2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct) | 979-8-3315-0691-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ISMAR-Adjunct64951.2024.00146

From Lab to a Long-Term Care Facility: Lessons Learned from Field Deployment of Augmented Reality Telepresence System as an Interactive Communication Technology

Mahrukh Tauseef*
Electrical Engineering
Vanderbilt University
Nashville, TN, United States

Akshith Ullal
Electrical Engineering
Vanderbilt University
Nashville, TN, United States

Alexandra Watkins
Mechanical Engineering
Vanderbilt University
Nashville, TN, United States

Zalen Ingram
Computer Science
Vanderbilt University
Nashville, TN, United States

Cathy Maxwell
School of Nursing
Vanderbilt University
Nashville, TN, United States

Judith Tate
College of Nursing
The Ohio State University
Columbus, Ohio, United States

Lisa Juckett
School of Health and Rehabilitation Sciences
The Ohio State University
Columbus, Ohio, United States

Lorraine C. Mion
College of Nursing
The Ohio State University
Columbus, Ohio, United States

Nilanjan Sarkar
Mechanical Engineering
Vanderbilt University
Nashville, TN, United States

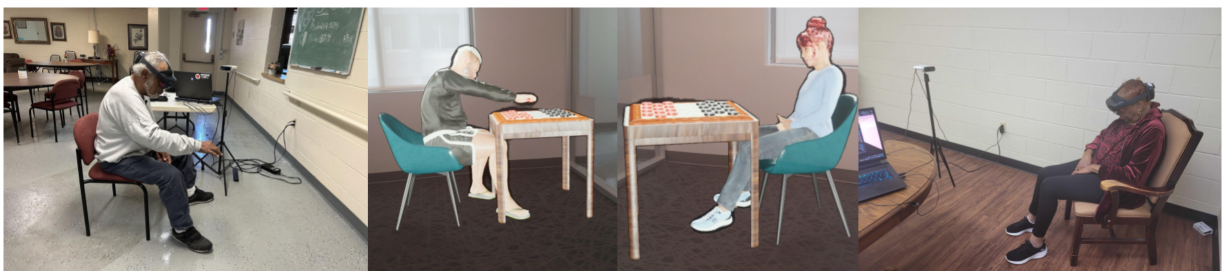

Figure 1: Two users residing in a long-term care facility playing checkers with each other remotely through their virtual avatars

Abstract

With new advancements in head-mounted display-based augmented reality (HMD-AR), networking capabilities, and software to support the embodiment of a user in a remote AR environment, AR telepresence can finally take its place as a viable Interactive Communication Technology (ICT). Recent studies and commercialization efforts provide evidence for HMD-AR's capabilities to host face-to-face interactions but with some limitations, such as the gap in knowledge about the requirements and considerations for the successful deployment of an AR telepresence system outside the lab. In this paper, we detail the development of our system and its iterative deployment over eight sessions in a long-term care facility with two older adults as participants. We present the lessons learned that can provide general guidelines for future deployment efforts of AR telepresence in the community.

Index Terms: Mixed/Augmented Reality, Interactive Communication Technology, Telepresence, Field Study

1 Introduction

Existing state-of-the-art 2D interactive communication technology (ICT) has been a reasonable compromise when face-to-face interactions are not possible, allowing for feasible audio and video communication among multiple people. Even though video can allow for an apt representation of a remote user, the 3D to 2D conversion often fails to maintain the physical scale of a person [10] or convey eye gaze and subtle body poses [6]. This can reduce the collective sense of co-presence, a vital communication component [10]. Video conferencing also forces users to adjust natural affordances [17] and may cause fatigue [8].

Modern AR can overcome some of the limitations affecting co-presence. Current head-mounted AR displays (HMD-AR) can seamlessly integrate virtual objects with the real world [9]. These virtual objects may as well be 3D photorealistic avatars of users. 3D cameras and tracking software can capture body pose and facial expressions, which, along with audio, can be mapped to an avatar. This capability defines an AR-based ICT that can create a face-to-face virtual interaction with conversation and task collaboration while simulating a sense of physical togetherness (see Figure 1).

Several studies have showcased HMD-AR-based ICT in a lab setting [12, 14, 15], but challenges in costs and logistics have inhibited its wide adaptation outside the lab. The high internet speeds needed for smooth interactions are costly, and the typical 360° camera rig used to generate high-fidelity avatars requires high computational resources [3]. Efforts for easier avatar creation and motion mapping for AR telepresence are being made by both academia [5, 7] and industry [4, 1], but have yet to achieve field-deployable methods for full body representation.

