Evaluating the Impact of Network Delay on user Quality of Experience of an Interactive Virtual Reality Industry 4.0 Application

David Concannon, BSc

Thesis submitted for Master of Science (MSc.) by Research

Department of Computer and Software Engineering, Athlone Institute of Technology

Supervised by: Dr. Niall Murray and Dr. Ronan Flynn

Submitted to Athlone Institute of Technology, September 2020.

ABSTRACT

The fourth industrial revolution is the name given to the paradigm shift towards further integrating the relationship between industrial technologies and the world of computing. The Industry 4.0 ecosystem includes a range of technologies such as the Internet of Things, robotics & automation, big data, machine learning and human-machine interaction. Industry 4.0 demands the development of new fundamental concepts, such as smart manufacturing, to improve efficiency and productivity in the industrial sector. Technologies such as virtual reality (VR) enable human-in-the-loop roles in Industry 4.0. Building human-centred interactive systems will enhance efficiency, human well-being and satisfaction, while at the same time counteracting potential health, safety and performance hazards. Critical to the success of such human-in-the-loop systems is the understanding of the user perceived quality of experience (QoE).

QoE is a discipline concerned with understanding the usability and utility of a technology in a particular context or environment. The focus of this MSc research is to understand the impact of network delay on a user's QoE whilst interacting with a VR application. In the environment, the user interacts with a virtual representation of a Fanuc injection moulding machine The Industry 4.0 aspect of interest for this research is teleoperation, which is the real-time control of a mechanical unit over a network. A user could, in theory, control a Fanuc across the Internet using a VR environment that is designed to replicate the operation of such a machine.

To evaluate the QoE of remotely operating a Fanuc, participants in this research carried out a basic, beginner-level operation task on the Fanuc in a virtual reality environment under both subjective and objective evaluation. The participants were provided with instructions in the VR environment, introducing the participants to the machine and assisting them via instructions as they completed the task. To understand the impact of network delay, one group experienced no network delay, while all other subject groups experienced artificially introduced network delays. Various objective implicit parameters were measured, such as time-to-task-completion, number of controller clicks and biometric measurements, such as heart rate, electro-dermal activity and eye gaze data. Subjective explicit metrics were captured post-test, using a questionnaire, to measure self-reported immersion and engagement with the environment. The results suggest that, although participants experience a mild drop in QoE as a result of network delay, they are tolerant of delays of up to 3000ms, with no significant deterioration in perceived usability of the virtual environment.

DECLARATION OF AUTHORSHIP

I hereby certify that this material, which I now submit for assessment on the program of study leading to the award of Master by Research program in the Department of Computer and Software Engineering at Athlone Institute of Technology is entirely my own work and has not been taken from the work of others and to the extent that such work has been cited and acknowledged within the text of my work.

I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this Institute.
- Where any part of this thesis has been previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the others, the source is always given.
- With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed: _____________________ Date: __________

ACKNOWLEDGMENTS

First and foremost, I would like to thank my Supervisors and Advisors, Dr. Niall Murray and Dr. Ronan Flynn at Athlone Institute of Technology. Without their guidance, words of encouragement and experience, I would never have been able to piece together this body of work. I thank them for all the opportunities they gave me from start to finish. I particularly wish to thank Dr. Murray for taking a chance at bringing me on as a self-funded student in the first place. Without this I would never have made it to where I am now, and I will be forever grateful.

I wish to thank Dr. Mark Daly and Dr. Jason Quinlan and all others who helped review this work. Their constructive comments and criticism were instrumental in upgrading the quality of this thesis to the best it could be.

I wish to thank all of my lab mates, of which there are almost too many to name, in the QoE research team in Athlone for all the help and pointers they have given, both on my work and the world of academic research. I also wish to thank them for being friends and people to talk to during the hardest, as well as best of times during both development, experimentation and writing for this project.

Many thanks to Christopher Hannevig who helped acquire funding part way through this thesis and allowing me to keep the ball rolling.

Lastly, I would like to thank my parents and family, as well as my ever-supportive girlfriend, all of whom lent their unconditional support for the last few years in making sure I would cross the finish line. Their help in supporting me, both physically and mentally, was critical in my completion of both researching and writing this thesis. This would never have been possible without them.

Thank you all.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

Section 1: INTRODUCTION

1.1 Background

Industry 4.0 is the term used to describe an evolution currently underway in the industrial sector as a result of the rapid pace at which technology has developed in the 21st century. First defined by the German government in 2013 [1], it has become a hot topic among those closely involved in areas such as manufacturing, agriculture, infrastructure and medicine. As implied by the term, industry has undergone three distinct evolutions prior to this one, primarily as a result of breakthrough technologies at the time, such as steam power and electricity [1]. These are expressed in this report as an initial industrial revolution in the 1800's with the innovation of steam powered machinery, a second revolution with the advent of electrically powered mass production, and a third in the form of introduction of computer guided automation. In the case of Industry 4.0, the driving technology behind the rapid pace of evolution is the advent of mechanical objects with cybernetic and network focused integration and computing capabilities. This has manifested in a concept known popularly as the Internet of Things (IoT), a framework describing a technology world where physical objects can communicate with each other through their cybernetic components [2]. One of the most promising technologies to come about as a result of innovations relating to Industry 4.0 are Cyber Physical Systems (CPS). Like the IoT, CPS refers to mechanical components that can communicate with each other over a network and form their own virtualized production chains. CPS involves the combination of many disciplines such as cybernetics, mechatronics, design and process science. In Industry 4.0, this is achieved through integrating CPS in the production process, forming automated self-regulating production lines and aggregating many of these elements together to form smart factories. Alongside the benefit of fully autonomous systems, CPS brings other benefits, such as modular capabilities, allowing machines from different systems to be integrated or removed according to needs [3] [4].

While CPS bring innovation to autonomous manufacturing, the fact remains that technology has not advanced yet to a point where fully human-independent systems are feasible. In addition to this, there is a fear that an overzealous approach to CPS integration could mean a loss of a source of labour for humans, leading to humans being outsourced in favour of more self-directing technology [5]. This has led to a push to ensure that human-centric designs for CPS are to the fore. While this has manifested itself in many ways, the focus in this body of research is on tele-operation [6] [7]. Teleoperation refers to the means by which a human may operate with a mechanical unit from a distance.

One of the challenges for the integration of teleoperation in CPS is the human interface. Traditional desktop interfaces, such as a PC, have established themselves as the norm. However, these modes of interaction do not take advantage of the rapidly evolving CPS technologies and trends. For example, training users on the operation of a mechanical unit through the use of an immersive virtual environment has been shown to have merit in terms of retention of knowledge and overall error rate [8]. As such, researchers have turned towards investigating more novel technologies through the medium of more immersive devices, such as augmented reality, using smartphones and tablets [5], and virtual reality [9]. At an everyday level, smartphones and tablets are now advanced enough to provide augmented reality experiences on the go. In addition, there has been a growth in more powerful specialised hardware that is specifically designed to provide these immersive experiences. Virtual Reality (VR) Head Mounted Displays (HMD), such as the HTC Vive [10], and augmented reality (AR) devices, such as the HoloLens [11], have gained traction across many domains as immersive devices. Network delay refers to the delay in response of a task or service over a network, more specifically to the time it takes for a packet of information to travel from one source to its intended destination, with delay manifesting when these packets fail to arrive or arrive late [12].

One of the major areas of research into developing the potential of these immersive multimedia devices is that of Quality of Experience (QoE). QoE as a body of research has gained recognition in industry and research communities as an effective way to measure a user’s expectations of a service or technology. Traditionally used in telecommunications to optimize user engagement with a service or product, QoE has expanded to other areas, such as education, medicine and tourism [13] [14] [15], as a way of measuring human experience. As user-centric requirements become more important in commercial products, QoE aims to maximize user engagement and quality perception [16]. QoE has traditionally been measured using subjective methods to capture human-centric perceptions by considering the influence factors of the technology, environmental context, and content. This would most commonly be achieved by using questionnaires during or after a test. Using the data gathered this way, QoE models are derived from a subjects Mean Opinion Score (MOS) [17], forming a framework whereby informative predictions of a user’s response to a technology or interface can be made. The study of QoE has also developed to include objective methods to measure user QoE; these include the monitoring of biological signals like heart rate, skin conductivity and eye tracking. In relation to network delay, QoE plays an important role as a user's subjective experience can be negatively impacted by delay and render a service or application unusable in their eyes, especially in the context of high activity contexts such as online gaming.

1.2 Research Question, Aim and Objectives

The focus of this MSc is the development of an application in a virtual reality environment with a view to understanding the impact that network delay has on a user's QoE. The application simulates the use of a real-world Fanuc injection moulding machine [10] suited for use in an Industry 4.0 context. The user QoE assessment uses wearable sensors to capture objective biometric data, along with a subjective QoE evaluation, to gain an insight into user perception of the usability and responsiveness of the environment.

The research question is:

In the context of operating a remote machine through a virtual reality environment, how does network delay impact a user's Quality of Experience of the application?

The aim of this research is to understand the effects that network delay has on a user's overall QoE and identify a delay threshold beyond which a user can consciously detect the network delay to the point of influencing reported QoE. To answer the research question and achieve the research aim, the following research objectives were defined:

- Research and critique current existing work for virtual reality industrial applications. This included the evaluation of existing studies into the impacts of network delay and an analysis of methods utilized within.
- Design and develop a virtual reality environment that allows a user to interact with a simulated machine, based on real-world specifications.
- Design a methodology that captures user QoE when applied to the developed virtual reality environment. This involves the integration of an eye tracking framework, VR interaction metrics and other biometric data capture.
- Perform a suite of user trials with participants to evaluate the QoE of the environment in the presence of varying degrees of simulated network delay.

1.3 Contributions of this work

This work adds to and expands upon the existing knowledge with respect to the evaluation of immersive virtual environments from a QoE standpoint. There exists research proposing frameworks and methodologies to evaluate QoE in immersive environments using physiological metrics alongside traditional QoE measurement methods. However, they do not consider such metrics in the context of network delay. In addition, this work also considers additional factors of measurement, such as interaction metrics via in-built trackers in the VR wands (user clicks) and tracking activations of the virtual elements within VR (button presses, time to completion of task). An eye tracking capture and display framework within the virtual environment was also developed to use captured eye-gaze as part of a more novel QoE evaluation. This was done to demonstrate eye-gaze as another possible vector to measure QoE with the aim of creating more robust and broader QoE evaluations.

1.4 Thesis Outline

The following gives a brief outline of the sections to follow.

Section 2: Literature Review

Section 2 gives an overview and critique of work related to the research presented. This involves examining the technologies and methods employed by other researchers in each area of interest to this study, including Industry 4.0, QoE and immersive multimedia. The literature review identifies gaps in the wider body of knowledge that the author used to define the research objectives for this work. Section 2 also informs the selection of the QoE metrics used and the development of the experiment design and methodology.

Section 3: Methodology and Experiment Design

The methodology and experiment design in Section 3 outlines the research method and technologies used for the work presented. This Section discusses the various design frameworks that constitute the testing methodology employed. Decisions and the rationale for the experiment framework are justified, including a discussion of third-party tools used. Existing QoE methodologies that this work employs are presented, along with the novel QoE developments, such as physiological measurements, that have been integrated into the experiments.

Section 4: Results Discussion and Analysis

Section 4 presents the experimental results with a detailed statistical analysis and discussion. Significant findings and trends identified in the experimental data are highlighted and interpreted. How the findings relate to the overall research objective of this work are discussed.

Section 5: Conclusion

Conclusions and potential directions for future work are outlined in Section 5.

Section 2: LITERATURE REVIEW

This Section provides an overview and critique of existing research in the areas of interest for this work. These include Industry 4.0, CPS and their relation to VR and AR, QoE and multimedia network delay.

2.1 Industry 4.0

2.1.1 Overview of Industry 4.0

Industry 4.0 is the term often used to describe advances in the areas of computing, industry and manufacturing, and how these technologies integrate to form the fourth industrial revolution. Introduced as a concept by the German federal government in 2013 [1], the term derives from the concept that industry has experienced three major revolutions in technology before this point. The factors that have influenced previous industrial revolutions through advances in technology are presented in [1] and illustrated in Figure 2.1. The first and second came about as a result of discovery of technologies such as steam power (the first industrial revolution) and oil and electricity (the second industrial revolution), the third and fourth industrial revolutions have come about through rapid developments in the areas of information technology (IT) and computing [1].

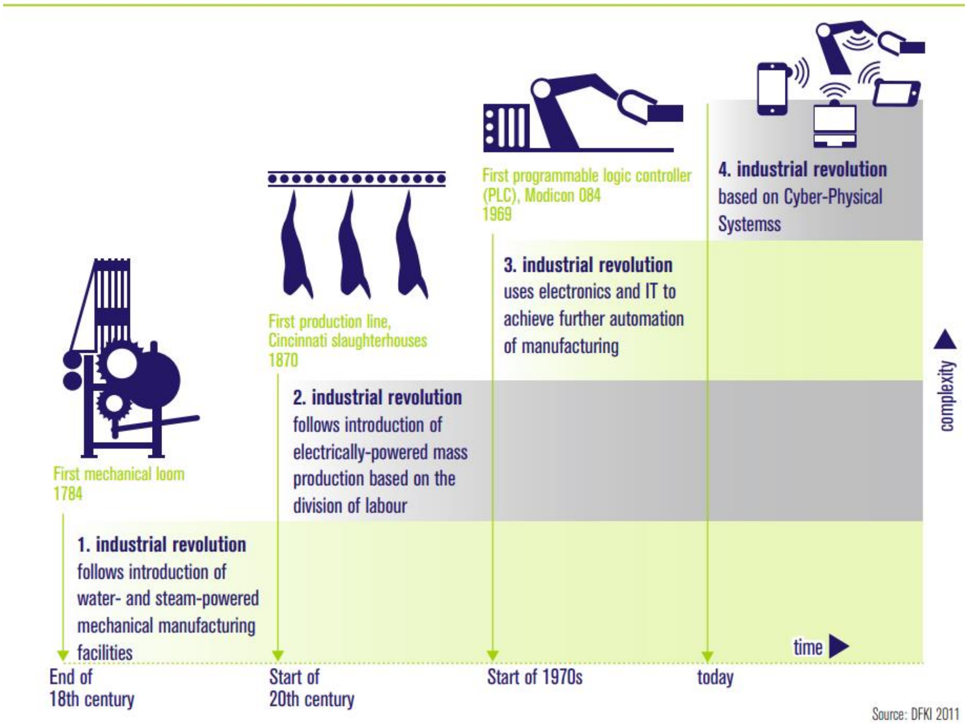


Figure 2.1: The four industrial revolutions in technology as defined in [1]

The fourth industrial revolution derives its origins from the advent of the Internet of Things. As Described by the ITU-T [2], "*the IoT is an infrastructure that enables advanced services by interconnecting physical and virtual things, based on existing and evolving interoperable information and communication technologies*". Through the exploitation of data capture and sharing, processing and communication, applications can interact to provide services not previously possible. Tangible objects, the things, exist in the physical world and are capable of being sensed and connected to others in its network. These things often have a virtual controller that facilities the communication and interaction with others in the environment. This infrastructure manifests itself in the context of Industry 4.0 as a technology known as Cyber Physical Systems (CPS).

