# Immersion Factors affecting Perception and Behaviour in a Virtual Reality Power Wheelchair Simulator

Abdulaziz Alshaer
University of Otago
Dunedin, New Zealand

Holger Regenbrecht
University of Otago
Dunedin, New Zealand

David O'Hare
University of Otago
Dunedin, New Zealand

Key words: human factors, sense of presence, virtual environments, navigation, simulation, affordance

**Corresponding author:**
Holger Regenbrecht
University of Otago
Dunedin, New Zealand.
Phone: (64)3-479-8322
Email: holger.regenbrecht@otago.ac.nz
www.hci.otago.ac.nz

# Abstract

Virtual Reality based driving simulators are increasingly used to train and assess users' abilities to operate vehicles in a controlled and safe way. For the development of those simulators it is important to identify and evaluate design factors affecting perception, behaviour, and driving performance. In an exemplary power wheelchair simulator setting we identified the three immersion factors display type (head-mounted display v monitor), ability to freely change the field of view (FOV), and the visualisation of the user's avatar as potentially affecting perception and behaviour. In a study with 72 participants we found all three factors affected the participants' sense of presence in the virtual environment. In particular the display type significantly affected both perceptual and behavioural measures whereas FOV only affected behavioural measures. Our findings could guide future Virtual Reality simulator designers to evoke targeted user behaviours and perceptions.

## 1. Introduction

Previous research has shown that virtual-reality-based applications are in increasingly wide-spread use for such tasks as driving. Driving simulators are used for, but not limited to, assessment, learning, rehabilitation, and entertainment [1]. One representative and exemplary class of such driving simulators are power (electrical) wheelchair (PWC) simulators [2]. They provide a risk-free environment that would allow users to drive efficiently in order to evaluate their ability at PWC driving. Despite their potential, existing PWC simulators have been found to be less usable than demanded and expected [3]. While previous research could show positive transfer effects from the virtual simulator to the real world, the users experienced difficulties in operating the simulator attributed to immersion factors like display characteristics [4, 5]. In addition, the use of Virtual Reality personal computer (PC) technology, i.e. interactive 3D desktop computer systems, for PWC simulation seems to be underdeveloped and under-researched. Only one software product is commercially available on the market: WheelSim [6], unfortunately unsuitable for training and assessment [3].

To close this gap in the availability of a suitable research platform for the investigation of design factors for this class of Virtual Reality (VR) simulators we developed a simple, but usable power wheelchair simulator which can be operated with different peripheral devices and can be configured to meet the needs of our research. We conducted initial interviews with professional experts (four occupational therapists) and consulted the appropriate literature, e.g. [3, 7, 8, 9] leading to the identification of system requirements. For instance, the ability of the users to drive accurately depends on how they perceive the scale of the space of the virtual environment (VE), which is a prerequisite for the validity as a training and/or assessment tool.

Another example is that the presentation of a self-avatar (a visual representation of the user's own body or body parts) in VEs in general has been shown to not only increase the sense of presence but also to improve size and distance judgments [10]. There is evidence [11, 12, 13] that a self-avatar could serve as a familiar size cue that provides scaling information and act as a frame of reference in the VE. Sun et al. [14] add that the presence aspect of the user's body can lead to significant effects on performance. In a PWC simulator, the visualisation of the virtual PWC itself would also provide scaling information about the dimensions of the virtual space and act as a usable frame of reference for spatial judgments. Also, another leading question in this investigation was whether a self-avatar would provide additional cues and would serve as a dual reference.

It is well researched that misperceptions of a simulation space can result in erroneous judgments that could alter the user’s behaviour [15, 16]. Therefore, it is important not only to measure users’ perception but also to differentiate behaviour. However, how to best measure the accuracy of space perception in VEs remains a difficult question [17]. Research in the past used verbal estimation, perceptually direct actions, and imagined action to estimate perceived distance in a VE [18]. In verbal estimation, perceived distance assessed through familiar units, such as meters. In perceptually direct actions, subjects would preform an action, such as blind walking or imagined action [18] [17], which only provide rather indirect measures.

In 1979, Gibson [19] introduced the concept of “affordance” which emphasizes the relationship between objects and their observers. For instance, a gap can afford passage if it is wide enough for the user. Many studies, since then, have demonstrated the practicality and usefulness of using affordance theory to measure user perceptions in VE [20] [21]. According to Geuss et al. [17], “affordance judgments may be especially useful as a perceptual measure of size in graphic displays because they require the user to see the space in terms of their own ability to act and therefore may be considered more task-relevant”. In our research presented here we use the affordance of "pass-ability" though wall-openings to measure perceived spatial size and distance.

Our research addressed the following questions: How accurately can PWC users make the right decisions when navigating a virtual environment? How do they perceive a particular gap as passable? How do different immersion factors (display type, field of view, and self-avatar presence) influence their behaviour, perception and sense of presence? Behaviour was measured through embedded actions (implicit performance); perception through self-report of the perceived size/distance in the VE (explicit judgment); and sense of presence through a standard questionnaire. The manipulated factors were self-avatar presence versus no self-avatar presence); a static field of view (FOV) versus a changeable FOV; and monitor display versus head-mounted display (HMD). This yielded a 2 (avatar presence) X 2 (FOV) X 2 (display type) mixed-design experiment. It is important to emphasise that the methodology involved both participants' self-report indication (whether a particular action can or cannot be performed) and behavioural decision-making (participant actually passed through or went around a particular gap).

We hypothesised that: 1) users’ implicit performance, explicit judgments, and sense of presence, would be better with the more immersive HMD regardless of field of view change or avatar presence; 2) users’ implicit performance, explicit judgments, and sense of presence would be better with the changeable field of view regardless of display type or avatar presence; 3) users’ implicit performance, explicit judgments, and sense of presence would be better with the presence of a self-avatar regardless of the display type, and 4) users’ implicit performance, explicit judgments, and sense of presence would be better with the changeable field of view, HMD display, and self-avatar.

## 3. Method

### 3.1 Participants

A pilot study with five participants was conducted to provide a formative evaluation of the procedures and instruments. This was followed by the actual experiment where 72 subjects participated. There were 46 males and 26 females with a mean age of 21.9 years (*SD* = 4.68, age range = 18 - 47), including students from the departments of Psychology and Information Science, of the University of Otago. Participants from the Psychology department were recruited via an online system and students were rewarded with class credits whereas participants from Information Science were recruited via personal connections and classroom announcements, and were rewarded with chocolate bars. All participants had normal or corrected-to-normal vision. Institutional ethical approvals were obtained from both departments.

### 3.2 Apparatus

#### 3.2.1 Virtual environment

All 3D models were built by using Google *SketchUp*. The virtual PWC, including the virtual joystick, was modelled on real PWC dimensions with an average width of 68cm. The VE used in this experiment was a high-fidelity 3D model of an abstracted (low distraction) hallway. The hallway consisted of walls, doorframes, and sets of two poles designed to represent gaps of varying widths throughout the hallway. To avoid participants' distractions and to remove cues to size and distance provided by familiar objects no furnishings or decorations were added. The hallway was wide enough throughout so that subjects could freely and easily navigate the environment (Figure 1). The virtual self-avatar was produced by "*MakeHuman*", a 3D character-building application.

