Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Civil and Environmental Engineering

Eirik Osnes

Thematic Maps in Virtual Reality

Accuracy of Visual Variables in an Immersive Virtual Environment

Master's thesis in Engineering and ICT
Supervisor: Terje Midtbø
June 2020

NTNU
Norwegian University of
Science and Technology

Eirik Osnes

Thematic Maps in Virtual Reality

Accuracy of Visual Variables in an Immersive Virtual Environment

Master’s thesis in Engineering and ICT
Supervisor: Terje Midtbø
June 2020

Norwegian University of Science and Technology
Faculty of Engineering
Department of Civil and Environmental Engineering

NTNU
Norwegian University of
Science and Technology

Master thesis

(TBA4925 - Geomatics, Master thesis)

Spring 2020
for
Eirik Osnes

Thematic Maps in Virtual Reality

Usability of Visual Variables in an immersive virtual environment.

BACKGROUND
What variables visual variables exist for use in thematic maps today, and which of these have been, or could be used for VR applications?
What effects does a 360 degree space have on awareness of ones surroundings and their ability to recollect information?
What are the technical characteristics of VR headsets, and the Oculus Quest in particular?

TASK DESCRIPTION

This project aims to ascertain the usability of different visual variables when the viewer is in an immersive virtual environment, facilitated by the use of an Oculus Quest. The task for the author is therefore to determine which variables to study, create the application for this purpose, have test persons perform these tests, and then review the results of this.

Determine a set of variables, and an experiment
Create an application for the Oculus Quest, in which testing will be undergone and data collected.

Specific tasks:

- Study related literature as noted in Background.
- Determine the specific variables to be tested in the experiment.
- Determine the statistical tests to be done, and the data needed to perform these.
- Create an application capable of testing and logging the required data.
- Run pilot tests to confirm the functionality of the application.
- Improve on any problems found during pilot testing.
- Run the final experiments on a number of test persons determined by the statistical tests.
- Run statistics and determine results.

ADMINISTRATIVE/GUIDANCE

The work on the Master Thesis starts on January 15th, 2020

The thesis report as described above shall be submitted digitally in INSPERA at the latest at June 11th, 2020.

Supervisors at NTNU and professor in charge:
Terje Midtbø

Trondheim, February, 2020

NTNU
Faculty of Engineering
Department of Civil and Environmental Engineering

Date	Our Reference
14.5.2020	
Your date	Your ref

Extraordinary circumstances due to the Corona pandemic

This Master thesis is based on work that was accomplished in the spring semester 2020. In this period the Corona pandemic was active and influenced the work of several master students. The grading of the thesis must take the pandemic situation into consideration.

If this Master thesis is affected by the Corona pandemic, the student will point out the influenced elements in the beginning of the report. More details about this may also be explained later in the thesis.

Terje Midtbø

Terje Midtbø
Professor

Address	**Org. no. 974 767 880**	**Location**	**Phone**	**Executive officer**
		Høgskoleringen 7 A	+47 73594640	Terje Midtbø
7491 Trondheim	postmottak@iv.ntnu.no			terjem@ntnu.no
Norway	www.ntnu.no/ibm			Phone: +4748606490

Abstract

With virtual reality (VR) becoming more commonplace in private and business alike, finding what possibilities exist for the technology becomes important. Cartography as a field has through centuries conformed to the technology available, to best convey the information confined in a map to its reader. This new paradigm of technological interfaces should be no different. The question remains however how this technology and the experience of being immersed and surrounded in the environment influences the ability to recognise the variables used to convey such information. In this paper, the visual variables of height and achromatic colour will be specifically selected as variables of interest.

Reviewing literature within the field of virtual reality, thematic mapping and human cognition, this paper aims to get an overview over the effects likely to affect the results of a visual search in a virtual environment. An experimental setup is proposed based on the *delayed match to sample* experimental paradigm, utilising an Oculus Quest VR headset. The user is first exposed to their target building, with its given height and colour, and shall so attempt to find and select this among a group of distractors, varying either height or colour. To find the effects of being surrounded, all tests are undergone both where the search set is fully within the field of view, and where evenly spaced a full 360 degrees around the user. The question is thus what effect this change of placement has on the accuracy of a visual search for the chosen variables.

With experiments impossible to perform due to the COVID-19 pandemic, no findings are presented. The only results presented are repeated tests undergone by the author and thus non-representative. These serve only to show the capabilities of the created application and the statistical methods that could be utilised to assess the data.

With a wide variety of potential variables identified in the paper, combined with a flexible and open-sourced application created for this purpose, the paper might serve as a starting point for further research within the field.

Sammendrag

Med virtuell virkelighet (Virtual Reality, VR) i stadig større grad vanlig både for privat og kommersiell bruk, blir det viktig å undersøke hvilke muligheter teknologien byr på. Kartografi som fagfelt har gjennom århundrer tilpasset seg tilgjengelig teknologi for å best mulig formidle informasjonen vist i kartet til leseren. Med dette nye paradigmet for digitale grensesnitt bør det samme være tilfellet. Spørsmålet som da stilles er hvordan denne teknologien og det å være omringet av det virtuelle miljøet påvirker evnen til å gjenkjenne variabler og informasjonen de gjengir. I denne oppgaven blir de visuelle variablene høyde og akromatisk farge sett på spesifikt som interessevariabler.

Ved å gjennomgå litteratur innen fagfeltene VR, tematiske kart og menneskelig kognisjon, vil denne oppgaven forsøke å gi et innblikk i effektene som trolig vil påvirke resultatene av visuelle søk i et virtuelt miljø. Et eksperimentelt oppsett er foreslått basert på det eksperimentelle paradigmet *delayed match to sample*, benyttende et Oculus Quest VR headset. Brukeren blir først vist målbygningen med dens høyde og farge, for så å forsøke å gjenkjenne og plukke ut denne i en gruppe distraktører hvor enten høyde eller farge varierer. For å finne effektene av at brukeren er omringet av distraktører, blir alle testene gjennomført både hvor hele søkesettet er i synsfeltet, og hvor de er spredt jevnt 360 grader rundt brukeren. Spørsmålet er dermed hvilken effekt denne endringen i plassering har på treffsikkerheten til de visuelle søkene for de valgte variablene.

Da eksperimentene ikke kunne gjennomføres som følge av COVID-19 pandemien, er ingen reelle resultater presentert i oppgaven. Noen repeterte tester gjort av forfatteren blir presentert, men er ikke representative. Disse fungerer kun for å vise funksjonalitet av applikasjonen laget for oppgaven, samt statistiske metoder som kan benyttes til å vurdere dataene.

Med mange ulike potensielle variabler identifisert, kombinert med en fleksibel og fritt tilgjengelig applikasjon laget for formålet, kan oppgaven fungere som et utgangspunkt for videre forskning innen fagfeltet.

Preface

This paper is written as the master thesis ending a 5-year study at The Norwegian University of Science and Technology (NTNU), where I have been studying Engineering and ICT, with Geomatics as the chosen speciality. Thanks go out to everyone who has made the completion of this study possible, with special thanks to Terje Midtbø, my supervisor, for his support.

Due to the COVID-19 pandemic neither the experiment nor pilot testing could be performed. Where this is relevant in the paper, it will be further noted.

- Eirik Osnes, the author

Contents

Figures

Tables

Abbreviations/Symbols

VR	Virtual Reality
VE	Virtual Environment
AR	Augmented Reality
MR	Mixed Reality
HMD	Head-mounted display
FOV	Field of View
EH	Eye-Height
DMS	Delayed Matching to Sample
RQ	Research Question

1 Introduction

Cartography is as ancient as human civilisation, and it is ever evolving. No longer is wall painting or rock carving the preferred method of recording. As the technology progresses, so does the field of cartography, with the maps becoming ever easier to create, while more informative and interactive, and through work from people like Bertin and MacEachren, the idea of thematic maps has become solidified. For many years now, digital maps have to an ever increasing degree come to stay – gone are the days where every household needed atlases to see a world map, or enormous road maps to find their way on a road trip. Today both of these tasks are relegated to digital technology such as Google Maps or a navigational GPS. This change is not limited to navigational maps, however, with thematical cartography also moving over to this new domain, allowing for easily showing viewers thematic information from the comfort of their own device, and by interaction exploring causes and effects. Therefore, when technology further progresses to a place today where Virtual Reality (VR) is affordable and increasingly widespread, the question remains how this new and exciting technology can be utilised in cartography.

VR is a technology that allows the user to be placed in an environment that seems real, although entirely virtual. Using a device such as a head-mounted display, all visual sense of the real world and only the virtual world can be viewed. This technology serves as an incredible advancement in interaction, with many usages within games, education, and sciences alike utilising this. Using VR one can view a fictional cityscape as if it were entirely real, and possibly shaping said cityscape as one sees fit through use of controllers or simply one's hands. The implications this has for the field of cartography are not fully known, and the possibilities for applications might be truly diverse.

Geographical Information Systems, or GIS, are today the main way of creating maps, allowing for easy iteration of design decisions. Today, with the ever increasing interest in VR, GIS providers are working to include VR functionality with the GIS directly, such as with ESRI CityEngine (Arisona, 2018). While such technologies can undoubtably create realistic looking city models and are also capable of producing thematic maps, little research seems to have been done with regards to determining the usability of the traditional thematic variables in the context of an immersive virtual environment. Contributing to this research is thus precisely the goal of this paper.

1.1 Research Questions

What influence does having targets placed in a 360-degree environment rather than within the user's field of view have on accuracy of a visual search

1. *when varying height?*
2. *when varying achromatic colour (greyscale)?*
3. *when both height and achromatic colour is considered?*

1.2 Reader Guidance

In this paper, theory will be presented, which will then be utilised to determine the design of the application used to gather data for the experiment, and the specific choices made for said experiment. This experiment was then intended to be performed, the results of which would then be discussed, and conclusions made. Unfortunately, due to the COVID-19 pandemic, performing these experiments was both unfeasible and deemed unethical. As a result, the results section is undergone using faux data as a means to determine functionality but allowing no conclusions to be made from this. All discussion are hypotheticals based on the theory presented. Little prior knowledge of cartography, VR or cognition is necessary for the reader, with relevant theory being presented in the paper.

2 Background

2.1 Virtual Reality

2.1.1 Specification

While virtual reality today is a well-known term, a multitude of definitions have been put forward. In Oxfords "A Dictionary of Marketing" the following definition was put forward:

> A lifelike artificial environment with various online applications such as computer games, simulations for training purposes (for airline pilots, for example), virtual tours, animations, architectural design, and advanced advertising. (Doyle, 2016)

Other definitions focus on the interactivity of a virtual reality application, such as the case is the definition from Merriam-Webster:

> 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 (Merriam-Webster, n.d.)

The term "Virtual Environment" or VE has also been widely utilised in academic literature, most times interchangeably with VR. As discussed by Bryson (2013), this debate was mostly ended in the 1990s, with VR being the preferred term. He further discussed how earlier definitions widely used the technology utilized to create the effect as part of the definition of said effect, and proposed the following definition:

> Virtual Reality is the use of computer technology to create the effect of an interactive three-dimensional world in which the objects have a sense of spatial presence (Bryson, 2013)

For the purpose of this paper, the latter definition shall be used, although the application created focuses little on interaction, but rather the spatial presence of objects.

Other immersive technologies also exist, Augmented Reality (AR) and Mixed Reality (MR) being most discussed. These utilise different amounts of a real environment in the experience, by overlaying virtual objects in a real-world environment (AR) or interacting with both real-world and virtual environments (MR) (Milman, 2018). Generally this can be seen in the context of a "virtuality continuum", being a scale between the polar opposites of the real environment and a fully virtual environment (Milgram and Kishino, 1994). A simplified visualisation of this can be seen in Figure 1 below.

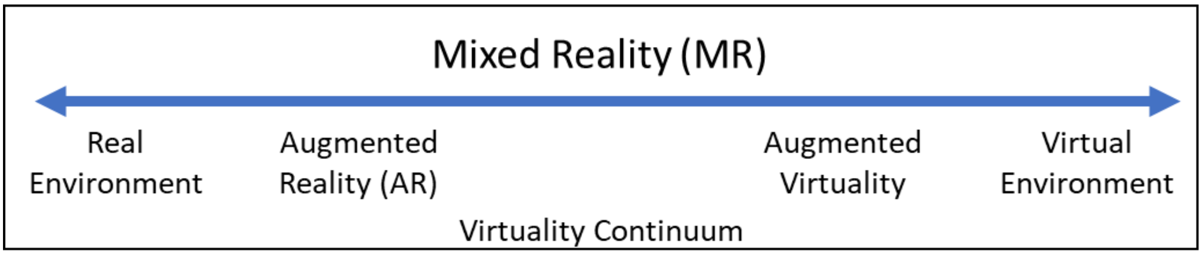


Figure 1 Simplified representation of a "virtuality continuum"

Looking only within the limited scope of fully virtual environments, there exists a general distinction defined by the means of which this effect is achieved, namely *immersive* and *non-immersive* VR (Suh and Prophet, 2018).

Non-immersive VR is achieved through the usage of traditional interfaces such as computer monitors, keyboards, and mice. The defining factor here being that the user is not fully immersed, by not interacting directly with objects as one would in real life and by viewing the world through a 2D or near-2D screen. This definition could include most digital interactable media. Examples of this are most computer games, for instance Minecraft (Suh and Prophet, 2018), and would also include traditional 3D maps used widely in cartography.

Immersive VR aims to increase immersion by use of more complex tracking systems. Generally, these technologies include a device tracking the users head movements and providing full peripheral vision of the virtual environment. This peripheral view can be achieved by using a stereoscopic screen surrounding the user, as with CAVE (Wijayasekara, Linda and Manic, 2011), or through limiting the peripheral view with a Head-mounted display (HMD), such as an Oculus Quest, Valve Index or HTC Vive. Interaction can be achieved by a multitude of tools, depending on the application. Commonly it is achieved through the use of handheld controllers tracking the location of the user’s hands. More advanced tracking is also available, with gloves or cameras tracking the exact movements of the user’s hands, allowing for interaction without a controller (Oculus, n.d.).

2.1.2 Oculus Quest

After years of development, millions of dollars in crowd- and venture funding and a buyout by Facebook, Oculus released in 2016 their first VR headset, the Oculus Rift. Since then several others have sprung up, both competitors and other Oculus headsets, each with their own price points, use cases and specifications. Overall the VR market was valued at 7.3 billion USD in 2018 with a projected value of 120 billion by 2026 according to Fortune Business Insights (2019).

In 2019 Oculus released a new headset, Oculus Quest. This headset offered immense freedom, being an untethered, standalone device. This was obtained through having the Quest be its own device running a version of the Android operating system, and using four angle sensors to spatially track the headset (Hardawar, 2019). Due to availability this was chosen as the preferred device for this experiment. This should not detract from the choice however, with as much as 49% of VR headsets sold in 2019 being Quests (Rogers and Lyons, 2020). As such the results of this paper should be usable to a wide range of users, based on the availability of the hardware.

The Oculus Quest’s specifications collated from different sources can be found in the table below (Oculus, n.d.; Hardawar, 2019; Higham, 2019):

Table 1 Oculus Quest specifications

Display	1440 x 1600 pixels OLED per eye
Refresh Rate	72 Hz
RAM	4GB
Storage	64GB / 128GB
Processor	Qualcomm Snapdragon 835
Movement	6 degrees of freedom
Weight	571 g
Battery Life	2-3 hours
Connection Interface	USB C

2.1.3 Current Applications

Many modern VR devices were developed with the express purpose of being a gaming device, but different fields have found their own uses for the technology. In fact only 43.8% of the overall VR market share in 2018 was attributed to the gaming industry (Fortune Business Insights, 2019). Suh and Prophet (2018) performed a literature analysis in 2018 attempting to map and aggregate the research that has been done using immersive technology. Their findings include overviews for domains of use; notably gaming, learning and training, psycho- and physiotherapy and virtual tours, as well as collating the stimuli and reactions commonly associated with the technology. They also note that the number of studies on immersive technologies is increasing. Other use cases can be found in literature, including fields like urban planning (Nguyen *et al.*, 2016), visual data mining (Nagel, Granum and Musaeus, 2014; Wijayasekara, Linda and Manic, 2011) and public health (Kamel Boulos *et al.*, 2017). Within the realm of cartography the use cases are also increasing, with programs such as ESRI CityEngine now capable of creating interactable VR cityscapes (Arisona, 2018) as well as Google releasing their VR version of Google Earth (Google, 2016).

2.1.4 Usability of Virtual Reality

There are several reasons for why one should use or avoid using immersive technology and knowing and designing around these is paramount to a successful application. Focusing on the responses found by Suh and Prophet (2018), and defined as the consequences of immersive technology use, we can find general trends for both positive and negative outcomes.

Positive responses

The positive outcomes of immersive technologies define the main factors for why one would wish to create an application of this kind. Maximising these effects or adding to them should be the goal of any application. These responses include learning effectiveness, learning engagement, learning attitude, task performance, intention of use, and in medical applications: reduced disease symptoms (Suh and Prophet, 2018).

Learning effectiveness refers to whether a person gains the intended skills or knowledge as part of a learning activity (Suh and Prophet, 2018). Research has shown that immersive technologies can increase learning effectiveness in AR (Ibanez *et al.*, 2016), and using desktop based, and thereby non-immersive VR (Merchant *et al.*, 2014).

Perceived learning effect can be increased through use of VR (Zhang *et al.*, 2017). Keehner *et al.* (2008) performed a study to ascertain if increased effectiveness in medical education using VR was due to interactivity within the virtual environment, or through viewing of key angles, by passively watching a recording of other another person's interaction with said environment, and found no significant difference. This result is disputed by Jang *et al.* (2017), who through a similar study found a significant increase in learning effectiveness through interaction. Other findings note that the perceived effectiveness of VR depend on the prior knowledge of the test subjects (Taçgın, 2020), both regarding the subject topic and the usage of the application, as well as their spatial abilities (Lee and Wong, 2014). As such the positive effect of VR for learning effectiveness is closely connected to the design of the application and its target audience. This observation is further strengthened by the work of Jou and Wang (2013), who found significant effects of learning in a Virtual Reality Learning Environment for some teaching objectives and not others when teaching technical skills.

Task performance is defined by Suh and Prophet (2018) as an increase in efficiency and accuracy. Most research done for this topic is related to enhancing a task through AR in order to reduce the error rate, typically in medical settings (Grobelski, Walczak and Pasieka, 2010; Zhao, Ong and Nee, 2016) or manual assembly (Radkowski, Herrema and Oliver, 2015); or increased real world task performance after training in a virtual environment (Lendvay *et al.*, 2013). Little research could be found regarding increasing the performance within the virtual environment itself, as would be the case when finding best interaction or visualisation options for a task performed fully within such an application. Lin *et al.* (2015) performed a study to measure task performance utilising different methods of VR displays, where HMD fared worse than both 3D displays, and projection displays for both movement time and throughput. Task performance can also be hampered if insufficient resolution is available to read task-significant text (Kim and Shin, 2018).

While learning engagement, learning attitude and intention of use are positive effects of using VR, this is more a question to *why* to create such an application, than a focus on the effects of it. This is absolutely of concern for any application wishing to apply findings from papers such as this, with regards to whether a VR application is the correct approach but is largely irrelevant to the scope of this paper.

Negative responses

The negative outcomes of immersive technologies define the main factors for why a user would not wish to use the application or why the user might fail to learn from or interact with the application to a satisfactory degree. Among these is factors like motion sickness and physical discomfort (Suh and Prophet, 2018).

Physical discomfort with HMDs are generally related to wearing a front-heavy apparatus on one's head, leading to strain on the neck (Kim and Shin, 2018), but can also relate to display configurations conflicting with the human visual system, leading to strain on the eyes (Koulieris *et al.*, 2017). Generally, however, physical discomfort is hard to mitigate without changing which hardware is used. As such it is important to recognise it as a problem, but more nuanced understanding is unnecessary.

Motion sickness is usually manifested as nausea and vomiting, and while commonly thought of with regards to sitting in a moving vehicle, such as an aeroplane, boat, or car, this is only one subset of motion sickness. A person can also feel motion sickness while they themselves are stationary. This can happen when there is a perceived motion from visual stimuli – such as when moving around a virtual environment. This is commonly called Visually Induced Motion Sickness, or VIMS (Hettinger and Riccio, 1992), and can lead to similar symptoms as traditional motion sickness. VIMS and motion sickness are also known under other names, with increasing specificity such as simulator sickness (Mourant and Thattacherry, 2000; Serge and Fragomeni, 2017), or cybersickness (Laviola, 2000; Rebenitsch and Owen, 2017), but these are largely relating to the same phenomena. For the purpose of this paper, these will be collectively referred to using the term motion sickness.

What specifically causes motion sickness in VR is disputed. One theory is Sensory Conflict Theory (Reason and Brand, 1975), where different senses are telling the brain conflicting information, that the body is not prepared to handle. Other theories include the Postural Instability Theory, where the subject cannot keep a comfortable posture in relation to the environment due to rapid changes (Laviola, 2000), and the Poison theory, where conflicting information is perceived by the body as hallucinogens, and thus poison to be ejected from the body – leading to nausea (Carvalho *et al.*, 2017).

Regardless of which theory is correct, if any, a large amount of discrete factors have been identified and tested in order to minimise the effect of motion sickness, with more than 40 proposed by Kolasinski (1995) and Renkewitz and Alexander (2007). These factors include user characteristics such as age, postural stability, and experience; system characteristics such as display refresh rate and system lags; as well as task characteristics like movement, visual content, and field of view; and interaction paradigms.

While many problems relating to discomfort and motion sickness is hardware-related, some problems can be mitigated through application design. Kim and Shin (2018) noted that increased need to rotate one's neck using an HMD was cause for physical discomfort, and thus having tasks not require this movement might also lessen the discomfort. Similarly motion sickness in VR seems to be tied to the simulation, with less movement within the application translating to less motion sickness – a VR video of a beach leads to less motion sickness than one where the user is on a roller-coaster (Somrak *et al.*, 2019). Field of view (FOV) has been shown to be a major factor for motion sickness, with a doubling of FOV leading to a doubling of induced motion sickness (DiZio and Lackner, 1997), and similarly a halving of FOV halving the induced motion sickness (Stoffregen *et al.*, 2008), and Rebenitsch and Owen (2017) observed the effect to be independent of the display screen size. Whether the subject is standing or sitting is also a seemingly important factor, with a standing person without a railing or similar to hold on to is more likely to experience motion sickness than a sitting individual (Merhi *et al.*, 2007; Moss and Muth, 2011).

