2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) | 978-1-6654-4057-8/20/$31.00 ©2021 IEEE | DOI: 10.1109/VRW52623.2021.00032

Towards Avatars for Remote Communication using Mobile Augmented Reality

Amarnath Murugan*
IDC School of Design
IIT Bombay

Rishi Vanukuru†
IDC School of Design
IIT Bombay

Jayesh Pillai‡
IDC School of Design
IIT Bombay

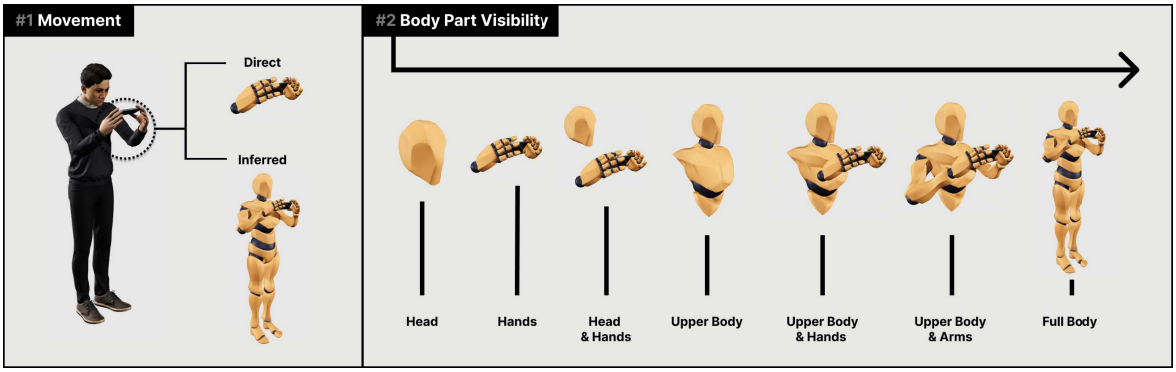

Figure 1: The Design Space of Mobile AR Avatars for Remote Communication

Abstract

Social experiences that use handheld mobile Augmented Reality with 6DOF tracking can potentially recreate the dynamics of in-person meetings, while still being more accessible than HMD based solutions. We describe a design space of possible avatars for such a platform along the dimensions of body part visibility and the type of movement. We propose a study that aims to measure the effect of five avatar types on social presence during group collaboration.

Index Terms: Human-centered computing—Mixed / Augmented reality; Human-centered computing—Collaborative interaction

1 Introduction

Among the many compelling applications of immersive media technologies, their ability to support remote communication has become particularly relevant in recent times. Around the world, in-person interaction has been reduced to a bare minimum as a result of the COVID 19 pandemic, and the use of telepresence systems across the reality-virtuality spectrum has skyrocketed. On one end, conventional video call services offer simple and accessible ways to connect with one another using everyday devices. At the other, HMD based telepresence systems are able to better recreate the spatial dynamics integral to in-person interactions. However, the requirement of specialised hardware means that few can actually use these platforms. This trade-off between device availability and interaction fidelity could potentially be addressed by Handheld Mobile Augmented Reality, given the increasing number of mobile and tablet devices capable of supporting 6DOF AR experiences.

The design of appropriate representations of remote users is an important consideration for telepresence platforms. Many studies have focused on defining the optimal characteristics of Avatars and Virtual Humans in order to increase realism and social presence in VR and Head-mounted AR systems. We believe that the modality of Mobile AR presents a few unique challenges that call for further research:

1. **Limited Pose Data:** The position and orientation of a user's head is perhaps the most important information required to control their remote avatar. In the case of Mobile AR, we only have access to the location and orientation of the mobile device, and by extension, the hands of the user. How do we best use the limited information available to control believable avatars?

2. **Multiple roles of the device:** In a remote social mobile AR experience, the same device might be used to view the other remote users, interact with AR content, and control the local user's avatar. How can we balance these competing requirements while designing avatars?

In this paper, we present an ongoing project where we attempt to address these challenges. Building upon recent studies on mixed reality collaboration and avatar representation, we explore a range of possible avatars for remote Mobile AR and describe their design. These include avatars that are directly based on the phone's relative position, and inferred avatars where a single source of information (the phone) is used to estimate how a user's body might behave. Focusing on 5 distinct avatar types with an increasing number of visible body parts, we present the design of a study that compares their effect on social presence during remote group activities.

