VR Invite: a Project-Independent Smartphone App for VR Observation and Interactivity

Jann Philipp Freiwald, Sünje Gollek, Frank Steinicke

VR Invite: a Project-Independent Smartphone App for VR Observation and Interactivity

Jann Philipp Freiwald[1], Sünje Gollek[1], and Frank Steinicke[1]

Universität Hamburg, Germany
{freiwald,6gollek,steinicke}@informatik.uni-hamburg.de

Abstract. Virtual Reality (VR) is a promising immersive technology, which provides users with place and plausibility illusions in a virtual environment (VE). However, current immersive experiences are often limited to those users wearing a VR head-mounted display (HMD). In this paper we present VR Invite, a project-independent smartphone app, which allows multiple non-immersive bystanders to observe and interact with the VE and the HMD users. Our system renders multiple view ports of the scene on a host computer, and transmits the data via wireless local network to the mobile devices. Furthermore, the position and orientation of the smartphones is tracked to change the viewpoints accordingly.
We conducted a user study with 26 participants in the context of rehabilitation for older adults in retirement homes, with a focus on bystander integration. In the study, a VR user had to play multiple rounds of a memory game, while a bystander provided support. We compared VR Invite with a TV-gamepad-combination as interaction medium for the support role regarding sense of presence, social presence, workload and usability, both with purely verbal and active assistance capabilities. The results indicate that the opportunity for direct interaction positively influences the bystander's sense of presence in the VE and the reported usability of the Smartphone app. However, social presence was rated higher in passive conditions in which the real person was the center of attention, as opposed to the avatar on the screen. Furthermore, users valued the comfort of sitting down over active participation and agency with room-scale movement.

Keywords: Virtual Reality, Smartphones, Collaborative Virtual Environments, Multi-User Mixed Reality Interactions, Social VR

1 Introduction

Current virtual reality (VR) setups seldom are a collaborative or social experience for a local group of people. This is due to the requirement of VR being rendered on a single head-mounted display (HMD), with a limited ability to include bystanders. Typically, only passive participation is offered by either mirroring the video feed or rendering the scene from a third person perspective to an external stationary monitor. To provide multi-user integration certain applications support the use of a number of HMDs connected via local area network

(LAN) or internet. However, those are tailored to a specific use case and require substantial amounts of technical equipment and expertise to use in a local environment.

In single HMD setups exploration of the virtual environment (VE) and social exchange with the HMD user are difficult for bystanders, as there is a need for input devices that are directly connected to the rendering computer, which in turn need to be provided in tandem with the VR setup. Such devices need to be tightly integrated and require an explicit implementation of camera control or other forms of interaction. However, the ubiquitous availability of smartphones allows for integrating bystanders into a social VR experience without the need for proprietary input devices with the "bring your own device" metaphor.

In this paper we introduce a project-independent smartphone app called *VR Invite*, which connects to a self-contained package for the Unity engine on a host computer or standalone VR headset via local wireless network. It allows bystanders hold a view port to the VE and observe the VR from any natural angle or position, as depicted in Figure 1. The package additionally supports transfer of touch inputs from the smartphone to the host computer. This enables an easy implementation of bystander interactivity for experiences that go beyond passive or verbal participation. Furthermore, we added visualizations of the additional view ports to increase the social presence of the HMD user. These visualizations provide a sense of spatial relation and participation between all users.

VR Invite can be used to extend existing projects with either active or passive participation capabilities, or build applications specifically tailored for an asymmetrical multi-user experience. To test the technical soundness of this library's first prototype, we conducted a user study in the scope of rehabilitation of older adults in a retirement home. Here, we measured the effect of the ability to freely move a handheld VR view port on both an HMD user and a bystander, primarily regarding sense of presence, social presence and usability.

To summarize, the contributions of this paper are:

- Development of VR Invite, a project-independent smartphone VR viewer app &
- a user study to compare bystander integration techniques for the support role in an asymmetrical rehabilitation scenario.

2 Related Work

Our work builds upon three areas of research: Collaborative Virtual Environments (CVEs), asymmetric mixed reality (XR) collaborations, and incorporation of bystanders into XR setups.

2.1 Collaborative Virtual Environments

CVEs enable collaboration and interaction in a shared VE between users who may either use the same physical work space or remotely connect through the

Fig. 1. The VR Invite app allows a bystander to observe and interact with the VE from any perspective.

internet [13]. In 1993 Carlsson et al. published DIVE, one of the first distributed interactive virtual reality systems [11]. They focused on multi-user interaction in VR environments and networking solutions for synchronized databases which reference virtual objects. CVEs now have a widespread use in VR applications, including rehabilitation, education, training, gaming and artistry. For example, Tsoupikova et al. used VR CVEs for rehabilitation in patients that suffered from a stroke [31]. They implemented motoric exercises for up to four patients with a focus on the social component of therapy. Kallioniemi et al. demonstrated how VR CVEs can be used to a help learning a foreign language [24]. In their CityCompass VR application two users move through a virtual city and alternate between the roles of tourist and guide. To reach a common goal, they have to communicate in a foreign language via headset.

Regarding co-located collaboration, Billinghurst et al. [6, 7] presented systems that let users perform a variety of interactions and visualizations in augmented reality collaborations. Their goal was to allow communication that is closer to face-to-face dialogue than screen-based interactions. Jones et al. proposed RoomAlive [23], following the concepts of Billinghurst et al. with an implementation of projection-based co-located collaboration. They explored the transformation of a living room into an interactive playing environment for multiple users. The diverse use of CVEs is also evident in areas such as security training [28], medicine [12] or project planning in architecture [22].

2.2 Asymmetric Mixed Reality Collaborations

VR Invite allows a number of bystanders to observe and interact with a virtual environment, using smartphones rather than tracked hand controllers. Hence, their form of interaction is asymmetrical to the HMD wearing user. Several prior works have researched asymmetric interactions in virtual or mixed reality [26, 16, 14, 22, 21, 15]. For instance, Oliveira et al. presented a distributed asymmetric CVE for training in industrial scenarios, using screen-based GUIs to guide an HMD wearing user [27]. Their goal was to teach a trainee to operate and repair faulty hardware through remote avatars.

