UNIVERSITY
OF OULU

FACULTY OF INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING
DEGREE PROGRAMME IN COMPUTER SCIENCE AND ENGINEERING

MASTER’S THESIS

Exploring the effectiveness of wall and floor mounted guiding light systems in an underground parking evacuation scenario

Author	Mikko Korkiakoski
Supervisor	Panos Kostakos
Second Examiner	Paula Alavesa

February 2021

Korkiakoski M. (2021) Exploring the effectiveness of wall and floor mounted guiding light systems in an underground parking evacuation scenario. University of Oulu, Faculty of Information Technology and Electrical Engineering, Degree Programme in Computer Science and Engineering. Master's Thesis, 75 p.

ABSTRACT

The usage of virtual reality (VR) in all kinds of applications has been on the rise for several years now. The technology and its applications have matured to a state where it is possible to create realistic and immersive simulations to easily and affordably train people for certain activities that could otherwise be dangerous or costly to implement. These activities include dealing with hazardous materials, evacuating structures or performing surgeries to name a few.

The focus of this thesis is on evaluating two types of differently placed guiding light systems using a VR simulation. The effectiveness of these light systems is examined using an evacuation scenario occurring in an underground parking garage during an emergency. The collected data consisted of heart rate measurements and various performance metrics such as completion time, walk distance, and average speed that were recorded during the simulation, and a questionnaire that was conducted before and after the simulation. The participants were divided into three groups consisting of a baseline group with no assistive lighting and two experiment groups with different types of assistive lights. The simulation was run using a VR HMD (head mounted display) in a glass-walled cubicle.

All the performance results and measurements are discussed and conclusions are made about the lighting system performances, user experiences and the heart rate measurements. The performance results as well as the heart rate measurements showed differences between the three groups. Furthermore, when comparing the participants by their gaming experience, the results showed significantly better performance for those with more gaming experience. Finally, the experiment as a whole is analysed and improvement suggestion are made to it as well as for possible further research into the topic.

Key words: Virtual Reality, underground parking, Head-Mounted Display, evacuation, guiding light system.

Korkiakoski M. (2021) Tutkielma seinä- ja lattia-asennettujen avustusvalojen vaikutuksesta maanalaisen parkkihallin evakuointiskenaariossa. Oulun yliopisto, tieto- ja sähkötekniikan tiedekunta, tietotekniikan tutkinto-ohjelma. Diplomityö, 75 p.

TIIVISTELMÄ

Virtuaalitodellisuuden (VR) käyttö kaikenlaisissa sovelluksissa on ollut kasvussa viime vuosien aikana. Itse tekniikka ja sitä käyttävät sovellukset ovat kehittyneet siihen pisteeseen, että niiden avulla on mahdollista luoda realistisia ja immersiivisiä simulaatioita, joilla voidaan helposti ja edullisesti kouluttaa ihmisiä aktiviteetteihin, jotka voisivat muuten olla vaarallisia tai kalliita toteuttaa. Tällaisia aktiviteetteja ovat mm. toimiminen vaarallisten aineiden kanssa, rakennusten evakuointi ja leikkausten tekeminen.

Tämän diplomityön fokus on kahden eri tavalla asennettujen avustusvalojen vertailu VR-simulaatiossa. Näiden valaistussysteemien vaikutusta arvioidaan maanalaisessa parkkihallissa tapahtuneen vaaratilanteen jälkeistä evakuointia mallintavan VR-simulaation avulla. Kerätty data koostui sykkeenmittauksesta ja useista suorituskykymittareista kuten suoritusajasta, kävellystä matkasta ja keskinopeudesta, jotka tallennettiin simulaation ajalta, sekä kysymysosioista, joista ensimmäinen täytettiin ennen simulaatiota ja toinen sen jälkeen. Testiin osallistujat jaettiin kolmeen ryhmään, joista kahdella testiryhmällä oli molemmilla apunaan toinen testattavista avustusvalojärjestelmistä ja verrokkiryhmään, jolla ei ollut minkäänlaisia avustusvaloja. Simulaatiossa käytettiin virtuaalitodellisuuslaseja ja ne tehtiin suljetussa lasikopissa.

Simulaatiosta suoriutuminen ja mittaustulokset käydään läpi ja niiden perusteella tehdään johtopäätökset valaistusjärjestelmien suorituskyvystä, käyttäjien kokemuksista ja sykkeenmittauksen tuloksista. Sekä suoritusten- että sykkeenmittauksen tuloksissa oli eroja ryhmien välillä. Lisäksi verrattaessa osallistujia näiden pelikokemuksen perusteella, osoittivat tulokset selkeästi parempaa suorituskykyä niillä, joilla oli enemmän kokemusta tietokonepeleistä. Lopuksi tutkimuksen onnistuminen kokonaisuudessaan analysoidaan ja tehdään parannusehdotuksia jo tehtyyn tutkimukseen sekä ehdotuksia mahdolliseen jatkotutkimukseen.

Avainsanat: Virtuaalitodellisuus, maanalainen parkkihalli, virtuaalitodellisuuslasit. evakuointi, avustusvalot.

TABLE OF CONTENTS

ABSTRACT
TIIVISTELMÄ
TABLE OF CONTENTS
FOREWORD
LIST OF ABBREVIATIONS AND SYMBOLS

FOREWORD

This thesis is the culmination of my long journey through my studies. There is no denying that it has taken me a lot longer to reach this point than I ever expected. But if the awful 2020 “covid-19 year” has anything positive to offer me, this is definitely it. Fortunately, I was able to find a subject for my thesis that I found very interesting. I was equally fortunate to have two excellent supervisors who pushed me and gave me space to work on the thesis at the right times. So, a big thank you goes to my supervisors Dr. Panos Kostakos and Dr. Paula Alavesa.

I must give thanks to my parents, Heli and Martti, for their patience and support through all these years. And I would also like to express my gratitude to my grandfather Eino, who unfortunately passed away before seeing me graduate. He helped me more than I could have asked for throughout his life and I that is something I will never forget.

And when it comes to the making of this thesis, I have to say I thoroughly enjoyed almost all parts of the process. The part I was not too keen on was the part where I broke my tooth and ended up having it removed completely. So, I hope your time reading through this thesis is at least a bit more enjoyable than that.

Oulu, December, 19th 2020

Mikko Korkiakoski

LIST OF ABBREVIATIONS AND SYMBOLS

3D	3-dimensional
AGE	Abrams/Gentile Entertainment
ANOVA	Analysis of Variance
API	Application Programming Interface
AR	Augmented Reality
bg	between-group
BMI	Body Mass Index
CBRN	Chemical, biological, radiological and nuclear
CRT	Cathode Ray Tube
CV1	Consumer Version 1
EDA	Electrodermal Activity
DK1	Development Kit 1
DK2	Development Kit 2
ECG	Electrograph
FOV	Field of View
GEQ	The Game Experience Questionnaire
HMD	Head Mounted Display
HUD	Heads Up Display
LCD	Liquid Crystal Display
LED	Light Emitting Diode
MIT	Massachusetts Institute of Technology
NASA	The National Aeronautics and Aerospace Administration
NES	Nintendo Entertainment System
Oculus Rift S	Virtual reality hardware
OLED	Organic Light Emitting Diode
PM system	Panorama Manifestation system
PPG	Photoplethysmogram
SKT	Skin Temperature
Unity	Engine used in creating scenarios and environments for VR
VIEW	The Virtual Interface Environment Workstation
VR	Virtual Reality
VRE	Virtual Reality Environment
wg	within-group

df	degrees of freedom
F	variation between sample means / variation within the samples
H_0	null hypothesis
M	mean
N	number of samples
p	significance
SD	standard deviation
Z	z-score / standardised score

1 INTRODUCTION

1.1 Motivation

When preparing for a disaster or crisis, it is vital to have in place a number of pre-planned measures for evacuating people out of buildings. Therefore, training the first-responders to act properly in different types of rescue and evacuation situations is of utmost importance. At the same time, it is important to investigate how to best guide the first-responders out of these structures in the most effective way. Training people in real-life drills requires a lot of planning, money and it is also very time consuming [1]. The same can be said about comparing different types of guiding systems. With the emergence of virtual reality (VR) applications and tools, the aforementioned drills and guiding system tests become an easier task to manage. Although the virtual simulations can look and sound quite realistic, people have different reactions when playing them. Some people feel they are very realistic while others see them as just games and play them as such. It must also be stated that doing virtual drills is in most cases the safer option for the participants, especially if the simulation deals with dangerous gasses or chemicals.

There have already been studies regarding wayfinding, evacuation and different guiding light systems. The function of a guiding light system is to help lead people somewhere, usually to the emergency exit of the building. Cosma et al. [2] studied guiding light strips installed onto the floor while Vilar et al. [3] compared horizontal and vertical lighting systems. Meng and Zhang [4] investigated how wayfinding in a relaxed setting and wayfinding during an emergency affects people's stress levels by measuring peoples' heart rate and skin conductivity during their experiments. The data revealed that people in the emergency situation exhibited higher levels of stress. Cho et al. [5] studied whether VR can provide accurate results regarding stress by recording three different physiological signals. They concluded that VR can indeed be effectively used to induce stress and that the methods used to record and analyze the stress levels were very accurate. The participants in this study also showed elevated levels of stress during the stress-inducing parts of the experiment. Drawing from these and other previous studies the aim for this thesis is to combine all these different areas into one large VR evacuation simulation and use it to assess the effectiveness of guiding light systems even further and whether their presence can lower the stress levels during an emergency.

1.2 Research questions

This thesis investigates two main research questions: what kind of effect does the presence of a guiding light system have in the effectiveness of evacuation and does a lighting system have a positive effect on the stress level of the person. In addition, different user groups such as male/female, gamer/non-gamer etc. are compared to see if there are significant differences in their performance metrics.

For the purposes of this thesis a VR simulation scenario was created as a research prototype. This scenario is situated in an underground parking and features two different types of guiding light configurations. The VR scenario requires the participants to maneuver through fire and water with toxic gas making visibility bad and the air unbreathable. As there has already been some research into different guiding light systems using VR [2, 3] and real-world drills [6], two best-performing configurations (wall and floor) were chosen. The color of the light was also chosen according to previous research [7] (green). In addition, the heart rate of the participants was measured during the simulation to investigate the question of whether the presence of guiding lights had any positive (or negative) physiological effect. Previous research into VR gaming and simulations has demonstrated that when people are experiencing VR content, their blood flow increases [8]. It has also been shown that photoplethysmogram (PPG) measurements are an accurate way of estimating the user's stress state [5]. So, if simply experiencing VR already increases blood flow, could something that is assisting in the task help decrease the stress state and blood flow. If the lighting systems could be shown to have a heart rate- lowering effect as well as an effect on the speed and precision of the evacuation, that would be a valuable find.

1.3 Objectives

The main objective was to investigate which lighting system (if either) is the most helpful in guiding people to the emergency exit and whether they are significantly better than having no lighting system at all. The working hypothesis regarding this objective was that at least one if not both of the lighting systems would perform significantly better than the control group with no lights. This hypothesis is based on earlier research into vertical and horizontal signage systems [2, 3], exit portal indicators [7] and effects of smoke on people's walking speed [6]. The secondary objective was to investigate if the guiding light systems affect people's stress levels during the experiment by analyzing the recorded heart rate data and the trends they might exhibit. Having an increased stress level can affect people's actions in various ways. Previous studies have already shown that people's stress levels can increase during VR simulations, just like they would in real life, if there are proper stress-inducing stimuli present [4, 5]. Again, the working hypothesis was that the two groups with the assistive lights would show significantly better results than the control group [3, 4, 9]. In other words, the groups with the assistive lights would exhibit downward trending heart rates while the control group would show upward trends.

1.4 Methods

The data gathered from the simulation consisted of user performance metrics such as completion time, walking speed, number of stops etc. and the collected heart rate data. The evaluation of the simulation itself was conducted with a user study. To record how the participants felt and experienced the simulation, The Game Experience Questionnaire [10] was used. The participants answered the questionnaire before and after the simulation. The questionnaire mostly consisted of questions about how the user experienced the simulation and how they felt after finishing it, but also some general questions about memory, pathfinding, user information etc. The data was analyzed using a mixed approach, depending on the quality of the collected material. The overarching theme for the collected metrics were drawn from previous studies and literature into the subject.

The lighting configurations were compared against each other and against a control situation with no lighting in an attempt to find out if one lighting system can be shown to perform significantly better than the other. The comparisons were done using the user performance metrics that were gathered during the simulation. The recorded heart rates were compared using the trends they demonstrated (upward, downward or none). It was done this way, because normal heart rate ranges vary from person to person. So, comparing how the heart rates change is a much better way of detecting whether playing the simulation had any effect on the participant. Different groups (male/female, gamer/non-gamer etc.) are also analyzed to see if there are relevant differences not just between the lighting types but also inside the configuration group. The answers from the gaming questionnaire were used to compare how the participants felt about the different scenarios and whether their experiences show any significant differences between the different scenarios.

The statistical method used in the analysis of the user performance metrics was ANOVA (analysis of variance), the questionnaire data was analyzed with Kruskal-Wallis H- test and Mann-Whitney U- test, and the heart rate trends were analyzed using the Mann-Kendall test, and Sen's slope method. After analyzing all the results, conclusions about the lighting systems and group differences are presented along with the effects that the different lighting configurations had on the participants' heart rates and whether the original hypotheses were correct. Lastly, this thesis gives some suggestions on how the experiments conducted here could be improved and what kind of topics further research into this field could investigate.

The thesis is organized as follows: Section 2 discusses the history on VR, VR equipment and simulations, and wayfinding simulations with and without the use of VR. Section 3 discusses the implementation of the simulation (definition, creation and data gathering methods). Section 4 provides information about the testing protocol, test setup, participants, the pre-and post-study questionnaire and discusses the different statistical methods used to analyse the results. Section 5 provides the test result data Section 6 discusses the results and makes conclusions about their significance. Section 7 provides concluding remarks about the thesis as a whole. Section 8 provides all the references used. Section 9 shows the two questionnaires (user- and general information and The Game Experience Questionnaire) used for this thesis.

2 RELATED WORK

2.1 History of VR

2.1.1 Early concepts and devices

Because of the fairly recent rise in the popularity, one might think that virtual reality is a relatively new thing. However, the concept of VR is older than one might expect. It could be argued that the history of VR had its start in the year 1838. It was Sir Charles Wheatstone [11] who introduced a concept called stereopsis. Stereopsis refers to the phenomenon where two eyes viewing the same picture or object from different points, creates a sense of depth and 3-dimensional structure. Wheatstone demonstrated this when he created the first version of stereoscope (Figure 1). It used a pair of mirrors at 45° angles to the user's eyes, each reflecting a picture located off to the side [11, 12].

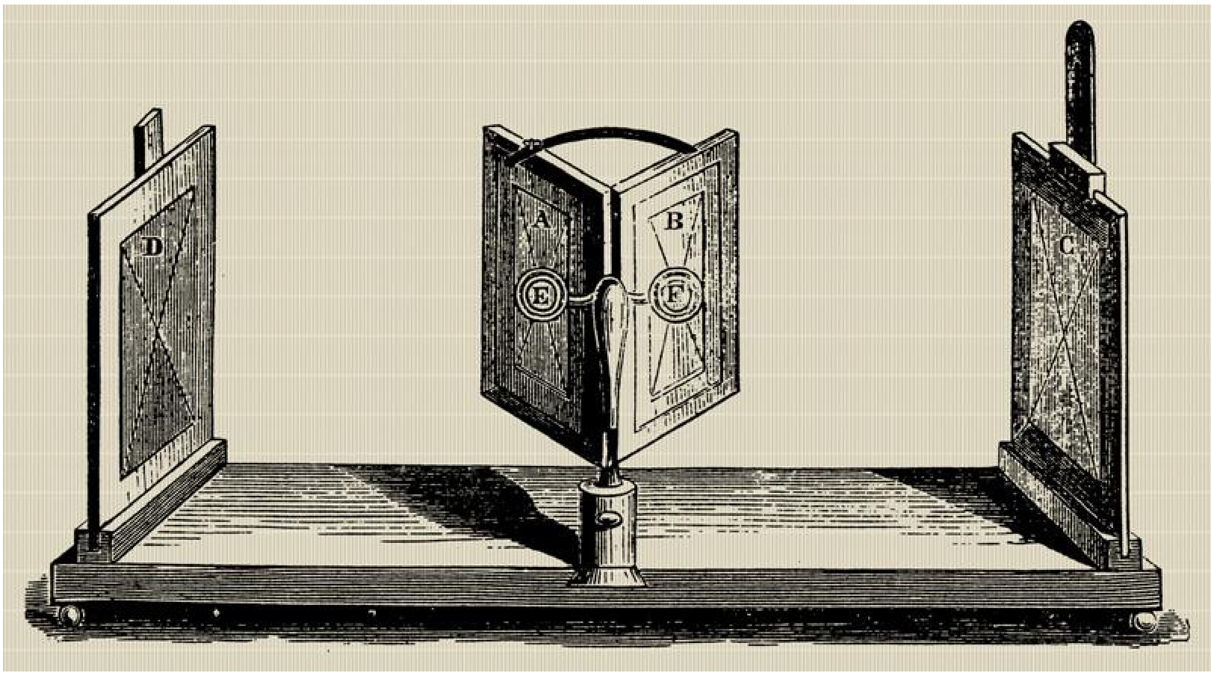


Figure 1. Sir Charles Wheatstone's stereoscope © Public domain

In 1927 Edward Link [13] started developing a flight simulator. It took him 18 months to build his first prototype and in 1929 he created the "Pilot Maker", an evolution of his simulator prototype (Figure 2). Link's flight simulator could be considered the next big step towards the VR as we know it today. The simulator was more commonly known as "Link Trainer" and it was the first commercial flight simulator that was completely electromechanical. The simulator used a motor-driven device that was able to mimic turbulence and other types of disturbances in the air. It had motors that linked to the rudder and steering column that allowed the user to control the pitch and roll of the apparatus, thus simulating a real aircraft in the air [13, 14, 15].

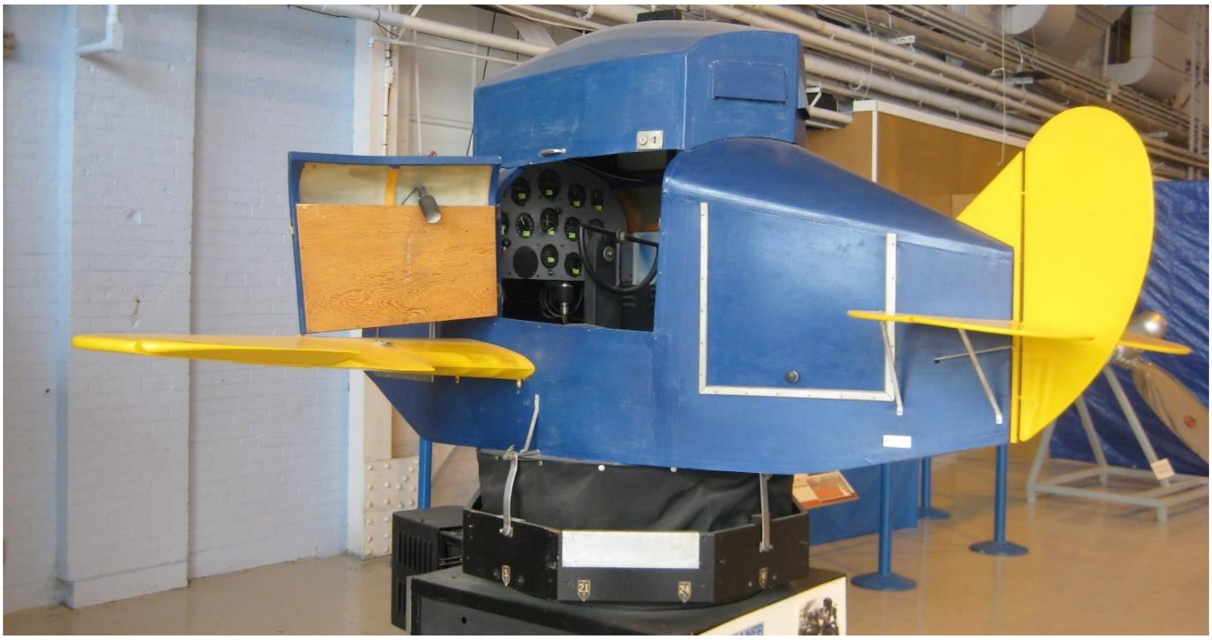


Figure 2. Edwin Link's "Link Trainer" flight simulator on display at the Western Canada Aviation Museum © Public domain

In 1956 the first virtual reality machine was created. This was done by cinematographer Morton Heilig [16], who named his invention "Sensorama". Sensorama (Figure 3) combined various different technologies in order to stimulate all five senses. It had full color 3-dimensional (3D) video, stereo audio speakers and a vibrating chair for vibration effects. In addition, the machine would also release odours and it had atmospheric effects such as wind simulation. The reason for creating the Sensorama was that Heilig wanted to immerse the viewers completely into his films. In the end only six short films were developed for the Sensorama. The machine itself was very advanced for its time and effects such as odour and wind simulation would be considered quite advanced even today [12, 16, 17].

