

The Impact of a Self-Avatar on Cognitive Load in Immersive Virtual Reality

Anthony Steed*

University College London, UK

Ye Pan†

University College London, UK

Fiona Zisch‡

University College London, UK

William Steptoe§

University College London, UK

ABSTRACT

The use of a self-avatar inside an immersive virtual reality system has been shown to have important effects on presence, interaction and perception of space. Based on studies from linguistics and cognition, in this paper we demonstrate that a self-avatar may aid the participant's cognitive processes while immersed in a virtual reality system. In our study participants were asked to memorise pairs of letters, perform a spatial rotation exercise and then recall the pairs of letters. In a between-subject factor they either had an avatar or not, and in a within-subject factor they were instructed to keep their hands still or not. We found that participants who both had an avatar and were allowed to move their hands had significantly higher letter pair recall. There was no significant difference between the other three conditions. Further analysis showed that participants who were allowed to move their hands, but could not see the self-avatar, usually didn't move their hands or stopped moving their hands after a short while. We argue that an active self-avatar may alleviate the mental load of doing the spatial rotation exercise and thus improve letter recall. The results are further evidence of the importance of an appropriate self-avatar representation in immersive virtual reality.

Index Terms: H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities;

1 INTRODUCTION

Head-mounted display (HMD) virtual reality systems surround the user's vision. One side-effect of this is that the user cannot see their own body. Thus it is common for such systems to include a virtual representation of a body that is depicted from the first-person perspective of the user's eyes (e.g. [35]). Such *self-avatars* have an obvious immediate benefit: they can give obvious cues about the user's location in the virtual world and the user's current body posture. Several potential benefits of having a self-avatar have been shown. As reviewed in more detail in Section 2, a self-avatar has been shown to have a positive benefit to interaction tasks, the sense of presence and perceptual judgements (e.g. [36]).

However, there are problems in providing a self-avatar. It is currently difficult to make a self-avatar look like the user's body and current attire, and to convincingly animate the self-avatar involves tracking multiple body parts [22]. There is a risk in generating a self-avatar in that due to mis-calibration or lack of detail in the rendering, the self-avatar might not sufficiently accurately represent the user's movements and appearance. This might then lead to the user finding it problematic to interact with the world or to the user experiencing a break in presence when they notice the discrepancies [34]. Perhaps for these reasons, amongst the current flurry of development for upcoming consumer HMD systems, relatively few applications include a complete

*e-mail: A.Steed@ucl.ac.uk

†e-mail: Y.Pan@cs.ucl.ac.uk

‡e-mail: fiona.zisch.12@ucl.ac.uk

§e-mail: W.Steptoe@ucl.ac.uk

self-avatar. Some applications include models of the hands, or the objects in the hands (e.g. [27]). In other systems, such as smartphone-based HMDs, where relatively little is known about limb placement because of the lack of position tracking, sometimes a static self-avatar is shown (e.g. [10]) but often there is no self-avatar (e.g. [40]).

Although the utility of a self-avatar has been shown for specific interaction tasks, in this paper, we ask whether the presence of a self-avatar can aid in general performance of tasks in a virtual world. In particular, whether the self-avatar can help with performance of tasks that can be completed through reasoning and imagination, not manipulation. Recent theories of enactive interaction and cognition make the argument that humans can off-load mental tasks into the physical world (see Section 2.2). Thus the acts of making sketches on paper or manipulating pieces in a game, help to make decisions (e.g. [5]). Further, gestures such as mimes can act as proxies for actual manipulation or movement [14]. If these theories are true, then the lack of a self-avatar might hinder this process of off-loading mental tasks, because there is no visual feedback in the virtual environment. For example, some people gesture when asked to mentally rotate an object. In this paper we do not delve into whether these gestures are a consequence of the mental activity or assist the mental activity; we simply expect that if an avatar is not shown, then these gestures may not happen and that the performance of these tasks will thus require more cognitive load.

To explore this issue, we designed an experiment where participants undertook a series of demanding tasks: memorisation of letter sequences, mental rotation of figures and recollection of letter sequences. We have three direct measures: success in rotating the figures, success in recalling the letters and movement of the hands. Our first three hypotheses are thus:

- Hypothesis 1 (H1): Performing the letter recollection tasks without a self-avatar will result in a lower performance than with a self-avatar.
- Hypothesis 2 (H2): Performing the mental rotation tasks without a self-avatar will result in a lower performance than with a self-avatar.
- Hypothesis 3 (H3): There will be significantly less hand movement for participants without a self-avatar than with a self-avatar.

Note that the first hypothesis is based on the assumption that if the cognitive load of doing the mental rotation is higher, then recollection of letters will be impaired.

To establish a link to prior work, we also add a second factor about whether participants are instructed to keep their hands still during the experiment so that they cannot gesture. The inability to gesture during explanations and mental rotation has previously been shown to be an impediment (see Section 2.2). Thus we add the following hypotheses:

- Hypothesis 4 (H4): Performing the letter recollection tasks with self-avatar when gestures are disallowed will result in a lower performance than when gestures are allowed.

- Hypothesis 5 (H5): Performing the mental rotation tasks with a self-avatar when gestures are disallowed will result in a lower performance than when gestures are allowed.

The experiment was a complete mixed 2 by 2 design, but we did not have a prior hypothesis about the difference between the no self-avatar gesture allowed and no self-avatar gesture not allowed conditions.

The experimental results demonstrate that hypotheses H1, H3 and H4 are supported. We do not find a performance difference on the mental rotation task (hypotheses H2 and H5), but note that the main effect of the lack of an avatar or the inability to move should be to increase cognitive load of the mental rotation task and this load would then interfere with the letter recollection task.

The remainder of this manuscript is organized as follows. In Section 2 we briefly review related work on self-avatars, embodiment and cognition. Section 3 describes the system implementation, scenario and methodology. Section 4 reports the main results. Section 5 discusses the results and then Section 6 concludes.