Oda et al. presented an asymmetric case study between a remote user and a local HMD wearing user [26]. They compared the use of traditional 2D interfaces

and VR headsets as medium for the remote user regarding their ability to explain a certain task. The results indicate that demonstrating a task with a VR headset was easier to understand than annotations through 2D interfaces.

Lindley et al. compared gamepad-based input to tracked body movements for asymmetric avatar interactions. They found that natural body movements elicited a higher social interaction when compared to predefined animations [25].

2.3 Bystanders in Mixed Reality

Gugenheimer et al. proposed two approaches regarding the incorporation of bystanders in a VR setting. Their FaceDisplay [18] is an extension for conventional HMDs, equipping the headset with three external touch-sensitive displays and a depth camera. This way bystanders can see the VE from the outside and trigger actions by touching the displays on the head. This of course comes with the problem that an active HMD user moves their head frequently, which makes observing the displays and interacting with them difficult. An exploratory user study showed that the FaceDisplay led to a high degree of dominance and responsibility of the bystander over the HMD user.

With ShareVR [17] Gugenheimer et al. investigated an asymmetric gaming scenario, with one HMD user and one bystander. The bystander is equipped with a Vive Wand controller that has a tethered screen attached, acting as a second view port. The scene is rendered from a second point of view on the VR computer and transmitted via standard HDMI cable. Also, a projector shows the VE from a top-down perspective on the floor. The setup showed an improvement in entertainment value, sense of presence and social interaction compared to the common TV-gamepad combination. The Master of Shapes commercial solution used a similar setup, but replaced the Vive Wand with a standalone Vive Tracker for positional tracking of a display [3].

Finally, Owlchemy Labs showed a prototype solution that uses a Smartphone for positional tracking and display of the view port [4]. As long as the front of the headset is in the smartphone's view, their relative position can be calculated. The view port is then rendered on the VR computer and transmitted via wireless LAN.

3 Implementation and Setup

Similar to the approaches of Gugenheimer et al. and Owlchemy Labs we designed a mobile view port solution for desktop and standalone VR experiences. Like Owlchemy Labs we implemented positional tracking on a smartphone and transmit the data to the VR host computer, which then renders the image and streams it back via wireless LAN. Contrary to the aforementioned solutions, we used Google's ARCore library [2] to define a world anchor point, which could be any kind of picture or easily recognized object. For simplicity's sake we used a QR code printed on a piece of paper, detected by ARCore's "Plane Detection",

"Augmented Images" and "Anchors" algorithms. The anchor can be placed anywhere in the physical room and is used to calculate the relative pose of a smartphone to the virtual origin point. Once the anchor is established by pointing the smartphone camera at it, ARCore builds and constantly refines a model of the physical room. The world anchor position can be defined in the engine's editor or by using the current position of an arbitrary tracked input device like a Vive Wand at runtime. This approach allows the smartphone user to look in any direction while we retain knowledge about the relative position and orientation to the VR scene's origin point, and transitively the VR player's position.

VR Invite was designed to be self-contained package for the Unity engine. When imported into an existing VR solution, the networking server doesn't need to be configured. At runtime, the server waits for smartphone app clients to connect to it, and then assigns them an in-engine camera object. The camera images are compressed to JPGs, serialized and sent via networking protocol to the client. Our current proof-of-concept prototype uses a simple TCP/IP implementation, which segments the image bytes into several chunks that can easily be reassembled on the client due to TCP/IP's guarantee of package order. Image resolution, compression and frame rate are parameters that can be adjusted depending on the number of clients. Figure 2 illustrates the networking scheme used in the current version of VR Invite.

The client within our smartphone app connects directly to a VR host computer by a given IP address and searches for a predefined anchor image. Upon identification of the world anchor point, the app streams its calculated relative pose to the anchor to the connected server. The camera's pose within the VR scene is now defined as the Vector from origin to world anchor position plus the vector from world anchor position to the smartphone, as determined by the app client.

This approach makes the combination of app and package project-independent, as the client only sends positional tracking data and receives a video stream in return, regardless of the scene's content or interactivity. It also allows a theoretically unlimited number of concurrent smartphone viewers, albeit at the cost of linearly increasing computational requirements for the host computer. The app also transmits touch inputs to the server, allowing optional implementation of project dependent interactivity. For example, touch positions can be interpreted as ray casts from the camera to highlight certain objects or to trigger events tied to virtual buttons or other interactive elements. To increase the VR player's social presence, an avatar should be displayed for each connected client. Obvious approaches are either displaying the image rendered for the view port as floating plane, or using a 3D model positioned where the view port is rendered from. We chose to display a 3D model of a smartphone imitating the pose of the real device to relay a sense of spatial relation.

The source code of both the VR Invite Unity package and smartphone app can be found on GitHub [1].

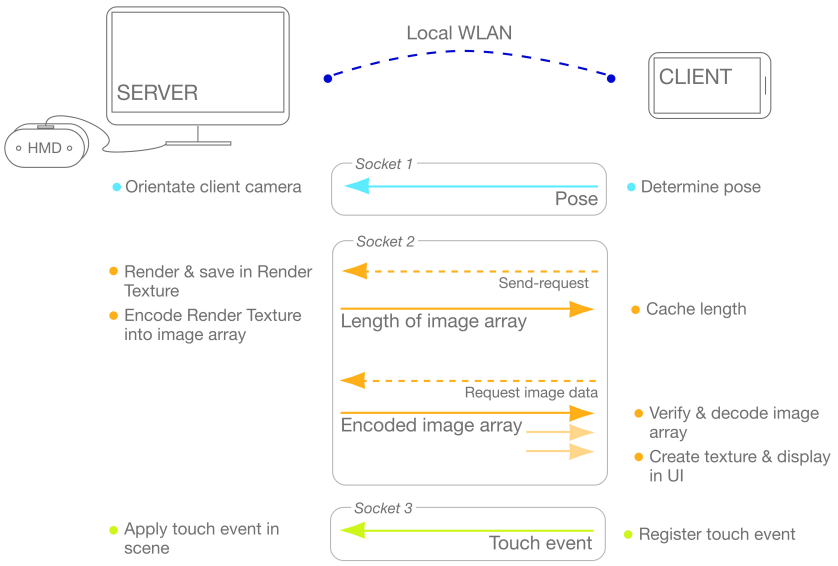

Fig. 2. The networking scheme of VR Invite.