Since there is a lack of work showcasing the feasibility of AR telepresence systems with full body representation outside the lab, the following question arises: given the current state of hardware and software available for AR telepresence, how feasible is it to deploy such a system in the field with minimum complexity, technical and cognitive overhead, and set-up constraints? To

*e-mail: mahrukh.tauseef@vanderbilt.edu

2771-1110/24/$31.00 ©2024 IEEE
DOI 10.1109/ISMAR-Adjunct64951.2024.00146

answer this, we conducted a study spanning eight visits with two elderly residents of a long-term care (LTC) facility. Literature has shown that older adults living in LTCs are particularly vulnerable to loneliness due to constraints affecting their ability to meet their loved ones [11, 13]. This makes them an ideal demographic since they can significantly benefit from the enhanced social connectedness provided by AR telepresence. Additionally, LTCs present multiple challenges, such as limited shared space and varying internet availability, making them ideal for testing the practicality of our system. By successfully deploying AR telepresence in such an environment, we provide valuable insights that could be applied to other settings and encourage future development and broader adaptation of HMD-AR in the real world.

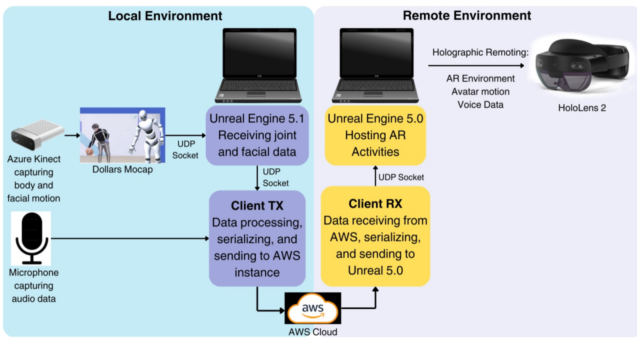

Figure 2: Architecture of the collaborative AR system used in the study

2 Methods

2.1 Data Communication

Figure 2 shows the architecture behind the communication of body, facial, and voice data from one environment to another. This architecture was duplicated for both users, allowing bidirectional communication. We ran the system on Alienware M18 laptops with 16 GB RAM and an NVIDIA RTX 4060 graphics card. Full-body and facial data was collected using a Microsoft Azure Kinect camera and Dollars Mocap software (see Figure 1), interfacing with Unreal Engine 5.1 via a UDP socket. We sent this data to a custom Python-based client that packaged, timestamped, serialized, and published avatar and audio data collected from a wearable microphone to a custom Python-based server deployed using Amazon Web Services (AWS). The AWS server handled aggregating data from multiple data streams and republishing data associated with each unique user to all connected clients. Data transfer was implemented using ZeroMQ (ZMQ), an open-source messaging framework with high speeds, low overhead, and efficient communication. The local user's audio, body, and facial data was serialized by Google Protocol Buffers (Protobuf), a widely available mechanism for storing and transferring structured data across networks. Remote clients subscribed to the AWS server's ZMQ publisher, deserialized received messages, and sent the avatar data to Unreal Engine 5.0 using a UDP socket, presenting a visualization of the remote avatar. For our implementation, two versions of Unreal Engine were required: Unreal 5.1 to meet requirements for the Dollars MoCap plugin and Unreal 5.0 since the Mixed Reality UX Tools plugin, used for hand interactions in AR, was not available for newer versions of Unreal Engine at the time of development.

Before deployment, we estimated our required network speed by taking the average size in bytes for the avatar data and voice packets sent to the network, multiplying it by the average number of each packet type sent per second, and then multiplying it by four as there is a receiver (RX) and transmitter (TX) stream for both users. We determined a minimum network speed of 60 Mbps.

2.2 AR Activities

A checkers game (total polygon count = 1.66 million) and a fireplace decoration activity (total polygon count = 1.77 million) were prototyped so that both users could interact with each other's remote avatars. We used direct hand manipulation- a standard mode of interaction in AR- for moving the checker pieces. Air tap gestures were used to manipulate items in the Fireplace activity. We used the Photon Cloud Multiplayer networking plugin to synchronize the movement of the interactive elements between the Unreal applications on either user's end. The participants' avatars were seated for both activities, across a table for checkers (see Figure 1 and beside each other for the fireplace decoration activity (See Figure 5 (a)). Some joint positions were fixed to keep the avatar sitting. However, most lower body motion was still allowed. The Microsoft Holographic Remoting application was used to run the AR applications on the HoloLens 2 to observe participant interactions.