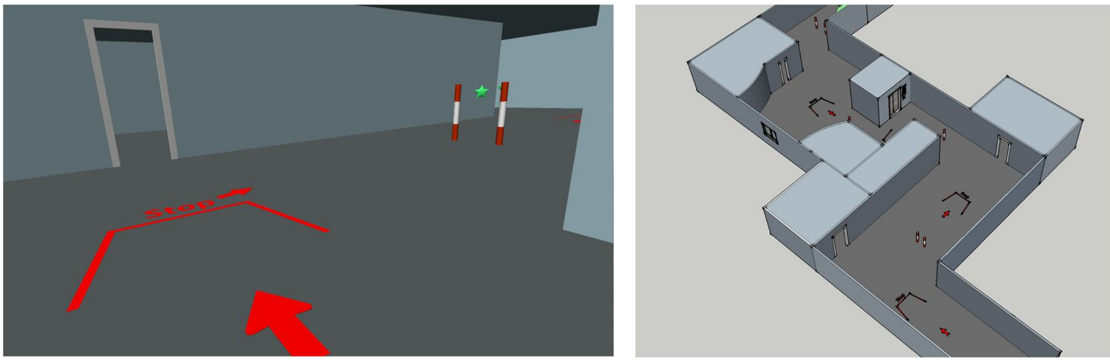


Figure 1: PWC simulator (hallway)

The hallway consisted of four doorframes of different widths distributed over the hallway. Similarly, there were four gaps of different widths between two poles, (see Figure 1) spread along the hallway. The doorframe and gap widths were differentiated based on the minimum clear gap width that the PWC could pass through, which is 76cm [22] . Two doorframes/gaps were passable (easy to pass = 76cm, hard to pass = 72cm) and two were not passable (hard to judge = 64cm, easy to judge = 60cm). Figure 2 shows the widths of all four gaps/doorframes and how they were associated with the PWC width.

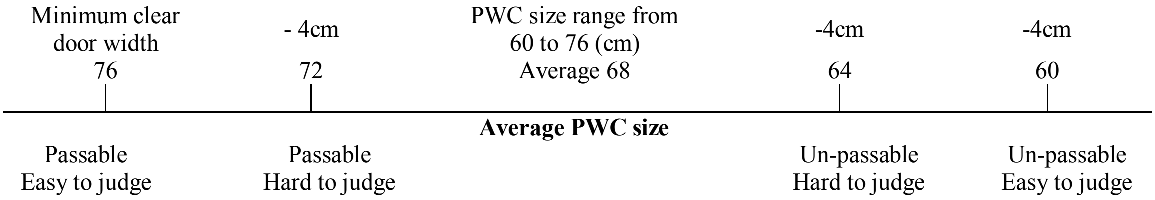


Figure 2: doorframes/gaps widths differences

### 3.2.2 PWC Simulator

The hardware apparatus used in this experiment involved a monitor, HMD-Oculus Rift DK2, a laptop, and joystick. The monitor size was 21.50" with a resolution of 1920x1080 pixels. The resolution of the HMD was 960x1080 pixels spread over two eye point displays with a 100° field of view. The Oculus Rift supported head position and orientation tracking. Head movements were tracked by a three-axis orientation sensing system integrated into the Oculus headset and used to continuously update the simulated viewpoint. A real time positional tracker attached to the top of the monitor was used to track participants' position. The system latency (delay between participant movement and updates in the HMD) was less than 20 milliseconds. For both monitor and HMD, the aspect ratio was 19:6. A gaming joystick (Logitech Attack3) was used to drive the virtual PWC and also to control the FOV in the monitor condition. A 17" Alienware high-end graphics laptop was used to run the simulator in both monitor and HMD conditions. The Unity 3D game engine was used to assimilate tracking and rendering. Two versions of the simulator were built with Unity 3D: one for the monitor display and the other for the Oculus Rift, due to the specific configuration required by the Oculus. A virtual hand and a virtual joystick were displayed in the environment to represent participants' hand movements. Figure 3 shows all the hardware components used in the experiment.


Figure 3: Experiment components including monitor, joystick, HMD, laptop, and HMD tracker placed at the top of the screen

### 3.2.3 Data recording

Simulator data were logged in a *txt* file on their occurrence. Data recorded were participants' name, condition name, condition order, attempted gap width, number of correct attempts (hits), number of incorrect attempts (false alarms), number of collisions with the poles, and time spent to complete the task.

### 3.3 Driving task

Participants were tasked with following directions (red arrows on the floor), stopping at stop signs (where they had to judge the pass-ability of doorframes), avoiding collisions with poles, and collecting stars (placed in the middle of each set of poles). The stars were used as an incentive to encourage participants to attempt to pass through any of the gaps they judged to be passable. To preserve as much realism as possible participants were not specifically told about the pass-ability of the doorframes or gaps. Moreover, the stars pulsed and rotated to prevent them from being used as a frame of reference to judge the gap width. The stars were placed at PWC user's chest height so that the participant had to drive completely though the gap to collect them. Once collected, the system provided visual and sound effects, signalling success. Figure 4 shows participants performing the task.

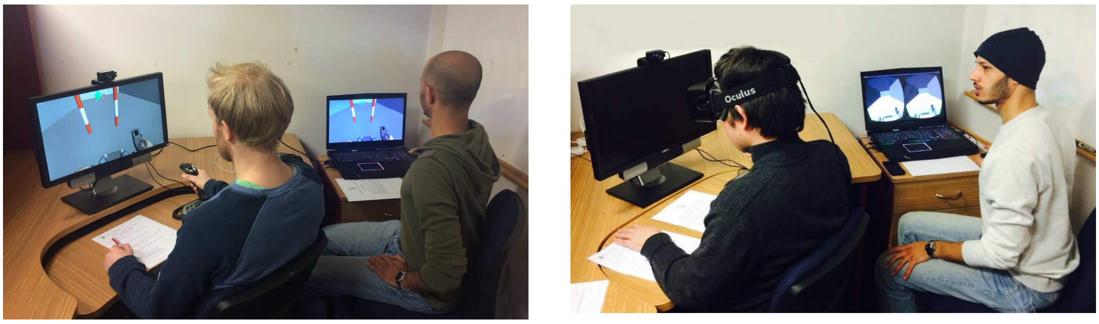


Figure 4: Monitor condition (on the left) and HMD condition (on the right)

### 3.4 Measures

#### 3.4.1 Primary explicit and implicit measures

Users' perception (explicit judgment measure of doorframes pass-ability) and embedded behaviour (implicit measure of gap pass-ability) were mainly based on participants' decision making which takes place in the presence of uncertainty. This was assessed in signal detection terms. A hit occurred when the participants' explicitly said "Yes" to passable doorframes or attempted passable gaps. False alarms, on the other hand, occurred when participants explicitly said "Yes" to un-passable doorframes or attempted un-passable gaps. Correct Rejections involve judging the un-passable doorframes as too small or avoiding trying to pass through the gaps that were impassable. Misses involved incorrectly judging passable doorframes as un-passable or incorrectly avoiding going through passable gaps.

