Interaction of Distant and Local Users in a Collaborative Virtual Environment

Adrian H. Hoppe[1], Roland Reeb[1], Florian van de Camp[2], and Rainer Stiefelhagen[1]

[1] Karlsruhe Institute of Technology (KIT)
{adrian.hoppe,rainer.stiefelhagen}@kit.edu, roland.reeb@student.kit.edu
[2] Fraunhofer IOSB
florian.vandecamp@iosb.fraunhofer.de

Abstract. Virtual Reality enables a new form of collaboration. It allows users to work together in the same virtual room regardless of their actual physical location. However, it is unclear which effect the physical location of the user has on task performance, the feeling of presence or immersion. We compared the collaboration of two users in the same local room and in remote rooms on the basis of a knowledge-transfer task. An instructor indicated different virtual objects using three different pointing gestures and a trainee selected the highlighted object. The results of a 28 participant user study show that the performance of the gestures in the local and remote setup is equal, according to NASA-TLX, rankings and time. Users feel equally co-present and tend to prefer the remote collaboration. The data presented in this paper shows that VR collaboration in a virtual room is independent from the physical location of the participants. This allows the development of VR applications without special consideration of the user's location. VR systems can use the advantages of a remote collaboration, like faster reaction times, no travel expenses and no user collision, or of local collaboration, e.g. direct contact between users.

1 Introduction

Virtual Reality (VR) technology becomes more ubiquitous and a main focus is user collaboration in the Virtual Environment (VE). The collaborative VE [3] allows multiple users to analyze and discuss information as well as interact with the VE and each other. One key advantage of VR with regard to cooperation is that users can meet in an interactive VE without actually being at the same physical location. This eliminates travel time and expenses and saves users the effort to discuss complex information via telephone or video conference by providing a hands-on experience. One disadvantage of a remote interaction can be a delayed interaction due to network latencies. As stated, VR allows remote collaboration but it has yet to be investigated if there are any differences between interacting with other users at the same location (*local*) or at different physical locations (*remote*) (see Fig. 1). Hence, the impact of the user's physical location has to
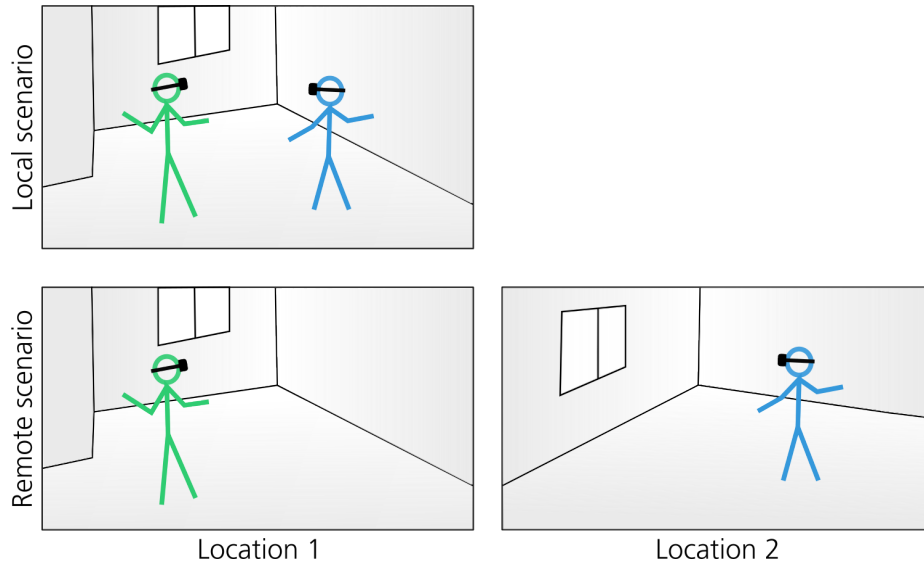
Fig. 1: In the local scenario both users are in the same physical room and the same virtual room. In the remote scenario, each user is in a separate room located in different buildings but both users are in the same virtual room.

