possibility is consistent with dual-process theories that distinguish between intuitive thinking (System 1) and deliberate, analytical cognition (System 2) (Kahneman, 2011). This motivates our research question: *Do the emotional components of VR simulations, designed to maximize immersion and affective empathy, actually help or hinder the analytical design process?*

To investigate this tension, this article explores CLT (Sweller, 2011) as an interpretive lens for understanding possible tradeoffs in empathy-driven VR experiences. We focus on the downstream effects of adding emotionally charged cues—such as sounds of frustration or impatient feedback—to an embodied VR simulation. We conducted a between-subjects experiment ($N=90$) in which novice designers experienced the challenges of people with vision impairment under three conditions: a passive video, an interactive but non-emotional VR simulation, and an emotionally charged VR simulation. Our findings reveal a nuanced pattern: both VR conditions increased empathy relative to passive video, and the emotional VR condition did not produce additional measurable empathy gains over the non-emotional VR condition. However, participants in the non-emotional VR condition produced a higher proportion of solutions aligned with the user-prioritized needs identified in our formative study than participants in both other conditions. We interpret this pattern as consistent with a CLT-based working hypothesis: emotionally charged features may add cognitive or affective demands that interfere with translating empathic understanding into user-need-aligned design ideas.

In summary, this research provides a critical and nuanced examination of VR's role in inclusive design, moving beyond the “empathy machine” metaphor to offer a more calibrated perspective. We present three primary contributions:

1. We empirically disentangle embodied VR from affective intensification by showing that both VR conditions increased empathy relative to passive video, while emotional cues did not yield additional measurable empathy gains over non-emotional VR.
2. We identify a downstream design cost associated with emotionally intensified VR. Non-emotional VR produced stronger user-need-aligned design ideation than both emotional VR and passive video, suggesting that emotional intensification does not necessarily improve downstream design outcomes.
3. We develop a CLT-informed interpretation and practical guidelines for calibrating VR empathy tools so that they support empathic understanding without imposing unnecessary demands on the cognitive work required for user-centered design.

Ultimately, this article provides both empirical evidence and a theoretical lens to guide the more thoughtful use of immersive technologies in inclusive design. By understanding the possible tradeoffs between empathic engagement and downstream cognitive demands, we can better engineer experiences that not only allow designers to feel for others but also help them think more clearly on users' behalf.

2. Related work

To situate our research, we draw on three domains of literature. First, we review research on VR as a tool for embodied perspective-taking and its role in empathy-oriented experiences. Second, we examine the complex role of empathy in design, including the distinction between its cognitive and affective components and the potential downsides of excessive emotional resonance. Finally, we introduce dual-process theory and CLT, which provide a conceptual lens for considering how emotionally evocative VR experiences may shape the cognitive demands of downstream design work. Together, these domains frame our examination of the practical tradeoffs involved in using VR to support empathy in design.

2.1. VR for perspective-taking

VR's capacity to foster empathy is often attributed to presence—the psychological feeling of being physically located in a virtual environment (Slater & Sanchez-Vives, 2016). Unlike traditional media that present experiences from a third-person viewpoint, VR facilitates first-person, embodied perspective-taking (Bevan et al., 2019; Chatain et al., 2023). This capability is grounded in the principles of embodied cognition, which posit that understanding emerges not from abstract thought alone, but through the body's direct interaction with an environment (Barsalou, 2008; Clark, 1998; Gallagher, 2005). By allowing users to experience another person's perspective from a first-person viewpoint, VR may support forms of understanding that are difficult to achieve through abstract knowledge alone (Ahn et al., 2013).

A recent meta-analysis and related empirical studies support VR's potential to foster empathy (Ventura et al., 2020). Studies have shown that immersive VR experiences can produce stronger and more enduring empathic responses and positive attitudes toward marginalized groups, including individuals experiencing homelessness (Herrera et al., 2018), older adults (Youn Oh et al., 2016), and children who are deaf or hard of hearing (Embøl et al., 2021), compared to traditional perspective-taking exercises. Interactivity and agency within a virtual environment can support embodied learning beyond passive observation (Bertrand et al., 2018). The goal of such interventions is often to bridge the “empathic accuracy problem,” where the fidelity of one's understanding of another decreases as their experiential differences increase (Packard & Burnham, 2021).

HCI research has moved beyond the foundational question of whether VR can enhance empathy to a more nuanced investigation of its application boundaries and design strategies. On one hand, studies on the social acceptability of deploying VR in real-world contexts, such as public transit, inform the situational boundaries for transitioning such tools from controlled labs to authentic environments (Williamson et al., 2019). On the other hand, at the level of internal experience design, narrative and interaction models are critical factors shaping the nature and intensity of empathy (Barbot & Kaufman, 2020; Bevan et al., 2019). This shift is also reflected in review and intervention-oriented work: reviews of VR-based empathy-training software and broader meta-analyses show that VR can support empathy and prejudice-reduction goals, but they also emphasize variability in design, measurement, and context (Christofi & Michael-Grigoriou, 2017; Khukalenko & Khanolainen, 2025; Schutte & Stilinović, 2017; Ventura et al., 2020). Consequently, the strategic design of VR experiences to support constructive empathy has become a central consideration across education, healthcare, and training contexts (Iqbal et al., 2024; Khukalenko & Khanolainen, 2025; Ventura et al., 2020).

At the same time, the “empathy machine” framing has been criticized for implying that immersion automatically produces ethically or practically useful understanding. Work on VR nonfiction and social-impact experiences cautions that such simulations can simplify others' lived experiences, over-emphasize spectatorship, or make emotional intensity seem like the primary evidence of understanding (Bevan et al., 2019; Hassan, 2020; Sora-Domenjó, 2022; Wimmer, 2020). Broader critiques of VR imaginaries similarly caution against treating immersive media as offering direct or frictionless access to others' experiences (Carter & Egliston, 2024). Recent HCI critiques also argue that empathy-centered design should be assessed beyond individual feeling states and should attend to social context, representation, and downstream action (Debnath et al., 2024).

In summary, prior work establishes VR as an important affordance for perspective-taking, while also showing that empathy-oriented VR requires careful design and evaluation. However, less is known about how VR-mediated empathy translates into performance on complex downstream tasks that require analytical problem-solving, such as design ideation. We address this gap by examining downstream design ideation alongside empathy outcomes.

2.2. Empathy in design: Cognitive versus affective components

Empathy is a cornerstone of human-centered design (Alessandro Gasparini, 2015; Mauri et al., 2022; Wright & McCarthy, 2008), yet its role is more complex than the simple axiom “more is better” would suggest. Psychological literature distinguishes between two primary dimensions of empathy: cognitive empathy (the ability to understand another’s perspective) and affective empathy (sharing another’s emotional state) (Davis, 1983; Decety & Jackson, 2004). These components are supported by distinct neural circuits (Eres et al., 2015; Shamay-Tsoory et al., 2009) and can have opposing effects on problem-solving. These distinctions highlight empathy’s potential to both connect designers with users and cloud judgment.

Recent work has begun to challenge the unqualified praise of empathy in creative and professional domains, suggesting that an over-reliance on affective empathy can be detrimental. High levels of affective resonance can lead to empathic distress, emotional exhaustion, withdrawal, or avoidance rather than constructive engagement, which can impair rational decision-making and even lead to inaction (Bloom, 2017; Decety, 2021; Sora-Domenjó, 2022). For instance, research in entrepreneurship found an inverted U-shaped relationship between empathy and new product development performance, suggesting that excessive empathy can introduce cognitive biases and obscure an objective assessment of user needs (Kurz et al., 2024). This is consistent with findings that being a strong emotional empathizer does not necessarily make one a better “mentalizer” or problem-solver (Kanske et al., 2016). The challenge for designers is to balance empathic understanding with systematic analysis to avoid such pitfalls (Coleman et al., 2008).

Within HCI and design research, scholars have examined these risks from both methodological and psychological perspectives. Research in experience design identifies emotion as central to meaningful experiences (Hassenzahl et al., 2010), while design scholars caution that empathic engagement does not necessarily ensure an accurate understanding of users’ lived experiences (Bennett & Rosner, 2019; Heylighen & Dong, 2019). Structured methods such as Dialogue-Labs can help designers process and translate experiential insights (Lucero et al., 2012). However, such methods primarily address what happens after an empathic encounter. Less attention has been paid to how the empathic stimulus itself can be composed to support subsequent design work without overwhelming it.

Thus, the existing literature suggests a potential tension between emotional intensification during empathic experiences and the analytical clarity required for design ideation. Heightened emotional states can interfere with mentalizing and emotion regulation, potentially reducing the analytical clarity available for subsequent tasks (Kanske et al., 2016; Thompson et al., 2022). Despite this, little research has experimentally investigated this tension within the context of VR-based design tools, where emotional immersion is often treated as a design goal. Our study addresses this gap by experimentally comparing emotional and non-emotional embodied VR experiences to examine how emotional intensification shapes subsequent user-need-aligned ideation.