2.2 Thematic Maps

In the project preceding this paper, a review of background information in regards to thematic maps was undergone (Osnes, 2019). This section will largely restate much of said information, updated and reviewed where new or updated information was found or made available. Of particular note for restate is subsections 2.2.1 and 2.2.2, as well as depth perception in subsection 2.2.3.

Thematic maps give a means to visualize spatial patterns of geographical features (Slocum *et al.*, 2009). This might be used to describe such patterns as population density, wind speeds, or other data not obviously tied to one specific location. This can be seen as a contrast to general-reference maps, where the location of a phenomena is central, such as maps used for navigation.

The usage patterns of thematic maps can be generally classified into three different groups (Slocum *et al.*, 2009):

- To provide specific information about locations
- To provide general information about patterns
- To compare patterns for different maps

All of these utilize the same tools to convey their information, albeit with different focuses. These tools are called graphic variables.

2.2.1 Graphic Variables

In the 1970's Jacques Bertin compiled a set of variables to convey statistical information in a map, these being *location, size, value, texture, colour, orientation* and *shape.* (Bertin, 1981) This was done based on theory from general statistics, where information could be found to greatly enhance the readability of statistical information. Since then a large group of cartographers have expanded and refined the visual variables proposed by Bertin. MacEachren is among the best-known cartographers among these, with his work formalising more static visual variables, as well as expanding the overall graphic variables to include *animation* variables (MacEachren, 1995). Further variables have also been proposed, and Table 2 collated by Judith Tyner (2010) intended to show the current *most agreed upon* variables. As a result, it gives a good overview of variables, but is not exhaustive.

Table 2 The graphic variables as collated by Tyner (2010)

Visual	Sound	Animation	Tactile	Haptic
Size	Location	Duration	Volume	Pressure
Shape	Loudness	Rate of change	Size	Spatial acuity
Hue	Pitch	Order	Value	Position
Lightness	Register	Display date	Texture	Texture
Saturation	Timbre	Frequency	Form	Hardness
Pattern	Duration	Synchronization	Orientation	Temperature
Texture	Rate of change		Elevation	
Location	Order			
Orientation	Attack/decay			

Although sound, animation, tactile and haptic variables undoubtably are relevant in the discussion of Virtual Reality, these will due to scope not be evaluated comprehensibly in this paper. Instead the paper will focus solely on the visual graphic variables.

2.2.2 Visual Variables

In a 2D map, only three different markings may be used to symbolise data, these being *points, lines* and *polygons* (Tyner, 2010). With modern technology, creation of maps in 3D and even 4D has been made relatively easy by displaying the map digitally. As such we naturally extend the previously mentioned markings to include a 3D *volume.* Being limited to the location and size of these markings to convey information are very limiting, and as such we combine these with additional variables to differentiate the symbols and shapes used (Slocum *et al.*, 2009). By restricting ourselves to only use the visual sense, as is common in maps, we are limited to using the visual variables. Working in three dimensions, it is natural to think that the environment should accurately portray reality where such a reality exists, as to create natural metaphors for navigation and interaction in said environment. This realistic representation may, however, not be suited for the purposes of exploring geospatial information. Rather, the rules for generalisation and abstraction of variables utilised in two-dimensional mapping might very well also apply in three (MacEachren *et al.*, 1999).

Static variables in thematic maps are the oldest and most commonly used within modern cartography and consists of the Visual variables listed in Table 2, among others proposed. These variables were designed as a means to distinguish and convey information in a traditional 2D paper map, but they have since been adapted into modern digital maps. Of these, the shape of the object is the primary distinguishing variable, whereas the remaining visual variables aim to give information about the phenomena symbolised (Tyner, 2010). As discussed in the preceding paper to this, location and shape are difficult variables to change for the purposes of signifying information, due to them being distinguishing variables and mandatory to find positional patterns (Slocum *et al.*, 2009). Likewise, size is difficult in a virtual environment if the object resized is to be perceived as a real object. This does not mean size is impossible to use. Both the perspective height described by Slocum *et al.* (2009) and the spatially semi-iconic VE by MacEachren *et al.* (1999) describe the idea that one axis in the virtual environment can be utilised to describe information. This thus gives a way of utilising size as a variable in the virtual environment.

2.2.3 Fidelity of Visual Variables in VR

In order to accurately ascertain the usefulness of visual variables in VR, and the ability to discern between two values, it is important to understand the related physical, psychological and technological characteristics.

Resolution

The technological characteristics of interest in this context are regarding the resolution of the display, both regarding the display resolution, but also the colour. The display resolution is generally defined as the number of pixels on a display, with the pixel size being determined by the resolution and the display size. Using the Oculus Quest, the display resolution is 1440 x 1600 pixels, with one of these displays per eye, this is listed in Table 1. Display resolution becomes important when the pixel sizes are large compared to the information it tries to show. Kim and Shin (2018) has shown that performance can be hampered if resolution is too low to read task-significant text, showing the need to understand this dimension. Using the Oculus Quest, this resolution should be enough to allow for readable text, as well as accurately portraying the visual variables not related to colour, so long as the tests and texts are made to accommodate.

The colour resolution, or *colour depth*, is harder to determine for the Oculus Quest. No official sources seem to disclose the colour depth, and with the testing of this being outside the scope of this paper, it has to be treated as an unknown. Commonly in computer screens, a 24bit colour depth is utilised. This means that for each of the colour dimensions available (RGB: Red, Green and Blue), 8 bits, and thus 256 distinct colours, can be created. Adding different levels of red, green, and blue, and one can create close to any colour – with a total of 16 777 216 distinct colours possible. As such, going by the assumption that the Oculus Quest can display 24bit colour depth, it is likely that the colour depth will not be the limiting factor in discerning colour differences.

Perceptual scaling

More likely the limiting factor will be a physical characteristic of the human biology. Humans are not equally capable of discerning between any two colours. Simplifying to only looking at the greyscale case, it is easier to discern between pure white (R,G and B all 255 in a 24bit display) and pure black (R, G and B all 0 in a 24 bit display), than it is to discern between two shades of grey. The closer these two shades are to each other, the harder it is likely to be discerning between them. However, it is not only the distance between the shades in absolute RGB steps that determine the similarity. The human perception of shades of grey is not linear, but logarithmic, meaning we are generally more likely to observe differences between two dark shades, compared to two light shades with equal distance, as described by the Weber-Fechner Law (Fechner, Howes and Boring, 1966). As such we can say that while the absolute difference of two shades might be constant, the similarity of the two are not. Having been first published in 1860, much research have succeeded this law, resulting in laws such as Steven’s power law, claiming the relation is a power rather than logarithmic, although the effect is similar (Stevens, 1957). These laws are not just for achromatic colours, however, as it is proposed that these holds true for changes in stimuli perceived by any sense.

In cartography, there are well known psychological effects in play when using size as a variable. In maps using proportionally scaled circles to symbolise values, humans tend to underestimate the size of larger circles (Flannery, 1971), thus following the general principle described in the Weber-Fechner law and Steven’s power law. As such the idea of perceptual scaling was proposed, where the larger symbols would be scaled so the values are *perceived* correctly, even if the absolute sizes are inaccurate. The scaling factor for this perceptual scaling was found through empirical experiments, and a visualisation of this scaling can be seen in Figure 2. This effect is not general for all shapes, however. Bars are found to be perceived linearly, meaning a mathematical (absolute) scaling can be applied, while volumes are more susceptible to underestimation (Flannery, 1971; Krygier, 2007; Jansen and Hornbæk, 2016). These can be seen as specifications of the Weber-Fechner law. A visualisation of the relation can be seen in Figure 3. This seems to be generalisable to three dimensions, where solid bars are encoded by length and spheres by surface area (Jansen and Hornbæk, 2016). Later works have called into question the usage of perceptual scaling, with authors like Edward Tufte effectively saying only absolute scaling should be used (Tufte, 2001). The fact remains, however, that this discussion is based on an idea of 2D cartography with traditional legends. In an immersive virtual reality application, this might not be applicable. Thus it remains to be seen if similar effects appear in such an application.

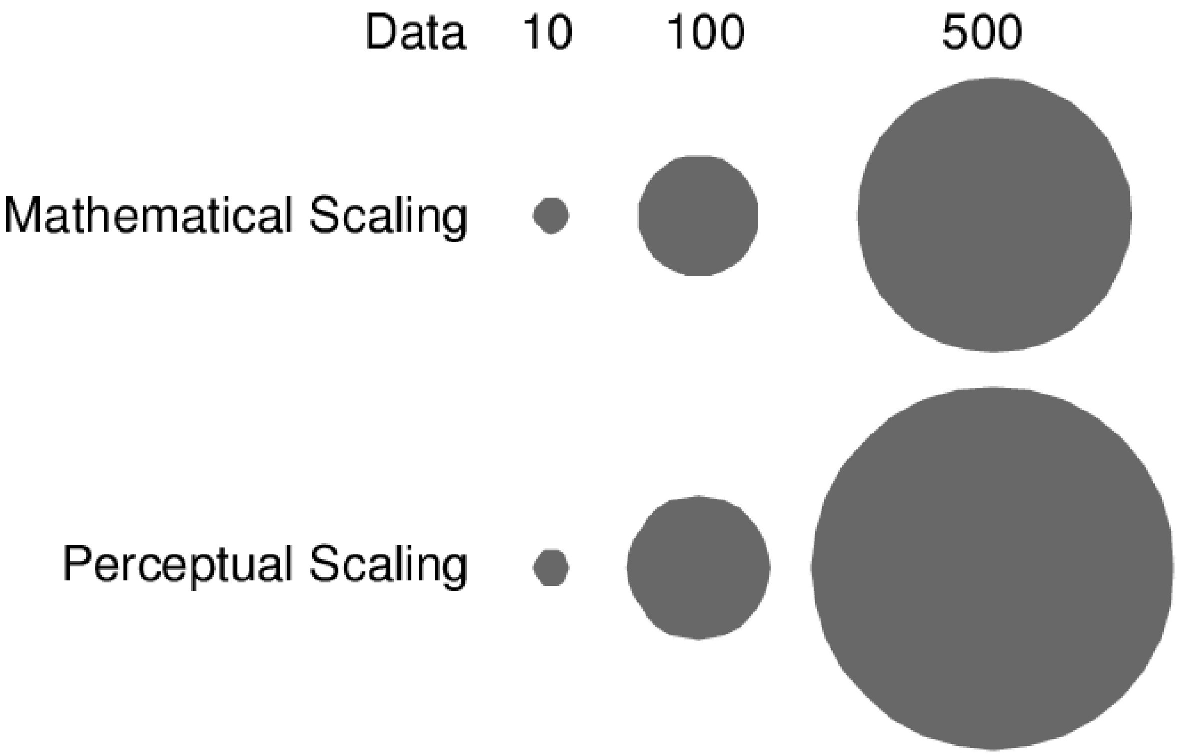


Figure 2 Mathematical vs. Perceptual Scaling of circles (Susumu, Chusi and Tsutomu, 2006)

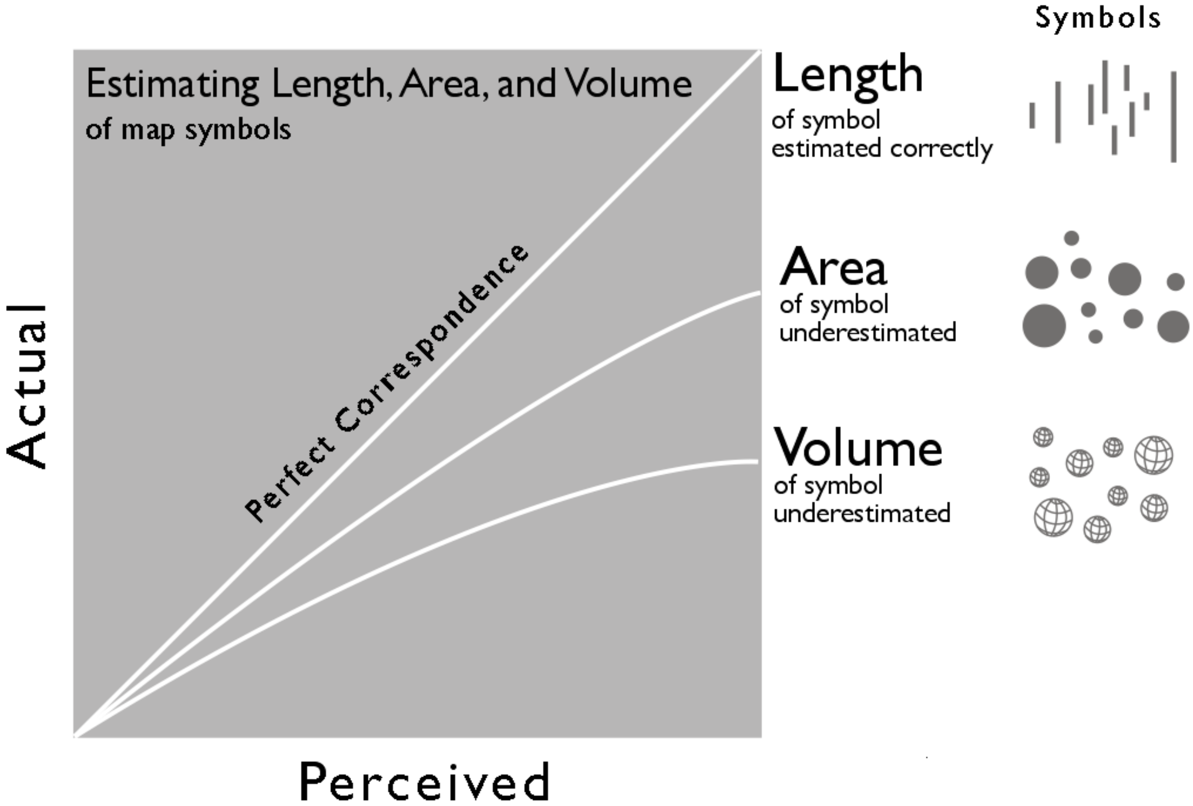


Figure 3 Perceived values for different dimensions. (Krygier, 2007)

Effects of Height

The estimation of height might be further complicated for taller buildings as a result of perspective. Humans have a limited vertical field of view, and as such cannot see the complete height of a building exceeding this view without tilting our heads. Doing this, especially in conjunction with moving closer to said building, allows us to view more of the building simultaneously, but lead to perspective distortions related to depth perception. As is visible in Figure 4, in this "worm's eye view" details higher on the building is compressed. Similarly, this might lead to the total height being harder to estimate, with heights of objects largely being overestimated, the degree to which is dependent on whether one views a height from the bottom (worm's eye view) or from the top (bird's eye view) (Stefanucci and Proffitt, 2009). The latter of these are shown to induce a significantly higher overestimation, possibly due to an innate fear of heights, though this is disputed (Stins *et al.*, 2013), with this effect being higher for higher viewing positions.

Another complication with regards to estimation of heights come from the idea of eye-height scaling. Eye-height scaling refers a tendency for humans to better estimate the height of an object, if said height is close to the eye-level of the observer (Wraga and Proffitt, 2000). This effect is similar for both sitting and standing individuals, with their respective eye-height. Wraga and Proffitt (2000) tried to determine the limits for which this effect is beneficial, finding a band between 0.2 and 2.5 EH (1 EH = eye-height of observer). Bingham (1993) had participants determine heights of trees between 2.7 EH to 16 EH and found limited use of eye-height scaling for taller trees, with tree form properties increasingly being the main factor for estimation. Further, this positive effect on estimation is only achieved if the viewer is situated on the ground (Wraga and Proffitt, 2000). This effect means that, by controlling the sizes of objects to be in a certain range, another tool for determining height is available to us, and as the objects deviate from this range, the more important other cues become. Twedt, Crawford and Proffitt (2012) corroborated the findings that users tend to push the perceived height away from their eye-height, by underestimating heights <1 EH and overestimating heights >1 EH. They further found this effect to be persistent in memory, meaning this effect might be important in a delayed search. While these under- and overestimation effects are mostly researched in regard to determining absolute height, it is not impossible that this effect might exist also when working with relative heights.

Figure 4 Perspective distortion occurs with tall buildings.

Depth Perception

As thematic mapping moves out of the 2D realm, and into a 3D world, depth perception becomes an important attribute of the human senses. In a study performed by Matatko et al. (2011), different factors was tested in a Virtual thematic cityscape in order to determine depth perception. Depth perception is driven by a set of depth cues and gradients, all of which work together to give the right impression of distance. These are listed in Table 3.

Table 3 Properties affecting depth perception, as mentioned by Matatko et al. (2011)

Oculomotor Cues	
Depth of Field	Objects further away from a fixation points are increasingly blurred.
Pictorial Cues	
Overlap	Topological alignment of depth as occluded objects appear further away.
Size in Field of View	A larger surface in the field of vision leads to a larger estimated object size.
Height in Field of View	Objects higher up in FOV are perceived further away.
Atmospheric Perspective	Particles in the air leads to reduced contrast on objects further away. Related to Contrast and Colour Gradients.
Familiar Size	An observer will from experience know the relation between distance and size, and thus can estimate one when knowing the other.
Linear Perspective	Parallel lines converge to a vanishing point. Related to Texture Gradients
Motion-Produced Cues	
Motion Parallax	Near objects appear to move quicker than far objects.
Deletion and Accretion	Objects at different distances will appear to move relative to each other. Deletion: Increased occlusion Accretion: Reduced occlusion
Binocular Disparity	Depth perceived from different visual stimuli from the different eyes – stereoscopic effect. Only applicable to near objects.
Depth Perception Gradients	
Texture	A combination of Size, Density and Form Gradients on textures.
Size	Size of objects become smaller with increasing distance.
Density	Objects appear to get closer when further away, thus increasing density.
Form	The shape of objects are distorted towards a vanishing point
Contrast	As distance increases, contrast of objects decreases, like a haze.
Colour	As distance increases, the colour of objects will be increasingly changed depending on the colours scattered by particles in the air.
Continuous Brightness	Gradual distortion and deformation through cast shadows and reflections.

Matatko et al. specifically showcase the effects of Texture Gradients – where having a ground texture can help gauge the distance to objects, Depth of Field – where blurring distant objects gave better accuracy to closer objects, but with users preferring a sharper image, and Contrast Gradient. For the latter results were varying, with a dense haze increasing accuracy in depth perception, but haze also clearly increasing the response time of participants. Further they comment that Motion Parallax is an essential tool to depth perception. Other research also suggest that pictorial cues, primarily familiar size, can lead to a significant improvement in depth perception (Ng, Chan and Lau, 2016). According to Renner, Velichkovsky and Helmert (2013) this is debatable, with only a weak influence on depth perception afforded by familiar size.

While depth perception, and the related task of estimating distance is likely to be more important in a navigational task where this information is of great importance, it can also be a significant factor to the usability of thematic maps. Seeing as an important task of thematic maps is to show spatial pattern, having an accurate idea of the distance between these objects could be important. Further, it ties implicitly in the understanding of the visual variable location, and referring to Table 3, all the visual variables listed in Table 2 will be affected by at least one property.

Besides inherently being necessary to determine the location of an object, depth perception plays an integral part in decoding information regarding size of objects (Bebko and Troje, 2019; Carlson, 1960; Dixon and Proffitt, 2002). Hornsey and Hibbard (2018) found this to hold true also with consumer VR using an Oculus Rift. Raddatz, Uhlarik and Jordan (2001) performed a series of experiments to determine how depth perception cues interact with *size constancy* (the ability to recognize two sizes as the same despite retinal sizes being different due to distance or other effects). Texturing the ground plane, they found foreshortening (lines parallel with the horizon getting closer the further away from the viewer, part of texture gradients) to be particularly effective in maintaining size constancy though only for vertical size.

Peripheral vision

Peripheral vision, as defined in the Oxford Dictionary of Media and Communication (Chandler and Munday, 2020) is:

> In relation to human eyesight, an area constituting all but the central 2% of the overall field of view around the point of fixation; it is relatively low resolution and thus offers a less focused image than foveal vision.

As such, any experiment utilising any significant portion of the field of view will be influenced by the effects resulting from this. Determining what effects this might be is therefore important when designing said experiments.

It is commonly said that one can only see in black and white in the peripheral, but this is not strictly true. There is a degradation of the ability to see colour contrast as the eccentricity (degrees from fovea, point of fixation) increases, but it seems to be gradual and still existing up to at least 50 degrees eccentricity (Hansen, Pracejus and Gegenfurtner, 2009). This effect is seemingly dependent on the size of the viewed object, with Abramov, Gordon and Chan (1991) finding that increasing the size of said object could lead to fovea-like colour vision at 20 degrees eccentricity, but not at 40 degrees. Others again have been able to find similar effects up to 45 degrees of eccentricity (Gordon and Abramov, 1977; Johnson, 1986). This effect is not equal for different hues however, with green being easier to determine at higher eccentricities than red or blue (Abramov, Gordon and Chan, 1991). A similar, though slower, degradation of visual acuity is also observed when testing for achromatic (greyscale) inputs (Anderson, Mullen and Hess, 1991). As such, one might expect a relation between head movement and accuracy in both cases, with more head movement resulting in less use of peripheral vision, although due to the FOV of the Oculus Quest being at only 90 degrees, thus giving a maximum eccentricity of 45 degrees, this effect might not be too pronounced. One would also expect this relation to be weaker in the achromatic case.

The perception of size is also affected by being viewed in the peripheral. Baldwin *et al.* (2016) found that objects viewed in the periphery is perceived as smaller than their perceived size in the fovea. This result held true for both horizontal and vertical peripherals, with objects being contracted in both dimensions, though more so in their congruent dimension. This held true even when participants were allowed to move their eyes freely with no time restriction. As such, it is possible that the height of buildings can be misread due to both vertical and horizontal peripheral interactions. To mitigate this effect, it might be important to allow participants to move not only their eyes, but also their heads freely, if these dimensions are to be measured.

2.2.4 Legends

In cartography, whenever one uses a variable that might not intuitively tell the information contained accurately, a legend might be utilised. A legend will generally be places on a blank space or a location with little important information and contain instructions to how the reader is to read phenomena through scales, symbols and the like (Tyner, 2010; Slocum *et al.*, 2009). In later years, a lot of work have been done with online web-based maps, some of which are three dimensional. A 3D atlas over Switzerland was launched in 2015, and a presentation was held describing their choices for their legends (Schnürer, 2014). In general, most scales and symbols translate directly to 3D, with little to no extra work. Worth noting is that size as a variable was not handled, except when in tandem with colour, such as in a precipitation map, where the colour was the variable shown in the legend, or when used on a 2D symbol or chart superimposed on the map. In the first of these cases, perspective height was used visually, but with no explanation through legends, and showing no extra information beyond what the colour encoded. As such, the question remains of how to effectively convey this information separately.