be determined. An obvious advantage of a local collaboration is the possibility to physically interact with the other user, e.g. to exchange tools. Furthermore, users can directly communicate by speech without using any additional communication device. However, local collaboration also has some disadvantages. Local users in VR need to be aware of the location of the other user to avoid collisions with them. Individual locomotion of local users poses additional challenges: The physical location of another user and his or her virtual location can be different. In that case, the virtual avatar of the other users cannot be used for collision handling. Also, direct speech communication can be misleading, since the direction of the user's avatar differs from the direction of the user's voice. In conclusion, local and remote collaboration have advantages and disadvantages. This paper addresses the differences between both setups regarding task performance and user experience.

The paper is structured as follows. Section 2 gives an overview of related work. Section 3 presents a setup for local and remote collaboration in VR. The proposed setups are evaluated and compared and the results of a user study are presented in Section 4. Results are further discussed in Section 5, and Section 6 draws a conclusion.

2   Related Work

Salzman et al. [11] developed a cooperative system where two users can work together in VR in a local setup. To achieve this, they used two Head Mounted Displays (HMD). The two users collaborated in an assembly task, where a windshield was inserted into a car. In the real world the windshield is a metal requisite which is tracked and virtually represented as a windshield. Users did not move in this setup, because the windshield was within reach of their start location. Furthermore this work focuses on requisite-based interaction which is not possible in a remote setup. Beck et al. [2] developed a immersive VR system that allows group-to-group telepresence. A group can view the virtual world on a projection screen with shutter glasses. A control device in front of the screen allows group locomotion. The group is captured with a camera and displayed in the virtual world. Two hardware setups allow two groups to interact with each other in VR. Other systems that allow collaboration in VR are Studierstube [13] and the PIT [1]. Kranstedt et al. [7] investigate collaborative pointing. Two users stand on opposite sides of a table, which has different parts of a model plane on it. One user, the *description giver*, sees an exact virtual representation of the table. His or her hand is tracked. The hand is represented in VR with an additional laser pointer, which allows the description giver to point on different locations on the table. The other user, the *object identifier*, is not in VR. He or she sees the pointing gesture of the description giver and identifies the pointed-at object with a pointer. Results of this study do not contain user collaboration since only the geometric pointing behavior was researched.

3   Collaboration in VR with Local and Remote User Locations

In order to compare local and remote collaboration in VR we determined the requirements for a VR system. To allow users to view the VE and interact with it independently, we installed three HTC Vive HMDs in two different locations. For the local scenario, we set up two HTC Vive, connected to two computers, in the same room. Since the HMD cables of the users in the local scenario could lead to users falling and a wireless connection for the HMDs was not available, the cables to the hmd were suspended from the ceiling of the room. The length of the cable was adjusted with a retractable dog leash. However, the cable of the third Vive in the remote room lies on the floor. For the remote scenario, we set up the HTC Vives in two different rooms respectively. The two rooms are located in different buildings but connected by a 1Gbps Ethernet network.

