Placement of Teleported Co-Users in AR

Jens Reinhardt[1], Marco Kurzweg[2], and Katrin Wolf[2]

[1] University of Applied Science Hamburg, Hamburg, Germany
jens.reinhardt@haw-hamburg.de
[2] University of Applied Science Berlin, Berlin, Germany
firstname.lastname@beuth-hochschule.de

Abstract. Teleportation and conversations with virtual representations of remote people have been made possible by recent developments in augmented reality (AR) technology. This paper aims at understanding how such AR telecommunication systems should be implemented by asking where to display 3D scans of potential remote users. As the perfect interaction design solution may be different while walking versus while staying in one place, we conducted a user study comparing both. We also varied the placement of the remote user in the co-user's field of view (FoV) and where the coordinate system in which the 3D scan is visualized has its origin. We found that remote users we talk to should, in general, be visualized in AR in front of us, but in situations in which the physical world requires attention a visualization in the periphery is better. Re-placing the co-user through gestures is not desired, but the ability to look away from them should be supported, which strongly supports placing virtual co-users in AR relatively to the user's body.

Keywords: augmented reality · hologram · teleportation · co-user · remote user visualization.

1 Introduction

Teleportation using Augmented Reality (AR) can enormously enrich telecommunication and realizes an old humankind's dream of talking to remotely located people and seeing them as if they would be in the same place we are. The affordability of AR glasses and recent technological advantages, e.g., to display remote people in real-time on AR glasses, enable us to talk to remote people more naturally than through traditional telecommunication technology. Besides a potential increase of sensed co-presence [22], seeing the face of a speaking person helps to understand the spoken word [27], and to see the lips of the speaker enables the listener to hear and thus better understand [25]. We believe that such hologram-like telecommunication will change the way we live and work together. We could better maintain relationships with friends and family when being away, and we would not need as many business travels as today if meetings using a technical solution like Microsoft's *Holoportation* [22] would become ubiquitous.

Holoportation displays a remote user on AR glasses so that they seem to stand beside us. While this seems to be perfectly fine when being at home while communicating with remote users, other (more mobile) solutions may be better suited for telecommunication on the go. Previous research on AR and VR video- or hologram based communication investigated technical solutions for recording, transmitting, and visualizing the remote conversation partners [6, 22, 32]. It has also been investigated how digital content, like text, has to be arranged and positioned when using smart glasses [8, 29, 21, 23]. Different aspects influence the placement of AR content. While AR content placed in the center may occlude important real-world content, peripheral vision is limited [26]. However, peripheral cues can also grab the user's attention [12] as humans can well perceive motion in their visual periphery [7]. Thus, the question of where to position a remote user's visual representation during AR telecommunication remains unclear and defines the research gap that we target in this paper.

In a user study, we varied the position of remote users' visual representation, placing them in the space beside the user (fixed world position), around the user in a way that the visualization would "follow" the user if they walk, or in the user's field of view (FoV). As peripheral visualizations versus centered visualizations lead to different attention [5] and perception [26], we also vary that attribute. Our results indicate no differences between scenarios, but indicate that placement gestures lack usability while a center placement and attaching the AR co-user to the user's body works best as it allows the user to look away if real-world content requires attention, while otherwise facing the co-user make the conversation most natural.

With this work, we contribute by recommending how we can display remote users we are talking to, which can improve the usability of next-generation telecommunication software for future AR glasses. Furthermore, it will enable us to stay in better contact with distant family and friends and reduce the need to travel, for example, to business meetings.

2 Related Work

This paper aims to improve communication with remote users displayed at AR head-worn optical see-through (OST) devices. Thus, we discuss works that previously explored AR telepresence systems; content presentation, arrangement, and positioning for AR glasses; and techniques to visualize remote co-users in AR.

2.1 AR Telepresence Systems

Billinghurst and Kato introduced an AR conferencing system that uses the superimposition of virtual images with the real world [13]. Remote co-workers are displayed on virtual monitors that can be freely placed in space by a user with the help of tracking markers.

Velamkayala et al. investigated AR collaboration with an interface that allowed the *HoloLens* user to see the face of the remote collaborator [30]. They compared the performance of teams in a navigation task performed while using the *HoloLens* and smartphone when using *Skype* [3]. Their results showed that the task completion time was longer when using the *HoloLens*, but the number of errors was lower when using that device. They concluded that the longer completion time was due to the design and unfamiliarity with the new technology. In contrast to our research focus, neither co-presence, telepresence, and social presence were evaluated, nor alternative placement of the remote user was explored. Lawrence et al. performed a pilot study to investigate the influence of video placement in AR conferencing for an assembly and a negotiation task [16]. They evaluated the different positions of the video window (fixed position in the head-mounted display's (HMD) field of view (FoV), fixed world-position, only audio) of Skype for the *HoloLens* for the different tasks. They evaluated communication, co-presence, and user preference. Their findings show that co-presence was rated significantly higher in the assembly task than in the negotiation task for the fixed world position. Moreover, they found that users preferred the two video conditions and rated the pure audio condition significantly lower. While this work varied placement of AR communication partners in static setups, we also compare the positioning of co-users' visualization depending on the mobility of the user (on the go versus staying in place).

2.2 Presentation and Arrangement of Content in AR Glasses

AR allows superimposing the real environment with additional information [10] that can be perceived while working on a real object and without losing focus on the object being worked on. A large body of research investigated how to present, arrange, and position this additional information.