2.3. Theoretical foundations: Dual-process and CLT

To develop a theoretical framing for the potential effects of emotional VR simulations, we draw on two complementary cognitive theories. Dual-process theory distinguishes between System 1, which is fast, intuitive, and emotional, and System 2, which is slow, deliberate, and analytical (Kahneman, 2011). An emotional VR simulation may engage intuitive and affective processing, whereas design ideation requires deliberate analysis of constraints and the structuring of solutions. From this perspective, strong intuitive or emotional responses may interfere with the more resource-intensive processing required for subsequent ideation.

CLT provides a more fine-grained lens for considering how such interference might occur (Sweller, 2011). CLT models working memory as a limited resource and suggests that effective learning and problem-solving depend on managing this load. Total cognitive load is commonly described in terms of intrinsic load (inherent task difficulty), extraneous load (demands imposed by the way information is presented), and germane load (resources devoted to learning and schema construction) (Kalyuga, 2008). In the present study, CLT helps specify a plausible account: negative emotional stimuli within a VR simulation may introduce additional demands that are extraneous to the subsequent ideation task. If present, these demands could operate similarly to anxiety by consuming finite working memory resources (Eysenck et al., 2007) and leaving fewer resources available for analyzing users' problems and generating effective solutions. Accordingly, we use this account as a conceptual basis for interpreting the observed pattern and for motivating future work that directly measures cognitive load alongside affective and attentional processes.

The HCI community has incorporated this theory into its research agenda (Hollender et al., 2010), and recent work offers converging evidence that task demands can compete for limited cognitive resources. For example, work in automotive HCI has demonstrated how situational demands and secondary tasks, such as gaze or speech input under driving pressure, deplete cognitive resources (Roider et al., 2017). This work provides a useful analogy for considering how emotionally evocative stimuli in VR may act like additional task demands that compete with the analytical resources required for design.

Together, these theories provide a conceptual lens for our investigation, moving beyond whether VR increases empathy to considering how emotional intensification may affect downstream design ideation. They allow us to articulate a testable account of the observed pattern. Because cognitive load was not measured directly, we treat this account as an interpretive framework and as motivation for future work that measures cognitive load alongside affective and attentional processes.

3. Formative study

To ground our experimental design in the authentic, lived experiences of individuals with vision impairment, we first conducted a formative study. This study employed a two-phase, mixed-method approach designed to first explore the problem space broadly and then validate our initial understanding. The first phase consisted of semi-structured interviews to identify key mobility challenges. The second phase involved co-design workshops to verify that our interpretation of these challenges resonated with the user community.

3.1. Participants

We recruited 10 adults (3 females, 7 males) diagnosed with Grade 1 low vision, as defined by the China Disabled Persons' Federation.¹ Participants were recruited through partnerships with local disability service organizations and community centers. Ages ranged from 18 to 60 ($M = 35.6, SD = 10.59$). The duration of their diagnosed low vision condition ranged from 5 to over 20 years. All participants provided informed written consent, and the study was approved by our institution's Institutional Review Board (IRB). Participants were compensated for their time.

3.2. Procedure and data analysis

One-on-one, semistructured interviews were conducted by two researchers, with each session lasting between 45 and 60 min. The interview protocol focused on three primary areas: navigation in familiar and unfamiliar environments, strategies for obstacle detection, and challenges encountered while using public transportation. Interviews were audio-recorded, professionally transcribed, and then analyzed using a thematic analysis approach. Two researchers independently coded the transcripts to identify recurring themes and pain points, meeting regularly to discuss codes and resolve discrepancies until a consensus was reached.

Based on the themes identified from the interviews, we developed a set of five preliminary user scenarios depicting common challenges. We then invited four of the original participants to a 90-minute

co-design workshop. In the workshop, we presented these scenarios and facilitated a discussion to gather feedback on their authenticity and to collaboratively refine details, ensuring our interpretations were accurate and representative.

3.3. Key findings

Our analysis revealed several findings that were decisive in shaping our subsequent research design. First, the interviews identified two core scenarios in daily mobility that were consistently described as the most challenging and emotionally evocative. The first of these is locating a specific shop. One participant described that even when navigation tools led him to the general vicinity, “the words on the shop signs are completely invisible to me, which is a huge obstacle.” He explained that this process of searching without being able to confirm the target filled him with a sense of “pain and urgency,” as well as “anxiety and confusion.” The second scenario is identifying a moving bus. A participant noted that due to his narrow field of vision, a moving bus remains in his sight for only a very brief time, causing “high tension.” This is compounded by difficulties in communicating with the driver, as he is unable to perceive non-verbal cues like a nod or a head shake. He described this situation: “If he just nods or shakes his head slightly ... I probably can’t perceive that,” which often leaves him in a state of “confusion and uncertainty.” These two recurring and emotionally charged scenarios became the foundation for our VR experiences.

Crucially, the study not only identified problems but also revealed an overwhelming user preference for a specific solution modality, which provided a basis for our subsequent evaluation metric. Across the co-design session, participants consistently positioned auditory solutions as the primary, and often only, viable option. One participant stated this explicitly during the co-design session, summarizing that “the audio solution is always the most perfect solution.” This sentiment was echoed by another, who emphasized that “all other information not related to sound is blur and unreliable.” This unanimous reliance on auditory information corroborates earlier findings (Zhao et al., 2020) and suggests that proposing an audio-focused solution can serve as a targeted indicator of whether designers recognized the user-prioritized sensory channel in these mobility scenarios.

3.4. Impact on subsequent research design

These findings were instrumental in shaping the methodology of our main experiment in two specific, interconnected ways. First, the pain points related to transit and obstacle detection directly informed the creation of our two VR scenarios; we simulated a bus-finding task in a terminal and navigation on a cluttered street to ensure our experimental environment was grounded in authentic, user-validated challenges. Second, and critically for our evaluation framework, the strong preference for auditory solutions provided a clear, ecologically valid metric for assessing user-need alignment in design ideation. We operationalized scenario-specific alignment with user needs as the generation of auditory-focused solutions. This allowed us to move beyond simple solution counts to a targeted measure of user-centered thinking derived directly from the stated preferences of our target user group.

4. User study

To examine how VR and emotional stimuli affect empathy and design ideation, we conducted a between-subjects laboratory experiment. The design compares three media experiences while keeping the experimental tasks grounded in the lived experiences of individuals with low vision, and it is structured to separate the effect of embodiment from the effect of emotional intensification. We describe below the participants, the apparatus and virtual environment, the experimental conditions, the session procedure, and our data collection and analysis.

4.1. Participants

We recruited $N = 90$ participants (46 females, 44 males; age $M = 23.2$, $SD = 2.6$) from the university community (Figure 1). All participants had received prior training in design thinking as part of their

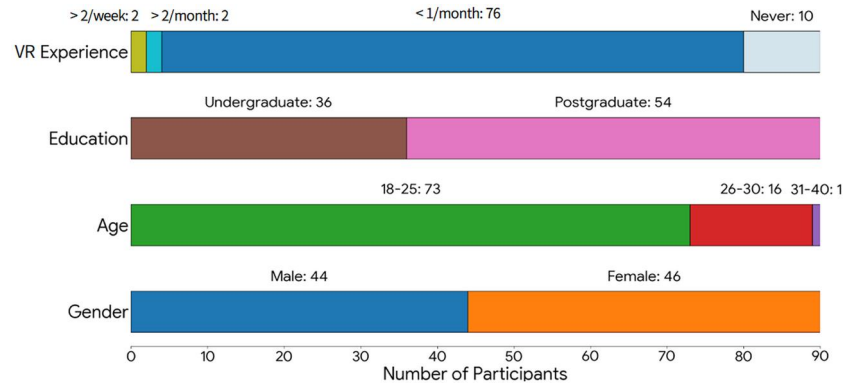


Figure 1. Participant demographics by gender, age group, education level, and prior VR experience.

curriculum, giving them a working understanding of user-centered problem-solving and positioning them as novice designers. All reported normal or corrected-to-normal vision and had no prior personal experience with low-vision conditions. Participants received a stipend for their time, and the study was approved by our institution's Institutional Review Board.

4.2. Apparatus and experimental scenario

The study was conducted in a controlled lab setting. Participants in the VR conditions used a Pico 4 Head-Mounted Display (HMD) with hand-held controllers for navigation and interaction. The video condition was administered on a standard desktop computer.

The experimental scenario, which was consistent across all conditions, simulated the daily mobility challenges of an individual with Grade 1 low vision. Informed by the formative study (Section 3), the scenario comprised two main tasks grounded in participants' reported pain points: (1) navigating a public transit hub to find a specific bus (Figure A1(a,b)) and (2) navigating a cluttered commercial street to locate a shop (Figure A1(c,d)). The virtual environment was populated with interactive non-player characters (NPCs), such as a bus driver and service attendants, with whom participants could engage for assistance.

4.3. Experimental design and conditions

The study employed a between-subjects design, with participants randomly assigned to one of three experimental conditions ($n = 30$ per condition). This design allows for a direct comparison between a traditional, passive medium (video) and an interactive, embodied medium (VR), as well as an isolated examination of the effect of emotional stimuli within the VR experience. Because the non-emotional and emotional VR conditions hold the embodied interaction and task-relevant information constant, their contrast isolates the effect of emotional intensification from the effect of embodiment itself.