Beyond the fact that specific variables might be hard to visualise in a legend, there is also the question as to where to localise the legend itself. In a 3D environment, especially an immersive 3D environment like in VR with an HMD, there might not be any blank spaces to fit a legend.

In the project leading up to this paper, four different strategies were identified: a static legend in the field of view, a static legend placed in the environment as a sign, pop-up information on demand, or a table-top model as an overview. The first two simply handles the problem of location but makes no progress in the question of representing three dimensional variables such as size. The pop-up labels on demand will surely help gather information about any specific object but fails to give a good overview. The table-top model is promising but requires the user to make any decisions based on variables such as size on said model rather than in the environment, questioning the need for an immersive environment in the first place.

As such there is no obvious solution to this problem. This paper will explore a fifth option, where an object of a given value is shown beforehand in true scale. The question then remains, however, as to how well this kind of visualisation can be remembered and recollected in an immersive virtual environment.

2.3 Memory

2.3.1 Working memory / short-term memory

In comparison tasks where two or more objects cannot be perceived simultaneously, such as when using a legend, you are reliant on your memory. This is due to having to commit to memory the specifics of the comparator, be it the symbol, size, colour, etc. Focusing on these kinds of tasks, there are more specific terms that can be utilised.

The probably best-known kinds of memory known to the common layperson would be long- and short-term memory. Using these terms, there is a separation between the learned knowledge and memories from throughout a lifetime, and the recent fleeting task-related memories related to the current goings-on, respectively. In an effort to understand what effects are truly tested, a deeper understanding is necessary.

Working memory describes the information one is thinking about at any particular moment, and the limited amount it is capable of holding is ever changing (Cowan, 2013). Working memory was developed as a theory largely through the work of George Miller, Alan Baddeley and Graham Hitch, and popularised through the publication of Baddeley's *Working Memory* in 1986 (Magnussen, 2013). The idea is that a multitude of cognitive tasks are being performed by the working memory, not necessarily fitting with the idea of short-term memory. These tasks include thinking, problem solving, reasoning language comprehension and production, and tracking of changing events, as well as the data required (Cowan, 2013). Working memory is thus the integrated system consisting of various components, all utilising the same cognitive resources. An example of these shared resources could for example be when focusing on remembering a long shopping list, one might forget not only parts of said list, but also other goals, such as dropping of a parcel.

The distinction between working memory and short-term memory is hard to define precisely, mostly due to the lack of consensus of the definitions for either term. Some refer to the working memory as the active recollection of relevant information, while short-term memory is related to the passive observations (Cowan, 2013). As an example, being asked whether you met any dogs on your walk once it's done would tap into the short-term memory, while counting the dogs you meet on your walk as you're walking would be working memory. As noted by Cowan, others see working memory as encompassing both active and passive cognition, thus effectively making short-term memory a subset of working memory. The discussion of further definitions has been discussed by several authors, including Cowan (2008).

One component of the working memory as defined by Baddeley includes a *visuo-spatial sketchpad* (Baddeley, 2012). This component is responsible for keeping a person's position in relation to other objects, so as to not bump into a chair behind them, thus spatial, but also the visual effects of maintaining in memory how the chair looks. It is also responsible for displaying and manipulating spatial and visual information held in long-term memory, and displaying this in "the mind's eye" (McLeod, 2012), such as the case if they were to attempt to visualise said chair. Magnussen (2013) makes a distinction between the two tasks, where maintenance of visuo-spatial information is the work of the sketchpad, while consciously producing an image is a separate process called *visual imagery*. This distinction is strengthened by findings where task-irrelevant noise might disturb visual imagery tasks, but maintenance of visual working memory seems unaffected.

Magnussen further describes a sub-division of the visuo-spatial sketchpad, where said component is divided into three separate sub-components. One subsystem is related to the processing of spatial characteristics, such as position in an image. Another subsystem is related to the processing of objects and object properties, such as size and colour. The last subsystem is related to the fidelity of the information retained. What distortions of colour, texture and orientation occur?

2.3.2 Visual Search

As defined by Proulx (2013), visual search is "the act of looking for an item" the target of which can be objects, features or events. Testing of visual search generally looks at two heuristics, namely speed and accuracy. Speed is related to the time it takes for the tester to make find their target or determine no such target exists, while accuracy is the degree to which the tester makes the correct decision. Commonly results are displayed as a graph mapping reaction time to number of items displayed, though accuracy based on display time is also examined (Proulx, 2013). Proulx further discusses the notion of *efficient* and *inefficient* searches, where an efficient search is where the reaction time increases slowly with increasing number of items, and conversely an inefficient search being one where the increase in reaction time is high.

The speed and accuracy in a visual search are dependent on several variables, and common variables to manipulate for this purpose are according to Proulx (2013):

1. The *set size*, being the number of items to search through, including the target.
2. The *defining feature* of the target, being the difference between the target and the other objects in the set.
3. The *degree of similarity* between the target and other objects. Generally, the more similar the target is to the other objects, the harder it will be to make determinations.
4. The *reported feature* of the target. If the tester is asked to make a judgement about the existence of the target or about a feature of the target, such as what colour it is.

The defining feature of the target need not be limited to a single feature, as it could be related to several simultaneously, such as searching for a yellow circle in a group of coloured geometrical shapes. However, if the defining feature is limited to a single basic geometric features (such as size, orientation and colour), the search might be faster as determinations can be made in an earlier stage according to *feature integration theory*, where all the items are processed simultaneously, rather than successively as is done in more complex searches (Treisman and Gelade, 1980). This is of course dependent on being able to view the whole set simultaneously, but large portions of the visual field can be utilised for this purpose (Wolfe, 1994).

Following the experimental setup of researchers like Treisman and Gelade (1980), and codified into experimental paradigms like *delayed match to sample*, or DMS (Cambridge Cognition, n.d.; Chudasama, 2010; Cools, 2010), a visual search can be divided into a few distinct phases. First of these is the *exposure* – where the user is shown the target(s) allowing them to commit them to working-, short- or long-term memory as wanted. The second phase is the *search*, where any determinations are to be made as to the target’s whereabouts, or in fact existence, among a group of distractors. These two phases need not be separate and can be done simultaneously, where the target is shown during the search time, such as the case with *the simultaneous matching to sample experimental paradigm* (Magnotti, Goodman and Katz, 2012). In the case where these phases are separate, a *retention interval* must be determined, defining the time between, in which the target is found only in the user’s cognition. Generally there is a tendency for increased exposure and search times to correlate with more accurate or faster searches, while the correlation is negative for the retention interval (Huang, Chang and Wei, 2010; Proulx, 2013).

3 Method

3.1 Application Design

For the purpose of testing, an application was developed using the Unity game engine (Unity, 2019) specifically for the use with an Oculus Quest. This section serves as an overview of the different functionality, but most specific decisions as to values of variables are found in section 3.2. This codebase is since released publicly on GitHub. A user's guide can be found in the appendix.

During the development of the application, multiple design decisions have been made to allow for the testing and recording of data to a satisfactory degree. Following on the research questions, the application must be able to facilitate visual search tasks both using a Narrow (field of view) and full (360 degrees) placement. As the context of this paper is that of a thematical map, it is natural to describe each object as a discrete object likely to be found in such an application. For the purpose of this paper, each object symbolises a building, and will be described as such.

The application is designed to test using an approximation of *Delayed Match to Sample* (DMS) (Cambridge Cognition, n.d.), modified to the immersive environment and the intended context.

3.1.1 Variables

As with any test, isolating the variables to test is of paramount importance. As such the first step in the application design should be to determine which variables are present in any given scenario and determine values to which each non-tested variable is to be held. A complete overview of the identified variables is described in Table 4. While any number of these variables could conceivably be utilised as independent variables in further experiments, the application at present allows for three: Height of Buildings, Colour of Buildings (specifically greyscale) and Distance of Buildings. With the latter, this change might be contrived in the context of a cityscape, with limited movement allowed while preserving locational patterns, but could be interesting for other psychophysical purposes.

The application is created to be flexible, as to easily iterate upon feedback during pilot tests. Therefore, each of the variables listed in Table 4 can be easily changed between the options listed. This also allows for this application to be used with different variables for other experiments.

Table 4 Controlled variables in the testing application.

Variable	Description	Options
Ground Texture	Pattern seen on the ground plane.	• Monocoloured • Tiled • Textured, not tiled
Known Objects	Objects known to the observer visible during the test.	• Yes • No • Toggled
Building Texture	Pattern seen on the building objects.	• Monocoloured • Realistic • Unrealistic
Visual Field of View Degrees	Degrees viewable at any one time in the HMD. Hardware and setup dependent.	Number of degrees > 0.
Number of Buildings	Number of Buildings to compare, including the correct one. Set size.	Any number > 0.
Height of Buildings	Height of the Building in meters.	Any height > 0
Colour of Buildings	Colour of the Building, primarily used with monocoloured Building Texture.	Full colour / Greyscale Any colour band 0-255.
Distance of Buildings	Spawn distance of Buildings, as seen from observer.	Any distance > 0
Building Model	Shape of the Building objects.	• Primitive • Simple • Realistic
Sampling algorithm	Sampling algorithm for active variables when creating tests.	• Even Spread • Uniform Sampling • Other
Placement Degrees	Spawn positions of Buildings in degrees.	• Within FOV • 360 Degrees
Exposure time	Time allowed to memorise correct answer.	• Unlimited • Any > 0 seconds.
Search time	Time allowed to search for correct answer.	• Unlimited • Any > 0 seconds.
Forced answers	Must observer select answer to continue?	• Yes • No

Table 4 Controlled variables in the testing application, continued.

Variables Known	When shown the target answer, is the observer told which variables will be active?	• Yes • No
Test Order	The order in which tests are undergone.	• Random • Controlled
Movement	Observer can move in the test environment.	• Allowed • Disallowed
User position	Physical positioning of user.	• Standing • Sitting

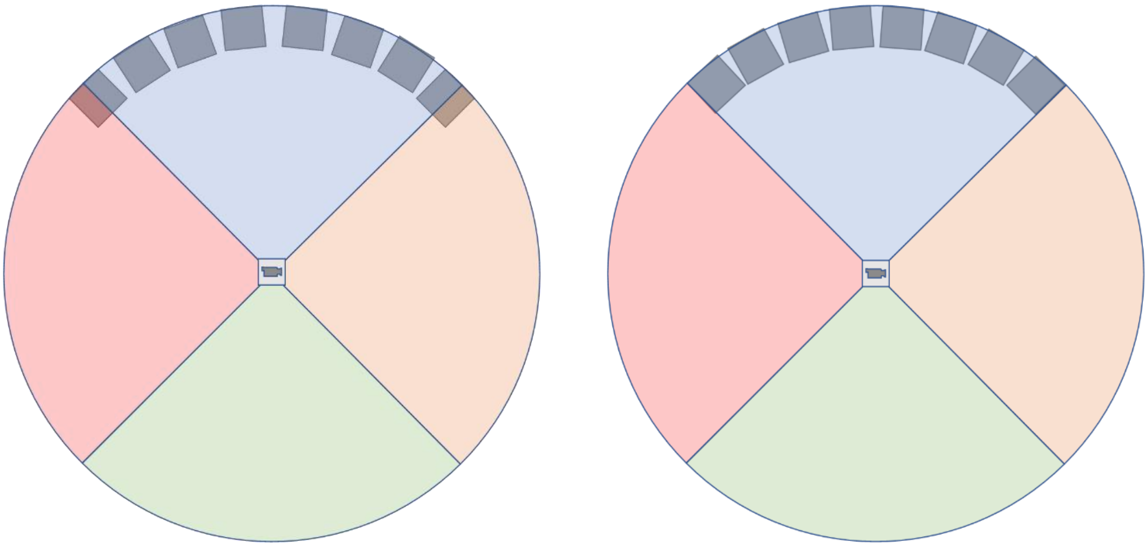
3.1.2 Placement of Buildings

The application allows for two kinds of spacing of buildings, described as Placement Degrees in Table 4. As such we have two possible cases: 1. All buildings are placed within the field of view (Narrow), and 2. The buildings can be placed anywhere around the user (Full).

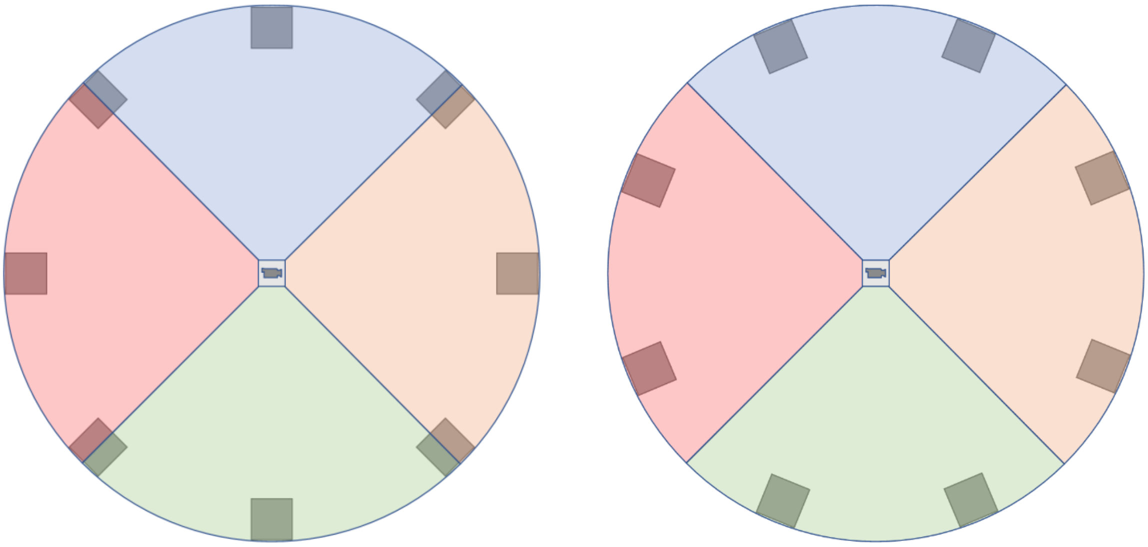
It was determined to be ideal to space the buildings evenly in the allowed space, in order to minimise possible biases due to variable spacing.

There are four factors responsible for the reducing the spacing: increased number of buildings, reduced distance of buildings, increased size of buildings (building model) and reduced degrees allowed. The reverse is also true, where the spacing can be increased. Knowing that the minimum FOV is at most 90 degrees due to hardware limitations of the Oculus Quest, the degrees allowed are mostly a known factor. Thus the remaining variables must be balanced as to not have building clipping into each other, with the exact values being determined based on the test performed. Assuming that these variables are held constant between the Narrow and the Full test scenarios, there will be a difference between spacing in these, where the distance between buildings will be smaller in the Narrow case. This difference in spacing is known to potentially produce biases, with Chen, Lin and Huang (2014) finding fewer errors in perceived height when the viewed objects were placed closer together. The most natural alternative, however, is the increasing the number of buildings in the 360-degree case, but similarly biases exist in this case, as increasing the set size is known to inhibit visual search (Wolfe, 2012).

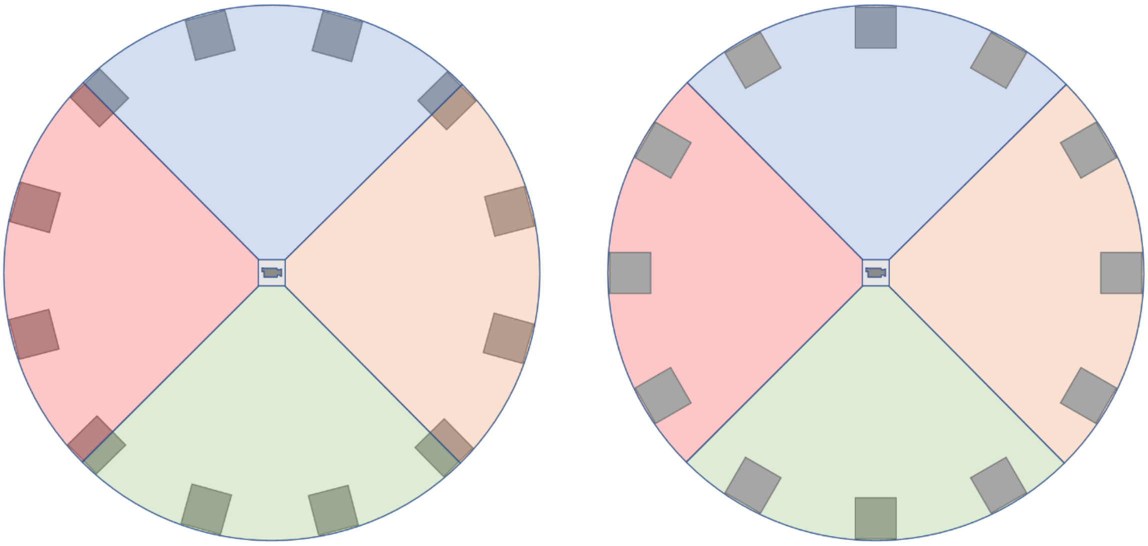
Having settled upon an even spacing, we are still left with the question of how to distribute the buildings. With the FOV of the Oculus Quest found to be 90 degrees, the idea of quadrants became natural in the creation of the tests, with one full quadrant being visible if facing its centre. These quadrants are marked as coloured semi-circles in Figure 5. This division was also proved useful in forcing generated tests to have answers spread semi-evenly along the 360 degrees available. The question then remains of how to place the buildings in quadrants. The main concern being to what degree the peripheral is utilised. For both Narrow and Full scenarios, the choice fell upon having the outermost buildings in any quadrant clip into the neighbouring quadrant, effectively being part of both, as can be seen in Figure 5c and e. This came from a wish to utilise the peripheral vision immediately in the Full scenario, while still maintaining even spacing. This does however mean that the outermost buildings are clipped, and only half is visible.



a) 8 Buildings in Narrow case with clipping **b)** 8 Buildings in Narrow case without clipping



c) 8 Buildings in Full case with clipping **d)** 8 Buildings in Full case without clipping



e) 12 Buildings in Full case with clipping **f)** 12 Buildings in Full case without clipping

Figure 5 Distribution of buildings in space.

This also lead to the same being true in the Narrow case, leading to the choice of using the layout shown in Figure 5a rather than b. Do note that for the purpose of the tests, the user would start each test facing in the middle of the upper (blue) quadrant, thus seeing only what is in that quadrant until they move their head.

3.1.3 Interactions

User interface design is a scientific field of its own, and the process of creating an intuitive and well-functioning interface is usually an iterative process utilising user tests. Unfortunately, this is was not possible during the design of this application, as a result of the COVID-19 pandemic. As such the interactions are solely based on the author's experience and therefore there are interactions that might need further implementation tweaks if tests are to be performed.

The case stands however, that the interactions available at present are capable of fully performing tests, and the controls are explained to the user through tutorials when opening the application. Any button available to be pressed at any given moment is also given a tooltip visible on the digital copies of the controllers, both in the testing environment, and during the tutorial, with all the options depicted in Figure 6. The interactions were also largely based on existing paradigms commonly used for VR gear in general (Somolinos, 2017) and for Oculus in particular, with point-and-click with a laser being the main one.

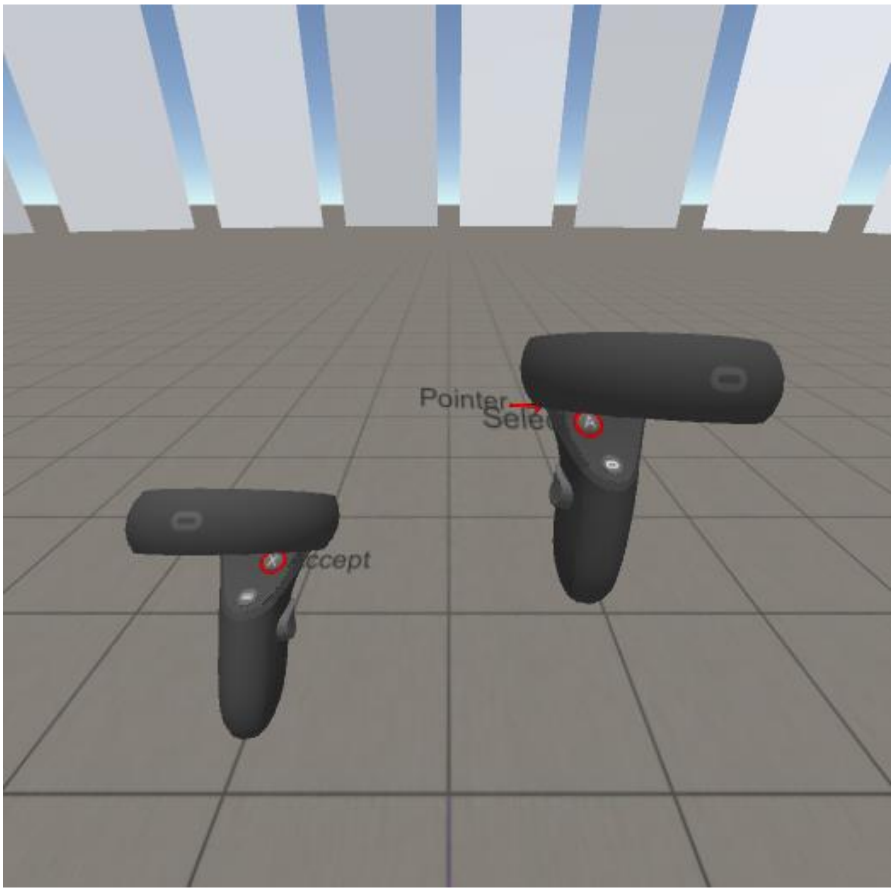


Figure 6 Tooltips are available on the controllers are always available.

To minimise the learning curve and confusion, only three buttons are used. The right index button provides a laser pointer, so the user can easily see where they are pointing, which can be seen in Figure 7. The A-button generally works to select whatever is pointed at if it is selectable. These are both locked to the right hand, to help subconsciously link them together. In fact, the application will not allow you to select a building if the laser pointer is not held. The X-button serves as an accept and proceed button, and as such any progression in the test environment as well as the tutorial is driven by this button.