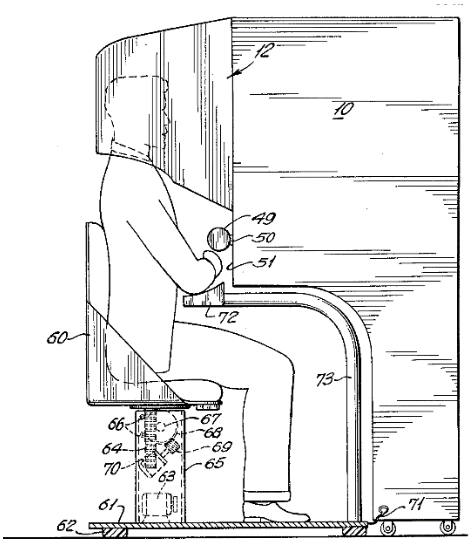


Figure 3. Morton Heilig's patent image for Sensorama simulator (image taken from US3050870) © Public domain

2.1.2 First wearable devices and continued development

Heilig was also the person to invent the first HMD (head mounted display). He patented his invention in 1960 and called it the Telesphere Mask (Figure 4). The device had no head tracking ability but it did have stereoscopic 3D, wide vision and stereo audio as well. Missing the head tracking ability, the Telesphere Mask could only be used for viewing media without any type of interaction [12, 15]. In 1961, only a year later, two Philco Corporation engineers, Charles Comeau and James Bryan [18] constructed the first HMD that had motion tracking. They called their invention “Headsight”. For displaying content, Headsight used two CRT (cathode ray tube) elements; one for each eye. For determining the position and direction of the user’s head, the Headsight used magnetic tracking. The CRT displays were attached to a remote camera via cables and the motion tracking would move the remote camera as the user moved their head. While neither device had any sort of computer integration nor image generation, the Telesphere Mask, the Headsight were still the first significant step towards the HMD devices we have today [12, 18].

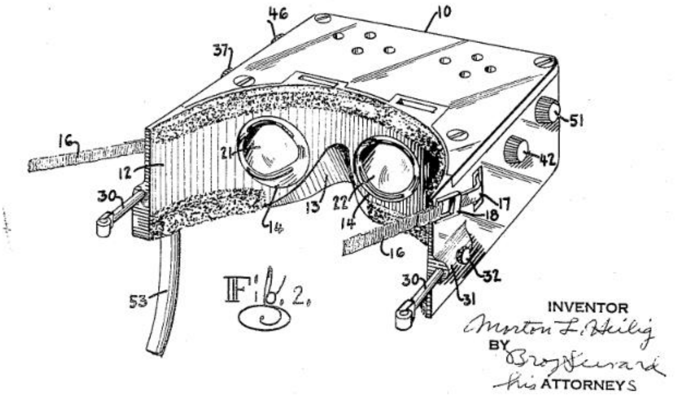


Figure 4. Figure 4. Morton Heilig’s Telesphere Mask illustration from his patent (image taken from US2955156) © Public domain

A few years later in 1965, Ivan Sutherland [18] presented a new concept to VR. He named his idea The Ultimate Display. In his concept, The Ultimate Display would be an HMD device, that would present to the user a virtual world, that would be completely indistinguishable from the real world. This virtual world would also have the user interact with different objects just like in real life. Sutherland being a computer scientist himself, he understood that something like this would require computer hardware to keep the virtual world functioning in real-time. The Ultimate Display concept was ground-breaking and very much ahead of its time and it is considered to be the blueprint for virtual reality [12, 15, 18].

Sutherland characterized his concept as “The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal. With appropriate programming such a display could literally be the Wonderland into which Alice walked” [19]. We are beginning to see his vision of “Alice in Wonderland” come true with today’s VR and its applications.

The next step in Sutherland’s vision was the creation of “Sword of Damocles”. Sutherland, with his student Bob Sproull, created and presented the first HMD device that was not connected to a camera but instead to a computer system. For today’s standards the device was still very primitive. It was very heavy and quite impractical to use because it had to be attached to the ceiling of the room. The name actually came from the fact that the rather large apparatus was hanging from the ceiling. The HMD itself displayed a 3D stereoscopic image output from a computer system. The presented virtual world was meant to be an immersive experience but in reality, the device could only display wireframe 3D models. The headtracking ability of the HMD allowed the change of perspective in the virtual world. The 3D wireframe models augmented the view to the real world so the user could change positions and see the models from different angles. So, the Sword of Damocles actually turned out to be the first AR (augmented reality) HMD. Although the device was never taken out of laboratory conditions and into the hands of the consumer, it was still a very crucial step towards commercial HMD devices [12, 15, 18].

The next important figure in the history of VR is Myron Krueger [18]. Krueger was a computer artist who coined the phrase “artificial reality”. By this he meant computer-generated environments that would respond and react to the people in them. Krueger had several projects that eventually led to the end product he titled “VIDEOPLACE”. This was in 1975, and it was the first interactive VR platform where people were able to communicate with each other in a responsive artificial, computer-generated environment despite the people being in different places, kilometres apart. While VIDEOPLACE used computer graphics it did not make use of any HMD device but instead it used projectors, video cameras, video displays and position-tracking technology. The room VIDEOPLACE was set up, was dark and the user was surrounded by big video screens on all sides. The user was being recorded with a camera and then a silhouette of them was computer-generated to mimic their movements and actions. The users from other VIDEOPLACE rooms could see the silhouettes of the others from their own video screens and they could interact with them within the same virtual world. This sparked the idea of communication inside virtual worlds between people in different locations [12, 15, 18].

Similar to VIDEOPLACE was the Aspen Movie Map. It was created in 1977 at MIT (Massachusetts Institute of Technology). The setup was similar to VIDEOPLACE in that there was no HMD device, only video screens. In the Movie Map a person could explore a virtual Aspen City in Colorado. Much like Google Street View today, the Movie Map used pictures taken from a vehicle moving through the streets of Aspen. The user could also choose between winter, summer or just polygon version of the map. Making a virtual map like this hinted that VR could in the future be used to explore faraway places without leaving your home [12, 15].

2.1.3 First VR gloves and HMD improvements

1977 also saw the creation of the first wired gloves to be used with VR. They were named “Sayre Gloves”. It was Thomas DeFanti and Daniel J. Sandin [20] who created the gloves but as the name might suggest it was not their own idea, but rather an idea from their colleague Richard Sayre. The way the gloves worked was that they used light based sensors with flexible tubes. On the other end of the tube there was a light source and on the other end a photocell. This was used to measure how much the user had bent their finger. The glove proved to be lightweight and quite inexpensive to produce. The finger tracking is often considered one of the starting points of gesture recognition research in computer science [12, 20].

In 1979 McDonnell-Douglas invented VITAL. It was an HMD that incorporated VR into the pilot’s helmet. Although quite primitive still, the HMD had headtracking and could follow the movement of the pilot’s eyes and match that with the computer-generated VR imagery. VITAL was probably the first HMD device ever to be in use outside of a lab [12, 15].

The wired glove concept was later improved by Thomas G. Zimmerman [21], who in 1982, patented an optical flex sensor mounted onto the glove to better measure finger bending. Zimmerman later worked with Jaron Lanier to incorporate ultrasonic and magnetic hand position tracking onto the glove. These improvements eventually led to the introduction of the first commercially available device called “Power Glove” (Figure 5), released by Nintendo in 1989. Although the product itself was not developed by Nintendo, it was an officially licensed product. The Power Glove was designed and developed by Grant Goddard and Samuel Cooper Davis [22] for Abrams/Gentile Entertainment (AGE). It was designed to be used with the 8-bit console, Nintendo Entertainment System (NES). The Power Glove had fairly basic movement- and finger flexion tracking, a few buttons and an 8-directional D-pad. The Power Glove never got popular because of its poor functionality and very few games that actually used it. It does however have a cult fanbase and is considered “legendary” (legendarily bad) by many older gamers. The Power Glove was ultimately discontinued in 1990 and it would be some time until Nintendo brought another motion- controlled device to the market [21, 22].

Figure 5. Nintendo's Power Glove © Public domain

As previously mentioned, some of the most important inventions in the field of VR have come from the flight industry. Like the Link Trainer and Vital helmet before the next important step in HMD development would again come from this field. It was the “Super Cockpit”, invented and developed in 1986, by Thomas Furness [23]. It was developed for a project with the same name for the U.S. air force. The goal for the project was to portray spatial information in a way that took advantage of the human’s natural perceptual mechanisms. The Super Cockpit HMD used computer-generated 3D-imagery, forward-looking infrared, radar imagery and avianotics data. The helmet also included head-tracking ability and voice-actuated controls. With the help of all these and some additional sensors, the pilot would be able to control the aircraft with eye movements, gestures and their voice. This 3D virtual space allowed for easier control of the aircraft as the pilot could fully concentrate on piloting and not constantly looking at dials and other information displayed in the real cockpit. The British Aerospace also used this technology from 1987 onwards [12, 23, 24].

The same year (1987) Jaron Lanier, who had been previously working with Thomas G. Zimmerman on the different wired glove concepts really popularised the name Virtual Reality. Lanier had teamed up with Zimmerman and together they founded VPL Research (Visual Programming Lab). Their company was the first to sell HMD devices and “data” gloves. The devices they sold included VR goggles interestingly named “EYEPhone 1” and “EYEPhone HRX”. The EYEPhone 1 at the time cost $9400 while the EYEPhone HRX was a whopping $49,000. Lanier did not actually invent the term Virtual Reality but through his company and products he was eventually the one to make it so popular that it became the name everybody started using for that type of products and research [12, 15].

2.1.4 NASA’s VR and CAVE

In 1989 NASA (The National Aeronautics and Aerospace Administration) started working on their own VR project titled VIEW (The Virtual Interface Environment Workstation). The system would comprise of an HMD device, DataGlove and DataSuit (Figure 6). This equipment was developed in partnership with the previously mentioned VPL Research. The HMD was a version similar to the EYEPhone. The HMD could display either computer-generated virtual imagery or video coming from remote cameras. The DataGlove was a version of the wired glove concept that could track the user’s finger movement with the use of several fibre optical cables. Coincidentally the DataGlove lead to the development of the previously mentioned Power Glove by Nintendo. The DataSuit was a full-body suit that could, with the use of multiple sensors, detect the user’s motions, bends, gestures and orientation. All these devices working together created possibly the most advanced VR experience to that date [15, 25]. But only two years later in 1991, Antonio Medina [15, 26] introduced a VR system to be used in controlling the robotic vehicles used in Mars exploration. The system was titled “Computer Simulated Teleoperation”. Due to the lag caused by the signal travel time from Earth to Mars the operation of the robot would never be in real time even though it was said to be “operated in real-time”. With the continued exploration of Mars, NASA has continued the development of their rover operating devices.

Figure 6. A woman wearing a head-mounted display incorporating Pop Optics goggles developed by NASA's Ames Research Center, and wired gloves © Public domain

CAVE (CAVE Automatic Virtual Environment) (Figure 7) could be considered an evolution of VIDEOPLACE and it was the next invention in the VR field. It was introduced in 1992 by Carolina Cruz-Neira, Daniel Sandin and Thomas DeFanti [27] at the University of Illinois. The early concept was characterized as a space roughly 3m x 3m x 3m made up of 3 rear projection screens for walls and a down projection screen for the floor. Projectors would project image to the screens and a computer-controlled audio source would provide atmospheric audio for the scene. The user's hand- and head movement would be tracked with 3D-glasses and hand sensors. The person using the CAVE could see objects in 3D floating in the air and could walk around them, getting a look from different perspectives. Initially this effect was created using electromagnetic sensors and alter the same effect was done by the use of infrared cameras. [27, 28].

Figure 7. A person inside one type of CAVE room in use © Public domain

2.1.5 VR comes to gaming

In 1991 the first mass produced VR gaming machine was released. It was called Virtuality and it was developed by The Virtuality Group. In Virtuality the user could play in a 3D virtual world. There were two versions available for Virtuality, one where the player stands up (SU-version) (Figure 8) and one where the player is sitting down (SD-version) (Figure 9). They each used an HMD called "Visette" that had two LCD (Liquid Crystal Display) panels with a resolution of 276x372 pixels and head-tracking ability. The Visette also featured 4 built-in speakers and a microphone. The SU version featured a hand-held controller device that had motion tracking. The SD version could have either a joystick, steering wheel or an aircraft yoke depending on the game that was being played. The first series of the equipment was called 1000 series (featuring 1000CS and 1000SD machines) and it used an Amiga 3000 computer to run the games. A total of 9 games were released for the 1000 series. The company later released improved models 2000- and 3000 series [12, 29].

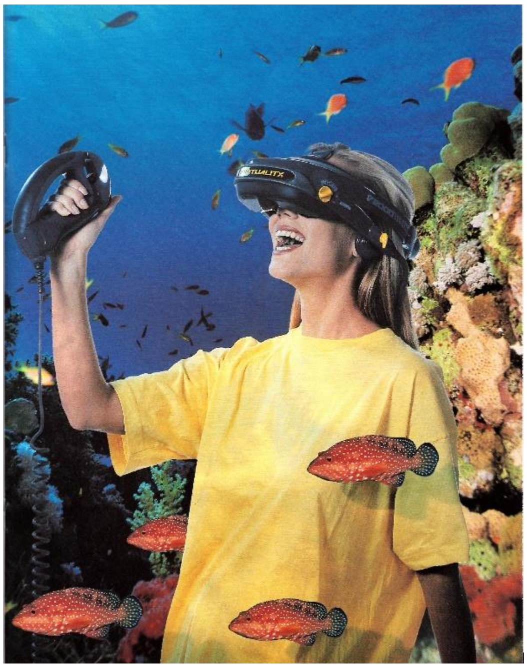


Figure 8. Virtuality SU unit © CC-BY 4 Dr. Jonathan D. Waldern/Virtuality Group

Figure 9. Virtuality SD unit © CC BY-SA 4 Yaraman [30]

In the mid-90s Sega and Nintendo tried to get into VR gaming with their new devices. Sega announced a Sega VR headset for the Sega Genesis home console in 1993. The headset was supposed to have head-tracking, LCD screens and stereo sound but development difficulties prevented the company from ever releasing the product. It was revealed that the system caused users to have severe headaches and motion sickness. The next year however, Sega did release The Sega VR-1, a motion simulator arcade. The simulator used an HMD with 3D polygon graphics and stereoscopic 3D. The simulator was only available at Sega World arcades amusement centres [15, 31].

Nintendo came out with its own VR console in 1995. It was called Virtual Boy (Figure 10 and Figure 11) and Nintendo called it the first ever portable home console that could display true 3D graphics. The Virtual Boy headset used two 1x224 linear arrays and rapidly scans the array across the eye's field of view using flat oscillating mirrors. It also had stereo speaker. The drawback of the device was that it could not show color, only monochrome. Eventually the device was a quickly deemed a failure. It was expensive, playing with it was not comfortable and the marketing campaign had also been a failure. Nintendo released the Virtual Boy in July 1995 and it was discontinued in March 1996 [15, 32].

Figure 10. Virtual Boy headset © Public domain

Figure 11. Virtual Boy controller © Public domain

However, there were some other affordable VR headsets that were released during the mid-90s. There were I-Glasses, the Victormaxx Cybermaxx and Fortevr VFX-1 [33]. But the technology just was not there at that point in time and after Sega and Nintendo failed to make an impact with their VR devices, the momentum died and it would be some time until VR would make its return into gaming.

2.1.6 Medical use and Google Street View

In 1997 two universities, Georgia Tech and Emory University, conducted studies into the use of VR to help Vietnam war veterans in dealing with PTSD (Post-traumatic Stress disorder). These experiments were called “Virtual Vietnam”. Scenarios were created using VR so the patients could experience some of their traumatic experiences again but this time in a safe environment. This is known as exposure therapy and is still in use today [12, 34, 35].

The next notable step in regarding VR would be Google’s Street View that made its debut in 2007. In Street View the streets were photographed using a dodecahedral camera that was attached on top a vehicle. The vehicle would drive around the city taking photos and creating the network of street images. The user could virtually move and look around the streets using the application. There was no HMD in use but one could still feel like they were exploring a place virtually that they had never been to. At first Street View would only feature a handful of cities but eventually more and more cities would become available. Three years later in 2010, Google released the 3D view in the application. Now users were able to see the buildings and their shapes in 3D [12, 15, 36].

Since 2010 Google has incrementally brought updates and improvements to Street View. In 2011 Indoor view of businesses was released and later improved in 2013. The users could look inside certain businesses and see what they looked like. In 2014 a “past view” for certain streets was released. Using this feature, one could see how a certain street looked like in the past. Since 2018 Japan’s streets can be viewed from a perspective of a dog. Google has taken the concept from the Aspen Movie Map and run with it. By May 2017 Google had taken more than 16 million kilometres of Street View footage in 83 countries. These include some panoramic views taken underwater [36].

2.1.7 Modern VR and return to gaming

Palmer Luckey [37], could be considered a VR pioneer of the 2010s. He thought the available devices were heavy, had too much lag, and their field of view was too narrow. So, in 2009, motivated by the bad quality of the current VR HMDs, he started developing his own HMD. He completed his prototype in 2010 and named it PR1. The PR1 had a 90-degree field of view, low latency and built-in haptic feedback. Luckey continued developing new devices and in 2012 he launched a Kickstarter crowdfunding campaign for his 6th generation prototype called "Oculus Rift". The campaign raised $2.4 million which was 974% of the original target amount. This allowed Luckey to hire new people and find a larger space for his offices [37, 38, 39].

In mid-2013 the first development kit for the Oculus Rift was released to the backers of the Kickstarter campaign. This first model was called Oculus Rift Development Kit 1 (DK1) (Figure 12 and Figure 13). It featured a 7-inch LCD screen with the resolution of 1280x800 at 24bit color depth. This means each eye had an effective resolution of half the width but same height; 640x800. The horizontal FOV (field of view) was 90+ degrees and horizontally 110 degree. This amounted to around double of what the other devices on the market had to offer. DK1 also had with lower latency and less motion blur than the earlier protypes [38, 39].

Figure 12. Oculus DK1 CC-BY 3 Sebastian Sabinger © www.pi23com

Figure 13. Oculus DK1 inside view CC-BY 3 Sebastian Sabinger © www.pi23.com

The next development model, Oculus Rift Development Kit 2 (DK2), was released a year later in mid-2014. It improved the HMD to full HD resolution of 1920x1080, also changing the aspect ratio to 16:9 from the 16:10 ratio in the DK1. The screen was also now an OLED (Organic Light Emitting Diode) instead of a regular LCD found in the DK1 model. The DK2 had a number of small improvements including higher refresh rate, detachable cable and the external control box (seen in Figure 14) was removed. The two initial releases were made available for developers so they could start making content for the upcoming commercial release [38, 39].

The final prototype device before the commercial release was title Crescent Bay. It improved the resolution, weighed less, added audio speakers and 360-degree tracking. The first commercially available Oculus product was released in 2016 r and it was called the Oculus Rift CV1 (Consumer Version 1) (Figure 14 and Figure 15). The CV1 had an increased resolution of 2160x1200 (1080x1200 per eye) with 90Hz screen refresh rate. As with previous models, it featured 360-degree position tracking and built-in audio. “Oculus Touch” handheld controllers (Figure 16) to be used with the CV1 were released in the end of 2016. The tracking for both the HMD and the controllers was done with a separated Constellation sensor (Figure 17). The CV1 can arguably be considered the device that started the current wave of VR devices [38, 40].

Figure 14. Oculus CV1 © Public Domain

Figure 15. Oculus CV1 inside view © Public domain

Figure 16. Oculus Touch Controllers © Public Domain

Figure 17. Constellation tracking sensor © Public Domain

Oculus released its first wireless HMD titled Oculus Go in 2017. The Oculus Go does not require any external hardware to run the applications but instead it runs the applications on its own integrated hardware. The Go again increased the display resolution, now to 2560x1440 but with lower refresh rate versus the CV1. It uses an Android OS (operating system), Qualcomm Snapdragon 821 chipset and internal flash storage to run and store the applications [41].

Oculus Quest was released in mid-2019 and it was basically an improved version of the Oculus Go. It had a bigger resolution, more storage space and a more powerful hardware in general. It was also fully capable of running the applications on its own hardware but unlike the Go, the Quest could also be attached to a computer or smartphone and let those run the applications with the Quest just being the display (like the CV1 for example) [42].

The Oculus Rift S is the successor of the CV1 and like the Quest, it is an all-round improvement to the original. The Rift S takes the CV1 as a base and implements the new improved aspects of Go and Quest. Unlike the CV1, Rift S does not require a stand-alone tracking sensor stand but the tracking is integrated onto the HMD, just like in Go and Quest. The Rift S can also be paired with the second-generation Touch controllers, like Quest [43].