Tanaka et al. investigated how information must be arranged on AR glasses focusing on viewability [28, 29]. Although the HMD can display information without interfering with the user's view, the information displayed may be complicated to see if the view behind the display is too structured, the background is too bright, or the contrast is too low. Their research focuses on the automated alignment of AR content based on the environment. Users' preferences are not taken into consideration. Chua et al. investigated the influence of the physical position of the display of monocular smart glasses on the performance in dual-task scenarios [5]. They compared nine display positions for displaying notifications and found that the notifications at the middle and lower-middle positions were noticed more quickly. Moreover, they found that the upper and peripheral positions were more convenient, less obtrusive, and favored. In particular, the best compromise of performance and usability in a dual-task scenario was obtained by the center-right position. The positioning of content at AR glasses often concerns textual information, which is targeted by research discussed in the following paragraphs. Orlosky et al. investigated a text management system for AR headsets in mobile outdoor scenarios that actively manages the movement of text in a user's FoV [21]. Camera tracking is used to identify dark areas in the user's FoV, which are then used to display textual content to maximize readability. Their

[3] https://news.microsoft.com/de-de/videos/microsoft-hololens-skype/

results indicate that dynamic text placement is preferred over text placement at fixed positions due to better text readability in dark areas. Rzayev et al. investigated text presentation on AR smart glasses and how walking, text position, and presentation type affect comprehension, reading and walking speed, and workload [23]. They found that text displayed in the upper right corner of smart glasses increases the subjective workload and reduces comprehension. Klose et al. investigated how the reading of text on AR glasses and the simultaneous execution of three real-world tasks influence each other [14]. They compared text placed in a head-locked coordinate system, a body-locked coordinate system as well as text placement at the top and bottom during the following tasks: a visual stimulus-response task, a simple walking task, and an obstacle course task. For the head-locked presentation, the text was fixed relative to the AR glasses FoV (top and bottom). For the body-locked presentation, the text was fixed at the height relative to the user (top and bottom). Their findings show that AR reading negatively influenced performance in all three tasks. The statistical analysis of user preferences showed no main effects of task, height, or coordinate system, but an interaction effect of task and coordinate system attachment was found. Participants preferred head-locked positions during the visual stimulus-response task. Simultaneous reading in AR during the obstacle course was significantly more distracting than when merely walking. Furthermore, the body-locked text was significantly more disturbing for the visual stimulus-response task than the head-locked text. At the same time, the anchoring of the coordinate system made no difference for the other two tasks.

While previous research examined the placement of textual and static content for AR in head-worn OST devices, we explored the placement of animated content, which is the co-user's visual representation. It is commonly known that moving objects are well perceivable in the periphery, while static objects are not [2, 26]. Hence, it is worth exploring where to best place moving content on AR glasses.

2.3 Co-User Visualization in AR

In the area of AR co-user visualization, research has been conducted on the realistic representation of gaze as well as on the virtual co-user integration in the local user's environment. Furthermore, work on partially obstructed transparent displays is related to this project and will be presented here.

Orts-Escolano et al. introduced *Holoportation*, a technique for augmented reality, which transfers a remote user into the augmented environment [22]. The work focuses on the capturing and 3D reconstruction of the remote user, but not on their placement. An immersive teleconferencing system, which seats participants at a virtual table while allowing each participant to gauge each other's gaze, was proposed by Zhang et al. [31]. Another approach to achieve a realistic meeting experience was introduced through *TELEPORT* by Gibbs et al. [11]. Here, a power wall is used to create a virtual extension of the meeting room wherein the communication partner is placed. *TELEPORT* is designed to give users a realistic physical context, and realistically represent the remote user's gaze. Nassani et al. explored if changes in visual representation, proximity, or a combination thereof, depending on social relations, influence how natural the experience is [19]. They discovered that visual representation is preferred over proximity, and that visual representation and the combination with proximity is preferred over the lack thereof. *Tracs*, introduced by Lindlbauer et al., is an approach on transparent displays, which allows for privacy in parts of the screen while facilitating collaboration by making areas of the screen transparent or shared with the other side [17, 18]. Lindlbauer suggests that this might promote collaboration with close spatial co-workers. Angos et al. presented an approach for dynamically manipulating scale, orientation, and the position of miniaturized co-users as holograms, which guarantees eye-contact [1]. In a preliminary study, they found that dynamic scaled remote people help obtain eye contact by considering differences in height, positioning, orientation, and the characteristics of the surrounding space.

Previous research investigated the visual representation, capturing, display, and transmission of the co-user. We are interested in the influence of positioning and spatial referencing of the co-user in different scenarios. Therefore we researched the placement in a static setup as well as when walking around.

2.4 Summary

Previous research has been focused on technical feasibility, capturing, transmission, and presentation of life-like 3D representations of co-users in AR. Moreover, while research on placement of content in AR, particularly on the placement of textual information and video call windows has been done, no previous work looked at the placement of 3D reconstructed remote users in head-worn OST AR devices. Research on the positioning of a life-like 3D representation of the co-user in AR distinguishing mobile and stationary scenarios has not yet been carried out. This research gap is addressed in this paper.

3 Method

To better understand how and where the virtual representation of a co-user has to be positioned in AR, we conducted an experiment. We compared three different positions of a virtual co-user in a controlled experiment with three different reference coordinate systems for a video call task in a walking and stationary situation.

3.1 Design

Our study had a 3x3x2 within subjects design with the independent variables *point of reference* (head-fixed, body-fixed, world-fixed), *placement* (center, periphery, interactive), and *scenario* (mobile, stationary). As a Latin square would not equally distribute 18 conditions across 18 participants, we had varied our 3x3 conditions in a Latin square order while having the last variable randomized within that order [3]. The dependent variables were *self reported co-presence*, *perceived others co-presence*, *social presence*, *telepresence*, and *usability.*

3.2 Measurements

Self reported co-presence, *perceived others co-presence*, *social presence*, and *telepresence* were measured by the Nowak and Biocca questionnaire [20].

Co-presence: is the feeling of belonging between two people [20]. Co-presence was measured on two different scales, one relating to the participants' perception of their partner's contribution to the interaction (*perceived others' co-presence*) and the other relating to the self-report of their contribution to the interaction (*self-reported co-presence*) [20].

Social presence: is the perceived ability of the medium to connect people [20].

Telepresence: is a measure of the feeling of "being there" to be "within" a virtual environment that a person has [20].

Usability was recorded using the System Usability Scale (SUS) questionnaire [4]. To better understand the quantitative data, we collected additional qualitative feedback in semi-structured interviews. Here, we were asking regarding the interactive *placement* condition:

- If you have the free choice to place the co-user to get the maximal usability, where would you place them?
- Why would you place them there?

Alternatively, we asked for the center, and periphery *placement* condition:

- Regarding the usability, what was beneficial about the co-user's position?
- Regarding the usability, what was obstructive about the co-user's position?

3.3 Participants

The experiment was conducted with 18 participants (6 females, 12 males) aged between 22 and 35 years and an average age of 27.8 years ($SD = 4.0$), recruited at our university campus. 17 (6 female) participants had experience with AR.