The number of hits and false alarms alone do not measure the diagnostic accuracy of response [23]. Optimal performance occurs when a participant indicates that a signal is present when the signal is actually present and absent when it is actually absent. Szalma et al. [23], proposed that overall performance could be best captured by the measures of Positive Predictive Power (PPP) and Negative Predictive Power (NPP). PPP is the proportion of "Yes" responses that are correct and was computed using the formula H/(H+FA), in which H is the number of correctly detected signals and FA the number of false alarms. A perfectly accurate participant would achieve a PPP score of 1. A score of 0 would indicate no correct detection or a complete inability to correctly discriminate between passable and un-passable gaps. NPP is the proportion of "No" responses that are correct and computed using the formula CR/(CR+M), in which CR is the number of correct rejections and M is the number of missed signals. Similarly to PPP, a participant who correctly rejected all non-signals and had no misses would yield a NPP score of 1.

### 3.4.2 Sense of presence and simulator sickness

The sense of presence was measured by a standard questionnaire, the Igroup Presence Questionnaire (IPQ) [24]. We were only interested in the general sense of presence and realism of the IPQ questionnaire, therefore, related questions were measured; each question took the form of a seven-point scale after each condition. Simulator sickness questions were part of the sense of presence questionnaire. We adapted five questions, each with a four-point scale from “none” to “severe”. This allowed for the measurement of the respondent’s physical well-being after each condition in group B (HMD group).

### 3.4.3 Post-driving questionnaire

This was designed to obtain subjective ratings of the simulator features (FOV and self-avatar). Participants were asked to rate the ease and comfort of each feature on a seven-point Likert-scale e.g., “ Do you think the self-avatar/controllability of the field of view made it easier to judge door/gaps in the virtual environment” (1 = Harder, 7 = Easier); “When the self-avatar/ field of view static was not there, did you feel more or less comfortable” (1 = Less comfortable, 7 = More comfortable). The last question of the post-driving questionnaire required participants to indicate which condition they preferred. Participant had to choose one of the four conditions generated by the combination of FOV levels and self-avatar levels.

## 3.5 Design

The design of this experiment was 2 (display type) * 2 (FOV) * 2 (avatar presence) mixed factorial in which display type was a between-subjects variable and FOV and self-avatar presence were within-subjects variables, yielding eight treatment conditions for both groups. In the within-subject variables, each variable consisted of low level of immersion (represented by X) and high level of immersion (represented by the first letter of each those levels). FOV - either static FOV (X) or changeable FOV, being able to look around (C). Self-avatar - either not present (X) or present (A). Table 1 depicts the mixed-subject factorial design. Measured variables included implicit performance, explicit judgments, sense of presence, opinions, and preference for the conditions.

Table 1: Mixed-subject factorial design

<table>
<tr><td></td><td></td><td colspan="2">Group A<br>Monitor</td><td colspan="2">Group B<br>HMD</td></tr>
<tr><td></td><td></td><td colspan="2">Self-avatar</td><td colspan="2">Self-avatar</td></tr>
<tr><td></td><td></td><td>No<br>X</td><td>Yes<br>A</td><td>No<br>X</td><td>Yes<br>A</td></tr>
<tr><td rowspan="2">FOV</td><td>No<br>X</td><td>X-X</td><td>X-A</td><td>X-X</td><td>X-A</td></tr>
<tr><td>Yes<br>C</td><td>C-X</td><td>C-A</td><td>C-X</td><td>C-A</td></tr>
<tr><td></td><td></td><td colspan="2">Subjects 1-36</td><td colspan="2">Subject 37-72</td></tr>
</table>

## 3.6 Counterbalancing

The mixed design was chosen to reduce the learning effect that would result from repeating the task eight times. To further control for any possible learning effects, 1) subjects were randomized in counterbalanced order, and 2) the combinations of the doorframes and gaps widths were also randomized across all four conditions in counterbalanced order. In addition, although participants repeated the tasks four times, they were generally unaware of the repetition. The participants followed one layout on a return path, which created a balanced set of comparable paths that the user could traverse without interruption. The absence of textures, decorations, furnishings etc. made it difficult for participants to predict what was coming next, e.g. it was hard for them to know which direction to travel next as the right turn became left when driving in the reverse direction. Moreover, the randomization of the gaps and doorframes across all condition made it impossible for participants to memorise which doorframes and/or gaps were passable and which were not.

### 3.7 Procedure

Upon arrival, participants read the information sheet and signed the consent form. This was followed by filling out a demographics questionnaire. Confounding variables such as prior experience with the joystick and/or HMD were controlled: Participants were asked questions prior to the experiment about their experience with the joystick and HMD. These questions determined how much information and training was needed before start. Participants used the same actual experiment setup for training, yet different versions of the simulator (monitor and HMD) depends on which group the participant was assigned to. The training version had no specific task (no doorframes or gaps were displayed, Figure 5) and used to provide the participants with a basic understanding of how to drive the virtual PWC using the joystick controller. In addition, a simple set of criteria were observed by the experimenter to make sure each participant was confident in using the joystick and HMD. Those criteria were: 1) driving forward/backward and turning right/left, 2) being able to follow the guiding arrows, 3) experience the orientation and position tracking of the HMD for those in the HMD group, and 4) experience the changeability of the FOV in the Monitor group using the joystick (hat switch).

After successful completion of the ‘training phase’, the participants were given the task description. Meanwhile, the experimenter started the actual experiment version. The order of the conditions was randomized beforehand. During each condition, the system automatically stopped participants at stop signs and corrected their position and orientation so that all participants judged doorframes from an exact distance and orientation. The experimenter then asked participants, “Can you pass through the door in front of you?” and recorded their “Yes” or “No” answer on a sheet of paper. After each condition, participants were given the sense of presence questionnaire. After the completion of all four conditions, participants answered the perceived comparison questionnaire. Finally, participants were debriefed and given a chocolate bar. The entire procedure took approximately 20 minutes per participant.

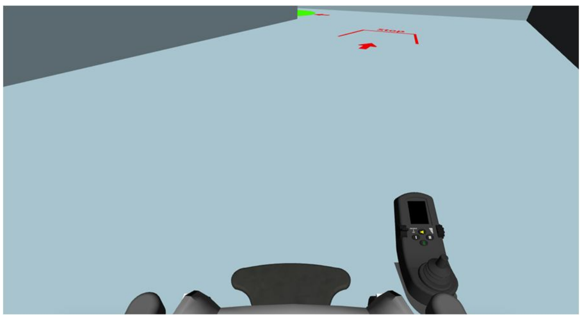

Figure 5: Training version used for demonstration, no gaps or doorframes added

## 4. Results

In this study, two ways of measuring performance in the VE were used: (1) implicit performance, where subjects had to judge pass-ability through embedded behaviour and (2) explicit judgments where subjects' judgments were obtained by self-report indications. In addition, we also measured participants' sense of presence, simulator sickness, opinion, and preference. The design was 2 (avatar presence) * 2 (FOV) * 2 (display type) mixed factorial, ANOVAs were run and the main interaction effects were examined.

## 4.1 Correct detection (Hit)

Implicit performance: The means of correct detection, together with standard deviations are reported in Table 2. The HMD group showed higher means in all conditions with the C-A condition being the highest (M = 2). For implicit performance, ANOVA confirmed significant interaction effects between FOV and display-type, $F(1,70)=4.84$ ,$p < .031$, $\omega 2 = .06$, and between FOV, self-avatar, and display-type, $F(1,70)=7.14$ ,$p < .009$, $\omega 2 = .09$ (Figure 6 shows significant interactions graphs). There was no significant interaction between FOV and self-avatar on users' behaviour. ANOVA also indicated a significant FOV main effect on users' behaviour, $F(1,70)=13.46$ ,$p < .000$, $\omega 2 = .16$, and a significant main effect for display type, $F(1,70)=25.52$ ,$p < .000$, $\omega 2 = .26$. The presence of the self-avatar was not statistically significant.