2.1.2 Cyber Physical Systems

Cyber physical systems are essentially network-connected (either internally or via the Internet) physical mechanisms that are controlled by de-centralized computing nodes to form their own virtualized production chains, and that can make informed data driven decisions. CPS are described in [3] as:

> *"Systems where physical and software components are deeply intertwined, each operating on different special and temporal scales, exhibiting multiple and distinct behaviours, and interacting with each other in a myriad of ways that change with context."*

The essential characteristics of CPS elements include a) cybernetic capability in every component, b) a high degree of automation, c) networking at multiple scales, d) integration at multiple temporal and spatial scales, and e) a modular dynamic capable of reorganization and reconfiguration.

CPS manifest themselves across many emergent technologies and examples of its implementations and uses can be seen in the field of robotics, autonomous manufacturing systems, assistive technologies and smart-environment buildings [18] [19] . CPS are expected to play a fundamental role in all facets of industry and therefore an emphasis has been placed on research into methods of integration and development in the area [4]. One of the important challenges set out for communities engaging in CPS research is the discovery of robust and standardised design techniques to ensure that the technology is optimised across all contexts during its integration into modern society [18] [20]. In [21], the author proposed a high level architecture (Figure 2.2) for CPS integration and provided an overview of how CPS designs can be tailored for industry. As seen in Figure 2.2, CPS implementations are divided into a five-level hierarchy, with each level representing the factors necessary for creation of a functional CPS. This serves to offer a more constructive framework for integration of a CPS to a real-world context, as opposed to the more abstract concept of a virtual-physical machine relationship [21].

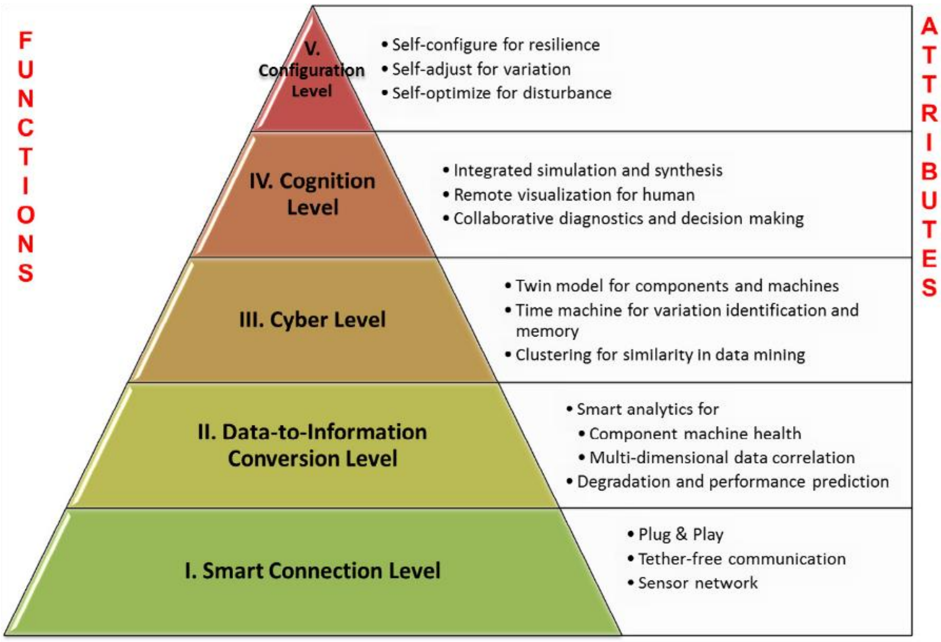


Figure 2.2: Levels of CPS Integration [21]

The concept of CPS is expanded upon in [5] to explain that CPS will require fundamental integration with human modalities as the technology expands to accommodate the new growth in industry, with a focus on the need to keep jobs and labour as automation becomes the norm. The work presented in [5] also explores the potential applications of novel technologies such as augmented and virtual reality.

A paradigm that has been proposed for human interaction with CPS is the concept of teleoperation. Teleoperation refers to the ability to control a mechanical system remotely. It is most commonly associated with robotics, particularly for military use, but the concept has been applied to other areas like medicine (robotic surgery) and industry (remote operation of heavy machinery). Users interacting with digital analogies of a real-world system via teleoperation reduces risk of injury by separating the tangible moving parts and the human from each other. It can allow for time saving by facilitating on-the-fly operation from a PC, as well as considerations for more naturalistic and intuitive interface designs. Teleoperation is an important concept in enabling human interaction with CPS due to their highly networked, decentralized nature. Allowing humans to interface with a cyber physical system would open more degrees of control to a system at the various layers of CPS integration. Research into different forms of interaction with mechanical systems has been performed in the past to demonstrate the benefits and restrictions of teleoperation [9] [22] [23].

One such architecture that allows for virtual reality teleoperation of an Industry 4.0-designed mechanism is the digital twin architecture [24]. The digital twin architecture involves development of a digital model of some tangible entity in a virtual environment, then bridging the gap between the two by a network connection, allowing for real-time exchange of information between the twins. This information can then be used to modify either entity, with changes in one being reflected in the other. This can be seen in successful real-world implementations, most notably with applications such as Google Earth [25]. As an industry-focused VR application, a digital representation of a real-world manufacturing unit can be created in a virtual environment, then connected to the real unit over a network. From this a user may interact with the virtual twin and any operations carried out within the virtual environment can be reflected in the real-world twin.

2.2 Immersive Technology

2.2.1 Multimedia and VR Technologies

As technology has advanced, users are engaged in a world where one can express themselves in a large variety of ways through the prism of multimedia. Smartphones have allowed users to interact with others and technology through text, video, sound, and images. Such interactions often require these handheld devices to capture, send and process large amounts of data. Virtual environments are those that allow for the visualization of cybernetic or analytic data in a way that allows humans to easily understand and interact with what is presented. Examples include video gaming, 3D modelling and data and network visualization software. Highly immersive devices that allow for different levels of interaction with media components in virtual environments have been developed. Devices such as virtual reality head mounted displays (VR HMDs) allow media to be experienced in a much deeper way, through immersive video and sound, and in some cases with haptic and olfactory elements [26].

Commercially, HMDs have become more widely available than before, with the average individual having greater access to immersive multimedia in their own homes, whereas before these would normally have been relegated to specialist use cases. Devices such as the Vive [10], Oculus Rift [27], and PlayStation VR [28] offer these experiences at a consumer level with minimal requirements. In addition, the implementation of controller interaction allows for an even deeper level of experience. Virtual environments are generally realized through computer generated graphical imagery displayed through an LED screen implemented in the HMD and allow a full 3D-presence experience for the user. The realism of these environments can be augmented and improved by simulating more natural forms of interaction via various external technologies, such as haptic gloves.

Though generally considered to be limited to the area of entertainment, significant work has been put into investigating the potential of VR for use in other sectors such as medicine [29], tourism [30] and industry [31].

2.2.2 Virtual Environment Development and Design

There are several tools available to the average consumer for virtual reality development. Typically, current state of the art VR environments are constructed using game engines designed for video game development. A game engine is a software development environment for people to develop video games for computers, consoles, or mobile devices. The core functionality provided by these engines are image rendering, physics calculations, sound generation, scripting, animation, artificial intelligence, and networking. There exist several high-fidelity engines that end users have access to for free, such as Unity [32], Unreal [33] and Godot [34].

Immersive VR environments are different from regular virtual environments as they potentially can completely immerse the user in the environment through the use of a HMD and other external olfactory, haptic or audio apparatus. Development of an immersive VR environment requires an understanding of unique factors and issues that come with the design of a full body immersive experience. These include issues such as visual fidelity, interaction methods [35], navigation [36] and safety issues such as tripping hazards and simulator sickness [37]. Researchers aiming to investigate the uses and applications of VR must be careful to keep these factors in mind when developing experiments for immersive VR environments. Such factors can be a large contributor of external biases and influences, especially in the case where the research pertains to areas such as quality of experience.

2.2.3 VR and Industry 4.0

Currently virtual reality in Industry 4.0 is mainly centred around the areas of simulation and virtual visualization. Application for these styles of virtual environments include visual manuals or operator guides, or digital twin style feedback systems. Over time, virtual visualisation has become more complex to match the growing industry infrastructure and technology as CPS networks and autonomous manufacturing lines develop. Most current state-of-the art implementations use less immersive methods of virtual simulation through the use of terminal-based displays like the desktop PC. As Industry 4.0. grows and the technologies involved improve, there is a need to modernize the interaction methods of humans with this environment [5].

Within Industry 4.0, heavy emphasis has been placed on the importance of robust human-machine interactive systems [38]. While the efficiency and autonomy of computer embedded mechanical technologies and CPS improve yearly, there is still a gap in the ability for these systems to be truly autonomous, self-maintaining and self-regulating. For these reasons, integration of Industry 4.0 concepts still considers the human element as crucial when designing these systems, from maintenance to high-level decision making [39]. Part of the challenge to these human-computer interactions is the facilitation of a natural interface, as CPS are often systems of high complexity with large volumes of data and information.

One of the proposed solutions to this problem is the introduction of virtual and augmented reality applications to bridge the gap between the virtual elements of a manufacturing process and the human operator. The use of immersive multimedia interaction applications could theoretically provide an environment that would be perceived as more natural and immersive for their human operators when compared to a terminal or desktop-based interaction system, allowing for a more engaging experience. An example of this currently used in industry is SmartFactory [40], an application that allows users to manage and interact with industrial machinery using their smartphone. In addition, uses of immersive virtual reality for training purposes are currently being investigated. In [8], the authors examine how users who had been trained in assembly operation by traditional means perform against those who had been trained using an immersive virtual reality system. This work found that there was a significant improvement in initial operator performance among those who had undergone the virtual reality training, though there were some issues reported within the study. These drawbacks related to the increased learning load introduced with operators learning how to use the VR system on top of the system being examined. Additional issues related to the limitation of movement and the confinement of the training simulation to a screen-sized area resulting in missed details and unclear instruction steps.

Immersive interaction systems, including VR frameworks, have been shown to be an effective method of controlling mechanical apparatus [9] [41]. By developing this type of interaction for established industrial CPS, a user is more likely to experience heightened engagement with a complicated physical system [41]. Work in [9] also highlights the complexity of designing such human-computer interaction systems, and how care must be taken to ensure that the user's perceived sense of presence and engagement is at a level to perform the task at hand. One of the particularly promising areas for virtual reality in Industry 4.0 is that of teleoperation by marrying VR virtualization with direct tele-operative interfaces. Work has already gone into the development and research of effective teleoperation methods for industry use in more traditional forms [42], and there is interest in adapting immersive technologies for these areas.

2.3 Networking

The Internet has become an integral component of all aspects of human life in the 21st century. Naturally, this has led to the Internet becoming an intrinsic aspect of industry as it becomes more integrated at all levels of operation. The fundamental driving force behind the Internet of Things and CPS is the desire to integrate this close connectivity with the Internet into the very fundamentals of production and mechanisation.

2.3.1 Network Characteristics

For perceptive quality studies of network conditions, most fundamental measurement techniques fall within the subject area of quality of service (QoS), a discipline related to quality of experience with a greater focus on measurement of system factors rather than human factors [43]. QoS has been widely used for network, cloud, and VoIP applications to understand how poor network connections impact these services. Some of the characteristics of interest for network analysis are that of latency, jitter, and packet loss. Latency is the most fundamental aspect of measuring network quality and refers to the end-to-end delay experienced by network packets travelling from one network node to another [12]. The greater the congestion in a network the more delay packets will experience due to increased processing and queueing times at network nodes, resulting in a skewed or lagged experience for real-time applications. For example, a high amount of delay in a VoIP application may result in one user experiencing another’s conversation with significant gaps in the dialogue, which can significantly influence a user’s QoE and their engagement with the call. Jitter refers to the variability in packet delay over a network [12]. This delay variation is largely attributed to changing traffic or congestion patterns in a network. As such closed networks with few endpoints will largely experience little or no jitter. Packet loss refers to the failure of a packet to reach its destination or the arrival of a packet out of order, as it travels between a sender and receiver on a network [12]. Network packet loss is usually as a result of buffer overflow due to high congestion and traffic. Packet loss is a particularly important issue for more interactive multimedia applications such as online games, with a larger impact on QoE on these types of applications than more passive ones like video streaming. Overall network health and quality is generally measured via the rate of successful packet arrival across a network, referred to as throughput [12].

It is important to state that there can be multiple types of delay not related to network characteristics when considering immersive virtual reality environments. Though all use the term delay, network delay, display delay and input delay are all different factors with their own set of unique influences on QoE. For this reason, network delay is often referred to as latency to distinguish it from other delay factors [44].

2.3.2 Network Design and QoS/QoE

While any given networked application is subject to the impacts of latency, jitter and packet loss, the type of application in question is an important factor when performing human centric quality evaluations of the application in question. This human-application relationship is what fundamentally divides QoS and QoE, as QoS measurements can be used for many types of services by focusing solely on network and system characteristics, while QoE examinations must consider the context, environment and human centric variables , such as expectations or expertise, on top of the systemic variables of an application. As such, QoE examinations can show very different results for different applications on the same network. For example, a VoIP application would consider QoE factors such as sound fidelity, conversation synchronization and audio cutting, while a real-time gaming application would consider issues such as consistent updates on other players status' and the responsiveness of the online session. In [45], the authors investigated the levels of delay acceptable from a subjective perspective in an online gaming experience. In this study the delay boundary of 120ms was identified as a cut-off point for user-perceived acceptability. In [46], the authors discuss that in usability evaluations, users report that delays of up to 100ms are effectively unnoticeable and regarded as instant, while latencies of 1s are reported as noticeable in typical subjective user reported evaluations. In [47], the authors performed a QoE assessment on a haptic media balloon popping game (see Figure 2.3) using a networked virtual environment with latency showing that users reported very negative MOS at delays of 500ms. There are many influencing factors on user QoE in networked environments that can result in a wide array of results, and so when designing a networked virtual environment careful consideration of the factors being examined must be taken.