With the VR devices improving every year, game developers have slowly but surely started to implement VR support into their titles. Racing simulators, flight simulators and first-person shooters are probably the genres where VR equipment can change the gaming experience most drastically. When previously you needed a multi-screen setup to get the most immersive experience in a simulator, now you can achieve an even more immersive experience with just an HMD device. Of course, there are games purposely built for an HMD and its peripherals. You can kill monsters by shooting arrows with a bow or hit incoming boxes with a lightsabre to the beat of the music for example. And the best yet is that we are only in the beginning of the VR revolution.

Currently there are several different HMD devices available on the market that offer similar types features. So, looking at all these devices individually would not be very fruitful as they are not so different to the Oculus devices already presented here. In this study the Oculus Rift S paired with two powerful, VR ready gaming laptops were used for running the simulations.

2.2 Real-world- and laboratory training for emergencies

There are many types of natural and man-made emergencies that require actions from different authorities and front-line practitioners like police, firemen, paramedics etc. Sometimes hazardous materials such as chemical, biological, radiological and nuclear (CBRN) agents are involved in these emergencies. Dealing with natural emergencies like hurricanes, earthquakes, tsunamis etc., can be easier than with man-made emergencies. Natural events can in most cases, be predicted and the impact estimated in advance. And with the right kind of pre-emptive actions the harm to people and infrastructure can usually be minimized if not almost entirely avoided. Emergencies caused by human error, terrorism etc., can be very difficult to prepare for because of their unpredictable nature [44].

Preparing to and dealing with ongoing emergencies and the aftermath can also be very expensive. Organizing training scenarios takes a lot of time and resources. And the bigger the scale of the scenario, the more expensive it becomes to organize and execute. This is especially true when it comes to organizing training simulations for CBRN-type of emergencies. In addition to cost and time, these agents are very hazardous to people and using real CBRN agents in training is also potentially very dangerous [44].

Preparing to almost any kind of an emergency usually involves some kind of an evacuation plan so simulations for evacuation research are important. The point of an evacuation is to get the people out of the building in a safe and quick manner. Sometimes these situations also involve fire and smoke (or toxic gasses) which makes it even more crucial that the evacuation process happens quickly. In many cases most deaths happen because of smoke/gas inhalation [45]. During an emergency evacuation, there are certain aspects of the human behaviour and responses to disruptions that can affect the process negatively, such as: [45]

- Unfamiliarity with the building and its exits [45]
- Noise of the alarm [45]
- Reduced visibility due to smoke and obstructed paths and/or exits [45]
- Effects of smoke/gas inhalation [45]
- Human behavioural responses such as panic, stress, decreased risk perception [45]

Training for emergency evacuation can be done in many different ways. Kinateder et al. [46] compared six different methods for this kind of training. in their paper. Three of these methods are set-up in the real-world while the other three are basically done in a laboratory-type of an environment. The methods based in real-world scenes are:

- Field study: Training scenarios re-enacted in a naturalistic setting [46]
- Case study: The trainees study a specific event and report their findings [46]
- Drills: Similar to field study but drills can be either announced or unannounced training events to train the personnel [46]

These real-world based methods offer either limited or no possibility for reproducing the training situation exactly as it was the previous time. The same goes for adjusting the setting of the simulation. These methods also have a high demand for time and cost and any kind of automated data collection is very difficult if not impossible to set up. The type of data that these kinds of methods produce is mostly human-behavioural data [46].

The three methods that are used in a laboratory environment are:

- Classical laboratory experiment: Real-world scenario that is transferred into a controlled laboratory environment [46]
- Hypothetical study: Participants do not partake in an actual training scenario but rather imagine a scenario or view videos of a real-world scenario and answer questions about how they would behave in these hypothetical/recorded scenarios [46]
- Virtual Reality experiment: Participants use VR equipment such as HMDs and controllers in a VR scenario [46]

These laboratory-type experiments have several benefits to the real-word-type experiments. The cost in doing experiments in a laboratory is very low compared to doing them in the field. The time taken to set up these types of experiments is again very low compared to having to construct and organize them in the real world. Data gathering is also made much easier because of the controlled setting. It can be automated with computers and practically any kind of data can be gathered especially when using VR equipment. The experiments done in a laboratory can also be completely replicated every time and the control over the experiment is much better than in real-world cases. In addition to behavioural data, laboratory experiments can also produce physiological data but in the “hypothetical study” case the data is only statements from the participants and the experts evaluating the scenario [46].

2.3 Using VR as the simulator for an emergency

With the help of virtual reality (VR) technology it is possible to create simulations relatively cheaply and with a lot more creative freedom and with nearly endless possibilities. Even if we cannot predict how a certain type of event will play out, it is possible to prepare for different types of scenarios much easier using VR. The hardware and software used for VR have advanced to a point where it is possible to create very realistic and complex scenarios to train different authorities in the handling of practically any type of emergency. Creating training scenarios for VR is also much more cost-effective than having the training done in real life. Although the initial costs for creating a VR simulation may be even higher than that of a live simulation, the costs after the simulation has been finished are almost non-existent. There might be some costs for equipment maintenance/upgrades or software updates for example. But for live exercises the costs always keep accumulating as you do more of them. So, in the long run, VR quickly becomes a much cheaper alternative [1].

The VR simulations can be done practically anywhere because the equipment required is relatively small and also easy to move unlike live exercises that usually require a lot of planning and infrastructure [1]. Simulations can also be run in large spaces where the participant(s) can physically move around the virtual environment. All you need is some computer(s) running the VR software and the VR hardware that usually consists of and HMD and either gloves or some other controlling devices. The quality of graphics nowadays is enough to create practically photorealistic surroundings. Using surround audio systems can immerse the user even deeper into the scene. There is also a possibility to go for an even deeper realism using some kind of odour, changing the temperature in the test room or even using some kind of wind machine to create a breeze. The possibilities really are endless.

But while VR simulations can be highly realistic there are still some things that cannot quite be simulated in VR like can be done in real life training simulations. Most commonly these are things related to human behaviour such as:

- Panic: An event that occurs in the human body both emotionally and physiologically decreasing the capacity individuals have in organizing their thoughts and in elaborating a more complex rational response [45]
- Stress: A generally uncomfortable emotional experience and it is often perceived by the biochemical physiological and behavioural changes in human beings [45]
- Risk perception: The first actions when facing a stimulus, such as fire, is to understand, recognize and think. Before the individual takes action, it is necessary for the individual to go through these three factors. This is known as risk perception [45]

In short, when training people for emergencies involving CBRN substances, the most realistic and therefore the best (if cost and time are not an issue) way is still to practise using actual CBRN substances. This is the only way to really train people for the physical, physiological and psychological impacts of dealing with the actual CBRN emergencies. Live agent training brings better confidence in tactics, in the equipment and can also reduce stress enormously in a real situation [44].

2.4 Measuring and evaluating VR wayfinding simulations

Ruddle and Lessels [47] suggested that there are three levels of metrics. These three metrics are task performance, physical behaviour, and cognitive rationale. These metrics “allow key features of the data to be presented in a concise form that allows the meaning of the data to be readily comprehended” [47].

The task performance metric is fairly self-explanatory. Any single task performance metric reports the measurements for a single task that the user has performed in the simulation. These tasks include time taken to accomplish something, some distance travelled, the number of completed tasks or the number of errors made during the simulation etc. [47]. When deciding which task performance metrics to use, it is first important to recognize, what kind of measurements are actually useful for drawing any kinds of conclusions from the simulation.

Physical behaviour metrics differ from the task performance metrics in that as the name suggests, they measure the behaviour of the user’s body (the virtual one) in the simulation rather than the tasks the body performs. These metrics are not as widely used as the task performance metrics though there is less consensus about what type of data is the most valuable. The physical behaviour metrics include things like time spent moving or not moving, the orientation of the virtual body and the path of movement inside the simulation etc. [47]. These metrics can be used to determine if the user visits the same places more than once or if they are possibly confused and moving in circles. They can also be used to measure if the user is stopping and looking around trying to find some references in the scenery or something to guide to the destination [47].

The last of the three metrics, cognitive rationale metrics, are an attempt to understand how the user's decision-making process works. The data used in this metric can be something the user is thinking aloud during the simulation, conducting interviews and questionnaires before and after the simulation etc. The problem with interviews and questionnaires done after the experiment is that the user has to have an accurate memory to remember what they did and why during the simulation. The thinking aloud method does not have the same problem, but it could in some cases impede the user's ability in the actual wayfinding, if they have to also be talking at the same time. In short, this metric relies more on the user's ability to evaluate their own performance rather than raw recorded data about their performance [47].

2.5 VR emergency- and wayfinding simulations

VR can be used to simulate practically every kind of an environment, but when dealing with the question of wayfinding it is usually some kind of an indoor space where the experiment takes place. These places can be for example road- or rail tunnels [2, 7], office buildings [3, 6], railroad stations, underground parking garages, metro tunnels [2] etc. Simulating emergency situations that happen indoors can make for an interesting but also a difficult subject. Creating the space to be used in the simulation is nowadays easy. The difficult part is to identify what in the participant's actions warrants investigation. Another question regarding wayfinding is, are there any methods or indicators that will help the participant find their way to the goal easier than without them.

Vilar et al. [3] studied how using vertical and horizontal signage helps in wayfinding. They hypothesized that people using either signage system would perform better than the ones trying to find their way without any help, and that the horizontal system would fare better than the vertical one. In addition, they hypothesized that there would be no differences between the genders. The horizontal signage system consisted of continuous coloured lines on the floor leading to the destinations while vertical system had signs on the wall that contained text and arrows pointing the directions. The neutral condition only had identifying names in the destinations but no signage systems. This simulation was conducted in an office-type environment in normal every-day conditions, so there was no sense of urgency unlike in emergency wayfinding simulations. The measurements included distance travelled, time spent in finding the destinations, number of times the participant stopped moving (for at least 2s) and average speed of the participant. All the data was analysed using ANOVA (Analysis of variance) with Scheffé post hoc test and two-way ANOVA when verifying the influence of gender. The conclusions from the study were that indeed the signage systems fared better than "no signage", but there was no significant difference between horizontal and vertical systems. Nor were there any significant differences in the results between the genders.

One very interesting method to investigate is the implementation of some kind of guiding light system to help the person reach the exit of the structure. Ronchi et al. [7] used the CAVE environment to investigate different types of guiding light configurations installed at the exit portal to be used in an emergency evacuation of a road tunnel. They investigated how the color of the light and its flashing rate affects in helping the participant in finding the exit. They also experimented with different types of light sources and different layouts for the lights at the exit portal. The colours in testing were green, blue and white and the light source was either a LED (Light Emitting Diode) or a strobe (or double strobe) light. The patterns in which the lights were installed were 3 lights with one on top of the exit and one on each side, one light on top of the exit and 2 bar-shaped lights on both sides of the exit. (Figure 18). illustrates the light configurations.

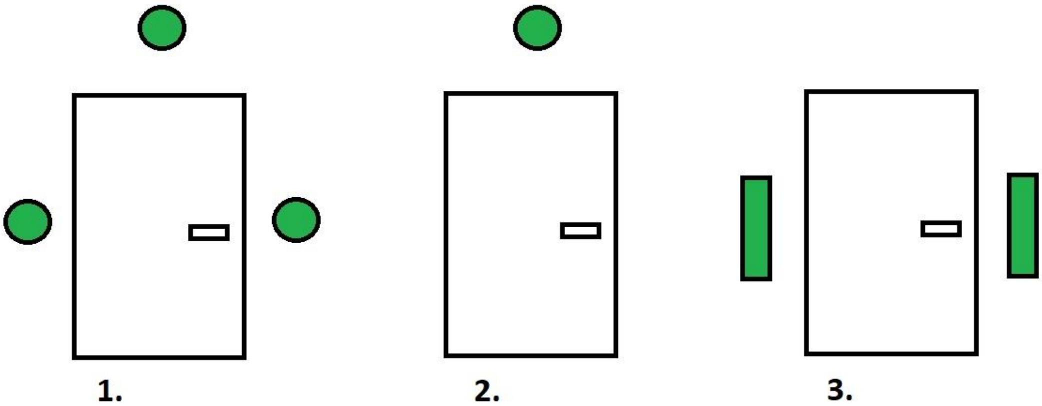


Figure 18. The light configurations used. Illustration based on [7] © Mikko Korkiakoski

The main takeaway from their study was that the flashing lights at the exit portal do have a positive impact in finding the emergency exit. Their control group that had no lights at all, received the lowest rank in their light installation ranking, suggesting that any kind of lights are better than no lights at all. Other interesting results were that the colours green and white performed better than blue and that the LED lights were preferred over strobe and double strobe configurations. They also suggested that the flashing rates of the light that work best are 1Hz and 4Hz.

In their paper, Cosma et al. [2] studied the use of green LED light strips installed onto the floor of the rail tunnel. This study is kind of a continuation of the previously mentioned study about the guiding lights installed at the exit portal. This is because the group consisted of three people out of which two were a part of that study into the exit portal lights and because this study into LED light strips relies on some of the conclusions made from the earlier one. Both of these studies also base their simulations in a similar indoor space where there is heavy and dense smoke, thus making visibility an issue, while also creating a situation where finding a way out is crucial for the survival of the person. The researchers hypothesized that due to the smoke, people could easily miss signs leading to the exit and possibly even the exit itself and thus. Having to deal with fire and smoke in an emergency will also create certain physical and physiological effects in people. These effects may then negatively affect the individual's decision-making, wayfinding, and just overall mental state. The test consisted of three different scenarios in which the participant had to navigate the railway tunnel to find the exit. In the control group scenario only had the standard railway tunnel lights but there were no guiding lights. The two scenarios with LED installations had either a continuous green LED line or 10m long green LED stripes guiding the way to the exit. The researchers measure the total time needed to find the exit, distance to exit at given time and the users' distance from the tunnel walls. The testing was conducted using the Oculus Rift HMD. Their results showed that both LED installations performed considerably better than no guiding lights and that almost all the participants reported noticing the green LED lights. Most of them also said that the color for the guiding LEDs was suitable for the evacuation purpose. Even though the two LED installations performed better than the control scenario, there was no real difference between the two of them in terms of performance.

Ruddle and Lessels [47] conducted an experiment in which they created a 10m x 10m room in VR that had 33 cylinders in it (Figure 19). The goal for the user was to find the 8 targets located on top the cylinders. There were also 8 decoys planted on top another 8 cylinders, while the rest 17 cylinders had nothing on them. They used 3 different levels of fidelity in the VR room to see if that made any difference in the results. The lowest quality had only basic shapes and no textures. The intermediate quality had some repeating brick wall textures and the highest quality had digital photograph textures in the walls of the room.

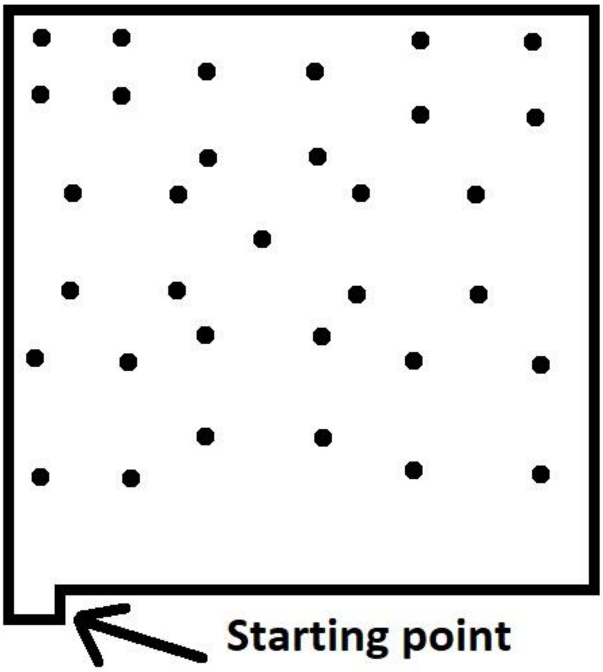


Figure 19. A rough layout of the VR room. Illustration based on [47] © Mikko Korkiakoski

The participants also completed the same experiment in a real-world room, similar to the one created in VR. In this real-world trial, 93% of the participants completed it by searching every target/decoy only once, which in other words is a perfect search. In the VR experiment however, some the participants completed the search the same way (perfectly), but for others the task proved difficult and they visited many places inside the room more than once. In fact, only 47% of the participants completed a perfect search. Some results were near-perfect, meaning the user only double checked a target/decoy only once. These results were part of a perfect search metric that fall under the task performance metric. The other evaluation metric used was the physical behaviour metric. This comprised of locomotion: macro-level search heuristic and movement breakdown, error classification and "looking around". The VR room was split into quadrants and then used in the macro-level search heuristic to list the sequence in which the user entered the quadrants (1-3-4-2-1-3 etc.). Movement breakdown consisted of the time the participant spent continuously moving from one target/decoy to the next and the percentage of time they spent stationary between targets and decoys. Error classification was divided in miss/local neglect/global neglect categories. A miss meant the user had previously touched the target but did not "see" it. A local neglect was recorded if the user had passed through a designated triangle in which there was a target but had missed it. And a global neglect was recorded for any other type of an error during the experiment. Looking around was simply the time the user spent changing their "view heading" separately while standing still and while moving. The results from this study showed that the participants used the same general heuristic search patterns in the real-world experiment as well as the VR room. However, the performance metrics were not similar at all. The difficulties and lower performance in VR was attributed mostly for the difficult controls (only moving in one direction at a time) in the VR system. The effect of the field of view (FOV) was investigated as well. The users often had a target inside their FOV, but they still did not react to it. They speculated that this could have been because the user already thought they had checked the target or that their focus was on some other part of the scene and they simply just missed it. So, it was not a problem caused by a low FOV because even in wider FOV the participants missed objects that were visible in the sider of their view. This issue could have been investigated better had the users been told to use the "thinking aloud" method, in which they are encouraged to talk aloud what they are thinking and doing. The authors suggest further study into the use of different kinds of FOVs in VR. The metrics used in this paper for wayfinding performance are very usable in future wayfinding research as well [47].

2.6 Non-VR emergency- and wayfinding simulations

Creating simulations for different emergencies and wayfinding research in real environments takes a lot of time and resources. Simulations in "real life" are of course the closest thing we can get to an actual emergency event, so we do not always want to just study things in VR.

One study investigated the usage of guiding lights installed on the walls and the floor of a smoke-filled corridor and compared their effectiveness to an overhead light system and just plain normal lighting. The main point of the study was to see if the different guiding lights would affect the walking speed of the person. The tested wall light configurations were a green electroluminescent continuous track installed 1m above the floor and a green LED lighting system with the lights installed at varying heights on the wall. The floor guiding light system was similar to the LED system used for wall lights, but it consisted of continuous and spaced lights. In this study, Wright et al. [6] concluded that the lights installed in the ceiling were very ineffective compared to all other tested light configurations. Even when increasing the illuminance of the overhead lights, the dimmer floor and wall guidance lighting systems still provided better results. They also concluded that there were no significant differences in walking speeds between the two different LED configurations and that the luminance of the lights does not have a significant impact on their effectiveness. More important seems to be where the lights are installed so they can provide visual cues that people easily understand.

2.7 VR experiments with heart rate measurements

There have a been a few studies that have incorporated heart rate measurements into an experiment using virtual reality environments (VRE). Meng and Chang [4] used a panorama manifestation system (PM system) as their VRE. The VRE consisted of six 47-inch liquid crystal displays (LCD) placed in a “circle” around the user creating a 360° view of the environment. They also used a video-based eye-tracker to track the user’s eye focus. The user’s heart rate and skin conductivity during the experiment was measured using a ProComp Infiniti device. The experiment consisted of a reading task and a wayfinding task. For the treatment group the wayfinding task included a simulated fire evacuation of the building. In addition to virtual hazards such as fire, explosions and a siren, real smoke was emitted by a smoke generator around the user’s environment to provide old factory and visual stimuli. The control group experienced none of the virtual or real-world stimuli but instead completed the wayfinding task in a more relaxed environment. Both groups showed increased heart rate and skin conductivity in the wayfinding task. The heart rate for the treatment group was only slightly higher than that of the control group, but the skin conductivity showed significant increase. Overall, the treatment group showed higher levels of stress during the experiment.

Cho et al. [5] performed an experiment with different levels of stress-inducing content. They recorded the users’ heart rates using a photoplethysmogram (PPG) sensor, electrodermal activity (EDA) and skin temperature (SKT). The actual experiment consisted of videos shown to the participants via Samsung Gear VR HMD. The videos were tested beforehand on people who did not partake in the actual experiment and were chosen on the basis of the stress levels they produced in the test subjects. The videos were classified as inducing mild, moderate and severe stress reactions. These videos were then shown in sequence during the experiment, with each video followed by a resting period to calm the subject before the next video. So, the experiment sequence was BASELINE, rest, MILD STRESS, rest, MODERATE STRESS, rest, SEVERE STRESS and final rest. Using the data from the previously mentioned three different physiological measurements, the authors found that indeed the physiological signals provided enough information to accurately (95% accuracy) classify a person’s stress level. The results from these two studies show that VR can be used in simulating real-world situations to produce stress levels that are accurately measurable, classifiable and that have correlation to their real-world counterparts.