Explicit judgments: Means and standard deviations are reported in Table 2. An ANOVA of the users' judgments indicated that explicit judgments did not significantly differ across any of the three immersion factors. The interaction between these factors also lacked significance.

Table 2: Correct detection means and standard deviations for implicit and explicit measures

| | | | **Monitor** Self-avatar No X | **Monitor** Self-avatar Yes A | | **HMD** Self-avatar No X | **HMD** Self-avatar Yes A | |
|---|---|---|---|---|---|---|---|---|
| Implicit Performance | **FOV** | No **X** | 1.25 (.69) | 1.56 (.73) | 1.40 | 1.88 (.39) | 1.86 (.42) | 1.87 |
| | | Yes **C** | 1.8 (.40) | 1.67 (.53) | 1.73 | 1.92 (.28) | 2 (0) | 1.96 |
| | | | 1.53 | 1.61 | **1.57** | 1.9 | 1.93 | **1.92** |
| Explicit Judgments | **FOV** | No **X** | 1.28 (.70) | 1.47 (.73) | 1.38 | 1.31 (.71) | 1.56 (.65) | 1.43 |
| | | Yes **C** | 1.44 (.65) | 1.36 (.72) | 1.4 | 1.64 (.59) | 1.67 (.53) | 1.65 |
| | | | 1.36 | 1.42 | **1.39** | 1.47 | 1.61 | **1.54** |

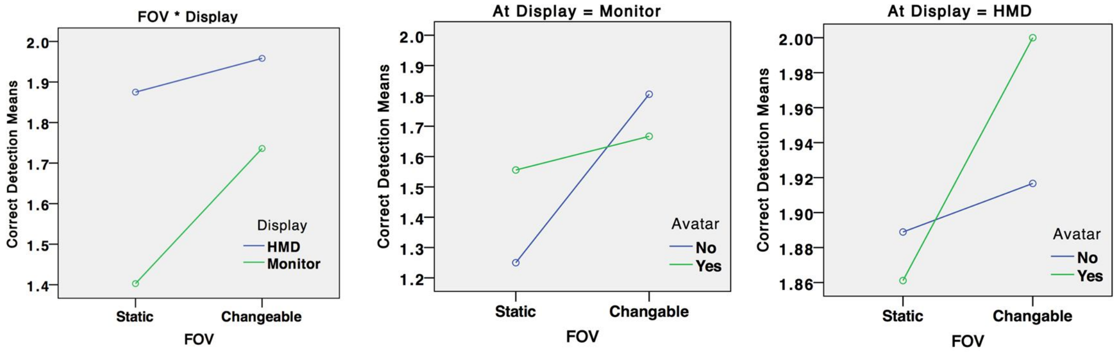


Figure 6: Left: interaction between FOV and Display factors. Centre: interaction between FOV and Avatar factors for Monitor condition. Right: interaction between FOV and Avatar factors for HMD condition.

## 4.2 False Alarm (FA)

Implicit performance: The means and standard deviations of false alarms for both implicit and explicit measures are reported in Table 3. Although means differ slightly between conditions, the two groups' overall scores were very close. However, ANOVA revealed a significant interaction effect between FOV and self-avatar factor, $F(1,70)=7.18$ ,p < .009, ω2 = .09 (Figure 7). There were no significant main effects or other interactions between factors.

Explicit judgments: Participants in the monitor group produced substantially more false alarms (M = .39) than those who used HMD (M = .09). In fact, 91% of all the false alarm scores in the HMD group were zeros. An ANOVA revealed a significant main effect for display type on users' judgments, $F(1,70)=16.17$ ,p < .000, ω2 = .19. The interaction between factors was not statistically significant.

Table 3: False alarm means and standard deviations for implicit and explicit measures

| | | | Monitor | | | HMD | | |
|---|---|---|---|---|---|---|---|---|
| | | | Self-avatar | | | Self-avatar | | |
| | | | No **X** | Yes **A** | | No **X** | Yes **A** | |
| Implicit Performance | FOV | No **X** | .92 (.81) | .97 (.77) | .95 | 1.08 (.73) | .94 (.83) | 1.01 |
| | | Yes **C** | .72 (.70) | 1.06 (.75) | .89 | .67 (.53) | 1.03 (.74) | .85 |
| | | | .82 | 1.02 | **.92** | .88 | .99 | **.93** |
| Explicit Judgments | FOV | No **X** | .53 (.73) | .31 (.62) | .42 | .17 (.44) | .11 (.32) | .14 |
| | | Yes **C** | .39 (.65) | .33 (.63) | .36 | .06 (.23) | .03 (.17) | .04 |
| | | | .46 | .32 | **.39** | 0.12 | .07 | **.09** |

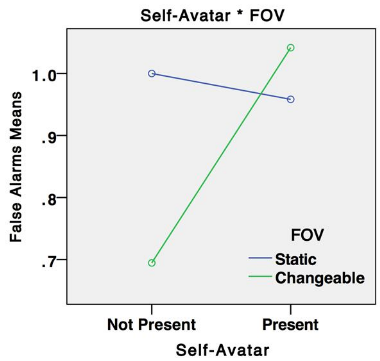


Figure 7: Interaction between FOV and avatar

## 4.3 Diagnostic Measures

### 4.3.1 Inclusion and exclusion criteria

To further evaluate participants' decision-making, PPP and NPP were used. Before analysing PPP and NPP, we set decision criteria at detectors and non-detectors (either participants discriminated between the stimuli or they did not). Non-detectors were participants who tried to go through every gaps regardless if they were passable or not or who avoided all the gaps regardless if they were passable or not. They were eliminated from that condition because the data was non-discriminating by obviously adapting a policy by going through or avoid. They were not providing any data about their ability to discriminate between singles.

In addition, Participants only removed from the condition where they were non-detectors considering that as a separate category for that condition. This is important since non-detectors by definition do not show any sensitivity to perceptual changes in that condition. Table 4 shows different numbers in each condition showing the different number of participants they were removed from that condition because they were not discriminating. The percentages of non-detectors for each condition are also shown in Table 4. The number of excluded (non-detectors) participants from analysis seemed to be reduced whenever participants were able to look around. The presence of the self-avatar appeared to increase the number of non-detectors especially with the monitor group. However, nonparametric tests revealed no statistical difference between FOV and self-avatar levels in each group and no statistical difference between the two groups (display type).