2 Background

2.1 Avatars in Augmented and Mixed Reality

The rendering style and level of body detail are two important dimensions for virtual avatars. Studies on perception in Virtual Reality have indicated that more realistic rendering styles (such as 3D scans of real humans instead of primitive 3D models) lead to increased realism and believability in others' avatars, despite the potential for uncanny valley effect [5]. Similar findings were hinted at in work by Nassani et al. [8] where realistic avatars were preferred to

*e-mail: amarnath2105@gmail.com
†e-mail: rishivanukuru@iitb.ac.in
‡e-mail: jay@iitb.ac.in

978-1-6654-4057-8/21/$31.00 ©2021 IEEE
DOI 10.1109/VRW52623.2021.00032

cartoon-like ones when representing close social contacts. More recently, Yoon et al. [14] conducted an experiment to evaluate the effect of avatar appearance on social presence in dyadic collaboration tasks. They considered 3 levels of body detail–head and hands, upper body, whole body– and 2 levels of rendering style–realistic and cartoon-like. Their findings revealed that increased body part visibility led to higher levels of social presence. No clear difference between the rendering styles was observed, and they suggest that the choice between the two styles be made on the basis of application context.

2.2 Inferred Avatars

A recent experimental study by Eubanks et al. [2] compared inverse-kinematic avatars influenced by four levels of increasing tracking fidelity. They demonstrated that the more the number of tracking points available, the better the sense of presence and embodiment. However, as is the case in mobile AR, the number of tracking points might be limited by technological constraints. Jung and Hughes [4] demonstrated for VR that providing information of body parts that were not directly available from motion tracking or controllers (such as legs and feet in VR setups where only the head and hands are tracked) did increase the level of presence and embodiment as well. Considering the behavioural realism of inferred movements, Herrera et al. [3] have compared inferred full body avatars with inverse kinematics-based arms against an avatar with floating hands and head in a dyadic setting; their findings suggest that participants who embodied the avatar with only a floating head and hands experienced greater social presence, than participants who embodied the full-bodied avatar.

2.3 Mixed Reality Remote Collaboration

While there has been much work towards remote collaboration using mixed reality, much of it has focused on dyadic interactions, where both parties use a head-mounted display (either AR or VR), and where the avatar rendering style is mostly realistic. Walker et al. [12] studied the effect of avatar size on collaborative dynamics such as leadership and attention while using realistic virtual avatars. Related projects by Piumsomboon et al. [9] and Teo et al. [10] also explore the use of realistic 3D reconstructed avatars for use at different scales.

2.4 2.4 Mobile AR Remote Collaboration

There has been limited work on the use of mobile AR as a collaborative tool in general. The CollabAR project by Wells and Houben [13] was among the first to study this in co-located settings. In two related studies, Muller et al. [6, 7] evaluated the usefulness of shared virtual landmarks and visual rendering of augmented scenes (overlay or pass-through) on collaboration in both co-located and remote settings. The setup involved did not consider the visual appearance of the remote user, and used a frustum to indicate their gaze direction. Similarly, a context-specific project by Datku et al. [1] about the use of remote mobile AR for distributed crime scene investigation looked into the interaction methods required for such a system, but not the avatar representation.

Based on this survey of prior work, we believe there is a need for research that studies remote collaboration in AR using mobile devices in group settings. Determining the appropriate visual representation of remote users in order to enhance social presence, particularly given the technical constraints involved, is an important first step.

3 Design Space of Avatars for Remote Mobile AR

We explore a Design Space of possible Avatars across two dimensions–Avatar Movement and Body Part Visibility. The range of possible avatars across these two dimensions is shown in Fig 1.

3.1 Avatar Movement

We explore two types of avatar movement based on the phone's translation and rotation - Direct and inferred. Furthermore, we would not be adding collisions between avatars, because in social AR offsetting the virtual position of an user would make it seem like all the virtual objects are drifting away, which might be perceived as a tracking issue. Retargeting techniques are also avoided so as to not introduce spatial inconsistencies that might hinder communication.

3.1.1 Direct Movement

As the name suggests, avatars with direct movement are those that consist of a single 3D model which merely translates or rotates according to the positional data received from the handheld device. The most accurate 'avatar' would be that of a mobile device being held by two hands, as that is the only true information available. However, this could be replaced with an avatar head, upper body (as is the case in vARa [14]), a full body, or other types of characters as well.

3.1.2 Inferred Movement

In contrast to the direct approach, we can use the single source of position and orientation to infer or extrapolate possible body movements of the remote user. Prior literature has demonstrated that avatars that display a larger range of body parts and body movements tend to increase social presence. While it may not be possible to derive highly accurate pose data, or attain a one-to-one mapping between the user's pose and their avatar's inferred pose, we believe that the intention of the remote user can be inferred and communicated to sufficient levels of accuracy.