3 Findings

3.1 Participant Information

We recruited two participants for this study for all eight sessions. Session activities are shown in Figure 4. During Session 1, participants provided informed consent, completed a questionnaire about their demographics and prior AR experience, and took a Self-Administered Gerocognitive Exam (SAGE)- a state-of-the-art tool used to evaluate thinking abilities [16] (see Figure 3). SAGE scores determined the likelihood of cognitive impairment (with a score below 17 reflecting a possibility of cognitive impairment). Both participants had normal physical abilities, i.e., they could grasp, pinch, walk, and stand easily, and they had a corrected-to-normal vision. Both participants consented to the use of their own and their avatar's images included in this paper. The approval for this study was obtained from Vanderbilt University's Institutional Review Board. Each participant was compensated with a $100 gift card.

3.2 Metahuman Creation

Metahuman avatars were created using Polycam software- a mobile application that reconstructs 3D scans using images of an object- to create 3D scans of the participants during Session 1. These scans were then used to make the first iteration of the Metahumans using Metahuman Creator (see Figure 3). This process took about 2 hours for each Metahuman. During session 2, we showed each participant iteration 1 of both Metahumans. We interviewed about the overall resemblance, facial features, skin color, hair, body shape, clothes, and their comfort with being represented by their avatar. Overall, both participants felt that P2's avatar bore a closer resemblance to him than P1's avatar's resemblance to herself. For iteration 2 of the avatars in Session 3, we decided to include the participants in the customization process. Neither participant requested changes during Session 2, but when given the opportunity to do so, they made significant changes to their Metahumans in Session 3 (see Figure 3). P1 changed her clothing and hair, while P2 changed his beard, hair, body proportions, and clothing. We again showed both customized avatars to each participant, and both were comfort-able with their own and their partner's representation. However, P1 mentioned having limited options for picking body proportions.

3.3 Testing bidirectional communication

We tested network speeds and performed short runs of the AR activities in different locations throughout the LTC. We used Google's internet speed test to measure network speeds. We considered the internet bandwidth suitable for the system if the participants observed or reported no noticeable delay or asynchronization. There

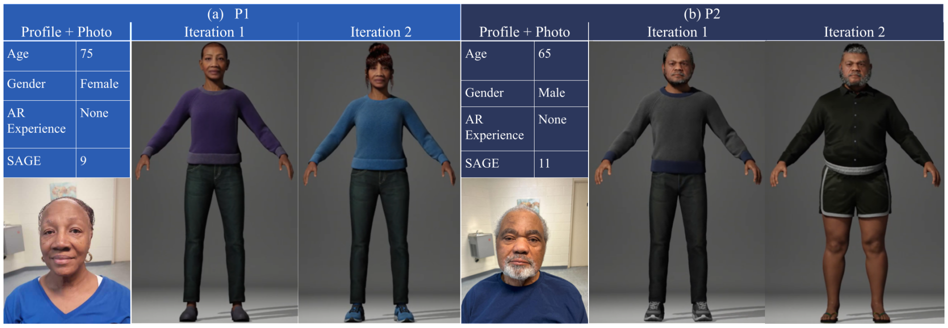

Figure 3: Participant profiles, photos, and Metahumans created during iterations 1 and 2

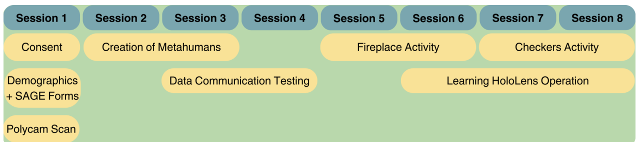

Figure 4: Deployment plan over eight sessions

were noticeable delays in avatar body movement in all except two rooms — closest to the LTC's single public router — in which we recorded network speeds of ≈ 80 Mbps.

3.4 Fireplace Decoration Activity

The initial spatial placement of the fireplace during Session 5 was set to allow both participants to sit side-by-side (see Fig. 5 (a)) and use 2D menus to decorate the fireplace with different objects and control the sound and lights in their shared virtual environment. Both participants were quick to learn the interface of the fireplace activity. During this session, we observed that the distance between the local participant and the virtual fireplace and Metahuman needed to be at least 10 feet in order to see both of them through the limited field-of-view (FoV) of the Hololens 2. Given the limited space, we reduced the distance between the virtual environment and the local user in the meeting room. As a result, the participants were unable to have both the fireplace and the Metahuman in their FoV at the same time (see Fig. 5 (b)). Consequently, we noticed that the participants were less aware of their partner's avatar and hardly turned to look at it during the interaction. For Session 6, we decided to adjust the spatial layout of the virtual environment such that both the participant and the fireplace appeared together in the local user's FoV (see Fig. 5 (c)). We observed increased verbal engagement and hand gestures, such as waving, between participants.

