Embodiment in Virtual Environments - Analyzing the Effects of Latency and Avatar Representation

Niklas Bockelmann*
University of Bremen
Bremen, Germany
s_u3t2ba@uni-bremen.de
*contributed equally

Roland Fischer*
University of Bremen
Bremen, Germany
r.fischer@uni-bremen.de

Gabriel Zachmann
University of Bremen
Bremen, Germany

***Abstract*—The way a user is represented in virtual reality is a key element for applications ranging from entertainment to immersive communication and remote collaboration. Having realistic and expressive avatars leads to a stronger sense of embodiment and presence. However, they also lead to increased system latency, which can cause several negative effects, including cybersickness and a reduced sense of embodiment. We conducted a user study to investigate the (interaction) effects of avatar representation/quality and latency on embodiment, task efficiency, and cybersickness in VR. Specifically, we compared a high-quality, personalized point cloud avatar with a lower-quality pre-modeled mesh avatar and latency settings between 150 and 300 ms. We found that the avatar quality had a greater effect on all components of embodiment than latency, and that the perception of the latter was influenced by the avatar representation. High-quality avatars were consistently and significantly rated superior and led to a less severe perception of latency. In contrast, avatar type and latency level had little effect on task efficiency and no notable one on cybersickness. Our work has practical implications for researchers and developers as it shows that having a high-quality avatar in VR is crucial, even at the cost of higher latency, as its benefits outweigh and reduce the negative effects of latency.**

***Index Terms*—Virtual Reality, Embodiment, Latency, Avatar, User Study, Point Cloud**

I. Introduction

Virtual reality (VR) plays an increasingly important role in our everyday lives as such systems allow for immersive exploration of arbitrary virtual environments and provide safe experiences as well as natural interactions. Applications range from video games, telepresence and immersive communication applications [1] to virtual (collaborative) training or simulation environments, i.e., virtual testbeds for space mission simulation [2]. The potential is huge, especially for collaborative and multi-user scenarios. A common and essential component for VR applications is the representation of the user through an avatar. Having highly realistic, high-fidelity, and expressive avatars is crucial regarding important aspects such as the sense of embodiment, and thus, eventually the immersion and presence [3], [4]. However, such avatars, e.g., those based on live-captured point clouds, require more time to process, possibly transmit, and render in real-time, resulting in increased latencies between the user's motion and the avatar's visual response. Low latencies are a critical factor for interactive 3D applications though, especially when experienced in VR. High latencies have been shown to negatively affect the sense of embodiment, task performance, and the overall user experience as well as to lead to increased motion sickness [5]. Ultimately, there must be a trade-off between avatar quality and required computational time/system latency, leading to the question of which aspect is more important and where the sweet spot lies to maximize the sense of embodiment. Previous work by Fribourg et al. [6] suggests that the avatar's appearance may not be the highest priority in terms of embodiment. Interestingly, in a non-VR game by Claypool and Claypool [7], the negative effects of latency have been reduced by altering the avatar's perceived speed, and Unruh et al. [8] found that the level of embodiment induced by different avatars can influence the perception of time. This leads to the question of whether the avatar quality can influence the perception of latency. Halbhuber et al. [9] found no interaction between latency and the perceived avatar fitness in fitness-oriented tasks in VR, but they did not consider avatar quality. It has not yet been investigated, how high latency can be with high-fidelity avatars before the embodiment significantly decreases, or the effects/tolerance of latency across avatars of different quality. Generally, this topic is severely under-researched.

We present the findings of a user study we conducted, in order to answer these questions and to investigate the effects of different avatar representations (of different quality, including the expressiveness) and latencies on embodiment in VR. Specifically, our research questions are how large the effects of avatar quality and latency are on embodiment and task performance, which factor is more important, whether avatar quality affects the perception and tolerance of latency, and if there is a threshold before negative effects become significant. For this, we designed a within-subject user study with 2 different avatar options, 3 latency levels, and 3 tasks with varying degrees of required motion. We examined the embodiment with its three sub-components (body ownership, agency, and self-location [10]), the task efficiency, and the level of cybersickness. With the results of this study, we provide valuable insights into this sparsely researched topic.

II. Related Work

Previous work showed that being represented by an avatar, among others, increases the sense of embodiment and presence [11], [12]. Also, the degree of (photo-)realism [13], [14] and personalization [15] of the avatars are important factors regarding the sense of embodiment and presence. Weidner et al. recently provided a comprehensive overview [4]. Salagean et al. [3] found that the combination of both factors leads to the best results. However, Fribourg et al. [6] discovered that users tend to prioritize factors such as the point-of-view and the degree of control over the avatar's appearance to maximize the sense of embodiment. Moreover, how a user is represented by an avatar in the virtual environment has a decisive impact on the user's general experience and behavior (Proteus effect), i.e., behaving more confident when embodying a more attractive avatar or acting less agile when embodying an avatar perceived as less fit [9], [16].

Latency and its negative effects is a well-researched topic. Regarding VR applications, high end-to-end latencies decrease the sense of body ownership [5], [17], agency [18] and presence [19]. Concretely, the levels of latency where negative effects were measured were 103 ms, 125-300 ms, and 90 ms, respectively. Other studies found that latency can cause or increase cybersickness, although the degree of required latency differed greatly between 63 ms [5], and not even at 350 ms [18]. Also, it was shown that (task) performance [20] and user experience [21] are negatively affected by latency, too. For instance, Hoyet et al. [22] reported a loss in user precision with latencies of 80 ms, Caserman et al. [5] observed significantly higher task completion times with latencies of 69 ms, and Toothman and Neff [23] recorded decreased user experiences with latencies of 300 ms and higher.