The three conditions were as follows:

- *C1: passive video.* Participants watched a first-person video recording of an individual completing the tasks in the VR simulation. They were instructed to imagine themselves in that person's role, and had no control over actions or viewpoint.
- *C2: non-emotional VR.* Participants were immersed in the interactive low-vision simulation, navigating freely with controllers and engaging with NPCs. The simulation presented the functional challenges of low vision without additional emotional cues.
- *C3: emotional VR.* This condition matched C2 in task structure, scene layout, task-relevant environmental cues, and task-relevant information, and added emotional framing through differences in NPC response wording and vocal tone. As a secondary cue, it also included brief behavior-contingent diegetic avatar feedback during specific scripted error events, such as a sigh after selecting the wrong bus; these cues provided no additional task-relevant information.

To ensure experimental control, NPC response scripts were standardized within each VR condition. When participants initiated the corresponding interaction, participants in the same condition received the same NPC response script, including the same task-relevant content and emotional tone. Across the non-emotional and emotional VR conditions, the task structure and scenario elements were held constant, including the bus route number, shop location, scene layout, task-relevant environmental cues, and the functional information provided by drivers and service attendants. The primary experimental manipulation concerned the emotional framing of NPC feedback, while task structure and task-relevant information were held constant.

For example, in English translation, when participants asked the driver whether they had found the correct bus, the emotional VR response was phrased as impatient feedback (e.g., “Of course it is Route 53. Can’t you see?”), whereas the non-emotional VR response provided the same information neutrally (e.g., “Yes, this is Route 53.”). Similarly, in English translation, when participants asked a service attendant at the wrong shop, the emotional VR response used annoyed wording (e.g., “Of course not. Can’t you see?”), whereas the non-emotional VR response provided the same information in a neutral manner (e.g., “Sorry, this is not the place you are looking for.”).

NPC-based emotional cues were embedded in scripted NPC interactions. Participants in both VR conditions were instructed to interact with the relevant NPCs, so the procedure was designed to standardize exposure to scripted assistance and feedback.

4.4. Procedure

Each study session lasted approximately 30–35 min and was facilitated by a single researcher to ensure consistency. The session was structured into these stages (Figure 2):

1. *Introduction and consent (5 min)*: The researcher explained the study’s purpose (framed as market research for a new application) and the procedure. Participants provided informed written consent.
2. *Pre-study questionnaire (5 min)*: Participants completed a questionnaire to gather baseline empathy scores, demographic data, and assess prior VR experience.
3. *VR System Tutorial (5 min)*: This stage was exclusive to participants in the two VR groups (non-emotional and emotional VR). They received a hands-on tutorial for the HMD and controllers to ensure proficiency with navigation and interaction before the main task. Participants in the video group did not undergo this stage.
4. *Task performance (10 min)*: Participants underwent their assigned experimental condition (video, non-emotional VR, or emotional VR). Participants in the two VR conditions were instructed to interact with task-relevant NPCs during the task, which was designed to expose them to the scripted assistance and feedback.
5. *Post-study data collection (10 min)*: Immediately after the task, participants completed the post-test empathy questionnaire. They were then given 10 min for a design ideation task. Participants were instructed to propose innovative methods or solutions for individuals with low vision within this timeframe. To quantify the output, they were asked to mark each distinct idea they generated and report the total number at the end of the 10-minute period.

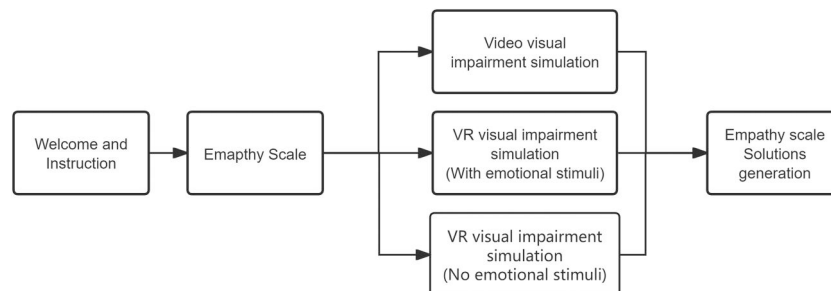


Figure 2. Overview of the experimental setup and procedure. Participants completed pre- and post-test empathy scales before and after being randomly assigned to one of three simulation conditions.

Once the main experiment concluded, we invited video-group participants to experience the non-emotional VR simulation, giving them a basis for comparison. They then took part in a semi-structured interview reflecting on the differences between the two media.

4.5. Data collection and metrics

We collected quantitative questionnaire data, coded design-ideation data, and qualitative reflections from the post-experiment interviews with video-group participants.

4.5.1. Empathy

To measure empathy, we used an 8-item scale adapted from the widely recognized Interpersonal Reactivity Index (Davis, 1983), administered pre- and post-intervention. It consisted of two 4-item subscales: cognitive empathy (e.g., “I could understand the challenges faced by people with low-vision”) and affective empathy (e.g., “I felt concerned for people with low-vision”). Responses were on a 4-point Likert scale. The scale demonstrated good internal consistency (Cronbach’s $\alpha > 0.78$). Our primary outcome was the change in the overall 8-item empathy score (post-test minus pretest). Changes in the cognitive and affective subscale scores were examined in exploratory analyses.

4.5.2. User-need alignment of design solutions

We operationalized participants’ understanding of user needs as the alignment between their proposed solutions and the needs prioritized by low-vision participants in our formative study. Because those participants consistently identified auditory information as critical for the mobility scenarios used in the experiment, we treated an “auditory-focused” solution as a targeted indicator of user-need alignment in this context.

To quantify this, we performed a content analysis of the 198 valid solutions. Following a predefined multi-label coding scheme (shown in Tables B1 and B2), two independent coders assigned one or more of five category labels—“auditory,” “tactile,” “visual,” “psychological,” and “other”—to each valid solution. Thus, a solution could be assigned to multiple categories when it contained more than one type of design intervention. For the reliability analysis, the solution-level labels were aggregated into participant-level category counts. Inter-coder agreement on these counts before disagreement resolution was high (quadratic-weighted Cohen’s $\kappa_w = 0.831$; exact agreement = 87.1%). Disagreements were subsequently resolved through discussion. Our primary metric for user-need alignment was the proportion of auditory-focused solutions among the total valid solutions proposed by each participant. This ratio controls for individual differences in idea quantity and provides a targeted measure of alignment with core user needs, rather than a comprehensive assessment of overall design quality.

5. Findings of the user study

This section reports the results of the experiment. We first present the quantitative analysis comparing the three conditions on empathy change and on the user-need alignment of design solutions, then a qualitative analysis of video-group participants’ reflections after they experienced the non-emotional VR simulation. The reflections are used descriptively to characterize the contrast between passive video and embodied VR.

5.1. Quantitative analysis

We used one-way analyses of variance to test the effect of condition (emotional VR, non-emotional VR, video) on empathy change and on user-need alignment, with Tukey HSD post hoc comparisons.

5.1.1. Effect on empathy development

We first examined changes in empathy scores, calculated as post-test minus pretest. Empathy-change scores were $M = 4.60$ ($SD = 3.63$) for emotional VR, $M = 5.00$ ($SD = 4.11$) for non-emotional VR, and $M = 2.07$ ($SD = 3.54$) for video. ANOVA residuals did not significantly deviate from normality

(Shapiro–Wilk $W = 0.978, p = 0.124$), and the ANOVA revealed a significant omnibus effect of condition, $F(2, 87) = 5.35, p = 0.006, \eta_p^2 = 0.109$.

Tukey HSD comparisons showed that both VR conditions produced greater empathy gains than video. The emotional VR group showed a greater increase than the video group (mean difference = 2.53, 95% CI [0.21, 4.85], Hedges' $g = 0.698$), as did the non-emotional VR group (mean difference = 2.93, 95% CI [0.61, 5.25], Hedges' $g = 0.755$). The two VR conditions did not differ (Hedges' $g = -0.102$). Taken together, both VR conditions produced greater empathy gains than passive video, while the emotional stimuli produced no additional measurable gain over non-emotional VR. Figure 3(a) shows both VR distributions shifted upward relative to video, with considerable overlap among all three distributions.

Exploratory subscale analyses, uncorrected for multiple comparisons, showed that affective empathy differed significantly by condition, $F(2, 87) = 4.70, p = 0.012, \eta_p^2 = 0.097$. Both VR conditions exceeded video (emotional VR–video: 1.23, 95% CI [0.06, 2.41], Hedges' $g = 0.718$; non-emotional VR–video: 1.37, 95% CI [0.19, 2.54], Hedges' $g = 0.705$), with no difference between the two VR conditions (Hedges' $g = -0.063$). Cognitive empathy showed no significant omnibus effect, $F(2, 87) = 2.73, p = 0.071, \eta_p^2 = 0.059$, and no pairwise comparison reached significance. These exploratory results suggest that the overall pattern was more evident for affective than cognitive empathy, while emotional stimuli produced no additional measurable benefit for either component.

