Avatar Type Affects Performance of Cognitive Tasks in Virtual Reality

Ye Pan
University College London
United Kingdom

Anthony Steed
University College London
United Kingdom

ABSTRACT

Current consumer virtual reality applications typically represent the user by an avatar comprising a simple head/torso and decoupled hands. In the prior work of Steed et al. it was shown that the presence or absence of an avatar could have a significant impact on the cognitive load of the user. We extend that work in two ways. First they only used a full-body avatar with articulated arms, so we add a condition with hands-only representation similar to the majority of current consumer applications. Second we provide a real-world benchmark so as to start to get at the impact of using any immersive system. We validate the prior results: real and full body avatar performance on a memory task is significantly better than no avatar. However the hands only condition is not significantly different than either these two extremes. We discuss why this might be, in particular we discuss the potential for a individual variation in response to the embodiment level.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**.

KEYWORDS

Avatar, Embodiment, Cognitive Tasks, Virtual Reality

1 INTRODUCTION

Current consumer systems provide a range of different capabilities of tracking of the user: from systems that only track the rotation of the head, through to systems that track the head and two hands [Murphy 2017; Steed et al. 2016a]. From a user experience point of view, there is an obvious tension between showing an avatar because it helps the user understand how their actions are or might affect the environment, and confusing the user because the don't recognize the objects or the avatar has behaviours that don't match what the user is doing [Pan and Steed 2017; Slater et al. 1995].

In this paper we expand upon previous work of Steed et al. that looked at how the representation of a self-avatar impacts cognitive load [Steed et al. 2016b]. In that paper the authors demonstrate that having an avatar, and being able to move it, improves performance on a memory task that is run alongside a spatial reasoning task. Users who had an avatar and could move performed significantly better than users who either did not have an avatar or had an avatar but were instructed not to move it. That drew on prior work in embodied and enactive cognition [Clark 2008; Noë 2004] that suggested that movement of the hands is a part of thinking (see also Section 2). However, the study was performed with only two variants of user representation: no body representation or a full body with articulated arms. We thus consider a user representation consisting of hands only as is commonly found in current applications. That study also overlooked the issue of comparison to a real benchmark: doing the task for "real" so the user can see their own body.

Thus we run the same protocol of Steed et al., [Steed et al. 2016b] but include four conditions: no body, hands-only representation, articulated body and real. Our expectation is that the performance on the memorisation and recall task will follow the order (low to high): no-body < hands-only < full-body < real.

The results partially support these hypotheses. They show a significant difference between no-body and the pair of conditions full-body and real, with performance on no-body being worse. However hands-only is not different from any of the other three. This is interesting and suggests that they could lie in the middle. We discuss the implication of this in light of previous work and suggest that this is evidence that individual reactions to these different avatars needs to be considered (see Section 7).

2 RELATED WORK

2.1 Avatars and Body Ownership

The illusion of ownership over one's body represented as a virtual avatar is now well-studied within virtual reality. For example, Sanchez-Vives et al. demonstrated ownership through participants observing visuomotor correlations [Sanchez-Vives et al. 2010] and Yuan and Steed showed that the association between virtual limb and body could be made by the participant engaging in an interactive task in a HMD-based virtual reality system [Yuan and Steed 2010]. These illusions suggest that the form of the self-avatar has an important impact on user response. Lugrin et al. investigated avatar types varying between humanoid, machine-like and cartoony. They found that each avatar generated a high level of ownership, but that the humanoid avatar was slighly lower [Lugrin et al. 2015]. Similarly, Jo et al. showed that a faithful reproduction of the participant's hand solicited a lower body ownership than a cartoony

avatar [Jo et al. 2017]. Argelaguet et al. showed that participants had a higher sense of ownership of a virtual hand that mapped directly to their real hand [Argelaguet et al. 2016]. Together these results suggest that behavioural fidelity might be more important than visual fidelity.