2 RELATED WORK

2.1 Self Representation

The provision of a self-avatar has been a feature or requirement of immersive virtual reality systems almost since their inception. In the late 1980s VPL Research was experimenting with non-realistic avatars (see discussion in [41]). The potential importance of a self-avatar for the sense of presence has been noted in early presence models such as that of Slater and Wilbur [37]. Slater et al. demonstrated that a virtual body had a significant impact on self-reports of presence during locomotion [36]. Other demonstrations, such as the UCL/UNC pit demonstration, showed that the virtual body can be an important part of a presence-generating experience [39]. Lin et al. showed that an avatar that was gender matched and calibrated to the height of the participant increased accuracy of height estimations and caused participants to be less likely to step off a ledge [21].

The utility of self-avatars has been studied broadly. Biocca provides an overview of the issues and impacts that a self-avatar can have [3]. These include that the self-avatar represents one's actions in the virtual world, but it may be difficult to identify with the self-avatar that one sees if it doesn't appear to represent the real body. However, humans appear to have a lot of flexibility in recognising and utilising a self-avatar in a virtual environment. Certainly in social situations, users can use avatars effectively to communicate [32]. For example, Dodds et al. have shown that a self-avatar is a useful resource in communicating with another person within a multi-user virtual reality system [7].

More relevant for this study is the process of *embodiment* within a self-avatar. When embodied within a self-avatar, in some ways the user treats the self-avatar as their actual body. Embodiment within self-avatars has now been extensively studied in virtual reality, partly because virtual reality has proved a flexible medium within which to control perceptual experience of experimental participants [18, 17]. This area stems from the seminal rubber hand illusion demonstration by Botvinick and Cohen [4], where a participant believes that a rubber hand is part of their body. Variations of this demonstration have been made in virtual reality demonstrating that virtual limbs can be experienced in some way as part of the body or as representative of the body [31]. 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 [42]. At the very least, these experiments show that the virtual body has an impact on how the person reacts to virtual stimuli. They also highlight that a self-avatar can be treated in some ways like a real

body. More recent work demonstrates that to some extent the brain treats the virtual avatar as if it was a real body [13] and that the user can feel agency through a virtual body [28].

Several studies have looked at the impact of a self-avatar on spatial awareness within a virtual environment. An early study by Draper was inconclusive, though this would have been conducted with rather limited equipment compared to more recent studies [9]. Later studies have shown that a self-avatar can improve distance estimation [24, 29]. Mohler et al. showed that an animated self-avatar was superior to a static self-avatar in distance estimation tasks [25]. However the effect of the self-avatar appears to be complex, with Streuber et al. not finding an effect of a self-avatar on a locomotion task and an object manipulation task [38], and McManus et al. finding mixed results for locomotion, stepping and manipulation tasks [23]. Banakou et al. have shown virtual reality embodiment within child-sized avatars impacts size judgements [1].

Overall, the prior work suggests that a self-avatar is important for motor-related tasks and for general interaction with the virtual environment. It is not so clear what the role of a self-avatar might be for abstract tasks that might require more cognitive processing and less physical action. The work on embodiment in avatars suggests that participants are able to understand and model, or at least adapt to, the visual feedback that a self-avatar provide. Thus it can be argued that if the body is useful in cognitive processing, so might a self-avatar. We expand this argument in the following section.

2.2 Cognition

2.2.1 Enactive Cognition

The study of the role of the body in everyday interaction and thought processes is a complex area that spans neuroscience, cognitive science and philosophy [2]. A recent development is that of a range of proposals around *enactive cognition* that suggest that cognition occurs as an interaction between the person and the environment, and that our ability to act is in constant and dynamic re-assessment depending on the capabilities of our bodies and the tools to hand. A very large literature has developed, but two influential texts are those of Noë [26] and Clark [6].

Related work in human-computer interaction has reflected on the success of interfaces that can exploit more of our previously learnt motor skills and common ways that we already interact with the world around us [8, 20, 16, 19, 11]. Thus interfaces can utilise known properties and affordances of objects and we can understand how our actions will affect the combined physical and digital system by reasoning about the physical, or simulated physical, interactions between our bodies and the objects that form the interface.

We can see an immediate link back to the potential utility of a self-avatar in a virtual reality: if done correctly, a self-avatar should be “invisible” as a mediator of interaction and we can act unconsciously through it. Slater has argued that the virtual reality is effective because it exploits known *sensorimotor contingencies* and the virtual body acts to reinforce the match between visual information and sensorimotor feedback and control [33].

2.2.2 Gesture and Cognition

To the discussion of the effect of a self-avatar, our paper brings in and extends a thread of work from linguistics and cognition on the use of gesture in cognitive tasks. Many people gesture when explaining and this appears to support the process of cognition. Hostetter et al. showed that in a task that required participants to describe ambiguous dot patterns, the need to describe more complex shapes solicited more complex gestures from participants [15]. The authors claimed that this supported their view that gestures play a role in conceptualising information when explaining. A subsequent review and framework discussion paper



(a) Actual lab situation.



(b) Virtual model showing background, headless male self-avatars and virtual table aligned to the real table.

Figure 1: Experiment setup

by Hostetter and Alibali suggests that some gestures are simulating actions that would transform the world or represent actions on mental imagery [14].

Goldin-Meadow et al. explored the role that gesture has in explaining mathematics [12]. They asked children and adults to remember letter sequences while explaining how they solved mathematical problems. When the participants were prevented from gesturing, they recalled significantly fewer letter sequences. The authors claim that gesturing appears to relieve some of the cognitive effort or resources required to do the explanation and thus allows participants to allocate more resource or effort to the recall task. Our study uses a similar paradigm, though with a spatial task rather than a mathematical task, see Section 3. We retain a condition of not allowing gestures, but extend the results to show how a self-avatars impacts a recall task.