5.1.2. Effect on user-need alignment of design solutions

The primary metric for analysis was the proportion of “auditory-focused solutions,” a targeted measure of alignment with the user-prioritized auditory needs identified in the formative study. Group-level auditory proportions were $M = 0.196$ ($SD = 0.353$, $Mdn = 0.000$, $IQR = 0.300$) for emotional VR, $M = 0.489$ ($SD = 0.409$, $Mdn = 0.500$, $IQR = 1.000$) for non-emotional VR, and $M = 0.246$ ($SD = 0.288$, $Mdn = 0.225$, $IQR = 0.458$) for video. A Shapiro–Wilk test indicated that the ANOVA residuals deviated from normality, $W = 0.904, p < 0.001$. However, a one-way ANOVA, which is relatively robust to moderate non-normality under the balanced design ($n = 30$ per group), revealed a significant omnibus effect of the experimental condition, $F(2, 87) = 5.91, p = 0.004, \eta_p^2 = 0.120$. Given the non-normal distribution, a Kruskal–Wallis test was conducted as a non-parametric omnibus sensitivity analysis and indicated an overall difference among the three conditions, $H(2) = 10.74, p = 0.005, \epsilon^2 = 0.100$. Group-specific differences were assessed using the Tukey HSD comparisons reported below.

A Tukey HSD post hoc analysis indicated significant differences between specific groups. Specifically, the non-emotional VR group generated a significantly higher proportion of auditory solutions than both the emotional VR group (mean difference = 0.29, 95% CI [0.08, 0.51], Hedges' $g = 0.758$) and the video group (mean difference = 0.24, 95% CI [0.03, 0.46], Hedges' $g = 0.679$).

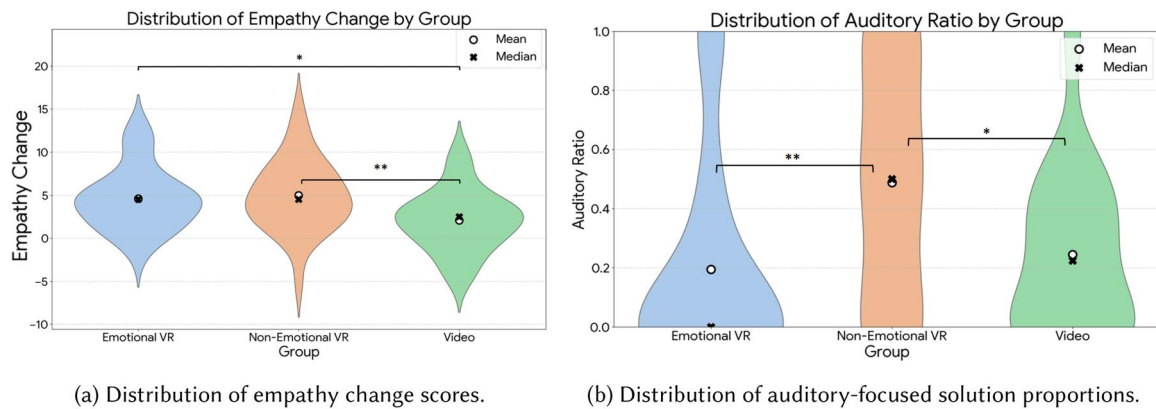


Figure 3. Violin plots of (a) empathy-change scores and (b) auditory-focused solution proportions by condition (the white circle with a black border represents the mean, and the black “X” marks the median). (a) Both VR groups display a visibly higher distribution of scores compared to the video group. (b) The non-emotional VR group displays a higher central tendency in auditory-solution proportion compared to the other two groups. Significance based on a Tukey HSD post hoc test is indicated by asterisks (* $p < 0.05$, ** $p < 0.01$).

Notably, there was no statistically significant difference between the emotional VR and video groups. This result indicates that, in this task, the non-emotional VR experience produced stronger alignment with user-prioritized auditory needs. The distribution in [Figure 3\(b\)](#) shows that both the emotional VR and video groups were concentrated at lower auditory-solution ratios, highlighting how difficult it was for many participants to generate solutions aligned with the low-vision needs identified in the formative study. The non-emotional VR group showed a higher central tendency in auditory-solution proportion, while also displaying participant-level variability. This variability suggests that individual differences may have shaped how participants translated the VR experience into design ideas.

5.1.3. Principal component analysis of solution space

To further explore patterns in the design solutions proposed by the three groups, we conducted a principal component analysis (PCA). While our ANOVA results highlighted the difference in the proportion of “auditory-focused solutions,” PCA provides an exploratory visualization of the broader solution space across the five coded categories: auditory, tactile, visual, psychological, and other.

Each participant was represented by five participant-level counts, corresponding to the number of valid solutions coded as auditory, tactile, visual, psychological, and other. Before PCA, each category was *z*-standardized across participants using its sample mean and sample standard deviation. These counts were not normalized by each participant’s total number of valid solutions; thus, the PCA summarizes variation in category counts, whereas the primary auditory-proportion analysis controls for individual differences in idea quantity.

The first two principal components (PC1 and PC2) accounted for a cumulative variance of 50.9% (PC1: 27.4%, PC2: 23.6%), providing a reasonable two-dimensional summary of the solution space. The component loadings suggested possible interpretations for each axis. PC1 was strongly and positively correlated with visual (0.62) and auditory (0.47) solutions, while being negatively correlated with other (−0.50) solutions. Thus, PC1 can be interpreted as a continuum from “Direct Sensory Interventions” (high PC1 scores) to “Abstract or Social Interventions” (low PC1 scores). PC2 was characterized by a strong positive loading for psychological (0.71) solutions and a negative loading for other (−0.51) and auditory (−0.36) solutions, suggesting a possible dimension of “interpersonal focus” versus “extrapersonal focus.” However, this interpretation is less clear-cut than PC1, given the heterogeneous nature of the “other” category.

One property of PC1 bears on how much weight it can carry. Because it combines auditory and visual loadings, it groups the modality our formative participants prioritized with the modality they described as unusable, and so cannot by itself distinguish user-aligned from user-misaligned sensory thinking. The auditory-proportion analysis remains our primary test of alignment; the PCA serves as an exploratory view of the surrounding solution space.

As visualized in [Figure 4](#), the conditions differed descriptively in their emphasis within the solution space, despite substantial overlap in their distributions. The non-emotional VR group had a more positive mean PC1 score than the emotional VR group, with the video group between them. Because positive PC1 scores were associated primarily with auditory and visual solutions, the solutions generated in the non-emotional VR condition showed greater emphasis on direct sensory interventions, consistent with the higher auditory-solution proportion reported above. The overlapping distributions also reflect substantial participant-level variability within each condition.

5.2. Qualitative analysis

To contextualize how participants perceived the difference between passive video and embodied VR, we examined reflections from the 30 video-group participants who experienced the non-emotional VR simulation after the main experiment. We use these reflections descriptively to illustrate recurring ways in which participants described the contrast between the two media.

The most prominent theme was the transformation from abstract, third-person knowledge to a more personal, first-person embodied experience. While the video provided factual information about the challenges faced by individuals with low vision, the VR simulation immersed participants within those challenges. This distinction was articulated by numerous participants. For example, P73 stated that their

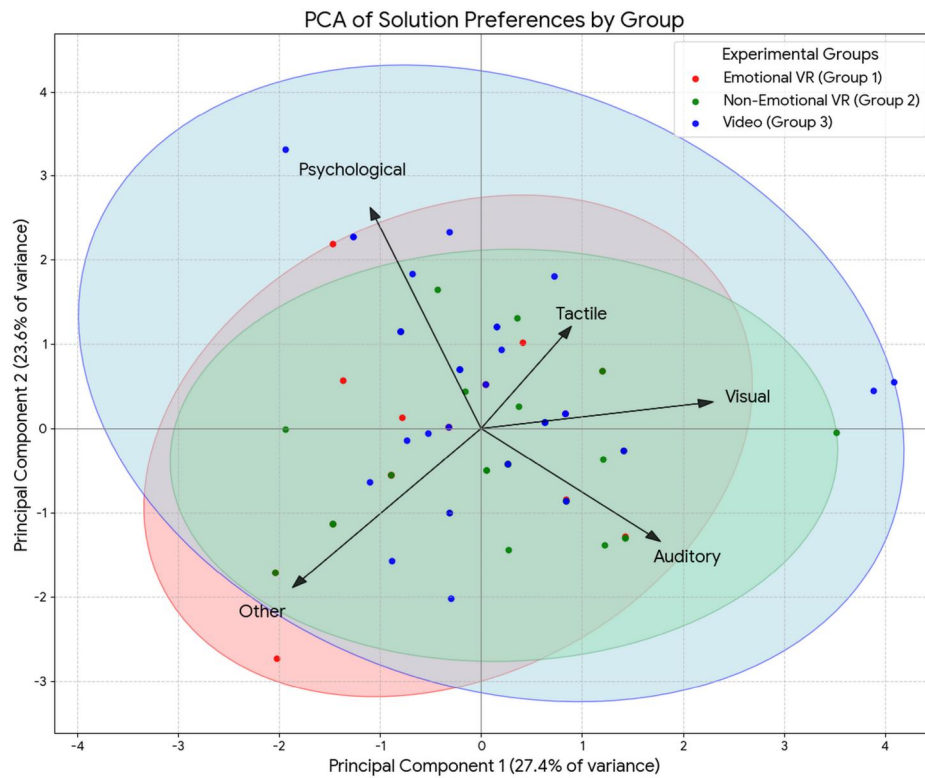


Figure 4. Two-dimensional exploratory PCA of the five z-standardized participant-level category counts. Dots represent individual participants by condition (red: emotional VR; green: non-emotional VR; blue: video), and arrows represent category loadings. Each shaded ellipse is a bounding region generated to enclose all participants in its group. Because the ellipses enclose every observation, their overlap partly reflects extreme individual cases rather than the typical spread within each condition.

prior understanding was merely “on a textual level, unable to truly experience the low-vision world.” P88 echoed this, describing their pre-VR knowledge as an “imaginary concept,” whereas VR provided a “concrete and real feeling” of the challenges, which in turn “greatly enhanced their empathy.” This shift from knowing about the user to feeling with the user helps contextualize the significant increase in empathy scores observed in the VR conditions relative to passive video.