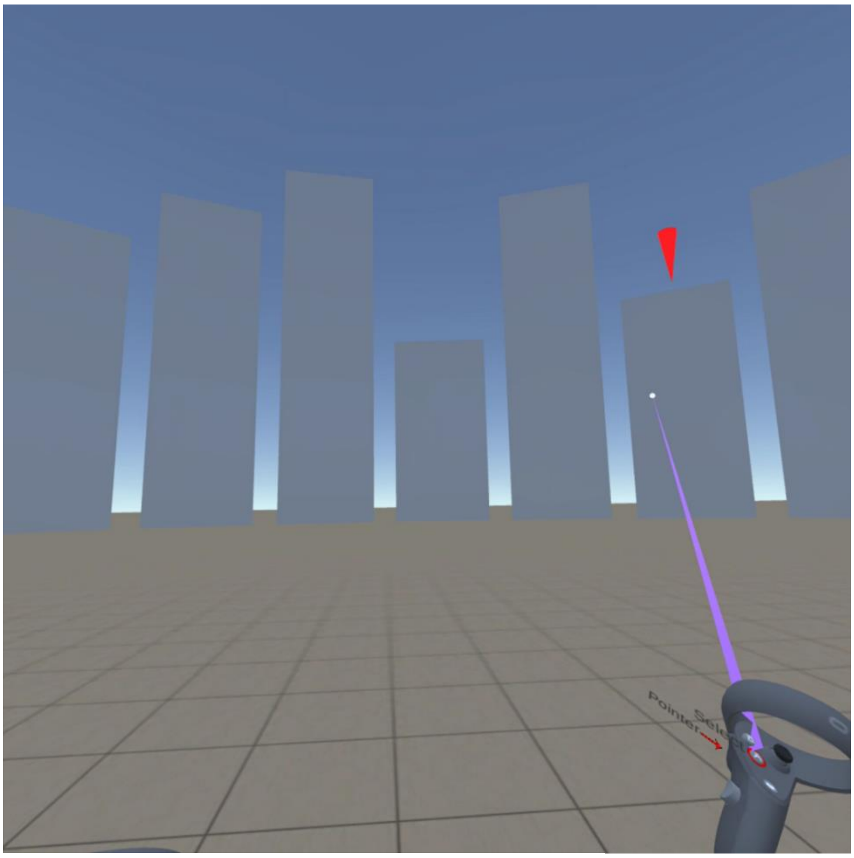


Figure 7 A laser pointer is used to select.

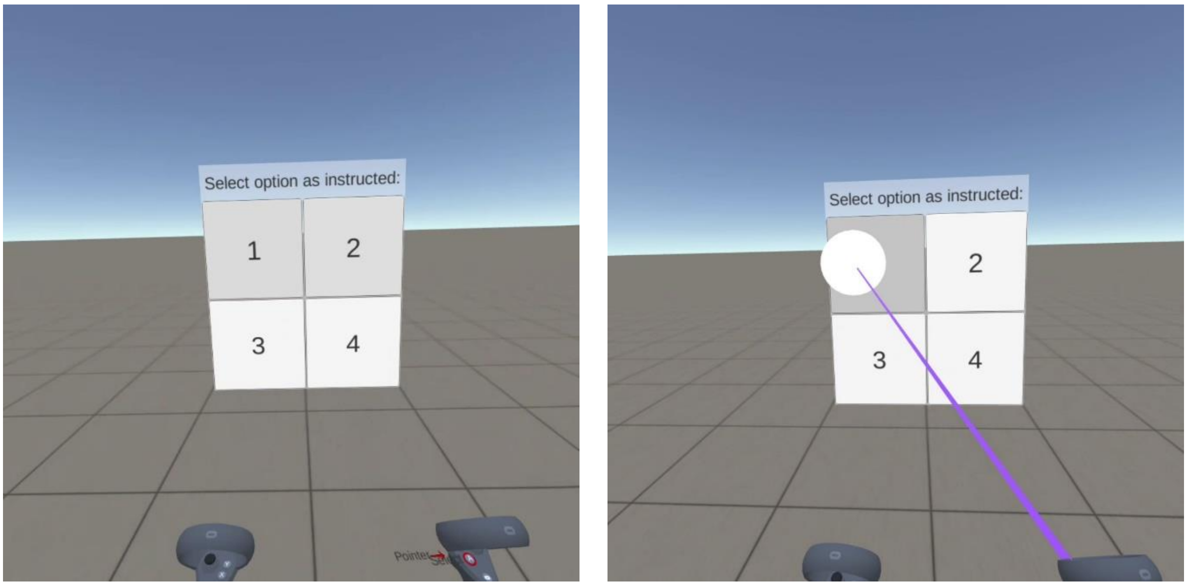
3.1.4 Test Flow

Each test person will face the same steps when interacting with the app. For the purpose of this, each individual visual search will be called a *test* while one run of all the tests are called a *test pass*. Each test person is to complete one test pass, containing the required tests, but for the purpose of eliminating bias, the tests and ordering thereof need not be identical, although the application also allows for this.

Each test pass is started with the option to select a number, 1-4, as shown in Figure 8. This number should be determined by the person overseeing the test, as this will determine the order of tests for this particular test pass, as per Table 5. More detail about test ordering will be discussed later.

Table 5 Order of tests from selected number.

Number	Order of tests
1	Narrow tests first, Height tests first
2	Narrow tests first, Colour tests first
3	Full tests first, Height tests first
4	Full tests first, Colour tests first



a) Selection screen

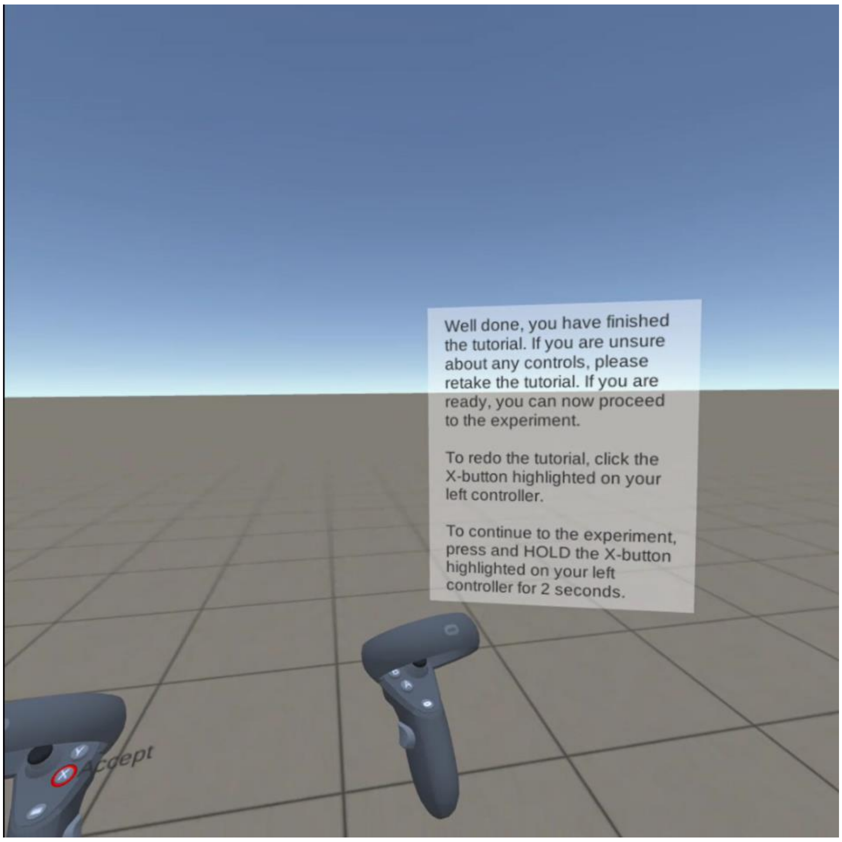
b) Selecting as with buildings, using the laser pointer and A button.

Figure 8 Selection screen for ordering tests.

Once an option is selected, an information screen is shown, before starting the tutorial, as seen in Figure 9. The tutorial aims to show the full set of interactions available, and how the tests are to be performed. As such the tutorial consists of a set of tests, performed identically as the actual tests later, except with some extra information displayed to the user, telling the user how to interact with the environment. Once all tutorial tests are performed, another view is entered, where the user is asked to redo the tutorial if anything is unclear, or proceed to the test pass if not, seen in Figure 10.


Thanks for partaking in this experiment. We will begin with a tutorial. To begin this tutorial, please press the X-button highlighted on your left controller

Figure 9 Start information screen.

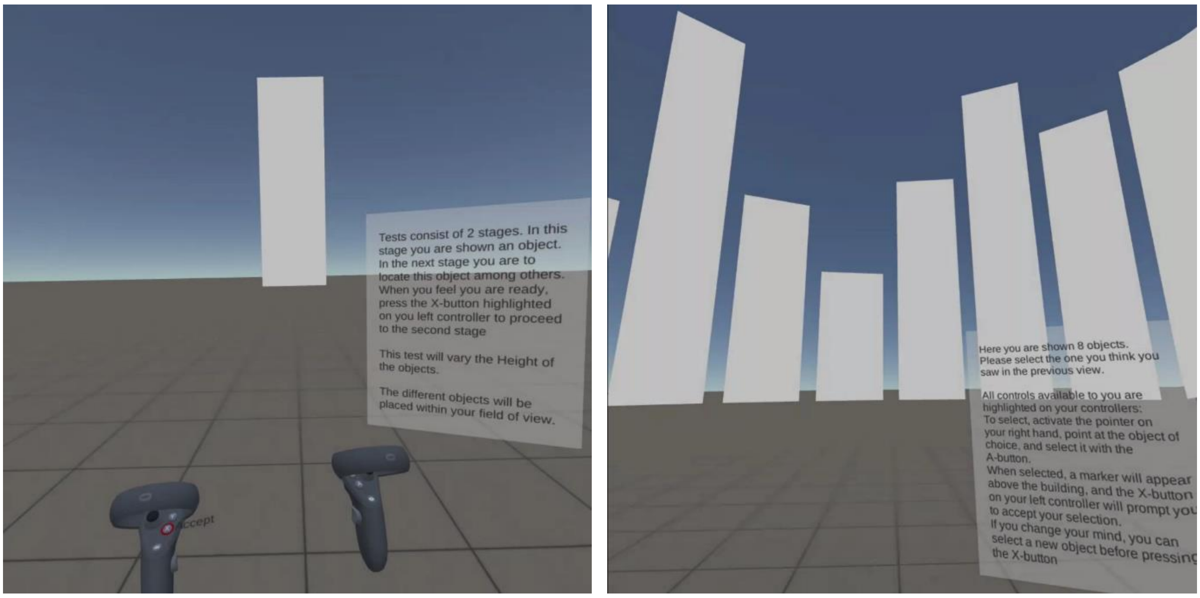

Well done, you have finished the tutorial. If you are unsure about any controls, please retake the tutorial. If you are ready, you can now proceed to the experiment.

To redo the tutorial, click the X-button highlighted on your left controller.

To continue to the experiment, press and HOLD the X-button highlighted on your left controller for 2 seconds.

Figure 10 End of tutorial screen.

Each test consists of two parts, as defined by visual searches. First, during the exposure phase, the user is shown the target, this being the building they are to locate in the following view, as seen in Figure 11a. This building will be identical is every way, with the exception of being located in the centre of their FOV and having no other buildings nearby. As such perceptual constancy is maximally maintained (Wagner, 2013). The user can have limited or unlimited time to examine the building, as wanted. Once the user chooses to proceed, or their allowed time is depleted, the user will then be shown the same building among a number of distractors, varying a single variable, representing the search phase of visual searches. These buildings are placed either in the initial FOV for a Narrow test, or utilising the entire 360 degrees around the user during a Full test, as visualised in Figure 5. Figure 11b shows how a Narrow case looks during the tutorial (thus having an info-screen available). The user is then to attempt to locate the building previously shown and select it. Upon selection, a marker will appear above the selected building, as shown in Figure 7. There are different ways for a test to be ended. Once a building is selected, the user is allowed to proceed to the next test. If the test design forces answers, this should be the only way to proceed. If not, the user is allowed to go to the next test at any time, skipping the current test without answering. Also for this step a maximum time allowed can be set, forcing the user to go to the next test. Worth noting is that visual searches can be done where the target does not exist in the search set (Wolfe and Horowitz, 2008; Wolfe and Van Wert, 2010), but the application is not currently able to handle such a case, as it is not relevant to the stated research questions.



Tests consist of 2 stages. In this stage you are shown an object. In the next stage you are to locate this object among others. When you feel you are ready, press the X-button highlighted on you left controller to proceed to the second stage

This test will vary the Height of the objects.

The different objects will be placed within your field of view.

Here you are shown 8 objects. Please select the one you think you saw in the previous view.

All controls available to you are highlighted on your controllers:
To select, activate the pointer on your right hand, point at the object of choice, and select it with the A-button.
When selected, a marker will appear above the building, and the X-button on your left controller will prompt you to accept your selection.
If you change your mind, you can select a new object before pressing the X-button

a) Part one of test: target shown.

b) Part two of test: search and select target.

Figure 11 The two parts of a test.

Once all tests are completed, the user is shown a view thanking them for their participation, and allowing no further interaction, seen in Figure 12.


The experiment is finished.

Thank you for participating!

Figure 12 End of experiment screen.

Between each view, the screen fades to black and back again over 2 seconds. This is done to hide artifacts from releasing and loading building models, but also aims to minimise any possible motion sickness resulting from rapid change objects in view. It also corresponds to the idea of *retention intervals* in visual searches and should be expected to influence the results thusly.

A flow chart describing all the states of the system is shown in Figure 13. Each arrow represents an active selection and has a corresponding screen fade.

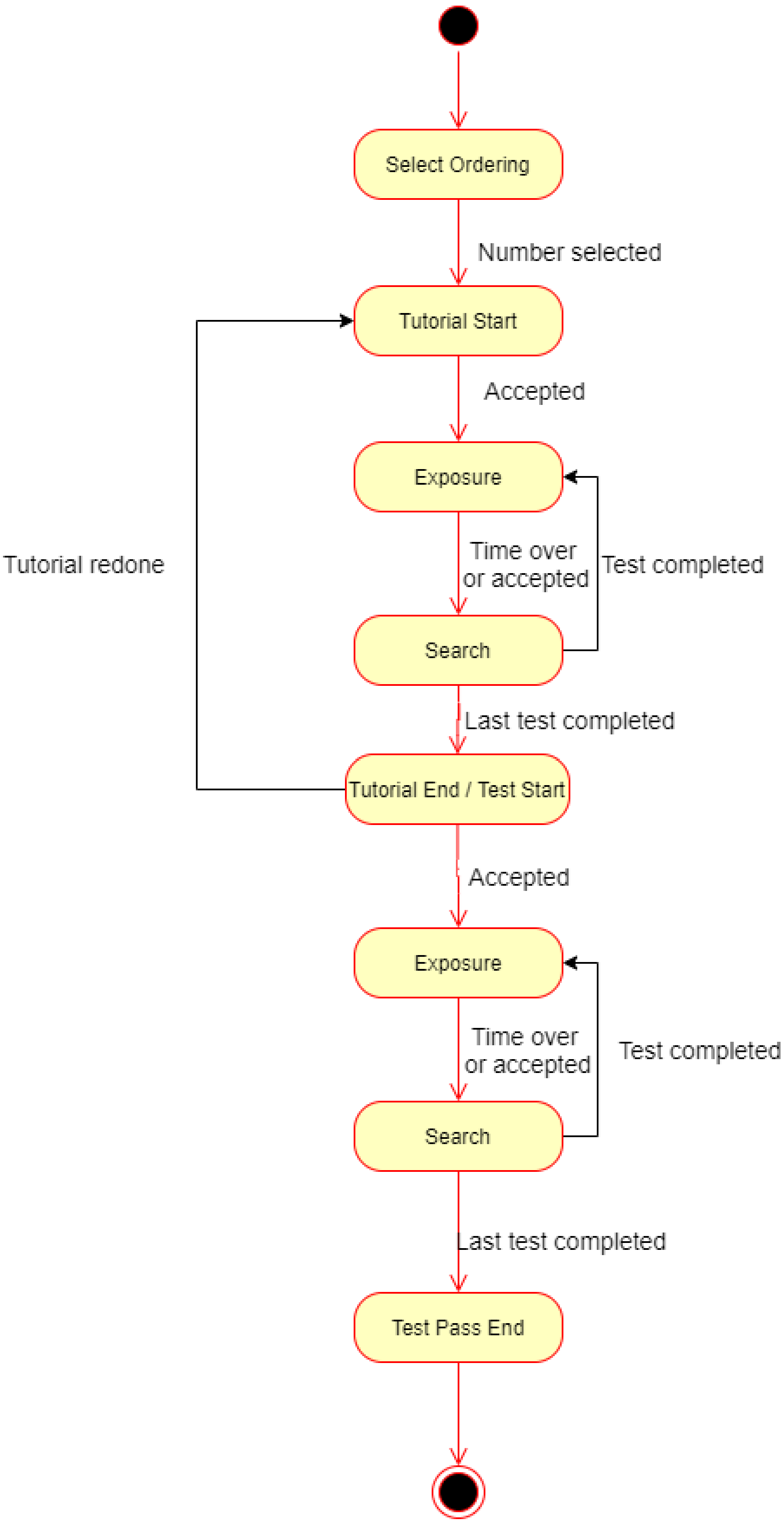


Figure 13 Flow chart of application states.

3.1.5 Test Creation

The application is made to easily be able to create new tests based on new criteria, so as to easily be able to test new things or react to feedback gotten during pilot tests. Generally, all variables listed in Table 4 are possible to change in less than a minute, should this be desired.

There are two ways of creating and running tests available in the app, depending on the randomisation wanted. One way is to manually create the tests and run these pre-generated tests to all users. This allows for greater control over how the variables will be tested, and makes test easier to compare, as each user will be faced with identical tests. This could of course also lead skewed results if the tests are not very carefully created.

The second way is generating the tests at runtime. For this purpose, a set of test parameters are created as a file (through the Unity interface) and added to the application. Then, as the application loads, the tests are created based on the test parameters. This means that each test pass will consist of tests with equal parameters, giving a unique set of tests for each user. The tests will be 1-to-1 comparable.

For each individual test, a set of parameters must be determined, besides the general case of determining the variables listed in Table 4. Most important is determining the Placement Degrees (Narrow or Full) for the given test, which variable will be the independent variable (Colour, Height) and what values this variable is allowed to vary between. In the case of Placement Degrees, an option of "Both" is also possible, where one of each is created with otherwise the same parameters. Further each test is given a quadrant where the correct answer should be placed. This is done to control the tests, so that each user is faced targets spread all around. This value is only utilised in 360-degree tests and is dependent of the number of buildings being divisible by 4, or else it is ignored. The quadrant can be set manually or randomised, and if using runtime generation of tests, tests with the same parameters can be created simultaneously for all 4 quadrants. When randomising quadrants, as a result of the shared quadrants of some buildings, seen in Figure 5c and e, these building positions would be more likely to be sampled, as such the sampling has been appropriately weighted to make each position equally likely. In the case of combining this feature with the creation of "Both" for placement degrees, an option is also available to create one Narrow test for each quadrant, as to have the same number of Narrow tests as Full tests.

Besides determining the span of values, the method of sampling within this span is also important. Two such methods are implemented: uniform sampling and even spread.

Uniform sampling is done by giving each building a random value within the span allowed. This is done both for the distractors and the target. Even spread given one random building the minimum value, one the maximum value, and the rest with even steps from the values between, guaranteeing unique values for all buildings in the test. In both cases the target is sampled the same way as the distractors.

3.1.6 Test Ordering

In order to control for biases due to the order in which a user is presented tests, the application is capable of sorting tests in multiple different, controlled, ways. This is done by varying six variables: "Narrow First", "Randomise Set Order", "Separate Narrow and Full", "Separate Test Sets", "Split by Test Type", "Reverse Test Type Order".

Test sets in this application refers to a group of tests that shares the same parameters and contains both the Full and Narrow tests. Tests created are divided into these automatically when creating them at runtime, or as an option when creating them manually. Having “Separate Test Sets” set to true therefore means that each test set should be completed in its entirety before moving on to the next. “Randomise Set Order” makes sure the order in which the test sets are shown is random.

If “Separate Narrow and Full” is set to true, all Narrow tests are shown before all the Full tests or vice versa, depending on whether “Narrow First” is set to true or false, respectively. If “Separate Narrow and Full” is true, at the same time as “Separate Test Sets”, the tests are split between Narrow and Full *within* the test set, and both are performed before moving to the next test set. If set to false, the application with show Narrow and Full tests interchangeably. If both this and “Separate Test Sets” are set to true, the tests are fully randomised.

“Split by Test Type” determines whether all tests with a given independent variable should be performed before moving on to the next. If true, all tests varying Colour will be done before all tests testing height or vice versa, depending on whether “Reverse Test Type Order” is set to true or false, respectively.

Worth noting is that the “Narrow First” and “Reverse Test Type Order” variables are set during runtime, by selecting an option within the application, as shown in Table 5.

3.1.7 Environmental Factors

While a number of variables are easily customised in the application, the perception of several might be related to environmental factors.

One such variable is colour, and to a lesser extent texture. The scene shown to the user in the application is lit by a light source located at a point. As such both shadows and direct light might influence the perceived colour of an object. In order to minimise the potential differences this might afford different buildings, this light source was placed directly in the centre of the scene, and thus directly above the position of the user. As such the light has the same distance to each of the buildings when colour is to be determined, as height and distance from centre here is held constant but could serve as a depth perception cue (Continuous Brightness).

The other depth perception cues are also related to this idea. As a number of the cues listed in Table 3 could be utilised to increase the depth perception in the scene, and thus allow for better perception of both height and distance, knowing the degree to which these are utilised might be imperative to understand the results. In the case of Motion-Produced cues, these are irrelevant if the user is restricted from moving but are present if not. Binocular Disparity is not practically relevant while utilising the metaphor of buildings, but in other cases could be used. The Oculomotor cues are anecdotally preserved in the application, and this will be a constant due to hardware limitations. Atmospheric perspective and thus contrast and colour gradients are possible in Unity, but not utilised in the application to minimise its effect on recognising colour. The remaining Depth Perception Gradients are available in the application, being directly related to geometry, with texture on both buildings and ground (thus utilising Linear Perspective) is available. Pictorial cues become most relevant when determining distance to buildings, or when utilising known objects during the test. These known objects can be any number of things whose size, combined with the distance, can be utilised to determine both distance and height of the target buildings. These objects should be of a familiar size, such as people, vehicles, and potentially trees. A version of this is implemented in the application, with multiple models being visible if toggled as seen in Figure 14. The usability of the specific models, and the placement algorithms is not determined, as this was intended to be implemented fully if found preferable during pilot testing.