3.4 Apparatus

Meta 2 glasses[4] were used as visual AR output device. We used the backpack computer (MSI VR ONE, i7 7820HK/2,9 GHz, 16GB RAM, GTX 1070) to run our prototype software, as the Meta 2 glasses have no data processing unit.

[4] https://www.metavision.com

Fig. 1. Screenshots of our prototype of the female (left) and male (right) representation of the co-user.

The apparatus was implemented in Unity3D[5] presenting the co-user as two randomly selected visualizations of a virtual phone caller (male/female) on the AR glasses. For these two virtual user representations, we recorded separate volumetric videos using Microsoft Kinect.

The user interface for the study participants was kept as simple as possible and solely displayed the virtual co-user (see Figure 1). For the virtual representation of the co-user (remote teleconferencing caller), we used three different reference points that define where the coordination system (in which the co-user is visualized) is attached to (or fixed to, see Figure 2). For the co-user's representation, we decided to use pre-recorded videos instead of live recordings and transmissions to keep the system's performance as high as possible. Moreover, this ensured that all participants would see the same performance of the callers' presentation, regardless of any errors in the live generation of the volumetric video or the transmission of it. The volumetric video consisted of a textured point cloud. To display the silhouette and surface of the co-user, the individual points of the point cloud were enlarged until the impression of a flat texture was created. To avoid gender biases, we recorded one female and one male person representing the callers.

[5] https://unity.com

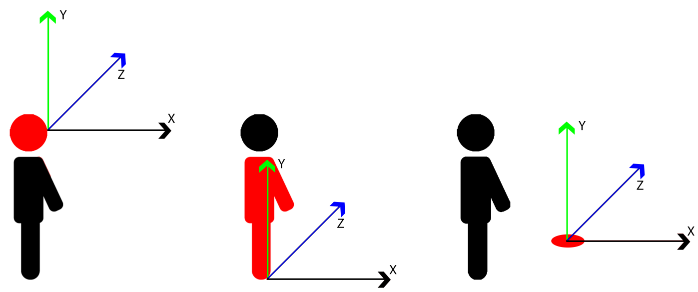

Fig. 2. *Point of Reference*: *head-fixed*, which means that the AR content is fixed to the head position and follows head movements (left), *body-fixed*, which means that the AR content is fixed to the user's feet and follows when the user walks or rotates their body (center), *world-fixed* means that the AR content stay in the real world even if the user looks or moves away (right).

Point of Reference We compared three different coordinate systems where the co-user's virtual visualization referenced (see Figure 2.)

Head-Fixed: The *head-fixed* coordinate system had its origin (or reference point) directly at the user's forehead. Therefore, the coordinate system is affected by head movements and the virtual visualization is following head rotations and tilts. Consequently, this coordinate system is representing the FoV (see Figure 2 (left)). As the origin, we used the origin of the virtual Meta 2 camera in the Unity scene with applied camera rotation to ensure that the virtual visualization follows the FoV.

Body-Fixed: The *body-fixed* coordinate system had its origin (or reference point) at the feet of the user. When moving through space, this coordinate system is carried by the user, but the origin remained unaffected when the user turned its head (see Figure 2 (center)). As the origin, we used the camera (head) position of the virtual Meta 2 camera in the Unity scene and subtracted the participant's height (which we measured at the beginning of the experiment). The origin was placed at users' feet to ensure the virtual visualization is always located at the ground, which was realized by subtracting the height of a participant from the origin.

World-Fixed The *world-fixed* coordinate system had its origin (or reference point) at the origin of the Unity scene (see Figure 2 (right)).

Placement Within the coordinate system that might be *head-*, *body-*, or *world-fixed*, we compare three possibilities where the co-user's virtual visualization is placed (see Figure 3).

Center: The virtual co-user was positioned centrally (see Figure 3 (left)). For the *head-fixed point of reference* the co-user was placed 2m in front of the user, while for the *body-fixed* and *world-fixed point of reference* the co-user was placed

Fig. 3. *Placement* conditions: *center* (left), *periphery* (middle), *interactive* (right).

3m in front of the user. The vertical alignment was three degrees downwards so that the co-user is fully visible and at eye level. The horizontal alignment was 0 degrees, exactly centered. We have chosen a smaller distance for the head-fixed coordinate system due to the walking course's nature. When participants walked and turned their heads in the direction of a wall bordering the course, the possibility of clipping the co-user into the wall possibly occurred. These clippings were greatly reduced by the decreased distance of the the co-user's representation.

Periphery: The virtual co-user was positioned in the right periphery (see Figure 3 (center)). For the *head-fixed point of reference*, the co-user was again placed 2m in front of the user and vertically aligned three degrees downwards. The horizontal alignment was 15 degrees to the right. This angle was chosen because the density of sensory cells on the human eye's retina decreases very strongly from 15 degrees eccentricity [9, 24]. For the *body-fixed* and *world-fixed point of reference* the co-user was placed 3m in front of the user. The vertical alignment was the same as for the *head-fixed point of reference*. The horizontal alignment was 25 degrees to the right.

Interactive: Here, participants had to interactively set the position of the co-user's visual representation through a pinch gesture, whose realization is described below (see Figure 3 (right)).

Although the Meta 2 has gesture recognition, it was too error-prone for our purposes. Therefore, we implemented a prototypical pinch gesture using Python[6] and OpenCV[7]. We equipped the index finger and the thump with a red and blue marker and implemented a color tracking algorithm to recognize the fingertips. If both fingertips touched each other, a pinch gesture, similar to the commonly known HoloLens gesture, was recognized. The user "clicked" on a position on the ground, as seen from its perspective, using the pinch gesture. The co-user then was placed on this location on the ground. The Python application ran as a separate application as a separate process. The communication with the Unity application was realized via network sockets.

Scenario For our experiment, we compared a stationary and a mobile scenario.

Stationary scenario: In the standing scenario, we placed every participant in the same position and with the same viewing direction into the room that an

[6] https://www.python.org
[7] https://https://opencv.org

unobstructed view into the room was ensured. The participant was not allowed to leave the position, but head and body movements were allowed for better observation of the caller.

