An Explorative Study of Interaction with Tracked Objects in a Virtual Reality Game

JONATHAN GOLAN
Abstract

In recent years, Internet enabled objects (Internet of Things) and their augmentation through virtual reality has become both technically possible and increasingly advanced. This paper explores how interaction between physical IoTs and their virtual and digital twins can be advanced. Three virtual reality games focusing on three different interactions were created. One game focused on rotating, another on pushing and pulling, and a third on lifting and dropping. All games revolved around the use of real tracked cardboard boxes which were represented in the games by a virtual box in the same relative position and rotation to the player, allowing participants to manipulate the virtual boxes by manipulating the physical boxes. 14 participants were asked to play the games and were afterwards interviewed in regards to the games, their interactions with them, their enjoyability and their thoughts regarding them and the concept in general. The results were acquired from game performance, the reflections of the participants, and the subsequent analysis of recorded audio and video. The study presents the difficulties, challenges and opportunities of such a system, while also providing insight into lessons learned from the creation of the system and the games. The main contributions of the paper are the lessons learned in creating the games and experiences in addition to a few specific areas of interest for future research on the area, namely the importance of ergonomic consideration and affordance evaluation.
Sammanfattning

An Explorative Study of Interaction with Tracked Objects in a Virtual Reality Game

Jonathan Golan
KTH, Royal Institute of Technology
Stockholm, Sweden
Golan@kth.se

Degree Project
Supervisor: Anders Lundström
Examiner: Rob Comber

ABSTRACT
In recent years, Internet enabled objects (Internet of Things) and their augmentation through virtual reality has become both technically possible and increasingly advanced. This paper explores how interaction between physical IoTs and their virtual and digital twins can be advanced. Three virtual reality games focusing on three different interactions were created. One game focused on rotating, another on pushing and pulling, and a third on lifting and dropping. All games revolved around the use of real tracked cardboard boxes which were represented in the games by a virtual box in the same relative position and rotation to the player, allowing participants to manipulate the virtual boxes by manipulating the physical boxes. 14 participants were asked to play the games and were afterwards interviewed in regards to the games, their interactions with them, their enjoyability and their thoughts regarding them and the concept in general. The results were acquired from game performance, the reflections of the participants, and the subsequent analysis of recorded audio and video. The study presents the difficulties, challenges and opportunities of such a system, while also providing insight into lessons learned from the creation of the system and the games. The main contributions of the paper are the lessons learned in creating the games and experiences in addition to a few specific areas of interest for future research on the area, namely the importance of ergonomic consideration and affordance evaluation.

Keywords
Virtual Reality, Passive Haptics, Tangible Interaction, Digital Twin, Computer Vision, Marker Tracking

1. INTRODUCTION

Two areas that have been in the spotlight in recent years are Virtual Reality (VR) and Tangible Interaction (TI). Most interactions in VR today use of either the standard keyboard and mouse most computers use, or custom hand controllers, depending on the VR hardware used. While these controllers might be more intuitive than a keyboard and mouse for VR, they are still artificial and lack the tangible qualities of real physical objects. For instance, the shape, weight, affordances, texture, and so on. The purpose of this study is to look closer at interactions with synchronized virtual and real objects in VR. The field of TI began with the paper “Tangible Bits” by Ishii and Ullmer [17]. The idea at the time was that digital interactions do not make full use of the richness of natural interactions with real objects, by interacting with for example household objects. [12] Ishii and Ullmer also argue that we should not limit ourselves to screens, but instead make use of our surroundings for rich meaningful interactions.

Such a system is heavily reliant on the accuracy of the virtual representation of reality when it comes to position, orientation, size, shape etc. It is the hope of this paper to study those issues and the interactions with multiple physical and virtual synchronized objects.

To study this area, three different tangible VR games were created, each focusing on different interactions with three to four large cardboard boxes that had both a physical and a virtual synchronized representation. In the setup, a camera tracks the boxes and sets the position and orientation of similar virtual boxes in VR. Hence, the virtual representations have the same relative positions in the virtual environment and in relation to the user. This allows the user to move the digital box by moving the real box, and find the real box by looking at the digital box. Note that the physical object is invisible to the user that wears a head mounted display (HMD).

Using these three games a study was set up with 14 participants who were asked to complete the tasks provided in these games. These sessions were video recorded and a short interview was conducted after each task, followed by a longer post-interview focusing on how the interactions compared to their expectations, their experience with and enjoyment of the games, and issues they had with the system in general. The videos and interviews were thematically analyzed in relation to the research question presented below.

Research question
What are the difficulties, challenges and opportunities in interacting with multiple tracked real objects in a Virtual Reality game?
2. RELATED RESEARCH

Below follow short introductions into the various subjects related to this paper.

2.1 Virtual Reality

VR has previously been defined as “a real or simulated environment in which a perceiver experiences telepresence” where telepresence is defined as “the experience of presence in an environment by means of a communication medium”, for example an HMD [28]. Steuer, however, admits that the definition does not mesh very accurately with typical uses of the term. Looking at modern popular definitions such as that of the online dictionary Merriam Webster, the term is defined somewhat differently. Merriam Webster defines VR as “an artificial environment which is experienced through sensory stimuli (such as sights and sounds) provided by a computer and in which one's actions partially determine what happens in the environment”. [24] The main difference between general definitions and scientific definitions is the perceiver’s ability to interact with the environment.