Table 1. Study information breakdown

Study	Type of system	Environment or location	Hazard(s)	Recorded metrics	Equipment/ Sensor	Statistical test(s)
[2]	VR	rail tunnel	none	time spent, movement patterns, coordinates, questionnaire, open discussion	HMD	analysis of means and standard deviations comparison, paired two-sample t-test
[3]	VR	office building	none	distance travelled, time spent, number of pauses, average speed	HMD, motion tracker	ANOVA, Scheffé post hoc test
[4]	PM system, with smoke generator	hotel	fire, explosions, siren, real smoke	heart rate, skin conductivity, travel distance, various time data	PM system, heart rate and skin conductivity, monitor, eye tracking	paired t-test
[5]	VR videos	videos	audio and video stimuli	heart rate, skin temperature, skin conductance	HMD, heart rate monitor, skin temperature, electrodermal activity sensor	ANOVA with Tukey's HSD
[6]	real-world experiment	various parts of a building	real smoke	time spent	none	ANOVA
[7]	VR/CAVE	road tunnel	none	questionnaire	projection and lighting systems	Wilcoxon signed rank test
[8]	VR/CAVE-like	maze	none	blood flow velocity (BFV), breaks in presence (BIP)	ultrasound probes	two-way ANOVA
[47]	VR	single room	none	task performance, locomotion: macro-level heuristic and movement breakdown, error classification, looking around	computer	ANOVA
This thesis	VR	underground parking structure	Toxic gas, fire, water leak, standing water	completion time, average speed, distance travelled, number of pauses, timestamps for pauses, number of turns, degrees of rotation, heart rate	HMD, heart rate monitor (PPG)	ANOVA, Kruskal-Wallis H test, Mann-Whitney U test, Mann-Kendall test, Sen's slope method

3 IMPLEMENTATION

3.1 The scenario

3.1.1 Backstory

There had been a terrorist incident prior to the starting time of the simulated scenario. The terrorists had acquired a dirty bomb and they also held a number of hostages. There had been a police chase before the terrorists ended up in the underground parking garage. The police had surrounded the terrorists’ van inside the parking garage and at this point the terrorists had detonated their dirty bomb. The explosion filled the parking garage with harmful gas and the resulting fires produced visibility impairing smoke. The explosion also caused a water leak causing flooding inside the garage. The user assumes the identity of a firefighter/first responder at the scene.

3.1.2 Defining the scenario

The objective was to create a scenario where the user is situated in an underground parking garage. The garage was filled up with toxic smoke to make the visibility lower and to make the user feel like the situation was dangerous. Various different sounds were added to increase the immersion. These sounds included such things as fire crackling, water flooding in and an emergency siren. The parking garage had three different configurations for different placements of the guiding lights. The lighting systems were placed either on the wall or on the floor and in one of the configurations there were be no guiding lights installed at all. The second half of the garage had two separate water leaks. The water was made to be flooding in from the pipes inside the garage. The floor on that side of the garage was covered in knee-deep water to again increase the immersion of danger. The scenario had a couple of objectives (not just an end goal) to make the experience more goal oriented and at the same time giving the user a sense of accomplishment and keeping them interested in the simulation. The VR HMD was used to show various kinds of information about the objectives so the users always knew what they should be doing next. Because of the toxic smoke, an oxygen tank and the level of oxygen in the tank were also simulated and the user was able to see their oxygen level via their VR HMD. The oxygen level served as a time limit for completing or failing the simulation.

3.1.3 Starting point/end point

The scenario starts with the player standing at the other end of the parking garage (Figure 20). The player is facing the garage area with the ramp behind them. The end point of the scenario is closer to the other end of the garage (Figure 21). The straight-line distance between the starting point and the end point is approximately 68m. However, there is no straight path from the starting point to the end point. The user must navigate through cars, rubble, fire, water and other objects to reach the end objective. The actual end point is the emergency exit that is slightly hidden out of view between a minivan and a concrete beam. Roughly in the middle between the starting point and the end point, there is a small maze of sorts, where the user must navigate their way through burning cars and concrete blocks.

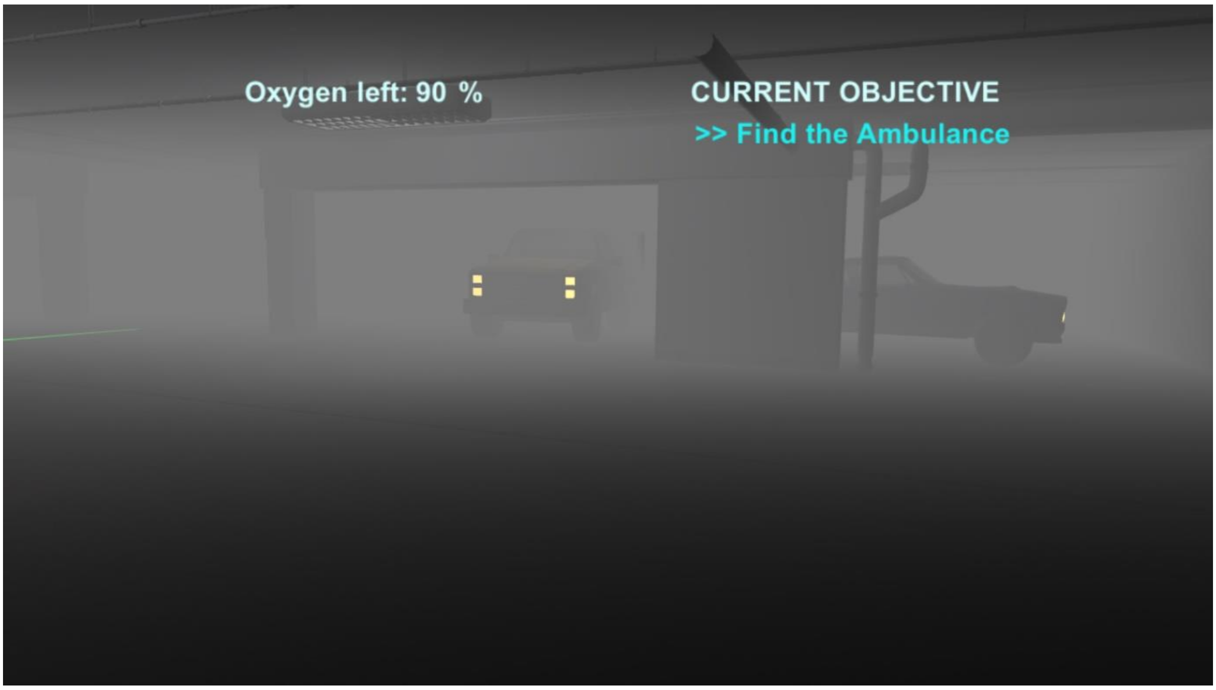


Figure 20. The player’s view from the starting point of the simulation © Mikko Korkiakoski

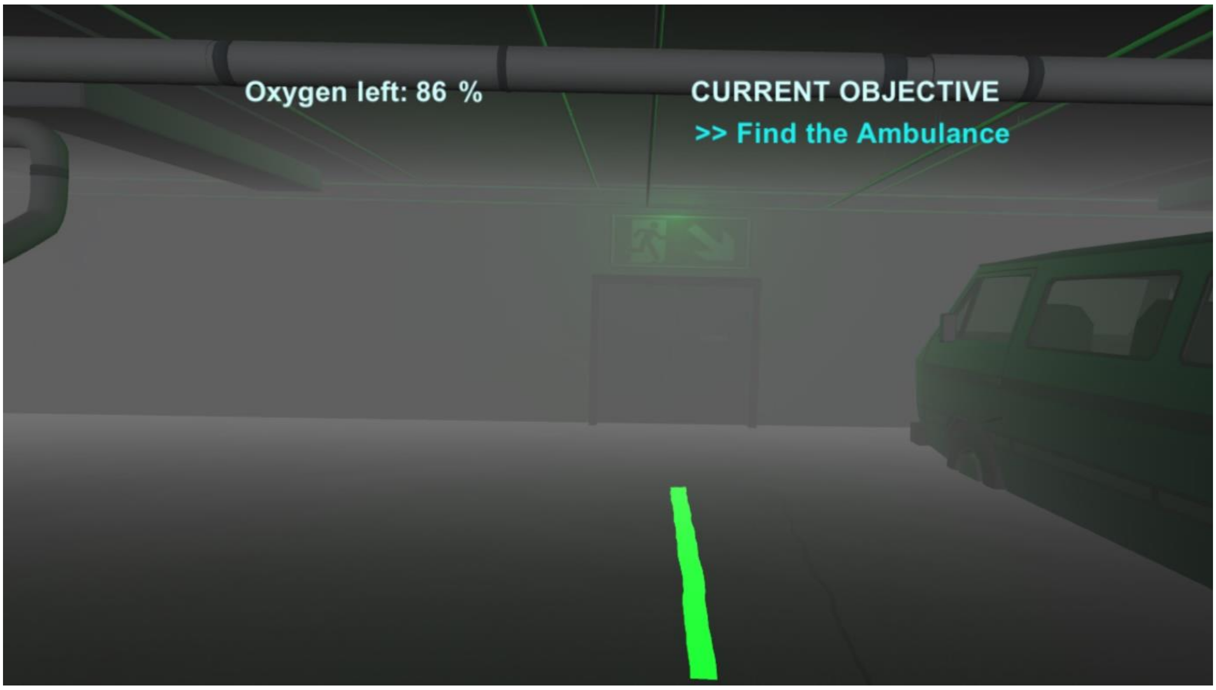


Figure 21. The player’s view of the end-point (emergency exit) of the simulation © Mikko Korkiakoski

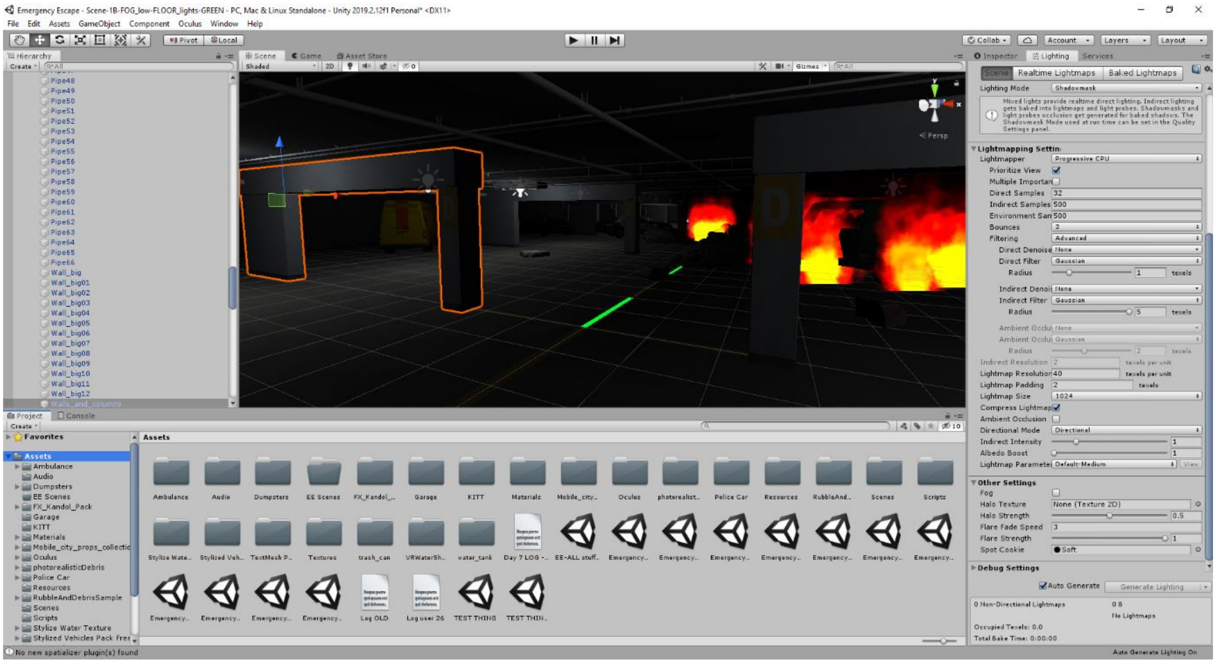
3.1.4 Game objectives

The player has two objectives in the scenario. The first objective is to find the ambulance located somewhere in “first half” of the garage. After the user has found the ambulance, they are given the second objective. The second objective is to locate the emergency exit that is located somewhere in the “second half” of the garage (Figure 22). However, the user can complete the scenario by only completing the second objective that is also the main objective of the scenario. To avoid the user endlessly wondering in the garage, a soft time limit was implemented. The game would not end if the time ran out, but instead the user would be given help in reaching the ambulance, the emergency exit or both if needed.

Figure 22. Pictured here are the starting point (1), first objective (2) and the final objective (3) of the simulation © Mikko Korkiakoski

3.2 Creating the scene

The 3D virtual environment simulation was created using the Unity game engine (Figure 23). Unity uses the C# programming language for scripting different types of actions and events (trigger point events, collision events, sound effects etc.). Scripts were used for logging and gathering data from players’ actions and movement. Scripted triggers were also used for logging data when the player completed an objective and also in creating special events. For playing the simulation, the Oculus Rift S HMD (head mounted display) and two handheld controllers were used.



Emergency Escape - Scene-1B-FOG_low-FLOOR_lights-GREEN - PC, Mac & Linux Standalone - Unity 2019.2.12f1 Personal* <DX11>

Lighting Mode: Shadowmask

Mixed lights provide realtime direct lighting. Indirect lighting gets baked into lightmaps and light probes. Shadowmasks and light probes occlusion get generated for baked shadows. The Shadowmask Mode used at run time can be set in the Quality Settings panel.

Lightmapping Settin
Lightmapper: Progressive CPU
Prioritize View ☑
Multiple Importan ☐
Direct Samples: 32
Indirect Samples: 500
Environment San: 500
Bounces: 2
Filtering: Advanced
Direct Denoise: None
Direct Filter: Gaussian
Radius: 1 texels
Indirect Denoi: None
Indirect Filter: Gaussian
Radius: 5 texels
Ambient Occlu: None
Ambient Occlu: Gaussian
Radius: 2 texels
Indirect Resolution: 2 texels per unit
Lightmap Resolution: 40 texels per unit
Lightmap Padding: 2 texels
Lightmap Size: 1024
Compress Lightmap ☑
Ambient Occlusion ☐
Directional Mode: Directional
Indirect Intensity: 1
Albedo Boost: 1
Lightmap Paramete: Default-Medium

Other Settings
Fog ☐
Halo Texture: None (Texture 2D)
Halo Strength: 0.5
Flare Fade Speed: 3
Flare Strength: 1
Spot Cookie: Soft

Debug Settings

☑ Auto Generate

0 Non-Directional Lightmaps 0 B
No Lightmaps
Occupied Texels: 0.0
Total Bake Time: 0:00:00

Auto Generate Lighting On

Figure 23. A view of the Unity application used to create the simulation © Mikko Korkiakoski

3.2.1 *Objects in the scene*

The virtual environment consists of a model of an underground parking garage and all the different types of objects inside the garage model. The original garage model was deemed too small for the purpose of this simulation, so a new, modified garage was created to make the scenario more challenging and to yield better results. The vehicle models inside the garage include six different types of civilian vehicles, a police car, and an ambulance. Other objects include such things as trash containers and different types rubble etc. To increase the immersion, some fire, water- and smoke particle effects were used, and they were paired with appropriate sound effects. The sound effects included ambient type sounds such as fire crackling and water splashing and flowing. A loud emergency siren sound was used to simulate a real fire/emergency alarm. The smoke effect was simulated using Unity’s built-in fog effect. With fog, it is possible to simulate situations where the visibility is decreased. Using a specific color and “thickness” for the fog it can be made to look like any kind of smoke. All these models and effects work together to make the simulation feel as realistic as possible. Two different sized LED emergency guiding lights were created to be used either on the floor or on the walls of the structure. The color green was used for the guiding lights as it has been shown to be universally the color most often associated with guiding and safety [7]. Some 3D-models of objects and effects used in the simulation are pictured below (Figure 24, Figure 25, Figure 26, Figure 27, Figure 28 and Figure 29).

Figure 24. 3D-model of a trashcan © Mikko Korkiakoski

Figure 25. Realistic rubble © Mikko Korkiakoski

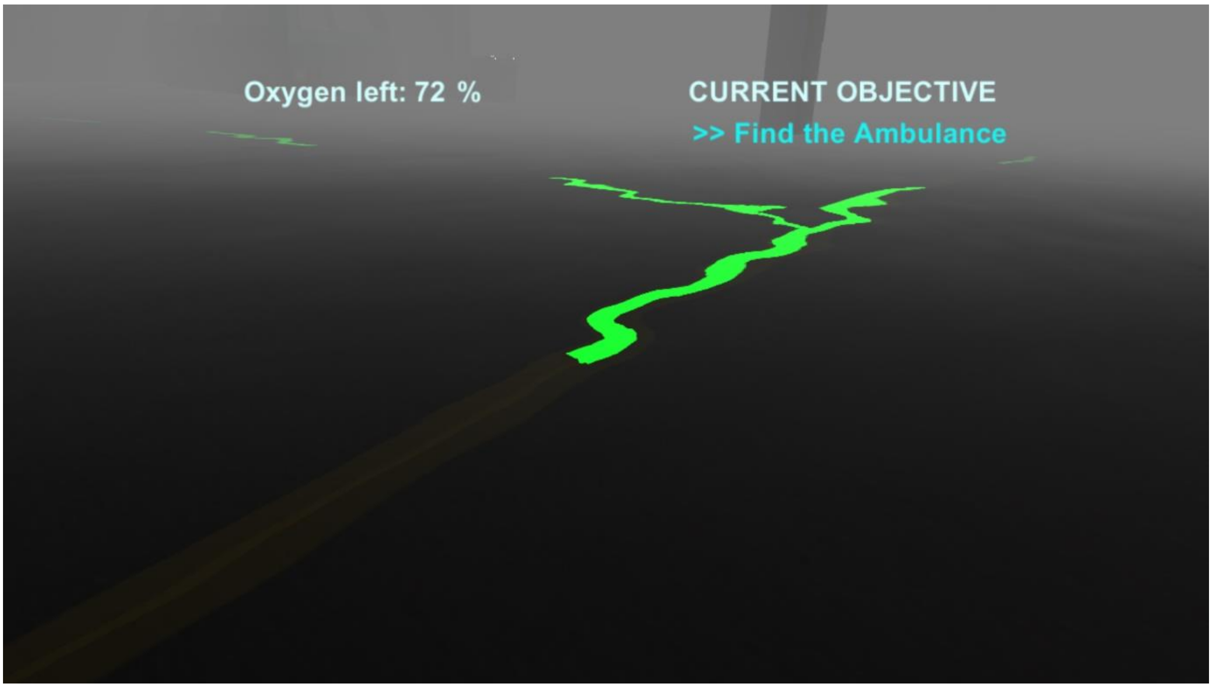


Figure 26. Standing water and distortion effects © Mikko Korkiakoski

Figure 27. 3D-model of an ambulance © Mikko Korkiakoski

Figure 28. 3D-model of a police car © Mikko Korkiakoski

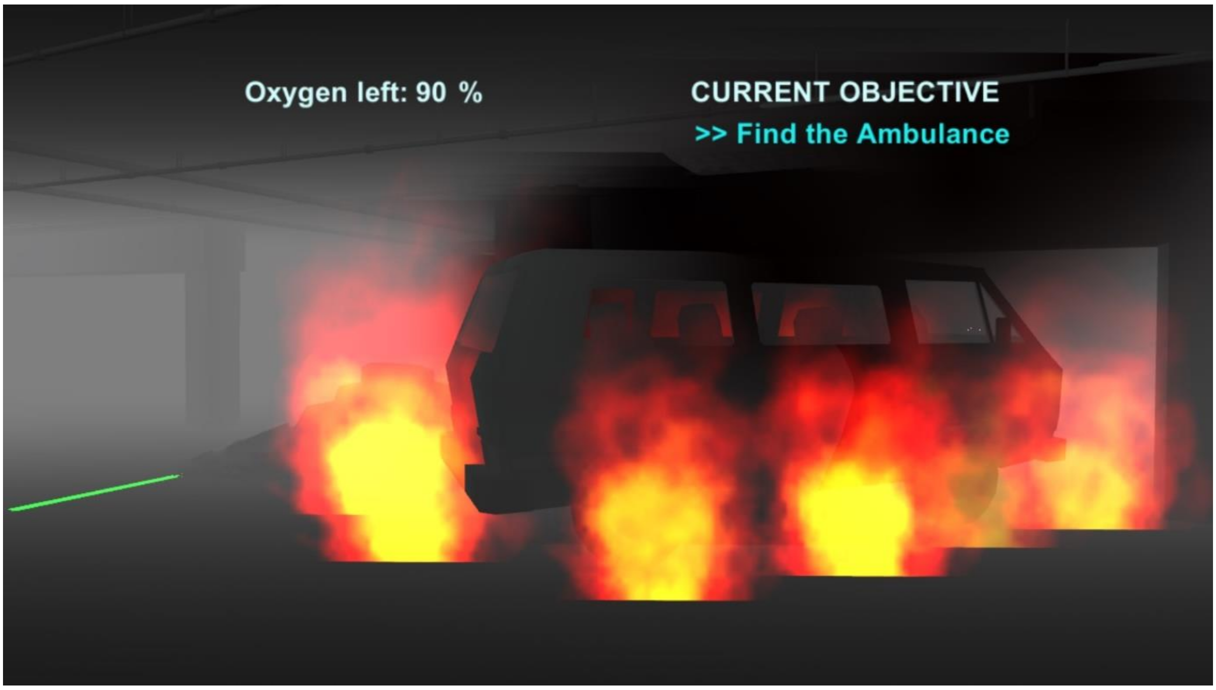


Figure 29. Burning van with fire and smoke effects (C) Mikko Korkiakoski

3.2.2 HMD HUD (heads up display)

A minimal HUD was used for information in the Oculus Rift S HMD (Figure 30). The player would see their current objective via their HMD at all times. Once the “find the ambulance” objective was completed the text in the HUD would change, telling the users the objective was completed and giving them their second objective. Once the player found the emergency exit (completed the simulation) the HUD would display a “mission completed” text and also show the player their simulation completion time and other logged information about their playthrough (Figure 31).