Previous works found that the perception of the latency is not only affected by its actual extent but also the activity that is performed in the virtual environment, i.e., Hoyet et al. [22] discovered a relationship between the movement speed of an object to be tracked with the hand and the perception of latency. Additionally, the negative effects of latency on the user can be reduced if he is not actively perceiving it [24]. Claypool and Claypool [7] discovered (in a non-VR game) that the adverse effects of latency can be reduced by altering the avatar to be perceived to be slower (and thus, the delayed responses were more in line with the expectations based on the avatar's appearance). Similarly, Halbhuber et al. [9] investigated if the avatar's appearance influences the perception or sensitivity of latency (in physically demanding tasks). In their case, avatar appearance refers to avatars being perceived as more or less fit, not the fidelity or visual quality. They found that the fitter avatar induced higher levels of embodiment and embodying it users had better task performance. They argued that this is caused by a higher resemblance to the actual participants and the Proteus effect. No interaction between the avatar's appearance and latency was found. In contrast, Unruh et al. [8] recently investigated the influence of embodiment on time perception in VR and did discover a relation. They used a full-body avatar, one consisting only of hands, and no avatar at all in front of a mirror with the task of estimating the delay when switching on a light. While there was no significant difference in delay estimation, they found that, independently of the activity level, the low-embodiment avatar (no avatar at all) led to the general perception of time passing slower than with the other two avatar variants with higher embodiment.

III. Implementation

To investigate the influence of varying latencies and avatar representations/qualities on embodiment in VR, we implemented a test environment in which participants can get represented with mesh-based or point-cloud-based avatars (see Fig. 2), and that allows us to set arbitrary motion-to-photon latencies, delaying the visual reaction of the avatars. We used the Unreal Engine 4 for our implementation. To add the chosen latency, we buffer the (sensor) input and apply the data with a corresponding delay. Additionally, we implemented three tasks that required various degrees of motion (see Fig. 1), a virtual questionnaire depicting the questions, one at a time, on a wall in front of the participant, and extensive logging functionality.

As representatives for the lower-quality avatar representation, we modeled a male and a female generic mesh avatar, including a skeleton for animation, similar to [15]. We refrained from using a higher-fidelity mesh as this avatar is supposed to be of lower (visual) quality. We also had to create slightly modified versions of the mesh avatars for which we removed parts of the head/face (such as the tongue and teeth). The reason for this is that in one of the tasks, participants often moved forwards and backward, and, with higher latencies, it occurred that the face could be seen from behind (e.g., eyes, teeth, etc.), which was highly uncomfortable. For the animation, we used 5-point tracking based on the HMD, the controllers, additional trackers at the ankles, as well as inverse kinematics. For the latter, we employed Unreal-Engine's animation blueprint-based workflow. However, we needed to tune and optimize the behavior to make it convincing. Particularly problematic were the knees and arms, which initially often bent in the wrong direction. We had another issue with odd-looking forearms when the participants twisted their wrists. Our solution was to add wristbands to the avatar to hide the stretched/twisted mesh parts. Also, to get an accurate rotation of the participant's body, we computed it not based on the HMD's rotation but the average rotation of the two feet/ankles. We added a dynamic scaling parameter to adapt the avatar's size to the individual participant's height. Still, the virtual hands were not accurately aligned with the real hand's positions, and elbows sometimes started to bend too early/late. The causes seemed to be the controllers' actual sensor position and the varying arm-length of the participants which scaled not perfectly with the height (see ape index). Thus, we implemented a t-pose test to measure each participant's arm length based on the position of the two controllers and added offsets to the controllers' positions at run-time (elongating the avatar's arms instead was reported to look unnatural). Eventually, we achieved a reasonably good level of tracking accuracy while keeping the desired distance in tracking/control fidelity to the higher-quality avatar.

To live-capture and create the point cloud avatars (as a high-quality avatar representative), we used a single RGB-D camera and the point cloud streaming and rendering pipeline

Fig. 1. Depictions of our 3 tasks: grabbing and placing spheres (left), popping bubbles (center-left), and imitating movements in front of the mirror (right). The model showing the movements is depicted in the second to right image.

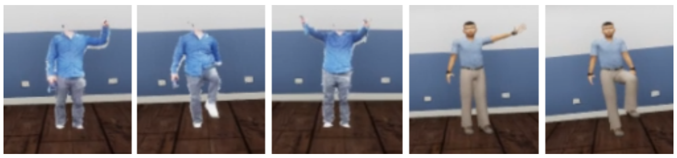

Fig. 2. The two avatar representations of our study as seen by the participants in the mirror during task 3. The left 3 images depict the point cloud avatar which more accurrately depicts the actual appearance and motion of the user and the right 2 the mesh avatar with less visual similarity and lower visual/motion quality.

proposed by Fischer et al. [25] with the registration method by Muehlenbrock et al. [26]. The benefit is that those methods are already implemented in Unreal Engine 4, focus on low latencies, and handle all steps from capturing the depth camera frames, pre-processing, and registration, to real-time rendering. The latter is done using splatting and Unreal Engine's particle system. To not obstruct the participant's view, the points representing the head/HMD were excluded from rendering. For more information, we refer to the respective publications.

IV. User Study

A. Hypotheses

This study aims to investigate the influence of avatar representation/quality and latency on embodiment in VR. Importantly, to get deeper insights, the three sub-factors that, according to previous research [10], build the sense of embodiment (body ownership, agency, self-location) were considered separately. Based on prior work about avatars, embodiment and its sub-factors, latency in virtual reality, and our own considerations, we defined the following six hypotheses to answer our research questions (see last paragraph of Sec. I).

High-quality avatars significantly increase body ownership [15] and embodiment in general [4]. However, according to Fribourg et al. [6], a high-quality avatar representation is not the strongest factor. Many studies found that latency reduces body ownership [5], [17] and agency [18], [27] significantly. Based on this and the common understanding of the three factors and their relationships, we assume the latency to be the more dominant factor for at least agency and self-location and formulate:

- H_1: The qualitative decrease in avatar representation reduces the body ownership more than the latency increase.
- H_2: The increase in latency reduces the agency more than the qualitative decrease in avatar representation.
- H_3: The increase in latency reduces the self-location more than the qualitative decrease in avatar representation.

Similarly, previous studies showed that increased latency [5], [18], [22] or lower-quality avatars [4] could significantly reduce task performance. Which factor is more dominant is not clear yet. However, we assume that:

- H_4: The increase in latency has a greater impact on task efficiency than the qualitative decrease in avatar representation.