However, according to the Milgram mixed reality continuum, Virtual Reality is at the end of the spectrum and only used for purely virtual environments. This can not be said for the environment used in this paper, as the real world boxes will be incorporated into the VE.

2.2 Substitutional reality

The research most relevant to this paper is the work done by Simeone et al. [26] They researched what they call Substitutional Reality. They define a Substitutional Environment as a “VE in which a substitution process has matched each virtual element with a physical element with varying degrees of mismatch”. The purpose of their research was to determine to what extent objects could be substituted in VR without affecting presence and immersion. They concluded that the most believable substitutions were those where the parts most likely interacted with were similar, eg a handle. Mismatch between VR and reality reportedly became significant when there were differences in tactile feedback, temperature and weight. So while their environment is similar to the one used in this paper, their goal and purpose are different in that they focused on how a substitutional reality affects perceived presence in a VE, this paper focuses on how pairing real objects to virtual ones affects the interactions with those objects.

Campbell et al. have also conducted similar research, with their exploratory study “that incorporates the design, implementation and study of a system that utilises virtual reality, tangible interaction and force feedback”. [5] They put test participants in a real wheelchair while in VR. Their movements in the wheelchair control a virtual wheelchair, which their avatar is seated in. The participant is then allowed to move around the VE. Whenever the virtual wheelchair bumps in a virtual object, brakes are applied to the real wheelchair to simulate a real collision. Their work points to the potential effectiveness of a dually tangible mixed reality. They intend to collect physiological data and measure presence in future work.

2.3 Passive Haptics

Providing tangible interaction would only require some tangible object to interact with, not replicas. However, doing so also provides something known as passive haptics.

Lindeman defined passive-haptic devices as “physical objects which provide feedback to the user simply by their shape, texture, or other inherent properties”. [21] This is opposed to active haptics where the feedback is controlled by some system. A clearer popular definition is “receiving feedback from touching a physical object that is registered to a virtual object”. [26]

Passive haptics have been shown to bring many improvements when added to VR systems. It has been shown to increase the realism of VEs, increase the feeling of presence in VEs, increase the sense of weight and solidity of other objects in the same VE and improving cognitive mapping of environments, among other things. [13][14][16][9]

2.4 Tracking

There were two options considered for the tracking of the boxes. The first one is the Vive Tracker, by HTC, made specifically for tracking additional objects beyond the controllers and the HMD. [32] The alternative is using a camera and computer vision to track markers, such as the Aruco module of OpenCV. [25] Aruco works by first printing markers which are then placed on objects or by themselves in the scene. A camera is then used to take a picture of the scene. The software will then find the marker and output its position and pose (angle and direction). This data gives us a good estimation of the marker’s real world position and pose.

The main strength of the Vive Tracker is the accurate tracking without additional setup, when already using the HTC Vive. As the Vive uses two sensors at opposite corners of the tracked space, there is fairly good protection against occlusion. Of course, putting the tracker under or inside a box will prevent it from working. The tracker also requires physical space and thus would for example prevent a cube from standing on that side if attached to it.

Figure 1. Left: a Vive Tracker. Right: An Aruco marker
Another downside of the Vive Tracker is the price tag of 120 euro per device. [32] As it is the intention of this paper to study interaction with multiple tracked objects, this price is severely prohibiting.

The only thing provided by the Aruco software is the image-to-3D-position of the marker. In order to avoid losing the position of a box, markers would have to be placed on all sides. Occlusion will still be a large problem with one camera. This can of course be alleviated by adding cameras, but this compounds the tracking as well. The markers are printed on paper and thus do not take up any space in practice. As markers can be printed freely the only thing limiting the number usable is the amount that can fit inside the viewing field of the camera. This amount depends on the size of the markers, their distance to the camera and the resolution of the camera. Speed is also an issue with markers. If the camera is slow, fast movements can make images blurry making it impossible to locate markers. Another factor worth considering is future potential. The area of computer vision is currently experiencing rapid development, and for example Facebook CEO Mark Zuckerberg hopes for the technology to be able to accurately describe images within 10 years time. [7] If that comes to pass, identifying and tracking objects, such as boxes, in an image could also be possible. The Augmented Reality (AR) technologies created by Google, ARCore, and Apple, ARKit, already have some markerless tracking capabilities. [2][1]. While it is easy to see the future strength of computer vision, it is harder to imagine where Valve will take their Lighthouse technology.

Ultimately, computer vision was chosen for this project, based mainly on price and the increased movements allowed by it.

2.5 Cybersickness

Whenever VR is involved, cybersickness is an important concern. Cybersickness has symptoms very similar to motion sickness and can appear both during and after a VR experience. While they are similar, cybersickness is still considered a distinct affliction. [20] A theory with high credibility for why cybersickness occurs is the mismatch between visual and vestibular systems (eg. seeing movement but not feeling it). However, cybersickness can be largely eliminated by using the tracking of the Vive system. Because the headset tracks the users head-position and direction, before showing the same movements in the VE, visual and vestibular data will match. As the tracking system will be used in the games created for this paper, no cybersickness is expected.