The HUD would also show a fictitious oxygen counter. This oxygen counter would serve as a soft time limit for the simulation, but it was also done so the participant might feel pressured because of the oxygen running out. The time limit was set to 500 seconds because after initial tests it was determined that the users would start to feel like they wanted to quit after spending around 10 minutes wondering around the garage and not being able to complete the objectives.

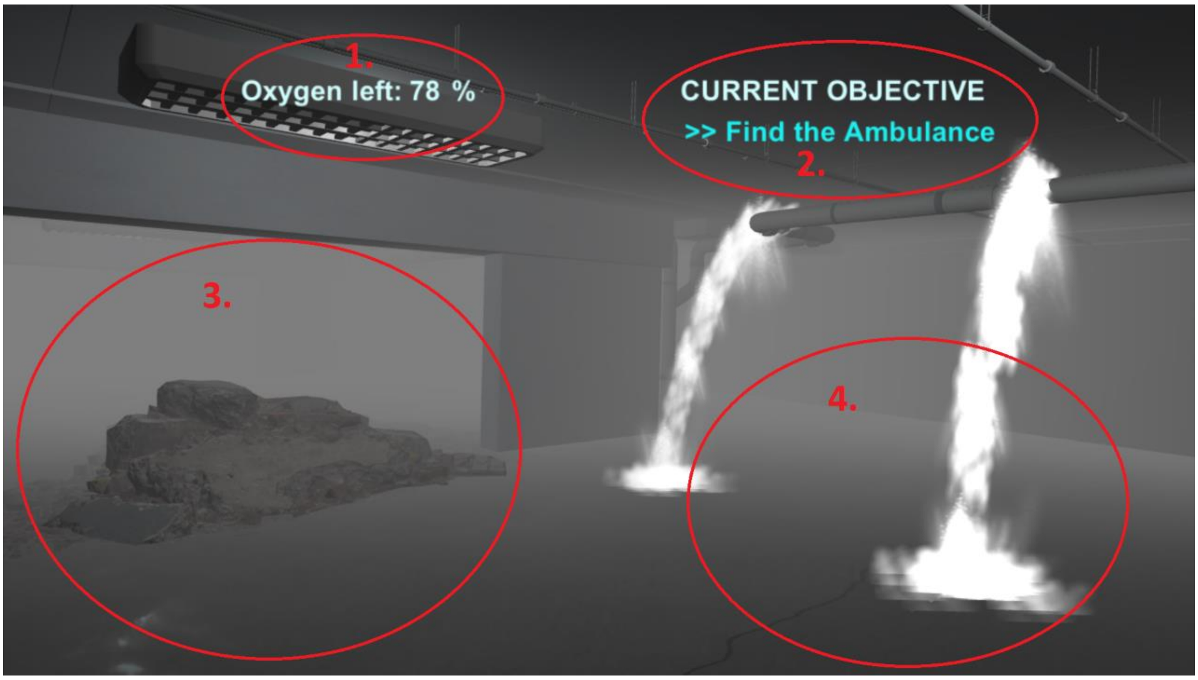
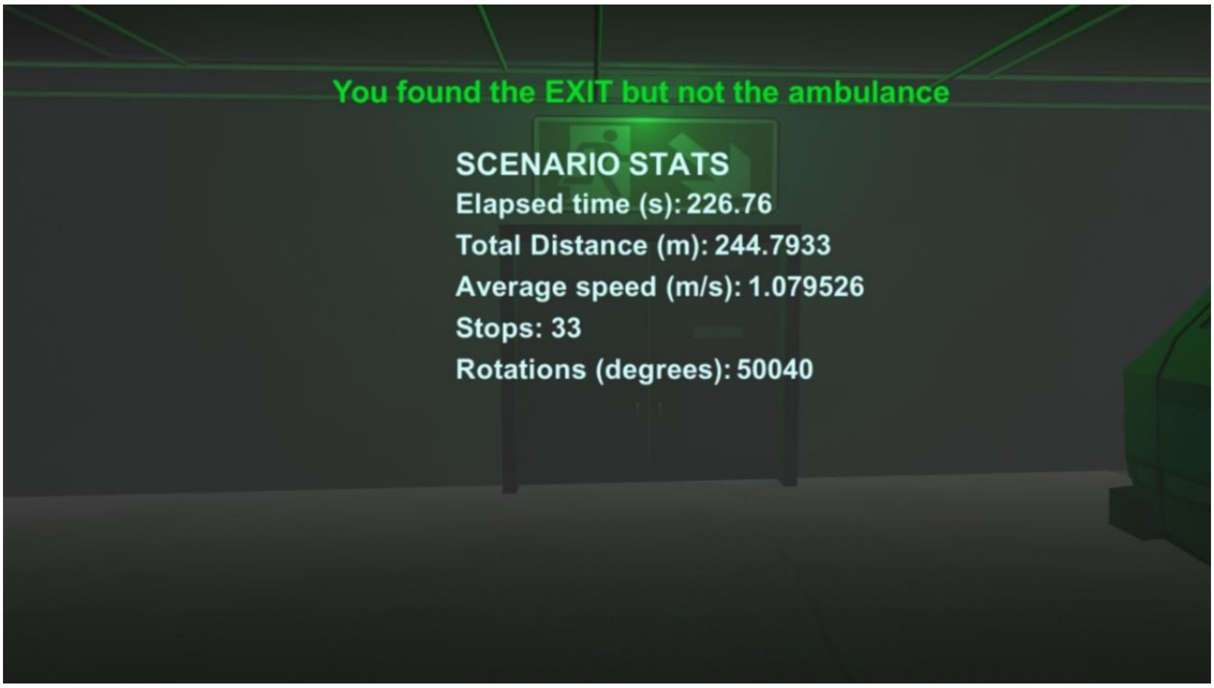


Figure 30. Pictured here are the oxygen counter (1), current objective (2), realistic rubble (3.), standing water and water leak effects (4.) and the surrounding smoke effect © Mikko Korkiakoski



You found the EXIT but not the ambulance

SCENARIO STATS
Elapsed time (s): 226.76
Total Distance (m): 244.7933
Average speed (m/s): 1.079526
Stops: 33
Rotations (degrees): 50040

Figure 31. The player's view of their statistics after finishing the simulation (in this case only the second objective was completed) © Mikko Korkiakoski

3.3 Data gathering

All the user data was logged and calculated using a centralized data gathering object. The object started running different scripts immediately at the start of the simulation. During the simulation everything of interest was logged. After the player reached the end of the simulation, all the logged raw data was exported in a structured format for further analysis of statistical significance. In a situation that the user was unable to complete the simulation, a manual logging option was implemented. This way even a partially completed simulation could be logged and saved by the conducting researcher. The in-game data was gathered during the simulation using the Unity scripting application programming interface (API) [48], and the methods used were as follows.

- **Completion time**: Uses the *Time.time* [49] function to measure the numeric value of seconds elapsed since the beginning of the simulation. The Unity game engine has a built-in function that logs the time the simulation has started. The *Time.time* function simply calculates the time difference between the logged start time and current time in seconds.
- **Distance travelled**: Uses the *Vector3.Distance* [50] to calculate player movement at each frame. A v*ector3* [51] variable stores a player's position vector (*x, y, z*). *Vector3.Distance* compares two *vector3* variables and calculates the distance between these two vector points. For each frame, the player position is captured and stored into a *vector3* variable. And then using *Vector3.Distance*, the distance travelled during the frames is calculated and stored. These distance fragments are then combined to get the aggregated Euclidian distance.
- **Average speed:** Calculated after the simulation has been completed simply by the ratio *distance travelled*/*completion time*.
- **Total stops**: The stops are checked comparing player position vectors (*vector3*). If the previous frame position vector is the same as current frame position vector it is calculated as a stop. The movement of the player character in the simulation is fluid in the sense that if the player releases the movement stick, it still takes a small amount of time for the player character in the game to come to a stop. So, using vector points to calculate the stops, paints an accurate picture of the "real" stops. Because if the player for example is moving hesitantly and keeps stopping for a tenth of a second every now and then, the player character usually does not come to a complete stop and thus, a stop is not calculated.
- **Number of rotations**: Uses the same principle as in total stops. The orientation of the player is checked each frame and compared to the orientation in the previous frame. If the frames have different orientation values, it is calculated as one 45° rotation. This because, the player rotation control stick turns the player 45° each time it is pressed. The rotation of the player is not always 100% accurate for a 45° change and there are instances where the rotation was not be completed inside a frame. Thus, a method that checks for incomplete (overflown) rotation and a rotation that is not exactly 45°, was built into the calculation script. With these extra checks, the rotations can be logged with 100% accuracy. The actual rotation values cannot be obtained from the raw player object data. The raw data first needs to be converted into usable rotation data by using the function *transform.eulerAngles* [52]. After the simulation is completed the calculated rotations are aggregated into one value.

- **Total rotation**: A simple calculation to convert the number of rotations into degrees. As previously mentioned, one rotation equals 45° so the total rotation is number of rotations * 45°.
- **Ambulance found:** (Objective 1): This is just a flag that is raised when the player reaches the ambulance and triggers the "objective completed" event. The flag also prompts a text to be displayed in the Oculus Rift S HMD HUD to let the user know that the first objective was completed.
- **Ambulance time**: The time taken until the ambulance is found and the first objective is completed. It uses the same *Time.time* function as the completion time script. When the player reaches the ambulance trigger, elapsed time at that moment is logged and stored into the data gathering script.
- **Exit found** (Objective 2): When the player reaches the emergency exit, it triggers the end of the simulation by raising the "exit found" flag. This event triggers all the calculations that need to be done to get the logged data into usable form. After calculating the data, they are then written into a separate log file. This event also prompts all the relevant simulation data to be displayed in the Oculus Rift S HMD HUD as well as a simulation completed message. This flag marks the end if the simulation.
- **Stop times** (additional): In addition to counting how many times the user stopped during the simulation, the times at which the stops happened, were also logged and saved. This data was not used in the data analysis but it could be useful if we were to investigate user behavior more closely. For example, it would be easy to investigate how confidently the user was moving around and did they get more confident as the simulation went on simply by looking at the stop count and stop times.
- **First barricade trigger times** (additional): A trigger point was created to the part of the garage just before the player would enter what was considered a "barricade tunnel". In this part of the garage, the player would have to walk a path that had burning cars on each side to make it through the police barricade and to the other side of the parking garage. The trigger was placed there to see if the player would hesitate because of the fires and possible danger. Again, as with the stop times, this data was logged and saved but not used in the data analysis.
- **Second barricade trigger times** (additional): As with the first barricade trigger, a second trigger was placed to the part of the garage from where the player would exit the "barricade tunnel". This data again was logged and saved but not analyzed further. It could be used in tandem with the first barricade trigger times to see how slowly the player walked through the "tunnel" or to see if they lost their bearings and walked back and forth between the two halves of the parking garage.

4 USER TESTING

4.1 Testing setup

All the user testing took place in the Tellus Arena located in the Linnanmaa Campus of the University of Oulu. This location has previously been used in numerous VR field trials [53, 54, 55, 56]. For the tests, a soundproof, glass walled cubicle was used. The cubicle itself was approximately $4m^2$ in area. During the test, only the test subject and the researcher conducting the test were present in the cubicle. Two laptop computers were used for the testing. One laptop had an Nvidia Geforce GTX 1060 (mobile version) graphics card and an Intel i7 processor, while the other one had an Nvidia Geforce RTX 2070 (mobile version) and an Intel i7 processor. Both laptops ran the simulation in Unity with identical settings and an identical framerate. So, despite the hardware being different the end user experience remained identical for all the participants. The VR equipment used during all testing was the Oculus Rift S HMD and two handheld controllers (Figure 32, Figure 33 and Figure 34).

Figure 32. The Oculus Rift S HMD © Mikko Korkiakoski

Figure 33. Oculus right-hand touch controller © Mikko Korkiakoski

Figure 34. Oculus left-hand touch controller © Mikko Korkiakoski

4.2 Participants

A total number of 41 tests were conducted for this experiment. 39 out of the 41 were deemed usable for the purposes of this thesis. The pilot user was excluded from the study because there were several issues while running the test. For the other user, the in-game statistics had to be excluded, because the test had to be aborted. For 30 out of the 39 participants, a Polar A370 heart rate monitor was used to record their heart rate during the simulation. All the participants were selected from a non-probabilistic sample of people entering or leaving the Tellus Arena or from people passing by our testing cubicle. This method, also known as convenience sampling, was used because the testing area was set to remain the same for all the tests and the area itself provided good opportunities for recruiting participants quickly and easily [57, 58]. The Tellus Arena has previously been used in many VR field trial [53, 54, 55, 56]. All participants were over the age of 18 and most of them were students at the Oulu University. 17 women and 22 men participated in the study. The age distribution for all participants was from 19 to 56 years of age with the majority of people being in the 20 to 30 years of age range (M=26.7, SD=6.95). An informed consent was gathered with two signed copies from each participant and of which the other was issued to the participants.

4.3 Heart rate monitoring

For monitoring and recording the heart rate of the participants, the Polar A370 heart rate monitor was used (Figure 35). The A370 uses a photoplethysmogram (PPG) (Figure 36) sensor which means that it optically measures the volumetric changes of an organ [59]. PPG is the signal that the optical heart rate solution measure, interprets and uses to calculate the user's heart rate. This means that, at least theoretically, you could measure the PPG signal anywhere from the body; for example, wrist, earlobe, finger, temple [59]. For this experiment the A370's PPG method was ideal, because the user setup was quick, easy and it did not require a chest strap unlike in monitors using the electrograph (ECG) method. Various Polar HR sensors have been used in several earlier studies and they have been shown to produce consistent results [60, 61, 62, 63, 64, 65, 66]. Previous research into VR gaming and simulations has demonstrated that experiencing VR content can increase the blood flow and affect PPG amplitude [8]. Studies have also shown that experiencing stress-inducing stimuli in VR can lead to elevated levels of stress [4, 5]. Gathering heart rate data from the simulation with different guiding light configurations, allows us to investigate whether the light configurations themselves have any effect, increasing or decreasing, on the participants heart rate. If there are significant changes observed, it could be used to further investigating how well a realistic VR simulation can induce human feelings such as panic, stress, and risk perception.

Figure 35. Polar A370 heart rate monitor © Mikko Korkiakoski

Figure 36. The PPG sensor of the Polar A370 heart rate monitor © Mikko Korkiakoski

4.4 Pre- and post- study questionnaire

Before starting the simulation, all participants filled the pre-study part of the questionnaire. In this part of the questionnaire participants would fill in their basic information such age, height, weight, profession, and gender. The height and weight would be used to calculate the participant's body mass index (BMI) to be used with the heart rate monitor results. The pre-study part also included questions about army service, volunteer firefighting, driver's licence, parking habits and video game and VR experience. Some general questions regarding the participant's wayfinding abilities and memory were included and would be answered using the five-point Likert scale [10].

In the post-study part of the questionnaire, the participants were asked about their experience with the simulation they had just played. This part included questions about their in-game experience and feelings as well as their state of mind immediately after the simulation had been completed and they had taken their VR headset off. All the post-study questions were answered using the five-point Likert scale [10].

4.5 Test protocol

Each test started with the researcher briefly explaining the consent form to the participant. They would then sign two copies of the form, one for them and the other for the researcher. If the heart rate monitor was to be used in the test, it would be given to the participant at this point. They would attach the monitor onto their left wrist and the recording would be started. The participant would be asked about their prior experience with gaming and VR to help determine how detailed the instructions would have to be regarding the simulation and controlling the character inside VR. The pre-study part of the questionnaire would also be done at this point. (Appendix 1.).

The experiment would continue with a brief explanation about the simulation. The participants were told that the simulation would happen in an underground parking garage and that they would assume the identity of a firefighter/first responder. The only other information the participants would receive about the situation was that there had been some kind of an incident in the garage. Next, the objectives of the simulation were revealed. Each participant was told that they would first need to look for the ambulance situated somewhere in the garage. The participants were instructed to walk within touching distance of the ambulance so that the objective completion would be triggered by the software. After finding the ambulance, the final objective for them was to find their way to the emergency exit. The HUD showing the objectives in the HMD was also explained at this point as well as the oxygen counter serving as the time limit. The time limit was explained to be "soft" and that if they could not complete the simulation before the time (oxygen) ran out, the researcher would help them in completing the objectives.

Next, the use of the handheld control devices and the Oculus Rifts S HMD were explained to the participant. The participant was then able to test the HMD and the controllers in a closed tutorial level. The tutorial level consisted of a small, well-lit portion of the garage and one parked car. During the tutorial the participants were asked if they felt any kind of motion sickness or other "weird" feelings. After they had learned the controls and reported feeling good and ready to start the actual simulation, the tutorial was terminated. At this point the participants were briefly reminded about the objectives in the simulation and that they could stop at any point if they felt sick or for any reason at all. It was also pointed out that they should not talk to the researcher during the experiment unless absolutely necessary. This was done so the immersion during the simulation would not break, but also so the participant would try their best and not ask for instructions at the first sight of trouble. However, if the researcher noticed the participant having a particularly difficult time with something, they would be given some hints about what to do.

After the simulation was completed or stopped, the post-study part of the questionnaire was filled. The questionnaire used was the “Game Experience Questionnaire (GEQ) [10] (Appendix 2.) and more specifically the core-, in-game- and post-game modules of the questionnaire. The data recorded from the simulation was marked, checked, and verified at this point. After the participant was done with the questionnaire, the heart rate recording was stopped, and the monitor was removed by the participants themselves. A coffee voucher worth 2€ was given to every person as a thank you for their participation. The testing situation usually lasted from 20 to 25 minutes.

4.6 Analysing the results

4.6.1 ANOVA

For analysing and comparing different performance metrics between different participant groups, one-way ANOVA (one-way analysis of variance) will be used. “ANOVA is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups” [67]. A statistically significant result means a result that unlikely happened by chance [70]. The null hypothesis here is that the means of the groups are exactly equal. The alternative hypothesis states that at least one group mean is different from the rest. The mathematical form of ANOVA can be written as [68]:

$$x_{ij} = \mu_i + \mathcal{E}_{ij}, \tag{1}$$

where x are the individual points (i and j denote the group and the individual observation), $\mathcal{E}$ is the unexplained variation, and the parameters of the model μ are the population means of each group. Each data point x_{ij} is its group mean plus error.

To test if the null hypothesis holds, we need to calculate the test statistic (F-ratio). Using the F-ratio we can find the probability (p value) of obtaining the data assuming the null hypothesis. If the p-value is significant (usually chosen as p < **0.05**) it means that at least one group mean is significantly different from the other group means [68]. The F-ratio is calculated as the ratio of “mean variation between groups and mean variation within groups”. So, in order to find the F-ratio, first we need to calculate these values. Starting with the mean variation between groups, the first step is to find the between-group variation. This is calculated by comparing the mean of each group with the overall mean of the data. For example, if we are comparing three groups (i = 1, 2, 3) then the equation goes as follows (bg = between-group) [68]:

$$bg\ variation\ = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + n_3(\bar{x}_3 - \bar{x})^2. \tag{2}$$

Here $\bar{x}_i$ is the mean for group i, $\bar{x}$ is the overall population mean, and n_i is the sample size. So, by adding up the square of the differences between each group mean and the overall population mean, multiplied by the sample size we get the between-group variation. But in order to get the mean variation between groups we need to divide the between-group variation by the number of degrees of freedom which is n-1 (sample size - 1).