Table 4: Non-detectors number and percentages for implicit and explicit measures

| | | | **Monitor** | | | **HMD** | | |
|---|---|---|---|---|---|---|---|---|
| | | | Self-avatar | | | Self-avatar | | |
| | | | No **X** | Yes **A** | | No **X** | Yes **A** | |
| Implicit Performance | FOV | No **X** | 9 25% | 10 27.9% | 26.5% | 12 33.3% | 10 27.9% | 30.6% |
| | | Yes **C** | 4 11.1% | 10 27.9% | 19.5% | 1 2.9% | 10 27.9% | 15.4% |
| | | | 18.1% | 27.9% | **23%** | 18.1% | 27.9% | **23%** |
| Explicit Judgments | FOV | No **X** | 6 16.7% | 6 16.7% | 16.7% | 4 11.1% | 3 8.33% | 9.7% |
| | | Yes **C** | 2 5.56% | 5 13.9% | 9.7% | 2 5.56% | 1 2.9% | 4.2% |
| | | | 11.1% | 15.3% | **13.2%** | 8.3% | 5.6% | **7%** |

### 4.3.2 Positive predictive power

Implicit performance: The means and standard deviations of PPP for both implicit and explicit measures are reported in Table 5. An ANOVA of the PPP revealed a significant interaction between FOV and avatar F(1,63)=7.22 ,p < .009, ω2 = .10 (Figure 8). None of the other interactions was statistically significant. ANOVA also indicated a significant main effect for FOV F(1,63)=4.85 ,p < .031, ω2 = .07.

Explicit judgments: An ANOVA revealed a significant main effect for display type, F(1,51)=13.7 ,p < .001, ω2 = .21. No significant effects were observed for FOV, self-avatar, or the interaction between factors.

Table 5: Means and standard deviations of PPP for implicit and explicit measures

| | | | **Monitor** | | | **HMD** | | |
|---|---|---|---|---|---|---|---|---|
| | | | Self-avatar | | | Self-avatar | | |
| | | | No **X** | Yes **A** | | No **X** | Yes **A** | |
| Implicit Performance | FOV | No **X** | .64 (.27) | .65 (.24) | .64 | .67 (.18) | .71 (.25) | .69 |
| | | Yes **C** | .78 (.20) | .66 (.21) | .72 | .78 (.18) | .70 (.18) | .74 |
| | | | .71 | .65 | **.67** | .73 | .70 | **.72** |
| Explicit Judgments | FOV | No **X** | .73 (.29) | .87 (.25) | .80 | .88 (.28) | .95 (.12) | .92 |
| | | Yes **C** | .81 (.28) | .84 (.28) | .82 | .99 (.08) | .98 (.09) | .98 |
| | | | .77 | .85 | **.81** | .93 | .97 | **.95** |

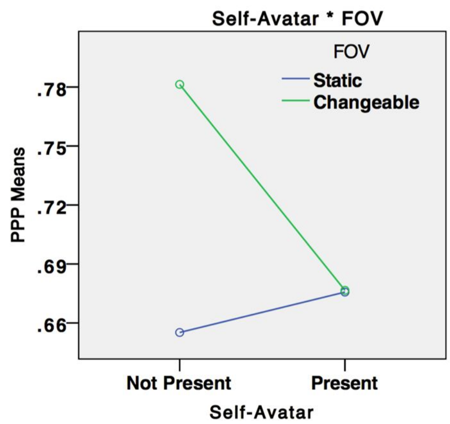


Figure 8: Interaction between FOV and self-avatar factors

4.3.3 Negative predictive power

Implicit performance: The means and standard deviations of NPP for both implicit and explicit measures are reported in Table 6. An ANOVA revealed significant main effects for FOV, F(1,33)=5.405 ,p < .026, ω2 = .14, and display type F(1,33)=26.8 ,p < .000, ω2 = .45. No significant effects were observed for the self-avatar factor, nor for the interaction between factors.

Explicit judgments: An ANOVA revealed a significant main effect for display type, F(1,46)=6.59 ,p < .013, ω2 = .09. No significant effects were observed for FOV, self-avatar, or the interaction between factors.

Table 6: Means and standard deviations of NPP for implicit and explicit measures

| | | | Monitor<br>Self-avatar<br>No **X** | Monitor<br>Self-avatar<br>Yes **A** | | HMD<br>Self-avatar<br>No **X** | HMD<br>Self-avatar<br>Yes **A** | |
|---|---|---|---|---|---|---|---|---|
| Implicit Performance | FOV | No **X** | .57 (.36) | .72 (.38) | .65 | .97 (.13) | .91 (.19) | .94 |
| | | Yes **C** | .86 (.27) | .75 (.33) | .81 | 1 (0) | 1 (0) | 1 |
| | | | .72 | .74 | **.73** | .98 | .95 | **.97** |
| Explicit Judgments | FOV | No **X** | .73 (.24) | .83 (.25) | .78 | .77 (.22) | .86 (.25) | .82 |
| | | Yes **C** | .79 (.26) | .75 (.3) | .77 | .89 (.17) | .9 (.17) | .90 |
| | | | .76 | .79 | **.78** | .83 | .88 | **.85** |

## 4.4 Sense of presence

For sense of presence, we were only interested in the general sense of presence and realism, which are only reported in this study. The means and standard deviations of the general sense of presence and realism are reported in Table 7.

Sense of presence: Participants' general sense of presence was obtained in response to the following question: In the computer-generated world I had a sense of "being there"? The question consisted of seven-point Likert-like item from -3 (not at all) to 3 (very much). An ANOVA indicated significant interaction effects between self-avatar and display type, $F(1,70)=11.88$ ,$p < .001$, $\omega2 = .14$, and between FOV, self-avatar, and display type, $F(1,70)=4.38$ ,$p < .04$, $\omega2 = .06$. Interaction graphs can be seen in Figure 9. Significant main effects were also revealed for all three factors (FOV, self-avatar, and display type) and were highly significant, $F(1,70)=18.32$ ,$p < .000$, $\omega2 = .21$, $F(1,70)=19.63$ ,$p < .000$, $\omega2 = .22$, $F(1,70)=17.78$ ,$p < .000$, $\omega2 = .20$, respectively.

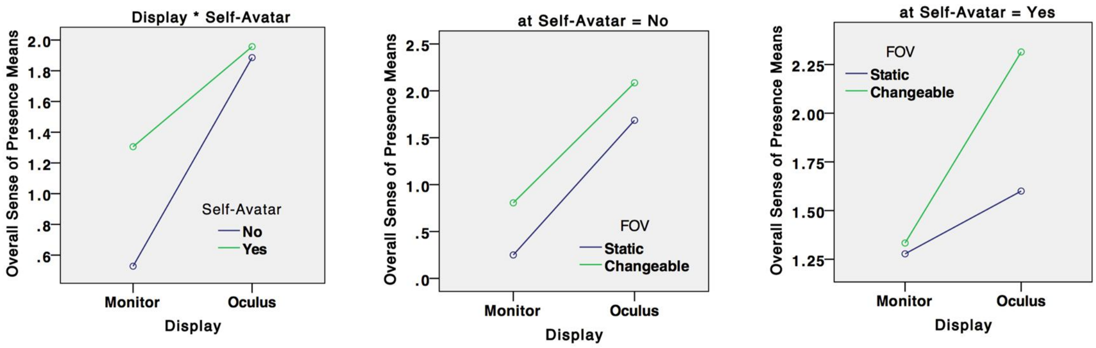


Figure 9: Left: interaction between Display and Self-Avatar factors. Centre: interaction between Display and FOV factors for Self-Avatar. Right: interaction between Display and FOV factors for Self-Avatar.

Realism: Two Likert-like items of the sense of presence questionnaire were used to measure realism: 1) "How much did your experience in the virtual environment seem consistent with your real world experience?", anchored with -3 (not consistent) and 3 (very consistent), and 2) "The virtual world seemed more realistic than the real world", anchored with -3 (fully disagree) and 3 (fully agree). The averages of these two questions were used to perform the analyses. An ANOVA revealed a significant main effect for the FOV factor, $F(1,70)=6.03$, $p < .017$, $\omega 2 = .08$, and the self-avatar factor, $F(1,70)=4.07$, $p < .06$, $\omega 2 = .08$. No significant effects were observed for display type, or the interaction between factors.