2.6 Internet of Things

An area that has received a lot of attention recently is the Internet of Things (IoT), which is the network of different appliances embedded with electronics, allowing them to collect and exchange data with each other. A Digital Twin (DT) is a digital replica of one of these physical assets, and is often used as a way of visualizing the asset and the data its sensors are generating. DTs have drawn the interest of large companies such as IBM and GE for its potential in industrial usage. [15] The reason for mentioning IoT is that the idea for this paper came from there. A possibly interesting application for the studied technology is for manipulating DTs. While the connection to IoT has been somewhat lost, it is still our hope that this research could say something about use cases for DTs as well.

3. Method

This project consists of two main parts. The first part was the development phase, which included the environment itself, the tracking and the games. The second part was a study aiming to answer the research question by video recording participants playing these tangible VR games and through semi-structured post-interviews. Three specific interactions were chosen for these games in order to learn more about issues related to different types of interaction techniques and interaction with multiple objects. This resulted in pushing and pulling, lifting and dropping and rotating boxes games.

14 participants (1 female) with a mean age of 25 years were recruited for the study. Seven of the participants were recruited from the Capgemeni office while the rest were acquaintances of the researcher. Two participants had no higher education (university or higher), six participants were current higher education students, two participants had finished higher education and four participants had an unknown level of education. In the following sections details about the study procedure, interviews, data analysis and implementation will be presented.

3.1 Study procedure

The procedure followed the following overarching structure: 1) presentation of the study, consent and right to withdraw participation, 2) demographic questionnaire, 3) introduction scene to get familiarized with the system followed by a short interview regarding first impressions, 4) push and pull game and a short interview, 5) rotation game and a short interview, 6) lift and drop game and a short interview, followed by 7) a final general interview.

The setup of the room can be seen in Figures 3 and 4. Figure 3 depicts a participant playing the Rotation Game. Figure 4 depicts the general setup of the surroundings.

All interviews were semi-structured. The short interviews that were conducted in between the gameplay sessions aimed to get feedback on the specific interactions and avoid mixing different impressions together. The semi-structured interviews averaged around 15 minutes in total per participant, resulting in 212 minutes of total recorded audio. All gameplay was recorded using a video camera.

The objects (boxes) used for interaction in this paper are not really unique enough to warrant their own name. However, as I have been unable to find an existing name that would describe them adequately, I’ve chosen to call them by their description, a paired real and virtual object, or Virtual-Real Pair (VRP).
Consent and demographics questionnaire

Prior to the study, the participants were asked to read and fill in a consent form informing them of the general purpose and the procedure of the study. They were also asked to fill in a form regarding their demographic information. They were also asked if video and audio recording was allowed and informed that they could choose to end the test at any point, for any reason.

Introduction Scene

First was an introductory scene, consisting of an otherwise empty room with four virtual boxes in front of the participant, each with its own colour. Under the boxes were the same colours, but in a different order. Participants were asked to first accustom themselves to the VR environment and then sort the virtual boxes according to their colours and the colours on the floor. The purpose was to get them acquainted with both the VR and the VRPs, while also setting up the boxes in the correct positions for the following game. They were allowed to stay as long as they wanted, but the aim would be a couple of minutes. After exiting the game they were asked a few basic questions regarding how they felt about the VR experience itself, for example concerning cybersickness and how they experienced interacting with the VRPs.

Push and pull game

The first game was a push and pull game. The scene consists of four virtual boxes placed on the ground in front of the participant. On the wall facing them there is a target number, current number, score and time left. The participant would push the boxes back and forth as switches for inputting the target binary number. Each correct configuration would award the participant one point. The reason for the additional VRP was in part to see how a larger number of VRPs affected the interaction, but also to make the binary counting of the pushing interaction game more challenging (a total of 16 states instead of 8). Participants were allowed to take as long as they wanted for inputting the first requested number after which they were again asked to play for 200 seconds. The time taken per requested number was recorded. This play session was followed by a short interview regarding their thoughts about the specific interaction that was used in that game and the game itself.

Rotating boxes game

Next was a rotation interaction game. The scene consists of three virtual boxes, with the appearance of differently patterned dice. A random pattern, their score and the time left are displayed on the wall in front of the participant. The participant then has to match the patterns on the boxes in front of them with the pattern on the wall by rotating the dice so that the correct side is oriented upwards. The player is allowed to take their time finishing the first pattern, after which a 200 second timer started ticking down. The time taken per pattern was recorded. This play session was followed by a short interview regarding their thoughts about the specific interaction that was used in that game and the game itself.