Next, we calculate the variation within groups (wg = within group). The within-group variation is the variation of each observation from the mean of the group. Again, assuming that we have 3 groups of data the equation goes as follow [68]:

$$wg\ variation = s^2_{group1}(n_{group1} - 1) + s^2_{group2}(n_{group2} - 1) + s^2_{group3}(n_{group3} - 1), \quad (3)$$

where we add up the variances (s^2) of the groups (i= 1, 2, 3) multiplied by the number of degrees of freedom of each group. As with the between-group variation we next divide the within-group variation by the total degrees of freedom to find the mean variation within groups. Lastly, we divide the values to find the F-ratio [68]:

$$F - ratio = \frac{mean\ variation\ between\ groups}{mean\ variation\ within\ groups}. \quad (4)$$

After finding the F-ratio, we can obtain the p-value using the F-distribution also known as the probability distribution of the test statistic (F-ratio) [70]. Now, if our p-value is > **0.05**, it means there is not a significant difference between the compared groups and the null hypothesis holds. However, if p < **0.05**, there is a significant difference between some of the groups and the null hypothesis is rejected [67, 68, 69, 70, 71].

4.6.2 Kruskal-Wallis H test and Mann-Whitney U test

For analysing the participants' answers to "The Game Experience Questionnaire" [10] the Kruskal-Wallis H method and the Mann-Whitney U test will be used. "The Kruskal-Wallis H test is a rank-based nonparametric test that is used to determine if there are statistically significant differences between two or more groups of an independent variable on a continuous or ordinal dependent variable" [72]. The null hypothesis for the Kruskal-Wallis H test is that the group populations have equal dominance [73]. This means that when one element is drawn from each group, the largest or smallest element is equally likely to come from any of the groups [73]. If the Kruskal-Wallis H test shows statistically significant differences, then the Mann-Whitney U test can be used to determine between which groups the significance lies.

The first step is to sort the data from all groups into ascending order to form a new combined set. Next, all the sorted data points are assigned a rank value. These assigned ranks are then assigned to the corresponding data points in the original groups and then added up to obtain the group rank sum r [74]. Now we can calculate the test statistic H [73]:

$$H = \left(\frac{12}{n(n+1)} \sum_{i=0}^{j} \frac{r_i^2}{n_i}\right) - 3(n+1), \quad (5)$$

where, j = number of groups, n_i= size of the i^{th} group, r_i is the rank sum for the i^{th} group and n is the total sample size. The obtained H values are then tested against the corresponding chi-square distribution values for j - 1 degrees of freedom and a given significance p. If the chi-square value is less than the obtained H value then the null hypothesis is rejected, otherwise null hypothesis holds. If the null hypothesis is rejected, we can use the Mann-Whitney U test in determining which group was found to have dominance over others [72, 73, 74, 75, 76].

Mann-Whitney U test is a nonparametric test that allows two groups to be compared without making the assumption that values are normally distributed [77]. The null hypothesis here is that the compared groups are equal. The first steps of this test are the same as in the Kruskal-Wallis H test. The datapoints from the two groups are combined, sorted into ascending order and given ranks. The ranks are then added up in the original groups. Next, we calculate the test statistics U_1 and U_2 [78, 79]:

$$U_1 = R_1 - \frac{n_1(n_1+1)}{2}, \tag{6}$$

$$U_2 = R_2 - \frac{n_2(n_2+1)}{2}, \tag{7}$$

where n_i is the sample size for sample i, and R_i is the sum of the ranks in sample i. The smaller of these U values is used in the next step where the test statistic is compared to the Mann-Whitney U critical value. The critical value is obtained from the Mann-Whitney U table and is determined by the sample sizes of the groups and the significance p (commonly **0.05**). If the test statistic value is less than the corresponding critical value, the null hypothesis is rejected and thus, the groups are not equal [77, 78, 79, 80, 81].

4.6.3 Mann-Kendall test and Sen's slope

Heart rate analysis will be done using the Mann-Kendall trend test to determine if there are any detected trends in the heart rate of the users and if these trends differ between the control group and the two experimental groups. The Mann-Kendall test is a non-parametric test that is used in detecting monotonic trends in a series of data. For example, it can be used for analysing environmental data, climate data or in this case, heart rate data. "The null hypothesis for the Mann-Kendall test is that the data come from a population with independent realizations and are identically distributed" [82]. In other words, there is no trend. The alternative hypothesis is that the data follows a monotonic (upward or downward) trend. The Mann-Kendall test statistic S is calculated as follows [82]:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn\left(X_j - X_i\right), \tag{8}$$

where X_i and X_j are the values of sequence i, j; n is the length of the time series and

$$sgn(\theta) = \begin{cases} 1 & if\ \theta > 0 \\ 0 & if\ \theta = 0. \\ -1 & if\ \theta < 0 \end{cases} \tag{9}$$

So, if $S = 0$, there is no trend in the data. If $S < 0$, there is a downward trend in the data. And if $S > 0$, an upward trend is present in the data. Basically, the Mann-Kendall test analyses the differences in signs between the data points. "A trend is present if the sign values tend to increase or decrease constantly" [83]. Every value is compared to the preceding value in the time series [83]. If a trend is present, it can be further analysed by using the Sen's slope test.

The Sen's slope test can be used to measure the magnitude of the trend (slope). "The test computes both the slope (linear rate of change) and intercept according to the Sen's method" [82]. First, the slopes for all the pairs of the ordered time points are calculated. This set of linear slopes is calculated as follows [82]:

$$d_k = \frac{X_j - X_i}{j - i}, \tag{10}$$

for $(1 \leq i < j \leq n)$, where d (median) is the slope, X denotes the variable, n is the number of data and i and j are indices. Next, the actual Sen's slope is calculated. It is calculated as the median from all the slopes that were calculated in the previous step. The intercepts are calculated for each timestep t as follows [82]:

$$a_t = X_t - b * t, \tag{11}$$

where $b =$ median d_k. The corresponding intercept is the median of all the intercepts [82]. Essentially, the value of Sen's slope gives us the magnitude of the detected trend in the data [82, 83, 84, 85, 86, 87, 88].

4.7 Notes about tests and users

Out of the 39 usable test results, six had something noteworthy about them. Three out of these six users reported having feelings of either some kind of motion sickness or dizziness. One of these users said that it was specifically the sound of the siren and seeing the fire that made them feel sick. Among this grouping was also one person that needed some guidance from the researcher on how to finish the scenario. Two other users also needed similar kind of guidance and while one of these two did not specifically ask for help, they had already run out of the allocated time, and so at this point they were basically given clear instructions on how to finish the test. A few other users also received some help but they were only small, general hints about what they can do and just reminders about what their current objective was.

The last of these six users was the only one that had to abort the test. In a discussion prior to the test, the user had said that they often get too much into the game they are playing and lose some sense of reality. They also said that they often get very anxious while playing games. The test itself started out fine but as soon as the user started seeing the fire and smoke, they became very anxious and eventually had to abort the test. It was evident that the user was very scared of walking up close to the fire, so they tried to avoid going near it. Unfortunately, the scenario requires the user to walk close and in between some burning cars and so in the end, this was simply not possible for the user in question. For this user, the scenario statistics are excluded but their answers to the questionnaires were still deemed usable.

5 RESULTS

5.1 Scenario performance metrics comparison

The tests were carried out for three different scenario conditions. The control group played in conditions with no assistive lights (14 users), the second group played the same scene but with assistive lights installed onto the floor of the structure (12 users) and the third group again played the same exact scene but with this time the assistive lights being installed onto the wall of the structure (12 users). All the other conditions, such as starting/ending position and objectives remained the same. The hypothesis is that the two groups with assistive lighting should perform better than the control group with no assistive lights.

Comparing four of the performance metrics (completion time, walking distance, avg. speed, and number of stops) captured, we can see that one group seems to perform different to the other two (Table 2). The control group and the group with assistive wall lights performed very similarly in all tests while the group with the assistive floor lights performed better in three out of the four metrics. Curiously, it was the average walking speed that was lower for this group, but at the same time the group's minimum average speed result was much higher than the other groups and standard deviation much lower than the control group and the "wall-lighting" group.

In the control group there was one instance that turned out to be a complete outlier in the results. One user with a substantial gaming experience "fluked" the simulation by accidentally running straight to the exit. While testing the game it was determined that the simulation could be completed in roughly 20 seconds if the user knew exactly where everything was and what to do. So, to do a blind run through of the scene in 32 seconds was extraordinary. Talking to the user after the experiment, they admitted they "just got lucky" and it was also very evident when following the simulation as well.

If we look at the metrics between just control group and the group with assistive lighting in the wall, they seem to be very close to each other. This is interesting, because the assumption was that the assistive lights would help the user regardless of where they were installed, but the results do not seem to reflect that. This trend was also noticed during the testing and so, during the post experiment discussions with some of the participants, they were asked if they noticed the green light strips on the wall and if they understood that they were there to guide people to the exit. The answers varied a lot with a few people saying they did not even notice them, some saying they thought they were just parking spot markers and only a couple saying they indeed thought they were guiding lights.

Comparing the floor light installation to the wall lights, we see a large difference in almost all the metrics. The floor lighting performed better across the board and users' behaviour during and answers after the experiment reflect this fact as well. It could be seen during the simulations that when the assistive lights were on the floor, many users looked downward when walking and tried to follow the green light strips. When they had to maneuver around obstacles and lost the light trail, they then tried to find the trail again once they had cleared the obstacles. Their answers after the experiment support the findings. Most people said they noticed the floor installation and that they understood that they would lead to the exit. The differences between the installation performances do make sense because the floor lights were situated in the middle of the structure floor, easily visible even with the smoke, while the wall lights were much harder to spot, especially because the starting point of the simulation was situated on the other side of the garage. People also tend to look more towards to floor when the visibility is lower, so the light strips on the wall would be harder to detect than their floor counterparts.

Unfortunately, none of these results show any real statistical significance (Table 3). The differences in all these metrics between the groups are well above our threshold significance (*p*) value of **0.05**. The *p* values for the metrics were: completion time **0.49**, walk distance **0.322**, average speed **0.867** and number of stops **0.512**. The raw numbers obviously show that the floor light installation performed best overall but it cannot be ruled statistically significant.

Finally, if we look at the number of turns the users made in the simulation (Table 4), we can see that again the floor lighting is leading the way but interestingly the wall lighting is the worst out of the three scenarios. Could this indicate that the wall light strips actually confused people instead of helping them? From the conversations with the users, it was clear that not many understood exactly what the wall lights meant. In some cases, the users followed the wall lights but into the wrong direction, so this could also partly explain the increased result in turns. But all in all, it is clear that the floor installation again produced the best results. The *p* value for the number of turns was **0.799 > 0.05**, that is again well above the significance threshold. So, again the floor light group is visibly the best performing group, but the results are not statistically significant.

Table 2. ANOVA statistics for completion time, walk distance, average speed and number of stops for all scenario types

						95% Confidence interval for Mean			
		N	**Mean**	**Standard deviation**	**Standard Error**	**Lower Bound**	**Upper Bound**	**Min**	**Max**
Completion time (s)	No lights	14	201.07	184.539	49.32	94.52	307.62	32	588
	Floor lights	12	142.42	87.943	25.387	86.54	198.29	59	362
	Wall lights	12	207.75	145.3	41.944	115.43	300.07	74	602
	Total	38	184.66	146.269	23.728	136.58	232.74	32	602
Walk distance (m)	No lights	14	554.86	341.27	91.208	357.81	751.9	152	1194
	Floor lights	12	432	195.886	56.547	307.54	556.46	209	900
	Wall lights	12	582.08	187.116	54.016	463.2	700.97	341	977
	Total	38	524.66	258.72	41.97	439.62	609.7	152	1194
Avg. speed (m/s)	No lights	14	3.5957	1.37786	0.36825	2.8002	4.3913	1.59	5.71
	Floor lights	12	3.3317	0.98293	0.28375	2.7071	3.9562	2.17	5.34
	Wall lights	12	3.4942	1.35003	0.38972	2.6364	4.3519	0.89	5.17
	Total	38	3.4803	1.22816	0.19923	3.0766	3.8839	0.89	5.71
Number of stops	No lights	14	49.79	67.48	18.035	10.82	88.75	3	220
	Floor lights	12	25.25	21.592	6.233	11.53	38.97	5	72
	Wall lights	12	49.33	74.15	21.405	2.22	96.45	4	263
	Total	38	41.89	59.199	9.603	22.44	61.35	3	263

Table 3. ANOVA significance for completion time, walk distance, average speed and number of stops for all scenario types

		Sum of squares	df	Mean square	F	Significance
Completion time (s)	Between Groups	31582.457	2	15791.229	0.727	0.49
	Within Groups	760014.095	35	21714.688		
	Total	791596.553	37			
Walk distance (m)	Between Groups	155365.922	2	77682.961	1.171	0.322
	Within Groups	2321266.631	35	66321.904		
	Total	2476632.553	37			
Avg. speed (m/s)	Between Groups	0.454	2	0.227	0.143	0.867
	Within Groups	55.356	35	1.582		
	Total	55.81	37			
Number of stops	Between Groups	4860.305	2	2430.153	0.682	0.512
	Within Groups	124805.274	35	3565.865		
	Total	129665.579	37			

Table 4. ANOVA statistics for the number of 45° turns made for all scenario types

					95% Confidence interval for Mean			
	N	Mean	Standard deviation	Standard Error	Lower Bound	Upper Bound	Minimum	Maximum
No lights	14	275.86	198.402	53.025	161.3	390.41	24	704
Floor lights	12	241	187.019	53.988	122.17	359.83	60	562
Wall lights	12	300.83	267.25	77.149	131.03	470.64	62	1075
Total	38	272.74	214.59	34.811	202.2	343.27	24	1075

5.2 Comparison between genders

Looking at the differences in performance metrics between males (22 participants) and females (16 participants) over the 3 groups, we can see there is some difference between the completion times (Table 5). Males on average were almost 80 seconds faster in completing the simulation. If we look at some of the answers in the questionnaire, there is a clear difference in how often people play videogames or use VR equipment. 8 out of the 22 males and only 2 out of 16 females reported playing video games daily. In addition, 9 males reported playing games weekly or monthly and again for females the number was much lower with only a single participant reporting playing games on a weekly basis. Regarding VR usage, 3 males used VR equipment monthly and two females said they used VR either weekly or monthly. For males, the daily gamers recorded 4 out of 5 fastest overall completion times. The fastest time for a female was also recorded by a daily gamer. Even though the statistics show that men in general performed better in the simulation the *p* value **0.104** of this is not below the threshold value of **0.05** and thus, this cannot be called a significant finding (Table 6).

Table 5. ANOVA statistics for completion times (s) between genders

					95% Confidence interval for Mean			
	N	**Mean (s)**	**Standard deviation (s)**	**Standard Error**	**Lower Bound (s)**	**Upper Bound (s)**	**Minimum (s)**	**Maximum (s)**
Male	22	151.73	144.757	30.862	87.55	215.91	32	588
Female	16	229.94	140.211	35.053	155.22	304.65	70	602
Total	38	184.66	146.269	23.728	136.58	232.74	32	602

Table 6. ANOVA significance for completion times (s) between genders

	Sum of squares	**df**	**Mean square**	**F**	**Significance**
Between groups	56661.251	1	56661.251	2.775	0.104
Within groups	734935.301	36	20414.869		
Total	791596.553	37			

When we look at the average walking speeds between the genders, we can see that there is quite a significant difference there (Table 7). For males, the average walking speed was nearly 3.9 m/s while females averaged almost one metre per second less. 5 males recorded an average speed that was over 5 m/s while only the fastest female achieved the same. However, walking fast does not always mean faster completion time, because even the person with the quickest completion time did not average over 5 m/s, in fact they were well below it with an average speed of 4.77 m/s. It is notable however, that males were clearly more confident in moving fast while exploring the area. The ANOVA statistics indicate that this difference in walking speed is in fact quite significant, with the p value being well below our threshold value **0.013 < 0.05** (Table 8).

Table 7. ANOVA statistics for average walking speed (m/s) between the genders

					95% Confidence interval for Mean			
	N	**Mean (m/s)**	**Standard deviation (m/s)**	**Standard Error (m/s)**	**Lower Bound (m/s)**	**Upper Bound (m/s)**	**Minimum (m/s)**	**Maximum (m/s)**
Male	22	3.8945	1.18195	0.25199	3.3705	4.4186	1.59	5.71
Female	16	2.9106	1.08029	0.27007	2.335	3.4863	0.89	5.33
Total	38	3.4083	1.22816	0.19923	3.0766	3.8839	0.89	5.71

Table 8. ANOVA significance for average walking speed (m/s) between the genders

	Sum of squares	**df**	**Mean square**	**F**	**Significance**
Between groups	8.968	1	8.968	6.892	0.013
Within groups	46.842	36	1.301		
Total	55.81	37			

The walking distance and the number of stops results show more of the same. Males were not only faster on average in completing the simulation but also walked faster during the simulation and the same is seen here with the walking distance and the number of stops made. The males walked around 100 meters less than their female counterparts and also made almost half the number of stops with ~30 for males and 58 for females respectively (Table 9). The maximum distance and stop values, however, are closer together than the averages and the minimums, so this could indicate that the differences in averages are not simply because of gender. The *p* values for these metrics (**0.259** and **0.155**) indicate that these findings are not statistically significant as they are well above the **0.05** threshold (Table 10).

Table 9. ANOVA statistics for average walking distance (m) and the number of stops during the simulation between genders

						95% Confidence interval for Mean			
		N	**Mean**	**Standard deviation**	**Standard Error**	**Lower Bound**	**Upper Bound**	**Minimum**	**Maximum**
Walking distance (m)	Male	22	483.77	249.724	53.241	373.05	594.49	152	1163
	Female	16	580.88	268.289	67.072	437.91	723.84	277	1194
	Total	38	524.66	258.72	41.97	439.62	609.7	152	1194
Number of stops	Male	22	30.18	50.02	11.091	7.12	53.25	3	220
	Female	16	58	66.166	16.542	22.74	93.26	10	263
	Total	38	41.89	59.199	9.603	22.44	61.35	3	263

Table 10. ANOVA significance for average walking distance (m) and the number of stops between the genders

		Sum of squares	**df**	**Mean square**	**F**	**Significance**
Walking Distance (m)	Between groups	87340.939	1	87340.939	1.316	0.259
	Within groups	2389291.614	36	66369.211		
	Total	2476632.553	37			
Number of stops	Between groups	7168.306	1	7168.306	2.107	0.155
	Within groups	122497.273	36	3402.702		
	Total	129665.579	37			

All in all, it is fair to suggest that in a VR simulation like this, gaming experience helps regardless of the gender. This is most likely because the controller and control scheme used to move the player in the simulation are similar to a gamepad and that gamers in general have more confidence in roaming around the simulation, again due to their gaming experience. So, the differences between genders might actually not be that significant and the difference in performance may simply be due to the male group having more members with extensive and/or ongoing gaming experience. Let us take a look at solely the gaming experience next to see if this assumption holds.

5.3 Gaming/VR experience impact

Looking specifically the gaming and VR experience of the participants and comparing the metrics, some interesting but probably not that surprising results can be found (Table 11). Comparing people who play videogames or use VR equipment at least weekly and people who only did these activities monthly or even seldomly, a clear distinction can be seen between the groups.

The gamers beat the non-gamers in all the important measured metrics. But more importantly the significance of these result is high, with four out of three *p* values (**0.009**, **0.004** and **0.017**) being well below the threshold value of **0.05** (Table 12). Only the walk distance metric falls a bit short of the threshold value with **0.056**. As was already seen when comparing the performance metrics between genders, gaming experience does indeed have a big impact in this kind of testing.

Table 11. ANOVA statistics for completion time, walk distance, average speed and number of stop in relation gaming/VR experience

						95% Confidence interval for Mean			
		N	**Mean**	**Standard deviation**	**Standard Error**	**Lower Bound**	**Upper Bound**	**Minimum**	**Maximum**
Completion time (s)	Daily/Weekly	18	120.78	57.443	13.539	92.21	149.34	32	226
	Monthly/less	20	242.15	177.103	39.601	159.26	325.04	59	602
	Total	38	184.66	146.269	23.728	136.58	232.74	32	602
Walk distance (m)	Daily/Weekly	18	440.61	143.921	33.922	369.04	512.18	152	689
	Monthly/less	20	600.3	314.804	70.392	452.97	747.63	209	1194
	Total	38	524.66	258.72	41.97	439.62	609.7	152	1194
Avg. speed (m/s)	Daily/Weekly	18	4.0594	1.16992	0.27575	3.4777	4.6412	1.79	5.71
	Monthly/less	20	2.959	1.05305	0.23547	2.4662	3.4518	0.89	5.17
	Total	38	3.4803	1.22816	0.19923	3.0766	3.8839	0.89	5.71
Number of stops	Daily/Weekly	18	18.22	16.861	3.974	9.84	26.61	3	61
	Monthly/less	20	63.2	74.575	16.675	28.3	98.1	7	263
	Total	38	41.89	59.199	9.603	22.44	61.35	3	263

Table 12. ANOVA significance for gaming/VR experience

		Sum of Squares	**df**	**Mean Square**	**F**	**Significance**
Completion time (s)	Between Groups	139558.892	1	139558.892	7.705	0.009
	Within Groups	652037.661	36	18112.157		
	Total	791596.553	37			
Walk distance (m)	Between Groups	241584.075	1	241584.075	3.891	0.056
	Within Groups	2235048.478	36	62084.68		
	Total	2476632.553	37			
Avg. speed (m/s)	Between Groups	11.472	1	11.472	9.315	0.004
	Within Groups	44.338	36	1.232		
	Total	55.81	37			
Number of stops	Between Groups	19165.268	1	19165.268	6.244	0.017
	Within Groups	110500.311	36	3069.453		
	Total	129665.579	37			

5.4 Participant pre-test self-evaluation

In the pre-test part of the questionnaire, the participants were asked to evaluate a few things about their route-finding and memory for example. The participants answered to the statements using a scale of 0-4, 0 being “completely disagree” and 4 being “completely agree” (Table 13). Let us take a brief look at a few examples of how this self-evaluation relates to performance in the simulation. Unfortunately, four out of the five metrics provided no significant results so let us take a look at the statement only statement that had some significance - “My sense of direction is very poor”.