This embodied perspective led to a recalibration of participants’ perception of the problem’s severity and was accompanied by a clear emotional impact. Many expressed shock, noting that the reality was “more severe than I imagined” (P79) and “much blurrier than I thought” (P87). This recalibration was not merely a cognitive update but an affective one. Participants reported feeling “inner frustration and anxiety” (P86) and a palpable sense of being “in danger and particularly anxious” (P63) when navigating the simulated environments.

Some reflections illustrate how embodied VR can broaden participants’ understanding of the solution space. The reflection from P90 captures this shift: “At first, I was focused on making the words bigger to help them see. After the experience, I realized that is almost impossible, and that I should focus on other aspects.” This “aha” moment suggests a pivot from attempting to remediate the visual impairment itself toward considering solutions that leverage other sensory modalities.

Taken together, these reflections descriptively illustrate how participants perceived the transition from passive video to embodied VR: as a shift toward more concrete understanding, stronger affective reactions, and reconsideration of possible solutions.

6. Discussion

This section interprets our findings in relation to the research question and the broader literature on VR, empathy, and inclusive design. The results show a dissociation: both VR conditions increased empathy relative to passive video, and emotional intensification produced no additional measurable

empathy gain; within the two VR conditions, the emotionally intensified condition showed weaker user-need alignment than non-emotional VR. We first discuss the respective contributions of embodiment and emotional intensification, then develop a CLT interpretation of the observed pattern, and finally outline implications for design practice and future research.

6.1. Embodiment versus emotional intensification

Our results support VR's value in bridging the empathy gap in inclusive design. Both VR conditions produced significantly greater empathy increases than passive video, consistent with prior research linking presence and first-person, embodied perspective-taking to empathic understanding (Ahn et al., 2013; Herrera et al., 2018; Slater & Sanchez-Vives, 2016). Reflections from video-group participants who later experienced the non-emotional simulation described a shift from a detached, "textual level" of understanding to a more visceral, embodied one. These accounts further illustrate how embodied VR can help designers engage with challenges that may be difficult to convey through passive media alone.

The experiment also identifies a boundary condition that qualifies the "empathy machine" claim (Heylighen & Dong, 2019; Sora-Domenjó, 2022). Adding emotional stimuli produced no additional measurable empathy gain over the non-emotional simulation. For user-need alignment, non-emotional VR produced a higher proportion of auditory-focused solutions than both emotional VR and video, while the latter two conditions did not differ significantly. Thus, within the two VR conditions, non-emotional VR preserved the empathy benefit of embodied VR while producing stronger user-need-aligned ideation. This pattern suggests that emotional intensification does not necessarily improve downstream design outcomes. One possible explanation, developed in the next subsection, is that emotionally charged cues may introduce demands extraneous to the subsequent ideation task.

In summary, VR's capacity for embodied simulation provides a valuable pathway to empathic understanding, yet the affective features used to intensify that experience may not always support subsequent analytical work. Taken together, the central design lesson is one of calibration: when VR is used as an empathy-building tool, adding emotional cues should be done with restraint. Additional emotional intensity may not produce measurable empathy gains beyond embodied interaction itself, yet it may still make subsequent ideation harder to align with user-prioritized needs. The next subsection develops this interpretation through CLT.

6.2. A cognitive-load-based interpretation of the performance pattern

CLT (Sweller, 2011) provides an interpretive lens for considering how emotionally charged cues might have contributed to weaker user-need-aligned ideation. It describes how limited working memory can be strained by presentation-related demands rather than by the problem itself.

A key result of our study was that the non-emotional VR group showed stronger user-need alignment, generating a significantly higher proportion of auditory-focused solutions than both the emotional VR and video groups. This outcome is unlikely to be explained solely by measured empathy gains, because both VR conditions increased empathy relative to passive video and did not differ significantly from each other. Together, these results point less to empathy level itself than to how emotionally intensified design features may shape the transition from empathic engagement to ideation. One plausible interpretation is that the audio-based stimuli in the emotional VR condition—such as diegetic sounds of frustration and impatient feedback from NPCs—may have introduced cognitive or affective demands whose effects carried over into the subsequent ideation task. In CLT terms, these presentation-related features may have functioned as extraneous demands unrelated to the core design problem itself (Sweller, 2011). If present, these demands could have left fewer cognitive resources available for analyzing users' functional challenges and generating solutions aligned with those needs.

Dual-process theory offers a complementary way to interpret this possibility (Kahneman, 2011). The visceral response to negative stimuli may have engaged intuitive, affective processing, while design ideation required deliberate, analytical cognition. The similar empathy gains across the two VR conditions suggest that future work should examine how the emotional condition may have shaped attention, arousal, or perceived task relevance. Qualitative reflections contextualize how participants perceived the

contrast between passive video and embodied VR. The behavioral outcomes provide the basis for interpreting the emotional versus non-emotional VR performance difference. The solution-space visualization suggested different ideation tendencies across conditions. These patterns are consistent with the CLT-based interpretation, although direct measures of cognitive load, emotional distress, attention, or distraction are needed to test the pathway more decisively.

Taken together, CLT provides a testable interpretation of the observed pattern. Emotionally charged cues may have introduced additional cognitive or affective demands, offering one possible explanation for why empathic understanding did not translate into equally user-need-aligned design ideas. This reframes the design question from whether VR increases empathy to how emotional intensity can be calibrated to support downstream design work.

6.3. Implications for design

Our findings carry implications for the practical use of VR in inclusive design workflows. For practitioners, the work cautions against adopting the “empathy machine” metaphor uncritically. The aim should not be the most emotionally intense simulation available, but an experience that supports empathic understanding and the design work that follows it. We recommend a calibrated approach spanning both the use and the construction of these tools.

For those using existing VR tools, this means a careful, procedurally-aware strategy:

- *Prioritize non-emotional simulations for core ideation.* During the early, divergent phases of design—such as problem-framing, needs-finding, and initial brainstorming—practitioners should consider interactive, non-emotional VR simulations. By reducing emotional demands that may be extraneous to ideation, these tools can foster deep, embodied understanding while potentially preserving resources for analytical thinking and creative problem-solving.
- *Use emotional stimuli strategically and with purpose.* Emotionally evocative VR experiences may still have value in other parts of the design process. For example, they could be used in later stages for solution validation, to communicate the urgency of a problem to stakeholders, or to help marketing teams develop more resonant narratives. However, their use should be deliberate and aligned with goals that are not purely generative.
- *Consider a reflective transition after emotive experiences.* A structured transition phase may help, although affective carryover should be treated as a hypothesis because it was not measured in this study (Lerner et al., 2004). If emotional engagement persists after the headset is removed, a structured interval could support emotional regulation and cognitive reappraisal before ideation (Cutuli, 2014). Future work should test whether such a transition improves subsequent ideation performance.

Furthermore, our research offers a critical implication for those who design these empathy tools:

- *Design for affective transparency with “emotional intensity indicators.”* Practitioners cannot apply a calibrated approach if the affective “dosage” of a tool is unknown. We propose that designers of empathy tools should build in features that promote user meta-cognition (Cabales, 2019). Drawing inspiration from “nutritional labels” for information (Abdul Latiff et al., 2016), a simple indicator, or avatar-related visualizations (Lee et al., 2022) could explicitly signal a simulation’s intended emotional load (e.g., “High Emotional Impact: Caution Advised for Ideation”). This acts as a metacognitive scaffold, prompting designers to reflect on how the tool might be influencing their cognitive state and to consciously guard against potential biases. This approach aligns with the broader call for transparency and responsibility in designing persuasive and assistive technologies (Grimpe et al., 2014).

This calibrated, two-pronged strategy—encompassing both mindful use and transparent design—allows teams to leverage VR’s strengths without overlooking the cognitive and affective demands that emotionally intense simulations may introduce. This task-sensitive approach also aligns with research on meaningful and eudaimonic experiences that moves beyond simple engagement metrics (Diefenbach et al., 2014; Mekler & Hornbæk, 2016).