Lift and drop game

Last was a lift and drop game. The scene consists of three virtual boxes with a road texture on top of them. This texture has no special properties, but is there to make the usage of the boxes clearer. In front of the participant there is a wall with six tunnels with roads coming out of them, a top row of three and a bottom row of another three tunnels and roads. All roads protrude from the tunnels for a distance before a gap followed by another piece of road entering tunnels on the opposite wall. When the boxes are resting on the ground their top side aligns in height with the bottom roads, but has to be lifted up to reach the top row. On the wall above the tunnels in front of the participant there is a count of the amount of remaining cars as well as their current score. Once the game starts cars start coming from all six tunnels, but one at a time and the participant has to lift the VRPs to form bridges for the cars to pass safely. Each car that exits the tunnels behind the participant awards them with a point. The cars come at set intervals and the number of created cars, 40, was decided so as to result in a total playing time of 200 seconds. The number of cars saved was recorded. This play session was followed by a short interview regarding their thoughts about the specific interaction that was used in that game and the game itself.

General Interview

After the interview for the lift and drop game, a more general interview about their cumulative experience and thoughts about the system followed.

3.2 Analyzing Quantitative Data

Thematic Analysis (TA) [31] was chosen for analyzing all the data gathered from the interviews. The main reasons for choosing TA was the relative ease of use, speed, and its proficiency for summarizing large sets of data. TA has a predefined order of doing things and it was followed quite closely:

1. The data was gone through thoroughly, selectively transcribing parts of interest to the research question. Participants were coded.
2. The transcribed data was coded.
3. Codes were collated into preliminary themes.
4. Themes were reviewed to check if they fit the research question, the relevance, as well as to deepen the analysis of each theme.
5. Finalization of themes, naming and defining them.
6. Final analysis and selection of illustrative examples bringing out the core aspects of the themes.

This process resulted in the following themes: enjoyment, affordance, physical exertion & ergonomy, ease of use, and use cases.

The video recordings of gameplay were analyzed in a similar fashion, however not strictly following TA. When the videos were analyzed, game strategy and interactions were recorded and divided into groups. Strategy was
divided into a main strategy for each game, and significant exceptions to those strategies, while the interactions were spread and added into the results found from the interviews.

3.3 Implementation

The implementation consists of four main parts, namely, the cardboard boxes, the VR hardware, the tracking, and the games.

The VR

The HTC Vive [32] was chosen for its accurate room-scale tracking. The Vive has a resolution of 1080 x 1200 per eye, 90 Hz refresh rate and a 110 degree field of view.

The Boxes

The core of the implementation is the tracked cardboard boxes. Three cardboard boxes (240 x 340 x 360mm) were placed in the area tracked by the HTC Vive tracking satellites. Six fiducials were placed on each box; one on each side. The boxes’ positions were then tracked using OpenCV’s [25] Aruco marker tracking software and fed into the games in order to set the equivalent positions and orientations of the virtual boxes.

The Tracking

The tracking used custom software based on the Aruco tracking software from OpenCV. The original software tracked the markers as separate entities while the custom version grouped the markers into groups of 6, as that is how they were placed on the boxes. The camera that was used for the tracking was the Logitech c920 [22] that has a maximum resolution of 1080p with 30 fps. It was chosen as it was the best camera available at the Capgemini office. It was apparent through testing that the c920 would be enough for the purposes of this thesis, but future work could take advantage of advances in camera hardware to increase both frame rate and resolution, which should increase the accuracy of the tracking substantially. A higher resolution also allows the use of the camera to track a larger area.

![Figure 3. Participant playing the Rotation Game.](image)

4. RESULTS
The results section contains descriptions of the participants’ strategies, mistakes, interactions and reflections followed by summarizations of the interviews, sorted under the themes found by the Thematic Analysis.

4.1 Participant interactions and gameplay

In general, the different types of games led the participants to similar interactions and gameplay within each game. However, some differences, misconceptions and exceptions were observed in how participants completed their tasks. There were also some observations of how the limitations of the tracking did not fully support the participants. In the following sections the general interactions, differences, misconceptions and exceptions will be presented for each game.

4.1.1 Introduction Scene

All participants completed the simple task without issues. No participants reported nausea or other similar symptoms. However, one participant (P5) reported minor discomfort at the beginning of the test and commented that it caused him to be more careful in the virtual environment.

As regards how it felt to interact with the boxes, half of the participants reported a mismatch in size between the real and the virtual box. Noteworthy was that most of them also had a hard time pinpointing what the difference was. For example, participant (P4) commented that:

They seem to be a bit wider in reality (P4)

While this was technically correct as the size of the real box was indeed a little larger, it is important to clarify that the actual shape of the virtual and the real box was identical, meaning that the real box was symmetrically a little larger in all dimensions, not only in width as perceived by P4. The remaining participants who commented on the size gave general comments on how the real boxes seemed to be larger than the virtual ones.

4.1.2 Push and Pull Game

Participants looked at the wall in front of them to see what number they have to construct followed by pushing and pulling the boxes into position and finally looking up at the wall in front of them for the next number and continuing the loop.

As the floor was covered in carpet, it was quite easy to simply push the boxes forward, letting them slide along the ground. Most participants did this from the start, with the exception of four participants (P1, P5, P12, P14) who instead lifted the boxes before placing them down again. However, by the end of the game they were also sliding the boxes forward instead of lifting.

When it came to moving the boxes back, techniques varied from person to person and time to time, most grabbing the box by the left and right top side before pulling the box back along the ground or lifting it through the air. See 5.3 Affordance for some specific exceptions.