A user is represented by an avatar which is aligned using the head and hand positions through inverse kinematics (see Fig. 2). The avatar is important for the feeling of co-presence [12]. The stylized representation has no significant difference to a human avatar [6] and avoids the uncanny valley [8]. Roth et al. [10] determined that non-realistic avatars handicap social interactions. However missing behavioral characteristics, like gaze or facial expressions, can be partially
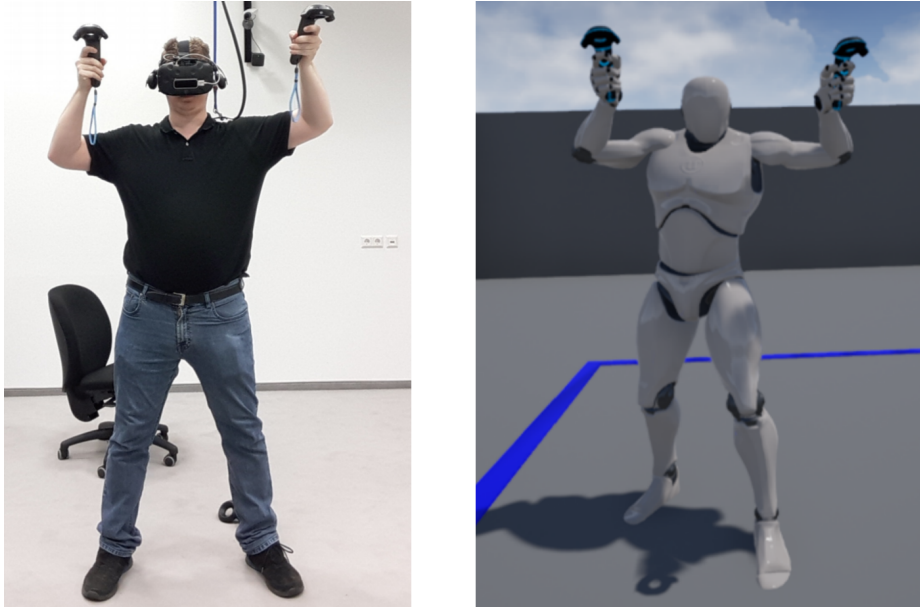
Fig. 2: Avatar representation of the user in VR.

compensated by using other behavior channels, like gestures. They concluded that a mannequin is a universal representation of a human, which is easy to reproduce and animate.

In both scenarios, users used a Logitech G930 headset to communicate via voice chat with each other. The direction of the audio signal of a speaking user is adjusted according to the location of his or her avatar.

To assure user collaboration we implemented a knowledge-transfer scenario where two users take different roles. One user, the *instructor*, highlights virtual objects for another user using a pointing gesture. The second user, the *trainee*, then needs to interact with the indicated object by selecting it using direct touch. To evaluate the effects of the two collaboration setups, three different pointing gestures were examined. The used pointing gestures are virtual hand, virtual pointer [9] and target marker (see Fig. 3). With target marker, the instructor has a virtual laser pointer attached to his or her hand. In addition to that the pointed-at object is highlighted. The trainee also sees the visual highlight but not the beam of the laser pointer, since it is not absolutely necessary. The virtual objects are represented by cubes, arranged in a grid of 3x3x3 as in [14]. To increase task difficulty the grid has a static and a rotating mode. In the rotating mode the whole grid rotates around two axes with different velocities.
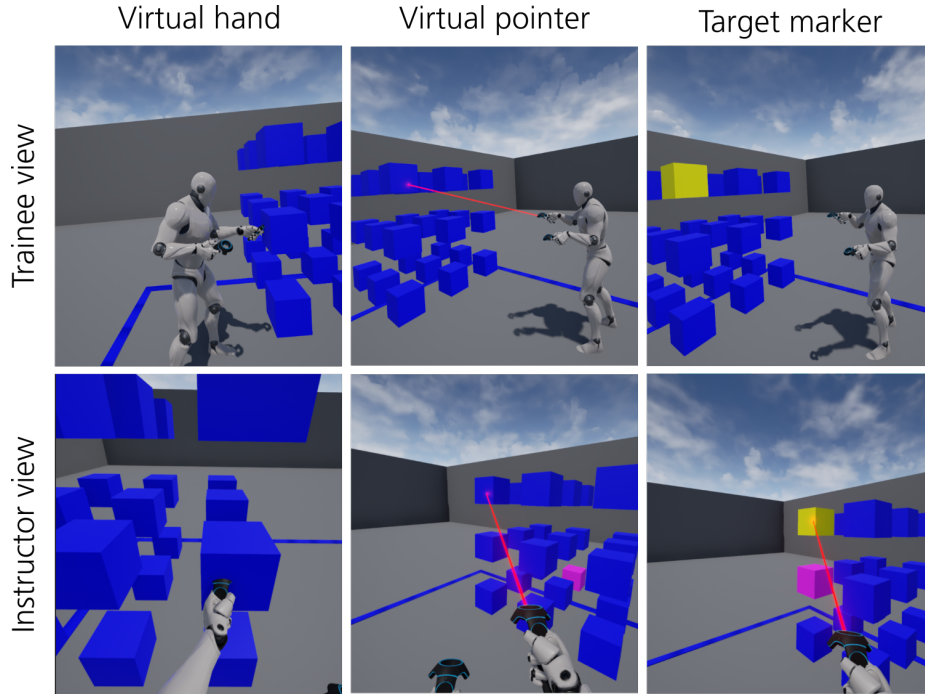
Fig. 3: All three pointing techniques from left to right with the trainee's view on top and the instructor's view on the bottom.