3.2 Body Part Visibility

The visibility of different body parts dictates the level of inference that must be present in the avatar movement. Direct movement might not be suitable for a full humanoid avatar, while inferred movement is not needed for an avatar that consists of just a phone model with two hands holding it. These two examples lie at the opposite ends of a spectrum of possible avatars, similar to the experimental study by Yoon et al. [14].

The positive relationship between body part visibility and social presence has been established in conditions where all the relevant body parts are being tracked through headsets and/or controllers. In the case of remote communication using Mobile AR, we are faced with an intriguing trade-off. Increased body part visibility might still lead to an increase in social presence, but avatars with more body parts would require more inference in their animations (such as the limbs and torso for a full humanoid avatar) which in turn might be perceived as uncanny and reduce social presence. We are particularly interested to study this interaction between avatar movement and body part visibility. As a result, we chose not to consider the avatar rendering style as a separate dimension, and instead use a consistent humanoid model with only the essential details of limbs and joints.

3.3 Final Avatars for Consideration

Of the many possible combinations of avatar movement and body part visibility, we chose to focus on 5 distinct avatars (highlighted in Fig 2). These are:

A **Hands Only, Direct:** This avatar is meant to function as the 'ground truth', as it is the most accurate representation of the spatial information available. Without a clear face or body to focus on, we are also interested to observe whether users are able to visualise the remote participants solely from this limited information.

B **Upper Body, Direct:** This avatar provides a clear target for users to focus on, has more visible body parts, but not so many

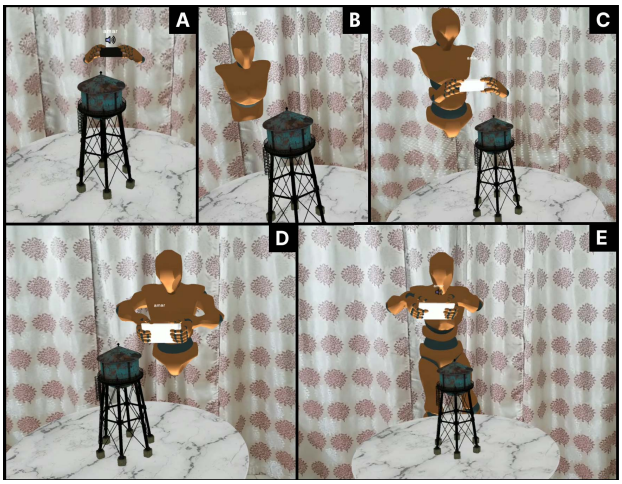

Figure 2: Screen captures of the prototype implementations for the 5 final avatars.

that the movement would appear to be obviously unnatural. vARa [11], the only other Social AR platform we have come across, uses low-poly characters of this nature, and so such an avatar would be a useful benchmark.

C **Upper Body + Hands, Inferred:** Here, we have a 'floating' avatar that displays all body parts from above the hip. The focus of the avatar's head, the hands, as well as the vertical position of the avatar are all inferred. We believe that this level of inference may work fairly well without introducing artefacts of unnatural movements.

D **Upper Body + Arms, Inferred:** This condition is similar to (C) however we also display the arms of the avatar. A recent study by Herrera and Bailenson [3] compared these two conditions (Only Hands, hands & arms) for VR avatars, and we would like to see if the results translate to handled AR.

E **Full Body, Inferred:** With the most number of visible body parts, this avatar also requires the most extrapolation for movement, and serves as the ideal goal for inferred avatars. We suspect that our current implementation might lead to less-than-ideal behavioural realism, and are including this configuration in the study to see if that is indeed the case.

The next section details out the implementation of avatars with inferred movement, after which we describe a proposed experimental plan to study the effect of these 5 variations on social presence and behavioural realism.

4 Implementing Inferred Movement

4.1 Estimating Position & Orientation

To represent the users better, we propose a calibration step to match the avatar's dimensions to better suit their arm reach and height. In order to make up for the lack of more points of reference in handheld mobile AR, we approach this step as follows. First, the user is asked to tap on a scanned area in the ground and stand on top of the same, this would serve as a proxy for the feet position. The distance between this and the phone's position projected onto the ground is used to get the user's arm reach. To get a better estimate of arm reach, the user is asked to extend their arm and hold the phone parallel to their chest, furthermore, their height is also estimated by using the phone's offset from the ground.