As noted previously, in many virtual reality simulations, there is only information about the head and hand locations, so other body parts need to be inferred. Jung and Hughes demonstrated that inferred parts of the avatar, specifically the lower body, impacted body ownership [Jung and Hughes 2016]. Kondo et al. [Kondo et al. 2018] tested whether body ownership could be induced to an invisible body using virtual socks and gloves synchronized with a participant's movement. They found that in the body ownership induced by only socks and gloves, observers perceived a complete body between socks and gloves, and the proprioceptive self-localization drift toward the invisible body was similar to the one observed in the full-body ownership illusion. These papers raise the interesting question of whether the presence of a partial body will impact cognitive strategies.

2.2 Cognition

From linguistics and cognition there are many studies that focus on the use of gestures in cognitive tasks. In particular, Goldin-Meadow et al. explored how children and adults performed mathematical tasks [Goldin-Meadow et al. 2001]. Participants were asked to memorize a set of letters, perform an explanation and then repeat the letters. They found that participants who were prevented from gesturing, by having them sit on their hands, recalled fewer letters. The main claim was that the letter recall task and explanation tasks both required significant cognitive resource, but participants were able to recruit their gestures to offset some effort for the explanation task, thus leaving more resource for the letter recall task.

3 TASKS

The protocol was as closely as possible a copy of that in [Steed et al. 2016b]. The two tasks employed were based on prior work, specifically a letter recall task from [Goldin-Meadow et al. 2001] and a mental rotation task similar to [Chu and Kita 2011]. Each trial consisted of both tasks, see the sequence of five images in Figure 2 which show the sequence of five cards that are shown on the table in front of the participant in the virtual environment (e.g., Figure 1(b)).

The letter recall task involves the participant being shown a card with the instruction to memorise four letter pairs (Figure 2a). They are then requested to recall these letters after completing the mental rotation task, see Figure 2d. Participants thus had to keep these letter pairs in mind for over 25 seconds without a visual prompt. They were allowed to look at this first card for 15 seconds.

The spatial rotation task is shown in the second and third images/cards. The participant is shown the second card for 15 seconds, Figure 2b. This card has a figure of some blocks and a row of four possible matching blocks underneath in different orientations. Two of the lower figures match the upper figure. After 15 seconds they are prompted to give their answer, see Figure 2c, and they have another 10 seconds before the figures disappear. After recalling the letters, there is a five second wait (Figure 2e) before the next trial.

4 TECHNICAL SETUP

The virtual reality system was built on a Windows 8.1 computer with an Intel Core i7 processor, 8GB ram and a GeForce TitanX graphics card. The head-mounted display was an Oculus Rift CV1. For head tracking we used the Oculus camera-based tracking. For hand tracking we used a Polhemus Fastrak tracker with two tracked receiver units. The Polhemus Fastrak is a 6 degree of freedom magnetic tracking system with an accurate range of approximately 1m. The limited effective range was not a problem for this study as participants sat at a table. The tracker transmitter base unit was placed on the table (small grey cube on the table in Figure 1(a)). The two tracker receiver units were attached near the wrist by Velcro on the back of weightlifting gloves that the participants wore. Different size gloves were available. A second Velcro strap on each arm kept the wire for the receiver units away from the table top so that it did not interfere with gestures. We used this tracker, rather than the Oculus Touch controllers to give a better impression of empty-handed interaction, to better reproduce the situation of [Steed et al. 2016b]. As the Fastrak is magnetic, it also worked well in a seated situation where the participant may put their hands under the table and thus out of sight of a camera.

The control system used the Unity 5.1.1 software. We used the MiddleVR 1.6 library to interface to the Polhemus device over VRPN. All scenes were rendered at 90Hz. The latency in tracking, updating the head position and rendering was approximately 18ms. The hand tracking had an end to end latency of approximately 60ms, but this was not noticed by any participant.

The virtual environment was a simple room with blue walls. There was a table holding the task materials, and a stool, see Figure 1(b). The full body avatars were either male or female assets from Rocketbox Studios. The self-avatars were animated using the Final IK plugin from RootMotion [RootMotion 2019]. See Figure 1(c) for a first person view. The hands only avatar was based on the hands from the VR Hands and FP Arms asset pack for Unity [FPArms 2019], see Figure 1(d).

5 METHOD

32 students and staff (15 female) at University College London, were recruited to take part in our user study. The median age was 27 (SD = 9.17). Each participant took part in one of four between-subject conditions: NoAvatar, HandsOnly, FullAvatar, and Real.