4 User Study

In the context of rehabilitation for older adults in retirement homes, we conducted a user study where a VR player had to play multiple rounds of a memory game, while a bystander supported them. The VR exergame *Memory* of the EXGAVINE project [29] was chosen as a case study. This interdisciplinary project is concerned with the development and evaluation of medically and therapeutically effective VR movement games for the treatment of patients with neurological diseases. While VR Invite can be incorporated into any VR scenario, we focused on use cases which usually have one or multiple bystanders, as opposed to single user or explicit multi user scenarios. The use case of the rehabilitation project is a scenario that not only fulfills this requirement, but could potentially greatly benefit from active bystander integration. Here, bystanders are currently not integrated at all, or only through a passive monitor setup. We expected an increase in sense of co-presence for the older VR users, and thus a positive influence on their enjoyment and engagement with the rehabilitation program. For these reasons, we decided to use this application as a test bed for our case study. Their *Memory* game is meant to be played by older adults with verbal support from family members or nursing staff for training physical and mental capabilities. The term "exergame" is a portmanteau of "exercise" and "gaming" and describes fitness driven game designs. Our goal was to extend the exergame with VR Invite as an uncomplicated and intuitive form of interactivity to increase engagement between players and bystanders. Using the smartphone like

a camera view finder was meant to be a concept that is easy to grasp without any knowledge of interactive video games or VR applications.

The focus of the study therefore was to determine if VR Invite is a suitable general purpose tool to observe and interact with a VR scene. To this end we compared VR Invite to a TV-gamepad-combination as input method for the support role regarding sense of presence, social presence, workload and usability, each with purely verbal or active assistance capabilities. In the chosen exergame the player is in a virtual park and has to solve a memory game with eight pairs of tiles (cf. Figure 3). Selecting a memory tile is done by throwing a virtual ball onto it, which simultaneously trains logical thinking and physical movement. We tested 2×2 combinations of independent variables: *Smartphone* versus *TV*, and *active* versus *passive* interactivity. Thus, each participant had to perform four trials. There was no focus on task performance due to the nature of this exergame's training and gradual self-improvement intent.

Passive TV represents the most common local VR scenario, where one or many bystanders can observe the real HMD user and have a TV screen or monitor showing the viewpoint of the VR user. *Active TV* adds input and a 3rd person perspective to the prior setup. Here, we gave the bystander a gamepad that could control the position and rotation of the TV screens's camera. Pressing any button while a memory tile is under a center crosshair visually highlights the tile in the VE.

The *Passive Smartphone* condition behaves the same as the *Passive TV*. It displays the VE from the HMD user's perspective on the smartphone screen, without the ability to move the camera or interact with the scene. Finally, *Active Smartphone* represents VR Invite. The smartphone can be used as a standalone view port, which can be freely moved. Touching a memory tile on the screen visually highlights it in the VE, as depicted in Figure 3. The avatar of the HMD user always consisted of a simple head with black HMD and a representation of the controllers. Active view ports were represented with a smartphone 3D model in the VE.

For each condition the workload was measured by the NASA Task Load Index (NASA-TLX) [19], the usability by the System Usability Scale (SUS) [10], the presence by the Slater-Usoh-Steed presence questionnaire (SUSP) [30] and the social presence by the Networked Minds Social Presence Inventory (NMSPI) [9, 8]. The NASA-TLX questionnaire consists of five 7-point Likert scales, and is used to subjectively assess physical and mental workload. The SUS likewise is a ten-item attitude Likert scale, which is used to subjectively assess usability of systems and interfaces. The SUSP again uses Likert scales, with 6 items revolving around the sense of being in a VE and the extent to which the VE becomes the subjective dominant reality. Lastly, the NMSPI is a tool to assess the social and co-presence, as well as psycho-behavioral interactions between multiple study participants. We used the NMSPI version 1.2, consisting of 34 7-point Likert scales. In addition to the questionnaires, we measured the distance covered by the bystander in the VE as well as the duration of visual contact during active conditions as an indication of active participation and social interaction. Here,

visual contact was determined by checking if the avatars are within the other participant's camera frustum.

Based on the above described criteria, the experiment was designed as within-subject, and the following hypotheses were formed:

- H_1: Active conditions are rated higher in sense of presence by the bystander.
- H_2: Active conditions are rated higher in social presence by the bystander.
- H_3: Active conditions are rated higher in usability by the bystander.
- H_4: The Smartphone causes a higher workload than the TV for the bystander.
- H_5: There is no difference in sense of presence between all conditions for the HMD user.
- H_6: Active conditions are rated higher in social presence by the HMD user.
- H_7: There is no difference in usability between all conditions for the HMD user.
- H_8: There is no difference in workload between all conditions for the HMD user.
- H_9: The moved distance with active smartphones is greater than with gamepads.
- H_{10}: The visual contact duration with active smartphones is greater than with gamepads.

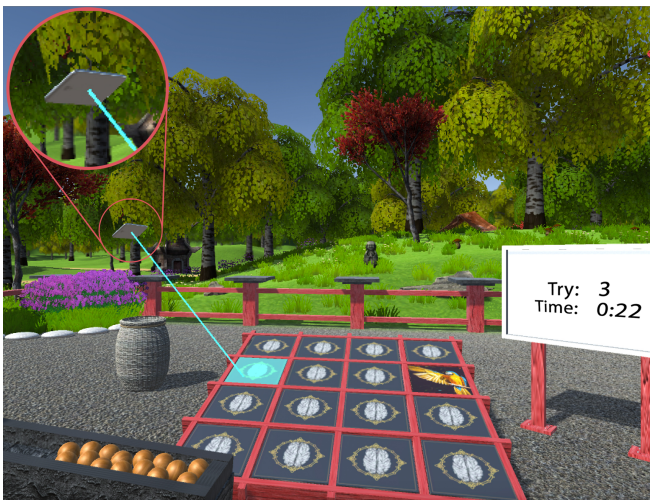

Fig. 3. The virtual environment of the VR memory game. A bystander with a smartphone highlights a tile.