6.4. Limitations and future research directions

Our study has boundaries that define a research agenda. The primary limitation is that cognitive load and related mediating processes were not measured directly. The observed pattern is consistent with a CLT-based interpretation, but the data cannot distinguish among several possible contributors: cognitive load, emotional distress, attentional narrowing, reduced self-efficacy, affective carryover, a mismatch between the emotional VR experience and the ideation task that followed, or emotionally charged cues functioning as task-irrelevant demands. These are not mutually exclusive, and we cannot establish which pathway, if any, best explains the performance differences. The participant-level variability observed in the non-emotional VR condition also suggests that the benefits of embodied, non-emotional VR may depend on unmeasured individual factors, such as prior design experience, familiarity with VR, spatial reasoning, or participants' ability to translate embodied experience into design-relevant insights. Testing these pathways directly is the next step, applying the methodological rigor found in high-stakes domains such as automotive HCI (Braun et al., 2021; Detjen et al., 2021). Future studies could incorporate instruments such as NASA-TLX, secondary-task measures, eye tracking, EEG (Kumar & Kumar, 2016), pupillometry (Beatty, 1982; Duchowski et al., 2018), and EDA (Nourbakhsh et al., 2012), alongside measures of emotional arousal or distress, attention, perceived task relevance, and self-efficacy. Such work would make it possible to test mediation pathways rather than inferring them from behavioral outcomes alone.

Furthermore, our design outcome metric should be interpreted as a targeted measure of user-need alignment; it does not provide a comprehensive measure of design quality. Auditory alignment was appropriate for the user needs and mobility scenarios identified in our formative study, where participants with low vision prioritized auditory information as a reliable sensory channel. Although task-relevant NPC information was delivered auditorily in both VR conditions, the emotional manipulation was implemented primarily through differences in spoken wording and vocal tone, together with behavior-contingent avatar feedback such as sighs. Because the primary outcome concerned auditory-focused solutions, a modality-specific response to negatively valenced auditory interaction cannot be ruled out. The observed pattern may therefore partly reflect reduced attention to, or avoidance of, auditory interaction; the present data do not establish a domain-general disruption to design ideation. Future studies should orthogonally manipulate emotional framing and sensory modality and use broader design-evaluation rubrics that include feasibility, novelty, creativity, usability, and direct user evaluation.

Although participants in both VR conditions were instructed to complete the same task-relevant NPC interactions, participant-level exposure to behavior-contingent emotional cues in the emotional VR condition was not recorded. Because cues such as avatar sighs were triggered by scripted error events, their number may have varied with participants' task behavior. Future studies should automatically log cue exposure or standardize the number of cues delivered.

Our participant sample also limits generalizability. Novice designers may respond differently from experienced professionals, and future studies should include expert designers to examine whether the observed pattern generalizes across levels of design experience. Broadening the demographic scope matters equally: understanding how these effects vary across designers from different cultural backgrounds, a central concern in cross-cultural HCI (Sturm et al., 2015), is essential to developing equitable and globally effective design tools.

Finally, this was a one-shot laboratory experiment built around a single low-vision simulation. Longitudinal studies should track how the possible cognitive costs and design benefits of empathic tools evolve with repeated use, and extending the paradigm to other accessibility contexts would test whether the cognitive-load account holds beyond this one scenario.

7. Conclusion

This article examined the role of VR in inclusive design, testing the assumption that more emotionally evocative simulations lead to better design outcomes. Motivated by the empathy gap between designers and users with low vision, we compared passive video, non-emotional VR, and emotional VR in terms of empathy change and subsequent user-need-aligned design ideation.

The three-condition design allowed us to distinguish the effects of interactive, embodied VR relative to passive video from the incremental effects of emotional intensification within VR. Both VR conditions produced greater empathy gains than passive video, while emotional VR added no measurable empathy gain over non-emotional VR. For user-need alignment, non-emotional VR yielded a higher proportion of auditory-focused solutions than both other conditions. The condition with the strongest user-need alignment thus combined interactive, embodied VR with no added emotional intensification, and its advantage was not accompanied by greater measurable empathy. This pattern suggests that emotional intensity is not a reliable proxy for downstream user-need alignment.

The primary contribution of this work is an empirical and theoretical reframing of VR empathy tools. Drawing on CLT, we propose, as a working hypothesis, that emotionally charged cues may introduce additional cognitive or affective demands that interfere with translating empathic understanding into design ideas. Because cognitive load was not directly measured, this account should be tested in future work with direct measures of cognitive load and related affective, attentional, and motivational processes.

For practitioners, the question is not only whether an immersive experience increases empathy but whether it supports the analytical work that follows. Our findings support calibrating emotional intensity to the ideation task while preserving the empathy benefits of embodied VR. VR empathy tools should therefore be evaluated in terms of both empathic engagement and subsequent design outcomes.

Note

1. The criteria for Grade 1 low vision adopted in this study are based on the Visual Disability Standards issued by the China Disabled Persons' Federation in December 2006, which defines Grade 1 low vision as best-corrected visual acuity equal to or better than 0.05 but worse than 0.1 (Reference: https://www.gov.cn/ztlz/gacjr/content_459939.htm).

Acknowledgments

All procedures involving human participants in this study received the ethics approval from the Human Subjects Panel of the Committee for Research Practices at the Hong Kong University of Science and Technology (Guangzhou), with the approval number HKUST(GZ)-HSP-2024-0101. Written informed consent was obtained from all individual participants included in the experiments.

The authors thank Mr. Rui Zheng, President of the Nanshan District Association of the Blind in Shenzhen, for his support during the pilot study and assistance in connecting the research team with individuals with low vision for interviews. The authors also acknowledge support from the Metaverse Joint Innovation Laboratory and the Base of Red Bird MPhil at the Hong Kong University of Science and Technology (Guangzhou).

Author contributions

CRedit: **Dawei Xiong**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Zhijun Ma**: Formal analysis, Methodology, Visualization, Writing – original draft; **Hao Ling**: Conceptualization, Data curation, Investigation, Writing – review & editing; **Xun Wu**: Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – review & editing; **Mingming Fan**: Methodology, Supervision, Writing – review & editing; **Xu Zhang**: Conceptualization, Formal analysis, Funding acquisition, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of generative AI use

During the revision and final preparation of this manuscript, the authors used OpenAI ChatGPT (GPT-5) for language refinement, consistency checking, and formatting. All outputs were reviewed and edited by the authors, who take full responsibility for the manuscript.

Disclosure statement

The authors report that there are no competing interests to declare.

Funding

Xu Zhang acknowledges funding support from the National Natural Science Foundation of China (No. 72203054), the Basic and Applied Basic Research Foundation of Guangdong Province (No. 2025A1515010549), the Yangcheng Scholars Research Project of Guangzhou (No. 2024312031), and the Guangzhou Municipal Science and Technology Project (No. 2025A03J3807, No. 2024A03J0630 and No. 2024A04J4549).

ORCID

Dawei Xiong  <http://orcid.org/0009-0006-2995-7385>
 Zhijun Ma  <http://orcid.org/0000-0002-8091-3760>
 Hao Ling  <http://orcid.org/0009-0007-6212-3267>
 Xun Wu  <http://orcid.org/0000-0002-7518-3951>
 Mingming Fan  <http://orcid.org/0000-0002-0356-4712>
 Xu Zhang  <http://orcid.org/0000-0001-7937-3827>

Data availability statement

The data that support the findings of this study are deposited in Science Data Bank under restricted access at <https://doi.org/10.57760/sciencedb.45738>. Because the dataset contains participant-level human-subject data, access requests will be considered by the corresponding author in accordance with applicable ethical and participant-privacy requirements.