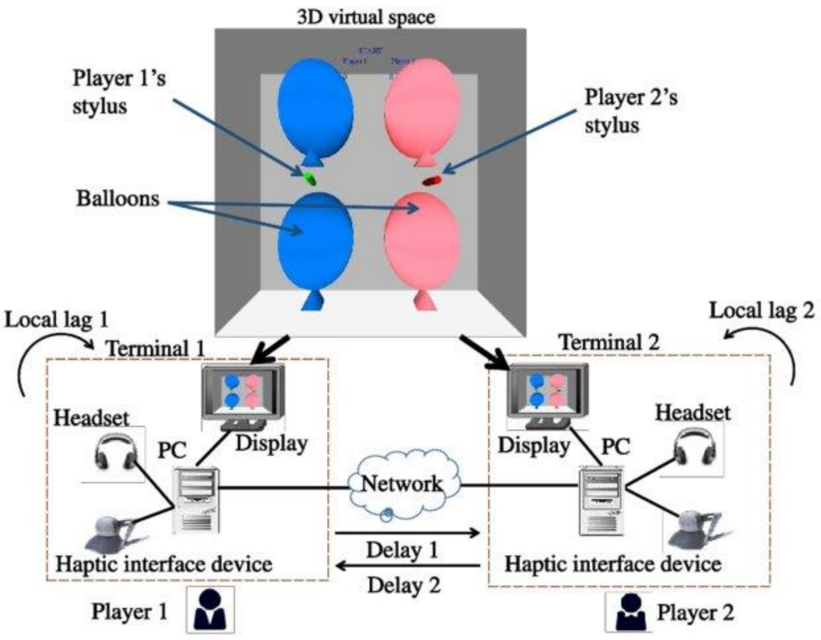


Figure 2.3: Peer to Peer network architecture as seen in [47]

A QoE examination of a networked application, or system, should consider both the systemic factors involved in quality perception (latency, jitter, packet loss) as well as the human and contextual factors specific to the application in question.

2.3.3 Networked Virtual Reality Frameworks

When evaluating QoS/QoE metrics on networked virtual environments, the style of virtual environment, the framework of the networked elements and how their combined design can influence the human factor must all be considered. Networks can be designed in many ways and will change in form and function depending on the application in question, especially so when considering virtual environments. For example, in [48] the authors examine how latency affects user QoE in a drum-playing environment in which one user attempts to mimic another's movements. In this experiment the networked environment was designed for 2 users and employed a peer-to-peer model to connect the users. The results of this study showed a very low tolerance for latency (between 30-60ms).

In [49], the authors employ a client/server-based framework to synchronize the position of an object between users to measure the effects that different methods of viewing a 3D environment would have on performance. The authors found that stereoscopic views increase performance even while subjects were under heavy network latency. While this study does not strictly relate to QoE investigations, it demonstrates possible frameworks for teleoperating via 3D media. With teleoperation as a use case in [50], authors use a one-way and two-way transmission style setup between a master and slave terminal to examine latency in a tele-operative setup from a QoE, with an architecture seen in Figure 2.4. The authors found that in one-way transmissions, QoE was heavily affected by both jitter and latency, whereas in two-way transmissions the effect of jitter was lessened, with the greatest effect on QoE coming from average latency. This study also demonstrated that subjects QoE was affected by delays from 50ms to 200ms in the context of haptic feedback. This method of tele-operative control can be seen in other applications that focus on virtual reality and telepresence as way of improving usability in teleoperation applications [9].

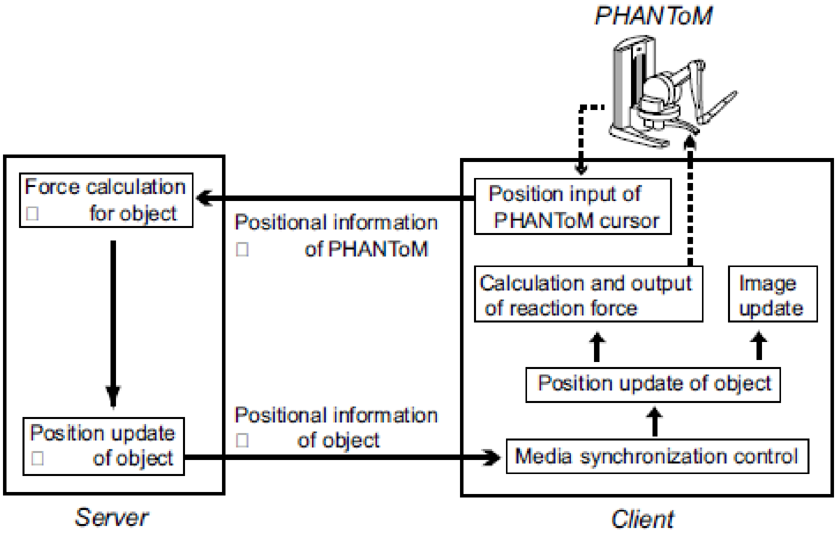


Figure 2.4: Client-Server network framework as seen in [49]

The effects of network delay on human interaction with a system can be measured via established standards and methods that researchers can use to structure studies of their own. Guidelines set out in ITU-T G.1010 [51] inform researchers how to evaluate quality of service measures under network delays, while also offering recommendations for delay values and classifications for various types of applications that are concerned with QoS. ITU-T G.1032 [52] has been set out to inform QoE-centric evaluations related to network elements pertaining to gaming perspectives (of which VR is a part). ITU-T recommendations P.10/G100 [53] serve to define many of the concepts related to QoS/QoE examinations of networking. By using these standards and methods, researchers can form a framework from which to structure their own networking experiments.

2.4 Quality of Experience

2.4.1 Background of QoE

In [54], the fundamental concept of quality of experience is defined as:

> *"The degree of delight or annoyance of a person whose experiencing involves an application, service, or system. It results from the person's evaluation of the fulfilment of his or her expectations and needs with respect to the utility and/or enjoyment in the light of the person's context, personality and current state"*

Essentially, QoE is the discipline concerned with understanding and measuring a user's subjective perception of a system given a certain context or environment. QoE has traditionally been centred on the area of telecommunications. However, as time has passed, the definition of QoE has come to encompass more than this and has extended to many domains in software and multimedia. QoE serves to measure a user's engagement with a system and determine limits within which content is delivered to maximize this engagement. Also, traditionally, QoE was a strictly systems-focused domain, however it has developed to include human influences and context factors as fundamental aspects of measurement.

Figure 2.5. shows how QoE is influenced by multiple factors. These factors can also be more broadly grouped into three categories: human factors, system factors and context factors [55]. Measurements of these factors are divided between quantitative and qualitative measures, making QoE a mixed measurement field of study.

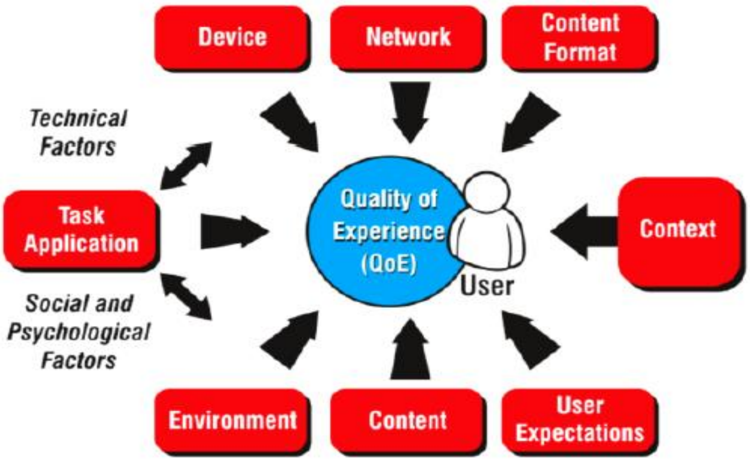


Figure 2.5: QoE Influence Factors, adapted from [16]

In QoE, factors of interest to the context being evaluated, such as immersion, usability and enjoyment, are generally measured using a post-experience questionnaire designed using a 5 to 9-point Likert scale [56]. This is designed to capture the subject's perceived feeling about their experience and create models representing QoE. This can be weighed against more objective measures, such as context sensitive in-experience metrics (for example: time-to-task completion, eye tracking). From the amalgamated explicit questionnaire results a mean opinion score (MOS) is formulated. This MOS is the general metric by which QoE is normally captured. As MOS is purely a subjective measure, and can be influenced by factors such as timing and recency biases, its role as a sole determinant of QoE has shown to be limited [57]. Therefore, the research community has started to expand the methods by which it measures QoE. This includes the introduction of quantitative physiological metrics, such as heart rate and electrodermal activity, through wearable sensors, such as the E4 developed by Empatica [58], to try and map biometric signals to a user's perceived sense of engagement with a system.

2.4.2 Explicit Measurement

Subjective evaluation of QoE is measured generally by breaking down the evaluation of an experience into broad interest factors (for example immersion, interaction, usability etc.) and questions are developed with the aim of evaluating these factors. Subjective evaluation is carried out according to standardised ITU-T guidelines, set in place to inform the research community of best practices for this measurement and to provide a uniform context for subjective data to be disseminated. In the case of VR, the recommendations set out in ITU-T P.910 [56] and ITU-T P.913 [59] are used to inform the evaluation of the subjective metrics used in the work presented in this thesis.

Absolute Category Rating

As laid out in [56], absolute category rating (ACR) is a method by which subjective evaluation data of a visual-audio stimulus can be collected. The requirements for this method of evaluation are as follows:

a) That the subject rate their experience of the stimulus on a 5-point Likert scale as in Table 2.1.

Table 2.1: 5-point Likert Scale for measuring Absolute Category Rating.

1	2	3	4	5
Bad	Poor	Fair	Good	Excellent

b) The Subject is only presented with one stimulus at a time and asked to rate their experience of the stimulus after having been exposed to the stimulus.

Questions are answered through categorical evaluation and the subject may be unaware of the numerical weighting of the categories. This rating method allows researchers to capture a large amount of subjective data in a small period. In certain cases, where a more discriminative rating is required, use of 7-point or 9-point Likert scale rating methods are acceptable.

Degradation Category Rating

Studies focused on the comparison of two stimuli of interest may use the Degradation Category Rating (DCR) system [56]. Using this rating system, the subject is introduced to the initial stimulus of interest, the reference stimulus, and following this a second stimulus is introduced. This allows for a comparative analysis of two experimental stimulus. This type of categorical rating is often used to evaluate differing visual aspects of interest. Like ACR, participants are asked to rate on a 5-point Likert scale as shown in Table 2.2.

Table 2.2: 5-point Likert Scale for measuring Degradation Category Rating

1	2	3	4	5
Very Annoying	Annoying	Slightly annoying	Perceptible but not annoying	Imperceptible

Mean Opinion Score

Mean opinion score, the traditional metric by which QoE is evaluated, serves to represent the overall user-perceived quality of the evaluated stimulus or system. MOS is rated using the same scale as the rating scale that the examiner chooses for their evaluation (ACR or DCR) and acts as an aggregate score for all recorded answers to that evaluation. MOS is commonly used to describe overall user opinion of a video or audio stimulus, however, it has been expanded to include several areas encompassed by QoE to provide insight into a user’s perception of novel systems and technologies.

2.4.3 Implicit Metrics

Though QoE has generally focused on subjective evaluation through the methods outlined in Section 2.4.2, the limitations of a wholly subjectively evaluated experience have been noted in recent studies [17] [57]. Through traditional subjective measures of QoE, users are often asked to fill in questionnaires after experiencing the system or environment of interest. However, this only factors in the user’s emotional state directly after the experience and any variation in emotional state during the experience is lost. These methods depend on conscious responses and often do not provide enough insight into underlying perceptual and cognitive processes [60].

Researchers have proposed various methods for, and evaluations on the relevance of using, biometric capture as an indicator of QoE [61]. Biometric sensors such as the Fitbit and the E4 armlet allow for capture of biometrics such as heart rate, temperature and electrodermal activity. These factors have been shown to correlate with a user's QoE in [61] and [62] act as indicators, along with subjective evaluations, to inform overall enjoyment or engagement assessments with a stimulus or technology. Additionally, research has been conducted into the viability of eye tracking in evaluating a user's experience of an environment or technology [63]. Immersive technologies such as VR and AR have been quick on the uptake of integrating eye-tracking technologies, for example in devices supplied by Tobii [64].

Physiological Metrics

Physiology is concerned with the measurement and capture of signals produced by the body as it undergoes its normal functions [65]. Much research has gone into the study of neurophysiology and the signals attributed to the brain and its relationship with physical responses. While we do not fully understand the complexities of cognition and the conscious neurological workings of the brain, brain functions and interconnections with physical processes around other parts of the human body are understood. These include such processes as increasing oxygen intake when excited, or temperature increase when embarrassed or aroused. By capturing these signals and processes, insight can be gained into aspects of the human experience.

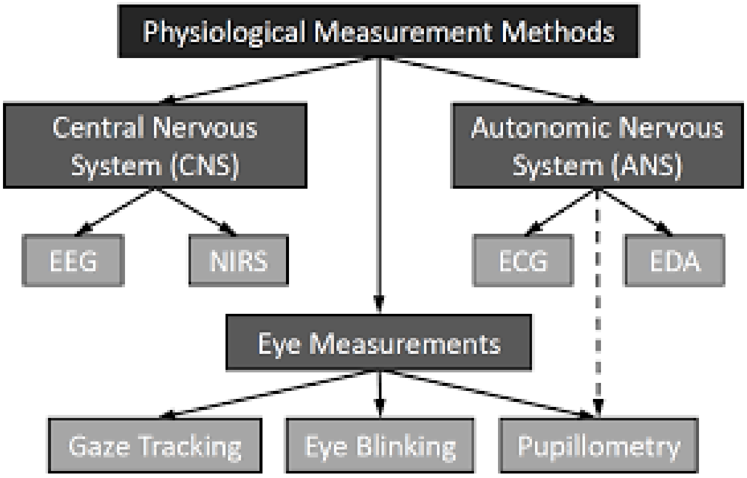


Figure 2.6: Physiological Metrics [66]

As regards QoE studies, a focus has been on the capture of the processes attributed to the central nervous system (CNS), the autonomic nervous system (ANS) and perceptual systems, such as eye measurements. As shown in Figure 2.6., each of these systems has different signals associated with each respective system and by capturing these signals through biometric scanners, researchers interested in QoE can gain an insight into the emotional state of a subject [66].

