Physical-virtual window: a social interaction resource for use in XR environments

Marcos Felipe Soares[1,2], Ana Júlia Tolêdo[1,2], Rafael Padilha[1,2], Guilherme C. Franco[1,2], Hugo Nascimento[1,2], Laurita R. Salles [2,4], Luciana de O. Berretta [1,2], Alércio R. Souza [2,3]

[1]Instituto de Informática (INF) - Universidade Federal de Goiás (UFG)

[2]Centro de Competência Embrapii em Tecnologias Imersivas (AKCIT)

[3]Faculdade de Informação e Comunicação (FIC) - Universidade Federal de Goiás (UFG)

[4]Programa de Pós-Graduação em Performances Culturais (PPGPC) - Universidade Federal de Goiás (UFG)

Abstract. *In immersive environments such as the metaverse, user interaction with the virtual space has been shown to increase engagement and a sense of belonging to the experience. By integrating a window from the physical world into the digital environment, not only are the means of interaction amplified, but a dynamic connection between multiple realities is also created, allowing them to intersect through this resource. In this paper, we propose the creation of a window that transmits spatial audio and video from the physical environment to a space in the metaverse and vice-versa. We also analyze the initial user perception and the limitations imposed by current technology.*

1. Introduction

The evolving concept of the metaverse builds upon the hypertext protocol's mission of creating an interconnected digital environment, expanding it into an immersive three-dimensional space. In this new paradigm, the perceptual asymmetry imposed by current-day virtual reality (VR) experiences is a major detrimental factor to the user experience. As such, narrowing this gap becomes key to the development of immersive environments. In this paper, we explore the concept of a physical-virtual window, a setup geared towards inter-reality interactivity that is made to enable two-way audiovisual communication between physical and virtual spaces, as a tool to alleviate the VR perceptual gap, in turn helping boost user satisfaction.

The overall goal of the study is to create one such window. To achieve this goal, the physical-virtual window must meet the requirements of having window-like dimensions, face-to-face interaction capability, bilateral and spatial audio functionality, and low-latency communication. Then, a pilot study, comprised of a communication experiment using the system followed by a questionnaire ascertaining user satisfaction, was conducted.

2. Related Works

Previous research on blending physical and virtual worlds shows positive impacts on social interaction [Andersson 2013], building on approaches by Bower, Cram and Groom [Bower et al. 2010]. Implementations include physical-physical portals like Benediktas

Gylys' Portal sculpture connecting cities [Guardian 2024] and Dublin-New York connections [Ireland.ie 2024], which demonstrated cultural exchange benefits [People 2024].

Bower et al. implemented one of the earliest comprehensive physical-virtual integration systems, creating a 'blended reality' classroom between physical space and Second Life. Their system achieved bidirectional audiovisual communication through Ustream video streaming and projection, testing social interactions between physical participants and virtual avatars. However, they faced severe limitations: 4-second communication latency, 2-3 fps video quality, audio comprehension problems, and orientation issues that created "parallel but separated environments" rather than unified interaction.

Our work differentiates itself by leveraging modern WebRTC protocols for lower latency communication and A-frame for web-based VR, ensuring wider availability. We also integrate spatial audio capabilities absent from their system and employ standardized web technologies designed for real-time interaction, addressing the technical and social gaps identified in previous research.

3. Methodology

For the purposes of our research, we thus define a physical-virtual window as two-part perceptual portal with audio-visual dimensions allowing two-way communication between physical participants and users inside an immersive experience. The physical part is a setup involving two digital device combos: a camera and a projector - for the visual channel -, and a speakers and microphone combo - enabling communication through the audio channel -. The virtual part is contained inside a host VR environment and consists of a simulated screen that displays, in real time, the video feed as it is captured by the camera. The audio feed is also made to sound as if coming from the screen using spacial audio technology. This setup is schematized in Figure 1.