Table 13. ANOVA significance for self-evaluation "My sense of direction is very poor”

	Answer	**N**	**Mean**	**Significance between groups**
Completion time (s)	0	14	107.79	0.117
	1	11	220.55	
	2	6	202.5	
	3	4	236.5	
	4	3	307	
	Total	38	184.66	
Walk distance (m)	0	14	434.43	0.264
	1	11	520.55	
	2	6	562.67	
	3	4	596.25	
	4	3	789.33	
	Total	38	524.66	
Avg. speed (m/s)	0	14	4.325	0.011
	1	11	3.0227	
	2	6	2.9267	
	3	4	2.52	
	4	3	3.6033	
	Total	38	3.4803	
Number of stops	0	14	20.93	0.503
	1	11	59.64	
	2	6	50	
	3	4	35.25	
	4	3	67.33	
	Total	38	41.89	

Just by looking at the results it would seem that the people who rated their sense of direction as very good, performed the best in all these metrics. However, the differences in all but one metric are not significant enough to conclude that people can always rate their sense of direction correctly, at least not when it is put to the test in this simulation. In average speed the difference between the groups was significant with a *p* value of **0.011** < **0.05** (Table 13). Further analysis shows that the significant differences were between groups “0 and 1” and “0 and 3”. It is interesting that people who rated their sense of direction as “very poor” still scored the second-best overall result. However, the sample size for groups 3 and 4 was rather small so one cannot draw concrete conclusions about the significance of this result. But it is noteworthy that in all the metrics, the people who replied "0" (strongly disagreeing with the statement) outperformed all the other groups by quite a substantial difference.

5.5 The Game Experience Questionnaire results

The Game Experience Questionnaire [10] lets the participant self-assess how they felt about the simulation and how they thought they performed in it. The questionnaire we used is divided into three parts. The first two parts (core module and in-game module), are about the experience during the simulation. The third part is about the participant's feelings after the simulation had ended. Note that for some of these results the number of participants is 37 and not 38, simply because one participant forgot to fill in the answers needed for the metric to be calculated.

Using the Kruskal-Wallis H test to analyse the core module (Table 14), we find that there is possibly a significant statistical difference between the groups in "negative affect". The *p* value for negative affect (**0.031**) is below our threshold of **0.05**.

Table 14. Kruskal-Wallis H significance for core module

	Competence	**Sensory and imaginative immersion**	**Flow**	**Tension/ annoyance**	**Challenge**	**Negative affect**	**Positive affect**
Kruskal-Wallis H	0.784	3.279	4.079	4.475	5.176	6.973	0.803
df	2	2	2	2	2	2	2
Asymp. Sig.	0.676	0.194	0.13	0.107	0.075	0.031	0.669

Table 15. Ranks for core negative affect

	Scene type	**N**	**Mean Rank**
Negative affect	No lights	14	13.79
	Floor lights	12	24.63
	Wall lights	12	21.04
	Total	38	

Because we do not know in between which groups the significant difference is, we use the Mann-Whitney U test to compare the three groups against each other (Table 15, Table 16 and Table 17). Using this kind of test will increase the error rate but we can account for this by dividing our *p* value with the number of comparisons we need to do. So, our new *p* value is 0.05 divided by the number of tests we need to run, which is 3 = **0.0167**.

Table 16. Ranks for no lights and floor light groups compared

	Scene type	**N**	**Mean Rank**	**Sum of Ranks**
Negative affect	No lights	14	10.32	144.5
	Floor lights	12	17.21	206.5
	Total	26		

Table 17. Mann-Whitney U significance for negative affect

	Negative affect
Mann-Whitney U	39.5
Wilcoxon W	144.5
Z	-2.417
Asymp. Sig. (2-tailed)	0.016
Exact Sig. [2*(1-tailed Sig.)]	.020^b

Comparing the control group with no lights and the group with the floor lights, we can see that the 2-tailed asymptotic significance value of **0.016** is just below our new threshold value of **0.0167** (Table 17). So, we can conclude that there is indeed a significant statistical difference between these two groups in this metric. The rank value of **24.63** for the floor lighting group was the highest of all the groups. And as shown in the Mann-Whitney U test, its rank value was significantly higher than the control group's **17.21** > **10.32**. Repeating the same Mann-Whitney U test for the control group and group with wall lights produced a *p* value of **0.053**. The *p* value between the floor light group and the wall light group was **0.318**. Both of these values are over our new threshold value of **0.0167** and thus, they cannot be considered statistically significant. It is interesting that participants in the group with the best of overall simulation results (group with floor lights) were also the group that felt the most negative about the simulation experience, at least when compared to the group that had no assistive lighting.

Moving on to the in-game module we see a similar trend with the Kruskal-Wallis H test results (Table 18). Again, the only metric to fall under our primary threshold value is the negative affect value of **0.043** < **0.05**.

Table 18. Kruskal-Wallis H significance for in-game module

	Competence	**Sensory and imaginative immersion**	**Flow**	**Tension**	**Challenge**	**Negative affect**	**Positive affect**
Kruskal-Wallis H	0.726	2.718	1.821	3.515	0.905	6.289	2.838
df	2	2	2	2	2	2	2
Asymp. Sig.	0.696	0.257	0.402	0.173	0.636	0.043	0.242

Table 19. Ranks for in-game negative affect

	Scene type	**N**	**Mean Rank**
Negative affect	No lights	14	13.89
	Floor lights	12	20.54
	Wall lights	11	23.82
	Total	37	

Using the Mann-Whitney U test again to determine where the differences are, we see that now the difference that is statistically significant is between the control group with no lights and the group with the wall lights. The *p* value (**0.001**) (Table 21) also falls under Mann-Whitney U threshold of **0.0167**. The rank score for the wall light group (**23.82**) is quite close to the floor light group (**20.54**) but significantly higher than the control group (**13.89**) (Table 19). And as shown by the Mann-Whitley comparison test the difference to the control group is significant with the rank being **17** > **9.86** (Table 20). For the comparison between the control group and the floor light group, the p value was **0.133** and between floor light group and the wall light group the value was **0.566**. So, for this in-game module the significant metric is the same as for the core module. These two modules are somewhat overlapping so this result is not that surprising. However, this time the significant difference was between different groups. Interestingly both of these results for negative affect suggest that the participants felt more negative about the simulation when there were assistive lights installed.

Table 20. Ranks for the compared groups

	Scene type	**N**	**Mean Rank**	**Sum of Ranks**
In-game negative affect	No lights	14	9.86	138
	Wall lights	11	<u>17</u>	187
	Total	25		

Table 21. Mann-Whitney U significance for negative affect

	Negative affect
Mann-Whitney U	33
Wilcoxon W	138
Z	-2.591
Asymp. Sig. (2-tailed)	<u>0.01</u>
Exact Sig. [2*(1-tailed Sig.)]	.015^{b}

Lastly, let us take a look at the post-game module results for the Kruskal-Wallis H test (Table 22). The post-game module differs from the other two modules in that it is all about the feelings immediately after the simulation has ended and not the actual simulation itself.

Table 22. Kruskal-Wallis H significance for post-game module

	Positive affect	**Negative affect**	**Tiredness**	**Returning to reality**
Kruskal-Wallis H	4.799	4.139	1.575	2.895
df	2	2	2	2
Asymp. Sig.	0.091	0.126	0.455	0.235

This time none of the metrics fall under our threshold value of **0.05** so there is no reason to investigate these stats further with the Mann-Whitney U test.

5.6 Heart rate measurements

For heart rate monitoring, the expected results were that the participants who played the simulation with assistive lights would possibly have a lower heart rate in general or have a downward trend compared to the control group that had no assistive lights. The heart rate was measured for 13 participants in the control group, for 9 participants in the floor light group and for 8 participants in the wall light group.

The results were analysed using the Mann-Kendall trend test. With this test it is possible to detect if there are any statistically significant trends (upward or downward) in the measure data. Kendall's tau (positive or negative) shows the direction of the trend while Sen's slope value gives us the magnitude of the trend. The null hypothesis H_0 is that there is no statistically significant trend in the data. So, if the significance value $\boldsymbol{p < 0.05}$ the trend is deemed significant and thus, the null hypothesis is rejected and the conclusion is that there is indeed a noticeable trend. Otherwise, the null hypothesis is accepted and there is no statistically significant trend.

The control group results were mixed (Table 23). For six participants there was a significant downward trend in their heart rate but at the same time there was an upward trend detected also for six participants, while only one result showed no significant trends. The slope magnitudes were also fairly low so it is difficult to draw any conclusions from this group's results alone.

Table 23. Mann-Kendall test results for the control group (no lights)

user#	HR (min - max)	Kendall's tau	p (Two-tailed)	H_0	Sen's slope value	trend
5	66 - 82	-0.302	<0.0001	reject	-0.050	downward
9	68 - 105	0.594	<0.0001	reject	0.279	upward
12	73 - 95	-0.182	0.006	reject	-0.66	downward
13	73 - 82	-0.18	0.001	reject	-0.007	downward
15	62 - 69	-0.323	0	reject	-0.026	downward
16	78 - 95	0.656	<0.0001	reject	0.059	upward
17	82 - 100	0.493	<0.0001	reject	0.046	upward
18	83 - 93	0.096	0.008	reject	0	upward
19	92 - 107	0.88	<0.0001	reject	0.286	upward
20	111 - 131	0.162	<0.0001	reject	0.009	upward
37	76 - 82	0.215	0.119	accept	0	none
39	79 - 96	-0.379	<0.0001	reject	-0.013	downward
40	112 - 119	-0.623	<0.0001	reject	-0.061	downward

The results for the floor light group are a bit different to the control group (Table 24). Just over half of the participants in this group (five out of nine) had a downward trend in their heart rate during the simulation. Two participants had an upward trend in their heart rate and another two had no statistically significant trend in theirs. While the downward trending heart rate is the most common result in this group, they only make up for just slightly over half (**56%**) of the results. So, the results for this group are mixed just like they were for the control group.

Table 24. Mann-Kendall test results for the group with floor lights

user#	HR (min - max)	Kendall's tau	p (Two-tailed)	H_0	Sen's slope value	trend
4	74 - 97	0.061	0.092	accept	0	none
7	61 - 71	-0.556	<0.0001	reject	-0.09	downward
8	74 - 84	-0.609	<0.0001	reject	-0.076	downward
14	85 - 91	-0.322	<0.0001	reject	-0.017	downward
28	63 - 86	0.023	0.592	accept	0	none
31	90 - 98	-0.665	<0.0001	reject	-0.1	downward
32	89 - 102	0.831	<0.0001	reject	0.188	upward
34	88 - 99	-0.191	0.001	reject	-0.014	downward
38	76 - 88	0.348	<0.0001	reject	0.033	upward

Lastly, we have the measurements for the wall light group (Table 25). In this group the majority of participants (six out of eight) had a statistically significant downward trend in their heart rate, while only one participant had an upward trend and in one case there was no significant trend. The results for this group were more in line with what was expected with **75%** (6/8) of the participants showing the expected downward trend in their heart rate.

Table 25. Mann-Kendall test results for the group with wall lights

user#	HR (min - max)	Kendall's tau	p (Two-tailed)	H_0	Sen's slope value	trend
6	82 - 95	-0.516	<0.0001	reject	-0.065	downward
10	69 - 77	-0.123	0.068	accept	0	none
11	65 - 80	-0.287	<0.0001	reject	-0.021	downward
29	83 - 94	-0.381	<0.0001	reject	-0.06	downward
30	77 - 87	-0.21	0.001	reject	-0.015	downward
33	85 - 130	-0.779	<0.0001	reject	-0.212	downward
35	54 - 78	0.897	<0.0001	reject	0.225	upward
36	65 - 81	-0.501	<0.0001	reject	-0.049	downward

Looking at the trends it can be seen that the experimental groups with the assistive lighting had relatively more downward trending results than the control group. As was previously noted, the wall light group had the largest share of downward trending heart rates with a share of **75%** (6/8). For the floor light group, the share was **56%** (5/9) and for the control group with no assistive lights **46%** (6/13). An upward trend was detected for **46%** (6/13) of the participants in the control group, for **22%** (2/9) in the floor light group and for **13%** (1/8) in the wall light group. No trend was detected for **8%** (1/13) for the control group, **22%** (2/9) of the floor light group and **13%** (1/8) of the wall light group

A bit surprisingly the wall light group scored the most expected results in this case. Many participants reported that they didn’t really know if the wall lights were actually there to guide them to the exit and some thought they were just there as parking space markers. It was very different for the floor lights with most people reporting that they understood that they were there to help them get to the exit. Reflecting on these user reports and results it is interesting that still the wall light group shows the most promising results here. The mix of upward- and downward trends detected in the control group is also quite surprising given that this group was usually the worst performing one. One explanation for the odd results could be that many participants became visibly bored or annoyed in this specific scenario after wondering around for a long period of time and not finding the exit. This might have skewed the heart rate trends in some way.

It must also not be forgotten that people who have previous experience with video games and/or VR environments will most likely remain calmer in these types of simulations compared to people for whom this is a completely new or a very rare experience. And for serious gamers and VR veterans it is probably just a normal or even an enjoyable experience so they can almost relax during the simulation. So, some of the variation in the results can definitely be attributed to some people simply having more experience in gaming and VR.

6 DISCUSSION AND CONCLUSIONS

6.1 Research problem

The main purpose of this thesis was to investigate how different guiding light placements affect peoples' ability of finding an exit in case of an emergency. There have been previous studies in this area [3, 6, 7] that have investigated similar conditions such as light placement, type and color. The goal for this thesis was to take the previous research results as a baseline and see if something new could be learned by applying them into a different scenario. Two different configurations of assistive lighting were tested against the control configuration with no lights. The light placements and the color were chosen on the basis of previous research that suggested that these were the best working solutions [3, 6, 7].

The hypothesis regarding the guiding lights was that they would perform significantly better than the control situation with no lights. Still, there was no specifically set goal, but instead the focus was to record how people performed during the simulation and then comparing the recorded performance statistics and different participant groups. The gaming questionnaire [10] was used as a part of the study to determine if peoples' experiences, feelings or previous activities affected or correlated with their performance statistics. Heart rate measurements were included to see if the different lighting configurations affected the participants in any meaningful physiological way. The starting hypothesis was that the guiding light systems would produce better results than the scenario with no lights. And although there have been previous studies in similar research that included heart rate measurements, there were no assumptions made beforehand about the results they would yield in this study.

6.2 Performance metrics review

Looking at the performance metrics gathered from the simulation it is clear that one group performed better than the other two. The group with the floor light installation scored the best results on all but one of the measurements. This result was expected as the previous research [3] pointed to this type of installation being the best. But when comparing the means of the results between the groups the differences are not significant enough to draw the conclusion that one installation type is better than the other.

The participants were asked about the light installations during the post-game discussions and when it came to the floor light installation, many replied that they did indeed understand their purpose. However, because there was an obstruction preventing the participants from following the floor installation path all the way to the exit, some participants said that they forgot the lights were even there while they were searching a way around the obstacle. Of course, the problem was the same for the group that had the wall light installation. But the majority of people in this group did not correctly understand the meaning of the wall lights.

The light installations in this simulation were simple, green light strips on the floor or on the wall. So perhaps future trials could test different types of markings like for example arrow shaped markings pointing the way with some text or imagery included as well. In a case of an obstruction there could be an alternative guiding system or possibly using both floor and wall installations at the same time.

6.3 Group comparisons

As was previously stated the differences with the main focus groups, the different light installations, were not statistically significant. Similar results were found when male and female performance metrics were compared. These metrics did show that males in general performed better than females in the simulation, but again the differences were not statistically significant. Only the walking speed metric showed males walking significantly faster than females but as the investigation into gaming experience showed this might not be gender related at all.

When the groups were compared according to their gaming experience, the results were interesting but also quite expected. In three out of the four main performance metrics (completion time, walking speed and number of stops) the comparisons showed statistically significant differences between the groups. The participants with the most gaming experience scored the best results in all these metrics and even the fourth metric (walk distance) was on the limit of the significance threshold.

As gaming and VR experience seem to be a quite relevant in regards to the performance of the user, it might be useful to separate experienced gamers and non-gamers into different groups when conducting this kind of a study. Comparing if male and female gamers have any differences in their performance, would also make for an interesting subject. Unfortunately, in this study there were several males but only one female who reported playing games regularly. So, the comparison could not be done between genders in this area. It is also worth noting that the larger number of gamers in the male group is very likely to have skewed the results at least partly in favour of the males.

6.4 Heart rate trends

For the heart rate measurements, the point of interest was whether there were any statistically significant trends within a certain group and how the groups compared to each other. The initial hypothesis was that the two groups with the guiding light installations would produce downward trending heart rates while the control group without the lights would either trend upward or show now trends.

The results for the control group were quite interesting with the participants having an equal amount of upward- and downward trending heart rates (**46%** of participants each). It was expected that one of the trends or even no trend would be the majority, so no conclusions can be drawn from this result. The guiding light installations performed a bit differently with the downward trend being the majority in both cases. The results for the floor light group were fairly similar to the control group with **56%** of participants exhibiting a significant downward trend, **22%** an upward trend and **22%** showing no significant trend. So again, an inconclusive result for this group. The wall light group however, showed the most consistent trend out of all the groups. A statistically significant downward trend was detected for **75%** of the participants while **13%** of the participants showed an upward trend. **13%** of the participants exhibited no trends.

7 CONCLUDING REMARKS

The major focus of this thesis was to investigate the effectiveness and differences of wall-and floor-mounted guiding light systems with the use of a VR simulation. The first part of the thesis went through the history of VR equipment and its applications while also taking a look into a few different simulations and studies that are similar to the research presented here. Naturally this thesis used these previous studies and research as a starting point. The implementation section explained how the main parts of the simulation were built, how the VR equipment was used, what kind of performance metrics were used and how the simulation worked overall. The next chapter explained how the user testing was conducted and what was measured outside of the actual simulation like the questionnaire and the heart rate measurements.

The results part showcased all the relevant behavioural data, performance metrics and questionnaire results. The results were analysed according to the hypothesis that the guiding light systems would yield better results than the baseline condition that used no lights. In most tests the results did indeed indicate that the guiding light installations performed better. Unfortunately, in the majority of the performance results, the differences to the baseline simulation case were deemed statistically insignificant. The one area that showed statistical significance, was when comparing non-gamer and gamers. The gamers performed significantly better across the board, only falling a little short of the significance value in the walking distance metric.

Regarding heart rate measurements, the results were inconclusive. In each test group both upward- and downward trends were detected, making any concrete conclusions difficult. The group that had wall-mounted assistive lights performed closest to the expectations with **75%** of the participants showing a downward trending heart rate. But due to the small size of the test group and the fact that the other **25%** was split equally into upward trend and no trend, it is difficult to declare the **75%** result significant.

Even though many of the results ended up being inconclusive, the fact that gaming experience and frequency was shown to have a significant impact on the result, is important for further studies. None of the other groups showed anything close to the significance detected in the gamer/non-gamer comparison. Thus, in future VR simulation studies, the participants' gaming experience should be a key issue as well as a major factor when forming the groups for the study.

8 REFERENCES

[1] Farra, S. L., Gneuhs M., Hodgson E., Kawosa B., Miller E. T., Simon A., Timm N., Hausfield J. (2019). Comparative Cost of Virtual Reality Training and Live Exercises for Training Hospital Workers for Evacuation. https://doi.org/10.1097/CIN.0000000000000540

[2] Cosma, G., Ronchi, E., Nilsson, D. (2016). Way-finding lighting systems for rail tunnel evacuation: A virtual reality experiment with Oculus Rift ®. Doi: https://doi.org/10.1080/19439962.2015.1046621

[3] Vilar E., Rebelo F., Noriega. (2012). Indoor Human Wayfinding Performance Using Vertical and Horizontal Signage in Virtual Reality. Doi: https://doi.org/10.1002/hfm.20503

[4] Meng, F., Zhang, W. (2014). Way-finding During a Fire Emergency: An Experimental Study in a Virtual Environment. Doi: https://doi.org/10.1080/00140139.2014.904006.