Here you are shown 8 objects.
Please select the one you think you saw in the previous view.

All controls available to you are highlighted on your controllers:
To select, activate the pointer on your right hand, point at the object of choice, and select it with the A-button.
When selected, a marker will appea above the building, and the X-butto on your left controller will prompt yo to accept your selection.
If you change your mind, you can select a new object before pressin the X-button

Figure 14 Known objects can be added to the tests.

3.1.8 Outputs

For the purpose of the experiment to be undergone by this paper, it was important that objective data was gathered. Therefore, the application tracks the user's movements and answers, and outputs these to files. The different outputs can be seen in Table 6. While only a small number of these variables are strictly needed to answer the research questions, namely "TestID", "TestType", "FieldOfView" and "CorrectChosen", the rest are available to provide more in-depth information.

The outputs are saved as JSON files directly on the Oculus Quest, in a folder structure made to easily identify test passes. As the tests are undergone, the outputs for each individual test is saved as they are completed, before all the tests are aggregated into a single file once the whole test pass is completed. This is done so that if an error occurs or a test person aborts the experiment, the results of hitherto completed tests are preserved.

Table 6 Outputs

Output name	Output values
TestID	Unique name
TimeUsed	Utilised search time in seconds
TimeViewingTarget	Time used to view the target before searching in seconds.
AggregatedRotation	Cumulative sum of all head rotation done during search in degrees.
DegreesUsed	Number of degrees viewed during search. Minimum degrees: Visual Field of View Degrees Maximum degrees: 360
TestType	Active variable for test: 0: Height 1: Colour 2: Distance
FieldOfView	Placement Degrees of test: 0: Narrow (Within FOV) 1: Full (360 degrees)
CorrectChosen	Is the correct answer chosen? True / False
CorrectValue	For the correct answer: The numeric value of the active variable.
ChosenValue	For the chosen answer: The numeric value of the active variable.
CorrectAngle	For the correct answer: The degrees away from initial orientation.
ChosenAngle	For the chosen answer: The degrees away from initial orientation.
TimeAfterObserved	Time used in seconds between observing the correct answer for the first time and selecting an answer. -1 if correct answer was never observed
MinValue	Minimum allowed value of active variable.
MaxValue	Maximum allowed value of active variable.

3.2 Test Design

In order to best answer the research questions, multiple variables needs to be controlled during testing, with the full list of possibilities available in Table 4. The full overview of chosen values set for each variable is noted in Table 7. Worth noting, however, is that multiple of these were supposed to be fully determined based on feedback during pilot testing.

Table 7 Variables chosen for test

Variable	Choice
Ground Texture	Tiled
Known Objects	No*
Building Texture	Monocoloured*
Visual Field of View Degrees	90
Number of Buildings	8
Height of Buildings	Active variable. Default 50m.
Colour of Buildings	Active variable. Greyscale. Default all colour bands 80/255.
Distance of Buildings	60m
Building Model	Primitive Cube. Base 10m x 10m.
Sampling algorithm	Even Spread.
Placement Degrees	Active variable.
Exposure time	Unlimited*
Search time	Unlimited*
Forced answers	Yes
Variables Known	Yes*
Test Order	Controlled
Movement	Disallowed
User position	Sitting*

* intended pilot test feedback

The choices of variables are based on the theory discussed previously in this paper, in an attempt to most accurately answer the research questions while reducing known sources of bias.

Ground Texture

For ground texture, a tiled texture was chosen. This was due to the findings of Matatko, Bollmann and Müller (2011) finding texture to be a good tool for depth perception in VR as well as Raddatz, Uhlarik and Jordan (2001) finding texture and foreshortening in particular to be effective in retaining size constancy.

Known Objects

There were no known objects intended to be used for the experiment, although pilot tests were intended to investigate this further. The reasoning for this is that while known objects are shown to help determine size and distance of buildings (Ng, Chan and Lau, 2016; Renner, Velichkovsky and Helmert, 2013; Raddatz, Uhlarik and Jordan, 2001), it could potentially distract from the core visual search, leaving less certainty to how much of an effect is resulting of the change between Full and Narrow views. The idea being that the probability of recalling size of a building could be related to which specific object is in front, and thus bias the result towards accuracy when the known object is similar to the one visible when previewing the target building. The counterargument goes that in the context of intended use, being a city-model, such objects might be natural to preserve a sense of realism.

Visual Field of View Degrees

The visual field of view is set to 90, as this is the widest the Quest seemed to be able to achieve from a few quick tests. It should be noted that reduced FOV is related to reduced motion sickness (Fernandes and Feiner, 2016), so keeping this to the maximum might lead to some motion sickness. However, in order to facilitate the further design of the experiment, the choice still remains with 90 degrees.

Spacing

As previously mentioned, spacing between buildings is defined by four factors, being distance, number, size, and degrees allowed. As the latter is an active variable in the experiment, the remaining is intended to be kept constant as to not produce further biases. Therefore, a set of values that does not result in the buildings clipping into each other is important. Starting with the metaphor of buildings, the size of these should be large enough to fit this metaphor. For this experiment, a 10m x 10m square is used as a base. Further, for the purpose of doing a visual search, there should be enough distractors to not have the users randomly guess the correct answers. Building on the idea of quadrants, and the inherent need for the test set to be divisible by four, this led to 8 or 12 being natural options. Using 8, a distance of 60 meters was found to give minimal, but still present, spacing in the Narrow case, while 80 meters was necessary for a test set of 12. Attempting to limit the distance, the decision was made to use 8, thus leading to the values described in Table 7.

Building Model

Besides determining the size of the building model, the shape and texture are also of interest. As was discussed in the project leading up to this paper, there are multiple ways of choosing these, while maintaining the possibility of encoding information through colour and height (Osnes, 2019). In the case of the model shape, there are three categories: primitive, where the full model is a cube, pyramid, cone, etc.; simple, similar with a primitive, but with smaller details; and realistic, which in this context would be a complete building model with details such as balconies. Similarly, texture could be applied to the buildings, with options varying between the primitive case of a monocoloured face, to a photorealistic surface. An option in between exists, where the texture itself is realistic, but with the colour changed to potentially unrealistic values, thus granting information. One design choice for this project was that no height information should be encoded in the texture or model directly, as this could mean that users would start counting floors by looking at windows, or looking at distortion of the bricks, rather than the height of the building as a whole. This experiment aims to find the effect of the height as a variable, rather than finding the best set of values to estimate said height. This kind of height-based texture is natural in the context of buildings but might well not be transferable to the general case and is therefore left out. Textures and shapes not containing any such information is still possible, however, and one such case was considered for the experiment shown in Figure 15, where a subtle but noticeable reflective glass texture was applied.

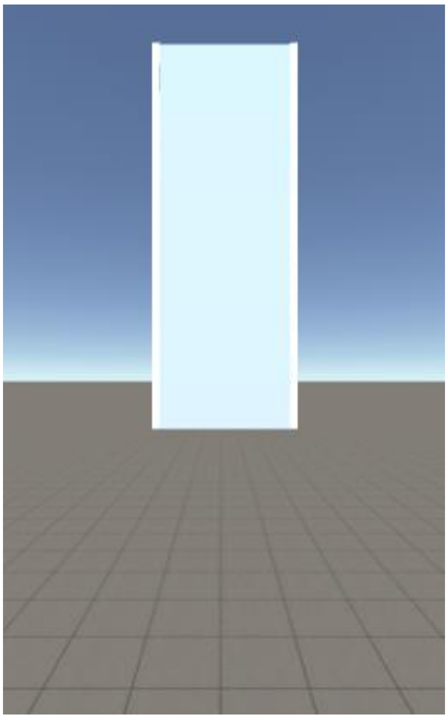


Figure 15 Textures can be added without granting inherent vertical information.

Visual Search Variables

Exposure time, search time, forced answers and variables known are all variables directly related to visual search. As previously discussed, performance in a visual search is commonly rated by one of two heuristics: accuracy or time. For this experiment the main goal is to ascertain the *accuracy* of the search, comparing the Full and Narrow cases. As such both time factors are set to unlimited, and users should be told to take the time they feel they need, not rushing their answers as this is related to decreased accuracy (Proulx, 2013). Increased exposure time is also tied to increased accuracy (Huang, Chang and Wei, 2010), and while this might lead to a skewed result in favour of users who naturally feel they need to use more time at this stage, this can allow for finding such effects. A case could be made however, that limiting these sets an upper limit on the total time used for a test pass. This could serve both as a way to make the data more homogenous by having less difference in time used, and also limiting the time immersed in the virtual environment – a factor known to correspond to induced motion sickness (Dużmańska, Strojny and Strojny, 2018). That said, the goal of the experiment is not to unreasonably rush the user, and any time limit should therefore be generous. Finding the appropriate time limits for not rushing the user, and potential motion sickness induced over time by the application would therefore be topics of interest during a pilot test. As the application is capable of logging not only if a correct answer is chosen, but the value of the chosen target, forcing answers can give insight to what degree the user is mistaken. Allowing skipping is likely to reduce the number of large errors, with users then skipping any question they are uncertain of. While this could lead to gaining insight as to how confident users are for different values, which of itself could make for interesting data, how well the information is maintained and recognised is of more interest in this paper, and thus the user will be forced to answer. Whether the variables are known or not, while having an impact on visual searches (Wolfe, 1994) was intended to be determined during pilot testing. To mimic the use case as a legend in a city model, where the dimension of interest would be known, they should be known also for this experiment. During pilot testing both cases would be attempted, and from there it would be determined if it had an obvious impact on the performance. If not, implementation would try to minimise user frustration and confusion.

Positioning

Movement and user position describe the physical variables related to the experiment. In order to limit the possibility of size constancy issues related to distance, the movement is disallowed. In practise this means that no controller input can move the user in the virtual environment, as well as the user being limited to a stationary boundary within Oculus Quest. Physically moving outside this boundary will warn the user as it happens, before shutting down the application. User position describes whether the user should be standing upright or sit down during the experiment. As found by Wraga and Proffitt (2000) both sitting and standing grants an advantage when determining height of objects, with the effect being stronger the closer to the eye-height of the user the height of the object is. Using the buildings, these are likely to be taller than the user standing, and thus much taller than the user sitting. Positive effects on visual searches for height, if any, is therefore more likely to occur while standing, but following the finding that this effect is only valid until 2.5 EH, constituting a height of less than 5 meters for most adults, this effect might be minimal, and therefore not necessary to optimise for in the current setup. Conversely, sitting has the positive effect of reducing motion sickness (Merhi *et al.*, 2007; Moss and Muth, 2011), and might therefore be preferable. The question remains however, of whether this poses a problem in regard to viewing the full 360 degrees around the user. The intended way to handle this would be to have the user sitting in a stationary chair capable of rotating freely or sitting on a stool. Pilot testing would aim to determine how well this functions, and whether standing up affords a more comfortable solution for the user.

Active Variables

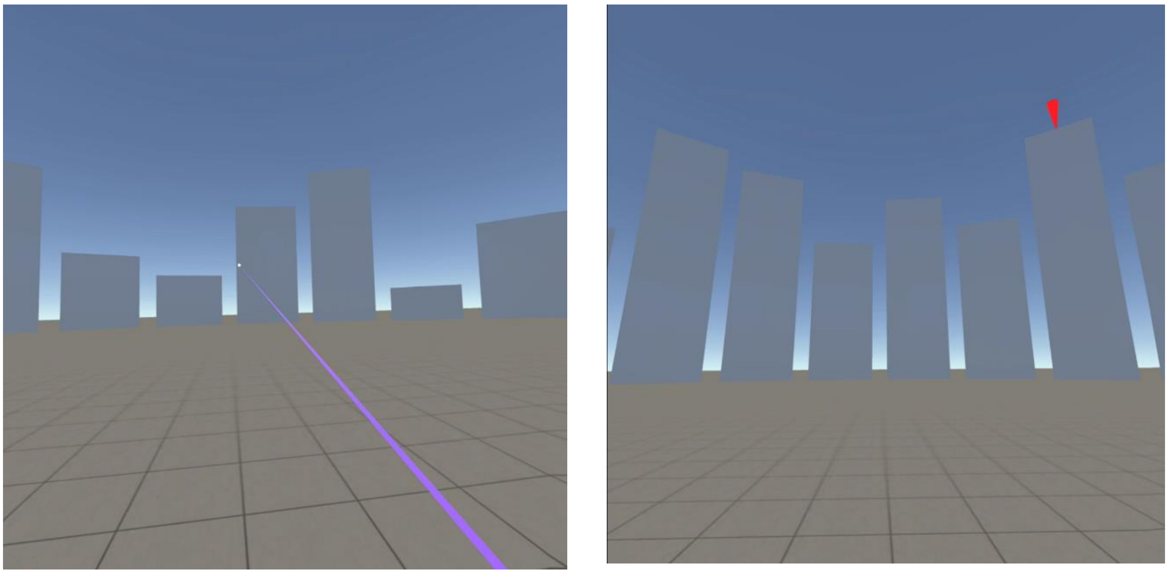
As active variables we have height, colour, and placement degrees. The latter of this is well defined as either being within field of view or 360 degrees. Being the factor by which results are to be compared to answer the research questions, all tests should be performed for both Narrow and Full cases. Height and colour are less binary. As defined earlier, each test is created using one active variable, maximum and minimum values, and a sampling algorithm. For height, these values represent the distance from the ground plane to the top of the building in meters, and as such is minimally bound to 0, but having to upper bound. Worth noting is that a building of height 0 will be un-selectable, having no face to target, while an unreasonably tall building quickly breaks the metaphor. The colour is bound both by a lower limit of 0 (completely black) and an upper limit of 255 (completely white), as commonly is the case in an RGB model. For this experiment the colour is greyscale, thus having each colour band (R, G and B) be set to the same value.

How many spans of values to test depends largely on the time used on tests, and time constraints related to motion sickness and fatigue. Lacking the insight a pilot test would grant in regard to this, the preferred case will be put forth. Each set of test parameters requires at a minimum two tests, one Full and one Narrow, so these can be compared. In order to make sure the target value is spread around the 360 degrees available, one test per quadrant is preferable, increasing the number to 5 tests per set of test parameters. As this is overly skewed towards Full tests, rather than Narrow, one Narrow test per Full test is implemented. This means a total of 8 tests are required to test any one span of values.

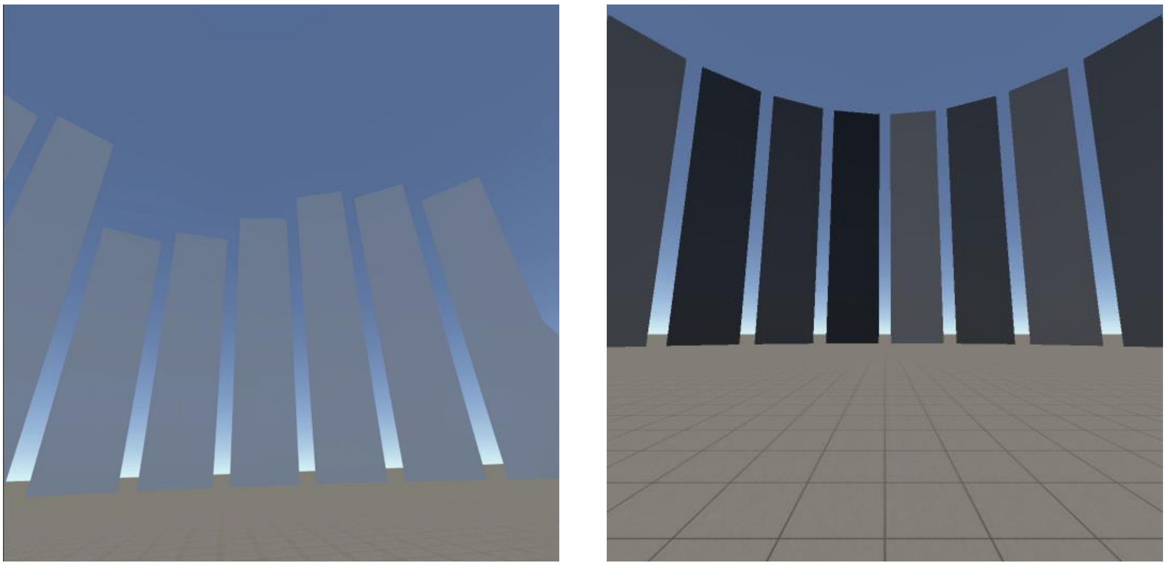
In order to control for the dimensions used in testing any variable, both variables should have values spanning different parts of their feasible values. For height, the three spans represent the smallest values possible while still being easily selectable in the application, buildings tall enough that the user must tilt their head to see the top of the buildings, and a third span in between. Similarly, the spans chosen for colour represents one set with the darkest values possible, one with the lightest and one in between. The size of the spans, as in the possible values they contain, are identical for all spans within one variable, but are largely arbitrary. All the spans are noted in Table 8 and shown as seen in the application in Figure 16, constituting a total of 48 tests. In order to guarantee the utilisation of the full spans, even spread was used as a sampling algorithm. This has the additional positive consequence of the colour values presented being closer to the common use case in cartography as discrete categorical values.

Table 8 Chosen value spans for active variables.

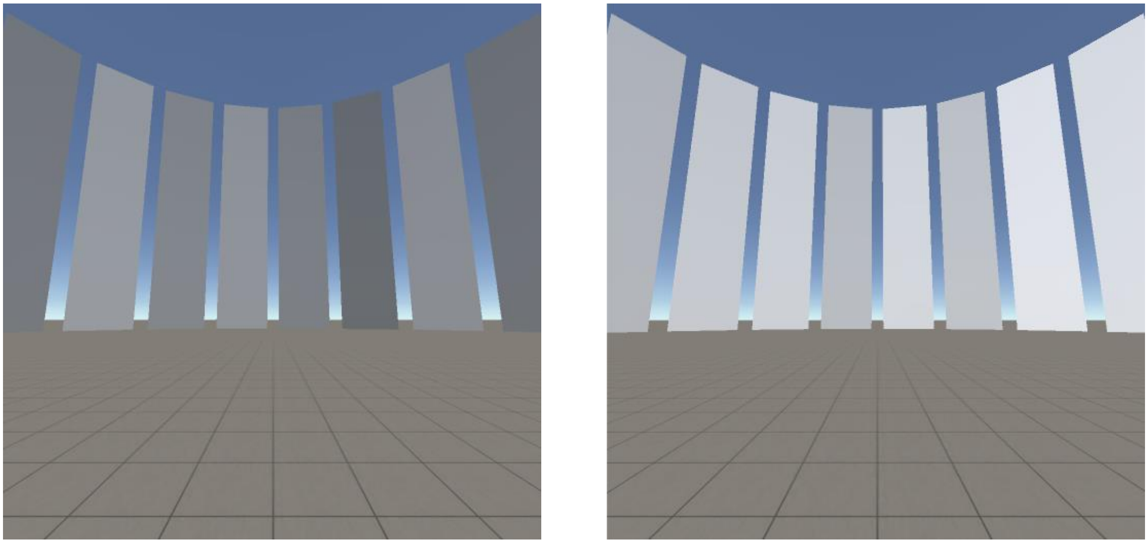
Variable	Spans
Height	5m – 25m
	20m – 40m
	40m – 60m
Colour	0-60
	100 – 160
	195 - 255



a) Height: 5m – 25m b) Height: 20m – 40m



c) Height: 40m – 60m. d) Colour: 0-60



e) Colour: 100-160 f) Colour: 195 - 255

Figure 16 Value spans as shown in the application.

Test Groups

The users should be split into 4 equal groups, balancing factors such as age and gender, with each group selecting a different ordering option, resulting as noted in Table 5. Groups should also be balanced based on their prior experience with both VR and cartography, this being known to impact results (Taçgın, 2020; Lee and Wong, 2014). This way any bias related to what variables are shown first is mitigated. All test persons should have normal or corrected-to-normal vision without the use of glassed, to maintain the FOV of the Quest. Each test pass is split into test sets, in which Narrow and Full sets are separated. Thus, each test pass is divided into 6 test sets (1 for each span), the first three of which will be all the height or colour tests of the test set, as defined by their group ordering, with the remaining three sets following. The order of the sets within each variable is random. Each set must be fully completed, both Narrow and Full, the order of which is determined by their group, before the next one is begun.

3.3 Ethical considerations

As with any experiments, ethical considerations should be taken. An unfortunate side effect of using immersive VR technologies is the motion sickness such an environment might induce. While care have been taken to design the experiment such that these effects are mitigated, they could still apply. With the effects of motion sickness becoming greater the longer the exposure to said inducing stimuli (Dużmańska, Strojny and Strojny, 2018), one must be aware of this. Informing the user of the potential for such a reaction and allowing for early termination of the experiment is therefore important.

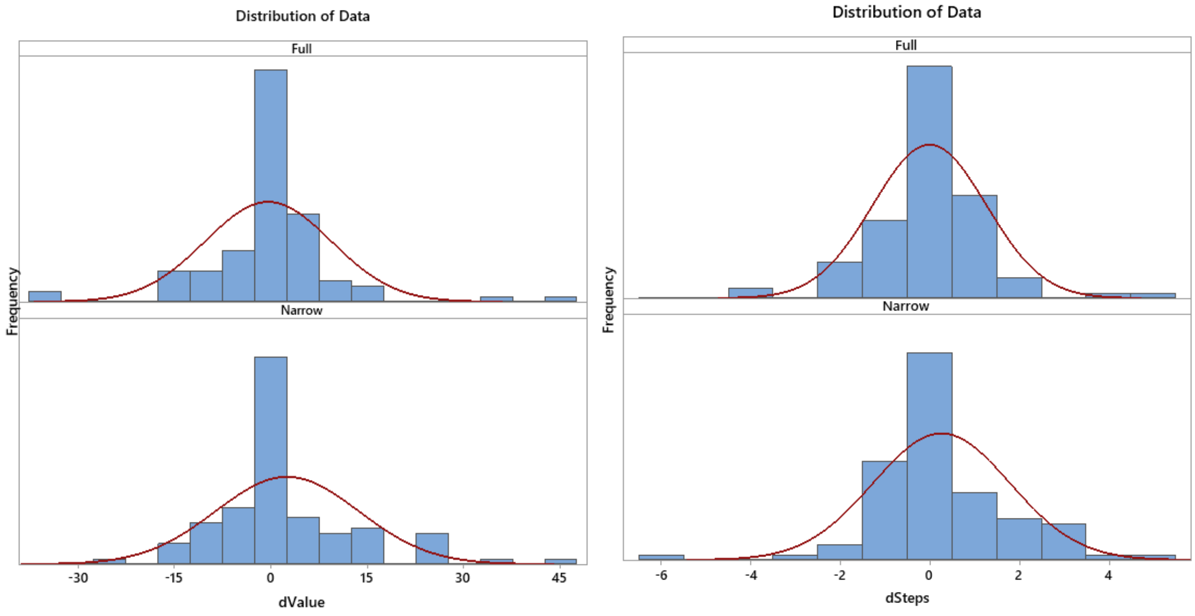
When doing user tests, it is important the experiment is to not induce unduly risks to the users. Therefore, when the COVID-19 pandemic started spreading into Norway, and restrictions were put in place to limit the spread of said disease, it was not deemed ethically defensible to perform experiments where one apparatus would be placed on multiple users faces, the faces being one of the main vectors of infection. With the pandemic not ending during the research period, no experiments could therefore be undergone.