4.1 Participants and Apparatus

26 participants (9 female, 1 diverse) took part in the experiment in pairs of two. In total, the study lasted about 90 minutes. Due to current health and hygiene

regulations, the study had to be performed predominantly with students rather than the intended demographic of older adults, their relatives and nursing staff (cf. section 'Limitations'). The age range was between 19 - 62 years with an average age of 26.81 years ($SD = 10.08$, $M = 24.5$). All participants had prior experience with stereoscopic displays through VR headsets or 3D cinema, while 57,7% had prior experience with studies in VR. On a scale of 0 to 4, participants reported a mean 3D gaming experience of 2.54 ($SD = 1.10$) and their mean gaming time was 11.17 hours ($SD = 15.99$) per week. An HTC Vive Pro with Wand controllers was used for the VR player and depending on the condition a Google Pixel 3XL or XBox One controller for the bystander. A computer equipped with Windows 10, an Intel Core i7-4930K, an NVIDIA GeForce RTX 2080 Ti and 16 GB of RAM was used to render both the VR and the VR Invite view ports.

4.2 Stimuli and Procedure

After giving their informed consent and filling in a demographic questionnaire, participants were introduced to the memory game as well as their assigned device. They were instructed to cooperatively solve the memory challenge. As described above, each trial consisted of a full game of memory, where matching images needed to be turned over consecutively. To select a tile, the VR user had to throw a virtual ball at it. Not hitting two tiles with the same image consecutively resets the last two tiles. The trial ended when all tiles were flipped over. While the VR user selects the tiles, the bystander could assist them purely verbally during passive conditions and additionally by highlighting a single tile during active conditions. The trials were arranged via latin square, each taking circa 5 minutes. Following each trial, both participants filled out the set of questionnaires. Once all conditions were completed, the pairs switched roles and repeated the experiment.

4.3 Results

In this section the results of the statistical analysis are presented. When the Shapiro-Wilk test showed normal distribution of the samples, a repeated-measure ANOVA and post-hoc paired t tests were used to test for differences between conditions. Otherwise, the Friedman test and Wilcoxon Signed Rank test were used. A 5% significance level was assumed, and only significant results are reported.

Presence For the sense of presence the SUSP score and the arithmetic mean of the SUSP results were considered. The wilcoxon test shows higher presence values for *Active TV* (score = 1.2) than the passive conditions, *Passive TV* (score = 0.3, p = .005) and *Passive Smartphone* (score = 0.2, p = .007). This is also evident in the arithmetic mean by paired t tests: *Active TV* ($M = 3.6$, $SD = 1.63$) is rated significantly better than *Passive TV* ($M = 2.8$, $SD = 1.33$,

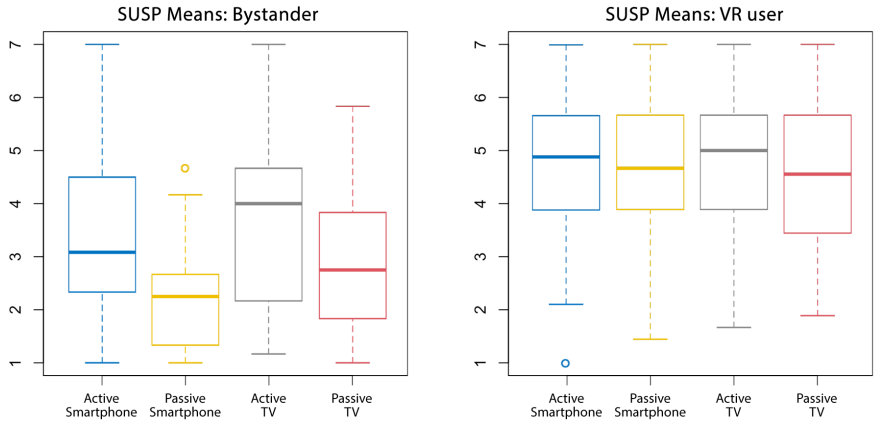

Fig. 4. Mean presence scores for Bystander (left) and VR user (right). Higher is better.

p = .003) and *Passive Smartphone* ($M = 2.3$, $SD = 1.04$, p < .0001). *Active Smartphone* ($M = 3.4$, $SD = 1.59$) is also higher than *Passive TV* (p = .007) and *Passive Smartphone* (p = 0003). In passive conditions, *Passive TV* performs significantly better than *Passive Smartphone* (p = .019). For the VR user role, no significant differences in sense of presence between the conditions could be observed either in the SUSP score or in the arithmetic mean. Figure 4 depicts the mean SUSP scores for both bystander and VR user.

Social Presence We found significant differences in the social presence of the bystander. The result of *Passive TV* ($M = 3.2$, $SD = 0.78$) is significantly higher than that of *Active Smartphone* ($M = 2.8$, $SD = 0.82$, p = .0004) and *Active TV* ($M = 2.7$, $SD = 0.87$, p = .0001). *Passive Smartphone* ($M = 3.4$, $SD = 0.89$) shows a significantly higher result than *Active Smartphone* (p = .0002) and *Active TV* (p = .0001). For the VR user role, *Passive TV* ($M = 3.3$, $SD = 0.90$) is higher than *Active TV* ($M = 2.7$, $SD = 1.02$, p = .002) and *Active Smartphone* ($M = 2.9$, $SD = 0.75$, p = .022). *Passive Smartphone* ($M = 3.2$, $SD = 0.79$) was rated significantly higher than *Active TV* (p = .007). Figure 5 depicts the NMSPI scores for both bystander and VR user.

Visual Contact We observed that some participants deliberately positioned themselves so that they always had the other participant in their view port's frustum, while others focused entirely on the memory tiles. Because of this, the deviations are of considerable size. On average, bystanders looked at the VR users 3.3 times with an *Active Smartphone* ($SD = 3.58$, $min = 0.0$, $max = 14.0$), and 11.3 times ($SD = 7.04$, $min = 1.0$, $max = 28.0$) with an *Active TV*. The average total time looking at the VR user was 11.0 seconds with an *Active Smartphone* ($SD = 13.34$, median=7.5, $min = 0.0$, $max = 59.0$) and 82.7 seconds with the *Active TV* condition ($SD = 77.01$, median=58.5, $min = 2.0$, $max = 367.0$).