Although a highly relevant topic, we are unaware of any previous studies that investigated or found any effects of the avatar's quality on the perception of latency in VR. However, previous work found that there are other factors influencing the perception of latency apart from its extent, see last paragraph of Section II. Hence, we assume that:

- H_5: The avatar representation affects the users' perception of latency.

Previous work found that high latencies and longer VR sessions can cause cybersickness [5], [28]. As our study includes high latencies and does take over 30 minutes, we formulate H_6:

- H_6: Over time, the experiment causes increasingly severe cybersickness symptoms.

B. Experimental Design

For the user study, a within-subject design was employed, such that each participant eventually experienced all avatar and latency combinations. To avoid ordering effects, the order was randomized. We decided on having and comparing two different full-body avatar representations, and three different latency levels, resulting in overall $2 \times 3 = 6$ conditions. As a more realistic (visual fidelity, degree of control/tracking, resemblance to user's appearance and motion), high-quality avatar that also usually takes more processing power and time (especially when aiming for maximal fidelity), a live-captured point cloud one was chosen. For instance, Yu et al. [29] found these superior to pre-modeled, tracked mesh avatars regarding co-presence, social presence, behavioral impression, and humanness. In comparison, a pre-modeled, 5-point tracked mesh avatar was employed as a less realistic (lower visual and motion fidelity, less visual similarity) but quick-to-compute option that is also commonly used. In the study, both avatar types were tested with the same 3 motion-to-photon latency levels for a fair comparison, even though the employed type of mesh avatar would normally be processed and rendered practically instantaneous. The lowest latency level, 150 ms, was selected as it is the minimum achievable latency with the point cloud avatar in our implementation. It was measured by simultaneously capturing the participant's and his avatar's motion on the screen using an external camera with 60 Hz. As previous studies reported a significant loss in embodiment and control with latencies of 300 ms and above [18], [23], [30], 300 ms was decided on as the highest latency setting and a third one of 225 ms was added in the middle of the range.

The perception of latency strongly depends on the task, according to Toothman et al. [23]. To comprehensively investigate the effects of the avatar representation and latency, we carefully designed 3 tasks that demand varying degrees of motion (and speed) and are similar to or inspired by already proven to be effective tasks in previous latency and avatar-focused studies [5], [15], [18], [23]. The three tasks are depicted in Fig 1. The first one consists of repeatedly grabbing the correct sphere out of eight differently colored ones that circularly move around in front of the participant and then placing it accurately in a designated spot in the center. This task does not require taking any steps but only moving the arms. The second task involves popping bubbles that rise before the participant by touching them. The bubbles appear quickly and randomly distributed inside a specified area before the participant. They rise randomly in wavy patterns; sometimes, small steps must be made to reach all of them comfortably. Throughout the study all participants experienced the same bubble movements (same seed). The goal is to pop as many as possible (of the 40) before they disappear again naturally (at a specific height). Since according to Schoenenberg et al. [31], some form of urgency is an important factor for effective latency evaluations, the task was designed in this way to create a sense of pressure. The third task consists of observing another avatar's motion (5 seconds) and then imitating it in front of a mirror (8 seconds), so that the participant sees his own motion. Six specific movements (see Fig. 2) were implemented: waving the hand (left/right), raising the leg up and down, stretching both arms forward and making big circles, stretching one arm sideways and making circles (left/right). Originally, the tasks were designed to take longer, but as the long time of the study was one main concern of a pre-study we conducted, the tasks were streamlined. Other results of the pre-study were also considered, such as hiding the other avatar of task 3 after observing its motion so the user can focus on his reflection in the mirror when replicating the movement.

C. Setup

For the study, a 3D office scene was created using the Unreal Engine 4. Although having high-quality rendering and lighting, the scene itself was kept rather simple and clutter-free to avoid distractions. To ensure smooth framerates of 90 fps and minimal default latencies, the HTC Vive Pro Eye HMD and a high-end PC with a Nvidia RTX 4090 and an AMD Ryzen 9 3900X CPU were used. To capture and create the point cloud avatar, the Azure Kinect depth camera (640×576@30 Hz) was employed. For the tracked mesh avatar, non-personalized models (male/female) and a 5-point-tracking approach in combination with inverse kinematics for the animation were chosen. Specifically, the HMD itself was used to track the head, the two controllers to track the hands as well as 2 additional Vive trackers strapped to the participant's ankles (the 3 most important factors for inverse kinematics according to [32]). See Sec. III for more details. Marker-based tracking systems such as OptiTrack were not used to keep the setup simple and prevent longer and more error-prone switching phases in between the rounds with different avatars.

D. Measures

Questionnaires were the main measure of the study, since embodiment is mostly subjective: A demographic questionnaire asked about age, height, gender, VR experience, and sensitivity regarding cybersickness. To evaluate the sense of embodiment and its individual components, a 10-question embodiment questionnaire was employed that we composed out of the "virtual embodiment questionnaire´´ by Roth and Latoschik [27] (body ownership, agency; q1-q8) and the "preliminary embodiment short questionnaire´´ by Eubanks et al. [12] (self-location; q9-q10). The reason for this combination is that Roth's well-tried questionnaire is missing questions about self-location and the more recent questionnaire by Eubanks included those (but altered the other questions in a way less suitable for our case). Answers to the questionnaire were given using a 7-point Likert scale. To quickly get the current level of cybersickness between rounds, the one-item "fast motion sickness scale´´ by Keshavarz and Hecht [33] was employed. Awnsers were given using a 7-point scale instead of the regular 20-point one for consistency between the questionnaires. Additionally to the questionnaires, the task performance and completion time for the relevant tasks were measured. Specifically, the time and precision (euclidean distance) for task 1 and the performance (number of popped bubbles) in task 2 were measured. For task 3, we found it not appropriate to measure the time or try to measure the accuracy of the imitation.