4 Results

With no experiments performed due to the COVID-19 pandemic, naturally no results were gathered from test persons. To demonstrate the capabilities of the application and the statistical methods available to facilitate these results, four test passes were done by the author, for which results will be presented. Measurements are intended to illustrate functionality and are in no way intended to be representative of results found if the experiment was undergone as planned.

A wide array of statistical test can be undergone on the data, relating directly to the research questions (RQs) or not. In order to determine results, a number for the error needs to be calculated, so that accuracy can be quantified. This can be done simply by taking the ChosenValue and subtracting CorrectValue into a difference: dValue.

Considering first the overall case of comparing accuracy of Full and Narrow cases, thus focusing on RQ3, a 2-Sample t test was done to determine difference in means between the Full and the Narrow tests, the distribution of which can be seen in Figure 17a. There is a strong tendency (P=0.055) for there being a *difference* between the two groups, with the mean of -0.42 and 2.52 for Full and Narrow tests, respectively. A 2-Sample Standard Deviation Test for dValue determined the standard deviations of the Full and Narrow tests to not be significantly different (P = 0.480). To determine whether the tests are more *accurate*, a 2-Sample t test was done using the absolute value of dValue, thus finding the absolute difference from the correct value. In this test, it was tested to see if the absolute error made in Full tests were less than Narrow, for which a non-significant tendency was found (P = 0.091).



a) Distribution of dValue. b) Distribution of dSteps.

Figure 17 Distribution of Data, Narrow and Full. Red curve is a fitted bell curve.

One issue when working with the data for both Colour and Height, is the fact that these are not directly comparable. Since the spans of possible values for each is of unequal sizes, any test considering these values as-is together is likely to be skewed towards the result of the variable with the larger spans, in this case Colour. This effect is exemplified by doing a 2-Sample Standard Deviation Test comparing the standard deviations of the Height and Colour data sets, where a significant difference ($P < 0.001$) was found, with Height having a standard deviation of 2.99 while Colour had 14.48. Finding a way to mitigate this is therefore important if the data for the two is to be combined or compared.

The values of error can be normalised into a common unit, using the size of the spans as the normalising factor. Dividing the dValue by this factor, 21 for Height and 61 for Colour, the errors are normalised into a span between -1 and 1. As the sampling in this experiment was "even spread", this can be further multiplied by 7 (Number of Buildings – 1) to get intuitive values. This new value, dSteps, represent the number of building values away the selected answer was, with the maximum value of 7 meaning the highest available value was chosen, while the lowest available value was correct, and vice versa for -7. The distribution of this value is shown in Figure 17b. Using this value, no significant difference between Full and Narrow means could be found ($P = 0.196$).

Considering next Height and Colour separately, thus considering RQ1 and 2 respectively for which the distributions are shown in Figure 18, 2-Sample t-tests were done to determine the effect of Narrow and Full views. In the case of Height, no significant difference in mean was found ($P = 0.282$), while a significant difference was found for Colour ($P = 0.029$). This helps explain the difference between the two tests when seen combined, as colour had more impact, and thus the test was closer to significant, when using non-normalised data. Interestingly, Full tests are not significantly more accurate than Narrow tests ($P = 0.075$), but rather the Full test tended to underestimate (mean = -0.146 steps) while Narrow tests tended to overestimate (mean = 0.604 steps). Similar effects were not found for Height.

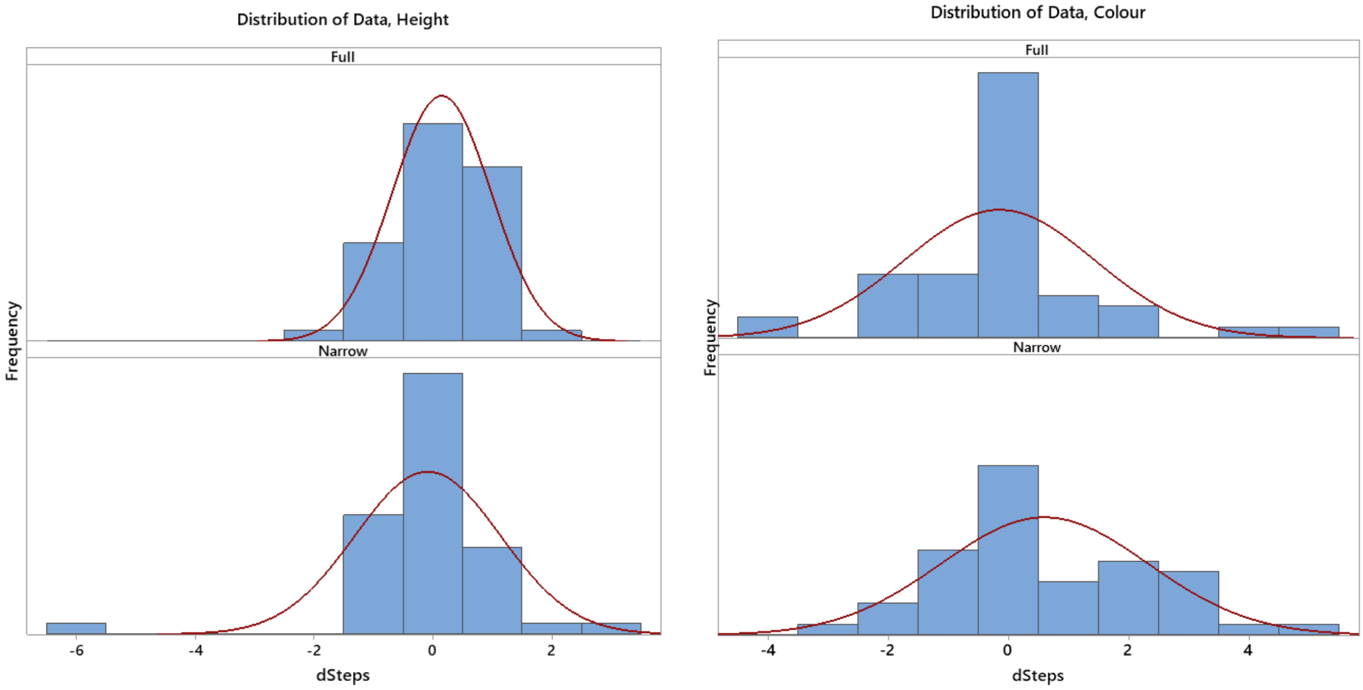


Figure 18 Distribution of Height and Colour data.

2-Sample Standard Deviation tests found no significant difference in standard deviations between Full and Narrow cases neither when viewing normalised Height and Colour together (P = 0.268) nor separately (P = 0.297 and P = 0.638, respectively).

While the above tests are sufficient to answer the research questions, a more nuanced view can be obtained. As an example, the accuracy in the different value spans defined in Table 8 can be determined. Distribution histograms for the different spans are shown in Figure 19, where each span is defined by its MinValue. Using One-Way ANOVAs for Colour and Height spans separately, differences between the means of the three spans can be determined. In both cases, no such differences could be found (P = 0.738 for Height, P = 0.073 for Colour). However, in both cases significant differences between standard deviations could be found using Standard Deviation Tests (Height: P = 0.050, Colour: P < 0.001). For Height, the only difference was between the lowest and the highest value spans, where the standard deviation of the former was lower than the latter. For Colour the lower span had significant differences to both the middle and the high span, with a much lower standard deviation (5.9375, 15.546 and 18.552 respectively). Further dissecting these variables, an interesting finding is that there is a significant (P = 0.004) difference between the span means for Colour in Narrow tests, while not found for any Heights nor Colour in Full tests.

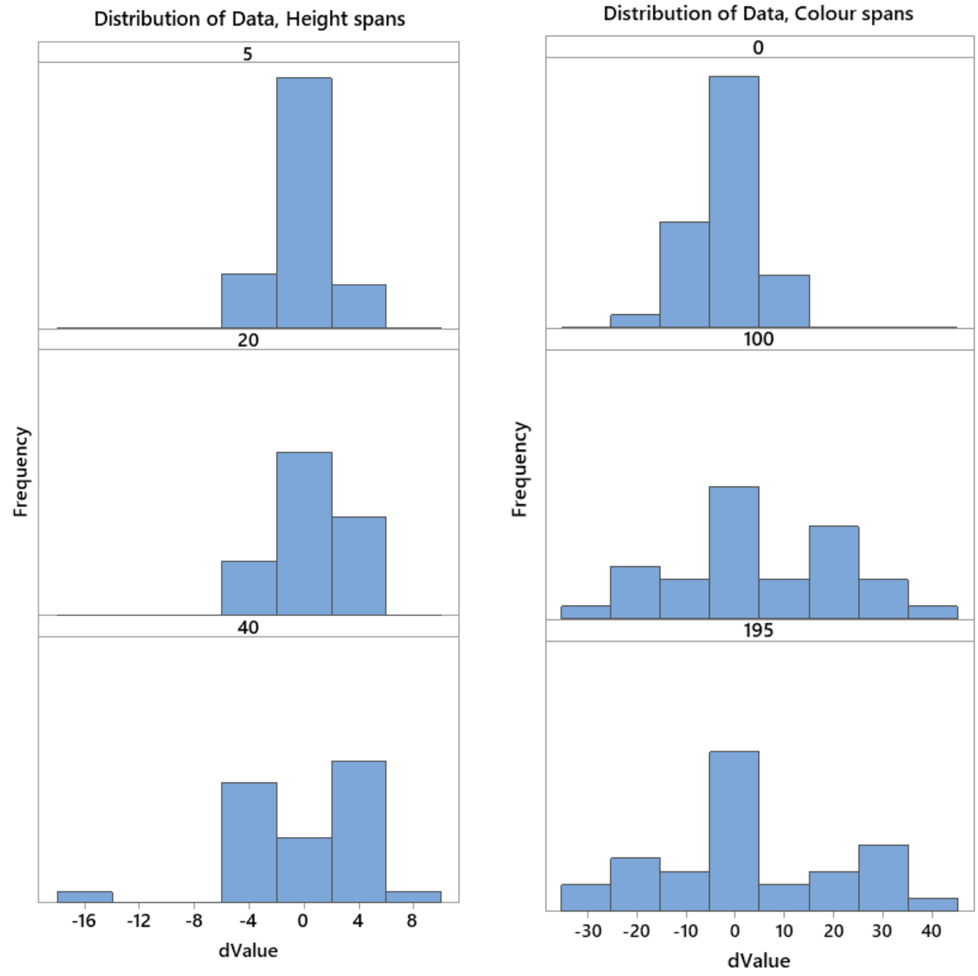


Figure 19 Distribution of Height and Colour span data.

As the application outputs a wide variety of data (see Table 6) there exists a multitude of factors that could conceivably be investigated, such as the influences of time, although the experiment is not design specifically for this purpose. Using regression tests, no relation was found using regression with absolute normalised error values as the dependent variable and plotting against TimeUsed (P = 0.170, r = -0.10). A negative relation was found however, when plotting against TimeViewingTarget (P = 0.041, r = -0.15) seen in Figure 20. This means error rates decreased, and thus accuracy increased, if the user used longer during the exposure phase. This effect was also found when looking at colour tests separately (P = 0.002, r = -0.31) while no such effect was present for height tests (P = 0.951, r = 0.01).

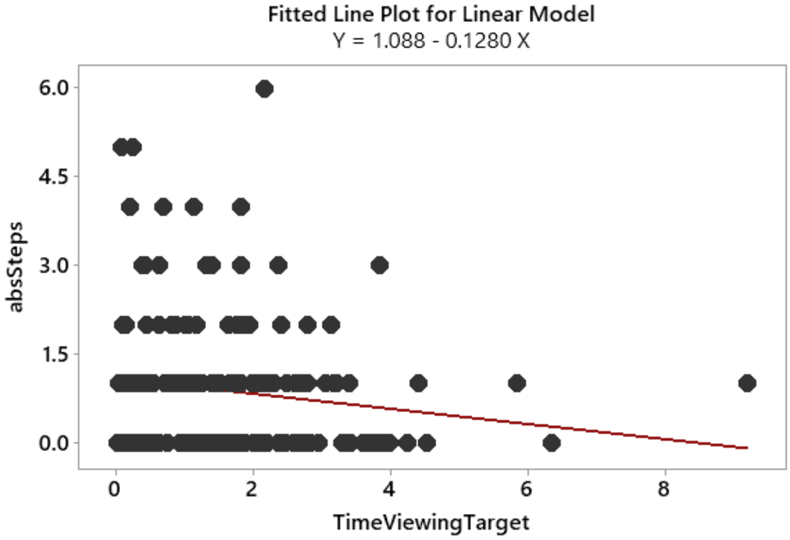


Figure 20 Regression, TimeViewingTarget.

Of interest could be to determine how much peripheral influences plays a part in the perception. If such influences exist, it is expected that moving the head, thus allowing buildings to be in focus, rather than periphery, would lead to a difference in accuracy. With this relation only being relevant for Narrow tests, a regression was done for the Narrow dataset as seen in Figure 21, where no relation was found (P = 0.313, r = 0.11). Using a similar regression for the Full dataset, a significant relation was found (P = 0.002, r = -0.32), and likewise a regression against DegreesUsed proved significant (P = 0.002, r = -0.32). These regression plots are shown in Figure 22. This means accuracy was increased from looking around, when the buildings are placed in the Full view, as expected.

Again, it is important to emphasize that these results are non-representative of expected data. The data is very likely to be biased, and this bias is not controlled for. Such biases could include both fatigue as all tests were done in succession, leading to impaired accuracy in later tests, while increased experience might provide the opposite. It is also likely that the experience gained while designing the application makes biases the results, but to what degree is impossible to determine.

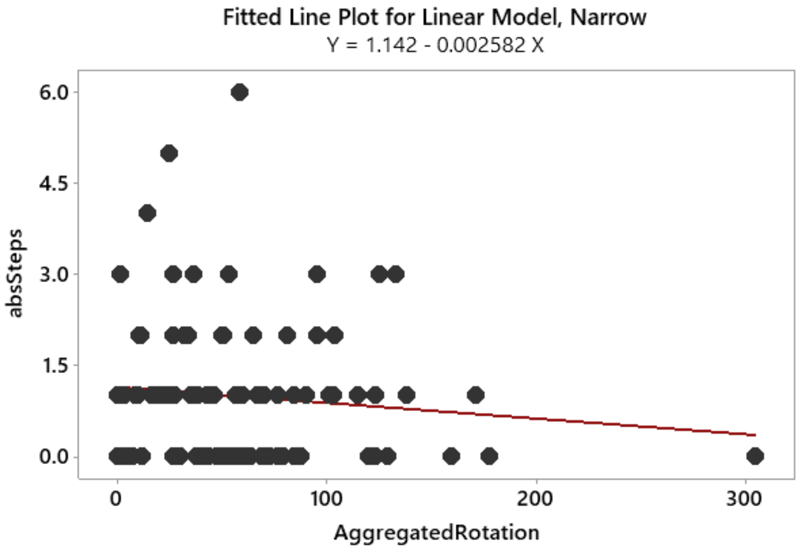


Figure 21 Regression, effect of peripheral.

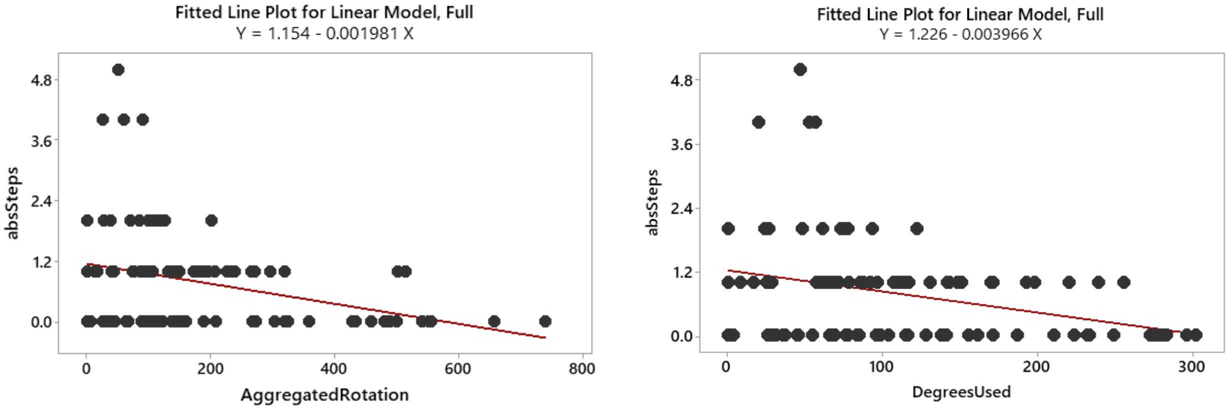


Figure 22 Regression, rotation in Full tests.

5 Discussion

As no experiments could be performed as a result of practical and ethical concerns due to the COVID-19 pandemic, analysing results and setting these in perspective proves impossible. A few test passes were undergone by the author as presented in the Results section, but these should not be seen as representative results. Therefore, any discussion will be speculation with no relation to feedback or results gathered from test persons but can serve help discerning the *expected* results.

Generally, it should be expected that a Narrow search would be more accurate. There are a few reasons for this. One is the idea that when only one dimension differentiates targets and distractors in a visual search, according to feature integration theory, all the items can be processed simultaneously (Treisman and Gelade, 1980). This effect, combined with not needing to physically turn, is likely to reduce time used on Narrow tests, which the experiment is not intended to test. It is also quite possible that it would lead to higher accuracy, if nothing else because of a shorter time since exposure, thus having the target dimension fresher in memory, similar to the effect of shortening the retention interval (Bosco *et al.*, 2008; Huang, Chang and Wei, 2010; Dosher, 1999). One interesting effect to look for, could be to see if there is any relation between the time spent searching and the accuracy, and seeing if this relation is the same in both Narrow and Full cases. While a more controlled experiment would be needed to ascertain such a relation accurately, finding a correlation, or lack thereof, could result in interesting further research. Similarly, it could be interesting note whether users that spend more time inspecting the target before the search prove to be more accurate, as previous research suggests (Huang, Chang and Wei, 2010).

Focusing on the possible test types, here Height and Colour, the expected behaviour in both cases is higher accuracy with lower values. In the case of Height, this expectation comes from many sources, including the theory behind perceptual scaling (Flannery, 1971; Krygier, 2007; Stevens, 1957), eye-height scaling possibly utilised in small buildings (Wraga and Proffitt, 2000), and overestimations common with tall buildings due to perspective (Stefanucci and Proffitt, 2009). The experiment is not designed to isolate the effects, so neither the degree to which each influences answers, nor what would be responsible for any potential disagreement with this expectation is necessarily determinable from this experiment. In the case of Colour, the Weber-Fechner law and Steven's power law both claim that darker colours are easier to discern (Fechner, Howes and Boring, 1966; Stevens, 1957), with dark colour being coded as low values.

Whether these expectations hold true remains to be seen, as does whether this is the case both in Narrow *and* Full tests. If so, this would further add to the credibility of the findings leading to these expectations. A potential next experiment could then be to attempt sampling the spans utilising perceptual scaling and seeing if such an effect can be mitigated. If these expectations do not hold, interesting conclusions could be drawn, where influences due to the immersive 360-degree environment have an impact. Theorising these consequences would be both difficult and unproductive in lieu of actual results, but should such results occur, it would open for further research for the generality of these effects.

While the expectation is that Narrow tests are more accurate, finding the opposite or no difference are certainly both possible outcomes. Any of these results would yield interesting implications to the usability of height and/or achromatic colour as thematic variables in such applications, maps or otherwise. In the case of the results following the expectation, it is likely that applications should refrain from coding information through any such variable not immediately visible to the user, even if the user is aware of having to turn around to see all information. If the inverse is true, this raises some interesting questions regarding how the information is stored in memory, and the search mechanisms in place to gather such a result but is similarly influential in the design decisions of applications using these variables. If no difference is found, it opens up these decisions to the designers fully, being free to immerse the user fully, with no evidence that this would harm the accuracy of the decisions they make.

6 Conclusions

This paper has proposed an experimental setup in order to answer the research questions. An application was made using Unity, so that this could be tested using an Oculus Quest as a VR device. Unfortunately, due to the COVID-19 pandemic and subsequent lockdowns and other related concerns, both legal and ethical, none such experiments could be performed, nor pilot testing done to refine said setup. Expected outcomes have been discussed based on theory from fields such as cartography, perception, and visual searches. Research is lacking for how these fields interact with a fully immersive environment, a topic this paper was meant to contribute to. The paper might therefore serve as a good starting point for further research, whether such research utilises the application created for this paper, being freely available online, or similar experiments undergone through other means.

7 Further Research

As the experiment put forth from this paper could not be performed, and thus no results could be gathered, an obvious avenue of further research would be to perform the experiment in a later study, when conditions allow. Likewise, few topics of further research can be found as a result of findings from this paper. However, this is not to imply no such paths can be recommended.

As a natural follow-up study to this study, would be to determine the *time spent* finding the answers, as opposed to the accuracy in finding it, these being the two primary heuristics of visual searches (Wolfe and Horowitz, 2008; Treisman and Gelade, 1980). Doing this is little more than a change of phrasing when presenting the experiment from "use as long as you need" to "as quickly as possible", or similar. Further, as an experiment akin to the one presented here could help identify if accuracy is different in an immersive environment, using experimental setups where the target might not be present when searching could help ascertain a threshold for the difference needed to not only find the closest, but to determine the exact equal.

Any number of variables held constant in the current experimental setup could be changed, with the overview of these variables found in Table 4. A multitude of other variables could also conceivably exist, even if not specified in this paper. Changing these variables and comparing findings could lead to interesting conclusions as to the generality of findings, and the effects of the changed variables. This could be done using said variable as an active variable in an experiment, or comparing between two separate experiments, if conditions are otherwise held constant. As the source code for the application created for this purpose is made available and changing of any number of variables can be done with little effort, this can be facilitated. One variable that naturally could be considered as an active variable in further experiments is chromatic colours, as the current setup is focused entirely on the use of achromatic colours.