More recent work has explored the role of gesture in spatial problem solving. Chu and Kita present three experiments where participant gestures were tracked and coded when performing mental rotation exercises. Participants were shown the rotation exercises on a desktop monitor. The most relevant of these experiments showed that when gestures were encouraged, participants performed better at the task [5]. We use a similar mental rotation exercise, but perform the study in an immersive virtual reality where the participant may or may not have a virtual body.

3 SYSTEM AND METHOD

3.1 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

DK2. For head tracking we used the Oculus's own head-tracker, mounted on a monitor and thus 0.4m off the table top (see Figure 1(a)). 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 a usable 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 underneath and aligned with the Oculus Rift DK2 tracker (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. The tracking system had no problems tracking the hands when the participant placed their hands on their knees under the table. No finger tracking was performed.

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 75Hz. 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.

3.2 Scene and Avatar

The scene provided three elements: a background scene, a self-avatars and the task materials.

The background scene was a model of the physical lab where the experiment took place. The physical lab is a small room with blue walls and a large black curtain. The virtual model copies these features. For this experiment we placed a table and chair in the lab, and modelled the table in the virtual scene in the same position. The chair was not modelled. This was because the chair was not visible to participants with a self-avatars and we did not want participants without a self-avatars to mis-understand the chair as a potential representation of themselves.

Some participants had a self-avatars. We provided both male and female avatars in generic clothing. The self-avatars were animated using the Final IK plugin from RootMotion [30]. We used three standard kinematic chains from this plugin. One chain connected the head tracking information to position the torso and invisible head, a second connected the shoulder and left arm joints to the left hand tracker and the third controlled the right arm in a similar manner. Collision objects were attached to both the avatars' hand geometries. Collision volumes were set up to match the table and to match the position of the knees of the avatar. Consequently, the virtual hands would not penetrate the table top or pass through the knees. Thus when the participant put their real hands flat on the table top the avatar's hands would appear to do the same with good registration between the real and virtual. The avatar's hand shape was static throughout the experience.

For rendering purposes, the self-avatars were modelled without a head as shown in Figure 1(b). Thus when the participant looked straight down they would see the torso and knees of the self-avatars. The self-avatars is visible from a first person point of view with hands on table in Figure 3(a) and with hands on knees in Figure 3(b).

The final elements of the scene, the task materials, are described in the next section.

3.3 Tasks

The two tasks were based on prior work, specifically a letter recall task from [12] and a mental rotation task similar to [5].

Each trial consisted of two tasks or tests, 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 (see, e.g., Figure 3(a)). The cards would appear approximately A3 size on the table. This was necessary to ensure that the text was easily visible in the HMD.

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. A pilot trial indicated that memorising three letter pairs was too easy for the majority of participants.

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.

3.4 Procedure

3.4.1 Participants

40 participants (21 female), students and staff at University College London, were recruited to take part in our user study. The median age was 26 ($SD = 14.01$).

3.4.2 Design

The study had a 2 self-avatar (self-avatar vs. no self-avatar) \times 2 gesture (allowed vs. not allowed) mixed design. Each participant took part in only one of the two self-avatar conditions, resulting in 2 between-subject conditions with 20 participants each. Each participant gave explanations under two gesture conditions: gesture allowed, in which their hands were unconstrained; and gesture not allowed, in which they were instructed to keep their hands still on their knees. Subjects were randomly assigned to the between-subject condition. The order of the within-subject condition was counter-balanced.

The five hypotheses of the study are listed in Section 1 and will be referred to as H1-H5 in the remainder of the paper.

The experiment was approved by University College London's Research Ethics Committee, project ID number 0439/001.

3.5 Procedure

On arriving at the laboratory, participants were presented an information sheet about the study. They were asked to read through this and then read and sign a consent form.

The experimenter explained the concept behind the tasks by showing a paper example of the letter pairs and the spatial rotation tasks. The participant did not practice these tests in the real world, but were asked if they understood the procedure and questions.

The experimenter then explained the equipment to be used and helped the participant put on the hand trackers and head-mounted display.

Each participant undertook 23 trials of the memory and spatial tests. The 23 trials were split into three sections: 1-3 were a practice section, 4-13 for section 1, 14-22 for section 2. The order of the latter two sections and the gesture conditions were counterbalanced, to reduce any confounding influence of the orderings such as learning effects or fatigue. During the practice session the participant could ask questions and was prompted to respond if they did not respond when the visual cue to respond occurred. In between the sessions, the participant was allowed to remove the HMD if they wished. All the memory tests and spatial tests were different so the participants could not learn the answers.

Upon completion of the experiment, the participants were paid £5 (approximately \$7.5) for their participation. The experiment took about 20 minutes.

3.6 Data Collection

Four measures were taken for later analysis: scoring of letter recall, scoring of spatial tasks, observation of participants making gestures, and measurement of hand movement.

3.6.1 Scoring of Letter Recall

Each trial involved the participants recalling four letter pairs. They verbally reported the letter pairs. The experimenter recorded them in an electronic form. Each correct letter pair in the correct position was given one mark. Thus the score was out of forty for each condition. This is reported as a percentage correct in the next section. Only recalling one of a pair of letters or getting the pairs out of sequence did not score anything.

3.6.2 Scoring of Spatial Tasks

Each trial involved the participant verbally reporting the two matching figures. The experimenter recorded the first two answers unless positively corrected by the participant. One mark was given for each correct figure. The score was out of twenty for each condition of ten trials. This was reported as a percentage correct in the following section.

3.6.3 Making Gestures

The experimenter observed when the participant made gestures in each condition. They would count how many trials the participant made a rotation gesture that was not obviously a simple inadvertent gesture such as scratching their arm or pushing the HMD. Each participant could thus be rated as gesturing up to 10 times, once for each trial, for each condition they undertook.