4   Evaluation

We conducted a user study with 30 participants which performed tasks in pairs of two. Due to technical problems one team was excluded, so the evaluation is based on the remaining 28 participants. 19 participants are male and 9 are female. Their average age was 22.5 years. On a five-point Likert scale from 0 (none) to 4 (high) the participants rated their experience with computer games with Ø 3.29 ± 0.81 and with VR with Ø 1.11 ± 1.10.

Each pair performed the tasks in both the remote scenario as well as the local scenario. The roles of the users were switched when the scenario changed. Both users performed all three gestures in the role of the instructor. To minimize the effects of learning and fatigue in the evaluation, the order of the scenarios, roles and gestures was randomized.

A task consists of three training rounds (two with static grid, one with rotating grid) and six timed rounds (three with static grid, three with rotating grid). One round contained one indication of the instructor and the interaction of the trainee with the virtual object. The round starts with both users standing on designated start positions and ends with the selection of the virtual object by the trainee.

To compare the two scenarios, qualitative and quantitative data was collected. Participants were asked how pleasant the collaboration with the partner was. Users rated the collaboration on a scale from 0 (very unpleasant) to 4 (very pleasant) with Ø 3.93 ± 0.26. This value shows that the pairs could work well together and the results are not negatively affected by a user's refusal to cooperate. Furthermore users were asked if they experienced nausea to check if the collected data could be negatively influenced. Users reported almost no nausea with Ø 3.79 ± 0.42 with 0 being strong nausea and 4 no nausea.
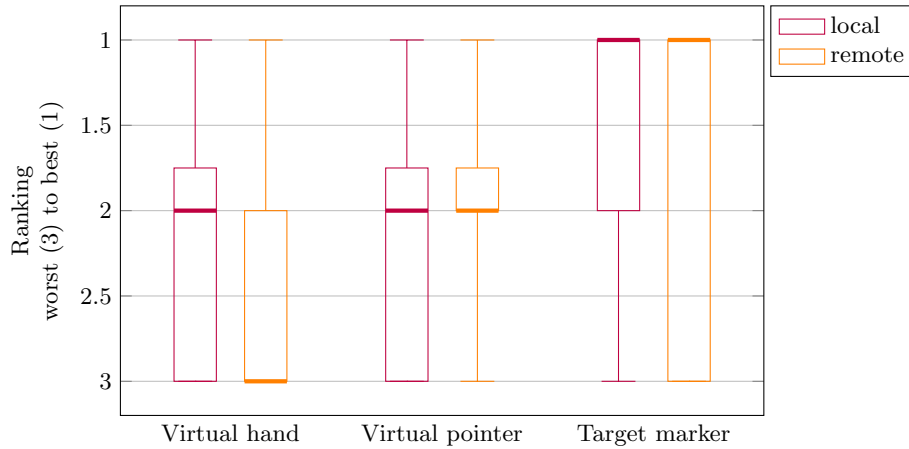

Fig. 4: Ranking of the pointing gestures with box-and-whisker plots performed by the instructor.

Figure 4 shows the users preferences for the different pointing gestures sorted by the room setup. Participants were asked to rank the gestures from first place (1) to last place (3). Since no user performed the gestures in the role of the instructor in both local and remote setup the sample is independent an the Mann-Whitney-U-Test is used to check for significant differences. The median of 1 of target marker shows that this technique did perform well. Pairwise comparisons between the two room setups show that there are no significant differences for any of the pointing gesture ($p = 0.378$ for virtual hand, $p = 0.120$ for virtual pointer, $p = 1.000$ for target marker).