Besides further experiments investigating the effects of different variables, work should be undergone to determine the effects this has on design decisions in practise. For example, if found that any particular variable is unsuited for use in a 360-degree environment when viewing the target values beforehand, other display options must be considered. As is commonly the case in digital media, options here are plentiful, and both interaction methods and display options can be considered and tested to find the optimal solutions for displaying thematic information in such an environment.

8 References

Abramov, I., Gordon, J. and Chan, H. (1991) Color appearance in the peripheral retina: effects of stimulus size, *J. Opt. Soc. Am.*, 8(2), pp. 404-414. doi: https://doi.org/10.1364/JOSAA.8.000404.

Anderson, S. J., Mullen, K. T. and Hess, R. F. (1991) Human peripheral spatial resolution for achromatic and chromatic stimuli: limits imposed by optical and retinal factors, 442(1), pp. 47-64. doi: 10.1113/jphysiol.1991.sp018781.

Arisona, S. (2018) *The CityEngine VR Experience for Unreal Engine: A Virtual Reality Experience for Urban Planning Applications* (Accessed: 15th of December 2019).

Baddeley, A. (2012) Working Memory: Theories, Models, and Controversies, *Annual Review of Psychology*, 63(1), pp. 1-29. doi: 10.1146/annurev-psych-120710-100422.

Baldwin, J. *et al.* (2016) The Perceived Size and Shape of Objects in Peripheral Vision, *i-Perception*, 7(4), pp. 204166951666190. doi: 10.1177/2041669516661900.

Bebko, A. O. and Troje, N. F. (2019) The size of objects in visual space compared to pictorial space, *Journal of Vision*, 19(10), pp. 16-16. doi: 10.1167/19.10.16.

Bertin, J. (1981) *Graphics and graphic information-processing*. New York: Walter de Gruyter.

Bingham, G. P. (1993) Perceiving the size of trees: Biological form and the horizon ratio, 54(4), pp. 485-495. doi: 10.3758/bf03211771.

Bosco, A. *et al.* (2008) Assessing human reorientation ability inside virtual reality environments: the effects of retention interval and landmark characteristics, *Cognitive Processing*, 9(4), pp. 299-309. doi: 10.1007/s10339-008-0210-6.

Bryson, S. (2013) Virtual Reality: A Definition History - A Personal Essay, *arXiv.org*.

Cambridge Cognition (n.d.) *Delayed Matching to Sample*. Available at: https://www.cambridgecognition.com/cantab/cognitive-tests/memory/delayed-matching-to-sample-dms (Accessed: 23rd May 2020).

Carlson, V. R. (1960) Overestimation in Size-Constancy Judgments, 73(2), pp. 199. doi: 10.2307/1419897.

Carvalho, P. *et al.* (2017) VR Rio 360: The Challenges of Motion Sickness in VR Environments, *Cham*. Springer International Publishing, pp. 495-504.

Chandler, D. and Munday, R. (2020) peripheral vision: Oxford University Press. doi: 10.1093/acref/9780198841838.013.2007.

Chen, C.-F., Lin, C.-C. and Huang, K.-C. (2014) EFFECTS OF SPACING BETWEEN ITEMS AND VIEW DIRECTION ON ERRORS IN THE PERCEIVED HEIGHT OF A ROTATED 3-D FIGURE 1 ,2, 119(1), pp. 215-227. doi: 10.2466/24.27.pms.119c12z6.

Chudasama, Y. (2010) Delayed (Non)Match-to-Sample Task, in Stolerman, I. P. (ed.) *Encyclopedia of Psychopharmacology.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 372-372.

Cools, R. (2010) Delayed Match-to-Sample Test, in Stolerman, I. P. (ed.) *Encyclopedia of Psychopharmacology.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 371-372.

Cowan, N. (2008) Chapter 20 What are the differences between long-term, short-term, and working memory? Elsevier, pp. 323-338.

Cowan, N. (2013) Working Memory *Encyclopedia of the Mind.* SAGE Publications, Inc, pp. 786-791.

Dixon, M. W. and Proffitt, D. R. (2002) Overestimation of Heights in Virtual Reality is Influenced more by Perceived Distal Size than by the 2-D versus 3-D Dimensionality of the Display, *Perception*, 31(1), pp. 103-112. doi: 10.1068/p3053.

DiZio, P. and Lackner, J. R. (1997) Circumventing side effects of immersive virtual environments, *Advances in human factors/ergonomics*, 21, pp. 893-896.

Dosher, B. A. (1999) Item Interference and Time Delays in Working Memory: Immediate Serial Recall, *International Journal of Psychology*, 34(5-6), pp. 276-284. doi: 10.1080/002075999399576.

Doyle, C. (2016) virtual reality (4 edn.): Oxford University Press.

Dużmańska, N., Strojny, P. and Strojny, A. (2018) Can Simulator Sickness Be Avoided? A Review on Temporal Aspects of Simulator Sickness, *Frontiers in Psychology*, 9. doi: 10.3389/fpsyg.2018.02132.

Fechner, G. T., Howes, D. H. and Boring, E. G. (1966) *Elements of psychophysics*. Holt, Rinehart and Winston New York.

Fernandes, A. S. and Feiner, S. K. (2016) Combating VR sickness through subtle dynamic field-of-view modification, *2016*. IEEE.

Flannery, J. J. (1971) THE RELATIVE EFFETIVENESS OF SOME COMMON GRADUATED POINT SYMBOLS IN THE PRESENTATION OF QUANTITATIVE DATA.

Fortune Business Insights (2019) *Virtual Reality Market Size, Share & Industry Analysis, By Offering (Hardware, Software), By Technology (Nonimmersive, Semi-Immersive), By Industry Vertical (Gaming & Entertainment Media, Healthcare, Education, Automotive, Aerospace & Defense, Manufacturing), By Application (Training & Simulation, Educational, Attraction, Research & Development) and Regional Forecast, 2019 - 2026* Available at: https://www.fortunebusinessinsights.com/industry-reports/virtual-reality-market-101378 (Accessed: 26th April 2020).

Google (2016) Google Earth VR. Available at: https://arvr.google.com/earth/.

Gordon, J. and Abramov, I. (1977) Color vision in the peripheral retina. II. Hue and saturation, *J Opt Soc Am*, 67(2), pp. 202-207. doi: 10.1364/josa.67.000202.

Grobelski, B., Walczak, D. A. and Pasieka, Z. (2010) New ways of visualization in laparoscopic surgery, *Videosurgery and Other Miniinvasive Techniques*, 3, pp. 120-128. doi: 10.5114/wiitm.2010.16425.

Hansen, T., Pracejus, L. and Gegenfurtner, K. R. (2009) Color perception in the intermediate periphery of the visual field, *Journal of Vision*, 9(4), pp. 26-26. doi: 10.1167/9.4.26.

Hardawar, D. (2019) *Oculus Quest review: VR freedom comes at a cost*. Available at: https://www.engadget.com/2019/04/30/oculus-quest-review-wireless-vr/?guccounter=1&guce_referrer=aHR0cHM6Ly9lbi53aWtpcGVkaWEub3JnLw&guce_referrer_sig=AQAAAKDK8_kK4NpkG_GZkLiOv_xftcwLY7RE2fFJPTdveEuC_wzmm1lHbVZMX10MpPd871x9LykUd6r8rMtPd8lyXqAJ-wMKDwpDHsNVMPvOcVrg-3TWkOuWbvO8G5WXXXkLqpSR-X6WkQ13nbW7X82gfx2fNbmMzrsyGo4STIJDa2YK (Accessed: 26th April 2020).

Hettinger, L. J. and Riccio, G. E. (1992) Visually Induced Motion Sickness in Virtual Environments, 1(3), pp. 306-310. doi: 10.1162/pres.1992.1.3.306.

Higham, M. (2019) *New Oculus VR Headsets Coming In Spring For $400: Rift S And Quest Details*. Available at: https://www.gamespot.com/articles/new-oculus-vr-headsets-coming-in-spring-for-400-ri/1100-6465711/ (Accessed: 26th April 2020).

Hornsey, R. and Hibbard, P. (2018) Shape and Size Constancy in Consumer Virtual Reality, *Journal of Vision*, 18(10), pp. 515-515. doi: 10.1167/18.10.515.

Huang, K.-C., Chang, W.-T. and Wei, W.-L. (2010) Effects of visual field, exposure time, and set size on icon search with varied delays using an LCD monitor, *Journal of the Society for Information Display*, 18(6), pp. 427. doi: 10.1889/jsid18.6.427.

Ibanez, M.-B. *et al.* (2016) Support for Augmented Reality Simulation Systems: The Effects of Scaffolding on Learning Outcomes and Behavior Patterns, *IEEE Transactions on Learning Technologies*, 9(1), pp. 46-56. doi: 10.1109/tlt.2015.2445761.

Jang, S. *et al.* (2017) Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment, *Computers & Education*, 106, pp. 150-165. doi: 10.1016/j.compedu.2016.12.009.

Jansen, Y. and Hornbæk, K. (2016) A Psychophysical Investigation of Size as a Physical Variable, *IEEE Transactions on Visualization and Computer Graphics*, 22(1), pp. 479-488. doi: 10.1109/TVCG.2015.2467951.

Johnson, M. A. (1986) Color Vision in the Peripheral Retina, *Optometry and Vision Science*, 63(2). Available at: https://journals.lww.com/optvissci/Fulltext/1986/02000/Color_Vision_in_the_Peripheral_Retina.3.aspx.

Jou, M. and Wang, J. (2013) Investigation of effects of virtual reality environments on learning performance of technical skills, *Computers in Human Behavior*, 29(2), pp. 433-438. doi: 10.1016/j.chb.2012.04.020.

Kamel Boulos, M. N. *et al.* (2017) From urban planning and emergency training to Pokémon Go: applications of virtual reality GIS (VRGIS) and augmented reality GIS (ARGIS) in personal, public and environmental health, *International Journal of Health Geographics*, 16(1). doi: 10.1186/s12942-017-0081-0.

Keehner, M. *et al.* (2008) Spatial Reasoning With External Visualizations: What Matters Is What You See, Not Whether You Interact, *Cognitive Science: A Multidisciplinary Journal*, 32(7), pp. 1099-1132. doi: 10.1080/03640210801898177.

Kim, E. and Shin, G. (2018) Head Rotation and Muscle Activity When Conducting Document Editing Tasks with a Head-Mounted Display, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1), pp. 952-955. doi: 10.1177/1541931218621219.

Kolasinski, E. M. (1995) *Simulator sickness in virtual environments*. US Army Research Institute for the Behavioral and Social Sciences.

Koulieris, G.-A. *et al.* (2017) Accommodation and comfort in head-mounted displays, *ACM Trans. Graph.*, 36(4), pp. Article 87. doi: 10.1145/3072959.3073622.

Krygier, J. (2007) *Perceptual Scaling of Map Symbols*. Available at: https://makingmaps.net/2007/08/28/perceptual-scaling-of-map-symbols/ (Accessed: 20th of May 2020).

Laviola, J. J. (2000) A discussion of cybersickness in virtual environments, *ACM SIGCHI Bulletin*, 32(1), pp. 47-56. doi: 10.1145/333329.333344.

Lee, E. A.-L. and Wong, K. W. (2014) Learning with desktop virtual reality: Low spatial ability learners are more positively affected, *Computers & Education*, 79, pp. 49-58. doi: 10.1016/j.compedu.2014.07.010.

Lendvay, T. S. *et al.* (2013) Virtual Reality Robotic Surgery Warm-Up Improves Task Performance in a Dry Laboratory Environment: A Prospective Randomized Controlled Study, *Journal of the American College of Surgeons*, 216(6), pp. 1181-1192. doi: 10.1016/j.jamcollsurg.2013.02.012.

Lin, C. J. *et al.* (2015) Effects of Displays on Visually Controlled Task Performance in Three-Dimensional Virtual Reality Environment, *Human Factors and Ergonomics in Manufacturing & Service Industries*, 25(5), pp. 523-533. doi: 10.1002/hfm.20566.

MacEachren, A. M. (1995) *How maps work : representation, visualization, and design*. New York: Guilford Press.

MacEachren, A. M. *et al.* (1999) *Virtual environments for geographic visualization: potential and challenges*. Unpublished paper presented at Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM internation conference on Information and knowledge management. Kansas City, Missouri, USA.

Magnotti, J. F., Goodman, A. M. and Katz, J. S. (2012) Matching to Sample Experimental Paradigm, in Seel, N. M. (ed.) *Encyclopedia of the Sciences of Learning.* Boston, MA: Springer US, pp. 2104-2106.

Magnussen, S. (2013) Visual Working Memory *Encyclopedia of the Mind.* SAGE Publications, Inc, pp. 765-769.

Matatko, A., Bollmann, J. and Müller, A. (2011) Depth Perception in Virtual Reality, in Kolbe, T. H., König, G. and Nagel, C. (ed.) *Advances in 3D Geo-Information Sciences.* Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 115-129.

McLeod, S. A. (2012) Working Memory, *Simply Psychology*. Available at: https://www.simplypsychology.org/working%20memory.html.

Merchant, Z. *et al.* (2014) Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis, *Computers & Education*, 70, pp. 29-40. doi: 10.1016/j.compedu.2013.07.033.

Merhi, O. *et al.* (2007) Motion Sickness, Console Video Games, and Head-Mounted Displays, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 49(5), pp. 920-934. doi: 10.1518/001872007x230262.

Merriam-Webster (n.d.) Virtual Reality *Merriam-Webster.com dictionary*. Available at: https://www.merriam-webster.com/dictionary/virtual%20reality (Accessed: April 18, 2020).

Milgram, P. and Kishino, F. (1994) A taxonomy of mixed reality visual displays, *IEICE TRANSACTIONS on Information and Systems*, 77(12), pp. 1321-1329.

Milman, N. B. (2018) Defining and Conceptualizing Mixed Reality, Augmented Reality, and Virtual Reality.(Ends and Means), *Distance Learning*, 15(2), pp. 55.

Moss, J. D. and Muth, E. R. (2011) Characteristics of Head-Mounted Displays and Their Effects on Simulator Sickness, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(3), pp. 308-319. doi: 10.1177/0018720811405196.

Mourant, R. R. and Thattacherry, T. R. (2000) Simulator Sickness in a Virtual Environments Driving Simulator, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 44(5), pp. 534-537. doi: 10.1177/154193120004400513.

Nagel, H. R., Granum, G. and Musaeus, P. (2014) Methods for Visual Mining of Data in Virtual Reality.

Ng, A. K. T., Chan, L. K. Y. and Lau, H. Y. K. (2016) Depth Perception in Virtual Environment: The Effects of Immersive System and Freedom of Movement, in Lackey, S. and Shumaker, R. (ed.) *Virtual, Augmented and Mixed Reality*, *Cham, 2016//*. Springer International Publishing, pp. 173-183.

Nguyen, M.-T. *et al.* (2016) Applying Virtual Reality in City Planning, *Cham*. Springer International Publishing, pp. 724-735.

Oculus (n.d.) *Oculus Quest*. Available at: https://www.oculus.com/quest/ (Accessed: 26th April 2020).

Osnes, E. (2019) *Thematic Maps in Virtual Reality – Proposals and Implications for Visual Variables and Non-Temporal Animations*. Department of Civil and Environmental Engineering, NTNU -- Norwegian University of Science and Technology.

Proulx, M. J. (2013) Visual Search *Encyclopedia of the Mind.* SAGE Publications, Inc, pp. 762-765.

Raddatz, K., Uhlarik, J. and Jordan, K. (2001) Perceived Size in Virtual Environments: The Role of Pictorial Depth Cues, 45(18), pp. 1404-1408. doi: 10.1177/154193120104501819.

Radkowski, R., Herrema, J. and Oliver, J. (2015) Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty, *International Journal of Human-Computer Interaction*, 31(5), pp. 337-349. doi: 10.1080/10447318.2014.994194.

Reason, J. T. and Brand, J. J. (1975) *Motion sickness*. Oxford, England: Academic Press.

Rebenitsch, L. and Owen, C. (2017) Evaluating Factors Affecting Virtual Reality Display, *Cham*. Springer International Publishing, pp. 544-555.

Renkewitz, H. and Alexander, T. (2007) *Perceptual issues of augmented and virtual environments*. FGAN-FKIE WACHTBERG (GERMANY).

Renner, R. S., Velichkovsky, B. M. and Helmert, J. R. (2013) The Perception of Egocentric Distances in Virtual Environments - A Review, *ACM Computing Surveys*, 46(2), pp. 23:21-23:40. doi: 10.1145/2543581.2543590.

Rogers, C. and Lyons, B. (2020) 2019 Year In Review: Digital Games and Interactive Media, *SuperData*. Available at: https://www.superdataresearch.com/2019-year-in-review.

Schnürer, R. (2014) *Creating Charts and Legends for 3D Atlas Maps - A Mashup of D3.js, osgEarth, and the Chromium Embedded Framework* FOSS4G, Open Source Geospatial Foundation (OSGeo). doi: 10.5446/31672.

Serge, S. R. and Fragomeni, G. (2017) Assessing the Relationship Between Type of Head Movement and Simulator Sickness Using an Immersive Virtual Reality Head Mounted Display: A Pilot Study, *Cham*. Springer International Publishing, pp. 556-566.

Slocum, T. A. *et al.* (2009) *Thematic cartography and geovisualization*. 3rd ed. edn. Upper Saddle River, N.J: Pearson Prentice Hall.

Somolinos, J. (2017) *How to design common/basic interactions in VR*. Available at: https://medium.com/@josesomolinos/how-to-design-common-basic-interactions-in-vr-f958cf160cfc (Accessed: 21th June 2020).

Somrak, A. *et al.* (2019) Estimating VR Sickness and user experience using different HMD technologies: An evaluation study, *Future Generation Computer Systems*, 94, pp. 302-316. doi: 10.1016/j.future.2018.11.041.

Stefanucci, J. K. and Proffitt, D. R. (2009) The roles of altitude and fear in the perception of height, *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), pp. 424-438. doi: 10.1037/a0013894.

Stevens, S. S. (1957) On the psychophysical law, *Psychological review*, 64(3), pp. 153-181.

Stins, J. F. *et al.* (2013) On the Role of Vertical Texture Cues in Height Perception, 25(4), pp. 357-368. doi: 10.1080/10407413.2013.842094.

Stoffregen, T. A. *et al.* (2008) Motion Sickness and Postural Sway in Console Video Games, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(2), pp. 322-331. doi: 10.1518/001872008x250755.

Suh, A. and Prophet, J. (2018) The state of immersive technology research: A literature analysis, *Computers in Human Behavior*, 86, pp. 77-90. doi: 10.1016/j.chb.2018.04.019.

Susumu, T., Chusi, K. and Tsutomu, M. (2006) Proportional Symbol Mapping in R, *Journal of Statistical Software*, 15. doi: 10.18637/jss.v015.i05.

Taçgın, Z. (2020) The perceived effectiveness regarding Immersive Virtual Reality learning environments changes by the prior knowledge of learners, *Education and Information Technologies*. doi: 10.1007/s10639-019-10088-0.

Treisman, A. M. and Gelade, G. (1980) *A Feature-Integration Theory of Attention* Cognitive Psychology.

Tufte, E. R. (2001) *The visual display of quantitative information*. 2nd ed. edn. Cheshire, Conn: Graphics Press.

Twedt, E., Crawford, L. E. and Proffitt, D. R. (2012) Memory for target height is scaled to observer height, *Memory & Cognition*, 40(3), pp. 339-351. doi: 10.3758/s13421-011-0166-0.

Tyner, J. (2010) *Principles of map design*. New York: Guilford Press.

Unity (2019): Unity Technologies. Available at: unity.com.

Wagner, M. (2013) Perceptual Constancy *Encyclopedia of the Mind.* SAGE Publications, Inc, pp. 580-582.

Wijayasekara, D., Linda, O. and Manic, M. (2011) CAVE-SOM: Immersive visual data mining using 3D Self-Organizing Maps, *The 2011 International Joint Conference on Neural Networks*, *31 July-5 Aug. 2011*. pp. 2471-2478.

Wolfe, J. and Horowitz, T. S. (2008) Visual search, *Scholarpedia*, 3, pp. 3325. Available at: http://www.scholarpedia.org/article/Visual_search.

Wolfe, J. M. (1994) Guided Search 2.0 A revised model of visual search, *Psychonomic bulletin & review*, 1(2), pp. 202-238. doi: http://dx.doi.org/10.3758/BF03200774.

Wolfe, J. M. and Van Wert, M. J. (2010) Varying Target Prevalence Reveals Two Dissociable Decision Criteria in Visual Search, *Current Biology*, 20(2), pp. 121-124. doi: 10.1016/j.cub.2009.11.066.

Wolfe, J. M. (2012) Saved by a Log:How Do Humans Perform Hybrid Visual and Memory Search?, *Psychological Science*, 23(7), pp. 698-703. doi: 10.1177/0956797612443968.

Wraga, M. and Proffitt, D. R. (2000) Mapping the Zone of Eye-Height Utility for Seated and Standing Observers, *Perception*, 29(11), pp. 1361-1383. doi: 10.1068/p2837.

Zhang, X. *et al.* (2017) How virtual reality affects perceived learning effectiveness: a task–technology fit perspective, pp. 1-9. doi: 10.1080/0144929x.2016.1268647.

Zhao, M. Y., Ong, S. K. and Nee, A. Y. C. (2016) An Augmented Reality-Assisted Therapeutic Healthcare Exercise System Based on Bare-Hand Interaction, *International Journal of Human–Computer Interaction*, 32(9), pp. 708-721. doi: 10.1080/10447318.2016.1191263.

Appendix

Appendix 1

Versioning

The application is developed using Unity version 2019.2.8f1 and was working for Oculus Quest build 16.0. It is likely to work in later versions of both programs, but this remains untested.

Download instructions

In order to access the application git, git LFS (Large File Storage) and unity must be downloaded and installed.

Git: *https://git-scm.com/downloads*

Git LFS: *https://git-lfs.github.com/*

Unity: *https://unity3d.com/get-unity/download* (Download Unity Hub)

Once all are installed, the application can be accessed from:

https://github.com/EirikOsnes/EirikMastersApp

NOTE: Git LFS breaks if the repository is downloaded, meaning all assets in the application breaks. To avoid this the repository MUST be *cloned*.