References

- Abdul Latiff, Z. A. B., Rezai, G., Mohamed, Z., & Amizi Ayob, M. (2016). Food labels' impact assessment on consumer purchasing behavior in Malaysia. *Journal of Food Products Marketing*, 22(2), 137–146. <https://doi.org/10.1080/10454446.2013.856053>
- Ahn, S. J., Le, A. M. T., & Bailenson, J. (2013). The effect of embodied experiences on self-other merging, attitude, and helping behavior. *Media Psychology*, 16(1), 7–38. <https://doi.org/10.1080/15213269.2012.755877>
- Alessandro Gasparini, A. (2015). *Perspective and use of empathy in design thinking* [Paper presentation]. The Eighth International Conference on Advances in Computer-Human Interactions (pp. 49–54).
- Barbot, B., & Kaufman, J. C. (2020). What makes immersive virtual reality the ultimate empathy machine? Discerning the underlying mechanisms of change. *Computers in Human Behavior*, 111, 106431. <https://doi.org/10.1016/j.chb.2020.106431>
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59(1), 617–645. <https://doi.org/10.1146/annurev.psych.59.103006.093639>
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>
- Bennett, C. L., & Rosner, D. K. (2019). The promise of empathy: Design, disability, and knowing the “other”. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–13).
- Bertrand, P., Guegan, J., Robieux, L., McCall, C. A., & Zenasni, F. (2018). Learning empathy through virtual reality: Multiple strategies for training empathy-related abilities using body ownership illusions in embodied virtual reality. *Frontiers in Robotics and AI*, 5, 26. <https://doi.org/10.3389/frobt.2018.00026>
- Bevan, C., Green, D. P., Farmer, H., Rose, M., Cater, K., Fraser, D. S., & Brown, H. (2019). Behind the curtain of the “ultimate empathy machine” on the composition of virtual reality nonfiction experiences. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1–12).
- Bloom, P. (2017). *Against empathy: The case for rational compassion*. Random House.
- Braun, M., Weber, F., & Alt, F. (2021). Affective automotive user interfaces – Reviewing the state of driver affect research and emotion regulation in the car. *ACM Computing Surveys*, 54(7), 1–26. <https://doi.org/10.1145/3460938>
- Bujić, M., Salminen, M., Macey, J., & Hamari, J. (2020). “Empathy machine”: How virtual reality affects human rights attitudes. *Internet Research*, 30(5), 1407–1425. <https://doi.org/10.1108/INTR-07-2019-0306>
- Cabales, V. (2019). *Muse: Scaffolding metacognitive reflection in design-based research* [Paper presentation]. Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1–6).
- Carter, M., & Egliston, B. (2024). *Fantasies of virtual reality: Untangling fiction, fact, and threat*. MIT Press.
- Chatain, J., Kapur, M., & Sumner, R. W. (2023). *Three perspectives on embodied learning in virtual reality: Opportunities for interaction design* [Paper presentation]. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (pp. 1–8).

- Christofi, M., & Michael-Grigoriou, D. (2017). *Virtual reality for inducing empathy and reducing prejudice towards stigmatized groups: A survey* [Paper presentation]. 2017 23rd International Conference on Virtual System & Multimedia (VSM) (pp. 1–8). IEEE.
- Clark, A. (1998). *Being there: Putting brain, body, and world together again*. MIT Press.
- Coleman, R., Clarkson, J., & Cassim, J. (2008). *Design for inclusivity: A practical guide to accessible, innovative and user-centred design* (1st ed.). Routledge. <https://doi.org/10.4324/9781315576626>
- Cutuli, D. (2014). Cognitive reappraisal and expressive suppression strategies role in the emotion regulation: An overview on their modulatory effects and neural correlates. *Frontiers in Systems Neuroscience*, 8, 175. <https://doi.org/10.3389/fnsys.2014.00175>
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44(1), 113–126. <https://doi.org/10.1037/0022-3514.44.1.113>
- Debnath, A., Lahana, A., Uğur Genç, H., Soubutts, E., Lahav, M., Horne, T., Meijer, W., Suen Pai, Y., Hsu, Y.-C., Barbareschi, G., et al. (2024). *EmpathiCH: Scrutinizing empathy-centric design beyond the individual* [Paper presentation]. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (pp. 1–7).
- Decety, J. (2021). Why empathy is not a reliable source of information in moral decision making. *Current Directions in Psychological Science*, 30(5), 425–430. <https://doi.org/10.1177/09637214211031943>
- Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3(2), 71–100. <https://doi.org/10.1177/1534582304267187>
- Detjen, H., Faltaous, S., Pfleging, B., Geisler, S., & Schneegass, S. (2021). How to increase automated vehicles' acceptance through in-vehicle interaction design: A review. *International Journal of Human-Computer Interaction*, 37(4), 308–330. <https://doi.org/10.1080/10447318.2020.1860517>
- Diefenbach, S., Kolb, N., & Hassenzahl, M. (2014). The 'hedonic' in human-computer interaction: history, contributions, and future research directions. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (pp. 305–314).
- Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., Raubal, M., & Giannopoulos, I. (2018). The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–13).
- Embøl, L., Hutters, C., Junker, A., Reipur, D., Adjorlu, A., Nordahl, R., & Serafin, S. (2021). HearMeVirtual reality: Using virtual reality to facilitate empathy between hearing impaired children and their parents. *Frontiers in Virtual Reality*, 2, 691984. <https://doi.org/10.3389/frvir.2021.691984>
- Eres, R., Decety, J., Louis, W. R., & Molenberghs, P. (2015). Individual differences in local gray matter density are associated with differences in affective and cognitive empathy. *NeuroImage*, 117, 305–310. <https://doi.org/10.1016/j.neuroimage.2015.05.038>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion (Washington, D.C.)*, 7(2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Gallagher, S. (2005). *How the body shapes the mind*. Oxford University Press.
- Grimpe, B., Hartswood, M., & Jirotko, M. (2014). Towards a closer dialogue between policy and practice: Responsible design in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2965–2974).
- Hassan, R. (2020). Digitality, virtual reality and the 'empathy machine'. *Digital Journalism*, 8(2), 195–212. <https://doi.org/10.1080/21670811.2018.1517604>
- Hassenzahl, M., Diefenbach, S., & Göritz, A. (2010). Needs, affect, and interactive products: Facets of user experience. *Interacting with Computers*, 22(5), 353–362. <https://doi.org/10.1016/j.intcom.2010.04.002>
- Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PLoS One*, 13(10), e0204494. <https://doi.org/10.1371/journal.pone.0204494>
- Heylighen, A., & Dong, A. (2019). To empathise or not to empathise? Empathy and its limits in design. *Design Studies*, 65, 107–124. <https://doi.org/10.1016/j.destud.2019.10.007>
- Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human-computer interaction. *Computers in Human Behavior*, 26(6), 1278–1288. <https://doi.org/10.1016/j.chb.2010.05.031>
- Iqbal, A. I., Aamir, A., Hammad, A., Hafsa, H., Basit, A., Oduoye, M. O., Anis, M. W., Ahmed, S., Younus, M. I., & Jabeen, S. (2024). Immersive technologies in healthcare: An in-depth exploration of virtual reality and augmented reality in enhancing patient care, medical education, and training paradigms. *Journal of Primary Care & Community Health*, 15, 21501319241293311. <https://doi.org/10.1177/21501319241293311>
- Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.
- Kalyuga, S. (2008). *Managing cognitive load in adaptive multimedia learning*. IGI Global.
- Kanske, P., Böckler, A., Trautwein, F.-M., Parianen Lesemann, F. H., & Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Social Cognitive and Affective Neuroscience*, 11(9), 1383–1392. <https://doi.org/10.1093/scan/nsw052>
- Khukalenko, I., & Khanolainen, D. (2025). The use of virtual reality training to enhance empathy: Software review. *Journal of New Approaches in Educational Research*, 14(1), 14. <https://doi.org/10.1007/s44322-025-00034-z>