3.6.4 Movement Measures

The participant's head, left hand and right hand movement were recorded throughout the experiment in a log file. The log file also recorded the timing of all card changes and trial numbers. Positions were written to the log file at 30Hz.

A script was written to extract the position logs for each condition for each participant. The length of this is the same for each participant. The script filtered the tracking data by taking a moving average with a box filter over 30 samples. Thus movements and noise under 1Hz were removed. This threshold removes jitter and very fast gestures, but preserves gross movements. A total movement for each hand was then calculated summing the Euclidean difference between the smoothed samples.

The total movement measure is sensitive to the width of the filter applied. We performed the statistical analysis of the total movement measure after applying several filter widths from 3 samples through to over 100 samples. Each of these analyses provided the same main effects as the chosen 1Hz (30 sample) filter that is reported in the results section.

In the next section, for completeness we report the left and right hand movement as well as the total hand movement, but the analysis is done on the latter. We have not analysed the head movement as we had no prior hypotheses about differences in motion that would occur. Participants had to pay close attention to the cards in front of them and thus kept their heads fairly still.

4 RESULTS

Table 1 gives a summary of the four measures taken.

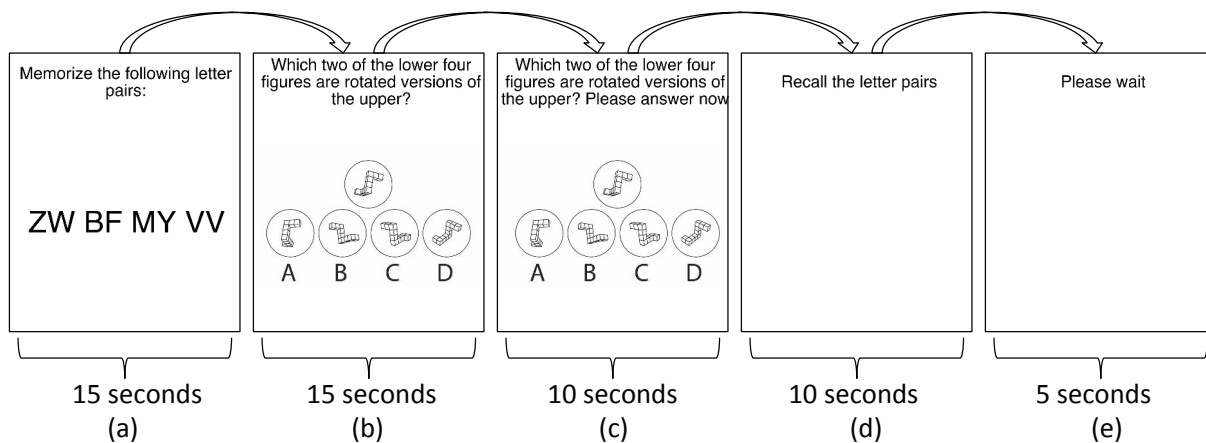


Figure 2: Examples of the five cards shown to participant in one trial. Each trial comprises two tasks or tests: a letter recall test and a mental rotation test

	Self-Avatar		No Self-Avatar	
	Allowed	Not allowed	Allowed	Not allowed
Letter Recall % correct (Mean)	.72	.44	.51	.49
Letter % correct (SD)	.15	.15	.09	.09
Spatial % correct (Mean)	.57	.57	.59	.62
Spatial % correct (SD)	.21	.11	.18	.15
Use gesture occurrences	200/200		33/200	
User movement metres (Mean)	19.88	6.57	6.5	5.45
User movement metres (SD)	10.81	5.86	7.58	7.02

Table 1: Summary of overall experimental results

4.1 Task Performance

4.1.1 Letter Recall

Figure 4 shows the mean proportion of correctly remembered letter pairs in each of the four experimental conditions. Participants in the self-avatar condition remembered 28% more letter pairs when gesturing than when not gesturing. In contrast, for participants in the without a self-avatar condition the difference between the two gesture conditions was not significant. Letter recall scores were normally distributed as assessed by a Shapiro-Wilk test ($p > .05$).

A mixed ANOVA was conducted on the mean proportion of correctly remembered letter pairs, with self-avatar as a between-subjects factor and gesture as a within-subjects factor. Firstly, the main effect of self-avatar was significant, $F(1,38) = 5.88, p = .02$, indicating that the mean score was significantly higher when participants had a self-avatar. Secondly, the gesture factor only has two levels, therefore the assumption of sphericity was not an issue. The mean score of the gesture allowed condition was significantly higher than that of the gesture is not allowed condition, $F(1,38) = 50.922, p < .001$. Thirdly, the self-avatar \times gesture interaction was significant, $F(1,38) = 40.617, p < .001$, indicating that the difference in mean score due to gesture (is or is not allowed) was present in the participants with a self-avatar condition but not the without a self-avatar condition. We note with reference to Figure 4 that the main effects appear to be driven by the interaction.

4.1.2 Mental Rotation

Figure 5 shows the mean proportion of correctly solved spatial tasks in each of the four experiment conditions. Participants' performances did not change substantially across these conditions.

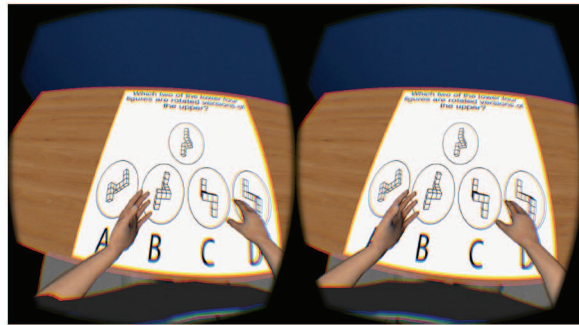
Mental rotation scores were normally distributed as assessed by a Shapiro-Wilk test ($p > .05$).