Assets

In accordance with their respective licenses, all environmental assets, as seen in Figure 14 have been removed from the repository.

All assets used for this experiment can be found on:

https://3dwarehouse.sketchup.com/model/01ae568b-4bef-4d23-b5b4-6738e3ab9b90/Deciduous-Trees-with-Realistic-Shadow

https://free3d.com/3d-model/bus-setia-negara-low-poly-557005.html

https://free3d.com/3d-model/automobile-v1--84248.html

https://www.mrcutout.com/78-cutouts/people-cutouts/6068-woman-with-a-smartphone-shopping-0008

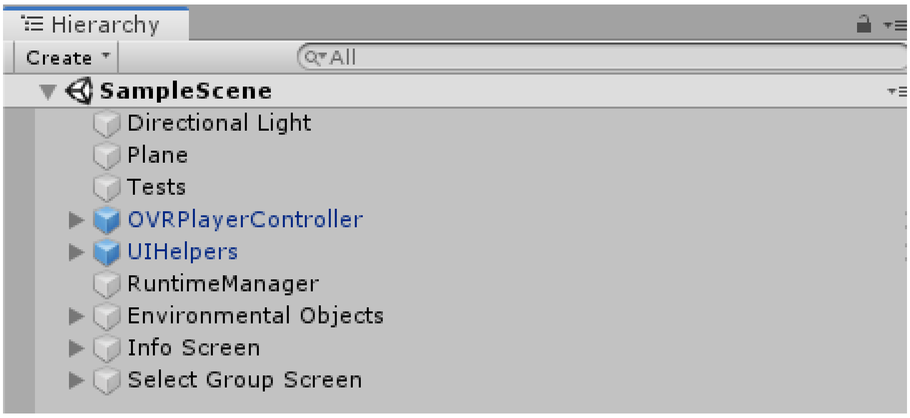
https://www.mrcutout.com/78-cutouts/people-cutouts/7317-businessman-with-a-smartphone-walking-0012

Overview

The application is developed using Unity, with some functionality ran through the Unity editor, and some governed by scripts (C#). For the purpose of creating and running tests – granted no functionality has broken due to software updates etc. – only the editor should be necessary, as long as tests are limited to the available variables. For further modification scripts must be created or edited. All scripts are documented in code and located under Assets/Scripts.

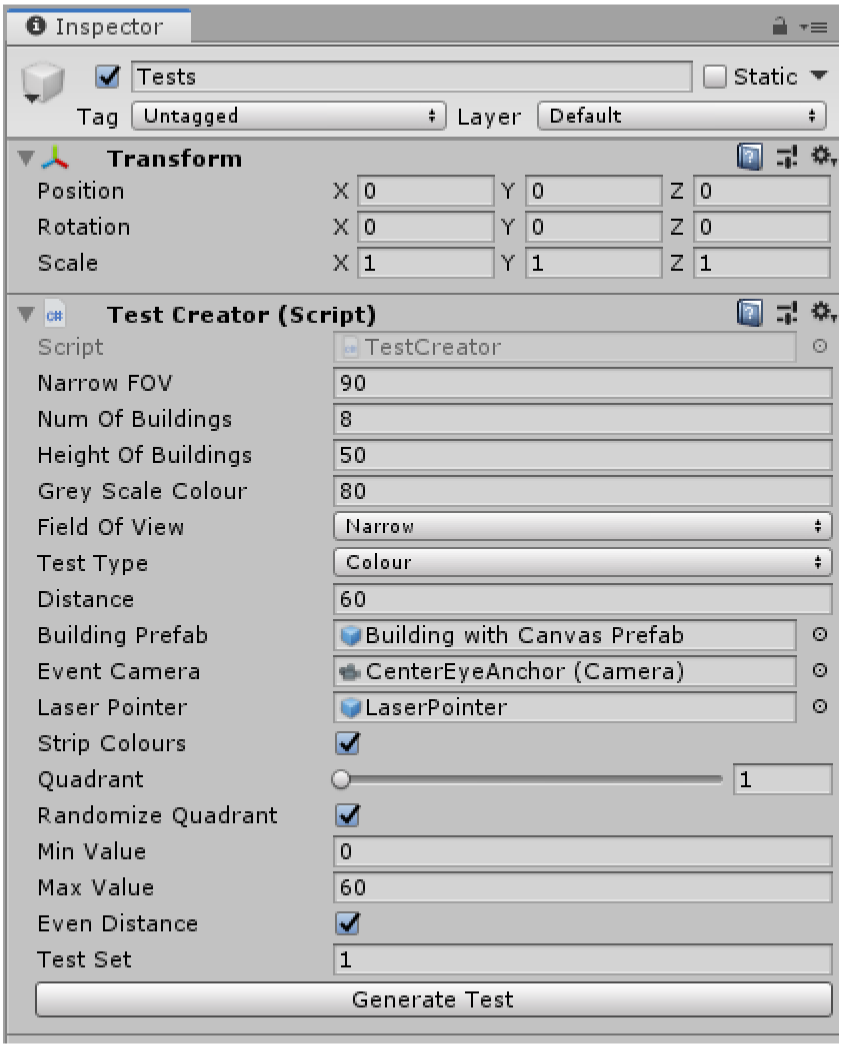
Creating tests

After opening the project in Unity, the hierarchy will look like this:



SampleScene
- Directional Light
- Plane
- Tests
- OVRPlayerController
- UIHelpers
- RuntimeManager
- Environmental Objects
- Info Screen
- Select Group Screen

Manual creation:
Selecting "Tests" should open the following view in the inspector:



☑ Tests

Tag: Untagged — Layer: Default

Transform

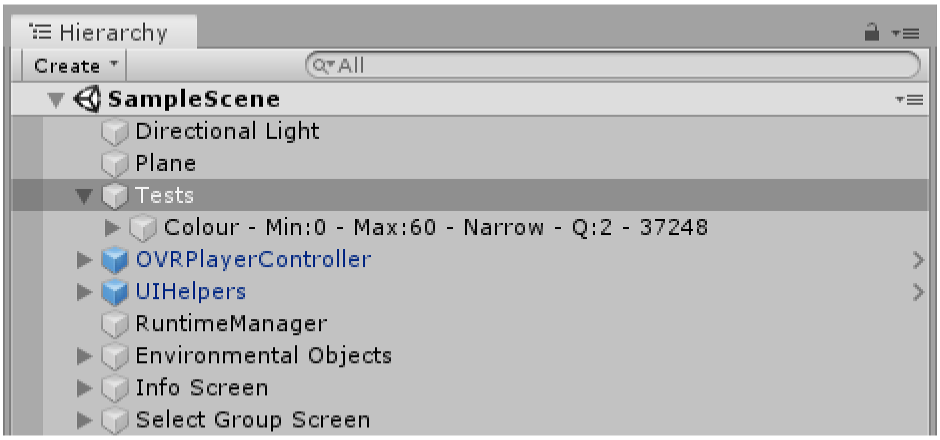
	X	Y	Z
Position	0	0	0
Rotation	0	0	0
Scale	1	1	1

Test Creator (Script)

Script	TestCreator
Narrow FOV	90
Num Of Buildings	8
Height Of Buildings	50
Grey Scale Colour	80
Field Of View	Narrow
Test Type	Colour
Distance	60
Building Prefab	Building with Canvas Prefab
Event Camera	CenterEyeAnchor (Camera)
Laser Pointer	LaserPointer
Strip Colours	☑
Quadrant	1
Randomize Quadrant	☑
Min Value	0
Max Value	60
Even Distance	☑
Test Set	1

By changing variables here as wanted, a single test can be created. Do note that Test Set must be set carefully if ordering based on test sets are to be done. If Field of View is set to "Both", a Narrow AND a Full test will be created. If "Randomize Quadrant" is checked, "Quadrant" is ignored. "Quadrant" is also ignored for Narrow cases. "Event Camera" and "LaserPointer" should not be changed.

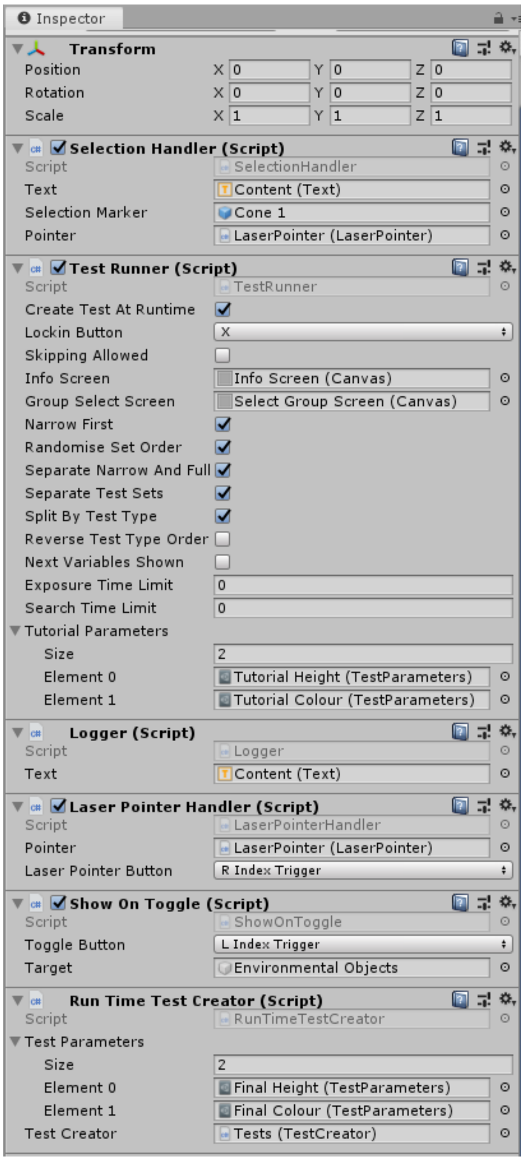
The test is created when “Generate Test” is pressed, at which point buildings will spawn in the view, and one test is listed in the hierarchy like so:



SampleScene
- Directional Light
- Plane
- Tests
  - Colour - Min:0 - Max:60 - Narrow - Q:2 - 37248
- OVRPlayerController
- UIHelpers
- RuntimeManager
- Environmental Objects
- Info Screen
- Select Group Screen

Runtime Creation

Selecting “RuntimeManager” opens in the Inspector a list of components necessary for the running of tests:



Transform
Position X 0 Y 0 Z 0
Rotation X 0 Y 0 Z 0
Scale X 1 Y 1 Z 1

☑ **Selection Handler (Script)**
Script SelectionHandler
Text Content (Text)
Selection Marker Cone 1
Pointer LaserPointer (LaserPointer)

☑ **Test Runner (Script)**
Script TestRunner
Create Test At Runtime ☑
Lockin Button X
Skipping Allowed ☐
Info Screen Info Screen (Canvas)
Group Select Screen Select Group Screen (Canvas)
Narrow First ☑
Randomise Set Order ☑
Separate Narrow And Full ☑
Separate Test Sets ☑
Split By Test Type ☑
Reverse Test Type Order ☐
Next Variables Shown ☐
Exposure Time Limit 0
Search Time Limit 0
Tutorial Parameters
  Size 2
  Element 0 Tutorial Height (TestParameters)
  Element 1 Tutorial Colour (TestParameters)

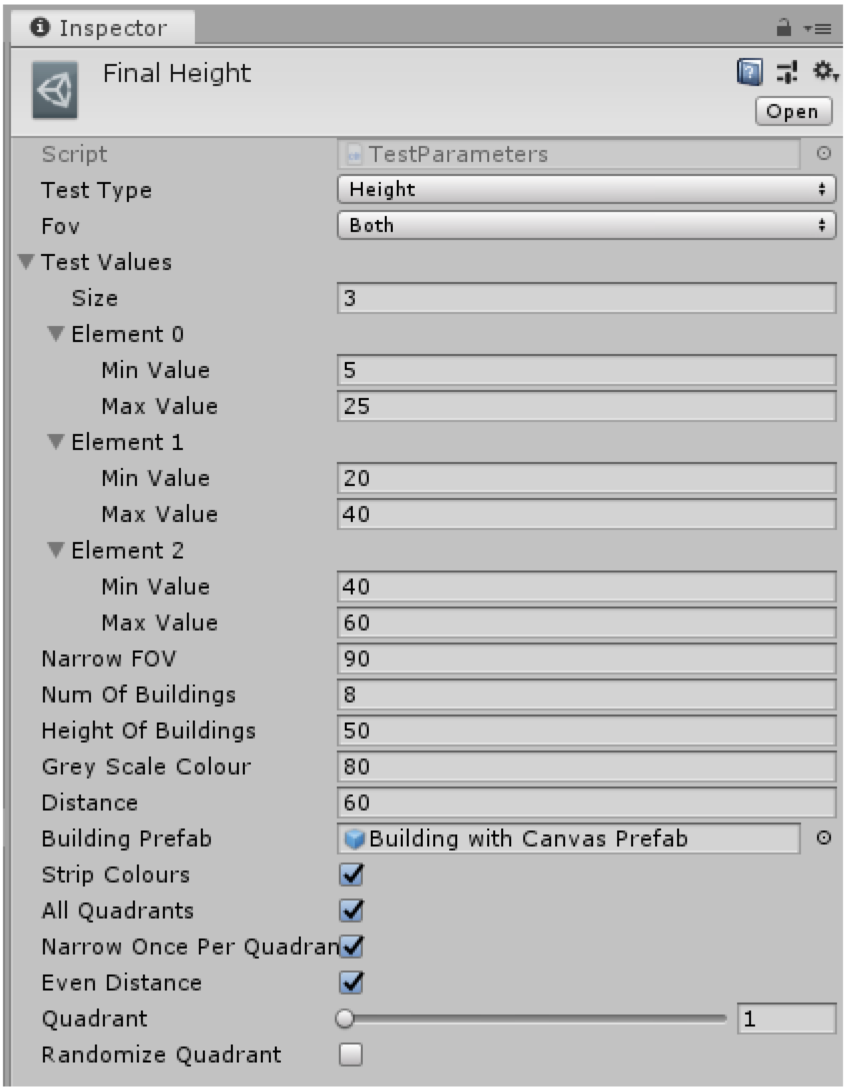
Logger (Script)
Script Logger
Text Content (Text)

☑ **Laser Pointer Handler (Script)**
Script LaserPointerHandler
Pointer LaserPointer (LaserPointer)
Laser Pointer Button R Index Trigger

☑ **Show On Toggle (Script)**
Script ShowOnToggle
Toggle Button L Index Trigger
Target Environmental Objects

Run Time Test Creator (Script)
Script RunTimeTestCreator
Test Parameters
  Size 2
  Element 0 Final Height (TestParameters)
  Element 1 Final Colour (TestParameters)
Test Creator Tests (TestCreator)

The “Run Time Test Creator” is responsible for the creation of tests at runtime. It can take in any number of TestParameters, for which tests will be created. TestParameters feature largely the same options as when creating tests manually. They can be created by right-clicking in the Project view – for example in the Test Parameters folder. In the pop-up menu shown, select Create > ScriptableObject > TestParameter after which options appear in the Inspector. The test parameters utilised in the paper is available, and for height look like so:



Final Height

Script	TestParameters
Test Type	Height
Fov	Both
Test Values	
Size	3
Element 0	
Min Value	5
Max Value	25
Element 1	
Min Value	20
Max Value	40
Element 2	
Min Value	40
Max Value	60
Narrow FOV	90
Num Of Buildings	8
Height Of Buildings	50
Grey Scale Colour	80
Distance	60
Building Prefab	Building with Canvas Prefab
Strip Colours	☑
All Quadrants	☑
Narrow Once Per Quadran	☑
Even Distance	☑
Quadrant	1
Randomize Quadrant	☐

Note that each TestParameter is capable of creating tests for only one active variable (Test Type), but can create any number of spans, determined by “Test Values”.

Once a TestParameters object is created, it can be dragged and dropped onto the “Run Time Test Creator” component or selected by clicking the little circle to the right of the elements in the list.

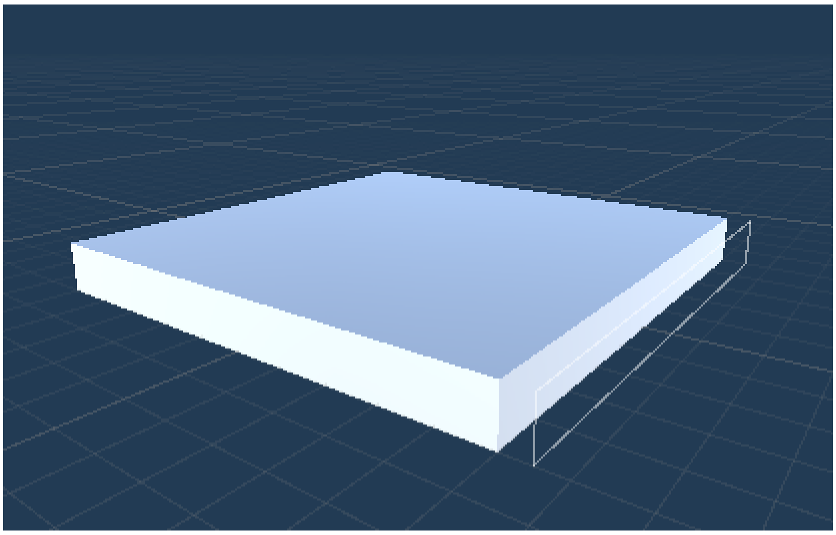
Creating tests this way does not create any objects in the editor, but rather only at runtime.

Changing target model

Selection works by having a canvas (an invisible plane) in front of each building that reacts to the LaserPointer and button clicks. As such to change models, a prefab (prefabricated model) must be created that includes such a canvas. This is done easiest by modifying the existing model.

The model used currently can be found under Assets > Models > Buildings > Building with Canvas Prefab. By selecting this, and dragging while holding CTRL, a copy can be created. By then double-clicking this copy, the scene changes to edit this prefab.

Import the wanted model by dragging it into the scene or the hierarchy. Then select it and set its position to 0,0,0. Move the model so that the bottom is flush with the ground. It should now look similar to the image below (here using a simple cube).

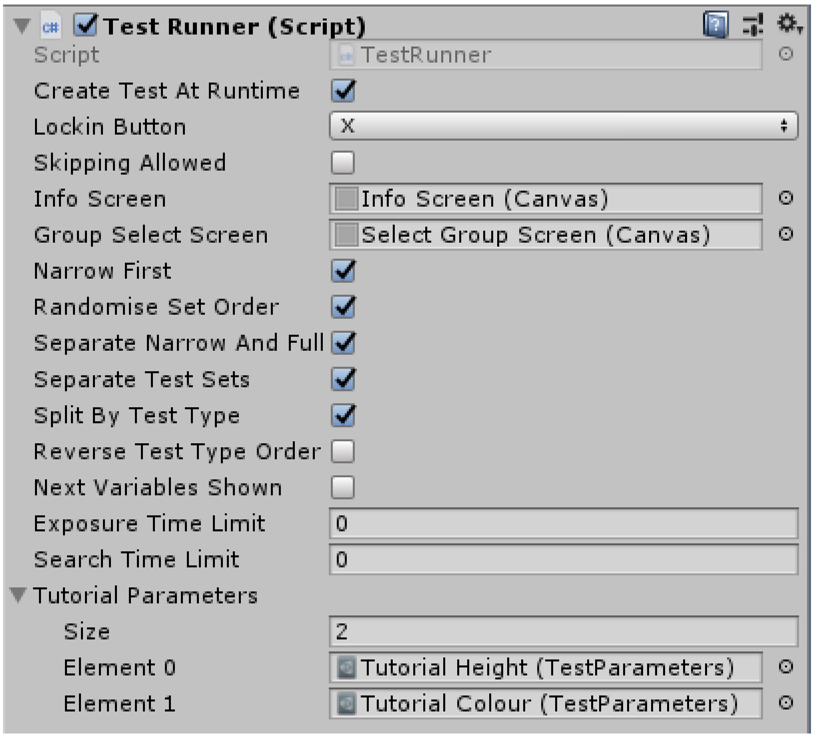


Select the Canvas and move and resize it to be equally wide and tall as your model, sitting in front.

NOTE: If the Height variable is to be used as meters, the model created must be exactly 1 unit tall in Unity. For centimetres, it must be exactly 0.01 units tall, etc.

Screens
The info screen and group select screen need not be changed to run but can be moved as wanted by selecting them in the Hierarchy view and changing their positions. The text shown on the info screen is determined in code (in the TestRunner Script).

Running Tests
Selecting “RuntimeManager” the “Test Runner” component is shown in the Inspector. This component contains the remaining variables that are to be kept consistent over the tests.



Test Runner (Script)

Script: TestRunner
Create Test At Runtime: ☑
Lockin Button: X
Skipping Allowed: ☐
Info Screen: Info Screen (Canvas)
Group Select Screen: Select Group Screen (Canvas)
Narrow First: ☑
Randomise Set Order: ☑
Separate Narrow And Full: ☑
Separate Test Sets: ☑
Split By Test Type: ☑
Reverse Test Type Order: ☐
Next Variables Shown: ☐
Exposure Time Limit: 0
Search Time Limit: 0
Tutorial Parameters
Size: 2
Element 0: Tutorial Height (TestParameters)
Element 1: Tutorial Colour (TestParameters)

"Create Test At Runtime" determines whether manually created tests or tests created at runtime should be utilised in the test pass. Note that if checked, TestParameters must have been added to the "Run Time Test Creator", while if unchecked, tests must exist in the hierarchy. The Lockin Button describes which button is used for changing states, and if changed, new thumbnails on the controllers should be created.

If a tutorial is wanted, TestParameters are added to the "Tutorial Parameters" list, like how it is done for runtime creation.

To run the test, the application must be built onto an Oculus Quest with developer mode activated. See *https://developer.oculus.com/documentation/native/android/mobile-device-setup/* for how to set up developer mode. In Unity, select File > Build Settings and make sure Android is selected. Then click "Player Settings…" and make sure Player > XR Settings > Virtual Reality Supported is checked.

Build to Quest by connecting it to the computer with an USB cable, and selecting File > Build and Run. At which point the application should open on the Quest. If no changes are to be made to the application, this only needs to be done once, after which the application can be found on the Quest in Library > Unknown Sources.

Retrieving data

During a test pass test results are saved to file as tests are completed, and once all tests are completed, a file containing the results of the full test pass is saved. By connecting the Quest to the PC these files can be located in Android/data and selecting the application's folder. All tests are stored in the Tests folder, where each test pass is saved under the year, month, day, and time of the test beginning. All data are saved in JSON files containing all the data shown in Table 6.

NTNU

Norwegian University of
Science and Technology