Mobile scenario: In the mobile scenario, the participant had to follow a marked rectangular path. All participants were placed on the same starting point. Head and body were allowed to be turned as desired while walking. They were not permitted to leave the path. The path was chosen in a way that there were no obstacles on the track. The side lengths of the chosen rectangle were 6m x 4m. Thus, a distance of 20m was completed per lap. The participants were free to choose their running speed. Thus, the walking distance varied from participant to participant.

3.5 Procedure & Task

First participants were welcomed and then asked to complete a consent form and fill in a demographic questionnaire. We also collected information about previous experience with AR. We measured the height of the participants to align the correct position of the ground plane.

At the beginning of the experiment, participants got an introduction to the Meta 2 device and were then shown how to execute the pinch gesture to place the co-user interactively. To avoid sequence effects, the order of conditions in which we varied *point of reference*, *placement* and *scenario*, followed a Latin square design. We equipped the participants with a VR backpack and the Meta 2 glasses. We asked the participants to proceed to the starting point marked on the ground, depending on the scenario.

The participant's task was to observe the caller. The participant should imagine a conversation situation with the remote caller and then evaluate their positioning and referencing. The caller informed the subscriber that a book borrowed from the participant was still in possession, that the book was very enjoyable, and that it would be returned soon. The spoken text of the virtual co-user text was chosen unemotionally to not bias the ratings. We have deliberately chosen not to engage in a real dialogue so as not to distract the participant.

Once the participants were in the right position, we showed them the caller. We used a Wizard of Oz design, so the experimenter actively performed the actions of the pre-recorded caller. The caller's visual appearance was previously communicated to the participants as a visual ringing and symbolized the incoming call. If the placement condition was interactive, the participant first had to place the caller through the interaction gesture. If the participants were ready to answer the call, they should loud and clearly say "Answer". The experiment leader had to start the recorded volumetric video by manually press a button in our prototype software. The volumetric video was placed according to the *placement* and referenced in the 3D space according to the *point of reference* conditions. In each condition, a male and a female co-user were shown simulating a video call. Both the co-user's appearances, the female and the male, were displayed in random order for each condition. Finally, the participants filled in

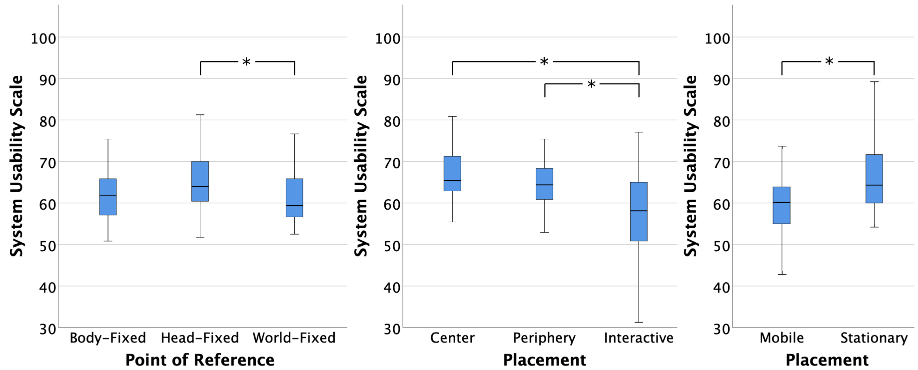

Fig. 4. *System usability scale for point of reference,placement, and scenario*

the questionnaires and repeated this procedure for all 18 conditions of our experiment. The experiment lasted between two and three hours, including the completion of the individual questionnaires, and several breaks.

4 Results

We used independent Friedman tests with *point of reference* (head-fixed, body-fixed, and world-fixed), and *placement* (center, interactive, and periphery) as independent variables to indicate significant effects on the ordinal data *self reported co-presence*, *perceived others' co-presence*, *social presence*, *telepresence*, and *usability*. Post-hoc analysis with Wilcoxon signed rank tests were conducted with a Bonferroni correction applied. Moreover, we used Wilcoxon signed-rank tests with the independent variable *scenario* (mobile, stationary) to indicate significant effects on the ordinal data *self reported co-presence*, *perceived others' co-presence*, *social presence*, *telepresence*, and *usability*.

4.1 Quantitative Results

Usability Descriptive statistics led to the following median values for the *SUS* scale sorted by *point of reference*: $Mdn_{\text{body-fixed}} = 61.875$, $Mdn_{\text{head-fixed}} = 63.384$, and $Mdn_{\text{world-fixed}} = 59.375$, by *placement*: $Mdn_{\text{center}} = 65.417$, $Mdn_{\text{periphery}} = 64.375$, and $Mdn_{\text{interactive}} = 58.125$, and by *scenario*: $Mdn_{\text{mobile}} = 60.139$, and $Mdn_{\text{stationary}} = 64.306$.

Independently performed Friedman tests indicate significant differences for the *usability* between the different coordinate systems for *point of reference*, $\chi^2(2) = 6.958$, $p = .031$, and for the *placement* conditions, $\chi^2(2) = 16.085$, $p < 0.001$, see Figure 4.

Post-hoc analyzes performed with a Wilcoxon Signed-Rank test show a significantly higher *usability* for the *head-fixed* compared to the *world-fixed* coordinate system, $Z = -3.291$, $p = 0.001$. However, Wilcoxon Signed-Rank tests did not show significant differences for the *usability* between the differently used coordinate systems, neither for the *body-fixed* compared to the *head-fixed*, $Z = -1.962$, $p = 0.05$, nor for the *body-fixed* compared to the *world-fixed* coordinate system, $Z = -.853$, $p = 0.394$. Moreover, Wilcoxon Signed-Rank tests show a significant higher *usability* when the co-user was placed in the *center*, $Z = -3.378$, $p = 0.001$, and also when the co-user was placed in the *periphery*, $Z = -3.420$, $p = 0.001$ compared to the *interactive placement*, see Figure 4. A Wilcoxon Signed-Rank test shows a significantly higher *usability* for the *stationary* compared to *mobile scenario*, $Z = -2.962$, $p = 0.003$, see Figure 4.

Self Reported Co-Presence Descriptive statistics led to following median values for the *self reported co-presence* sorted by *point of reference*: $Mdn_{\text{body-fixed}} = 18.083$, $Mdn_{\text{head-fixed}} = 18.833$, and $Mdn_{\text{world-fixed}} = 18.583$, by *placement*: $Mdn_{\text{center}} = 18.167$, $Mdn_{\text{periphery}} = 17.917$, and $Mdn_{\text{interactive}} = 19.000$, and by *scenario*: $Mdn_{\text{mobile}} = 18.389$, and $Mdn_{\text{stationary}} = 18.444$.