4.1.3 Rotation Game

Participants looked at the wall in front of them to see what patterns and numbers they had to match, followed by picking one of the boxes up and rotating it into the correct orientation before putting it down somewhere other than from where they picked it up or where it is going (since there is still a box there). They then moved the box that was in the way, allowing them to place the first box in the correct position. They then repeated this process with the second and third boxes.

While finding the correct side of the boxes was somewhat difficult, it only caused some extra time spent rotating the boxes. What proved to be the real challenge was switching the positions of the boxes. As there is only one camera tracking the boxes in front of the participants, they more or less only have a 2D space to move the boxes in. What many participants wanted to do was to move one box back, and then move another box past it. This however blocked the camera from seeing the box behind, and thus lost tracking of it temporarily. Many participants still chose to use this method and keep track of the box through touch and memory instead. This resulted in occasional fumbling for boxes. Many also chose to temporarily stack the boxes on top of each other to save space, which did not remove the box from the virtual view.

4.1.4 Lift and Drop Game

Participants looked up at the roads in front of them to see where the next car is coming from. If a car was coming on one of the higher levels, the participant would lift the box placed below that lane to fill in the gap in the road and allow the car to pass safely. Afterwards, the participant would place the box back on the ground to fill in the ground level gap in the road. When cars were hidden behind other objects in the scene, such as roads or boxes, a red outline would be visible through those objects, so that the user wouldn’t lose track of the cars’ positions. If a car was coming on the ground level, participants would try to let the cars pass on their own, but sometimes grab the box in order to adjust its position slightly. This would then be repeated for each following car. However, the most efficient observed method, used by (P11) for most of the game, was to place two boxes at two ground level roads and use the middle box for the remaining four roads (three mid air and one ground level). Most participants used this to some extent at some point (using a box for other roads it was not intended for). Some users tried pushing the limits of the game. P13 commented that:

You want to try some tricks after a while but when I tried to move it from one and move it in parallel to the side to make it go on another road it felt like it got locked (P13)

The user wanted to lift a car along with the box horizontally from one road to another parallel road. “Locked” here refers to that the car did not move sideways,
as a result of how the game was programmed, only allowing cars to drive forward.

In general the participants missed very few cars. Many missed their only point on the first car as there was no practice round unlike the other games. P6 among a few others missed a point because of tracking inaccuracies, causing the box to unexpectedly shift, dropping the car.

4.2 Enjoyment

When asked about the relative enjoyment of the three games, the game most often reported as most enjoyable was the Car game, with 10 out of 14 participants rating it as the best game out of the three games. Users reported the interactive aspects as being the reason for their enjoyment, as compared to the other games which largely consisted of following instructions, without much room for freedom. Users in general had very little critique for the games themselves, beyond some lacking game elements, such as feedback, e.g. clearer information on when an attempt is incorrect or correct.

While opinions about the rotation and push and pull games were quite similar to each other, with participants ranking the push and pull game just slightly above the former, some participants were quite enthusiastic about the lift and drop game. P3 commented regarding the lift and drop game that:

> It was a lot more fun than I thought it would be (P3)

The same participant also commented that:

> I felt like I was getting lost in it, in the sense that, you know, I was looking in the wrong area kind of thing but it was nice. I really felt immersed in that little world, just adding a few colors for example and the motion really made me feel like I was immersed in it. (P3)

The second comment indicates a high perceived presence, a metric not measured in this report. It does however have a positive correlation with enjoyment. [30]

Another participant had different reasons for favouring the same game:

> It was more fun than the others, there was more interaction. Or, there was a lot of interaction in the others too, but there was more time pressure because of the moving objects here. (P2)

This seems to indicate that this user liked the reactive nature of the game, or was perhaps motivated by the possibility of failure, which was not present in the same way in the previous games.

4.3 Affordance

The boxes in VR are made out of a quite nondescript material. See Figure 5 for the different VR boxes and the actual cardboard box. When informally asked after the interviews, about what the material of the virtual boxes appeared to be, P6 answered “something like plastic”. The properties of the real and virtual boxes differ in many ways. The real boxes are hollow and made from cardboard, with an A4 paper taped onto each side, causing the texture across the box to be dividing into three types: the rough cardboard, the smooth paper and the glossy tape. Meanwhile, the virtual boxes’ surface was uniformly smooth, with the exception of the top road piece in the lift and drop game, which was still completely flat. The sides were also perfectly straight. For the real boxes, rather than being precisely cuboid, their sides bulged somewhat, and as the tests went on, the edges became increasingly frayed. This indicates that cardboard is not a viable option for lasting physical interaction as there is too much physical strain in the kinds of interactions tested here.

The affordances of the real boxes resulted in some interactions with the virtual boxes which were incongruent with their visible properties. A lot of the handling of the boxes might have been impossible had the boxes been heavier, but perhaps more significantly, participants used two unexpected ways to grab the boxes, that were reliant on the uneven shapes of the cardboard boxes. Both these cases were seen in the push and pull game, where participants might be expected to either put their hands on one side of the box in order to push or pull it in the other direction, or putting their hands on opposite sides of the box in order to lift the box. Instead, in the first case, P9 and P12 were observed pulling the boxes towards themselves by grabbing them by their frayed edges during the binary game and P8 reported doing this as well during the final general interview. Another method was to grab a box by its corner by applying grip pressure to it, as one might hold a basketball with one hand. P1 was observed doing this. P1 is, anecdotally, tall and the owner of large hands. This type of handling requires quite high friction from the lifted object, such as the rubber of a basketball, which is why this was surprising. Furthermore, the texture of the virtual boxes gives the impression of a completely flat surface without imperfections, a surface with seemingly low friction.