3.5 Checkers

Unlike the air tap interaction used in the fireplace activity, we decided to use the HoloLens 2's pinch and grab technique for picking up checker pieces to mimic real-world physical behavior [2]. During Session 7, both participants initially struggled with picking up the pieces despite having normal dexterity. They could not tell when to start the grab and pinch motion to get a hold of a piece. Their focus shifted from playing checkers against each other to learning how to move pieces by picking and putting back the pieces out of turn. There was no verbal interaction between the participants during this period. Instead, they were more often asking for assistance from the researchers conducting the study to help them move the pieces. Both participants were able to acclimate to this interaction modality towards the end of Session 7. For Session 8, we let participants practice picking up the pieces before starting the game. After a few failed attempts, we observed that both participants could grab and put pieces in the right place. We reset the checker's board and asked both participants to play against one another. They maintained a turn-by-turn play with increased verbal interaction.

3.6 Learning HoloLens Operation

During the last three sessions (6-8), we assessed whether participants unfamiliar with HMD-AR could successfully navigate the device from turning it on to entering the correct application. We created a tutorial video with step-by-step instructions. Both participants could turn on the HoloLens 2, enter a login PIN, and run the Holographic Remoting application within 20 minutes during Session 6, but they needed verbal instructions to be able to do so. During Session 7, the participants opted to view the tutorial again, asked for minimal assistance to run the app, and completed the task in less time. During Session 8, both participants opted to run the app without the video tutorial. P2 required no assistance; meanwhile, P1 asked for minimal assistance. In the end, both of them reported feeling comfortable using the headset.

4 Discussion and Conclusion

The deployment of our AR telepresence system in a low-income LTC presented challenges and insights into the field deployment of AR technologies. Even though both participants were physically in the same facility, the nature of our data communication system required round-trip travel to a remote server, emulating interactions between two remote users. Our field test highlighted several deployment limitations, such as the need for appropriate space to set up our system hardware —a dedicated PC, motion capture hardware, and HMD-AR device —and a reliable internet connection —a minimum of 80 Mbps was required for satisfactory performance. Addressing these issues is not a trivial endeavor and will require continued adoption of new advances in reality capture technologies and optimizations in network usage.

Participants' choices regarding their avatar's appearance often differed from their actual physical appearance, such as choosing different hair or beard styles. This opens the possibility of AR telepresence systems and avatar creation as a method of self-expression for users who cannot do so in real life. Enhancements in avatar customization are a critical need, potentially through the use of modeling software like Blender, to allow for greater personalization and variety in avatar creation.

Both participants experienced difficulty focusing on their partner's avatar while engaging in an activity, likely due to the limited field of view provided by the HoloLens 2. This difficulty underscores the importance of designing AR activities that accommodate the hardware's restricted FoV to ensure that key visual elements are within the user's sight. This also motivates designing tasks and AR experiences that can be quickly reconfigured to accommodate the

Figure 5: (a) Initial avatar placement concerning the fireplace and the user. (b) FoV and space constraints were cutting off the avatar with initial placement. (c) Updated placement of the avatar as an optimal solution for space constraint and limited FoV

space and shape constraints of a user's workspace. In more complex interactions, there was a noticeable decline in the participants' awareness of each other's avatars, even when the hardware's FoV was accommodated. A possible solution is more intuitive cues and simplified interactions to help maintain focus and social connection.

Contrary to common stereotypes regarding older adults' competency with new technologies, our participants overcame their initial unfamiliarity with the system, learned the built-in gestures of the HoloLens 2, and effectively performed the presented activities. This provides optimism that AR technologies can effectively target diverse or cognitively challenged populations. That said, a major limitation of this study is its small and homogeneous participant group, which restricts the broad generalizability of our findings. Our future work aims to collect insights from a larger and more diverse audience.

This study provides a snapshot of bringing AR telepresence systems out of the lab and into non-optimal environments. We believe the insights gathered can be a stepping stone for creating more inclusive, accessible, and empowering technologies for diverse populations and their equally diverse environments.

5 Acknowledgments

We gratefully acknowledge the support of the National Institute of Health (award R21AG078480), and the National Science Foundation (award 2225890) provided for this research.