A-frame [Supermedium 2016], a three.js-based web framework for building VR experiences, was chosen for the immersive virtual environment. Along with A-frame, other tools were employed in tailoring the 3D support environment to the project's needs. These include male and female avatars sourced from Hackweek Avatar Maker [Mozilla 2023], and multiplayer functionality provided by the Networked A-frame library [Networked 2023] using the Janus WebRTC server [JanusAdapter 2022]. Figure 2 illustrates our implementation of these tools.

To test the physical-virtual window, a pilot study was conducted with six users who accessed the environment either through the physical space, the virtual environment, or in some cases both alternately. Users were instructed to first introduce themselves to each other and then play rock-paper-scissors against one another. The choice of this game was because it is simple to perform, can be played using only hands (both real user hands and virtual avatar hands), and requires concentration and network communication speed. Figure 3 shows two users interacting via the system. Each virtual world user accessed the environment using a Meta Quest 3 device. Physically, a projection screen and camera were installed to present the virtual user's image and audio, and to capture the image and audio of the user in the physical environment.

After approximately 5 to 10 minutes of users introducing themselves and playing the game multiple times, they completed a questionnaire to report their experiences. The

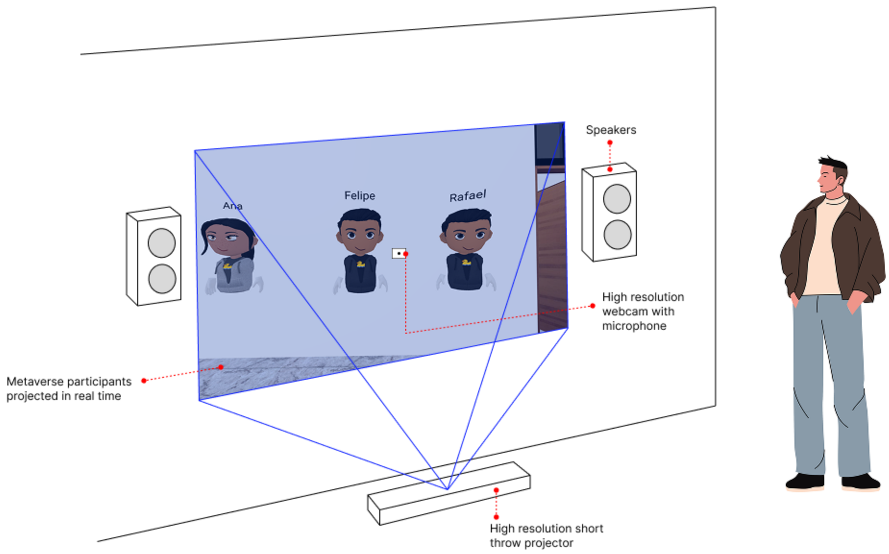

Figure 1. Schema detailing the physical device setup

questionnaire is comprised of three parts: collection of personal data such as age and name, rating of assertions pertaining to the immersive digital environment or the physical environment on a 5-point Likert scale, and optional open-ended questions, accepting both short and long feedback.

4. Results

Our evaluation involved 6 participants, with ages ranging from 19 to 27 years.

Participants reported moderate to high levels of immersion (around 3.5), while virtual environment realism averaged 3.2. Navigation controls scored above 4 points, and window usability scored 4.4. Communication quality averaged 3.2, comparable to face-to-face interaction. No significant adverse effects were reported (headache 1.4, nausea 1.2). Physical window realism scored 2.7, while technical difficulties averaged 2.7, indicating persistent stability issues. Despite technological advances over Bower et al.'s work, results show persistent challenges. While we eliminated their severe audio and latency problems, moderate realism scores (2.7-3.2) and technical difficulties (2.7) indicate fundamental user experience barriers remain at similar levels to their findings. User feedback appreciated spatial audio improvements but still requested reduced latency and enhanced visual realism, echoing concerns from the earlier study.

Regarding potential applications, participants identified education, entertainment, telemedicine, training, and remote meetings as promising areas. One participant specifically noted the system's potential for helping individuals with social interaction difficulties by providing a gradual introduction to social contact.