A mixed ANOVA was conducted on the mean proportion of correctly solved spatial task, with self-avatar as a between-subjects factor and gesture as a within-subjects factor. Results revealed that the main effect of self-avatar, the main effect of gesture and the self-avatar \times gesture interaction were not significant, with $F(1,38) = .676, p > .05$, $F(1,38) = .629, p > .05$, and $F(1,38) = .482, p > .05$, respectively. This indicated that the mean scores were not significantly different across all four experimental conditions.

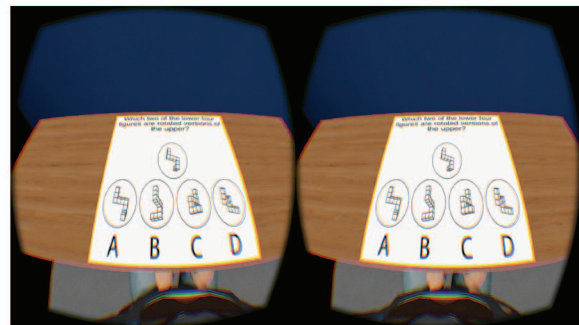
4.2 Gesture and Hand Movement

We observed participants' hand movements in the gesture allowed conditions during the experiment. There are a total of 200 potential occurrences of gestures, 10 for each of the 20 participants in each of the self-avatar and without a self-avatar conditions. For the participants with a self-avatar, all participants used gesture in all trials (thus 200 out of 200 occurrences of gestures, see Table 1). However, for participants without a self-avatar, only 3 out of 20 participants used gestures consistently (10 occurrences each). We further note one participant who only used gesture for one trial and another who used gestures in two trials. Thus the total number of occurrences of gestures in the no self-avatar condition was 33 out of 200.

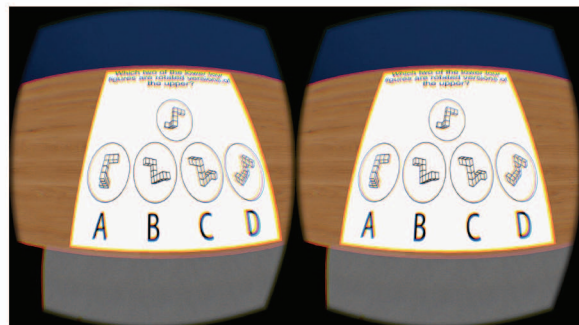
Figure 6 shows the mean for left hand, right hand and total hand movements in four experimental conditions. Note this represents the actual movement of the user's hands, not the movement of the self-avatar's hands. The raw distance measures were not normally distributed as assessed by Shapiro-Wilk test ($p > .05$). Because of the skew characteristics, a square root transformation was applied,



(a) Self-Avatar with gesture allowed



(b) Self-Avatar with gesture not allowed



(c) No self-avatars with and without gesture allowed

Figure 3: First person stereo views of the different experimental conditions

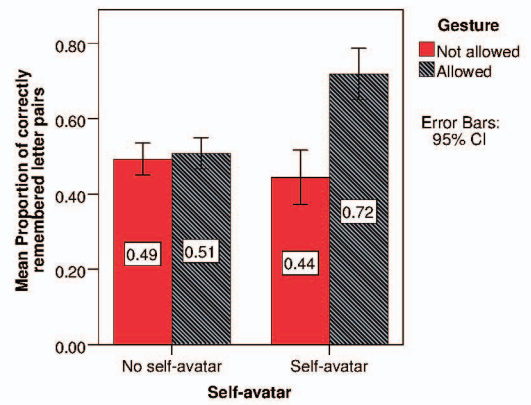


Figure 4: Proportion of correctly remembered letter pairs

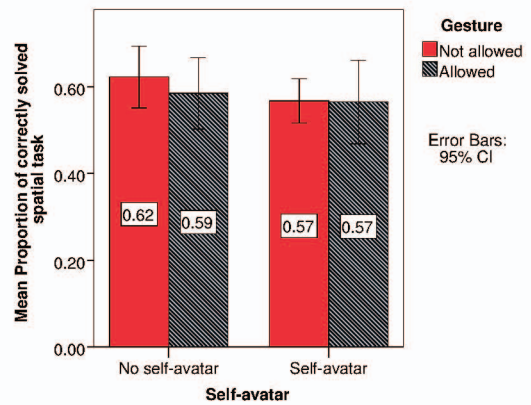


Figure 5: Proportion of correctly solved spatial tasks

in which case the data were normally distributed as assessed by a Shapiro-Wilk test ($p > .05$) with the exception of the no-self avatar, gesture not allowed condition ($p = .013$). However, the data were normally distributed as assessed by the Kolmogorov-Smirnov test ($p > .05$). We believe that the analysis is robust for several reasons; visual inspection of the apparently large effect in Figure 6; the overall robustness of the mixed ANOVA method; the strong significance of the mixed ANOVA; the observational data above; the main hypotheses not overly relying on the specific condition that was not normal; and finally to expand the previous point we ran separate t-test analyses of pairs of conditions without the no-self avatar gesture not allowed data condition.

A mixed ANOVA was conducted on the square root of the mean distance of total hand movement, with self-avatars as a between-subjects factor and gesture as a within-subjects factor. Firstly, the main effect of self-avatars was significant, $F(1, 38) = 13.984, p = .001$, indicating the overall mean distance was significantly higher when participants had a self-avatars. Secondly, the overall mean distance of the gesture allowed condition was significantly higher than that of the gesture is not allowed condition, $F(1, 38) = 22.511, p < .001$. Thirdly, the self-avatars \times gesture interaction was significant, $F(1, 38) = 16.737, p < .001$, indicating that the difference in mean distance due to gesture (is or is not allowed) was present in the participants with a self-avatars condition, but not the without a self-avatars condition. We note with reference to Figure 6 that the main effect appears to be driven by the interaction.