Table 7: Means and standard deviations of participants' general sense of presence and realism

| | | | Monitor | | | HMD | | |
|---|---|---|---|---|---|---|---|---|
| | | | Self-avatar | | | Self-avatar | | |
| | | | No X | Yes A | | No X | Yes A | |
| General sense of presence | FOV | No X | .25 (1.62) | 1.28 (1.34) | .76 | 1.61 (1.02) | 1.6 (1.20) | 1.60 |
| | | Yes C | .81 (1.47) | 1.33 (1.26) | 1.07 | 2.08 (.84) | 2.31 (.71) | 2.19 |
| | | | .53 | 1.31 | **0.92** | 1.85 | 1.94 | **1.90** |
| Realism | FOV | No X | -0.73 (1.06) | -0.36 (1.23) | -0.55 | -0.47 (1.25) | -0.54 (1.39) | -0.51 |
| | | Yes C | -0.43 (1.22) | -0.22 (1.31) | -0.33 | -0.23 (1.47) | -0.14 (1.49) | -0.19 |
| | | | -0.58 | -0.29 | **-0.44** | -0.35 | -0.34 | **-0.35** |

## 4.5 Simulator sickness

Simulator sickness is usually associated with immersive VEs, such as HMDs, and as a confounding variable was measured using a standard simulator sickness questionnaire (SSQ). Subjects, in the HMD group, had to answer the SSQ as part of the sense of presence questionnaire. Because simulator sickness is not the focus of this study, only selected symptoms (general discomfort, difficulty concentrating, dizziness, difficulty focusing, and nausea) out of 16 (original SSQ), were measured. Each subject had to rate each symptom from 0 (none) to 4 (severe). The percentage of the number of participant who actually felt sick and their average ratings are reported in Table 8. As expected, more participants experienced simulator sickness symptoms when the FOV was static (60%) and the number dropped almost to half when they were able to look around (33%). However, their symptoms were slight (the average rating varied from 1.1 to 1.4 for each condition) and did not affect their ability to complete the study.

Table 8: percentage and average rating of participants' simulator sickness

| | | HMD | | |
|---|---|---|---|---|
| | | Self-avatar | | |
| | | No X | Yes A | |
| FOV | No X | 58% 1.3 | 63% 1.1 | 60% |
| | Yes C | 36% 1.2 | 30% 1.4 | 33% |
| | | 47% | 46% | **46%** |

## 4.6 Experience

The comparative questionnaire was answered only once by each participant after completing all conditions. Four questions, consisting of seven-point Likert-like scale items, were developed to measure user experience in each group. First two questions correspond to the self-avatar factor as follow: Q1) "Do you think the virtual body (self-avatar) made it easier to judge door/gaps in the virtual environment?", and Q2) "When the avatar was not there, did you feel more or less comfortable?" Participants, in both groups, found it easier to judge doorframes/gaps when the self-avatar was present with both means above mid-point (Monitor group: M = 4.69, SD = 1.56; HMD group: M = 4.61, SD = 1.40). However, they felt less comfortable when the self-avatar was not present (Monitor group: M = 3.42, SD = 1.79; HMD group: M = 3.50, SD = 1.30). An independent-sample t-test was conducted to compare ease of judgment (Q1) and comfort (Q2) between monitor and HMD groups. No statistically significant differences between means were found.

The last two questions (Q3 and Q4) corresponded to the FOV factor. Question 3 asked: "Do you think the controllability of the field of view made it easier to judge door/gaps in the virtual environment?" Similar to self-avatar presence, participants believed it was easier to judge doorframes/gaps when they could look around with both means above mid-point (Monitor group: M = 5.14, SD = 1.51; HMD group: M = 5.47, SD = 1.42). No statistical significant was found between the two groups. Question 4 asked: "When the field of view was static, did you feel more or less comfortable?" An independent-sample t-test indicated that there was a significant difference between the scores for the monitor group (M = 3, SD = 1.74) and the HMD group (M=2.22, SD=1.40). A boxplot of all 4 questions can be seen in Figure 9.

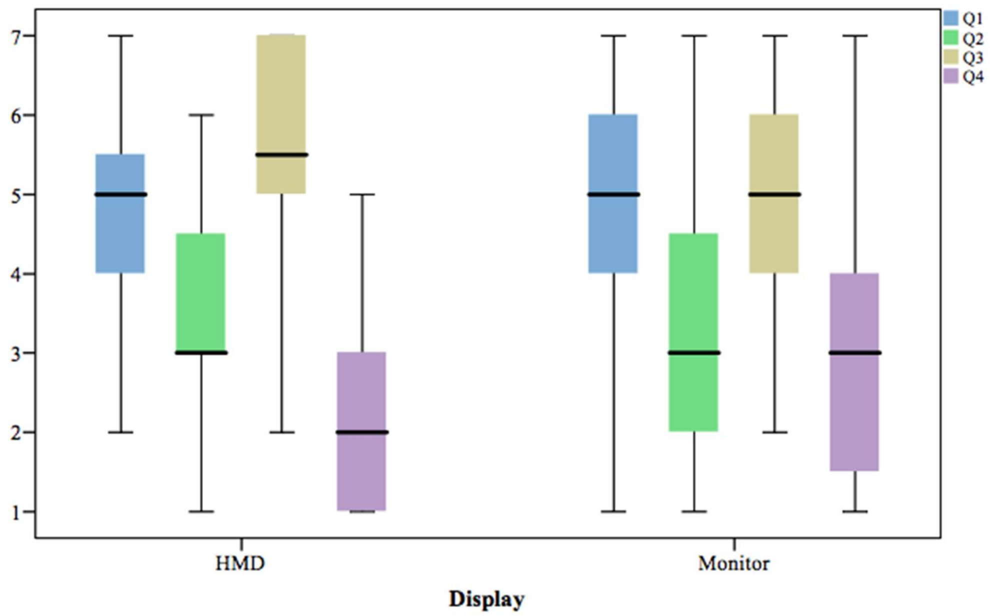


Figure 9: Boxplot of all 4 comparative questions: Q1: self-avatar ease of judgement, Q2: self-avatar comfort, Q3: FOV ease of judgement, Q4: FOV comfort

## 4.7 Preference

Participants were asked about their preference in which they had to choose one of the four conditions in each group. In both group, the C-A (controlled field of view, with self-avatar present) condition was the most favoured with 63% of the response in the monitor group and 68% in the HMD group. The following graphs represent the responses of subjects to questions five (Figure 10).