- Kumar, N., & Kumar, J. (2016). Measurement of cognitive load in HCI systems using EEG power spectrum: An experimental study. *Procedia Computer Science*, 84, 70–78. <https://doi.org/10.1016/j.procs.2016.04.068>
- Kurz, K., Bock, C., & Hanschur, L. (2024). Flip the tweet – The two-sided coin of entrepreneurial empathy and its ambiguous influence on new product development. *Journal of Business Venturing*, 39(2), 106378. <https://doi.org/10.1016/j.jbusvent.2023.106378>
- Lee, S., El Ali, A., Wijntjes, M., & Cesar, P. (2022). Understanding and designing avatar biosignal visualizations for social virtual reality entertainment. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (pp. 1–15).
- Lerner, J. S., Small, D. A., & Loewenstein, G. (2004). Heart strings and purse strings: Carryover effects of emotions on economic decisions. *Psychological Science*, 15(5), 337–341. <https://doi.org/10.1111/j.0956-7976.2004.00679.x>
- Lucero, A., Vaajakallio, K., & Dalsgaard, P. (2012). The dialogue-labs method: Structuring process, space and materials to spark dialogue in co-design events. *CoDesign*, 8(1), 1–23. <https://doi.org/10.1080/15710882.2011.609888>
- Mack, K., McDonnell, E., Jain, D., Lu Wang, L., Froehlich, J. E., & Findlater, L. (2021). What do we mean by “accessibility research”? A literature survey of accessibility papers in CHI and ASSETS from 1994 to 2019. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (pp. 1–18).
- Mauri, A., Hsu, Y.-C., Brambilla, M., Huang, T.-H. K., O’kane, A. A., & Verma, H. (2022). *Empathy-centric design at scale* [Paper presentation]. CHI Conference on Human Factors in Computing Systems Extended Abstracts (pp. 1–6).
- Mekler, E. D., & Hornbæk, K. (2016). *Momentary pleasure or lasting meaning? Distinguishing eudaimonic and hedonic user experiences* [Paper presentation]. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery (pp. 4509–4520). <https://doi.org/10.1145/2858036.2858225>
- Motahar, T., Brown, N., Stampfer Wiese, E., & Wiese, J. (2023). Building “design empathy” for people with disabilities: An unsolved challenge in HCI education. In *Proceedings of the 5th Annual Symposium on HCI Education* (pp. 68–71).
- Nourbakhsh, N., Wang, Y., Chen, F., & Calvo, R. A. (2012). *Using galvanic skin response for cognitive load measurement in arithmetic and reading tasks* [Paper presentation]. In Proceedings of the 24th Australian Computer-Human Interaction Conference (OzCHI ‘12) (pp. 420–423). Association for Computing Machinery. <https://doi.org/10.1145/2414536.2414602>
- Packard, M. D., & Burnham, T. A. (2021). Do we understand each other? Toward a simulated empathy theory for entrepreneurship. *Journal of Business Venturing*, 36(1), 106076. <https://doi.org/10.1016/j.jbusvent.2020.106076>
- Righi, V., Sayago, S., & Blat, J. (2017). When we talk about older people in HCI, who are we talking about? Towards a ‘turn to community’ in the design of technologies for a growing ageing population. *International Journal of Human-Computer Studies*, 108, 15–31. <https://doi.org/10.1016/j.ijhcs.2017.06.005>
- Roider, F., Rümelin, S., Pfleging, B., & Gross, T. (2017). *The effects of situational demands on gaze, speech and gesture input in the vehicle* [Paper presentation]. Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI ‘17) (pp. 94–102). ACM. <https://doi.org/10.1145/3122986.3122999>
- Schutte, N. S., & Stilinović, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and Emotion*, 41(6), 708–712. <https://doi.org/10.1007/s11031-017-9641-7>
- Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: A double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain: A Journal of Neurology*, 132(Pt 3), 617–627. <https://doi.org/10.1093/brain/awn279>
- Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI*, 3, 74. <https://doi.org/10.3389/frobt.2016.00074>
- Sora-Domenjó, C. (2022). Disrupting the “empathy machine”: The power and perils of virtual reality in addressing social issues. *Frontiers in Psychology*, 13, 814565. <https://doi.org/10.3389/fpsyg.2022.814565>
- Sturm, C., Oh, A., Linxen, S., Nocera, J. A., Dray, S., & Reinecke, K. (2015). How WEIRD is HCI? Extending HCI principles to other countries and cultures. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2425–2428).
- Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, 37–76). Elsevier.
- Thompson, N. M., Van Reekum, C. M., & Chakrabarti, B. (2022). Cognitive and affective empathy relate differentially to emotion regulation. *Affective Science*, 3(1), 118–134. <https://doi.org/10.1007/s42761-021-00062-w>
- Van Loon, A., Bailenson, J., Zaki, J., Bostick, J., & Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PLoS One*, 13(8), e0202442. <https://doi.org/10.1371/journal.pone.0202442>
- Ventura, S., Badenes-Ribera, L., Herrero, R., Cebolla, A., Galiana, L., & Baños, R. (2020). Virtual reality as a medium to elicit empathy: A meta-analysis. *Cyberpsychology, Behavior and Social Networking*, 23(10)2020, 667–676. <https://doi.org/10.1089/cyber.2019.0681>
- Williamson, J., McGill, M., & Outram, K. (2019). *PlaneVR: Social acceptability of virtual reality for aeroplane passengers* [Paper presentation]. Proceedings of the 2019 CHI Conference on Human Factors in Computing

- Systems (CHI '19). Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3290605.3300310>
- Wimmer, J. (2020). The “ultimate empathy machine” revisited: Challenges of augmented and virtual realities from an ethical perspective. In *Games and ethics: Theoretical and empirical approaches to ethical questions in digital game cultures* (pp. 129–142). Springer. https://doi.org/10.1007/978-3-658-28175-5_9
- World Health Organization. (2023). Blindness and vision impairment. WHO Fact Sheets (August 2023). <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>
- Wright, P., & McCarthy, J. (2008). Empathy and experience in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 637–646).
- Youn Oh, S., Bailenson, J., Weisz, E., & Zaki, J. (2016). Virtually old: Embodied perspective taking and the reduction of ageism under threat. *Computers in Human Behavior*, 60, 398–410. <https://doi.org/10.1016/j.chb.2016.02.007>
- Zhao, Y., Kupferstein, E., Rojnirun, H., Findlater, L., & Azenkot, S. (2020). The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14).

About the authors

Dawei Xiong is a PhD student in the Innovation, Policy and Entrepreneurship Thrust, Society Hub, at the Hong Kong University of Science and Technology (Guangzhou). His research integrates behavioral public administration with emerging technologies, focusing on AI agents, virtual reality, and inclusive governance.

Zhijun Ma is a PhD student in Computational Media and Arts at the Hong Kong University of Science and Technology (Guangzhou). He holds a B.Eng. in Landscape Architecture from Southeast University and an MPhil from HKUST(GZ). His research focuses on human-centered spatial intelligence for complex physical tasks.

Hao Ling is a PhD candidate in the Innovation, Policy and Entrepreneurship Thrust, Society Hub, at the Hong Kong University of Science and Technology (Guangzhou). His research focuses on behavioral and experimental economics, examining how virtual reality and AI-mediated environments shape trust, empathy, accessibility, and inclusive design.

Xun Wu is a professor of Innovation, Policy and Entrepreneurship Thrust at the Hong Kong University of Science and Technology (Guangzhou). His research focuses on public policy and management, with interests in technology policy, water resource management, and anti-corruption.

Mingming Fan is an associate professor at The Hong Kong University of Science and Technology (Guangzhou and Hong Kong). He leads Augmenting People via Empowering X (APEX) Group to explore the design principles, interaction techniques, and applications of Human-AI Collaboration and VR/AR/MR, especially for addressing aging, accessibility, and education/learning challenges.

Xu Zhang is an assistant professor of Innovation, Policy, and Entrepreneurship at the Hong Kong University of Science and Technology (Guangzhou). His research is grounded in behavioral and experimental economics, with strong interdisciplinary connections to human-computer interaction.

Appendix A: VR environment

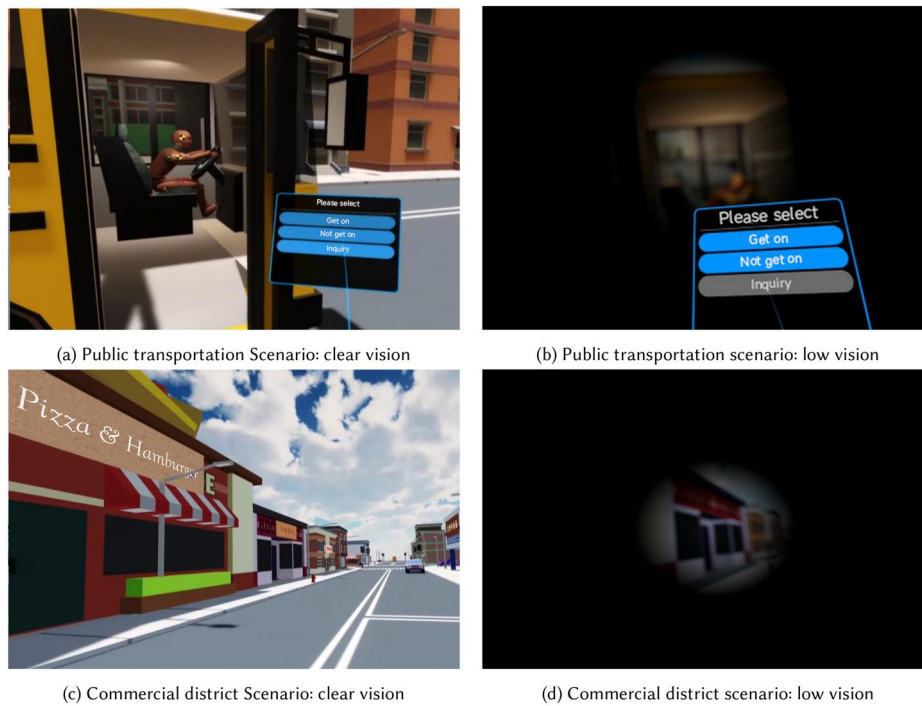


Figure A1. VR environment. Images (b) and (d) depict the first-person low-vision perspective VR simulation experienced by participants during the experiment, while images (a) and (c) provide clear-vision reference views to help readers understand the simulated scenarios.

Appendix B: Coding criteria and examples of valid solutions

Table B1. Tagging rules for text analysis.

Category label	Keywords
Auditory	Voice, sound, audio, broadcast, hearing, volume, alert, horn, noise, say, speak
Tactile	Blind path, Braille, handrail, vibration, tactile, touch, texture, shake, oscillate
Visual	Color, traffic light, signal, illumination, signboard, observe, route marker, conspicuous mark, large font, LED
Psychological	Fear, patience, care, attitude, trouble, anxiety, sense of security, pressure, nervousness, worry, self-confidence, panic

Table B2. Examples of coded solutions.

Solution	Category label
Setting blind path in the commercial street	Tactile
Offering clear broadcasts when the bus is coming	Auditory
Voice navigation at the bus station reminds low-vision persons when their bus is coming	Auditory
Mobile phone can vibrate or make a voice announcement when the vehicle is reached	Auditory, tactile
Braille signs can be adopted for the shop	Tactile
Add voice prompts to traffic lights	Auditory, visual
Add more indoor Braille maps or audio guided maps	Auditory, tactile
Provide training for bus drivers to improve their attitude toward people with low vision	Psychological