E. Procedure

The study procedure is depicted in Figure 3. First, the participants were informed about the study and its purpose and had to consent (including the anonymous data collection). The experiment strictly followed the university's ethical guidelines. After that, they had to fill out the demographics questionnaire, put on the HMD (including the additional trackers on the ankles, which stayed attached for the whole study), and make a t-pose for the calibration procedure for the mesh avatar. Then, followed time to familiarize themselves with the VR environment and, eventually, the participants had to complete all six rounds of the study. Each round included all 3 tasks and

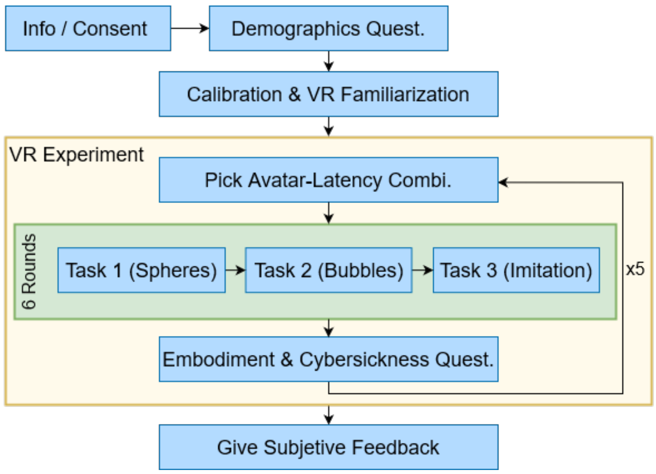

Fig. 3. Flow diagram of the procedure of the user study. Over 6 rounds, each participant experienced all avatar-latency combinations in randomized order.

one latency-avatar combination. The order of the rounds was randomized to avoid ordering effecs. The participants were allowed to decide when to start each task within a round, and after each round, they had to complete the cybersickness and embodiment questionnaires directly in VR. For these, the participants were presented with the questions, one at a time, on a virtual wall and were asked to answer verbally. We decided on in-VR questionnaires, based on prior research and our pre-study results that found it to be less distracting and time-consuming, hence, generally preferable. Finally, after all rounds, the participants were given time to give subjective feedback. The study took 35 minutes on average.

V. Results

In this section, we present the results of our questionnaires and the measured task performance and completion times. First, we tested the data for normality using the Shapiro-Wilk test. As the test was negative for most data, we assumed not to have normal distributions. Additionally, our study consisted of multiple factors, and, in the case of the questionnaires, ordinal data, hence, we decided to employ ART-ANOVA for significance and interaction testing. If the test reported significant differences, we employed pairwise Bonferroni-corrected Wilcoxon signed-rank post-hoc tests. We assumed a significance threshold of $p = 0.05$, as usually done. In addition, to provide an easy-to-understand overview of the data, we also calculated the means and standard deviations for all data (we found the medians to be less meaningful with a 7-point Likert scale). In the interest of readability, we only report the most relevant data directly. However, additional plots, and all descriptive data can be found in the supplementary material. In the following, we abbreviate the conditions with combinations of M/P for mesh/point cloud, and 1/2/3 for 150/225/300 ms of latency.

A. Participants

For our study, we recruited $n = 33$ participants of which 85 % were male and 15 % female. The age range was between 21 and 45 with an average of 27.3 years. Moreover, 33.3 % of the participants reported having never experienced VR before, 30.3 % experienced it once or twice, 21.2 % stated to use it from time to time, and 15.2 % even regularly.

B. Embodiment

We depict the results of the individual factors of embodiment in Fig 4. The overall embodiment can be calculated by averaging the three factors (for corresponding data and plots, we refer to the supplementary material). We can see that the results for the three factors are generally similar: The point cloud avatar is rated (most often significantly) superior to the corresponding mesh avatar for all latency levels, and with increased latency, the ratings decrease for both avatars. Moreover, the decrease mostly follows a roughly linear trend over the observed latency range. However, with the mesh avatar, the relative overall score loss is consistently higher than with the point cloud avatar. We found differences especially often to be significant for the mesh avatar and when comparing the first and third latency levels. The latency matrix can be seen in Fig. 5. Interestingly, the difference is the highest for the body ownership with -20.2/-4.4 % (the point cloud avatar's score loss is especially small). For agency, we measured a relative overall score loss of -24.2/-13.5 % (especially high for the mesh) and for self-location -16.8/-13.5 %). All means and standard deviations are listed in Tab. I. Lastly, although individually significant, we did not find any significant interactions between the two factors of latency and avatar representation.

TABLE I
AVERAGE AND STANDARD DEVIATIONS REGARDING EMBODIMENT, TASK EFFICIENCY AND LATENCY PERCEPTION.

Factor / Metric	Condition					
	M1	**P1**	**M2**	**P2**	**M3**	**P3**
Bod. Own. / μ	4.46	5.51	4.18	5.30	3.56	5.27
Bod. Own. / σ	1.61	1.37	1.68	1.42	1.55	1.27
Agency / μ	5.50	6.14	4.78	6.03	4.17	5.31
Agency / σ	1.36	1.08	1.87	1.11	1.93	1.52
Self Loc. / μ	4.71	5.83	4.37	5.43	3.92	5.04
Self Loc. / σ	1.62	1.18	1.63	1.36	1.76	1.63
Time1 / μ	30.79	29.07	32.91	32.14	38.19	32.30
Time1 / σ	10.52	9.74	11.56	10.83	14.63	11.02
Perf.2 / μ	36.06	34.87	35.30	33.72	32.30	33.39
Perf.2 / σ	4.25	4.98	4.24	5.13	6.82	4.99
Lat. Perc. / μ	4.84	5.87	3.72	5.42	3.27	4.33
Lat. Perc. / σ	1.54	0.96	2.09	1.41	1.85	1.61

C. Task Efficiency

The results regarding the completion times of task 1 are depicted in Fig. 6 (left). We found that higher latency tends to lead to slightly higher completion times for both avatar variants. Also, completion times tended to be generally slightly lower when using the point cloud avatar. However, we found no significant differences, and only the pair of M3 and P1 was near the threshold with $p = 0.06$. When looking at the task performance of task 2 (see Fig. 6 (middle)), we can observe the same trend of the performance again decreasing