After the model is instantiated, the limits of comfortable arm reach movement is estimated by using the arm reach value. We found limits between 60% and 80% of arm reach to work well. When the phone's position falls outside this range, it usually implies that the avatar must be moved forward or backward. It is also possible that the user has moved to their side when the position exceeds the upper limit, as the distance is calculated between the proxy feet and the projected phone positions, the direction of this vector must also be considered to determine the new position of the avatar. Thus, lateral movement is added after measuring the angle between this vector and the vector pointing in the direction the avatar is facing i.e the forward vector, and if it is found to be greater than the lateral deviation threshold (20 degrees in our case). For vertical movement, the phone's distance from the ground is compared against the avatar's chest offset from the ground; if they are greater than 10-20 cm, the avatar is shifted to keep the difference within this range.

These heuristics were derived after observing people use our Social AR platform. They tend to hold the phone close to half their arm reach in front of them, and move their body to navigate and orient themselves. To view the phone comfortably, the device is not moved or rotated by itself to a larger extent, additionally, this also implies that that the user is usually facing the same direction as the phone.

When the phone's position exceeds either the comfortable arm reach limits or the lateral deviation threshold, the user is assumed to have moved from their position and new position and orientation values are estimated. With the observations mentioned above, we set the new position of the avatar half its arm reach away from the phone but in the opposite direction of the phone's forward vector projected on the ground. The avatar is interpolated to move to this position and rotated such that its forward vector is pointing in the same direction as the project forward of the phone.

4.2 Responsive Body Movement

The above mentioned steps are used for all inferred avatar types to get their position and rotation, but depending on the avatar type we also animate different body parts to make them appear responsive. In both inferred upper body only and the full body avatars, the head is rotated to face the phone at all times.Additionally, when the phone's orientation switches between portrait and landscape mode, we snap the avatar upright for direct type, and for inferred type we switch the pose of the fingers to correspond with the way the phone would be held at that orientation.

For the full body avatar, the limbs have to be animated along with its translation and rotation as well. We approach this with a semi-procedural system, the arm movement is obtained through Inverse Kinematics (IK) with the phone's position as a target. For the leg movement, we use two different animation sets. The first set consists of keyframed walking animations for four directions and an idle animation, and the second set consists of similar animations but for crouched stance. The final animation is obtained by first interpolating between the first and second set depending on the vertical offset of the mobile, and then interpolating within each set based on the direction in which the avatar must move.

There are multiple drawbacks to this system. The step length is controlled by interpolating between the idle and walking pose, but it is difficult to obtain an exact step length that would correspond to the velocity of the model while translating. Any changes to the extreme poses must be done by modifying the keyframes, which further reduces the control over the pose in run-time. We hope to use a completely procedural approach to gait to improve the responsiveness of the avatars.

5 Proposed Experimental Study

5.1 Design & Measures

We propose a 5x2 within-subjects study, where the independent variables are as follows:

- *Avatar Type:* Hands Only - Direct, Upper Body - Direct, Upper Body + Hands - Inferred, Upper Body + Arms - Inferred, Full Body - Inferred
- *Task Type:* Social Communication, Artefact-centric Communication

We are primarily interested in the social presence evoked by the avatar types and their differences across tasks, for which we will be using the Networked Minds questionnaire and the Social Presence Scale. Through observations and post-task interviews, we will record subjective feedback regarding the level of presence, nature of communication, and degree of behavioural realism of the avatar movements. We will also use this feedback to determine the contexts in which each avatar is better suited.

5.2 Apparatus

The experiment application was built in Unity, and networking was implemented using Photon. Currently the application is targeted for Android devices with ARcore support. Inverse Kinematics is implemented using Unity's animation rigging package, and the models and animation sets of the full body avatar are taken from Mixamo.

5.3 Method

The experiment would be conducted remotely, so participants will be screened by device availability, and are expected to use their personal device. After consenting to participate, they are requested to use specific AR applications in the market, to become familiar with the possible interactions in the platform. Participants will be reminded that anonymous usage logs will be collected during the course of the experiment. Where possible, and if participants provide consent, we will also ask them to record their movements using a camera or mobile device for further analysis. Following this, they join the multiplayer lobby, and wait for the experimenter to start the experience. Each session would have five concurrent users, of which four would be experimenters. The same set of four would be present for all participants, in order to keep the movement behaviour seen by them consistent.

After everyone joins the lobby, the AR experience is started for all users. Once the floor is scanned, the user is taken through the calibration process mentioned in section 4.1, and are asked to place a central artefact (a table in this case) in front of them; to establish a reference point around which other users' avatars would be instantiated. With this completed for all users, they go through the social task for each of the five avatar types (counterbalanced). Participants fill the questionnaires after completing the task for each avatar type before proceeding to the next. Finally, they repeat the same with the artefact-centric task.