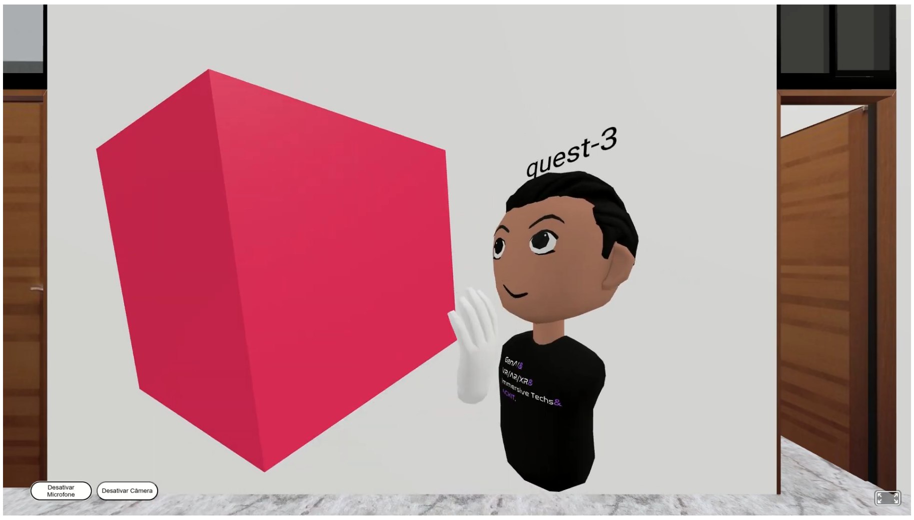

Figure 2. A user interacts with a grabbable cube within the virtual environment

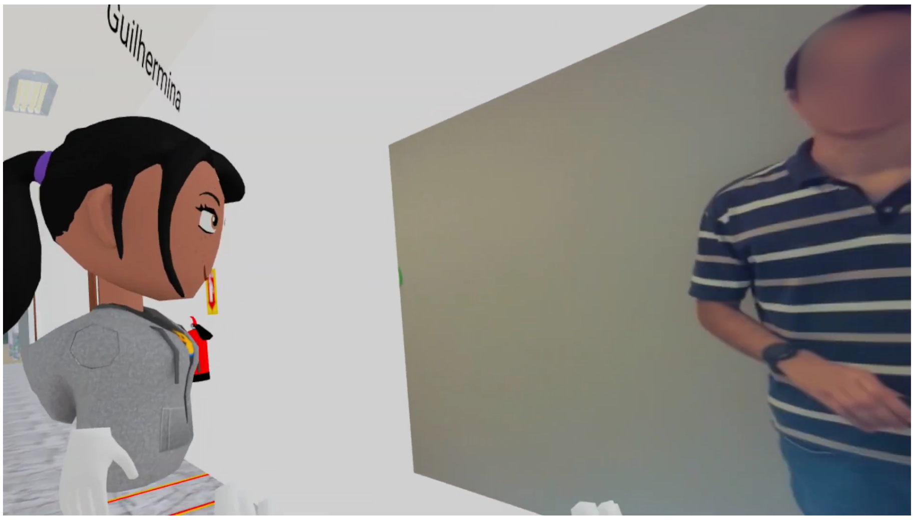

Figure 3. Two users interact via the physical-virtual window

5. Discussion

Our implementation of the physical-virtual window demonstrates significant progress in bridging physical and virtual realities. The successful integration of A-frame, WebRTC, and other technologies created a functional bidirectional communication system, as highlighted by the high usability scores and positive comments on the present of spatial audio. This feedback suggests that our technical approach effectively addresses core interaction challenges identified in related work. The system successfully facilitated meaningful social interactions, as evidenced by participants' ability to engage in conversations and play games through the window.

However, reported technical difficulties indicate that some challenges persist. Communication latency, mentioned by a few participants, remains a significant barrier to seamless interaction. This echoes limitations observed in earlier work by Bower, Cram and Groom [Bower et al. 2010], suggesting that despite technological advances, real-time inter-reality communication still faces infrastructure constraints. One potential avenue for improvement would be platform-specific optimization, making fuller use of the latency-reduction tools provided within each environment. The moderate realism scores highlight the ongoing challenge of achieving visual fidelity that matches participants' expectations. The camera and display used, while functional, do not provide the resolution expected to be fully convincing. Additionally, the unrealistic 3D models currently in use create a level of disconnect. This issue might be mitigated by integrating a more advanced, pre-existing avatar editor into the support environment, such as Ready Player Me [ReadyPlayerMe 2024], in order to improve both realism and user identification.