We verified the difference for mean distance of total hand movement by performing independent-samples t-test on no self-avatar versus self-avatar gesture allowed conditions and a paired-samples t-test on the gesture not allowed versus gesture allowed for the self-avatar condition. Both were significant ($t(38) = -4.531, p < .001$ and $t(19) = -5.338, p < .001$ respectively). Note that these two analyses do not use the condition (no self-avatar gesture not allowed) which has non-normal data.

5 Discussion

The results on the letter recall task support H1 and H4. The results of the gesturing and movement analysis support H3. H2 and H5 are not supported, but this is not surprising in retrospect. In the experimental design, the purpose of the mental rotation task is to induce cognitive load that interferes with the letter recall. Performance on the mental rotation task was not significantly different across the conditions and there is not a ceiling effect. This suggests that participants were able to do these tasks reasonably well, but in combination with results on letter recall, it suggests that indeed the cognitive load was higher when the participant didn't have a self-avatar or wasn't allowed to gesture.

We explore the difference in hand movement across the conditions in Figure 7. This figure overlays the patterns of movement for all participants in each of the four conditions. In these figures, Z movement is right (+ive) and left with respect to the participant and X movement is away from (+ive) and towards the torso. X and Z define the plane of the table top. The origin is set to an initial head reference point that was used in Unity to calibrate the tracker positions, but the head is then usually tracked around -0.3m on the Z axis when the participant is sat upright. These figures clearly show the differences in behaviour. In the gesture not allowed conditions there are still small movements, but as discussed previously, the total movement is less. We observe that both hands are used in gesturing. This has important implications for system design as some virtual reality systems provide only a single hand tracker. The figures suggest that two-handed tracking will be important.

Figure 8 shows the smoothed motion of the left and right hand over the two within-subject conditions that participant id 8 undertook. This participant did the self-avatar gesture allowed and self-avatar gesture not allowed conditions. It can be clearly seen that this participant gestured frequently with their right hand when gesture was allowed and not at all with either hand when gesture was not allowed. A similar figure for the Y direction would show the hands mostly remaining on the table during gesture allowed condition, but fixed on the knees for the gesture not allowed condition.

In designing a controlled experiment we had to make several choices about the situation of the user, the tracking and other hardware system choices, the representation of the self-avatar for the user and the tasks provided. All of these could affect the results and are interesting avenues to explore. Each of the male and female avatars we used was a static size. Although no-one commented on this, this would mean that the distance between the eyes and the chest and the articulation of the arm could be subtly wrong. The hand was tracked, but was always shown in a fixed open pose. Any or all of additional calibration, matching the body shape of the user and tracking the body and fingers more accurately could change the results. We believe that the tendency would be to strengthen the results as the gestures would be conveyed more accurately. We also refer back to the work on embodiment in VR (see Section 2.1), where variations of the rubber hand illusion have been shown for avatars that do not match the user's body (e.g. [42]). Further, Banakou et al. have shown virtual reality embodiment within child-sized avatars [1].

Another issue to investigate is the self-avatar's pose. In our conditions, if the self-avatar was visible it was always tracked. In the gesture not-allowed condition the participant was instructed to keep their hands still. It is interesting to ask what would happen if the avatar was simply drawn as a static model. This would mirror some common conditions of virtual reality applications where hand tracking is not available. Given that when there was no avatar but gestures were allowed, relatively few gestures were made (see Section 4.2) and some participants gestured at the start, but then stopped, it is interesting to ask whether seeing static hands and arms would make participants more or less likely to gesture.

We note two issues related to the instructions for the participants in the gesture not-allowed condition. We asked participants in the gesture not-allowed conditions to put their hands on their knees. This is similar to the previous work [12, 5]. It was done so that the self-avatar would be well out of the way of the task materials. However, if the participant placed their hands on the table their performance might be different. Perhaps they would find it harder to keep their hands still or perhaps the visual presence of the arms would affect recall. We do not expect that this would be a large effect, because in our study although the hands are on the knees, they are still visible. It is related to a fair objection to the protocol that can be levelled at this paper and the prior work: that there may be a cognitive overhead of keeping the hands still. We note that this only affects one of the main effects in the study. The self-avatar gesture allowed versus self-avatar gesture not-allowed is most similar to the previous work. Simply being able to create a similar effect in virtual reality is interesting.

A final issue is the difficulty of the task. We might expect that for easier and harder tasks the impact of the gestures and thus the self-avatar would be more or less. In Section 2.2.2 we noted the work of Hostetter et al. [15]. In their study participants performed more complex gestures when explaining more complex problems. This might mean that with harder problems requiring more complex gestures, a self-avatar would be less effective. Similarly for simple tasks the self-avatar might not be necessary. This is an important route for future work in order to demonstrate the effectiveness of immersive virtual reality for applications that might involve a range of task difficulties such as training scenarios. Finally, although we found no impact of body and gesture on the mental rotation task performance, a harder task might solicit more gesture and thus we might expect performance on that task to degrade.

6 Conclusion

The experiment reported in this paper shows that the use of an active self-avatar can enhance a user's ability to perform certain cognitive tasks. Specifically, we demonstrated that participants with an active self-avatar were better at recall of letter pairs when undertaking mental rotation tasks than participants with a self-avatar but not allowed to move the hands or participants without a self-avatar whether or not they were allowed to move their hands. The results extend previous work that looked at the role of a self-avatar in tasks where a self-avatar provides more direct information such as manipulation or communication.