The ANS is the subset of the peripheral nervous system (PNS), which serves as a way for the CNS to relay information and stimuli to the rest of the body. The ANS is concerned with controlling automatic or involuntary responses exhibited by the human body.

As regards research into QoE, the subset known as the sympathetic nervous system (SNS) is of particular interest. This subset is controlled with physiological responses generally associated with changes in emotional state, for example rising body temperature and heart rate in a state of fear or excitement. This system can influence pulse, skin conductance, blood flow and eye dilation. Previous research into QoE has made use of capturing these types of bodily reponses to help measure influences on emotional arousal [61]. Arousal is a physiological and psychological state whereby sensory organs are stimulated to a point of perception. The metrics that the work presented in this thesis has focused on are those that concerned with the ANS, namely electro dermal actvity and heart rate. These metrics were chosen as they have been shown to have strong links to a person's emotional state and previous works have demonstrated correlations between these signals and a user's reported subjective experience [61] [62] [67].

Electrodermal activity (EDA) or galvanic skin response (GSR) refers to a change in the electrical resistance of the skin that is a physiochemical response to emotional arousal that increases SNS activity [68]. Typically, as a human's emotional state heightens, the conductivity of their skin also increases. These changes in conductivity are divided into two distinct categories, namely, tonic and phasic. Tonic changes are slower, more subtle changes, while phasic changes are identified by more erratic peaks and valleys. Tonic changes are indicative of the general emotional state over a period of time and, therefore, are difficult to use for pinpoint measures of emotion during a stimulus. Phasic skin changes are observed during moments of stress or pain and can be used to indicate sudden changes in emotional state [69]. By measuring changes in the conductive state of skin, one can find an indication of arousal. EDA is measured through placements of electrodes on a subject's skin. These electrodes then produce a small imperceptible current, which indicate the level of conductivity of the skin between the electrodes. By examining the changes in electrical conductance over a period of time, changes in arousal and emotional state can be inferred. While measurement of EDA allows a researcher to pinpoint changes in arousal over time, it does not give insight into the types or specifics in a change of emotional state. As a result of this, a researcher must correlate the EDA responses of a subject to specific stimuli or contexts involved in the test at hand, or combine the results of EDA measurement with other physiological signals to gain a more robust prediction of a subject's QoE. As shown in [67], participants who were exposed to aggressive outbursts showed an increase in EDA and heart rate. The results demonstrated that EDA was an indicator of a heightened sense of anxiety.

Heart rate is an extensivley researched method of determining a human's emotional state, in addition to other indications such as general health, fitness and alertness [70]. Both excitement and stress have been demonstrated to elicit an increase in heart rate. In addition, it has been demonstrated that measuring heart rate variablity is another metric that can provide insight to physiological responses [71]. As a result, it has become a widespread method in the measurement of human emotional response. Changes in heart rate are controlled by impulses from the ANS and increased heart rate arise as a result of the SNS. Decreased heart rate is controleld by the parasympathetic nervous system (PSNS), another subset of the ANS. Methods to measure heart rate are numerous. Electrocardiography is the most recognized method, often used in medical contexts to monitor a patient's heart activity. Data capture can be carried out through placement of electrodes on a subject's body, which will detect miniscule electrical pulses generated by heartbeat. Non-invasive means, such as finger-worn monitors are often used in research environments.

In both the case of EDA and heart rate, QoE researchers prefer the use of non-invasive technologies to capture biometric data. Invasive methods of measuring these signals result in skewed data, as these will often cause unwanted sources excitement or distress on the part of the subject. Non-invasive methods, such as the E4 activity armband, are ideal measurement devices as they can provide accurate and robust biometric data with minimal distress.

Eye gaze is another method by which evaluators may gain an insight into the human emotional state [66]. Eye gaze is the co-ordinated motion of the eyes and head to track visual information that the eyes receive and process every moment. There is a wide body of research to support the observations that eye movement and gaze are linked to cognition and psychological responses. Eye gaze has been used as an indicator of attention and fixation, alongside electroencephalographic measurements [72], while blink rate and dilation indicate visual fatigue and heavy cognitive load [73]. Thus, eye gaze and movement measurements can give a unique insight into human cognitive activity not generally available by other means. Eye movement is a challenging metric to measure as it is often rapid and erratic. Cameras directed towards the eyes can elicit a stressful or agitated response in a subject. This is a situation that is preferably avoided when engaged in the evaluation of QoE. With the improvement in VR technologies, head mounted displays are now come equipped with eye tracking capabilities. An example of such is the Tobii Pro, a modified version of the HTC Vive, with inbuilt eye tracking sensors and technology. This has the benefit of having the eye tracking methodology completely hidden from the subject, which eliminates unwanted changes in emotional state and arousal.

Eye gaze is typically measured in terms of fixations and saccades. Fixations refer to densely populated points of eye gaze in an area being evaluated using eye tracking. Fixations are generally classified by areas of interest or density centroids [74]. Saccades refer to the rapid, jerky eye movements that occur between saccades. Identifying and filtering these rapid eye movements from fixation points is an important aspect in eye tracking analysis. Fixations are generally separated from saccades by applying a velocity threshold to each gaze point in a dataset. Gaze points falling under this threshold are considered fixations and the remaining points are considered as saccades. Threshold determination can vary by context, but generally an angular velocity of <100deg/sec can be considered to indicate a fixation [74]. The stricter a velocity threshold is, the more susceptible the resulting data is to noise, but also allows for a sharper distinction between fixation and saccades. Eye gaze that shows high levels of saccades can suggest a high level of distraction or stress in a person, while high densities of fixations on certain aspects or areas of a stimulus can suggest arousal or interest. Generally, when discussing eye tracking, fixation groupings and densities are broken down into areas-of-interest (AOI), which refer to specific areas in a stimulus that are designated as particularly important or relevant to an experiment.

2.4.4 Interaction Metrics and QoE

One of the novel aspects of virtual reality is the integration of unique forms of interaction with an environment. VR allows a more naturalistic and true-to-life method of interaction, such as someone using their hands to pick up and interact with virtual objects. In [75], the authors outline the basic concepts behind interactions with a system and how it affects a user's QoE of that system. The authors formulate a framework (Figure 2.7) that outlines the various effects of interactivity on not only perceived user QoE, but also on the various environmental influences of interactivity, as well as the consequences that can arise from changes in the various interaction influences. Of note from this study is the concept of flow and technological acceptance, indicating that when users perceive the interaction methodologies within a system to be easy to use, it generally has a net positive QoE impact. Flow refers to the ideal state of use of a system whereby the perceived difficulty of interaction is equal to the optimum amount of concentration exhibited by the user, thereby achieving a state of maximum enjoyment of interacting with a system. This indicates interaction being a heavily influential metric on QoE and, as such, should be considered as part of any holistic QoE evaluation model.

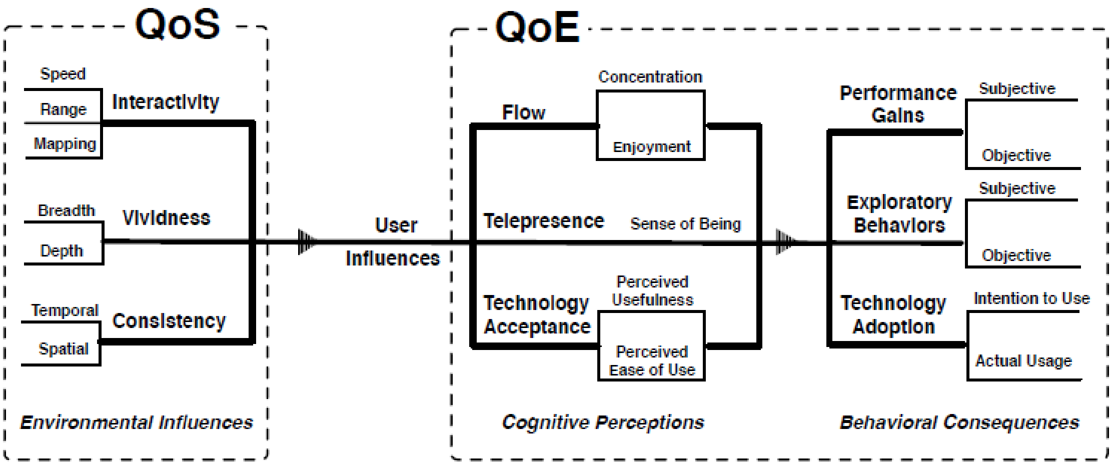


Figure 2.7: The influence factors and perceptions of interaction on QoE and QoS [75]

In [62], the authors developed their own novel methods of measuring user interactions from a QoE viewpoint. Here interaction metrics specific to the task at hand (such as time to task completion, error rates and accuracy) were considered as part of performance measurement. The authors implemented several system level metric capture methods, using immersive technologies, to capture metrics unique to the task at hand. The authors noted that as user’s response times increased, they also showed a correlative increase in EDA. This serves to link previously established emotional indicators, like EDA, to user interactions through scenarios such as stress over being unable to select a correct answer in an AR test (as was the context behind this study). While response times were shown to have a link to increases in physiological metrics, no significant link between wrong answers and misclicks were found.

2.5 Summary

This Section examined and critiqued existing work related to Industry 4.0., applications of QoE and research methodologies utilised in QoE evaluations. There is a large amount of research in support of using immersive technologies as an interaction methodology with CPS and its various paradigms. In addition, there exist many examples of implementations of virtual reality environments for interaction with and manipulation of physical manufacturing units. Several architectures suited to the research presented in this thesis, such as smart factories, Homunculus-style machine operating environments and digital twin-based teleoperation frameworks, were identified. QoE studies have been carried out to examine how users respond to using VR as an interface versus more traditional methods of control, such as a desktop PC or a machine-based terminal. Although there are QoE examinations of tele-operative applications, they largely focused on how they influence a user's performance and learning capabilities. No research could be identified in the way of an examination of network delay, a fundamental aspect of CPS and teleoperation, and its effect on QoE in that regard. In this study, a client-server framework was designed to emulate the behaviour of a master-slave transmission setup, whereby the server acts as the slave terminal simulating a virtual tele-operative machine and the client acts as the master terminal.

Current state-of-the-art trends in QoE evaluation were discussed. Explicit metrics that traditionally made up the bulk of QoE evaluations, such as Likert questionnaires for obtaining MOS, are limited in their scope. Implicit methodologies for QoE assessment are increasingly being employed. QoE researchers now elect to use more physiological-based examinations, using signals like heart rate and EDA to gauge emotional arousal. In addition, other methods are also being employed, such as eye-gaze and pupillometry measurements.

For our evaluation we decided to use a HTC Vive as our HMD of choice, due to it being a good balance between price and performance. It represents a strong midrange device in terms of performance while maintaining a price affordable by an average consumer. It also has 3rd party modification available for it for the purposes of research, which we used in the form of an eye-tracking addon supplied by Tobii. This allowed for a robust 3D eye tracking framework that is largely unexplored and novel. To develop the VR environment, we chose to use Unity as the engine. Unity is both free and easy to learn for beginner level users who are looking to begin entry level game development. It also comes with in build VR support along with toolkits supplied by Valve (proprietors of the HTC Vive) allowing for an easy plug-in and go approach to VR development without having to deviate wildly from an established Unity knowledge base. Unity also has plugin support for Tobii devices allowing us to easily integrate eye tracking frameworks directly into the VR environment. For our physiological measurements, we elected to use an Empatica E4, as it provides a very wide range of signal measurements (including EDA and Heart Rate) while also remaining unobtrusive as a wrist worn device, for reducing the bias in subjective measurements. For our network framework, we used a client to server communication with a server acting as a digital twin stand-in for the Fanuc. The client would interact with the machine in an environment running client side, while the server would receive, and relay commands issued by the client to reflect changes in the VR environment.

Section 3: TECHNOLOGY, METHODOLOGY AND VIRTUAL ENVIRONMENT DESIGN

This section provides an overview of the technologies employed as part of this research and discussion on the various elements that went into the design of the virtual environment. In addition, the research methodology and experiment protocol implemented to answer the research question is presented.

3.1 Equipment and Technology

3.1.1 Virtual Reality Hardware

The HTC Vive [10] is a virtual reality headset developed by the HTC Corporation in collaboration with Valve Corporation. The Vive was the headset chosen as the VR HMD for use in this work. We chose the HTC Vive over other commercially available headsets as it represented a good choice for performance, support and feature set while remaining within a reasonable cost. The Vive has very high tracking accuracy and high-quality equipment that is easy to use. It consists of three different hardware components: the HMD headset, the controllers and the tracking base stations. This configuration allows the Vive to give a room scale VR experience with relative freedom for the user to move around a limited space. The headset and controllers for the Vive can be seen in Figure 3.1.

Figure 3.1: HTC Vive, including native wand controllers

The Vive headset has a refresh rate of 90Hz and a field of view of 90 to 110 degrees, based on adjustment. The headset display consists of organic LED (OLED) panels in a ‘goggle’ like configuration. Each OLED panel has a resolution of 1,080x1,200 pixels. The headset also contains a front-facing camera that allows for a number of features, such as projection of ‘safety walls’ or identifying static real-world objects, for the user’s safety when moving around in the room. The headset uses infrared sensors dotted around the frontal shell to detect and interact with the base stations allowing for high accuracy tracking within the defined VR zone. The headset is fully adjustable for comfort and stability, the outer shell can be adjusted to accommodate glasses or other eyewear required for the user. The headset also includes several auxiliary features such as an inbuilt gyroscope, accelerometer, and proximity sensor.