The social activity is a game of 20 questions, where a player chooses a word and the other players ask yes or no questions to figure out the word. This game was chosen for two reasons, firstly, it keeps the focus of the users on the avatar of another, second, it is likely to foster both one on one (asking questions and answering) and group conversation (discussing to figure out the word). We propose group tasks of five people in a session as opposed to dyadic interactions to better study the influence of multiple avatars, and since prior literature has mostly studied avatar perception in dyads.

The artefact-centric activity is intended to have the users work towards a common goal based around an object in AR. Such an activity points to more structured use-cases such as classroom learning and work meetings. Furthermore, we wanted the activity to have frequent context-switches between the artefact and the avatars, while nudging participants to walk around in the space. This is likely to provide insights on how the participants perceived the avatar when the focus shifts between them and an AR object, and the movement of the avatar itself. We plan to use short tasks that involve the inspection of and comparison between complex 3D solids, similar to those used in the CollabAR study [13]. The participant would be presented with 2 different 3D models with a surface made of multicoloured tiles. Together with the experimenters, the participants will need to inspect both models, and identify the colours that do not appear on one of them. Such a task would require all members to move around and inspect the objects, while also discussing among themselves to corroborate their results. For each of the 5 avatar types, we will be presenting one such comparison task. Participants will be presented with short questionnaires to measure social presence and behavioural realism after each task-avatar pair.

Once both task sets are completed, we will conduct a short interview with the participants to understand their experience and gain more subjective feedback on their perception of the avatars. We will then analyse the movement logs and video recordings (where available) to gain a better understanding of how well the inferred movements matched the real ones.

6 Conclusion

Remote collaboration has become integral to everyday life as a result of the COVID-19 pandemic, and there is a pressing need for teleconferencing solutions that recreate the dynamics of in-person meetings. Social experiences that use handheld mobile Augmented Reality with 6DOF tracking can potentially offer a more spatial experience in remote settings, while still being a more accessible than expensive HMD based solutions. However, with just one point of reference, the avatar representation for such a platform brings forth new challenges. In this work-in-progress paper, we explore a design space of avatar representations for remote collaboration in handheld mobile AR. We describe the prototypes developed for these avatars, and propose a tentative plan to study the effect of 5 distinct avatar types on the social presence and behavioural realism during remote group communication. We hope to use the workshop to seek feedback on our avatar prototypes and experimental design, and discuss a plan of action for further research on avatars and virtual humans in mobile AR settings.

References

[1] D. Datcu, S. G. Lukosch, and H. K. Lukosch. Handheld augmented reality for distributed collaborative crime scene investigation. In *Proceedings of the 19th International Conference on Supporting Group Work*, pp. 267–276, 2016.

[2] J. C. Eubanks, A. G. Moore, P. A. Fishwick, and R. P. McMahan. The effects of body tracking fidelity on embodiment of an inverse-kinematic avatar for male participants. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 54–63. IEEE, 2020.

[3] F. Herrera, S. Y. Oh, and J. N. Bailenson. Effect of behavioral realism on social interactions inside collaborative virtual environments. *PRESENCE: Virtual and Augmented Reality*, 27(2):163–182, 2020.

[4] S. Jung and C. E. Hughes. The effects of indirect real body cues of irrelevant parts on virtual body ownership and presence. In *Proceedings of the 26th International Conference on Artificial Reality and Telexistence and the 21st Eurographics Symposium on Virtual Environments*, pp. 107–114, 2016.

[5] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, pp. 1–10, 2017.

[6] J. Müller, R. Rädle, and H. Reiterer. Remote collaboration with mixed reality displays: how shared virtual landmarks facilitate spatial referencing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 6481–6486, 2017.

[7] J. Müller, J. Zagermann, J. Wieland, U. Pfeil, and H. Reiterer. A qualitative comparison between augmented and virtual reality collaboration with handheld devices. In *Proceedings of Mensch und Computer 2019*, pp. 399–410. 2019.

[8] A. Nassani, G. Lee, M. Billinghurst, T. Langlotz, and R. W. Lindeman. Using visual and spatial cues to represent social contacts in ar. In *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*, pp. 1–6. 2017.

[9] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pp. 1–13, 2018.

[10] T. Teo, L. Lawrence, G. A. Lee, M. Billinghurst, and M. Adcock. Mixed reality remote collaboration combining 360 video and 3d reconstruction. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–14, 2019.

[11] vARa. vARa: A Social XR Metaverse. `https://www.varanow.com`, 2020.

[12] M. E. Walker, D. Szafir, and I. Rae. The influence of size in augmented reality telepresence avatars. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 538–546. IEEE, 2019.

[13] T. Wells and S. Houben. Collabar–investigating the mediating role of mobile ar interfaces on co-located group collaboration. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.

[14] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinqhurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 547–556. IEEE, 2019.