Participants were first given an information sheet, asked to read through this and sign a corresponding consent form. The experimenter explained the task using a paper example and then helped the participant don the equipment. All participants first put on the hand trackers. If the participant was undertaking the task in the HMD they then donned it. Participants in the Real condition thus did not use the HMD but still wore the hand trackers in case this affected their use of their hands (e.g. due to the hand trackers being connected by a wire). Each participant undertook 3 practice trials where they could ask questions. They were asked to confirm that they could carry on without further questions. They then undertook 20 measured trials.

The participants were paid £5. The experiment took about 20 minutes. The experiment was approved by University College London Research Ethics Committee.
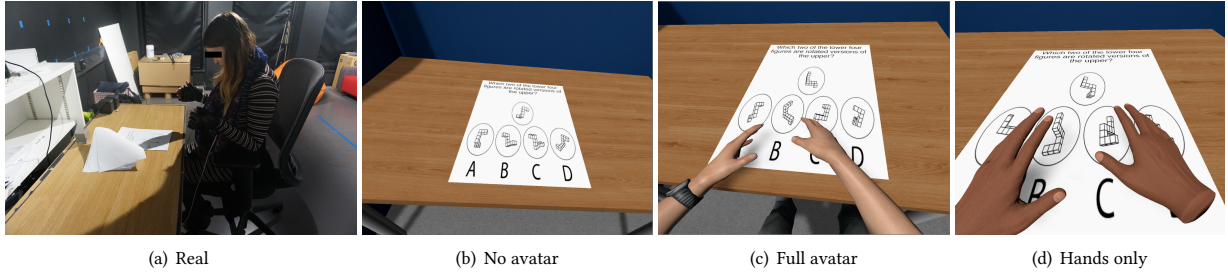
(a) Real   (b) No avatar   (c) Full avatar   (d) Hands only

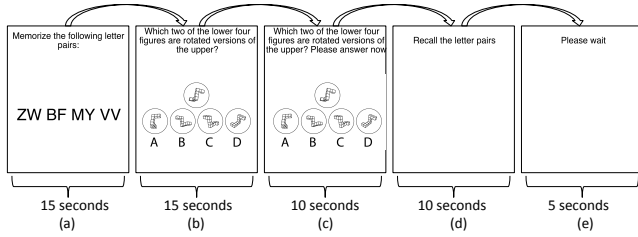**Figure 1: The four conditions for the experiment**

Figure 2: Examples of the five cards shown to participant in one trial. Each trial comprises two tasks: a letter recall test and a mental rotation test. From [Steed et al. 2016b] with permission.

Table 1: Summary of results, including mean and SD for % correct answers of the letter recall task and spatial task, and the number of occurrences of using gestures.

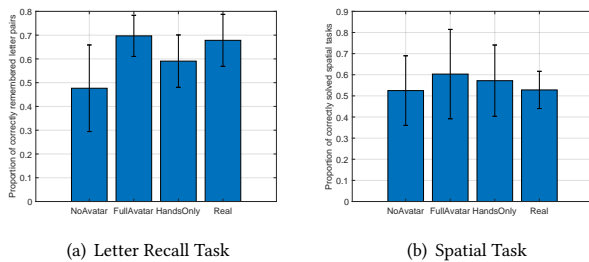| Gesture	NoAvatar	Hands	Full	Real
Letter Recall (M)	.48	.59	.68	.7
Letter Recall (SD)	.18	.11	.11	.65
Spatial (M)	.52	.57	.6	.52
Spatial (SD)	.16	.17	.21	.09
Use gesture	27/160	74/160	160/160	160/160

(a) Letter Recall Task   (b) Spatial Task

Figure 3: Proportion of correctly remembered letter pairs and solved spatial tasks. The error bars show SD.

6 RESULTS

We measured number of letters recalled, number of correct answers on the spatial task and observation of participants making gestures. Table 1 gives a summary of the results.

6.0.1 Letter Recall. Each of the 20 trials involved the participants recalling four letter pairs. The experimenter records each letter verbalised, and gives a single mark for each correct letter. Thus the maximum achievable score is 80, but we report as a percentage correct. Figure 3(a) shows the mean proportion of correctly remembered letter pairs in four conditions.