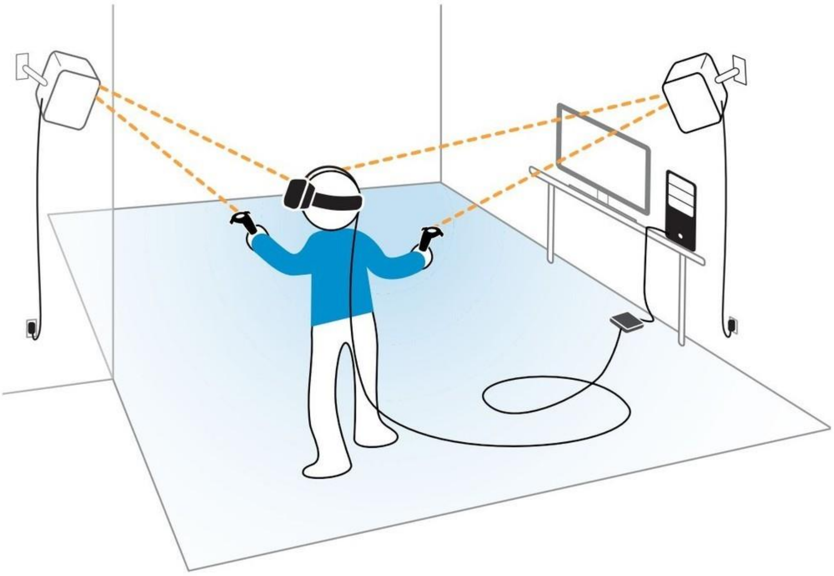


Figure 3.2: Base Stations for the HTC Vive showing how they track in an area.

The Vive base stations, otherwise known as the Lighthouse system, seen in Figure 3.2, are stand or wall-mounted boxes that emit infrared pulses in a zone between them, allowing for 360-degree tracking in an area of up to 5x5 meters between the two boxes. These boxes have a pulse frequency of 60 times a second allowing for a very high degree of positional tracking of the headset when between two of the lighthouses. The base station and headset system are designed in such a way that allows for minimal presence of wires within the Virtual Reality tracking zone, save for the connection of the headset to a local device. This can be eliminated through accessories available for the Vive but was not used for this study.

In addition to the headset and lighthouses, the Vive comes with controllers designed with the tracking system in mind, colloquially known as ‘wands’ depicted in Figure 3.3. Each wand has 4 distinct methods of input that can be freely programmed by VR developers to suit their needs or the architecture of their virtual reality environment.

- A trigger located on the underside of the wand designed to be operated with the index finger.
- A circular track pad located on the top of the controller and operated with the thumb.
- Two grip pads located on the neck of the controlled that can be operated by squeezing with the palm of the hand.
- Two conventional buttons labelled button 1 and 2 located above and below the circle pad, respectively.

This allows for a high array of potential interaction methods with the virtual reality environment while also maintaining a naturalistic control scheme with minimal complexity allowing for easy learning and adaptation to the interaction method. Each wand contains a number of IR sensors much the same as the headset that allow the controllers to be spatially tracked in the same way.

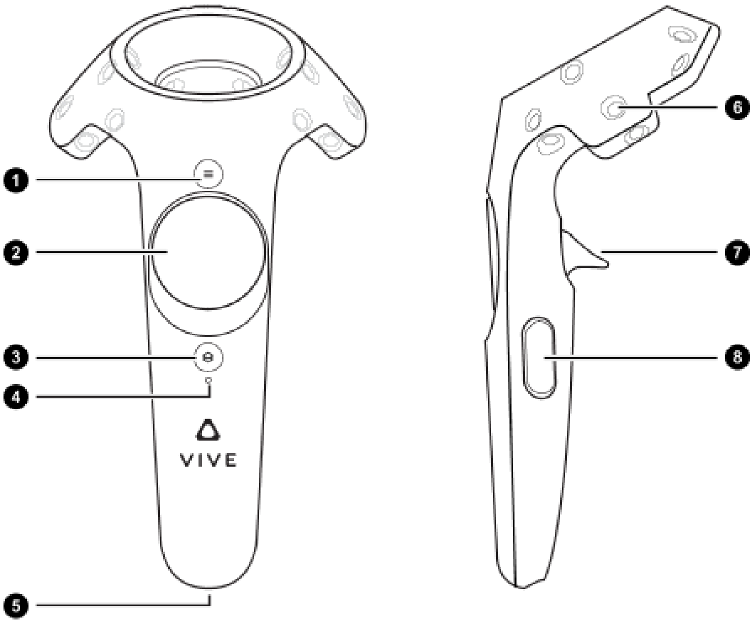


Figure 3.3: Vive controllers as seen in [76]. Buttons shown are 1) Menu button (button 1), 2) Circular trackpad, 3) Button 2, 4) Power status LED, 5) Mini-USB charging port, 6) Infrared Tracking ring, 7) Trigger, 8) Grip buttons.

For the work presented in this thesis, the Unity platform was chosen for development of the VR environment, in particular for its ease of integration with Vive. Unity facilitates significant control over user position, object interaction and the controller scheme for the Vive. Another advantage of Vive was the ready availability of 3rd party modifications for the hardware.

3.1.2 Physiological Sensors and Eye-Tracking

Figure 3.4: Empatica E4 for measuring Heart Rate and EDA [58]

Physiological metrics of heart rate and electrodermal activity were captured using an E4 wristband, developed by Empatica [58] and shown in Figure 3.4. The E4 is a non-invasive method by which biometric capture can be achieved. This device is worn on the wrist and contains sensors that can measure several metrics. Metrics of interest for QoE research are electrodermal activity (EDA) or galvanic skin response (GVR), heart rate, blood pressure, acceleration, and temperature. Of particular interest for the research presented are EDA and heart rate, as they are considered strong indicators of emotional arousal in humans [69]. Heart rate is measured via blood volume pulse (BVP), which serves to measure the variability of heart rate. The E4 measures heart rate via a PPG sensor consisting of 2 red and green LEDs and a photodiode, located on the underside of the main body of the device. This sensor detects the subject’s BVP at a rate of 64Hz in a 15.52mm^2 area. From this, subject heart rate and blood pressure can be determined with a high degree of accuracy.

In the case of our environment, we displayed Heart Rate in normalized time frames based on work presented in [62] referred to as slides. Due to the variability in time to task completion across all participants, it was ideal to present comparative visual data of entire groups by normalizing the length of the test when displaying results. To do this we identified slides 1-3 as the baseline phase for each participant, with slides 4-10 representing the remaining portion of the test. This allowed for a clearer view of change in heart rate over time as participants progressed through the experiment.

Skin conductivity serves as the metric for measuring EDA, and the change in conductivity over time is known as the galvanic skin response (GSR). The E4 measures phasic and tonic changes in skin conductivity via 2 silver electrodes located on the end of device strap, intended to rest on the underside of the subject's inner wrist. These electrodes measure the conductivity of the skin in a range of 0.01 micro-Siemen to 100 micro-Siemens at a frequency of 64Hz. The synchronization of the measurement of EDA and BVP allows for accurate tracking of a subject's change in physiological activity over time. Recorded EDA was represented visually similarly to heart rate, using a normalized slide measure to show change over time.

A modification for the HTC Vive that was implemented in this research was the addition of the Tobii Pro eye tracking system, developed by Tobii Technology [64]. In addition to the ability to project a user's gaze within the virtual environment, the Unity Tobii SDK provides access to the raw data generated by the tracking device. In the virtual reality environment, eye tracking data was collected to see where participants looked most frequently over the course of the test. Eye Gaze data was recorded using a modification for the VIVE, called Tobii Pro, developed by Tobii [64]. The Tobii pro package consists of a hardware modification to an existing HTC Vive, in the form of miniature cameras placed around the rim of the OLED screens on the inside of the headset (Figure 3.5). The eye gaze capture system operated at 120Hz and has an accuracy margin of error of 0.5 degrees. These cameras track a user's eye gaze in real time and the data collected can be used for interaction methodologies within a virtual reality environment via an SDK provided by Tobii for certain platforms such as Unity, Python and MATLAB. Within Unity, the native Tobii SDK was used to calculate and capture head pose

Table 3.1: Picture of the eye tracking camera rims around the OLED lenses [91]

and gaze direction. From the gaze calculations, a unity tool was designed and developed that converted Tobii gaze data to an interactive event object. The Tobii generated data was then converted into world space coordinates in Unity terms, which allowed for the calculation and recording of ray casts that represented a user's eye gaze in real time.

3.1.3 Lab Environment

The experiment was run in a virtual reality laboratory in Athlone Institute of Technology. This lab was designed with space and safety in mind for running tests using immersive multimedia. The equipment consisted of a computer running the VR application, connected via Ethernet to a laptop running the Photon Server software, a software designed to let developers host game servers for free (Expanded in Section 3.2.6). A desk where the participant was asked to sit for baseline measurements and questionnaire completion, further explained in Section 3.3.1, and the space where the virtual reality equipment was setup. The virtual reality testing space was approximately 1.5x1.5 m^2 in which the lighthouses were set up. Within this area was a safety rig as shown in Figure 3.6 that was designed to ensure that the risk of injury due to loss of balance or tripping would be reduced. The safety rig had padded barriers surrounding the subject on all sides except for the entrance. participants were strapped into a safety harness suspended by surgical tubing that ensured they were safely secured. Finally, participants were placed in front of a set of handlebars they could grip in the event of loss of balance. The

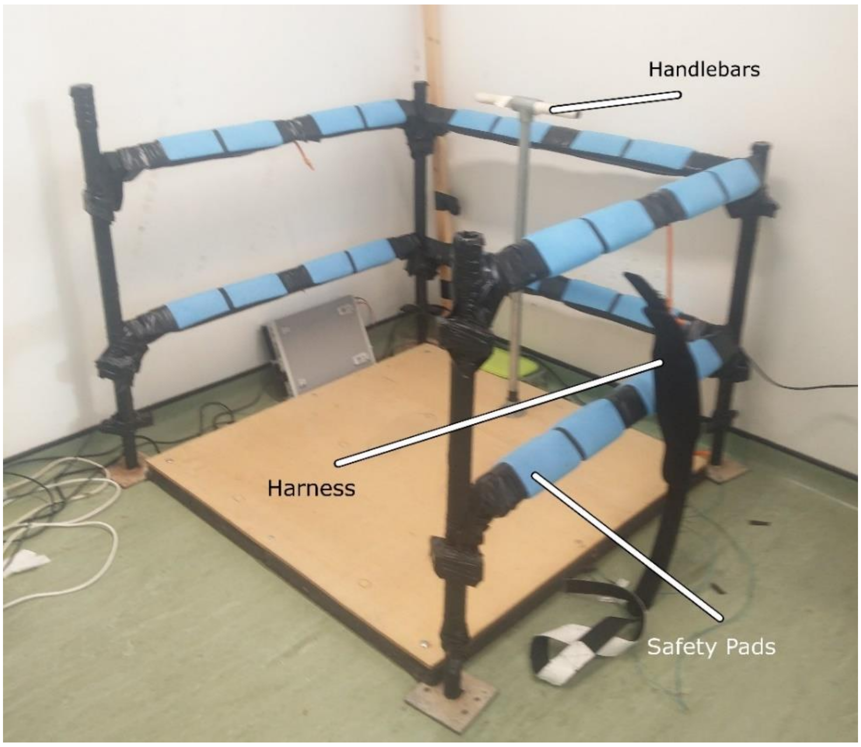


Figure 3.5: Safety Rig used to ensure that participants did not suffer any injury.

environment was always well lit during the experiment to ensure proper visibility and the temperature was maintained at a comfortable level, though during the test users' vision was completely obscured by the headset.

The computer used to run the virtual environment was composed of an Intel I7-8700 3.20 GHz CPU, an NVIDIA GTX 1080, 16GB of RAM and ran the Windows 10 Home edition. The version of Unity used to run the VR environment was Unity 2017.4. The computer used to run the server had similar specifications and again was running Windows 10 with a copy of Photon's Server software, while also running an instance of Network Emulator for Windows Toolkit (NEWT). NEWT is a tool that allows for emulation of various network conditions depending on the user's needs. It can influence lag, jitter, and packet loss as well as other features. For our purposes we used it solely for increasing the delay between the server and the client.

3.2 The Virtual Environment

3.2.1 Development Engine

The virtual reality environment for this experiment was developed in Unity 2017.4. Unity is a game development engine aimed at allowing developers access to a comprehensive set of tools for designing and creating games and virtual experiences for free. Unity is an ideal platform for developing virtual reality experiences for a few reasons. Unity provides a strong set of initial tutorials allowing newer developers to get a handle on the engine quickly. In addition, all scripting and coding of processes and methods are handled in the programming language C#, using a library native to Unity. As C# is a widely used programming language in other areas, such as mobile and Android development, it offers a good entry point for developers.

Aside from the ease of learning and use, Unity offers a high level of support for development and distribution of third-party modifications and tools. Users who develop personal plugins and toolsets for Unity are able to distribute them on the Unity store for others to import into their own projects. This allows for the availability of a vast array of tools without the need for immediate development of these tools for each individual project. Among these tools are a set of plugins that allow for adapting regular virtual environments in Unity into VR in a plug and play manner. Unity provides in-built first party support for many of the common consumer headsets, such as the HTC Vive, alongside Unity specific SDKs provided by the developers of these headsets.

For modification of the 3D models used in Unity, the 3D modelling software Blender [77] was used. Blender allows for the comprehensive creation, modification and animation of 3D models and environments. It allows for concise control over textures, lighting and physics simulation.

3.2.2 Toolkits

For the development of the virtual environment, two separate SDKs designed for the purpose of allowing simple VR development in Unity were used. The first was the SteamVR SDK [78] (provided by Valve) for the purpose of allowing Vive development in Unity. This SDK provides all the functions and Unity objects (in-engine asset) needed to run a virtual environment in Unity, as well as tools for configuring control schemes and virtual position tracking models for the Vive controllers.

The second toolkit used was Virtual Reality Toolkit (VRTK) [79]. VRTK is a third-party toolkit with a collection of scripts and assets to aid with building and executing a virtual reality environment in Unity. VRTK offers a number of streamlined modules that help handle locomotion and interactions such as touching and grabbing. It also provided virtual reality UI elements and other physics-based assets such as gravity affected hinges, rotators, and buttons.

3.2.3 Digital Twin Architecture (Fanuc Roboshot)

Figure 3.6: Fanuc Roboshot S50ia Injection Moulding machine [80]

The purpose of this work was to investigate the effects of network delay on a user’s QoE while using a teleoperation application. For this reason, it was important to design the application in such a way that reflected a real-world example of this type of environment, such as one that could conceivably be used for Industry 4.0 purposes. The following factors were considered:

- complexity – how difficult or easy is it to use the machine?

- safety - does real-world operation of the machine present a potential hazard?
- intercommunicative ability - can the machine be teleoperated?
- feasibility of virtualisation - is it possible to create an accurate VR model of the machine?