[5] Cho, D., Ham, J., Oh, J., Park, J., Kim, S., Lee, N.-K., & Lee, B. (2017). Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine. *Sensors, 17*(10), Article 2435. https://doi.org/10.3390/s17102435

[6] Wright M., S., Cook G., K., Webber G., M., B. (2001). The effects of smoke on people's walking speed using overhead lighting and wayguidance provision. URL: https://www.reading.ac.uk/kqFINCH/nhe/research/eewp/bstnspeedfin.pdf. Accessed 06.02.2021

[7] Ronchi E., Nilsson D., Kojić K., Eriksson J., Lovreglio R., Modig H., Walter A. (2015). A virtual reality experiment on flashing lights at emergency exit portals for road tunnel evacuation. Doi: http://dx.doi.org/10.1007/s10694-015-0462-5

[8] Rey, B., Parkhutik, V., and Alcañiz, M. (2011). Breaks in presence in virtual environments: an analysis of blood flow velocity responses. *Presence 20*, 273–286. 10.1162/PRES_a_00049. Doi: https://doi.org/10.1162/pres_a_00049

[9] Hidayetoglu M., Yildirim K., Akalin A. (2011). The effect of color and light on indoor wayfinding and the evaluation of the perceived environment. Doi https://doi.org/10.1016/j.jenvp.2011.09.001

[10] IJsselsteijn, W. A., De Kort, Y. A. W., and Poels, K. (2013). The Game Experience Questionnaire. Eindhoven: Technische Universiteit Eindhoven 3–9.

[11] Stereopsis. URL: https://en.wikipedia.org/wiki/Stereopsis. Accessed 06.02.2021

[12] Barnard, D. (2019, August 6). History of VR – Timeline of events and tech development. URL: https://virtualspeech.com/blog/history-of-vr. Accessed 06.02.2021

[13] The Link Flight Trainer – A Historic Mechanical Engineering Landmark. (2000, June 10). URL: https://www.asme.org/wwwasmeorg/media/resourcefiles/aboutasme/who%20we%20are/engineering%20history/landmarks/210-link-c-3-flight-trainer.pdf. Accessed 06.02.2021

[14] Link Trainer. URL: https://en.wikipedia.org/wiki/Link_Trainer. Accessed 06.02.2021

[15] History of virtual reality. URL https://www.vrs.org.uk/virtual-reality/history.html. Accessed 06.02.2021

[16] Morton Heilig. URL: https://en.wikipedia.org/wiki/Morton_Heilig. Accessed 06.02.2021

[17] The History of Virtual Reality: An Escape to Different Worlds. (2018, May 7). URL: https://www.colocationamerica.com/blog/history-of-virtual-reality. Accessed 06.02.2021

[18] Steinicke, F. (2016). Being Really Virtual. Immersive Natives and the Future of Virtual Reality. Doi: https://doi.org/10.1007/978-3-319-43078-2

[19] Sutherland, I. (1965). The Ultimate Display. URL (cached article link): http://citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.3720. Accessed 06.02.021

[20] Daniel J. Sandlin Invents the Sayre Glove. URL: http://historyofinformation.com/detail.php?entryid=4080. Accessed 06.02.2021

[21] Wired Glove. URL: https://en.wikipedia.org/wiki/Wired_glove. Accessed 06.02.2021

[22] Power Glove. URL: https://en.wikipedia.org/wiki/Power_Glove. Accessed 06.02.2021

[23] Furness A., T. (1986). The Super Cockpit and its Human Factors Challenges. Doi: https://doi.org/10.1177/154193128603000112

[24] Lowood, H. Education and Training. URL: https://www.britannica.com/technology/virtual-reality/Entertainment. Accessed 06.02.2021

[25] The Virtual Interface Environment Workstation. URL: https://www.nasa.gov/ames/spinoff/new_continent_of_ideas/. Accessed 06.02.2021

[26] Virtual Reality. URL: https://psu.pb.unizin.org/ist110/chapter/6-3-virtual-reality/. Accessed 06.02.2021

[27] Cruz-Neira, C., Sandin, D., DeFanti, T. (1992). Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. Doi: https://doi.org/10.1145/166117.166134

[28] CAVE automatic virtual environment. URL: https://en.wikipedia.org/wiki/Cave_automatic_virtual_environment. Accessed 06.02.2021

[29] Virtuality (product). URL: https://en.wikipedia.org/wiki/Virtuality_(product). Accessed 06.02.2021

[30] Virtuality SD-model photograph. URL: https://en.wikipedia.org/wiki/Virtuality_(product)#/media/File:Ford_Galaxy_Launch_by_Imagination_1995.jpg. Accessed 09.02.2021

[31] Sega VR. URL: https://en.wikipedia.org/wiki/Sega_VR. Accessed 06.02.2021

[32] Virtual Boy. URL: https://en.wikipedia.org/wiki/Virtual_Boy. Accessed 06.02.2021

[33] 1995 Virtual IO I-Glasses. URL: https://vrtifacts.com/1995-virtual-io-i-glasses/. Accessed 06.02.2021

[34] Rothbaum, B., Hodges, L., Alarcon, R., Ready, D., Shahar, F., Graap, K., Pair, J., Hebert, P., Gotz, D., Wills, B., Baltzell, D. (1999). Virtual Reality Exposure Therapy for PTSDVietnam Veterans: A Case Study. Doi: https://doi.org/10.1023/A%3A1024772308758

[35] Rothbaum, B., Hodges, L., Alarcon, R., Ready, D., Shahar, F., Graap, K., Pair, J., Hebert, P., Gotz, D., Wills, B., Baltzell, D. (1998). Virtual Vietnam: A Virtual Environment for the Treatment of Vietnam War Veterans with Post-traumatic Stress Disorder. URL: http://gotz.web.unc.edu/files/2013/10/icat.pdf. Accessed 06.02.2021

[36] Google Street View. URL: https://en.wikipedia.org/wiki/Google_Street_View. Accessed 06.02.2021

[37] Palmer Luckey. URL: https://en.wikipedia.org/wiki/Palmer_Luckey. Accessed 06.02.2021

[38] Oculus Rift. URL: https://en.wikipedia.org/wiki/Oculus_Rift. Accessed 06.02.2021

[39] Oculus (brand). URL: https://en.wikipedia.org/wiki/Oculus_(brand). Accessed 06.02.2021

[40] Oculus Rift CV1. URL: https://en.wikipedia.org/wiki/Oculus_Rift_CV1. Accessed 06.02.2021

[41] Oculus Go. URL: https://en.wikipedia.org/wiki/Oculus_Go. Accessed 06.02.2021

[42] Oculus Quest. URL: https://en.wikipedia.org/wiki/Oculus_Quest. Accessed 06.02.2021

[43] Oculus Rift S. URL: https://en.wikipedia.org/wiki/Oculus_Rift_S. Accessed 06.02.2021

[44] Alexander Stolar (2012), Live CBRN agent training for responders as a key role in a safe crisis recovery. *NATO Science for Peace and Security Series - E - Human and Societal Dynamics Volume 94* (2012). Doi: https://doi.org/10.3233/978-1-61499-039-0-58

[45] Bernardes, S., Rebelo, F., Vilar, E., Noriega, P., Borges, T. (2015). Methodological approaches for use virtual reality to develop emergency evacuation simulations for training, in emergency situations. *Procedia Manufacturing 3* (2015) 6313 – 6320. Doi: https://doi.org/10.1016/j.promfg.2015.07.946

[46] Kinateder M., Ronchi E., Nilsson D., Kobes M., Müller M., Paul P., Mühlberger A. (2014). Virtual Reality for Fire Evacuation Research. Doi: https://doi.org/10.13140/2.1.3380.9284

[47] Ruddle, R. A. & Lessels, S. (2006). Three Levels of Metric for Evaluating Wayfinding. *Presence: Virtual and Augmented Reality, 15(6),* 637-654. https://doi.org/10.1162/pres.15.6.637

[48] Unity Scripting API. URL: https://docs.unity3d.com/ScriptReference/. Accessed 06.02.2021

[49] Time.time. URL: https://docs.unity3d.com/ScriptReference/Time-time.html. Accessed 06.02.2021

[50] Vector3.distance. URL: https://docs.unity3d.com/ScriptReference/Vector3.Distance.html. Accessed 06.02.2021

[51] Vector3. URL: https://docs.unity3d.com/ScriptReference/Vector3.html. Accessed 06.02.2021

[52] Transform.eulerAngles. URL: https://docs.unity3d.com/ScriptReference/Transform-eulerAngles.html. Accessed 06.02.2021

[53] Alavesa, P., Pakanen, M., Niemelä, A., Huang, W., Väinämö, S., Haukipuro, L., Arhippainen, L., Ojala, T. (2018). Mobile Augmented Reality Client as a UX Method for Living Lab's User Involvement Tool. *In Proceedings of the 22nd International Academic Mindtrek Conference (Mindtrek '18)*. ACM, New York, NY, USA, 135-142. Doi: https://doi.org/10.1145/3275116.3275147

[54] Florea, C., Alavesa, P., Arhippainen, L., Pouke, M., Huang, W., Haukipuro, L., Väinämö, S., Niemelä, A., Cortés Orduña, M., Pakanen, M., & Ojala, T. (2019). Extending a User Involvement Tool with Virtual and Augmented Reality. *IEEE*

Virtual Reality, IEEE VR, Osaka, Japan. Doi: https://doi.org/10.1109/VR.2019.8798299

[55] Kostakos, P., Alavesa, P., Oppenlaender, J., Hosio, S. (2019). VR ethnography: a pilot study on the use of virtual reality "go-along" interviews in Google street view. *In Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia (MUM '19)*. Association for Computing Machinery, New York, NY, USA, Article 53, 1–5. Doi: https://doi.org/10.1145/3365610.3368422

[56] Pakanen, M., Lovén, L., Alavesa, P., Gilman, E., Terken, J., Eggen, B., & Pirttikangas, S. (2018). Design challenges of wellbeing supporting smart environment in collaborative use situations. *In UbiComp/ISWC 2018 - Adjunct Proceedings of the 2018 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2018 ACM International Symposium on Wearable Computers* (pp. 688-692). Association for Computing Machinery, Inc. Doi: https://doi.org/10.1145/3267305.3267691

[57] Trochim, W. (2020, March 10). Nonprobability Sampling. URL: https://conjointly.com/kb/nonprobability-sampling/. Accessed 06.02.2021

[58] Convenience Sampling. URL: https://en.wikipedia.org/wiki/Convenience_sampling. Accessed 06.02.2021

[59] What is Optical Heart Rate Tracking. (2020, September 4). URL: https://www.polar.com/blog/optical-heart-rate-tracking-polar/. Accessed 06.02.2021

[60] Bai, Y., Welk, G. J., Nam, Y. H., Lee, J. A., Lee, J. M., Kim, Y., et al. (2016). Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc 2016;48(1):151-158. Doi: https://doi.org/10.1249/mss.0000000000000727

[61] Brooke, S. M., An, H. S., Kang, S. K., Noble, J. M., Berg, K. E., Lee, J. M. (2017). Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 Apr, 31(4):1097-1106. Doi: https://doi.org/10.1519/jsc.0000000000001571

[62] Fokkema, T., Kooiman, T. J., Krijnen, W. P., van der Schans, C. P., de Groot, M. (2017). Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017 Apr, 49(4):793-800. Doi: https://doi.org/10.1249/mss.0000000000001146

[63] Giles, D., Draper, N., Neil, W. (2016). Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. Eur J Appl Physiol 2016 Mar;116(3):563-571. https://doi.org/10.1007/s00421-015-3303-9

[64] Hernández-Vicente, A., Santos-Lozano, A., De Cocker, K, Garatachea, N. (2016). Validation study of Polar V800 accelerometer. Ann Transl Med 2016 Aug;4(15):278. Doi: https://doi.org/10.21037/atm.2016.07.16

[65] Mercer, K., Li, M., Giangregorio, L., Burns, C., Grindrod, K. (2016). Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016;4(2):e40. Doi: https://doi.org/10.2196/mhealth.4461

[66] Müller, A. M., Wang, N. X., Yao, J., Tan, C. S., Low, I., Lim, N., Tan, J., Tan, A., & Müller-Riemenschneider, F. (2019). Heart Rate Measures From Wrist-Worn Activity Trackers in a Laboratory and Free-Living Setting: Validation Study. JMIR mHealth and uHealth, 7(10), e14120. Doi: https://doi.org/10.2196/14120

[67] One-way ANOVA in SPSS Statistics. URL: https://statistics.laerd.com/spss-tutorials/one-way-anova-using-spss-statistics.php. Accessed 06.02.2021

[68] Hindle, A. (2013, April 16). Research Point: ANOVA Explaind. URL: https://en-author-services.edanz.com/blogs/statistics-anova-explained. Accessed 06.02.2021

[69] Geert van den Berg, R. ANOVA – Super Simple Introduction. URL: https://www.spss-tutorials.com/anova-what-is-it/. Accessed 06.02.2021

[70] F Statistic/F Value: Simple Definition and Interpretation. URL: https://www.statisticshowto.com/probability-and-statistics/f-statistic-value-test/#Fdist. Accessed 06.02.2021

[71] ANOVA Test: Definition, Types, Examples. URL: https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/anova/. Accessed 06.02.2021

[72] Kruskal-Wallis H Test using SPSS Statistics. URL: https://statistics.laerd.com/spss-tutorials/kruskal-wallis-h-test-using-spss-statistics.php. Accessed 06.02.2021

[73] Zaiontz, C. Kruskal-Wallis Test. URL: https://www.real-statistics.com/one-way-analysis-of-variance-anova/kruskal-wallis-test/. Accessed 06.02.2021

[74] Kruskal-Wallis H Test: Definition, Examples & Assumptions. URL: https://www.statisticshowto.com/kruskal-wallis/. Accessed 06.02.2021

[75] Kruskal-Wallis one-way analysis of variance. URL: https://en.wikipedia.org/wiki/Kruskal%E2%80%93Wallis_one-way_analysis_of_variance. Accessed 06.02.2021

[76] Forthofer, R., Lee, E., Hernandez, M., (2007). Kruskal-Wallis Test. URL: https://www.sciencedirect.com/topics/mathematics/kruskal-wallis-test. Accessed 06.02.2021

[77] Mann-Whitney U Test Calculator. URL: https://www.socscistatistics.com/tests/mannwhitney/. Accessed 06.02.2021

[78] Mann-Whitney U Test. URL: https://www.statisticshowto.com/mann-whitney-u-test/. Accessed 06.02.2021

[79] Mann-Whitney U Test (Wilcoxon rank sum test). URL: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_nonparametric/bs704_nonparametric4.html. Accessed 06.02.2021

[80] Mann-Whitney U Test. URL: https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test. Accessed 06.02.2021

[81] Mann Whitney Test – Simple Example. URL: https://www.spss-tutorials.com/spss-mann-whitney-test-simple-example/. Accessed 06.02.2021

[82] Pohlert, T. (2020, September 17). Non-Parametric Trend Tests and Change-Point Detection. URL: https://cran.r-project.org/web/packages/trend/vignettes/trend.pdf. Accessed 06.02.2021

[83] Mann Kendall Trend Test: Definition, Running the Test. URL: https://www.statisticshowto.com/mann-kendall-trend-test/. Accessed 06.02.2021

[84] Wang, F., Shao, W., Yu, H., Kan, G., He, X., Zhang, D., Ren, M., Wang, G. (2020, February 6). Re-evaluation of the Power of the Mann-Kendall Test of Detecting Monotonic Trends in Hydrometeorological Time Series. URL: https://www.frontiersin.org/articles/10.3389/feart.2020.00014/full. Accessed 06.02.2012

[85] Zaiontz, C. Mann-Kendall Test. URL: https://www.real-statistics.com/time-series-analysis/time-series-miscellaneous/mann-kendall-test/. Accessed 06.02.2021

[86] Zaintz, C. Sen’s Slope. URL: https://www.real-statistics.com/time-series-analysis/time-series-miscellaneous/sens-slope/. Accessed 06.02.2021

[87] Sen’s Slope/Theill-Sen estimator. URL: https://www.statisticshowto.com/sens-slope-theil-sen-estimator/. Accessed 06.02.2021

[88] Appendix C: Trend Analysis Tests. URL: https://www.waterloohydrogeologic.com/help/aquachem/aq4_manual_ap_a4.htm. Accessed 06.02.2021

9 APPENDICES

Appendix 1	Participant information and pre-study questionnaire
Appendix 2	The Game Experience Questionnaire. Core-, in-game- and post-game modules

Appendix 1 Participant information and pre-study questionnaire

1. Age:
2. Height:
3. Weight:
4. Gender: Female Male
5. Are you: Student University Employee Visitor
6. What is your major or profession?
7. Have you done military service, is so in which division?

8. Have you participated in voluntary rescue or firefighting activities?

9. Do you have a driver's license? Yes No
10. Do you own a car? Yes No

11. How often do you park to an underground parking space?
    Daily Weekly Monthly Seldom/Never

12. How often do you play video games?
    Daily Weekly Monthly Seldom/Never

13. How often you use virtual reality equipment?
    Daily Weekly Monthly Seldom/Never

Please indicate for each of the items, on the following scale:

not at all	slightly	moderately	fairly	extremely
0	1	2	3	4
< >	< >	< >	< >	< >

I am very good at giving directions.	0	1	2	3	4
I think it is important to find new routes in the environment.	0	1	2	3	4
I have a poor memory for where I left things.	0	1	2	3	4
I like to travel.	0	1	2	3	4
I am very good at judging distances.	0	1	2	3	4
My ''sense of direction'' is very poor	0	1	2	3	4

Appendix 2 The Game Experience Questionnaire. Core-, in-game- and post-game modules

Please indicate how you felt while playing the game for each of the items, on the following scale:

not at all	slightly	moderately	fairly	extremely
0	1	2	3	4
< >	< >	< >	< >	< >

1	I was interested in the game's story	0	1	2	3	4
2	I felt successful	0	1	2	3	4
3	I felt bored	0	1	2	3	4
4	I found it impressive	0	1	2	3	4
5	I forgot everything around me	0	1	2	3	4
6	I felt frustrated	0	1	2	3	4
7	I found it tiresome	0	1	2	3	4
8	I felt irritable	0	1	2	3	4
9	I felt skilful	0	1	2	3	4
10	I felt completely absorbed	0	1	2	3	4
11	I felt content	0	1	2	3	4
12	I felt challenged	0	1	2	3	4
13	I had to put a lot of effort into it	0	1	2	3	4
14	I felt good	0	1	2	3	4

Please indicate how you felt while playing the game for each of the items, on the following scale:

not at all	slightly	moderately	fairly	extremely
0	1	2	3	4
< >	< >	< >	< >	< >

1	I felt content	0	1	2	3	4
2	I felt skilful	0	1	2	3	4
3	I was interested in the game's story	0	1	2	3	4
4	I thought it was fun	0	1	2	3	4
5	I was fully occupied with the game	0	1	2	3	4
6	I felt happy	0	1	2	3	4
7	It gave me a bad mood	0	1	2	3	4
8	I thought about other things	0	1	2	3	4
9	I found it tiresome	0	1	2	3	4
10	I felt competent	0	1	2	3	4
11	I thought it was hard	0	1	2	3	4
12	It was aesthetically pleasing	0	1	2	3	4
13	I forgot everything around me	0	1	2	3	4
14	I felt good	0	1	2	3	4
15	I was good at it	0	1	2	3	4
16	I felt bored	0	1	2	3	4
17	I felt successful	0	1	2	3	4
18	I felt imaginative	0	1	2	3	4
19	I felt that I could explore things	0	1	2	3	4
20	I enjoyed it	0	1	2	3	4
21	I was fast at reaching the game's targets	0	1	2	3	4
22	I felt annoyed	0	1	2	3	4
23	I felt pressured	0	1	2	3	4
24	I felt irritable	0	1	2	3	4
25	I lost track of time	0	1	2	3	4

26	I felt challenged	0	1	2	3	4
27	I found it impressive	0	1	2	3	4
28	I was deeply concentrated in the game	0	1	2	3	4
29	I felt frustrated	0	1	2	3	4
30	It felt like a rich experience	0	1	2	3	4
31	I lost connection with the outside world	0	1	2	3	4
32	I felt time pressure	0	1	2	3	4
33	I had to put a lot of effort into it	0	1	2	3	4

Please indicate how you felt after you finished playing the game for each of the items, on the following scale:

not at all	slightly	moderately	fairly	extremely
0	1	2	3	4
< >	< >	< >	< >	< >

1	I felt revived	0	1	2	3	4
2	I felt bad	0	1	2	3	4
3	I found it hard to get back to reality	0	1	2	3	4
4	I felt guilty	0	1	2	3	4
5	It felt like a victory	0	1	2	3	4
6	I found it a waste of time	0	1	2	3	4
7	I felt energized	0	1	2	3	4
8	I felt satisfied	0	1	2	3	4
9	I felt disoriented	0	1	2	3	4
10	I felt exhausted	0	1	2	3	4
11	I felt that I could have done more useful things	0	1	2	3	4
12	I felt powerful	0	1	2	3	4
13	I felt weary	0	1	2	3	4
14	I felt regret	0	1	2	3	4
15	I felt ashamed	0	1	2	3	4
16	I felt proud	0	1	2	3	4
17	I had a sense that I had returned from a journey	0	1	2	3	4