NASA-TLX rankings of the instructor show a low mental, physical and temporal demand with high performance and low effort and frustration levels (see Fig. 5). After applying an ANOVA, the results show no significant differences in a pairwise Dunn-Bonferroni test, except for the ranking of the physical demand of the gestures. The virtual hand interaction is more physically demanding with $p < 0.023$ in a pairwise comparison. The results of the NASA-TLX questionnaire for the trainee are similar to those of the instructor. However, no significant differences between the pointing gestures occur. This is as expected since the in-
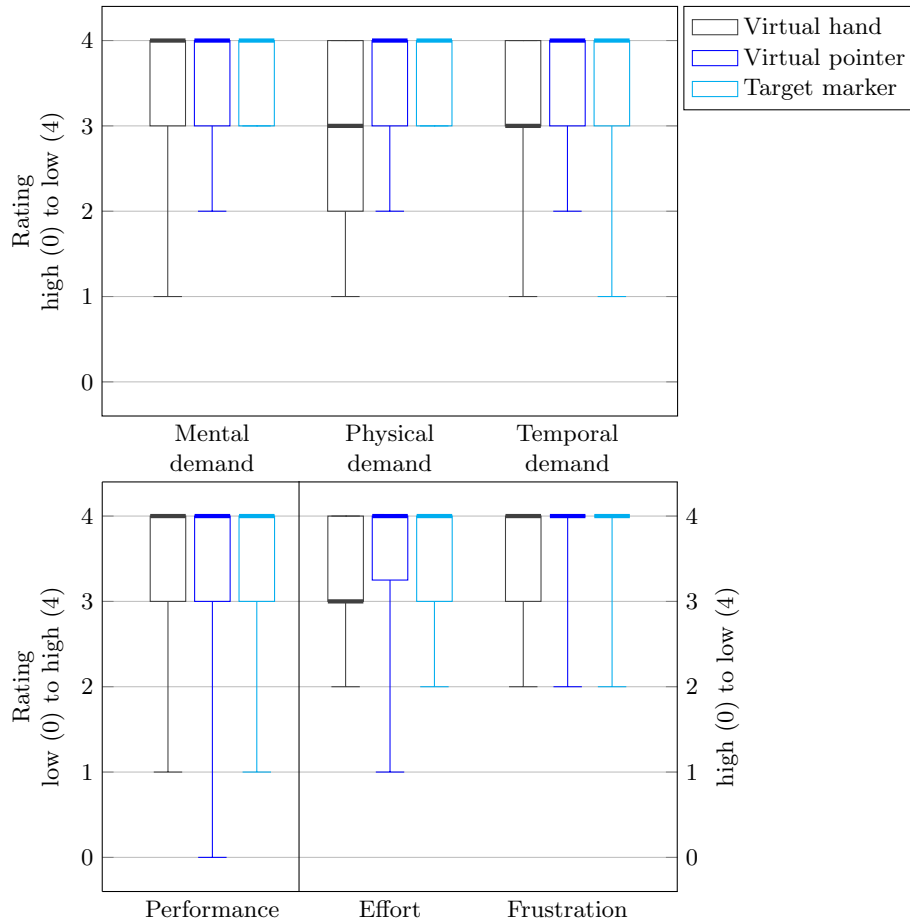
Fig. 5: NASA-TLX with box-and-whisker plots for pointing gestures performed by the instructor.

teraction of the trainee did not change, when the instructor changed to another pointing gesture.