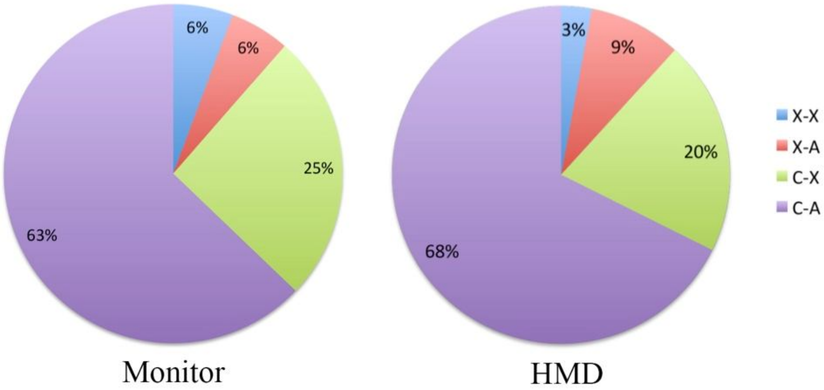


Figure 10: Participants' preferences for each condition within each group

## 6. Discussion

This study investigated how different immersion factors influenced participants' perception, behaviour, and sense of presence while driving a virtual PWC. Findings suggest that while the main effects of display type (Monitor v HMD) were strong and persistent on most participants' behaviours and perceptions, the main effects of FOV were only strong on participants' behaviours, whereas the self-avatar had no main effects at all. Furthermore, all immersion factors significantly affected participants' sense of presence.

### 6.1 Correct Detection and False Alarms

Correct detection: Participants showed significantly more accurate behaviour in detecting passable gaps with the HMD and changeable FOV than when using the monitor display or a static FOV. The effects of the FOV were different for the two displays. For example, the difference between having the FOV changeable or not did not make much difference when participants used the HMD compared to those who used the monitor. In addition, with the static FOV, participants showed much better detection scores on the HMD, whereas when having the changeable FOV there was no such performance difference between the displays. The effects of the three factors were found for the participants' behaviour in manoeuvring between or around the gaps but did not affect how participants explicitly judged the pass-abilities of the doorframes in the VE. One possible explanation is that accurate detection was easier from close distance to the virtual gaps. In contrast, the explicit doorframe judgments were made from a fixed distance (3 units in Unity = 3 meters relatively to the VE).

The significant three-way interactions between the immersion factors were different across the display factors. For the monitor group, the best and worst detection scores were when the avatar was not present (best with the changeable FOV and worst with the static FOV). For those using the HMD, on the other hand, the best and worst detection scores were when the avatar was present (best with the changeable FOV and worst with the static FOV). This could be due to the fact that participants were able to look around and see their whole body, which could have facilitated better judgment rather then just seeing the hands and parts of the legs with the static FOV. This is consistent with other research findings, for example, Mohler et al. claims that participants make less errors in judging distance in immersive VE if they can fully explore a self-avatar of themselves [25].

False alarms: The effects of the display factors were significant only on participants' explicit perceptual judgments. With the HMD, participants significantly reduced the number of false alarms and were better able to correctly reject/perceive doorframes' pass-ability. However, neither FOV nor self-avatar had an effect on participants' perceptions or behaviours. This may be due to the fact that HMD improve depth perception therefore enhancing spatial sensitivity, which enabled more accurate judgment of distance in the VE [26]. Interestingly, the interactions between FOV and self-avatar were significant on behaviour. In particular, participants made fewer false alarms when they were able to look around and the self-avatar was not present. With the avatar presence, participants seemed to better detect pass-ability but they also made more false alarms.

In summary, only display type and FOV affected participants' behaviours in detecting passable gaps. In addition, HMD worked better in both FOVs. The changeable FOV was more effective when the self-avatar was present in the HMD and not present in the monitor display. The use of the HMD also improved perceptual sensitivity and reduced the number of false alarms in judging passable doorframes, in particular, with the changeable FOV and the absence of the self-avatar. The self-avatar did not play a large role in detecting passability, in fact, it reduced sensitivity to some degree as shown with the number of false alarms.

## 6.2 PPP and NPP

There were a number of participants who failed to discriminate between the stimuli in each condition. In terms of behaviour, the number of non-discriminators was exactly the same in both display groups. In terms of perception, the number of non-discriminators in the monitor group was almost twice the number in the HMD group. However, these differences were not statistically significant. The proportion of correct "Yes" responses and "No" responses were calculated for the remaining discriminators to yield the measures of positive (PPP) and negative (NPP) predictive power.

The main findings were that display type affected PPP for the perceptual judgments and NPP for both behavioural and perceptual measures. The FOV factor affected both PPP and NPP for the behavioural measures whereas self-avatar had no main effects at all. With the HMD, participants had significantly better PPP than those using the monitor display in judging passable doorframes. It was clear that with the HMD and changeable FOV most participants were better able to avoid all un-passable gaps while driving regardless of the self-avatar presence.

## 6.3 Sense of presence, comparison and preference

All the immersion factors affected participants' general sense of presence. Their sense of presence was increased when using the HMD, changeable FOV, or self-avatar. Realism, on the other hand, was only affected by the changeable FOV. The interaction between the display and self-avatar was better with HMD regardless of the self-avatar presence. The self-avatar did increase the sense of presence when a monitor was used. In the three-way interactions, self-avatar presence did not have effects on the monitor group but it did in the HMD group. The changeable FOV increased participants' sense of presence in both groups.

Although participants thought that the self-avatar presence made it easier to judge passability, the self-avatar factor did not have effects on either participants' perception or behaviour across all measures. Furthermore, participants also felt less comfortable when the avatar was not present. Unlike the self-avatar, changeable FOV affected participants' behaviour while participants thought it made it easier for them to judge passability. It was found that not having the changeable FOV features made participants significantly less comfortable when using the HMD compared to the monitor display. The preference result was quite similar in both groups in which participants preferred the changeable FOV with the self-avatar presence in both groups. Changeable FOV with no self-avatar presence was the second most preferred in both groups.

## 6.4 Implication for simulator design and use

The present study suggests that an effective simulator for PWCs should at least include the changeability of the FOV as a design feature, particularly if no HMD is used. The HMD display, in this case the Oculus Rift DK2, improved participants' perceptions and behaviours on most of the measures, especially with regards to the accuracy of detecting passable and unpassable doorframes/gaps. The introduction of a self-avatar could be considered - although it did not have direct main effects on participants' perception and behaviour, it had some significant interaction effects with the HMD and/or the FOV.

The numbers of correct detections and false alarms and the associated positive and negative predictive power measures used here to measure participants perception and behaviour accuracy in the VE could also be useful as assessment and/or training measures in future applications. For example, these values could determine the user's risk levels, indicating users' weaknesses and strengths in decision making and judging. Such outcomes would also help to determine users' spatial memory and navigation abilities. The result of this research could also benefit other vehicle simulation systems, in particular towards navigational interaction in VR systems in general. For instance, car, airplane, or bicycle simulators all require good perception of the VE to accurately make the right decision while navigating the VE.

### 6.5 Limitations and future works

Our study was conducted with (mainly) college-age students. This sample has likely experienced virtual reality applications before, has good cognitive function, and is more likely to have used computers and computer games before. Also, they are not the main targeted user group for this particular vehicle. These factors might have influenced users' perception and behaviour. Future research should consider a broader and more targeted sample or even actual PWC users.

A potential methodological concern is the possible influence of a learning effect generated by repeating the task. The experiment was designed, however, to greatly minimise such potential effects by: 1) mixing between-subject and within-subject designs; 2) randomising and counterbalancing, and; 3), the unawareness of the participants of the task repetitions because of our particular task design.

Another limitation is that the experiment tested only three factors with two levels each thus did not fully reflect all the simulator features that may influence perception or behavior. However, we based our choice on issues reported by therapists as well as by comprehensive literature review. Future research could investigate other factors such as VE colours, different avatar appearances, and different visualization techniques, for example bird's eye views etc.