Individual Friedman tests did neither indicate significant differences for the *self reported co-presence* between the differently used coordinate systems for *point of reference*, $\chi^2(2) = 1.701$, $p = .427$, nor significant differences between the different *placement* conditions, $\chi^2(2) = 5.382$, $p = .068$.

A Wilcoxon Signed-Rank test did not show significant difference for the *self reported co-presence* score between the *scenario* conditions *mobile* and *stationary*, $Z = -.047$, $p = .962$.

Perceived Others' Co-Presence Descriptive statistics led to following median values for the *perceived others' co-presence* sorted by *point of reference*: $Mdn_{\text{body-fixed}} = 39.250$, $Mdn_{\text{head-fixed}} = 40.417$, and $Mdn_{\text{world-fixed}} = 39.583$, by *placement*: $Mdn_{\text{center}} = 39.667$, $Mdn_{\text{periphery}} = 38.917$, and $Mdn_{\text{interactive}} = 40.167$, and by *scenario*: $Mdn_{\text{mobile}} = 38.889$, and $Mdn_{\text{stationary}} = 40.000$.

Individual Friedman tests did neither indicate significant differences for the *perceived others' co-presence* between the different coordinate systems for *point of reference*, $\chi^2(2) = 2.257$, $p = .323$, nor significant differences between the different *placement* conditions, $\chi^2(2) = 3.200$, $p = .202$.

A Wilcoxon Signed-Rank test did not show a significant difference for the *perceived others' co-presence* score between the *mobile* and *stationary scenario*, $Z = 1.111$, $p = .267$.

Social Presence Descriptive statistics led to following median values for the *social presence* sorted by *point of reference*: $Mdn_{\text{body-fixed}} = 9.398$, $Mdn_{\text{head-fixed}} = 9.373$, and $Mdn_{\text{world-fixed}} = 9.469$, by *placement*: $Mdn_{\text{center}} = 9.183$, $Mdn_{\text{periphery}} = 9.660$, and $Mdn_{\text{interactive}} = 9.142$, and by *scenario*: $Mdn_{\text{mobile}} = 9.409$, and $Mdn_{\text{stationary}} = 8.930$.

Individual Friedman tests did neither indicate significant differences for the *social presence* between the different coordinate systems for *point of reference*, $\chi^2(2) = 0.778$, $p = .678$, nor significant differences between the different *placement* conditions, $\chi^2(2) = 5.944$, $p = .051$.

A Wilcoxon Signed-Rank test did not show a significant difference in the *social presence* score for the *scenario* conditions *mobile* and *stationary*, $Z = -2.069$, $p = .039$.

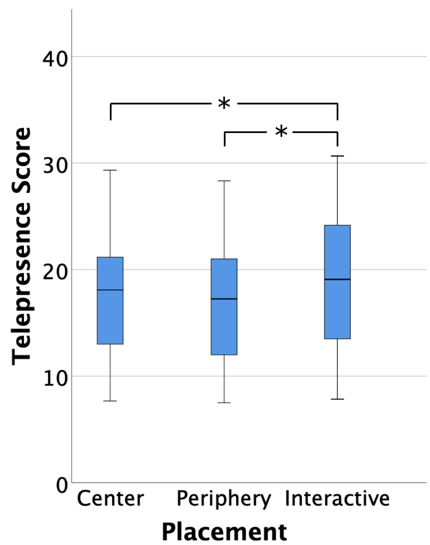

Fig. 5. *Telepresence score for placement condition (center, periphery, interactive)*

Telepresence Descriptive statistics led to following median values for the *telepresence* sorted by *point of reference*: $Mdn_{\text{body-fixed}} = 16.417$, $Mdn_{\text{head-fixed}} = 17.250$, and $Mdn_{\text{world-fixed}} = 19.000$, by *placement*: $Mdn_{\text{center}} = 18.083$, $Mdn_{\text{periphery}} = 17.250$, and $Mdn_{\text{interactive}} = 19.083$, and by *scenario*: $Mdn_{\text{mobile}} = 17.167$, and $Mdn_{\text{stationary}} = 18.889$.

Whilst a Friedman test did not indicate significant differences for *telepresence* between the different coordinate systems for *point of reference*, $\chi^2(2) = 4.324$, $p = .115$, a Friedman test indicated significant differences between the different *placement* conditions, $\chi^2(2) = 17.333$, $p < .001$, see Figure 5.

Post-hoc analysis with Wilcoxon Signed-Rank tests showed a significant higher *telepresence* when the co-user was placed *interactively* by the user, compared to the placement in the *center*, $Z = -2.813$, $p = .005$, and also a significant higher *telepresence* when the co-user was placed *interactively* by the user, compared to the placement in the *periphery*, $Z = -3.050$, $p = .002$. No significant difference could be found for placing the co-user in the *center* compared to the *periphery* condition, $Z = -1.767$, $p = .077$, see Figure 5.

Furthermore, a Mann-Whitney U test did not show significant difference in the *telepresence* score for the *scenario* conditions *mobile* and *stationary*, $Z = 0.523$, $p = .601$.

4.2 Qualitative Results

We analyzed the qualitative data collected during semi-structured interviews through closed coding. The categories were structured according to our independent variables to find explanations for our quantitative analyses' results.

Point of Reference Our quantitative results did indicate a significant favor for the *head-fixed point of reference* of the coordinate system in which the co-user should be visualized. The qualitative results indicate that a majority of our participants (15) prefer to have the co-user always in view. This is a strong argument for the *head-fixed* coordinate system:

– *"The caller was always in my view and I could focus on the object of the conversation"* (P14, *center, mobile*).

Reasons given for the *head-fixed* favor were:

– *"The conversation felt more direct."* (P8, *center, mobile*),
– *"Position right in front of me, I was able to keep eye contact with the caller"* (P17, *center, mobile*),
– *"I was able to change my head position for comfort without risking loosing eye contact"* (P16, *center, stationary*).