4.4 Physical Exertion & Ergonomy

A recurring theme among participants was their surprise at the physical strenuousness of the tasks. Many reported sweating and increased pulse. As participants were not motivated to score higher, other than by competitiveness, expended effort varied greatly among participants. One participant commented that:
General consensus seemed to agree that the physical effort required was more than expected.

While multiple users reported fatigue and pain in their back and neck, the level of physical exertion was not thought to be the direct cause of this, although a contributing factor. Instead, afflicted participants attributed the pain and fatigue to poor working posture while completing the tasks, specifically bending over to manipulate the boxes on the floor, while bending their neck up to see their task, score and time. As participant height was not a collected metric, nothing can be said for sure, but the afflicted participants seemed to be taller. These issues were mainly reported during the Binary game. Notably, participants P8 and P10 preferred to kneel as opposed to standing and did so for most of the test. P10 and P12 suggested placing the boxes on a table for easier and more comfortable access, with P10 commenting that:

*I might have preferred doing this on a table with smaller boxes, especially when you were supposed to switch the positions of the boxes in the binary example. I had to reach over a box to reach another box. If you had been on a table and been able to lift a box with one hand or something, it would have been a much smoother interaction* (P10)

Some participants reported that the Rotation game, which allowed them to take a pose of their own choosing while rotating the VRPs, was a nice change of pace from the otherwise quite bent over poses.

### 4.5 Ease of Use

The clear majority of reported issues with interacting with the VRPs were caused by tracking inaccuracies. In the Binary game this was mainly seen when users did not reach far enough to grab the boxes, and hence missed or fumbled with grasping them. In the Rotation game users sometimes bumped boxes into each other when trying to move them around. Users also complained about delay in the rotational movements of the boxes. In the Car game, the bridge building required users to hold the pieces somewhat still while the car passed, but this was made hard by the fact that the boxes would jitter a little back and forth, which sometimes caused the car to slip off and fall to the ground, without any error from the participant.

As for issues not caused by the tracking (although they were perhaps exacerbated by it) rotation was the most problematic, with multiple users complaining about it being difficult and cumbersome to swap and rotate boxes around, especially because of limitations of the system such as not being able to place boxes behind each other and being limited to the tracked area. Getting a hold on far away boxes in the Binary game also proved challenging for some, because of the outstretched pose in addition to the somewhat hard to grab texture. However, this is a natural problem when using the body and could be viewed as a part of the gameplay.

It was more of a workout than I thought. (P3)

![Figure 6 Push & Pull Game. N is the number of participants with an entry, as faster participants managed to finish more tasks. Seconds is the average time it took for all participants to finish task number x. The thin blue line is a trend line for the average time.](image)

![Figure 7 Same as Figure 6 but for the Rotation Game](image)
were not able to see real objects in their surroundings either. Although the immediate area had been cleared for safety, there was still a desk with the computer just outside the area, as can be seen in Figure 3. Although the HTC Vive system warns users when they get close to the edge, by showing a grid wall, this warning was sometimes missed when the participants were looking in another direction. In fact, two participants accidentally came in contact with the desk, but fortunately only lightly, causing no harm.

An interesting comment made by P7 and P14 was that they wished the boxes had been cube-shaped (equilateral, instead of the used oblong boxes), especially for the Rotation game, as real dice are cube-shaped. They thought that would have made the game easier.

4.6 Use Cases

There were five main recurring themes that were brought up when the participants were asked about use cases for the technology. First, the relatively high levels of physical exertion experienced by many participants led them to believe that using VRPs could be an effective means of making exercise more fun or interesting. Secondly, with additional tracking users suggested that they could receive more in-depth feedback on for example their posture. Thirdly, the participants suggested that VRPs could be used for precise instructions without real, present instructors, for example as instructor movements could easily be compared to instructee movements. Fourthly, mainly based on their experience with the Binary game, some participants suggested using the technology for teaching. Either, as with the Binary game, to teach somewhat abstract subjects in a more concrete way or more generally as gamification. P6 commented that:

*In the same manner that minecraft is used to teach kids, you could give a first impression of binary numbers and addition. (P6)*

Fifthly, it could be used for simulation, which is a common use for VR. However, the participants focused especially on construction and architectural/design simulation. Likely because of the similarity between the boxes and building blocks. While some meant it like literal building blocks, others thought more along the lines of the box representing a module, such as building. The user could then construct neighborhoods or even cities from these modules.

5. DISCUSSION

5.1 Enjoyment

While the lifting and dropping game was a clear favourite, it is difficult to provide a conclusive explanation as to why that is the case. The lifting or dropping interactions themselves do not seem to be the cause, as participants credited the increased interactivity and reactive gameplay. This would point towards the interaction between the VRPs and the cars which was a component that was absent in the previous games. While the cars only drove straight forward, their interactions with the boxes could be quite unpredictable, allowing the participants the chance to explore and experiment more than they could before. Another factor was the sense of urgency and possibility of failure as there was a limited number of cars, and any car that fell off the road was permanently lost.