Future research should also consider assessment measures beyond traditional simple task-based measures. Simulators are limitless and with the right implementation, alternative values and measure can be used. We used the number of hits and false alarms to assess participants' perception and behaviour, and we extended our analysis by calculating probabilities of making the right/wrong decisions. An investigation on how to use these values to provide other assessment methods would be of great value.

This study investigated three properties of driving simulators – display type, field of view changeability, and self-avatar presence– and their effects on participants' perception and behaviour. The findings provide strong evidence for the potential benefits of using a head-mounted display HMD, such as an Oculus Rift, and the powerful effects of being able to look around the VE. Our contribution lies in the fact that this experiment probed how accurately PWC users were able to behave and perceive action possibilities in the VE which is a necessary pre-requisite of transferable training and assessment. The results provide some potential design guidelines for future PWC simulator design.

**Acknowledgments**

The authors wish to thank all participants and the HCI group including Mohammad Alghamdi, Johnny Collins, Chris Heinrich, Simon Hoermann, and Carl Leichter. The first author is sponsored by the Saudi Arabian Ministry of Education.

## 7. References


[1] J. R. Wilson, "Virtual environments applications and applied ergonomics," *Appl. Ergon.*, vol. 30, no. 1, pp. 3–9, Feb. 1999.

[2] R. A. Cooper, D. Ding, R. Simpson, S. G. Fitzgerald, D. M. Spaeth, S. Guo, A. M. Koontz, R. Cooper, J. Kim, and M. L. Boninger, "Virtual reality and computer-enhanced training applied to wheeled mobility: an overview of work in Pittsburgh," *Assist. Technol. Off. J. RESNA*, vol. 17, no. 2, pp. 159–170, 2005.

[3] A. Alshaer, S. Hoermann, and H. Regenbrecht, "Influence of peripheral and stereoscopic vision on driving performance in a power wheelchair simulator system," in *2013 International Conference on Virtual Rehabilitation (ICVR)*, 2013, pp. 164–152.

[4] A. Harrison, G. Derwent, A. Enticknap, F. D. Rose, and E. A. Attree, "The role of virtual reality technology in the assessment and training of inexperienced powered wheelchair users," *Disabil. Rehabil.*, vol. 24, no. 11–12, pp. 599–606, Jan. 2002.

[5] P. S. Archambault, S. Tremblay, S. Cachecho, F. Routhier, and P. Boissy, "Driving performance in a power wheelchair simulator," *Disabil. Rehabil. Assist. Technol.*, vol. 7, no. 3, pp. 226–233, May 2012.

[6] P. Abellard, I. Randria, A. Abellard, M. M. Ben Khelifa, and P. Ramanantsizehe, "Electric Wheelchair Navigation Simulators: why, when, how?," in *Mechatronic Systems Applications*, A. M. D. Di Paola and G. Cicirelli, Eds. InTech, 2010.

[7] J. Kjeldskov, "Combining interaction techniques and display types for virtual reality," presented at the OZCHI: Annual Conference of the Australian Computer-Human Interaction Special Interest Group, Australia, 2001, pp. 77–83.

[8] C. Harrison, P. Dall, P. Grant, M. Granat, T. Maver, and C. Bernard, "Development of a wheelchair virtual reality platform for use in evaluating wheelchair access," in *Proceedings of the Third International Conference on Disability*, Alghero, Italy, 2000.

[9] T. Pithon, T. Weiss, S. Richir, and E. Klinger, "Wheelchair simulators: A review," *Technol. Disabil.*, vol. 21, no. 1, pp. 1–10, Jan. 2009.

[10] U. Schultze, "Embodiment and presence in virtual worlds: a review," *J. Inf. Technol.*, vol. 25, no. 4, pp. 434–449, Dec. 2010.

[11] B. Ries, V. Interrante, M. Kaeding, and L. Anderson, "The effect of self-embodiment on distance perception in immersive virtual environments," 2008, p. 167.

[12] S. Lim and B. Reeves, "Being in the Game: Effects of Avatar Choice and Point of View on Psychophysiological Responses During Play," *Media Psychol.*, vol. 12, no. 4, pp. 348–370, Nov. 2009.

[13] T. Dodds, B. J. Mohler, S. de la Rosa, S. Streuber, and H. H. Bülthoff, "Embodied interaction in immersive virtual environments with real time self-animated avatars," in *Workshop Embodied Interaction: Theory and Practice in HCI*, New York, NY, USA, 2011.

[14] H.-M. Sun, S.-P. Li, Y.-Q. Zhu, and B. Hsiao, "The effect of user's perceived presence and promotion focus on usability for interacting in virtual environments," *Appl. Ergon.*, vol. 50, pp. 126–132, Sep. 2015.

[15] S. H. C.-R. Betty J. Mohler, "The Effect of Viewing a Self-Avatar on Distance Judgments in an HMD-Based Virtual Environment," *Presence*, vol. 19, no. 3, pp. 230–242, 2010.

[16] D. Henry and T. Furness, "Spatial perception in virtual environments: Evaluating an architectural application," in *, 1993 IEEE Virtual Reality Annual International Symposium, 1993*, 1993, pp. 33–40.

[17] M. Geuss, J. Stefanucci, S. Creem-Regehr, and W. B. Thompson, "Can I Pass?: Using Affordances to Measure Perceived Size in Virtual Environments," in *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, New York, NY, USA, 2010, pp. 61–64.

[18] M. Rébillat, X. Boutillon, É. Corteel, and B. F. G. Katz, "Audio, Visual, and Audio-visual Egocentric Distance Perception by Moving Subjects in Virtual Environments," *ACM Trans Appl Percept*, vol. 9, no. 4, pp. 19:1–19:17, Oct. 2012.

[19] J. J. Gibson, *The Ecological Approach to Visual Perception : Classic Edition*. Hoboken: Taylor and Francis, 2014.

[20] L. S. Mark, “Eyeheight-scaled information about affordances: a study of sitting and stair climbing,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 13, no. 3, pp. 361–370, Aug. 1987.

[21] W. H. Warren and S. Whang, “Visual guidance of walking through apertures: body-scaled information for affordances,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 13, no. 3, pp. 371–383, Aug. 1987.

[22] Access 2000, “Design for Access and Mobility – Buildings and Associated Facilities.” New Zealand Standards/, 2001.

[23] J. L. Szalma, P. A. Hancock, J. S. Warm, W. N. Dember, and K. S. Parsons, “Training for vigilance: using predictive power to evaluate feedback effectiveness,” *Hum. Factors*, vol. 48, no. 4, pp. 682–692, 2006.

[24] H. Regenbrecht and T. Schubert, “Real and Illusory Interactions Enhance Presence in Virtual Environments,” *Presence Teleoperators Virtual Environ.*, vol. 11, no. 4, pp. 425–434, Aug. 2002.

[25] B. J. Mohler, H. H. Bülthoff, 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*, New York, NY, USA, 2008, p. 194–.

[26] E. Heineken and F. P. Schulte, “Acquiring Distance Knowledge In Virtual Environments,” in *What Is Essential For Virtual Reality Systems To Meet Military Human Performance Goals?*, Hague, Netherland, 2000, pp. 17–1–17–5.