Although the majority of the participants favored the *head-fixed point of reference*, advantages for a *body-fixed* coordinate system were also named. A real advantage of the *body-fixed* referenced coordinate system was that (while having the benefits of the *head-fixed* position, such as eye contact, etc.), one could spontaneously focus on other things in the environment:

– *"Center position but fixed through room tracking is a good way to focus on the caller, but it still feels natural to move the head around."* (P18, *center, stationary*)
– *"The direction of the caller was more or less clear and I could control if I want to look at them or not"* (P13, *center, mobile*).

Placement Our quantitative results indicate a significantly higher *usability* if the co-user is placed in the *center* and in the *periphery* than if *interactively* placing the co-user.

The majority (12) of the participants commented positively on the *placement* in the *center* of the FoV and stated that the co-user would be easy to focus, for example:

– *"Was always in view and easy to focus on"* (P7, *body-fixed, stationary*),

- *"The caller was always in my view and I could focus on the object of the conversation"* (P14, *head-fixed, mobile*),
- *"The caller was in the middle of my view.... so I don't have to turn my head or move my eyes"* (P7, *head-fixed, stationary*).

Moreover, participants indicated when they preferred the *placement* in the *periphery*. There were also comments explaining when the *center placement* lacks usability, and the *periphery placement* works better. The *center placement* can, for example, be obstructive and can cause occlusion, especially during *mobile* scenarios and when the coordinate system was *head-fixed*:

- *"The caller was always in the center of my view, so could not see my surroundings as good as I wanted sometimes"* (P2),
- *"Always in view. Can not look away. Obstructing the middle of the FoV. Due to not being always on eye level felt more like looking at a screen instead of a real person."* (P3),
- *"It would take the focus out of my normal vision. This would be bad in some situations (driving a car, shopping...)"* (P8).

For the *interactive placement*, the responses showed that users tended to place the caller in the center position (11), which confirms the preference for the *placement*:

- *"In the center of my view."* (P18, *head-fixed, interactive, mobile*),
- *"Directly in front of me or slightly to the side."* (P7, *world-fixed, interactive, stationary*), or
- *"More or less in front of me but a little bit out of the middle axis.*(P11, *head-fixed, interactive, mobile*).

Also, the *interactive placement* was often put in the center, which was, e.g., done for the following reason:

- *"So I can see their reactions and gestures"* (P2, *head-fixed, stationary*).

The alternative and only slightly less favored placement was the *periphery*, which was confirmed by seven positive comments, such as:

- *"On my right-hand side, like a good friend who walks a few steps with me"* (P15, *head-fixed, interactive, mobile*).

Scenario Our quantitative results show a significant favor for the *stationary scenario*. Participants (7) stated that if the scenario was *mobile*, it was irritating and felt uncomfortable communicating with a *world-fixed* co-user:

- *"The caller was an obstacle in my route, so i had to pass through him/her. Not really comfortable."* (P2, *periphery*),
- *"Sometimes you go through the person, weird feeling"* (P5, *periphery*).
- *"... moving away from the person, while listening to the caller, seemed counter-intuitive"* (P14, *center*).

5 Discussion

We structured and discussed the obtained quantitative and qualitative results according to our independent variables.

5.1 Point of Reference

From a *usability* point of view, the *head-fixed* condition was rated significantly higher by the users. This is in line with our qualitative results that indicate that the participants preferred head-fixed referencing. Participants preferred such referencing as it supports a sense of the remote user's presence by enabling face-to-face conversation and maintaining eye contact. This can serve as an explanation for the *usability* favor of always seeing the co-user, which is also in line with psychological research on non-verbal communication cues, such as [15] as the understanding of the co-user is increased when the face of the co-user [27] or lips are visible [25]. Moreover, participants perceived the *world-fixed* and *body-fixed* referenced co-user equally. In the *stationary* condition, both references seemed nearly identical. We did not measure differences in any presence score, which could be because the co-user was theoretically always and under each condition visible within the FoV. As our walking circle was relatively small, the *mobile* scenario allowed the user to see the *world-fixed* person still when walking "away". Hence, a mobile scenario in which users walk further away and leave the room could show the lack of *presence* when communicating with a *world-fixed* virtual co-user. The qualitative data shows that participants, in general, like the *body-fixed* reference system as well as the *head-fixed* system, but from time to time, the *head-fixed* system can lead to occlusion of real-world content through the co-user's visualization. Especially if the co-user is placed in the center, the *head-fixed* system does not allow for looking away as even when turning the head, the co-user would always be in the center of the view. The *body-fixed* system allows for looking away and focusing on real-world content, which would be especially crucial in multi-task scenarios or when walking through big cities with traffic.

5.2 Placement

Considering the *usability*, the *center* and the *periphery* conditions were rated significantly higher compared to the *interactive placement*. The answers in the semi-structured interviews provide more information about the intends of our participants. The participants considered whether the user's placement directly in the *center* or in the *periphery* was advantageous, depending on the situation. If the focus is on the conversation, the center placement is preferred. Like mentioned before, the placement in the *center* of the FoV corresponds to everyday face-to-face conversations where the conversation partners face each other, and eye contact can be maintained. If the surroundings must be taken into account, peripheral placement is preferred. Not having the co-user in the center of the FoV means the possibility to see the surroundings better and to be able to spend attention on other activities besides the conversation. With regard to *usability*, the interactive placement seems to have proved to be a greater effort. Users had chosen a placement in the *center* or the *periphery* when they placed the co-user interactively. The qualitative data show this. Thus the same *placement* led to an increased effort, which is reflected in the lower usability. Interestingly, the sense of *telepresence* for the *interactive placement* of the co-user was rated significantly highest. While costing cognitive effort and time, one can interpret the data in a way that the free placement of the co-user increases the feeling of being in the virtual environment, as the co-user is positioned in the position that suits the individual needs of users which leads to an increased sense of *telepresence*. We did not measure differences in the other presence scores except the *telepresence*, which could be because the co-user was theoretically always and under each condition visible within the FoV.