The Fanuc Roboshot [80], shown in Figure 3.7, is an injection moulding machine designed to create accurate, high quality plastic moulds that can be reproduced and replicated on a mass production scale. The Fanuc is designed for ease of use and sustainability and contains many of the characteristics typical of mechanical units designed with Industry 4.0 principles in mind. Athlone Institute of Technology has a Fanuc Roboshot a-S150ia on site and it represented a good fit for the four factors considered above. For these reasons it was chosen as the use-case for the experimental investigation.

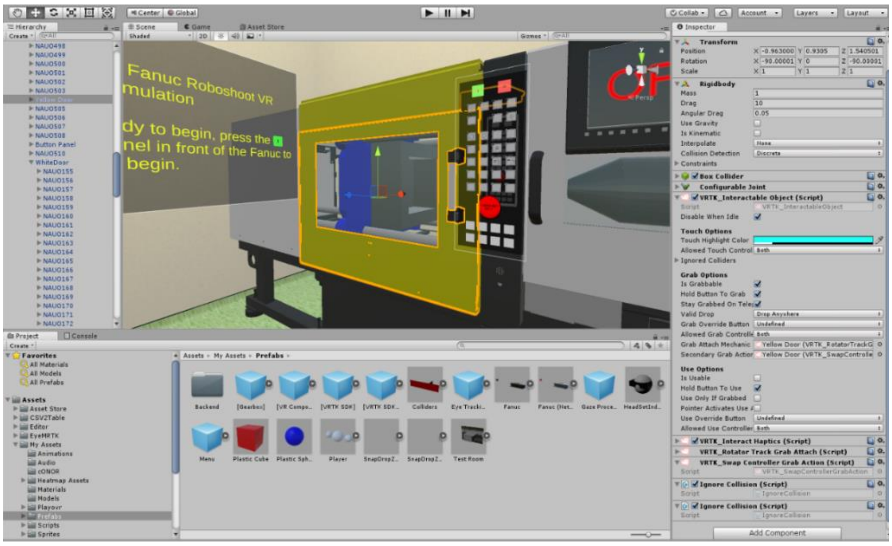


Figure 3.7: Developer UI of Unity environment. Figure shows some of the toolsets and viewports available in Unity, as well as prefab objects used to create interactive elements

A basic virtual 3D model of the Fanuc was sourced from JL Goor [81], a company specializing in plastic moulding. Blender was used to modify this model in order to add a rudimentary analogy of the real-world control panel (a user interface panel), injection unit and nozzle, as well as a basic simple mould that would produce basic polygonal objects such as a sphere. When the model reflected the real-world Fanuc visually and proportionally, it was imported into Unity where further work was done to replicate many of the moving parts that the Fanuc model would require to operate in the virtual environment. This included adding physics properties to the doors and mould area, such as sliding rails and mechanical movement. Unity physics methods and scripts were used to achieve this and model movements specific to different parts of the Fanuc (for example, the opening of the Fanuc's doors would move and operate in a different manner to the movement of the mould or injection unit).

To allow user-control of a Fanuc within the VR environment, a press-and-play style approach was used to implement a basic Fanuc operation cycle. The steps in this cycle were (i) turning on the Fanuc, ensuring the plastic injection unit was correctly in place and locked to ensure no manufacturing failure, and (ii) engaging a single mould production cycle. The real-world button presses for this task on the Fanuc control panel were replicated on the virtual reality UI panel, which was modelled on the control panel. The two panels can be compared in Figure 3.9. Buttons that were not required for the task were left unlabelled on the virtual user interface panel. Some buttons on the user interface panel were changed to accommodate certain VR specific tasks (such as changing the mould within the Fanuc, done by hand normally).

The virtual user interface panel was designed in such a way that allowed for on-the-fly programming of functions and scripts. Each button on this panel, including the ones not used for the task, were set to be allowed to run a script assigned to it. In addition, buttons could be set to perform toggles or co-routines. This flexibility was built in to allow for future modifications without requiring a complete rebuild of the environment.

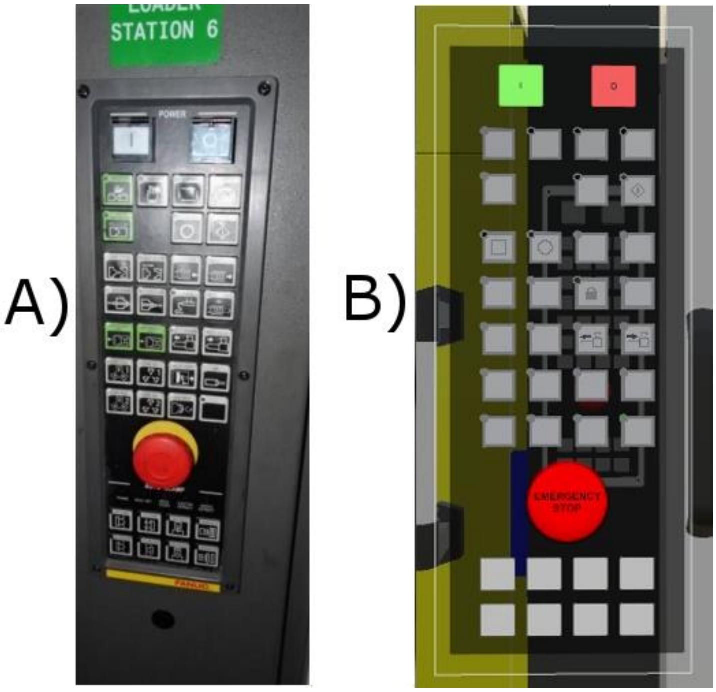


Figure 3.8: A) Real control panel on the Fanuc machine panel and B) VR panel developed within the virtual environment

The essential buttons required for this task were then programmed to do what was needed to complete the task. The real-life series of processes and operations of the Fanuc were constructed into a logic table. These were then translated into checks and triggers in Unity to ensure that participants could operate the Fanuc correctly. Checks were also implemented to ensure that participants could not progress the test until they had completed the necessary steps in the correct sequence. These triggers and checks also served to signal to the environment when the instructions provided to the user were to change, with new sets of instructions appearing when the user successfully completed a sub-task needed to progress the test. All logic checks were handled by an object set to enable and disable Unity objects as a series rudimentary I/O checks.

The task selected was designed to be simple to implement and understand for any user of the environment, under the assumption that the user had not previously used VR and had not operated a unit such as the Fanuc. The basic steps required by the user to complete the task were as follows:

1. Turn the machine on.
2. Navigate the injection unit into the correct place and lock it there to ensure safety.
3. Select the type and shape of object they would like to produce from the mould.
4. Engage the moulding process.
5. Extract the object once completed.
6. Turn off the Fanuc.

The simplicity of the task selected was intended to reduce the cognitive load associated with learning how to interact with the VR environment and to operate the Fanuc. It is known that heavy cognitive load can cause stress and irritation. Thus, reducing these factors was intended to eliminate any negative effect on the QoE a user might have had. This was done so that there would be minimal influence from these factors and the influence of network delay could be more easily assessed.

3.2.4 Auxiliary Features

Some of the important aspects of VR design include the addition of haptic, auditory and visual feedback for actions taken by the user over the duration of the test [82] [83]. To help immerse the user within the virtual environment, auditory feedback was generated when a button on the user interface panel was pressed. In addition, the controller emitted a rumble to indicate the pressing of a button or activation of an interaction, adding a layer of haptic feedback to the environment. The doors and injection nozzle made sliding noises as they moved, as well as an auditory locking sound when the injection nozzle was placed in the correct spot at that point in the task. When the Fanuc was activated, the sounds of machinery operating were heard by the user. A chime sounded to signal to a user that they had completed a step correctly. These elements provided clarity of use and feedback from the Fanuc, as well as improved immersion within the virtual environment. Certain sounds also served to reflect the network delay during the test. For example, the sound of a button clicking would not play until the server had received the signal from the script and communicated back to the client to play the relevant sound, similarly for the chime and operation sounds. Rumble was not delayed to indicate that the user had indeed correctly pressed the button on the physical hardware correctly. This was done to avoid masking the effects of delay on QoE, as in pilot testing it was noted that without auditory feedback participants would not notice delay nearly as much, nor would they understand when they were expected to proceed with the next phase of the test.

Visual feedback was given to the user in a number of ways. The LED screen normally used on a real world Fanuc for displaying data related to the moulding process information was repurposed in the VR environment to give feedback to the user about the state of the machine (such as whether it was ON or OFF and if there were any actions that the user needed to take to proceed with the current step) as seen in Figure 3.8. In addition, parts of the Fanuc that were to be handled, would be highlighted in strong contrasting colours when the user was able to grasp the object. For example, when the user moved the controller into the virtual door handle, it would glow a bright blue alongside a very minor controller rumble to indicate to the user that it could now be handled. When operating the Fanuc from a distance, a laser was emitted from the virtual model of the controller that allowed for precision button pressing and interaction from a distance. This laser could also change colour depending on if the user was hovering over an object that could be interacted with (red if not, green if yes). This was done to improve clarity of interaction and to alleviate some cognitive load of learning a new interaction style.

3.2.5 Environmental Design

The virtual environment was split into two parts, the training playground and the test environment. This is in line with the experimental protocol in Section 3.3. The training playground was designed as a simple abstract space where the user could be introduced to the control scheme without any external distracting factors. This environment consisted only of a white space with a table and a virtual keyboard hovering in the air, illustrated in Figure 3.10.

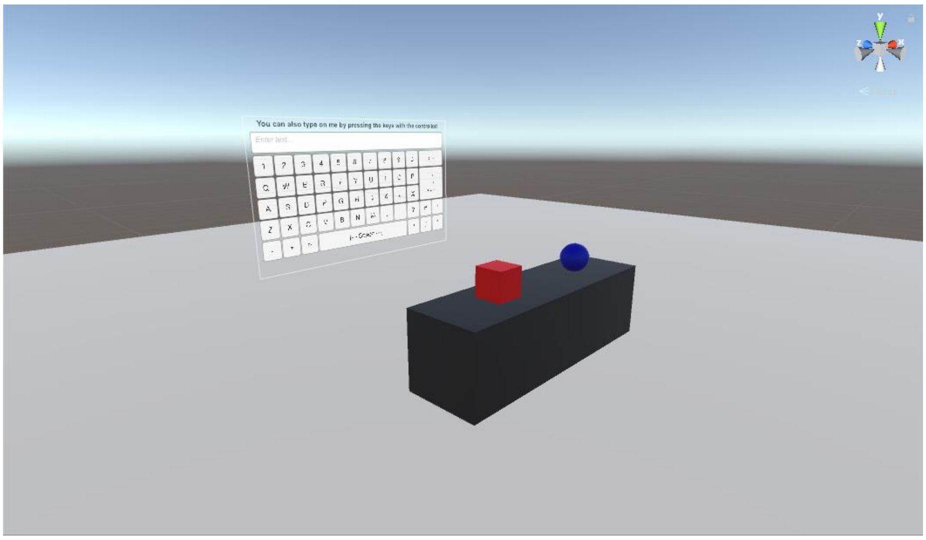


Figure 3.9: Training Environment with basic interactable objects and a keyboard to introduce the concept of grabbing and button pressing to the user

On the table was placed a sphere and a cube, each of which could be interacted with and influenced by the physics engine within Unity. These were designed to introduce participants to the concept of objects that could be grasped and to set expectations on how virtual objects would likely act within the physics of the testing environment. It also served to introduce the basic elements of the control scheme: press the trigger to interact; press the circle pad to activate the distance pointer. The UI keyboard served to introduce more accurately the type of interaction expected of the participants in the actual test.

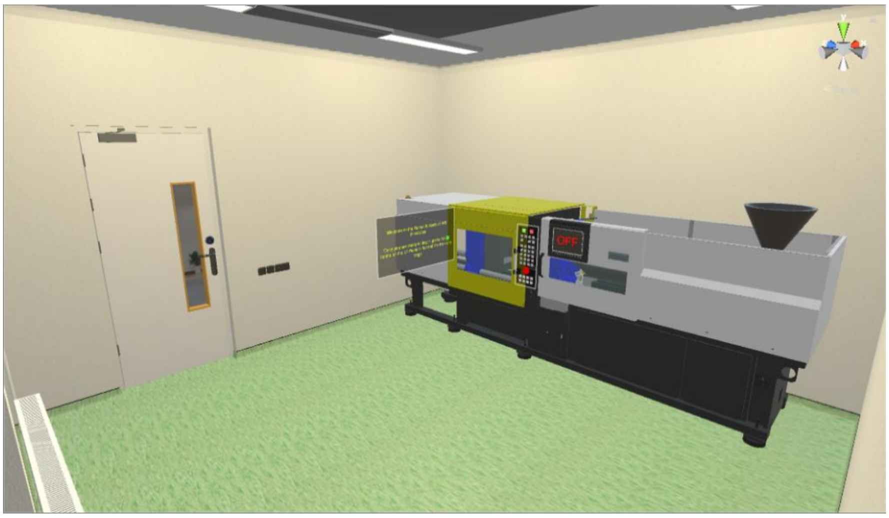
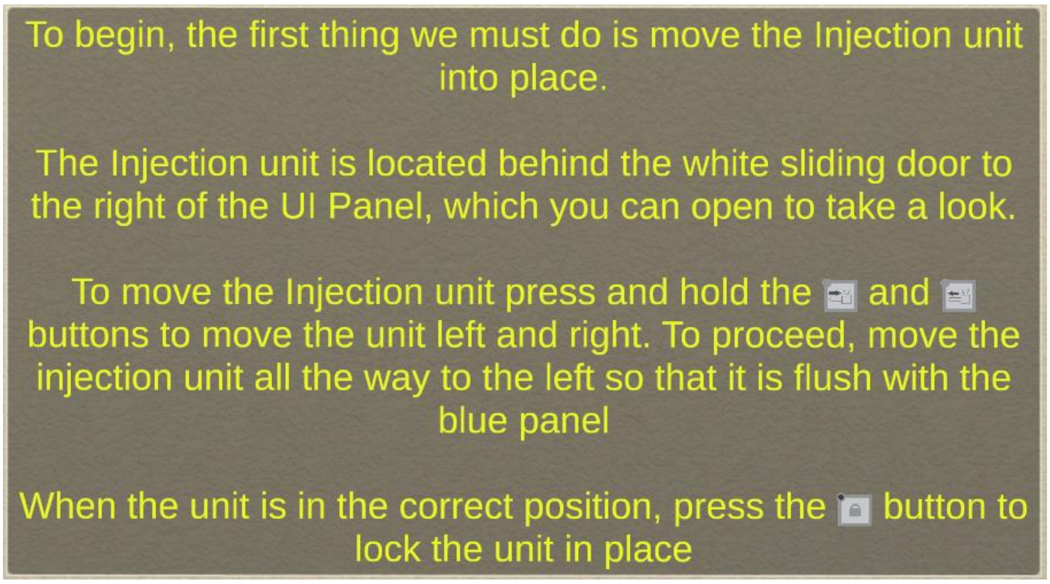


Figure 3.10: Testing Environment displaying the virtual Fanuc machine within a virtual copy of the testing laboratory.