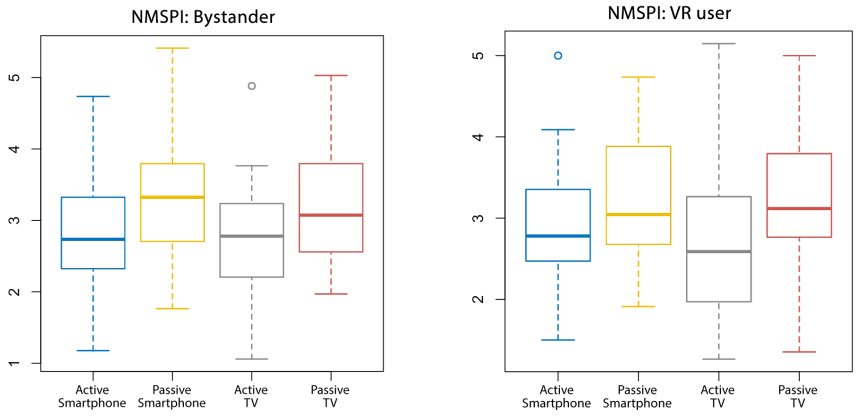

Fig. 5. Social presence scores for Bystander (left) and VR user (right). Higher is better.

Wilcoxon tests showed that both *time* ($p < .0001$) and *number* of eye contacts ($p < .0001$) are significantly higher for *TV* overall than for the *Smartphone*. On the other hand, the VR user looked at the Bystander an average of 11.5 times in the *Active Smartphone* condition ($SD = 10.09$, $min = 2.0$, $max = 39.0$) and 8.0 times when using the *Active TV* ($SD = 6.12$, $min = 0.0$, $max = 23.0$). The average total time looking at the bystander was 65.1 s during the *Active Smartphone* condition($SD = 63.38$, $min = 7.0$, $max = 309.0$) and 45.6 s when using the *Active TV* ($SD = 64.70$, $min = 0.0$, $max = 320.0$).

Usability On the SUS, bystanders rated the usability of *Active TV* ($M = 89.5$, $SD = 11.66$) significantly better than all other conditions: *Passive TV* ($M = 81.3$, $SD = 13.75$, $p = .009$), *Active Smartphone* ($M = 81.0$, $SD = 14.95$, $p = .02$) and *Passive Smartphone* ($M = 68.9$, $SD = 15.05$, $p < .0001$). The usability value in the *Active Smartphone* is significantly higher than in the *Passive Smartphone* ($p = .0002$). The result of *Passive Smartphone* is significantly lower than that of *Passive TV* ($p = .003$). No significant differences in usability were found for the VR user role. Figure 6 illustrates the SUS score distribution for both bystander and VR user.

Workload We found a significantly higher workload for the bystander in the condition *Passive Smartphone* ($M = 29.5$, $SD = 14.50$) than in *Active TV* ($M = 19.2$, $SD = 11.56$, $p = .0006$) and *Passive TV* ($M = 21.4$, $SD = 14.23$, $p = .008$). *Active Smartphone* ($M = 25.9$, $SD = 16.07$) is significantly higher than *Active TV* ($p = .02$). There was no significant difference between the workload results for the VR user. Figure 7 show the NASA-TLX scores for both bystander and VR user.

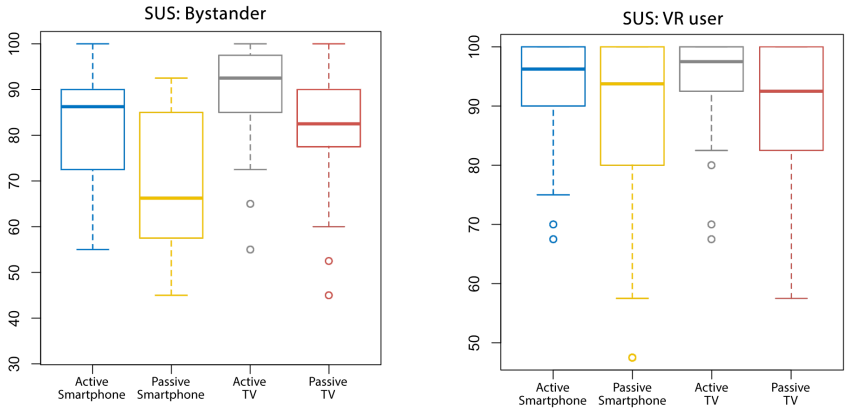

Fig. 6. Usability scores for Bystander (left) and VR user (right). Higher is better.

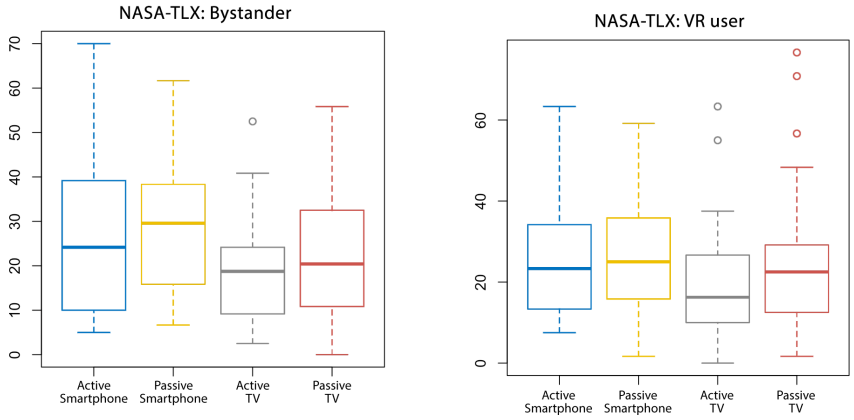

Fig. 7. Workload scores for Bystander (left) and VR user (right). Lower is better.

Bystander travel For the calculation of bystander travel two participant pairs had to be removed from the evaluation. Due short losses of ARCore's tracking, the measurements showed huge momentary spikes in covered distance. All other trials showed no tracking data inconsistencies.

A significantly longer distance was covered in the *Active TV* condition than in the *Active Smartphone* condition ($p < .0001$). The average distance in *Active Smartphone* is 15.9 m ($SD = 12.46$, $max = 64.1$) with a minimum value of 6.0 m. In *Active TV* this value is 92.1 m ($SD = 75.43$, $max = 368.0$) with a minimum value of 8.2 m. Only one participant used the provided tools for movement excessively, traveling 64.1 m for *Active Smartphone* and 368.0 m for*Active TV*. When excluding this outlier, the maximum recorded values become 35.5 m for *Active Smartphone* and 177.1 m for *Active TV*. This however has no impact on the significance of the difference between the conditions.