Average interaction times per round are about four to five seconds, as shown in Fig. 6. The differences between the two setups are significant in the case of the virtual pointer ($p = 0.003$) and target marker ($p = 0.012$) according to a sign test. The effect size [5] can be described as medium ($r = 0.321$) and low ($r = 0.274$) respectively. Since the tasks are identical in both setups further investigations were performed. Both the time it took the instructor to point to the target object the first time and the number of correctly and incorrectly indicated virtual cubes are not significantly different. As a result, the speed of the trainee seems to be different in the local and remote setup.
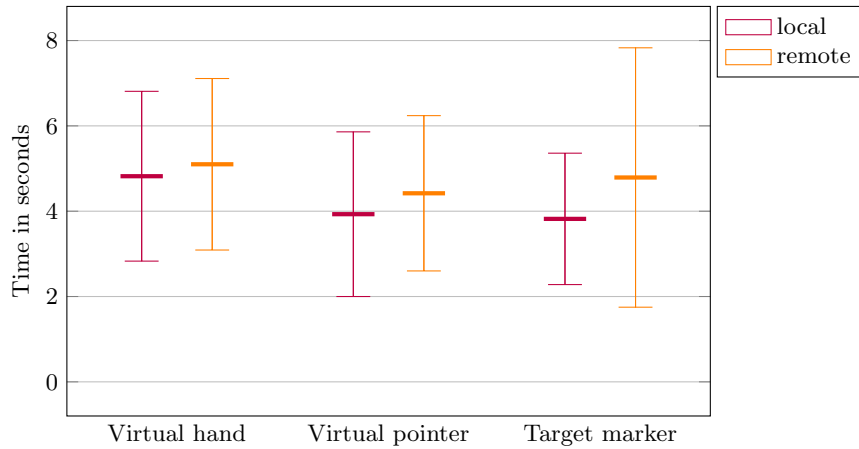
Fig. 6: Timings with average and standard deviation, sorted by local and remote setup.

When asked 50% of the users did not prefer either one setup. 11% preferred the local interaction and 39% liked the remote interaction better. Ten out of eleven users explained their preference for the remote setup by saying that they did not need to worry about any collisions with the other party while working remotely. The other user was impressed by the capabilities of the collaboration via network. From the three users, who preferred the local setup, two said the collaboration is more realistic and one said that he did perceive the other user more as a human rather than a robot.

Users were asked how much they depended on speech communication while solving the tasks on a scale from 0 (not at all) to 4 (very much). A sign test for the ratings of Ø 0.43 ± 0.02 in the local setup and Ø 0.39 ± 0.02 in the remote setup shows no significant difference between the two scenarios ($p = 0.774$).

Furthermore, the users rated the amount of co-presence they experienced with the other user, while performing the tasks of the user study. Co-presence was assessed on a scale from 0 (users feel like they are in different rooms) to 4 (users feel like they are in the same room). In the local setup users rated co-presence as Ø 2.97 ± 0.03 and as Ø 2.82 ± 0.03 in the remote setup. A sign test shows that the difference is not significant ($p = 0.092$).

5 Discussion

The results of the user study show that in general all gestures are suitable for the given task. No gesture outperformed any of the other gestures in all aspects. This result conforms with the conclusion of Bowman et al. [4] that all interaction techniques in VR have their strength and weaknesses and that there is no best technique.

A comparison of the two scenarios, local and remote, shows no significant differences in task performance or user rating except for the interaction time needed. The paired users were less than a second slower in the remote setup. User commentary indicates that the parquet flooring in the remote room was more slippery than the carpet in the main/local room which resulted in users being more careful in their movements. In addition, users dragged the cable behind them in the remote room, since the cable was not suspended from the ceiling. As a result, the speed difference might just be an environmental factor and not a factor of the user's location.

All pointing gestures performed well enough for users to consider the ability to talk to each other as a surplus. Qualitative ratings show that users feel equally co-present regardless of their actual, physical location. However, even the local setup did not achieve full ratings of co-presence from every user which could be explained by the reduced field of view of current VR headsets that limit the environmental awareness in VR compared to the real world or by the fact that the VR HMD and audio headset immerse the user so much that he feels as a part of the virtual world and tunes out reality. The participants show a slight preference for the remote setup since the do not need to worry about collisions. For applications that do not depend on direct contact between humans it might be therefore advantageous for users to collaborate from different locations.