Once the user indicated they were comfortable with the interaction and control schemes, by verbally informing the researcher present that they were ready to proceed, the training environment was set to transition to the testing environment as seen in Figure 3.11. The testing environment was designed as a 1:1 scale model of the same laboratory that the real world experiments were conducted in. The Fanuc was placed on the north wall of the lab opposite the windows. Ambient noise was added to match the atmosphere of the actual lab. Lighting was set to be bright and clear to help participants see and distinguish any visual elements. The colours of the UI and Fanuc were deliberately chosen to be bright and contrasting with as little visual clutter as possible. This was done to maximise the sense of presence of the user within the environment and to reduce simulator sickness by grounding the user in a familiar place, as well as giving the user a sense of scale compared to the real world.

A set of instructions was also placed within the environment to guide the participants through the test without input from the examiner, illustrated in Figure 3.12. These instructions were linked to the logic object of the Fanuc and progressed as the machine's state changed. Text colour on the instructions was selected as a bright yellow on a grey semi-transparent background to allow for a high level of contrast. In addition, all buttons that the user was required to press by the instructions were displayed as a sprite image on the text. This was to improve visual clarity and allow for greater ease of use as well as easing possible language barriers to the instructions. The instructions also provided audio feedback when they changed throughout the test to indicate to the participants that they had progressed to the next step.



To begin, the first thing we must do is move the Injection unit into place.

The Injection unit is located behind the white sliding door to the right of the UI Panel, which you can open to take a look.

To move the Injection unit press and hold the → and ← buttons to move the unit left and right. To proceed, move the injection unit all the way to the left so that it is flush with the blue panel

When the unit is in the correct position, press the 🔒 button to lock the unit in place

Figure 3.11: Screenshot of a panel of the instructions provided to participants throughout the experiment. 2D Sprites were used to give a visual indication of which button to press instead of using text descriptions.

3.2.6 Network Simulator

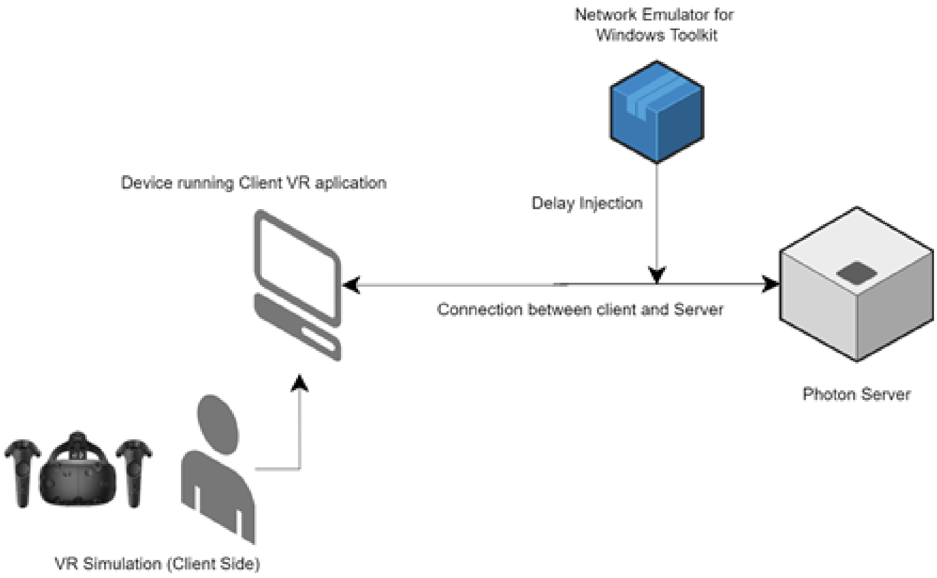


Figure 3.12: Architecture of the Network Simulation. The user interacted with the client VR terminal which connected to a Photon Server handling unity action calls. A network emulator injected artificial delay between the two

To simulate network activity, the Photon Network architecture was applied to the virtual reality environment [84]. Photon is a third-party asset for the Unity Engine that allows developers to add multiplayer capabilities to their projects. Photon also offered a method by which the server for the VR environment could be hosted locally, thus eliminating unwanted network effects such as jitter. Photon also supplies developers with the means to simulate different network conditions, however, for the experimental setup used, it was easier to emulate poor network connections using a third-party application, such as Network Emulator for Windows Toolkit [85], in between the local user and server.

The client-server configuration used in the experimental setup is illustrated in Figure 3.13. An instance of the virtual environment was run on the server side, while participants interacted with a client-side mirror of the virtual environment on a second computer. The server, which was running the same instance of Unity and the environment as the client, without the VR elements, handled the machine state of the Fanuc and reflected its status on the client-side machine. This is how a pseudo-digital twin analogy was achieved, with the virtual server Fanuc acting as the real-world Fanuc in this framework. When the user interacted with the client Fanuc and initiated any processes or scripts using the user interface panel, a remote process call (RPC) was sent to the server, which indicated to it that it needed to run a process. The server would then run the required process and tell the client Fanuc to update its state to match that of the server. By injecting delay into the connection between the client and the server, and delaying the RPCs, an emulation of network latency was achieved. Each process that was required to be performed as part of the task was designed as an RPC script so that Photon would recognise them as a networked element.

Unity's handling of audio was also routed through RPC calls. When a process was initiated that required a sound to play (such as clicking a button or the whirring of machine noises) the call to initiate that sound was sent through an RPC, then the server would ask the environment to play the sound once it had received it. This allowed for sounds to be delayed under the same latency conditions as the processes and state updates of the Fanuc. Sound was deliberately delayed in this way, so that it was synchronised with the visual and interaction delays. Client movement and controller interactions were not affected by this client-server communication process for two reasons. The first was that the client's movements and position were not required to be mirrored on the server for the environment to function since a digital twin framework would not need to consider the position of the user. Secondly, even minor delays or lags in visual feedback or positional attention in virtual reality can cause severe cybersickness and it was preferable to avoid this.

3.2.7 Eye Tracking

As mentioned previously, eye tracking was achieved through the Tobii Pro modification for the HTC Vive. One of the major challenges for this environment was designing a method by which eye tracking data could be isolated, captured, recorded and analysed. Tobii supplied an SDK with their hardware that allowed for use of eye gaze to act as a visual indicator and interact with the environment. However, this SDK was not designed with raw eye gaze data capture in mind. As such, a system was designed that collected raw data from the Tobii SDK and stored it in text files. The SDK had a system of labelling and mapping gaze in terms of pose and position. The combination of these two elements served as a starting point for the gaze and the direction the eyes were pointing too. This gaze data was calculated separately for each eye and then the SDK performed a calculation to combine each eye's data into a combined gaze point. The raw pose and position data of the combined gaze were collected and stored in a separate text file, alongside validity checks and the timestamp that each gaze data point was collected at. In addition, the information from the Tobii SDK was used to create raycasts (a Unity construct that allows for the creation of a vector from a single point out along a direction of some determined length, like a ray of light) so that eye gaze data could be converted directly into readable Unity co-ordinates.

Using the data collected, the raw Tobii data was read back into Unity post-test and the timeline of a user's gaze data was stored in a list object. By mapping each element in these reconstructed lists to a frame timestamp of the test, it was possible to reconstruct the gaze of each individual subject for replay and analysis. Each recorded gaze element was then converted into a raycast of 10m length and checked for the individual collisions of each ray with some object in the environment. By checking these collisions and mapping them to the objects they were located on it was possible to create an areas-of-interest map to show what objects participants most focused on in a test and when these objects were focused on. Before the test, the control console, the face of the Fanuc and the user instructions were defined as the most important areas of interest and would be the ones of most interest in the analysis. Areas of interest were also made large and encompassing of broad areas of the Fanuc to not have user height or other varying factors be a large influence on how eye gaze data was collected.

Once gaze had been reconstructed and collisions calculated, fixations in the eye gaze data were filtered from saccadic movements. This involved setting a velocity threshold between points. If the velocity between two points exceeded the defined threshold, the later gaze point was labelled as saccadic and excluded from analysis. In addition, eye gaze data that did not fall within any of the defined areas of interest was labelled as saccadic and excluded from analysis. As mentioned in the literature [74], the threshold generally accepted for identifying fixations (influenced by environment and context) is <100deg/s. For the analysis carried out in this thesis a velocity of 40 deg/s was defined as the threshold as it offered the best compromise between data noise and clear definition of fixation groups. This threshold value was determined based

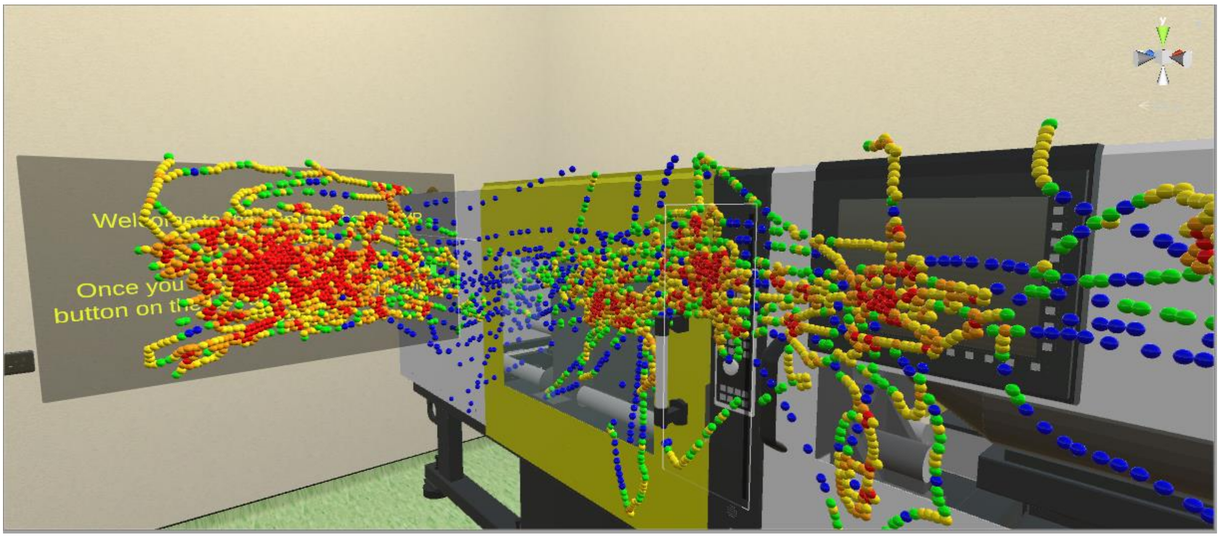


Figure 3.13: Example of heatmap functionalities developed for the eye gaze framework. Red nodes indicate areas of most visual raycast density with blue nodes representing areas of least density (scale: Red-Orange-Yellow-Green-Blue.). This indicated areas users were most focused on during the testing phase.

on pilot testing of varying thresholds on captured eye gaze data when using the virtual reality environment.

The analysis of eye tracking data in these experiments was based on [74]. This allowed for a thorough analysis of eye-tracking data by applying a velocity threshold analysis alongside an area-of-interest methodology. The eye gaze analysis framework was also developed to allow for more accurate recording, such as individual point location within the areas of interest.

As seen in Figure 3.14, a rudimentary heatmap was generated to visualize eye gaze within the virtual environment. Though there exist many implementations for VR for 3D surface gradient heatmaps, many are only available through proprietary software with significant costs involved in the additions of these technologies to an environment. To allow for a visual element to the gaze data collected in the experiments, and to avoid the cost of proprietary software, a different method was implemented that was inspired by [86] [87]. In [86], eye gaze was visualized by generating spheres at gaze points recorded during the study. These spheres were then coloured based on the distance between the gaze point and the point of gaze origin. In [87], a proof-of-concept density map was generated in 3D space that demonstrated that the sphere method could be used to visualize gaze data.

Gaze points were read from the stored data and remapped to the virtual environment by checking their collision point and point of origin. This resulted in a series of points mapped to their exact position in 3D space when they were recorded. Gaze points not falling within the predefined area-of-interest within the environment were eliminated and the remaining gaze points were assigned a Unity sphere object. This object would then run a calculation to change its colour based on the frequency count of nearby other identical objects. This resulted in a collection of spheres that were assigned a higher heat based on the density of nearby spheres. Spheres that overlapped a single point multiple times and spheres within a very small radius (roughly half a radius of the sphere object from the centroid point) were eliminated to avoid overcrowding and to achieve a smoother heatmap effect. Though this method is a crude way of calculating the heat intensity, the end result did allow for a useful visualisation of where a user looked in the environment as indicated by higher densities of spheres within the area-of-interest zones.

3.3 Research Methodology

The research method employed in this work is experimental. As outlined in Section 1.2, the focus of this research was to develop a virtual reality application that could be used to understand the impact of network delay on the user QoE. In this context, the experiments focused on the relationship between virtual reality technologies and the network behaviour. The variable considered in this work between the user or subject groups is network delay. One of the subject groups (as part of the between-group design) experienced the ideal zero delay state. The other groups experienced the VR environment subjected to artificial network delay of varying levels. Based on pilot tests to gauge subject reactions to network delay, the delay values selected for this experiment were 0s, 1s, 2s and 3s. In our pilot tests, it was found that users did not react or notice a delay below 1 second, despite indication in the literature review that delays of far lower would be noticeable. This led us to further develop our delay thresholds beyond that. Once it was found that above 1s of delay was still being reported as acceptable we decided on 1s intervals to test extreme levels of delay.

Results from the experiment were captured and evaluated in different ways. Subjective data pertaining to QoE was captured via a post-test questionnaire. Objective physiological data and biometric signals, such as heart rate and EDA were captured using the Empatica E4 over the duration of the test. Additionally, eye gaze and head movement were captured through the VR HMD. Furthermore, gaze points were analysed in terms of fixations and saccades. The novel combination of these elements was used to gain insight into the participants perceived QoE of a VR application in the presence of varying levels of network delay.