5.3 Scenario

Our results did show significantly higher usability for the *stationary scenario*. When the study participants were standing in one place, the covering of the co-user's view was not perceived as disturbing, which had a positive effect on usability. Users could focus on the caller and the talk. When walking in the mobile condition, each condition of the co-user placement resulted in parts of the view being obscured. Walking and simultaneous focusing on the co-user was perceived as disturbing and obstructive. On the one hand, the path to be walked could not be entirely perceived due to the environment's partial occlusion and led to uncertainties when moving. On the other hand, distractions seem to occur because simultaneous visual focusing on the caller and walking is a cognitive load. If one has to be careful when walking, for example, to pay attention to traffic, a *center placement* might be disturbing. Furthermore, if walking further away from the co-user's visualization, results for the *mobile scenarios* might show that the *world-fixed* condition lacks usability.

5.4 Design Recommendations

Scenario: We do not recommend scenario-dependent placements as we address the scenario dependency shown in our results through the recommendation of a dynamically changeable *reference system*.

Reference System: When focusing on the conversation, we recommend using the *head-fixed* referenced coordinate system. Keeping the co-user in view supports the remote user's presence by enabling face-to-face conversation and maintaining eye contact as an important social signal. From time to time, the *head-fixed* system can lead to occlusion of real-world content through the co-user's visualization. Especially if the co-user is placed in the *center*, the *head-fixed* system does not allow for looking away as even when turning the head, the co-user would always be in the *center* of the view. The *body-fixed* system allows for looking away and focusing on real-world content, which would be especially crucial in multitasking scenarios or when walking through big cities with traffic.

If the conversation is not in focus and other tasks require attention, we recommend using the *body-fixed* referenced coordinate system. This referencing leaves the control of content placement to the user through head-turning, for example during a multitask situation when attention sharing and cognitive load matter.

Placement: Placing the co-user in the *center* is often favored as this positioning represents natural face-to-face communication. Such natural communication is supported when placing the co-user in the *center* of the FoV. For tasks that require little attention and where the co-user has an instructing function, we recommend the placement in the *periphery.* Suppose the task has a high cognitive load and requires much attention. In that case, we recommend the *center* or *periphery* placement with the *body-fixed* referencing of the coordinate system. This allows the user to place the co-user through head rotations, representing a natural visual attention switch.

Future work While a face-to-face set up, just like in natural inter-social communication was always given in the head-fixed coordinate system, it was partly missing in the body- and floor-fixed version due to users' movements and rotation. Such difference in communication quality due to partly missing eye contact could be compensated through automatic avatar rotations, especially for the body-fixed coordinate system. If such rotation is desired might be worth investigating in future work.

6 Conclusion

Exploring where visualizations of remote communication partner should be placed in AR, we focused on three research questions:

(1) Where shall we display co-users we talk to in AR relative to us: somewhere in the room, somewhere around us or in our FoV? In other words, where shall the reference of a coordinate system be in which the co-user is visualized: world attached, body attached, or attached to our FoV?

(2) Where within such coordinate system shall the co-user's visualization be placed: in the center, in the periphery, or chosen by the user (through gesture interaction)?

(3) How are placement and coordinate systems favored in different scenarios comparing a mobile and a stationary scenario?

Through a user study, we found that remote users we talk to should be visualized in AR in front of us, but in situations in which the physical world requires attention, visualization in the periphery is better.

In short, for pure communication situations, AR co-users should be placed in the *center* of the FoV using a *head-fixed* referencing. For task situations, AR co-users visualizations should be referenced *body-fixed* to allows the user to place the co-user through head rotations.

Acknowledgements

This work is funded by the German Ministry of Education and Research (BMBF) within the GEVAKUB project (01JKD1701B) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 425869442 and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

References

1. Anjos, R.K.d., Sousa, M., Mendes, D., Medeiros, D., Billinghurst, M., Anslow, C., Jorge, J.: Adventures in hologram space: Exploring the design space of eye-to-eye volumetric telepresence. In: 25th ACM Symposium on Virtual Reality Software and Technology. VRST '19, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3359996.3364244, https://doi.org/10.1145/3359996.3364244
2. Basler, A.: Über das sehen von bewegungen. Archiv für die gesamte Physiologie des Menschen und der Tiere **115**(11), 582–601 (1906). https://doi.org/10.1007/BF01677292, https://doi.org/10.1007/BF01677292
3. Bradley, J.V.: Complete counterbalancing of immediate sequential effects in a latin square design. Journal of the American Statistical Association **53**(282), 525–528 (1958). https://doi.org/10.1080/01621459.1958.10501456, https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1958.10501456
4. Brooke, J.: Sus: A quick and dirty usability scale (1996)
5. Chua, S.H., Perrault, S.T., Matthies, D.J.C., Zhao, S.: Positioning glass: Investigating display positions of monocular optical see-through head-mounted display. In: Proceedings of the Fourth International Symposium on Chinese CHI. pp. 1:1–1:6. ChineseCHI2016, ACM, New York, NY, USA (2016). https://doi.org/10.1145/2948708.2948713, http://doi.acm.org/10.1145/2948708.2948713
6. Du, R., Chuang, M., Chang, W., Hoppe, H., Varshney, A.: Montage4d: Interactive seamless fusion of multiview video textures. In: Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games. pp. 5:1–5:11. I3D '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3190834.3190843, http://doi.acm.org/10.1145/3190834.3190843
7. Finlay, D.: Motion perception in the peripheral visual field. Perception **11**(4), 457–462 (1982)
8. Fiorentino, M., Debernardis, S., Uva, A.E., Monno, G.: Augmented reality text style readability with see-through head-mounted displays in industrial context. Presence: Teleoperators and Virtual Environments **22**(2), 171–190 (2013)
9. Fischer, B.: Blick- Punkte. Neurobiologische Prinzipien des Sehens und der Blicksteuerung. Huber, Bern (1999)
10. Gabbard, J.L., Swan, J.E., Hix, D., Schulman, R.S., Lucas, J., Gupta, D.: An empirical user-based study of text drawing styles and outdoor background textures for augmented reality. In: IEEE Proceedings. VR 2005. Virtual Reality, 2005. pp. 11–18 (March 2005). https://doi.org/10.1109/VR.2005.1492748
11. Gibbs, S.J., Arapis, C., Breiteneder, C.J.: Teleport— towards immersive copresence. Multimedia Syst. **7**(3), 214–221 (May 1999). https://doi.org/10.1007/s005300050123, http://dx.doi.org/10.1007/s005300050123