6   Conclusion

VR allows users to collaborate in a virtual environment regardless of their physical location. The presented results of the user study show that an immersive VE enables collaboration regardless of the actual location of the different users. For the executed task almost no significant differences in a VR collaboration between two users in the same and separate physical locations could be found. User's performance, preferences and capability to collaborate seem to be equal in both setups.
The evaluation shows that a basic technical setup can already achieve this effect. The immersion into the virtual world and the feeling of co-presence might be increased even more with an improved and even more realistic avatar, haptics and environment. This opens up novel applications for collaboration in application domains such as education, design, diagnostics, and support. While these results show what is possible with regard to collaboration in VR, a key factor is the speed and latency of the network connection as it will severely influence the quality of the perceived co-presence. Further work is necessary to determine the minimum speed and maximum latency requirements to set a baseline for bringing collaboration in virtual reality to real-world applications. In addition, new techniques need to be developed that allow direct contact between humans in remote setups or collision avoidance mechanisms for local VR collaboration.

References

[1] Kevin Arthur et al. "Designing and Building the PIT: A Head-Tracked Stereo Workspace for Two Users". In: *2nd International Immersive Projection Technology Workshop*. 1998, pp. 11–12.

[2] Stephan Beck et al. "Immersive group-to-group telepresence". In: *IEEE Transactions on Visualization and Computer Graphics* 19.4 (2013), pp. 616–625.

[3] Steve Benford et al. "Collaborative virtual environments". In: *Communications of the ACM* 44.7 (2001), pp. 79–85.

[4] Doug A. Bowman et al. "An Introduction to 3-D User Interface Design". In: *Presence: Teleoper. Virtual Environ.* 10.1 (Feb. 2001), pp. 96–108. ISSN: 1054-7460. DOI: `10.1162/105474601750182342`. URL: `http://dx.doi.org/10.1162/105474601750182342`.

[5] Jacob Cohen. "A power primer." In: *Psychological bulletin* 112.1 (1992), p. 155.

[6] Michael Gerhard, David J Moore, and Dave J Hobbs. "Continuous presence in collaborative virtual environments: Towards a hybrid avatar-agent model for user representation". In: *International Workshop on Intelligent Virtual Agents*. Springer. 2001, pp. 137–155.

[7] Alfred Kranstedt et al. "Measuring and reconstructing pointing in visual contexts". In: *Proceedings of the brandial* (2006), pp. 82–89.

[8] Masahiro Mori, Karl F MacDorman, and Norri Kageki. "The uncanny valley [from the field]". In: *IEEE Robotics & Automation Magazine* 19.2 (2012), pp. 98–100.

[9] Ivan Poupyrev et al. "Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques". In: *Computer graphics forum*. Vol. 17. Wiley Online Library. 1998, pp. 41–52.

[10] D. Roth et al. "Avatar realism and social interaction quality in virtual reality". In: *2016 IEEE Virtual Reality (VR)*. 2016, pp. 277–278. DOI: `10.1109/VR.2016.7504761`.

[11] H. Salzmann, J. Jacobs, and B. Froehlich. "Collaborative Interaction in Co-located Two-user Scenarios". In: *Proceedings of the 15th Joint Virtual Reality Eurographics Conference on Virtual Environments*. JVRC'09. Lyon, France: Eurographics Association, 2009, pp. 85–92. ISBN: 978-3-905674-20-0. DOI: `10.2312/EGVE/JVRC09/085-092`. URL: `http://dx.doi.org/10.2312/EGVE/JVRC09/085-092`.

[12] Anthony Steed and Ralph Schroeder. "Collaboration in Immersive and Non-immersive Virtual Environments". In: *Immersed in Media*. Springer, 2015, pp. 263–282.

[13] Zsolt Szalavári et al. "Studierstube: An environment for collaboration in augmented reality". In: *Virtual Reality* 3.1 (1998), pp. 37–48.

[14] Chadwick A Wingrave, Doug A Bowman, and Naren Ramakrishnan. "Towards preferences in virtual environment interfaces". In: *EGVE*. Vol. 2. 2002, pp. 63–72.