A one-way ANOVA was conducted to determine if the mean proportion of correctly remembered letter pairs for participants with different levels of embodiment. There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test (p > .05); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .253). Data is presented as mean and standard deviation. There was a significant main effect of our four conditions, $F(3, 28) = 5.002, p = .007$. The mean increased from the NoAvatar (47.66% ± 18.22%), to HandOnly (59.06% ± 11.01%), to Real (67.81% ± 10.93%) to Full avatar (69.69% ± .65%) groups, in that order. Tukey post hoc analysis revealed that the mean score in the NoAvatar group was significantly lower than the FullAvatar group ($p = .009$) and Real group, ($p = .018$), but no other group differences were statistically significant.

6.0.2 Mental Rotation. Each of the 20 trials involved the participant give two letters corresponding to the two matching figures. The experimenter records the first two answers and gives a single mark for each if correct. Thus the maximum achievable score is 40, but we report as a percentage correct. Figure 3(b) shows the mean proportion of correctly solved spatial task in four experimental conditions.

A one-way ANOVA was conducted to determine if the mean proportion of correctly solved spatial task was different for groups with different levels of embodiment. Participants were classified into four groups: NoAvatar, HandsOnly, FullAvatar, and Real. There were no outliers, as assessed by boxplot; data was normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .104$). Results revealed that the

differences between these groups was not statistically significant, $F(3, 28) = .415, p = .743$.

6.0.3 Gesturing. The experimenter also counted how many trials each participant gestured during while the experiment was in progress. The gesture had to be a rotation gesture or similar, but not self-touching gesture such as scratching. We only report the total number of gestures made by all subjects. With eight participants in each condition, there is a maximum of 160 gesture occurrences.The observation results on gesturing are shown in the last line of Table 1. There were 27 occurrences of gestures in the NoAvatar condition, 74 in HandsOnly and 160 in both FullAvatar and Real conditions.

7 DISCUSSION

First, as in previous work, we did not expect to find a difference in the mental rotation scores (see [Goldin-Meadow et al. 2001; Steed et al. 2016b]). Thus the main effects are seen in the letter recall study.

On this, the main interesting finding is that there there was no significant difference between full avatar and real task performance. This was a gap in the previous work [Steed et al. 2016b], and it suggests that for this specific task, the presence of a full avatar creates a situation which is effectively similar to the real task situation. We would still expect that this would not generalize across a very broad range of tasks, but it opens up the interesting question of delineating the scope of those tasks for which the performance on cognitive tasks is similar.

We found no significant difference between hands only and full avatar. This is in line with recent literature that suggests that a full body illusion can be induced in partial body representations [Kondo et al. 2018]. Thus we can't exclude the possibility that the representation of the arms and torso is unnecessary in this specific task. However, the work on appearance of the hands and arms also suggests that more realism may be counterproductive (e.g. [Jo et al. 2017].). We also note that the hands only condition is not significantly different than the no avatar condition. This suggests that there might be a significant participant-dependent effect of this condition. This might be due to the appearance of the hands; either because it is different to the participant's own hands (noting it is a between subjects experiment), or because the lack of arms draws attention to the appearance of the hands.

We note that the no avatar group was significantly lower than the full avatar. This reproduces one of the key findings of the study of Steed et al. [Steed et al. 2016b] that we based our study on.

Finally, we note the dramatic differences in occurrences of gestures in the different conditions. There are very few gestures in the no body condition (27/160) compared to the full body and real conditions (160/160). A similar differences was found just for the no avatar versus full body avatar in Steed et al. [Steed et al. 2016b]. Interestingly the hands only condition is between these two extremes (74/160). This lends credence to the suggestion that the hands only avatar are not quite the same as either no avatar or full avatar conditions.

8 CONCLUSION

Our results confirm and extend a prior work that looked only at presence or not of an avatar [Steed et al. 2016b]. We reproduce

one of their results (equivalent to no body < full body), and extend it with an additional virtual condition (hands-only) and also the comparison to real. These additional results suggest that the hand-only avatar is not a poor solution, but it is not obviously as good as a full-body avatar. Our current hypothesis is that there is a significant variation in individual reaction to the hands-only avatar. This is consistent with related work on the importance of visual and behavioural fidelity for ownership and agency over the virtual body.