12. Gruenefeld, U., Löcken, A., Brueck, Y., Boll, S., Heuten, W.: Where to look: Exploring peripheral cues for shifting attention to spatially distributed out-of-view objects. In: Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications. p. 221–228. AutomotiveUI '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3239060.3239080, https://doi.org/10.1145/3239060.3239080
13. Kato, H., Billinghurst, M.: Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In: Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality. p. 85. IWAR '99, IEEE Computer Society, USA (1999)
14. Klose, E.M., Mack, N.A., Hegenberg, J., Schmidt, L.: Text presentation for augmented reality applications in dual-task situations. In: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). pp. 636–644 (March 2019). https://doi.org/10.1109/VR.2019.8797992
15. Krauss, R.M., Chen, Y., Chawla, P.: Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In: Advances in experimental social psychology, vol. 28, pp. 389–450. Elsevier (1996)
16. Lawrence, L., Dey, A., Billinghurst, M.: The effect of video placement in ar conferencing applications. In: Proceedings of the 30th Australian Conference on Computer-Human Interaction. pp. 453–457. OzCHI '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3292147.3292203, https://doi.org/10.1145/3292147.3292203
17. Lindlbauer, D., Aoki, T., Höchtl, A., Uema, Y., Haller, M., Inami, M., Müller, J.: A collaborative see-through display supporting on-demand privacy. In: ACM SIGGRAPH 2014 Emerging Technologies. pp. 1:1–1:1. SIGGRAPH '14, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2614066.2614095, http://doi.acm.org/10.1145/2614066.2614095
18. Lindlbauer, D., Aoki, T., Walter, R., Uema, Y., Höchtl, A., Haller, M., Inami, M., Müller, J.: Tracs: Transparency-control for see-through displays. In: Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology. pp. 657–661. UIST '14, ACM, New York, NY, USA (2014). https://doi.org/10.1145/2642918.2647350, http://doi.acm.org/10.1145/2642918.2647350
19. Nassani, A., Lee, G., Billinghurst, M., Langlotz, T., Lindeman, R.W.: Using visual and spatial cues to represent social contacts in ar. In: SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications. pp. 15:1–15:6. SA '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3132787.3139199, http://doi.acm.org/10.1145/3132787.3139199
20. Nowak, K.L., Biocca, F.: The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. Presence **12**(5), 481–494 (Oct 2003). https://doi.org/10.1162/105474603322761289
21. Orlosky, J., Kiyokawa, K., Takemura, H.: Dynamic text management for see-through wearable and heads-up display systems. In: Proceedings of the 2013 International Conference on Intelligent User Interfaces. pp. 363–370. IUI '13, ACM, New York, NY, USA (2013). https://doi.org/10.1145/2449396.2449443, http://doi.acm.org/10.1145/2449396.2449443
22. Orts-Escolano, S., Rhemann, C., Fanello, S., Chang, W., Kowdle, A., Degtyarev, Y., Kim, D., Davidson, P.L., Khamis, S., Dou, M., Tankovich, V., Loop, C., Cai, Q., Chou, P.A., Mennicken, S., Valentin, J., Pradeep, V., Wang, S.,

Kang, S.B., Kohli, P., Lutchyn, Y., Keskin, C., Izadi, S.: Holoportation: Virtual 3d teleportation in real-time. In: Proceedings of the 29th Annual Symposium on User Interface Software and Technology. pp. 741–754. UIST '16, ACM, New York, NY, USA (2016). https://doi.org/10.1145/2984511.2984517, http://doi.acm.org/10.1145/2984511.2984517
23. Rzayev, R., Woźniak, P.W., Dingler, T., Henze, N.: Reading on smart glasses: The effect of text position, presentation type and walking. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 45:1–45:9. CHI '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3173574.3173619, http://doi.acm.org/10.1145/3173574.3173619
24. Schandry, R.: Biologische Psychologie. Beltz (2011)
25. Schwartz, J.L., Berthommier, F., Savariaux, C.: Seeing to hear better: evidence for early audio-visual interactions in speech identification. Cognition **93**(2), B69 – B78 (2004). https://doi.org/https://doi.org/10.1016/j.cognition.2004.01.006, http://www.sciencedirect.com/science/article/pii/S001002770400054X
26. Strasburger, H., Rentschler, I., Jüttner, M.: Peripheral vision and pattern recognition: A review. Journal of Vision **11**(5), 13–13 (12 2011). https://doi.org/10.1167/11.5.13, https://doi.org/10.1167/11.5.13
27. Sumby, W.H., Pollack, I.: Visual contribution to speech intelligibility in noise. The Journal of the Acoustical Society of America **26**(2), 212–215 (1954). https://doi.org/10.1121/1.1907309
28. Tanaka, K., Kishino, Y., Miyamae, M., Terada, T., Nishio, S.: An information layout method for an optical see-through hmd considering the background. In: 2007 11th IEEE International Symposium on Wearable Computers. pp. 109–110 (Oct 2007). https://doi.org/10.1109/ISWC.2007.4373791
29. Tanaka, K., Kishino, Y., Miyamae, M., Terada, T., Nishio, S.: An information layout method for an optical see-through head mounted display focusing on the viewability. In: Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality. pp. 139–142. ISMAR '08, IEEE Computer Society, Washington, DC, USA (2008). https://doi.org/10.1109/ISMAR.2008.4637340, https://doi.org/10.1109/ISMAR.2008.4637340
30. Velamkayala, E.R., Zambrano, M.V., Li, H.: Effects of hololens in collaboration: A case in navigation tasks. Proceedings of the Human Factors and Ergonomics Society Annual Meeting **61**(1), 2110–2114 (2017). https://doi.org/10.1177/1541931213602009, https://doi.org/10.1177/1541931213602009
31. Zhang, C., Cai, Q., Chou, P., Zhang, Z., Martin-Brualla, R.: Viewport: A distributed, immersive teleconferencing system with infrared dot pattern. IEEE MultiMedia **20**(1), 17–27 (Jan 2013). https://doi.org/10.1109/MMUL.2013.12, http://dx.doi.org/10.1109/MMUL.2013.12
32. Zhang, L., Amin, S.O., Westphal, C.: Vr video conferencing over named data networks. In: Proceedings of the Workshop on Virtual Reality and Augmented Reality Network. pp. 7–12. VR/AR Network '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3097895.3097897, http://doi.acm.org/10.1145/3097895.3097897