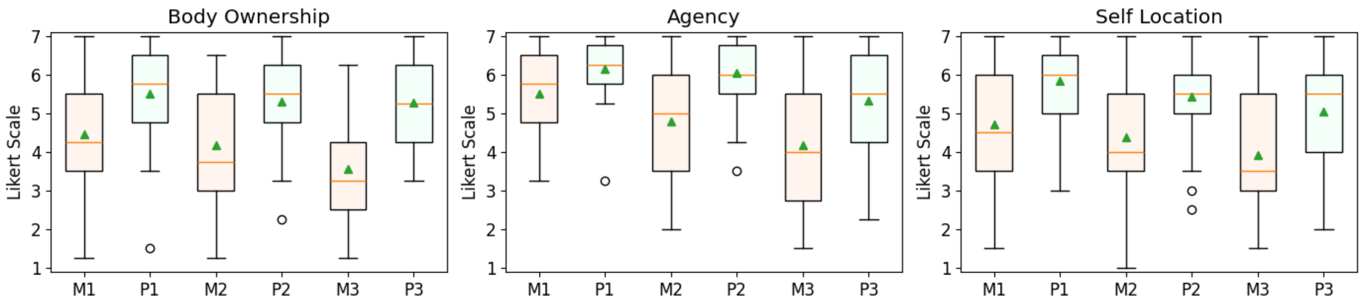

Fig. 4. Influence of avatar representation (M=mesh, P=point cloud) and latency (1=150, 2=225, 3=300 ms) on the 3 sub-factors of embodiment.

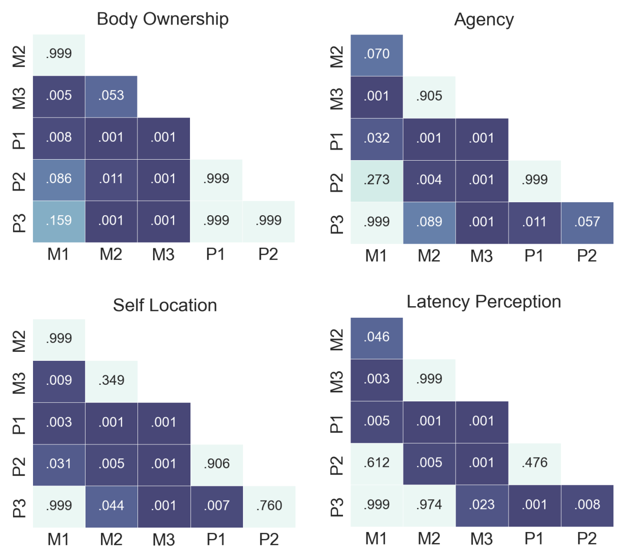

Fig. 5. Significance matrices between avatar type (M/P) and latency (1/2/3). Sign. diff. in dark blue (threshold $\leq$ .005, clamped at .999/.001).

slightly with higher latency. This time, however, at least for the lower and medium-high latency settings, the performance with the point cloud avatar tended to be slightly worse than with the mesh variant, while the behavior flips somewhat in the highest latency setting. Here, we found M3 to be significantly lower than M1 ($p = 0.003$) and M2 ($p = 0.047$). The means and standard deviations can be seen in Table I. For the task performance of task 1, we found no notable differences at all between any conditions (for plots and detailed data, we refer to the supplementary material). Overall, all these effects regarding task performance and completion times are very weak (especially in task 1) and mostly not significant.

D. Latency Perception and Cybersickness

To investigate the participants' perception of the latency depending on the avatar, we analyzed the answers to the eighth question of the embodiment questionnaire ("The movements of the virtual body were synchronized with my own movements"). The results are depicted in Fig. 6; averages and standard deviations can be seen in Table I. The point cloud avatar is generally perceived as significantly more synchronized with the own body than the mesh avatar at the corresponding latency level. Furthermore, with the point cloud avatar, the perceived synchronicity decreases less severely with increased latency and for the most part only later at higher latency levels. Concretely, with the mesh avatar the relative score losses with successively increased latency are -23.1 % and -12.1 %, while the point cloud avatar loses only -7.7 % and -20.1 %. Accordingly, significant differences exist mostly between M1 and M2/M3 and P1/P2 and P3, see Fig. 5 (bottom-right). However, we found again no interaction between latency and avatar representation.

The average cybersickness level increased from 1.18 ($SD = 0.39$) before the study with each successive round to 1.18, 1.21, 1.36, 1.39, 1.51 and eventually 1.52 ($SD = 0.87$) after the study. This steady increase is notable but does not quite reach the significance threshold with $p = 0.054$ between the pre- and post-study scores. The cybersickness was generally low, even after the study most participants had no symptoms at all, and no participant had such severe symptoms that he had to quit the study. Moreover, the relative increase in cybersickness was similar between the conditions.

VI. Discussion

First of all, the results showed that both factors, avatar quality and latency, significantly affect the sense of embodiment, which is consistent with prior work [5], [14]. Actually, we found that both factors significantly affect all three sub-factors of embodiment and that these are affected rather similarly. The higher quality point cloud avatar is consistently rated significantly higher than the mesh avatar regarding body ownership, agency, and self-location and these decrease mostly somewhat linearly with increased latency.

The difference between avatars is the most pronounced for body ownership, as the mesh avatar scores are particularly weak in this regard, and the point cloud avatar's body ownership is much less affected by an increase in latency. For all latency levels, switching to the better quality avatar increases the body ownership more than reducing the latency. Hence, we can confirm our hypothesis H_1. We observed a similar overall behavior for the agency, although the difference in avatar variants is slightly smaller, and the mesh scores are not as bad as with body ownership. Also, the increase in latency has a stronger effect on both the avatars' agencies. However, the point cloud avatar was still superior, and contrary to our