One would think that the increased interactivity in the rotation game over the push and pull game, picking up boxes and handling them more thoroughly, would give it an edge in popularity. As that was not the case however, the increased challenge and required thinking of the push and pull game might have helped its popularity. The fact that a large number of the participants were students of math and computer science related topics probably helped the math based game a lot. It seems that game design won out in this particular case.

5.2 Affordance

For the games used in this project, the users interacting with the VRPs beyond the apparent affordances of the real virtual boxes (e.g. grabbing them by their frayed edges) did not cause any issues. The differences between the real boxes and the virtual boxes were also relatively small to begin with. It is however not hard to imagine scenarios where this mismatch causes problems, especially when the mismatch is larger. One such example could be the scenario found in Simeone’s work where an umbrella was used as a substitute for a lightsaber. [26] If a user, either on purpose or unwittingly, opened the umbrella, it would become significantly harder to use for its purposes in the game (quickly swinging it around).

Another issue, which was also present in this study, is that unintended or unaccounted for usage could damage the physical props. For example, when users grab the cardboard boxes by their frayed edges, the damage to the edges is likely to increase.

In the cases mentioned, the user has taken advantage of affordances in the physical object not available in the virtual one, but opposite cases might be even more serious. If a real glass ball for example is given as the physical prop for a virtual rubber ball, attempts at bouncing it could even lead to injury.

5.3 Physical Exertion and Ergonomics

Participants were unfortunately not asked why they found the level of physical exertion surprising, so that remains unclear. A theory would be that they are unaccustomed to lifting relatively heavy objects while playing games. Even those with VR experience usually only lift the HMD and hand controllers around. Another contributing factor is the heat generated by the HMD quite easily causing sweating in the face. While the interviews did not ask participants about their feelings in regard to physical exertion, no strong negative emotions were expressed. While important
for health, many categorize exercise as a chore, and find it
hard to motivate themselves to do it. It has been shown
before that gamification can lead to increased motivation
for exercising. Most of these works however take a more
indirect approach, along the lines of recording exercise and
awarding points to participants for exercise completed.
[3][18] As the usage of a VRP is a more direct application
of gamification to exercise, the effects of these differences
would have to be studied to a greater extent.

All activities require careful consideration of ergonomic
aspects, but the consequences of failing to do so are much
greater when the activity is performed for a longer duration
or with a higher degree of physical exertion. VEs
incorporating VRPs are good examples where it is especially important to do so, as the addition of physical
objects can increase strain easily. Beyond taking
ergonomics into consideration in the design, this kind of
interaction might make it necessary to inform users about
correct and incorrect ways of interaction in the system in
order to avoid unnecessary strain and injury. This is a
negative side of this kind of tangible system and in the
cases tested here this posed serious issues similar to
ergonomic consideration in real life, although they used the
system for a very limited time.

However, while the specific interaction plays a role, much
can be done to avoid these problems through design as the
largest factor was probably the game design itself. One
remedy is to avoid bending interactions, as it appears as if
it was not the lifting itself that caused the discomfort, but
the bending over and bending one's neck up to look at the
task instructions from that position. Perhaps one suggestion
could be to make healthy posture a core part of the game
mechanics and hence necessary for mastery. As mentioned
in 4.6 Use Cases, additional sensors could make identifying
the user's posture possible and thus make it possible to for
example force or incentivize correct poses or punish
incorrect poses.

Of course, the physical object used for the VRP is also of
vital importance. Although it would probably be optimal to
determine the appropriate object on a case by case basis,
using smaller objects is an easy way to decrease physical
strain. However, there might be experiential drawbacks of
using small and light objects in conjunction with heavy
looking virtual objects that may affect how people perceive
the objects. In this regard it appears as if there might be a
design tension between using higher weight tangible
objects from an experiential and ergonomic perspective.

Concerning some participants, P8 and P10, deciding to
stand on their knees, it is hard to say what caused them to
do so. While no height data was collected, both of them
were approximately 180 cm tall, which was taller than
most of the other participants. Height could be the reason,
but the small number of participants further complicates
the making of any conclusions. One benefit of this posture
was the fact that a lower vantage point made it easier to
both look at the boxes and the instructions on the wall
without bending the neck up and down as much. Another
reason could be the increased stability as the differences
between the VE and reality can cause discomfort other than
nausea, as seen from the comment of participant P5. One
suggestion might also be to look into how the user’s height
could be used to adapt the games from an ergonomic
perspective. Especially since height data could be
automatically collected using the tracking system of the
HMD. This might however also require a different
configuration in the physical world depending on the game
mechanics.

5.4 Ease of Use

Based on user actions and reports, the specific interaction
itself seemed to be of little importance. Rather, it was a
combination of the surrounding factors that affected the
experience more as a whole. Tracking accuracy was a
factor in all cases, but for example for the Binary game,
distance and the VRP’s texture were problematic. For the
Rotation game it was instead the size of the boxes and the
large movements it required that was problematic at times.