5 Limitations

As described above, relatives and nursing staff were supposed to use the VR Invite mobile app to interact with elderly participants, who are immersed in a VR rehabilitation training environment. However, due to current health and hygiene regulations, it was only partially possible to perform the study with representative participants and only single bystanders. We therefore had to predominantly rely on university students for both roles of bystander and VR user, which will undoubtedly have skewed our collected data in regards to gaming experience and familiarity with input devices like gamepad controllers. Yet, there is an overlap of the tested and targeted demographic for the bystander role, which would have included family members and nursing staff. Nevertheless, the collected data shows strong significances which cannot solely be attributed to the selection of test subjects. VR Invite itself is a general purpose VR cooperation tool, which can be used in a variety of scenarios. Therefore, we believe that the general trends we found are representative for a wide range of use cases, but not necessarily for the intended elderly VR users of the EXGAVINE project.

The version of VR Invite that was used during the experiment is an in-development prototype and is not yet ready to be released. While the tracking was consistent throughout the study, we received feedback that the frame rate on the smartphone was not as smooth as the participants had liked it to be. We attribute this to the naive serialization and network transportation implementation, which we chose to use for our proof of concept prototype. While the tested solution was not optimized, we deemed it good enough for initial testing and thus conducted the experiment. The performance of the image transfer also had no impact on the rendering for the VR headset, as the transfer is handled in a separate thread. To make VR Invite production ready, the networking solution will have to be exchanged for a more sophisticated image compression and transfer stack. This in turn could possibly have a positive impact on the user ratings for usability and possibly sense of presence as well. We don't however expect the reported data to drastically shift with improved frame rates, as the results are quite unambiguous.

6 Discussion

Sense of presence The results of the study highlight the connection between active participation in a collaborative experience and an increase in the self-reported sense of presence. When looking at the arithmetic mean, both *Active Smartphone* and *Active TV* received favorable ratings when compared to their passive counterparts. As expected, the ability to independently move and interact with the virtual environment is a major factor in feeling involved in a CVE, confirming hypothesis H_1 (Active conditions are rated higher in sense of presence by the bystander). However, the overall average sense of presence was at a relatively low value of 3.025 out of 7. This can be explained by the fact that the bystanders perceived the CE only via screens and always had the real environment in their field of vision. As mentioned in the limitations section, technical

improvements of VR Invite could further shift the bystanders sense of presence in the active conditions favor. We found no significant difference between the conditions for the VR user's sense of presence, which confirms hypothesis H_5 (There is no difference in sense of presence between all conditions for the HMD user). The displayed camera model was perceived equally as present when it was in a fixed location as to it following the bystanders real position. This is most likely due to the fact that bystanders had a tendency to remain standing in the same position for the majority of the trials. They reported an initial wow-effect and tried to move the view port around the tracking space. Once they found a position that was out of the VR user's range from which they could observe the memory tiles, they remained stationary. Because of that, there was no significant difference between the camera model's movement from the VR user's perspective.

Social Presence Contrary to the hypotheses H_2 (Active conditions are rated higher in social presence by the bystander) and H_6 (Active conditions are rated higher in social presence by the HMD user), the sense of social presence was rated significantly higher by both bystander and VR user in the passive conditions. This might seem counter intuitive at first, but can be explained by two factors. One, the bystander's attention is primarily focused on the VR user in the real world. Instead of looking at a smartphone or TV screen, and thus an avatar, bystanders tended to spend a longer time observing the real human being. Two, when not able to directly interact with the virtual environment, bystanders had to go through the VR user as an intermediary by verbally communicating with them. In the active conditions the pointing interaction was used frequently and reported as joyful. However, the pairs spoke noticeably less with each other. This leads us to believe that verbal communication is a more important factor for sense of social presence than visual communication through cues and individual agency. Thus, autonomy in bystander integration appears to negatively correlate with social presence. Prior studies have found that current VR meetings do not reach the same sense of social presence as a real in-person meeting, which can be applied here as well [20].

Visual Contact Regarding visual contact, we found that the *Active Smartphone* was used significantly more often to interact with the memory tiles, while the *Active TV* condition showed more visual contacts with the VR user, disconfirming hypothesis H_{10} (The visual contact duration with active smartphones is greater than with gamepads). The *Active Smartphone* incited the bystander to behave autonomous and solve the memory game on their own, and in turn giving more visual cues to the VR user than the *Active TV*. In a sense, *Active Smartphone* users were more engaged with the experience, but less so on a social level. On the other hand, the VR user looked at the Bystander more often and for longer periods of time during the *Active Smartphone* condition, confirming they are not within their reach to accidentally hit them with the Vive Wands (11.5 times versus 8.0 times and 65.1 seconds versus 45.6 seconds). Thus, VR users were actively aware of the bystander's location.

We also made the following noteworthy observations regarding visual contact and collected demographic data. Male VR users held visual contact with the *Smartphone* significantly longer than female users. This behaviour was inverted during the *TV* conditions. Overall, while the bystanders used a *Smartphone* as opposed to a controller, they looked at the VR user less often, but the VR user looked at them more often. Vice versa for trials where the bystander used a controller. There was a negative correlation between hours played per week and duration of visual contacts during the *Active Smartphone* condition. A positive correlation was found between age and duration of visual contacts during the *Active Smartphone* condition.

Usability Bystanders rated the usability higher in the active conditions. *Active TV* achieved the highest rating, followed by *Active Smartphone*. This confirms hypothesis H_3 (Active conditions are rated higher in usability by the bystander). From participant comments we deduced that control over the view port and the interaction made the game easier and more enjoyable, which aligns with the findings of Gugenheimer et al. [17]. The mean values for *Active TV*, *Active Smartphone* and *Passive TV* are above 80 points, a result close to or exceeding the "Excellent" usability category according to the SUS evaluation guidelines of Bangor et al. [5]. *Passive Smartphone* was rated lowest with 68.9 points, which Bangor et al. classify as borderline between "OK" and "Good" usability on the adjective rating scale.