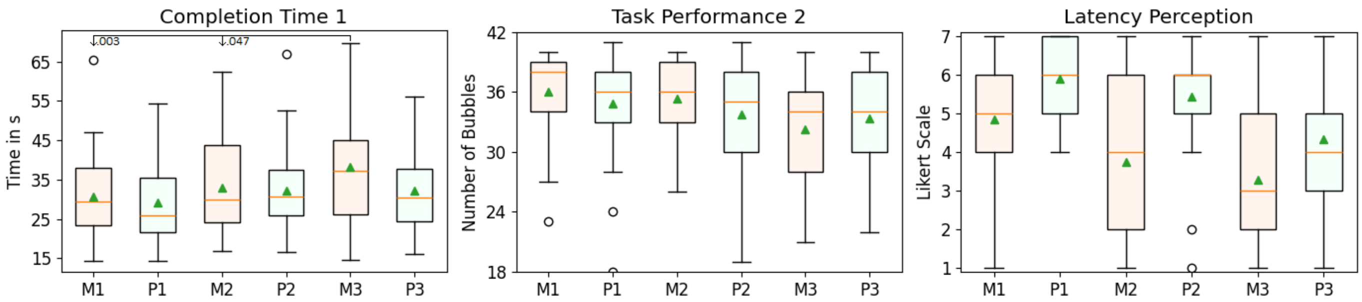

Fig. 6. Influence of avatar representation (M=mesh, P=point cloud) and latency (1=150, 2=225, 3=300 ms) on task efficiency and latency perception.

expectation, the difference between avatars was mostly higher than between latency levels for the same avatar. For instance, only with latencies of 300 ms had the point cloud avatar an agency rating as low as the mesh with 150 ms. Therefore, we falsified our hypothesis H_2. With self-location, the overall behavior is again similar to that of the other factors; the point cloud avatar got higher ratings, with increased latency the decrease in ratings is weaker than with the mesh, and differences between avatars are higher than between latencies. Thus, we also falsified our hypothesis H_3. The higher quality avatar being superior and more impactful for all three factors means that the same is true for the overall sense of embodiment. This result is highly relevant and somewhat intriguing. A potential reason for this is that the point cloud avatar not only is visually superior but also regarding the control aspect which is also relevant for agency and self-location. These findings show that higher-quality real-time avatar representations should be developed and that they can be employed beneficially even though they cause higher latencies. For instance, from our results, there is no significant loss in embodiment when going from 150 to 225 ms of latency with the point cloud avatar. These additional 75 ms could then be spend in additional processing to qualitatively enhance the avatar.

Similar to the embodiment, we found that increased latency led to decreased task efficiency for both avatars, which is also consistent with prior work [22]. Interestingly, however, the decrease was less severe than we would have expected, especially for task 1. This may be due to the task being relatively easy and being without too much pressure (which was found to be a relevant factor [31]). Comparing the two avatar variants, we also found only small, inconclusive differences. The high-quality point cloud avatar performed slightly better in task 1, the mesh one in task 2 for the lower to medium latencies. Generally, the differences regarding task efficiency were quite small (less than expected), both, between avatars and latency levels, and most often not statistically significant. Thus, we cannot confirm or falsify hypothesis H_4.

Looking at the latency perception results, we see that the point cloud avatar is again significantly superior for all latency levels. The results decrease notably with latency for both avatar types, but less so for the point-cloud avatar. Moreover, for the latter, most of the loss occurs later at higher latencies than for the mesh. From this we can conclude that the latency is perceived as less severe with the higher quality avatar, confirming our hypothesis H_5. This shows that the avatar quality is another factor influencing the perception of latency, which can be utilized in future VR development.

Cybersickness levels tended to increase over the course of the study, but not significantly and less than expected given the duration and presence of latency. This confirms the conflicting literature [5], [18]. Potentially, the increase in cybersickness would have been higher with more complex scenes and tasks that require more moving around. Because the levels of cybersickness were generally low and the differences so small, we found no significant differences between conditions. Overall, despite the trend toward increased cybersickness, we cannot confirm hypothesis H_6. An overview over all our findings regarding our hypotheses is given in Tab. II.

TABLE II
OVERVIEW AND RESULTS OF OUR HYPOTHESES.

Hypothesis	**Result**
H_1: Avatar decr. body own. more than latency	confirmed
H_2: Latency decr. agency more than avatar	falsified
H_3: Latency decr. self-loc. more than avatar	falsified
H_4: Latency decr. task eff. more than avatar	inconclusive
H_5: Avatar affects latency perception	confirmed
H_6: Cybersickness increases over time	inconclusive

VII. LIMITATIONS

One limitation is that the latency range was rather small, especially the lower end was quite high, even though we optimized the point cloud avatar's pipeline to minimize the minimal achievable latency. Also, we only tested two avatar variants and measured the difference in latency perception based on a one-item questionnaire, which might not be as reliable as a more comprehensive questionnaire.

VIII. CONCLUSIONS AND FUTURE WORK

With this work, we presented a user study with $n = 33$ participants to investigate the effects of latency and avatar quality on embodiment in VR. The goal was to examine which factor is more dominant, if the perception of the latency is affected by the avatar('s quality), and how much latency is tolerable. For our study, we implemented higher and lower quality avatars (live-captured point cloud and generic mesh, respectively) and three artificial (motion-to-photon) latency settings between 150 ms and 300 ms. We evaluated the sense of embodiment, task efficiency, and the level of cybersickness based on 3 tasks with varying degrees of required motion. The results showed that both avatar representation and latency significantly affected all components of embodiment, that the high-quality avatar was consistently and significantly superior, and that the avatar representation had a consistently higher influence than the latency level. We also found that avatar representation affects the perception of latency, specifically that latency is perceived as less severe/negative with higher quality avatars. The effects of avatar type and latency on task efficiency and cybersickness were surprisingly small and mostly not significant. We found no interactions between avatar type and latency. Our results indicate that higher-quality avatars should be prioritized over lower-quality ones, even when causing higher latencies, as they still lead to a higher sense of embodiment (all 3 sub-factors) and even reduce the perceived negative impact on latency.

In the future, we plan to conduct studies with greater latency ranges, especially below 150 ms. Also, it would be interesting to compare multiple point cloud and real-time reconstructed mesh avatars with different fidelity settings each to obtain more comprehensive and applicable comparisons. Another important step would be to expand the evaluation to multi-user scenarios and investigate social and co-presence.