5.5 Use Cases

Perhaps the biggest advantages of passive haptics is the
increased perceived presence in VR, while tangible
interaction provides the intuitivity accompanied with
handling familiar objects. These aspects were however not
brought up by participants at any point during the tests.
This is quite understandable as an increase in perceived
presence can be quite hard to notice, while also not being
very relevant to the game itself. Such an increase in
presence is in general more useful in situations where the
goal is to make the user suspend their disbelief or elicit
more natural responses to stimuli, for example in phobia
training in VR. [8] It has also been shown that increased
presence can lead to increased enjoyment of a VE. [30]
There was the mentioned comment from P3 regarding an
increased sense of presence in the Lift & Drop game, but
that was unrelated to the haptics, at least from the
participants own telling. Furthermore, the increased
intuitivity of handling familiar objects is another boon. But
it is still in passive haptics and tangible interaction the
VRPs have their advantage over regular VR.

Thus, cases making use of the increase in presence could
be any case where increased enjoyment is desirable.
Perhaps the most fitting case, which has already been
mentioned, is that of gamifying physical exercise or
rehabilitation, where the existence of the VRP also can
allow for the incorporation of tools of various weights and
sizes.

As for teaching, the technology has potential beyond the
suggested uses. It has been shown that tool appropriation in
VR translates to the real world, meaning that someone
could practice in the usage of a tool before using it in
reality, when practicing with the real tool isn’t an option.
[23]

The other suggested usage, as a tool for construction and/or
design is also supported in that using boxes as building
blocks is very intuitive. However, the building block
similarity is quite limiting and does not lend itself well to
other interactions. A change in VRP could nonetheless
open up these options greatly.
5.6 Limitations

The perhaps most limiting factor in the implementation of this project was tracking accuracy. While the study itself did not require perfect accuracy, the lack thereof affected the participants’ experiences greatly, both in regard to positional/rotational precision and delay. This lack of accuracy was caused by multiple factors. First in the chain are the markers themselves. They were attached by hand to the boxes. The boxes were also not precisely cuboid, resulting in a slight tilt at some sides at all times. Next is the camera used for tracking the markers. While it was good for a web camera, higher resolution and photos taken per second will yield better tracking. The output from the tracking algorithm itself also seemed to be somewhat unstable, although it is hard to determine if this was caused by the algorithm or the input provided. Finally there was the software calibration, which was done manually as there was not enough time to implement automatic calibration. Due to the manual aspect, the calibration had to be redone each time the camera was set up, resulting in not only some inaccuracies, but also different errors each time.

Another limiting factor is the fact that only one participant was female. As women have been shown to be more sensitive to cybersickness it could have led to lower reports of it. [19] However, as there were no reports of any cybersickness whatsoever and as no cybersickness was expected with the use of the Vive, that is unlikely to have been a factor. Still, it could have affected the tests in other unforeseen ways.

5.7 Future Research

This exploratory study brought attention to the many issues regarding VRPs that would require more quantitative research to properly understand. This includes but is not limited to how using VRPs affects enjoyment, effectiveness, efficiency and immersion. What are the effects on the user by a mismatch between the real and the virtual objects used for a VRP?

A regrettable issue with this study is the encompassing issue with accuracy. Replicating the study with more accurate tools might be able to bring smaller but nonetheless important issues to the surface. Changing other aspects such as the physical objects, virtual objects or the interactions themselves could also lead to interesting results.

6. CONCLUSION

Because of the exploratory nature of this paper, none of the findings have been more thoroughly researched, e.g. to compare with non-VRP systems or to check for significance in effect sizes. It is instead the hope of the researcher to inspire future research within the subject, and to provide a clear path as to where to go from now. It also became clear that when less than 10 minutes of actual usage led to complaints about strain, ergonomics have to take a central position in any design.

Ultimately, the purpose of this study was to answer the question:

What are the difficulties, challenges and opportunities in interacting with multiple tracked real objects in a Virtual Reality game?

Designing experiences that take into account good task design, affordance and ergonomics is certainly both difficult and challenging. It is only when these systems are smooth, comfortable and add value that they will have a chance to come into more widespread use. The technology is also not quite there yet, but is closing in at a very fast pace. Each new year brings great improvements upon previous systems, and the solutions to the biggest problem with the VRPs, the seamless tracking of the physical objects, without occlusion or stuttering, is surely not far off.

We are however already at a stage where a little more investing into the hardware and software used could give you a fully functional system capable of utilizing VRPs. As a VR system it is already able to provide great entertainment, but with the added possibilities of the VRP such as possible increase of immersion and easier integration with physical exercise. The final and perhaps most exciting opportunity is the ability to bring more intuitive interactions to the VR space, past keyboards, mice and controllers.

7. ACKNOWLEDGMENTS

The author would like to thank his advisors, Anders Lundström at KTH and Johan Hallberg at Capgemini for their support throughout the project. Thanks also to Ken Olling and Ludwig Sidenmark for additional advice and help.

8. REFERENCES


[29] Unity 3D. Retrieved June 15 from [https://unity.com/](https://unity.com/)