As mentioned above, the technical shortcomings of our implementation could have had an impact on the usability rating. We don't however expect changes to the networking stack to overcome the gap in ratings that presented itself. Not having to actively move within the real space or holding up a smartphone was rated significantly more usable than the opportunity for natural view port manipulation and freedom of movement. The convenience of sitting down outweighed the increase in precision and naturalness of movement.

We found no significant difference for the reported usability of the VR user between the conditions, confirming hypothesis H_7 (There is no difference in usability between all conditions for the HMD user). The visual cues did neither increase or decrease the usability of the entire setup or the memory game as a whole. Looking at the mean values, the active conditions at over 90 points are slightly higher than the passive ones at an average of 88.55 points. All values indicate a satisfactory usability. In the qualitative feedback, especially the visual cues and the avatars were found to be helpful in cooperating with the bystander. Overall this did however not alter the way the VR user interacts with the VE.

Workload As expected, we also could not find a significant difference for the VR user's workload. While additional visual guidance from the bystanders was perceived as helpful, it also had to be mentally processed and combined with verbal communication and their own memory of the tiles. These two factors effectively canceled each other out, while there was no impact on the physical aspect of the game. This confirms hypothesis H_8 (There is no difference in workload between all conditions for the HMD user). On the Bystanders' side, the workload in the *Smartphone* conditions is reported higher than for the *TV* con-

ditions. This was to be expected; holding up a phone screen is more strenuous than sitting down with a controller. Similarly, actively walking through the real world with a screen in hand less comfortable than sitting down and using the joystick movement. This confirms hypothesis H_4 (The Smartphone causes a higher workload than the TV for the bystander.). Interestingly, bystanders reported a higher workload for holding the *Passive Smartphone* to their face than actively participating with an *Active Smartphone*. It is possible that the smaller screen size compared to the TV resulted in a higher cognitive and physical challenge. However, the average workload values do not exceed 29.5 points in any condition and are thus in a similar range to the values of the VR users.

Travel Distance A significantly longer distance was covered in the *Active TV* condition than with an *Active Smartphone*, with a factor of 5.8 (92.1 m versus 15.9 m). We expected participants to make use of the freedom of movement, which they did not. This disconfirms hypothesis H_9 (The moved distance with active smartphones is greater than with gamepads). Only one participant used the provided tools for movement excessively, traveling 64.1 m with the *Active Smartphone* and 368.0 m with the *Active TV*. Bystanders tried to avoid the movement radius of the VR user, and were content with a position that let them comfortably see the entire memory board on their screen.

We found several correlations between the traveled distance and demographic data. Participants that were not familiar with gamepad controls generally moved less than those experienced with gamepads, and rated VR Invite more positively overall. There was a negative correlation between *Active TV* and age. Older participants moved significantly less. There was a positive correlation between *Active TV* and hours of playtime per week. More hours of playtime led to a significant increase in travel.

In short, there is an influence of 3D gaming experience on virtual scene exploration. There was no such effect for the *Active Smartphone*. This indicates that VR Invite is an intuitive tool that can be used independently of experience with virtual scenes. The familiarity with smartphones in general and the metaphor of the portable window led to a quick adoption of the technique.

Convenience Over Agency After the study participants were asked to indicate which input method they preferred overall. Here, *Active TV* was chosen 18 times and *Active Smartphone* 8 times. Participants had an initial "wow-effect" with freedom of motion through VR Invite, but ultimately valued the comfort of sitting down with a controller in hand over room-scale movement and user agency. Screen size and having to hold up the phone over long periods of time were quoted as reasons for preferring the TV condition. It appears that there has to be a balance between convenience and interactivity, where avoidable movement is seen as a cost. The payoff or the incentive for movement seemingly has to be disproportionately big to outweigh the loss of convenience when there is an alternative form of interaction. To summarize, convenience predominates interactivity and agency if the payoff is not disproportionately big.

7 Conclusion and Future Work

We proposed and evaluated a prototype of a project-independent smartphone viewer app, which enables bystanders to explore and interact with PC or standalone VR applications with unexpected and interesting results. In the utilized memory VR exergame, the bystanders in particular reported a higher sense of presence and higher usability of the active conditions compared to the passive ones. However, the self-reported social presence was significantly lower during active than in passive conditions. Bystanders showed a higher independence and agency when using VR Invite, focusing on the task at hand rather than the VR user. However, agency appeared to negatively correlate with social presence in the explored bystander scenario. This indicates that VR Invite is best suited for implementations with active participation rather than pure observation, where a TV screen was preferred thanks to the convenience it provided. Overall, participants preferred the *TV* over VR Invite for single bystander scenarios, quoting convenience as main driving factor. This demonstrates that while user agency and interactivity are welcome and improve the experience over passive participation, convenience cannot be understated as a central factor of interaction paradigms.

VR Invite was rated to have satisfactory usability, and showed promising results that highlight the potential of individual view ports in use cases with multiple bystanders, where a single TV screen is not sufficient for multi user interactivity. The active component of the *TV* condition can also not be applied to multiple users, as the screen has either to be split or only one bystander can have control. Contrary, impromptu social sessions can make use of VR Invite's ability to quickly join an experience, providing multiple viewing angles. TV and VR Invite can of course be combined, with a neutral perspective on the TV and individual view ports for each bystander for personal agency and interactivity. Regarding agency, Gugenheimer et al. reported that mobile systems could significantly increase the enjoyment of a collaborative game [17]. Their findings indicate that VR Invite would be best suited for deliberately designed multi-user applications, making use of the provided touch interactivity.

Our next goal is the optimization of VR Invite with regards to image serialization and networking stack to make it more pleasant and versatile to use. This in turn will enable us to perform further studies with multiple bystanders to investigate social presence in groups of actively supporting users and one or multiple VR users. Furthermore, we plan to extend VR Invite to support augmented and mixed reality. In XR mode, the host computer or standalone XR device should only render the virtual objects of interest on a transparent background instead of the entire VE. The smartphone passes through its built-in camera feed to its display, and layers the video stream received from the rendering host on top. Implementing and testing these capabilities remains as a target for future work.