Acknowledgment

This work was partially supported by grant 50RK2253B (VaMex-VRN) of the German Aerospace Center (DLR).

References

[1] A. Schäfer, G. Reis, and D. Stricker, "A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems," *ACM Comput. Surv.*, vol. 55, no. 6, dec 2022.

[2] R. Weller, C. Schröder, J. Teuber, P. Dittmann, and G. Zachmann, "VR-Interactions for Planning Planetary Swarm Exploration Missions in VaMEx-VTB," in *AeroConf.* IEEE, Mar. 2021.

[3] A. Salagean, E. Crellin, M. Parsons, D. Cosker, and D. Stanton Fraser, "Meeting your virtual twin: Effects of photorealism and personalization on embodiment, self-identification and perception of self-avatars in virtual reality," in *Proc. CHI*. ACM, 2023.

[4] F. Weidner, G. Boettcher, S. Arévalo Arboleda, C. Diao, L. Sinani, C. Kunert, C. Gerhardt, W. Broll, and A. Raake, "A systematic review on the visualization of avatars and agents in ar & vr displayed using head-mounted displays," *IEEE Trans. Vis. Comput. Graphics*, vol. PP, pp. 1–11, 05 2023.

[5] P. Caserman, M. Martinussen, and S. Göbel, "Effects of end-to-end latency on user experience and performance in immersive virtual reality applications," in *Entertainment Computing and Serious Games*. Springer, 2019, pp. 57–69.

[6] R. Fribourg, F. Argelaguet, A. Lécuyer, and L. Hoyet, "Avatar and sense of embodiment: Studying the relative preference between appearance, control and point of view," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 5, pp. 2062–2072, 2020.

[7] M. Claypool and K. Claypool, "Latency can kill: precision and deadline in online games," in *Proc. MMSys*. ACM, 2010, p. 215–222.

[8] F. Unruh, D. Vogel, M. Landeck, J.-L. Lugrin, and M. Latoschik, "Body and time: Virtual embodiment and its effect on time perception," *IEEE Trans. Vis. Comput. Graphics*, vol. PP, pp. 1–11, 05 2023.

[9] D. Halbhuber, M. Kocur, A. Kalus, K. Angermeyer, V. Schwind, and N. Henze, "Understanding the effects of perceived avatar appearance on latency sensitivity in full-body motion-tracked virtual reality," in *MuC*. ACM, 2023, p. 1–15.

[10] K. Kilteni, R. Groten, and M. Slater, "The sense of embodiment in virtual reality," *Presence Teleoperators & Virtual Environments*, vol. 21, 11 2012.

[11] E. A. McManus, B. Bodenheimer, S. Streuber, S. de la Rosa, H. H. Bülthoff, and B. J. Mohler, "The influence of avatar (self and character) animations on distance estimation, object interaction and locomotion in immersive virtual environments," in *Proc. APGV*. ACM, 2011, p. 37–44.

[12] J. C. Eubanks, A. G. Moore, P. A. Fishwick, and R. P. McMahan, "A preliminary embodiment short questionnaire," *Front. VR*, vol. 2, 2021.

[13] A. Maselli and M. Slater, "The building blocks of the full body ownership illusion," *Front. human neuroscience*, vol. 7, p. 83, 03 2013.

[14] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch, "The effect of avatar realism in immersive social virtual realities," in *Proc. VRST*. ACM, 2017.

[15] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik, "The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response," *IEEE Trans. Vis. Comput. Graphics*, vol. 24, no. 4, pp. 1643–1652, 2018.

[16] N. Yee and J. Bailenson, "The proteus effect: The effect of transformed self-representation on behavior," *Human Communication Research*, vol. 33, pp. 271 – 290, 07 2007.

[17] S. Kasahara, K. Konno, R. Owaki, T. Nishi, A. Takeshita, T. Ito, S. Kasuga, and J. Ushiba, "Malleable embodiment: Changing sense of embodiment by spatial-temporal deformation of virtual human body," in *Proc. CHI*. ACM, 2017, p. 6438–6448.

[18] T. Waltemate, I. Senna, F. Hülsmann, M. Rohde, S. Kopp, M. Ernst, and M. Botsch, "The impact of latency on perceptual judgments and motor performance in closed-loop interaction in virtual reality," in *Proc. VRST*. ACM, 2016, p. 27–35.

[19] M. Meehan, S. Razzaque, M. Whitton, and F. Brooks, "Effect of latency on presence in stressful virtual environments," in *Proc. IEEE VR*, 2003, pp. 141–148.

[20] J. Martens, T. Franke, N. Rauh, and J. Krems, "Effects of low-range latency on performance and perception in a virtual, unstable second-order control task," *Quality and User Experience*, vol. 3, 10 2018.

[21] R. Zhou, S. Shao, W. Li, and L. Zhou, "How to define the user's tolerance of response time in using mobile applications," in *IEEE IEEM*, 12 2016, pp. 281–285.

[22] L. Hoyet, C. Spies, P. Plantard, A. Sorel, R. Kulpa, and F. Multon, "Influence of motion speed on the perception of latency in avatar control," in *IEEE AIVR*, 2019, pp. 286–2863.

[23] N. Toothman and M. Neff, "The impact of avatar tracking errors on user experience in vr," in *IEEE VR*. IEEE Computer Society, mar 2019, pp. 756–766.

[24] J.-P. Stauffert, F. Niebling, and M. E. Latoschik, "Effects of latency jitter on simulator sickness in a search task," in *IEEE VR*, 2018, pp. 121–127.

[25] R. Fischer, A. Mühlenbrock, F. Kulapichitr, V. N. Uslar, D. Weyhe, and G. Zachmann, *Evaluation of Point Cloud Streaming and Rendering for VR-Based Telepresence in the OR*. Springer International Publishing, 09 2022, pp. 89–110.

[26] A. Mühlenbrock, R. Fischer, C. Schröder-Dering, R. Weller, and G. Zachmann, "Fast, accurate and robust registration of multiple depth sensors without need for rgb and ir images," *The Visual Computer*, vol. 38, 05 2022.