The results fit with previous works in virtual reality and embodied interaction that have argued that the body and gesture are important parts of our cognitive processes. Prior work has shown that the inability to move one's body during explanation tasks impairs performance (e.g. [12] and see Section 2.2.2). We have shown that a similar result still holds in immersive virtual reality, but also that the self-avatar is necessary in order to support performance on the specific tasks we chose. Thus our results indicate that a tracked self-avatar with sensorimotor contingencies appears to reduce cognitive load on certain tasks.

A lot of work needs to be done on the importance of the impact of the self-avatar. Obvious questions include the impact of the visual



Figure 6: Mean hand distance travelled

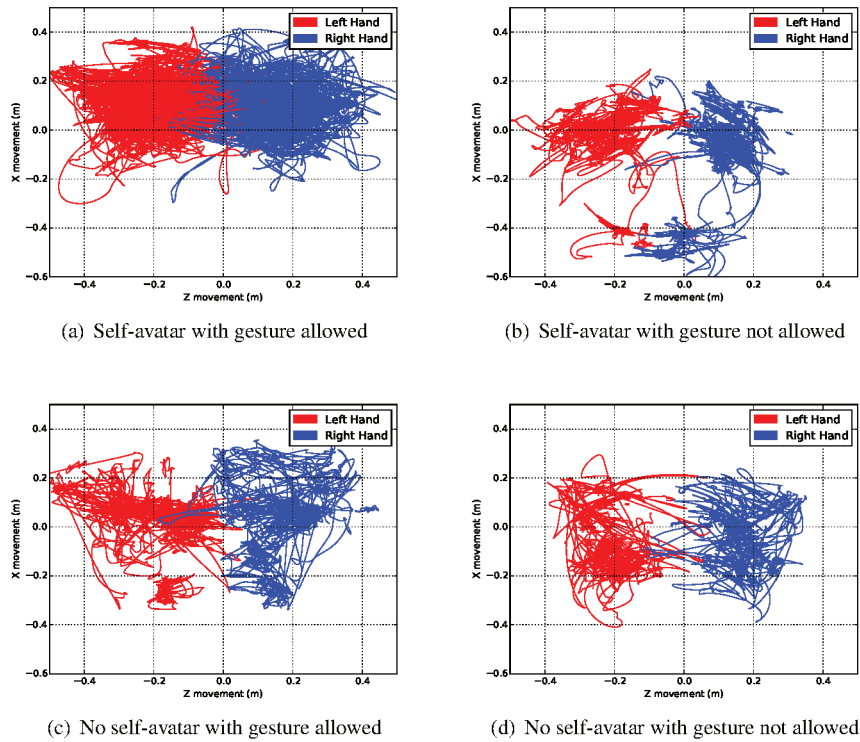
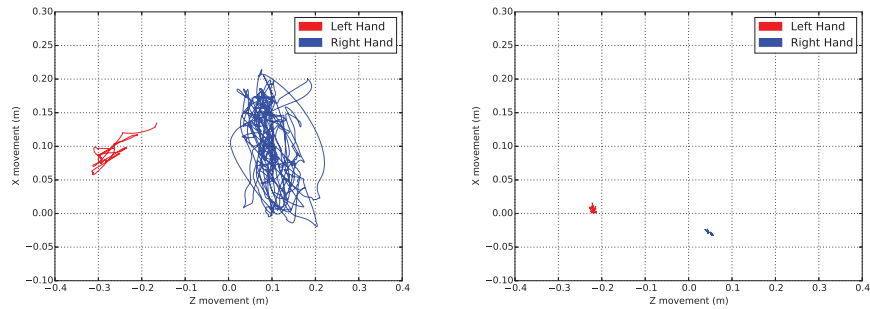


Figure 7: Plots of the movement of the hands of all participants in the four conditions



(a) Participant id 8, self-avatar with gesture allowed (b) Participant id 8, self-avatar with no gesture allowed

Figure 8: Plots of the movement of the hands of one participant with a self-avatar, when gesture is or is not allowed

quality of a self-avatar, the necessary accuracy of the tracking, hand gestures to accompany hand movements, latency of movement and visual quality of the display. It would be interesting to compare performance inside the HMD against doing the task with paper in the physical world. However, the focus of the presented experiment has been on informing the design of the most effective immersive virtual reality systems, rather than demonstrating that performance of specific cognitive tasks in immersive virtual reality systems can approach that of the real world.

To conclude, our results have important implications for the design of virtual reality systems: they suggest that a self-avatar is important, not just for direct manipulation tasks but also to reduce the cognitive overhead of performing a broader class of tasks that involve cognitive processing.