References

1. Github: VR Invite Source Code. https://github.com/uhhhci/VRInvite (2021), last accessed: 18.01.2021
2. Google: ARCore Overview. https://developers.google.com/ar/discover (2021), last accessed: 18.01.2021
3. Master of Shapes: Mobile Room Scale. https://masterofshapes.com/work/cover-me/ (2021), last accessed: 18.01.2021
4. Owlchemy Labs: Mobile Spectator. https://owlchemylabs.com/owlchemy-mobile-spectator-ar-spectator-camera/ (2021), last accessed: 18.01.2021
5. Bangor, A., Kortum, P., Miller, J.: Determining what individual sus scores mean: Adding an adjective rating scale. Journal of usability studies **4**(3), 114–123 (2009)
6. Billinghurst, M., Kato, H.: Collaborative augmented reality. Commun. ACM **45**(7), 64–70 (Jul 2002). https://doi.org/10.1145/514236.514265, https://doi.org/10.1145/514236.514265
7. Billinghurst, M., Poupyrev, I., Kato, H., May, R.: Mixing realities in shared space: An augmented reality interface for collaborative computing. In: 2000 IEEE international conference on multimedia and expo. ICME2000. Proceedings. Latest advances in the fast changing world of multimedia (Cat. No. 00TH8532). vol. 3, pp. 1641–1644. IEEE (2000)
8. Biocca, F., Harms, C.: Networked minds social presence inventory:—(scales only, version 1.2) measures of co-presence, social presence, subjective symmetry, and intersubjective symmetry (2003)
9. Biocca, F., Harms, C., Gregg, J.: The networked minds measure of social presence: Pilot test of the factor structure and concurrent validity. In: 4th annual international workshop on presence, Philadelphia, PA. pp. 1–9 (2001)
10. Brooke, J., et al.: Sus-a quick and dirty usability scale. Usability evaluation in industry **189**(194), 4–7 (1996)
11. Carlsson, C., Hagsand, O.: Dive a multi-user virtual reality system. In: Proceedings of IEEE Virtual Reality Annual International Symposium. pp. 394–400. IEEE (1993)
12. Cecil, J., Ramanathan, P., Rahneshin, V., Prakash, A., Pirela-Cruz, M.: Collaborative virtual environments for orthopedic surgery. In: 2013 IEEE international conference on automation science and engineering (CASE). pp. 133–137. IEEE (2013)
13. Churchill, E.F., Snowdon, D.: Collaborative virtual environments: an introductory review of issues and systems. virtual reality **3**(1), 3–15 (1998)
14. Coninx, K., Van Reeth, F., Flerackers, E.: A hybrid 2d/3d user interface for immersive object modeling. In: Proceedings Computer Graphics International. pp. 47–55. IEEE (1997)
15. Dedual, N.J., Oda, O., Feiner, S.K.: Creating hybrid user interfaces with a 2d multi-touch tabletop and a 3d see-through head-worn display. In: 2011 10th IEEE International Symposium on Mixed and Augmented Reality. pp. 231–232. IEEE (2011)
16. Duval, T., Fleury, C.: An asymmetric 2d pointer/3d ray for 3d interaction within collaborative virtual environments. In: Proceedings of the 14th international Conference on 3D Web Technology. pp. 33–41 (2009)
17. Gugenheimer, J., Stemasov, E., Frommel, J., Rukzio, E.: Sharevr: Enabling co-located experiences for virtual reality between hmd and non-hmd users. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 4021–4033 (2017)

18. Gugenheimer, J., Stemasov, E., Sareen, H., Rukzio, E.: Facedisplay: towards asymmetric multi-user interaction for nomadic virtual reality. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2018)
19. Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: Advances in psychology, vol. 52, pp. 139–183. Elsevier (1988)
20. Hodge, E.M., Tabrizi, M., Farwell, M.A., Wuensch, K.L.: Virtual reality classrooms: Strategies for creating a social presence. International Journal of Social Sciences **2**(2), 105–109 (2008)
21. Holm, R., Stauder, E., Wagner, R., Priglinger, M., Volkert, J.: A combined immersive and desktop authoring tool for virtual environments. In: Proceedings IEEE Virtual Reality 2002. pp. 93–100. IEEE (2002)
22. Ibayashi, H., Sugiura, Y., Sakamoto, D., Miyata, N., Tada, M., Okuma, T., Kurata, T., Mochimaru, M., Igarashi, T.: Dollhouse vr: a multi-view, multi-user collaborative design workspace with vr technology. In: SIGGRAPH Asia 2015 Emerging Technologies, pp. 1–2 (2015)
23. Jones, B., Sodhi, R., Murdock, M., Mehra, R., Benko, H., Wilson, A., Ofek, E., MacIntyre, B., Raghuvanshi, N., Shapira, L.: Roomalive: magical experiences enabled by scalable, adaptive projector-camera units. In: Proceedings of the 27th annual ACM symposium on User interface software and technology. pp. 637–644 (2014)
24. Kallioniemi, P., Ronkainen, K., Karhu, J., Sharma, S., Hakulinen, J., Turunen, M.: Citycompass vr-a collaborative virtual language learning environment. In: IFIP Conference on Human-Computer Interaction. pp. 540–543. Springer (2019)
25. Lindley, S.E., Le Couteur, J., Berthouze, N.L.: Stirring up experience through movement in game play: effects on engagement and social behaviour. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 511–514 (2008)
26. Oda, O., Elvezio, C., Sukan, M., Feiner, S., Tversky, B.: Virtual replicas for remote assistance in virtual and augmented reality. In: Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology. pp. 405–415 (2015)
27. Oliveira, J.C., Shen, X., Georganas, N.D.: Collaborative virtual environment for industrial training and e-commerce. IEEE VRTS **288** (2000)
28. Passos, C., Da Silva, M.H., Mol, A.C., Carvalho, P.V.: Design of a collaborative virtual environment for training security agents in big events. Cognition, Technology & Work **19**(2-3), 315–328 (2017)
29. Rings, S., Steinicke, F., Dewitz, B., Büntig, F., Geiger, C.: Exgavine - exergames as novel form of therapy in virtual reality for the treatment of neurological diseases (2019)
30. Slater, M., Usoh, M., Steed, A.: Depth of presence in virtual environments. Presence: Teleoperators & Virtual Environments **3**(2), 130–144 (1994)
31. Tsoupikova, D., Triandafilou, K., Solanki, S., Barry, A., Preuss, F., Kamper, D.: Real-time diagnostic data in multi-user virtual reality post-stroke therapy. In: SIGGRAPH ASIA 2016 VR Showcase, pp. 1–2 (2016)