It is important to acknowledge that our findings are based on a limited sample size of six participants, which constrains the generalizability of our results. Larger-scale studies with more diverse participant populations would be necessary to draw broader conclusions about the effectiveness of physical-virtual window systems. To better evaluate our concept, we plan to conduct a follow-up study in which the physical installation will remain accessible for an extended period in a commonly frequented area of the university, and the support virtual environment will be hosted on an internet domain, enabling broader participation. To facilitate feedback, a QR code will be displayed at the site, allowing passersby to provide both positive comments and constructive criticism.

6. Conclusion

This study successfully demonstrates the feasibility and potential of physical-virtual windows as tools for enhancing social interaction in metaverse environments. Our implementation achieved its primary goal of enabling bidirectional audiovisual communication between physical and virtual spaces, with participants reporting moderate to high levels of satisfaction and usability, and represented meaningful advancement over earlier approaches, addressing specific technical limitations identified by Bower et al.

At the same time, our findings highlight persistent challenges in achieving seamless inter-reality communication. Communication latency and visual realism remain significant barriers, underscoring broader XR research concerns around maintaining presence, immersion, and natural interaction across heterogeneous platforms. These results suggest that while the physical-virtual window concept is sound, technological infrastructure must continue to evolve to support it.

Future research should not only address these technical limitations but also situate physical-virtual windows within ongoing XR debates on scalability, inclusivity, and long-term engagement. Larger-scale studies with more diverse participant populations will be essential for understanding the broader social and psychological impacts of sustained physical-virtual interaction, and for positioning this approach as a meaningful contribution to the development of next-generation XR systems.

7. Acknowledgements

This work was supported by The Advanced Knowledge Center in Immersive Technologies (AKCIT), with financial resources from the PPI IoT/Manufatura 4.0 / PPI HardwareBR of the MCTI grant number 057/2023, signed with EMBRAPII, and by the Fundação de Apoio à Pesquisa do Estado de Goiás (FAPEG), Brazil.

References

Andersson, L. (2013). Blending realities: Social impacts of merging physical and digital spaces. Master's thesis, Blekinge Institute of Technology. Accessed February 28, 2025. DOI:10.14236/ewic/HCI2011.50.

Bower, M., Cram, A., and Groom, D. (2010). Blended reality: Issues and potentials in combining virtual worlds and face-to-face classes. *ResearchGate*.

Guardian, T. (2024). Portal art exhibit new york dublin. Accessed: June 23, 2025. https://www.theguardian.com/us-news/article/2024/may/17/portal-art-exhibit-new-york-dublin-benediktas-gylys.

Ireland.ie (2024). The portal connecting dublin and new york city in real time. Accessed: June 23, 2025. https://www.ireland.ie/en/usa/the-portal-connecting-dublin-and-new-york-city-in-real-time/.

JanusAdapter (2022). Github - networked-aframe/naf-janus-adapter: networked-aframe janus network adapter. Accessed February 25, 2025. https://github.com/networked-aframe/naf-janus-adapter.

Mozilla (2023). mozilla/hackweek-avatar-maker: Hubs team hack week project. standalone avatar creator. Accessed February 23, 2025. https://webrtc.org/.

Networked, A. (2023). Networked-aframe. Accessed February 24, 2025. https://github.com/networked-aframe/networked-aframe.

People, D. (2024). Dublin and nyc students use the portal to make maths magic across the atlantic. Accessed: June 23, 2025. https://dublinpeople.com/news/dublin/articles/2024/06/13/54465/.

ReadyPlayerMe (2024). Ready player me - integrate an avatar creator into your game in days. Accessed February 25, 2025. https://readyplayer.me/.

Supermedium (2016). A-frame – make webvr. Accessed February 25, 2025. https://aframe.io.