References

[1] Apple vision pro user guide. https://support.apple.com/guide/apple-vision-pro/capture-your-persona-beta-dev934d40a17/visionos. Accessed: 2024. 1

[2] Direct manipulation with hands. https://learn.microsoft.com/en-us/windows/mixed-reality/design/direct-manipulation, 2022. 3

[3] A. Bartl, S. Wenninger, E. Wolf, M. Botsch, and M. E. Latoschik. Affordable but not cheap: A case study of the effects of two 3d-reconstruction methods of virtual humans. Frontiers in Virtual Reality, 2:694617, 2021. 1

[4] C. Cao, T. Simon, J. K. Kim, G. Schwartz, M. Zollhoefer, S.-S. Saito, S. Lombardi, S.-E. Wei, D. Belko, S.-I. Yu, Y. Sheikh, and J. Saragih. Authentic volumetric avatars from a phone scan. ACM Trans. Graph., 41(4), jul 2022. doi: 10.1145/3528223.3530143 1

[5] T. Combe, R. Fribourg, L. Detto, and J.-M. Norm. Exploring the influence of virtual avatar heads in mixed reality on social presence, performance and user experience in collaborative tasks. IEEE Transactions on Visualization and Computer Graphics, 30(5):2206–2216, 2024. doi: 10.1109/TVCG.2024.3372051 1

[6] P. Cukor, L. Baer, B. S. Willis, L. Leahy, J. O' LAUGHLEN, M. Murphy, M. Withers, and E. Martin. Use of videophones and low-cost standard telephone lines to provide a social presence in telepsychiatry. Telemedicine Journal, 4(4):313–321, 1998. 1

[7] G. C. Dobre, M. Wilczkowiak, M. Gillies, X. Pan, and S. Rintel. Nice is different than good: Longitudinal communicative effects of realistic and cartoon avatars in real mixed reality work meetings. CHI EA '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491101.3519628 1

[8] N. Döring, K. D. Moor, M. Fiedler, K. Schoenenberg, and A. Raake. Videoconference fatigue: A conceptual analysis. International journal of environmental research and public health, 19(4):2061, 2022. 1

[9] W. Fang, L. Chen, T. Zhang, C. Chen, Z. Teng, and L. Wang. Head-mounted display augmented reality in manufacturing: A systematic review. Robotics and Computer-Integrated Manufacturing, 83:102567, Oct. 2023. doi: 10.1016/j.rcim.2023.102567 1

[10] J. Hauber, H. Regenbrecht, M. Billinghurst, and A. Cockburn. Spatiality in videoconferencing: trade-offs between efficiency and social presence. In Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work, CSCW '06, p. 413–422. Association for Computing Machinery, New York, NY, USA, 2006. doi: 10.1145/1180875.1180937 1

[11] J. Holt-Lunstad, T. F. Robles, and D. A. Sbarra. Advancing social connection as a public health priority in the united states. American psychologist, 72(6):517, 2017. 2

[12] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P. A. Chou, S. Mennicken, J. Valentin, V. Pradeep, S. Wang, S. B. Kang, P. Kohli, Y. Lutchyn, C. Keskin, and S. Izadi. Holoportation: Virtual 3D Teleportation in Real-time. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST '16, pp. 741–754. Association for Computing Machinery, New York, NY, USA, Oct. 2016. doi: 10.1145/2984511.2984517 1

[13] H. M. O'Rourke, L. Collins, and S. Sidani. Interventions to address social connectedness and loneliness for older adults: a scoping review. BMC geriatrics, 18:1–13, 2018. 2

[14] T. Pejsa, J. Kantor, H. Benko, E. Ofek, and A. Wilson. Room2Room: Enabling Life-Size Telepresence in a Projected Augmented Reality Environment. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, CSCW '16, pp. 1716–1725. Association for Computing Machinery, New York, NY, USA, Feb. 2016. doi: 10.1145/2818048.2819965 1

[15] T. Rhee, S. Thompson, D. Medeiros, R. dos Anjos, and A. Chalmers. Augmented Virtual Teleportation for High-Fidelity Telecollaboration. IEEE Transactions on Visualization and Computer Graphics, 26(5):1923–1933, May 2020. Conference Name: IEEE Transactions on Visualization and Computer Graphics. doi: 10.1109/TVCG.2020.2973065 1

[16] D. W. Scharre, S.-I. Chang, R. A. Murden, J. Lamb, D. Q. Beversdorf, M. Kataki, H. N. Nagaraja, and R. A. Bornstein. Self-administered gerocognitive examination (sage): a brief cognitive assessment instrument for mild cognitive impairment (mci) and early dementia. Alzheimer Disease & Associated Disorders, 24(1):64–71, 2010. 2

[17] J. Zubek, E. Nagorska, J. Komorowska-Mach, K. Skowronska, K. Zielinskí, and J. Rczaszek-Leonardi. Dynamics of remote communication: Movement coordination in video-mediated and face-to-face conversations. Entropy, 24(4):559, 2022. 1