Overall the results suggest that the design of the avatar is a very important consideration, not only for purposes of supporting interaction and engendering a sense of ownership, but because the specific form of the avatar may affect a user's performance on cognitive tasks.

REFERENCES

Ferran Argelaguet, Ludovic Hoyet, Michael Trico, and Anatole Lécuyer. 2016. The role of interaction in virtual embodiment: Effects of the virtual hand representation. In *2016 IEEE Virtual Reality (VR)*. 3–10. https://doi.org/10.1109/VR.2016.7504682

Mingyuan Chu and Sotaro Kita. 2011. The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General* 140, 1 (2011), 102.

Andy Clark. 2008. *Supersizing the mind: Embodiment, action, and cognitive extension: Embodiment, action, and cognitive extension.* Oxford University Press.

FPArms 2019. VR Hands and FP Arms. Retrieved September 22, 2019 from https://assetstore.unity.com/packages/3d/characters/humanoids/vr-hands-and-fp-arms-pack-77815

Susan Goldin-Meadow, Howard Nusbaum, Spencer D Kelly, and Susan Wagner. 2001. Explaining math: Gesturing lightens the load. *Psychological Science* 12, 6 (2001), 516–522.

Dongsik Jo, Kangsoo Kim, Gregory F. Welch, Woojin Jeon, Yongwan Kim, Ki-Hong Kim, and Gerard Jounghyun Kim. 2017. The Impact of Avatar-owner Visual Similarity on Body Ownership in Immersive Virtual Reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. ACM, New York, NY, USA, 77:1–77:2. https://doi.org/10.1145/3139131.3141214

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

Ryota Kondo, Maki Sugimoto, Kouta Minamizawa, Takayuki Hoshi, Masahiko Inami, and Michiteru Kitazaki. 2018. Illusory body ownership of an invisible body interpolated between virtual hands and feet via visual-motor synchronicity. *Scientific reports* 8, 1 (2018), 7541.

Jean-Luc Lugrin, Johanna Latt, and Marc Erich Latoschik. 2015. Avatar anthropomorphism and illusion of body ownership in VR. *2015 IEEE Virtual Reality (VR)*, 229–230. https://doi.org/10.1109/VR.2015.7223379

Dooley Murphy. 2017. Bodiless embodiment: A descriptive survey of avatar bodily coherence in first-wave consumer vr applications. In *2017 IEEE Virtual Reality (VR)*. IEEE, 265–266.

Alva Noë. 2004. *Action in perception.* MIT press.

Ye Pan and Anthony Steed. 2017. The impact of self-avatars on trust and collaboration in shared virtual environments. *PLOS ONE* 12, 12 (Dec. 2017), e0189078. https://doi.org/10.1371/journal.pone.0189078

RootMotion. 2019. FinalIK. Retrieved September 22, 2019 from http://root-motion.com

Maria V. Sanchez-Vives, Bernhard Spanlang, Antonio Frisoli, Massimo Bergamasco, and Mel Slater. 2010. Virtual Hand Illusion Induced by Visuomotor Correlations. *PLoS ONE* 5, 4 (04 2010), e10381. https://doi.org/10.1371/journal.pone.0010381

Mel Slater, Martin Usoh, and Anthony Steed. 1995. Taking Steps: The Influence of a Walking Technique on Presence in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 2, 3 (Sept. 1995), 201–219. https://doi.org/10.1145/210079.210084

Anthony Steed, Sebastian Friston, María Murcia-López, Jason Drummond, Ye Pan, and David Swapp. 2016a. An In the Wild Experiment on Presence and Embodiment using Consumer Virtual Reality Equipment. *IEEE Transactions on Visualization and Computer Graphics* 22, 4 (April 2016), 1406–1414. https://doi.org/10.1109/TVCG.2016.2518135

Anthony Steed, Ye Pan, Fiona Zisch, and William Steptoe. 2016b. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)*. 67–76. https://doi.org/10.1109/VR.2016.7504689

Ye Yuan and Anthony Steed. 2010. Is the rubber hand illusion induced by immersive virtual reality?. In *Virtual Reality Conference (VR), 2010 IEEE*. IEEE, 95–102.