REFERENCES

- [1] D. Banakou, R. Groten, and M. Slater. Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences*, 110(31):12846–12851, 2013.
- [2] L. W. Barsalou. Grounded cognition. *Annual Review of Psychology*, 59(1):617–645, 2008. PMID: 17705682.
- [3] F. Biocca. Connected to my avatar. In *Social Computing and Social Media*, pages 421–429. Springer, 2014.
- [4] M. Botvinick, J. Cohen, et al. Rubber hands’ feel’touch that eyes see. *Nature*, 391(6669):756–756, 1998.
- [5] M. Chu and S. Kita. The nature of gestures’ beneficial role in spatial problem solving. *Journal of Experimental Psychology: General*, 140(1):102, 2011.
- [6] A. Clark. *Supersizing the mind: Embodiment, action, and cognitive extension: Embodiment, action, and cognitive extension*. Oxford University Press, 2008.
- [7] T. J. Dodds, B. J. Mohler, and H. H. Bühlhoff. Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments. *PLoS one*, 6(10):e25759, 2011.
- [8] P. Dourish. *Where the action is: the foundations of embodied interaction*. MIT press, 2001.
- [9] M. Draper. Exploring the influence of a virtual body on spatial awareness. Master’s thesis, University of Washington, Seattle, Washington, 1995.
- [10] DrashVR LLC. Titans of Space. <http://www.titansofspacevr.com/>, 2015. Accessed 6-December-2015.
- [11] S. Gallagher and R. Lindgren. Enactive metaphors: learning through full-body engagement. *Educational Psychology Review*, 27(3):391–404, 2015.
- [12] S. Goldin-Meadow, H. Nusbaum, S. D. Kelly, and S. Wagner. Explaining math: Gesturing lightens the load. *Psychological Science*, 12(6):516–522, 2001.
- [13] M. Gonzalez-Franco, T. Peck, A. Rodriguez-Fornells, and M. Slater. A threat to a virtual hand elicits motor cortex activation. *Experimental Brain Research*, 232(3):875–887, 2014.
- [14] A. B. Hostetter and M. W. Alibali. Visible embodiment: Gestures as simulated action. *Psychonomic bulletin & review*, 15(3):495–514, 2008.
- [15] A. B. Hostetter, M. W. Alibali, and S. Kita. I see it in my hands eye: Representational gestures reflect conceptual demands. *Language and Cognitive Processes*, 22(3):313–336, 2007.
- [16] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-based interaction: a framework for post-wimp interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 201–210. ACM, 2008.
- [17] K. Kilteni, R. Groten, and M. Slater. The sense of embodiment in virtual reality. *Presence: Teleoperators and Virtual Environments*, 21(4):373–387, 2012.
- [18] K. Kilteni, A. Maselli, K. Koerding, and M. Slater. Over my fake body: body ownership illusions for studying the multisensory basis of own-body perception. *Frontiers in Human Neuroscience*, 9(141), 2015.
- [19] D. Kirsh. Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 20(1):3, 2013.
- [20] S. R. Klemmer, B. Hartmann, and L. Takayama. How bodies matter: five themes for interaction design. In *Proceedings of the 6th conference on Designing Interactive systems*, pages 140–149. ACM, 2006.
- [21] Q. Lin, J. J. Rieser, and B. Bodenheimer. Stepping off a ledge in an hmd-based immersive virtual environment. In *Proceedings of the ACM Symposium on Applied Perception, SAP ’13*, pages 107–110, New York, NY, USA, 2013. ACM.
- [22] N. Magnenat-Thalmann and D. Thalmann. *Handbook of virtual humans*. John Wiley & Sons, 2005.
- [23] E. A. McManus, B. Bodenheimer, S. Streuber, S. de la Rosa, H. H. Bühlhoff, and B. J. Mohler. The influence of avatar (self and character) animations on distance estimation, object interaction and locomotion in immersive virtual environments. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization, APGV ’11*, pages 37–44, New York, NY, USA, 2011. ACM.
- [24] B. J. Mohler, H. H. Bühlhoff, W. B. Thompson, and S. H. Creem-Regehr. A full-body avatar improves egocentric distance judgments in an immersive virtual environment. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization*, page 194. ACM, 2008.
- [25] B. J. Mohler, S. H. Creem-Regehr, W. B. Thompson, and H. H. Bühlhoff. The effect of viewing a self-avatar on distance judgments in an hmd-based virtual environment. *Presence: Teleoper. Virtual Environ.*, 19(3):230–242, June 2010.
- [26] A. Noë. *Action in perception*. MIT press, 2004.

- [27] Oculus. Toybox Demo for Oculus Touch. <https://youtu.be/iFEMiyGMa58>, 2015. Accessed 6-December-2015.
- [28] G. Padrao, M. Gonzalez-Franco, M. V. Sanchez-Vives, M. Slater, and A. Rodriguez-Fornells. Violating body movement semantics: Neural signatures of self-generated and external-generated errors. *NeuroImage*, 124, Part A:147 – 156, 2016.
- [29] B. Ries, V. Interrante, M. Kaeding, and L. Anderson. The effect of self-embodiment on distance perception in immersive virtual environments. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, VRST '08, pages 167–170, New York, NY, USA, 2008. ACM.
- [30] RootMotion. FinalIK. <http://root-motion.com/>, 2015. Accessed 6-December-2015.
- [31] M. V. Sanchez-Vives, B. Spanlang, A. Frisoli, M. Bergamasco, and M. Slater. Virtual hand illusion induced by visuomotor correlations. *PLoS ONE*, 5(4):e10381, 04 2010.
- [32] U. Schultze. Embodiment and presence in virtual worlds: a review. *Journal of Information Technology*, 25(4):434–449, 2010.
- [33] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.
- [34] M. Slater and A. Steed. A virtual presence counter. *Presence*, 9(5):413–434, 2000.
- [35] M. Slater, M. Usoh, and A. Steed. Depth of presence in virtual environments. *Presence*, 3(2):130–144, 1994.
- [36] M. Slater, M. Usoh, and A. Steed. Taking steps: The influence of a walking technique on presence in virtual reality. *ACM Trans. Comput.-Hum. Interact.*, 2(3):201–219, Sept. 1995.
- [37] M. Slater and S. Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and virtual environments*, 6(6):603–616, 1997.
- [38] S. Streuber, S. de la Rosa, L. Trutoiu, H. H. Bühlhoff, and B. J. Mohler. Does brief exposure to a self-avatar affect common human behaviors in immersive virtual environments? *Eurographics 2009*, pages 1–4, 2009.
- [39] M. Usoh, K. Arthur, M. C. Whitton, R. Bastos, A. Steed, M. Slater, and F. P. Brooks, Jr. Walking >walking-in-place >flying, in virtual environments. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '99, pages 359–364, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.
- [40] Ustwo Games. Lands End. <http://www.landsendgame.com/>, 2015. Accessed 6-December-2015.
- [41] A. S. Won, J. Bailenson, J. Lee, and J. Lanier. Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication*, 2015.
- [42] Y. Yuan and A. Steed. Is the rubber hand illusion induced by immersive virtual reality? In *Virtual Reality Conference (VR), 2010 IEEE*, pages 95–102. IEEE, 2010.