[27] D. Roth and M. E. Latoschik, "Construction of the virtual embodiment questionnaire (veq)," *IEEE Trans. Vis. Comput. Graphics*, vol. 26, no. 12, pp. 3546–3556, 2020.

[28] P. Bockelman and D. Lingum, "Factors of cybersickness," in *HCI International – Posters' Ext. Abs.* Springer, 2017, pp. 3–8.

[29] K. Yu, G. Gorbachev, U. Eck, F. Pankratz, N. Navab, and D. Roth, "Avatars for teleconsultation: Effects of avatar embodiment techniques on user perception in 3d asymmetric telepresence," *IEEE Trans. Vis. Comput. Graphics*, vol. PP, pp. 1–1, 08 2021.

[30] R. Doherty and P. Sorenson, "Keeping users in the flow: Mapping system responsiveness with user experience," *Procedia Manufacturing*, vol. 3, pp. 4384–4391, 12 2015.

[31] K. Schoenenberg, "The quality of mediated-conversations under transmission delay," Ph.D. dissertation, TU Berlin, 01 2016.

[32] J. C. Eubanks, A. G. Moore, P. A. Fishwick, and R. P. McMahan, "The effects of body tracking fidelity on embodiment of an inverse-kinematic avatar for male participants," in *ISMAR*, 2020, pp. 54–63.

[33] B. Keshavarz and H. Hecht, "Validating an efficient method to quantify motion sickness," *Human factors*, vol. 53, pp. 415–26, 08 2011.

APPENDIX

TABLE III
AVERAGE AND STANDARD DEVIATIONS.

Factor / Metric	**Condition**					
	M1	**P1**	**M2**	**P2**	**M3**	**P3**
Bod. Own. / μ	4.46	5.51	4.18	5.30	3.56	5.27
Bod. Own. / σ	1.61	1.37	1.68	1.42	1.55	1.27
Agency / μ	5.50	6.14	4.78	6.03	4.17	5.31
Agency / σ	1.36	1.08	1.87	1.11	1.93	1.52
Self Loc. / μ	4.71	5.83	4.37	5.43	3.92	5.04
Self Loc. / σ	1.62	1.18	1.63	1.36	1.76	1.63
Embod. / μ	4.89	5.83	4.44	5.59	3.88	5.21
Embod. / σ	1.59	1.24	1.74	1.34	1.76	1.48
Perf.1 / μ	52.85	51.62	51.93	51.76	60.90	56.64
Perf.1 / σ	29.91	36.05	39.00	32.22	45.32	47.49
Time1 / μ	30.79	29.07	32.91	32.14	38.19	32.30
Time1 / σ	10.52	9.74	11.56	10.83	14.63	11.02
Perf.2 / μ	36.06	34.87	35.30	33.72	32.30	33.39
Perf.2 / σ	4.25	4.98	4.24	5.13	6.82	4.99
Lat. Perc. / μ	4.84	5.87	3.72	5.42	3.27	4.33
Lat. Perc. / σ	1.54	0.96	2.09	1.41	1.85	1.61
Factor / Metric	**Round**					
	1	**2**	**3**	**4**	**5**	**6**
Cybers. / μ	1.18	1.18	1.21	1.36	1.39	1.52
Cybers. / σ	0.39	0.46	0.48	0.74	0.70	0.87

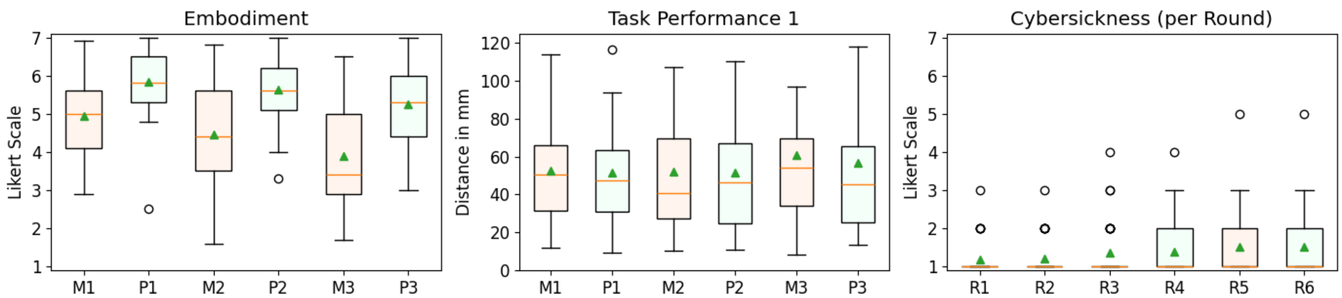

Fig. 7. Additional box plots.

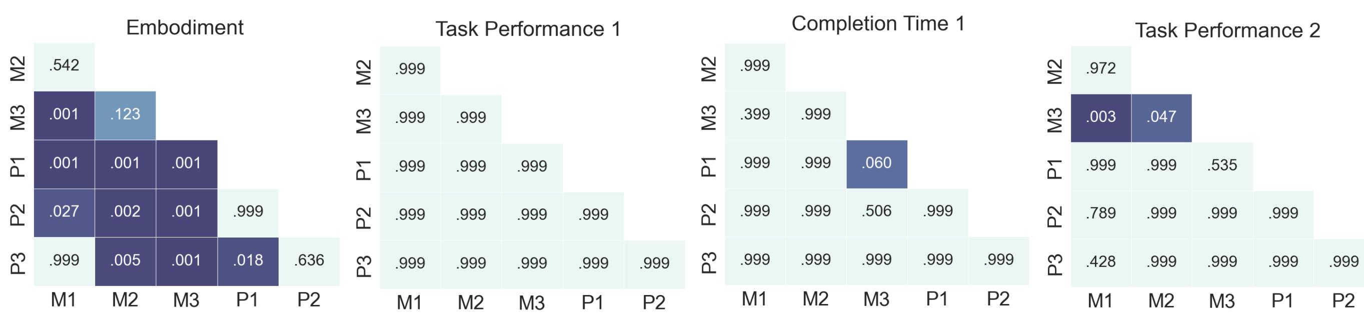

Fig. 8. Additional significance matrices.