3.3.1 Research Methodology Phases

Information and Screening Phase	Resting Phase	Training Phase	Testing Phase
Information	5 Minute Resting	Training Description	Virtual Reality Test
Visual Acuity Test	HR Baseline	Training Activity	HR, EDA and Eye Gaze Captured
Color Perception Test	EDA Baseline	HR and EDA Capture	Post-Test Questionnaire

Figure 3.14: Outline of research methodology adapted from [62]

The research methodology included four key phases, shown visually in Figure 3.15. First, the information and screening phase is where the participants were informed about what would be taking place in the test. As part of this phase, participants were provided with an information sheet outlining the project, its aims and what they should expect over the duration of the test. Participants were introduced to the technology used during the test, namely the Empatica E4 and the Vive HMD. Participants were given a consent form and could declare if they would like to withdraw for any reason. As a VR HMD is a heavily visual-based technology, participants were screened for good visual health. As part of this process participants were evaluated for visual defects that could influence perception. Visual acuity was evaluated using the Snellen test [88]. Participants were permitted to use their glasses if needed to ensure visual acuity. Colour perception was evaluated using the Ishihara test [89] to ensure participants were not red-green colour-blind. Any participants who did not pass these tests were considered ineligible.

In phase two, the resting phase, participants had the Empatica E4 placed on their wrist of their non-dominant hand and asked to wait for a 5-minute recess. This served to obtain a baseline measure of the subject’s heart rate and EDA. Due to the novel nature of immersive technologies such as VR and the nature of experiencing new technology, participants could experience a degree of excitement or anxiety. This would introduce the potential to alter physiological measurements in a way that would introduce bias to the evaluation of QoE metrics. For this reason, participants were subjected to this rest period in phase two.

Phase three introduced participants to the technologies involved in the test. Participants then went through a brief practice environment that allowed them to familiarize themselves with the VR technology and interaction methods, the training phase. After a brief introduction and explanation of the control scheme, the HMD was fitted, and it was ensured that all the hardware was comfortable for each subject. Participants were then placed in an interactive environment that allowed them to pick up and interact with primitive objects. Upon completion, an opportunity was provided to ask questions regarding the test and the control methods. Participants could request an extension to the time of this training phase if necessary.

In the final testing phase, participants interacted with the actual test environment. They were directed to interact with the digital representation of the Fanuc moulding machine. Participants were guided via instructions (See Figure 3.12) within the virtual environment and were informed at each step, by means of a visual aid, the actions they are required to take. Actions taken within this phase of this test were based on a beginner’s introduction to the physical Fanuc machine and the test was designed to emulate that experience as closely as possible. Audio and visual feedback was implemented to help user engagement. Heart rate and EDA were measured throughout. In addition, user eye gaze was also measured and stored during this phase. Upon completion of the testing, the subject was asked to complete the subjective questionnaire.

3.4 Questionnaire

A post-experience questionnaire was designed to gather information from all participants in each group (0s delay and the three other delay groups). The absolute category rating system was used to measure these subjective values on a 5-point Likert scale [59] to provide a MOS rating. Inspired by subjective evaluations from [61] [62] [60], ten questions were developed to measure various factors of user QoE in terms of immersion, usability and interaction. A breakdown of the questionnaire is given in Table 3.1.

Table 3.2: Questions labelled by factor: Immersion (Blue), Interaction (Red) and Usability (Green). Immersion refers to how comfortable and engaged the users were with the environment. Interaction refers to how natural the control scheme and interaction paradigms were to use and learn. Usability is how easy users felt the task in the virtual environment was.

Q1: I was immersed in the virtual environment	Q2: I felt a strong sense of presence in the virtual environment	Q3: The virtual environment did not feel realistic
Q4: Using the system was natural	Q5: I did not enjoy interacting with the virtual environment	Q6: I was not restricted in my movements
Q7: The system was slow and unresponsive	Q8: The task was not easy to complete	Q9: The instructions were not useful
	Q10: There was a delay in the virtual environment	

The questionnaire was designed to assess three factors of interest, namely, immersion, interaction, and usability. All questions were randomized for each individual subject to eliminate ordering effects. In addition, questions were asked with either a positive or negative syntax, so that participants would not be influenced into answering overwhelmingly positive or negative for every question.

The first factor, queried by Questions 1, 2 and 3 in Table 3.1, aimed to test how immersive the virtual environment was, which is an important factor in measuring QoE. These questions largely asked the participants if they felt that the environment was engaging and if taking part in the task helped them feel disconnected from the world outside of the virtual reality environment.

The second factor queried by the questionnaire was interaction (Questions 4, 5 and 6 in Table 3.1), which focused on the methods of interaction in the VR environment and if the participants found them acceptable or enjoyable. Since the VR environment had to be controlled using a native controller, coupled with the relative complexity in translating the task of operating a machine to a virtual environment, it was important to examine from a QoE perspective how well participants reacted to the method of control.

The last factor was usability (Questions7, 8, 9 and 10 in Table 3.1) and could be considered the most important factor relating to the use of VR for an Industry 4.0 supported task. Here participants were asked to rate how suited the virtual reality environment was to the task at hand, as well as aiming to query how the participants felt when the variable of delay was introduced.

3.5 Summary

In this section, a discussion of the technologies used in the experiment as well as the methodologies used was provided. We aimed to discuss and demonstrate the reasoning and justifications that guided the development of this experiment. We also discussed the variables that were of primary interest for this work and why they were chosen. We also aimed to reproduce the foundation of the construction of the virtual environment to guide readers towards reproducing this work for their own future use. In the next section we will discuss the results that we recorded from this work as well as a discussion and analysis of the results that were found.

Section 4: Results and Discussion

This section presents the results from experiments performed using the methodology outlined in the previous section. Of interest for this section is:

1. The perceived impact of network delay on user QoE of a virtual reality environment.
2. How QoE could be inferred from interactions and bio signals in this type of applications.
3. How the design of this QoE Capture framework contributes to research in this area.
4. The impact of inclusion of eye-tracking to a QoE study in immersive multimedia.

4.1 Population Overview

The subject recruitment for this work was based on convenience sampling. The participants were largely sourced from the local student population, with volunteers from outside the Institute also included. Out of a total of 40 participants, 18 were female and 22 were male.

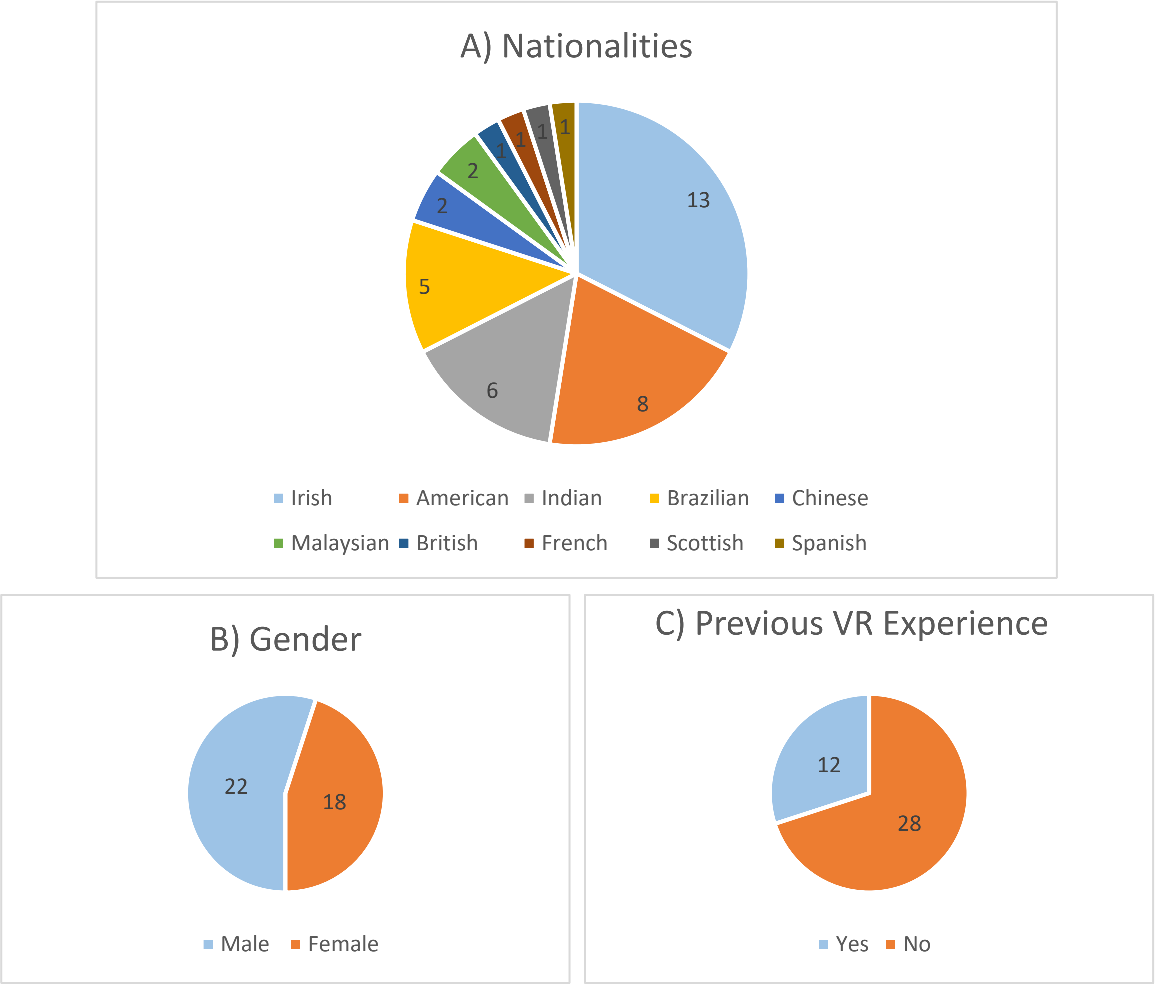


Figure 4.1: Figure showing info related to the participant population. A) Nationalities of the participants., B) the gender divide of the participants and C) Whether or not the participants had previous experience with VR before undertaking the test

There was a diverse range of nationalities represented within the sample. In terms of subject familiarity with the technology and subject at hand, none were familiar with the operation of a Fanuc Injection Moulding machine. Of the 40 participants 28 had never used VR before, whereas 12 were familiar with the operation of an HTC Vive or other similar VR HMD device. This does indicate that the results could show a bias of first-time novel use of an immersive technology within the sample size. A visual breakdown of the population can be seen in Figures 4.1.A, 4.1.B and 4.1C.

4.2 Statistical Analysis

As all subjects were tested individually and no data from any one subject was considered to influence the data from another. On the same note, subjects were not selected based on any other variable via a convenience sampling. Since the bearing of age, gender or experience with VR was considered as a filtering variable, the analysis was conducted under the assumption that the data was independent.

As we were performing analysis on the different between multiple groups (>2 groups) and normality tests (Kolmogorov-Smirnov and Shapiro-Wilk) were performed to see if parametric or nonparametric analysis was appropriate, the two main statistical tests performed were OneWay ANOVA and Kruskal Wallis H tests.

4.3 Explicit Results

Table 4.1 and Figure 4.2 present the results of the ACR MOS data captured for those who participated in this test for delay values of 0, 1, 2 and 3 seconds. Participants were asked to complete the 10 questions from Table 3.1 on a 1-5 Likert scale. This served to form an overall MOS for each subject, indicating the overall perceived QoE of the VR environment. The MOS was then tallied and averaged by delay group. For those who experienced no delay, the overall average MOS for the entire group was 4.33, those who experienced 1s of delay had a MOS of 4.17, those who experienced 2s of delay had a MOS of 4.24 and those with a 3s delay had a MOS of 4.32. Overall, MOS remained consistently high for all groups, remaining within the range of 4-5. This indicated that overall, even when considering the extreme delay value of 3s, all groups perceived the environment positively.

Population standard deviation in the groups remained similar between each, most questions showed a deviation between 0.5 and 0.7 with the closest agreements for Question 4 (“Using the system was natural”) in the 2s delay group and Questions 1 (“I was immersed in the virtual environment”) and 10 (“There was a delay in the virtual environment”) in the 3s delay group. Large standard deviations can be seen for Questions 6 (“I was not restricted in my movements”) and 8 (“The task was not easy to complete”) in the 1s delay group, indicating a measure of disagreement between the participants reporting. Question 6 showed a particularly high variance as it was the only question in testing where a 1 was scored on the Likert scale indicating active displeasure from a subject. While this was an outlier in terms of the mean of this question, it was determined not to be reporting error by the subject at hand and was considered in the results.

Table 4.1: Subjective Questionnaire Results with statistical significance values (p<0.05). Sig. is the significance found using the Kruskal Wallis H test and df is the degrees of freedom (Sample Groups N-1) Standard deviation based on Population.

	No Delay		1000ms		2000ms		3000ms		Total		
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	df	Sig.	Kruskal-Wallis H
Question 1	4.5	0.52704628	4.4	0.51639778	4.4	0.6992059	4.1	0.316227766	3	0.850	0.796
Question 2	4.1	0.56764621	4.3	0.67494856	4.2	0.91893658	4.7	0.483045892	3	0.163	5.129
Question 3	4.1	0.73786479	3.8	0.91893658	4.2	0.63245553	4.1	0.737864787	3	0.776	1.104
Question 4	3.9	0.56764621	3.8	0.78881064	3.9	0.31622777	4.1	0.567646212	3	0.677	1.523
Question 5	4.4	0.96609178	4.5	0.97182532	4.6	0.51639778	4.7	0.674948558	3	0.821	0.918
Question 6	4.2	0.78881064	4.1	1.28668394	4.3	0.8232726	4.4	0.699205899	3	0.958	0.311
Question 7	4.4	0.6992059	4	0.66666667	4.2	0.63245553	4.4	0.516397779	3	0.444	2.680
Question 8	4.4	0.6992059	4.3	0.8232726	4.2	0.91893658	4.5	0.707106781	3	0.879	0.677
Question 9	4.6	0.6992059	4.2	1.22927259	4.5	0.70710678	4.4	0.699205899	3	0.875	0.691
Question10	4.7	0.48304589	4.3	0.67494856	4	0.66666667	3.9	0.316227766	3	**0.013**	**10.821**

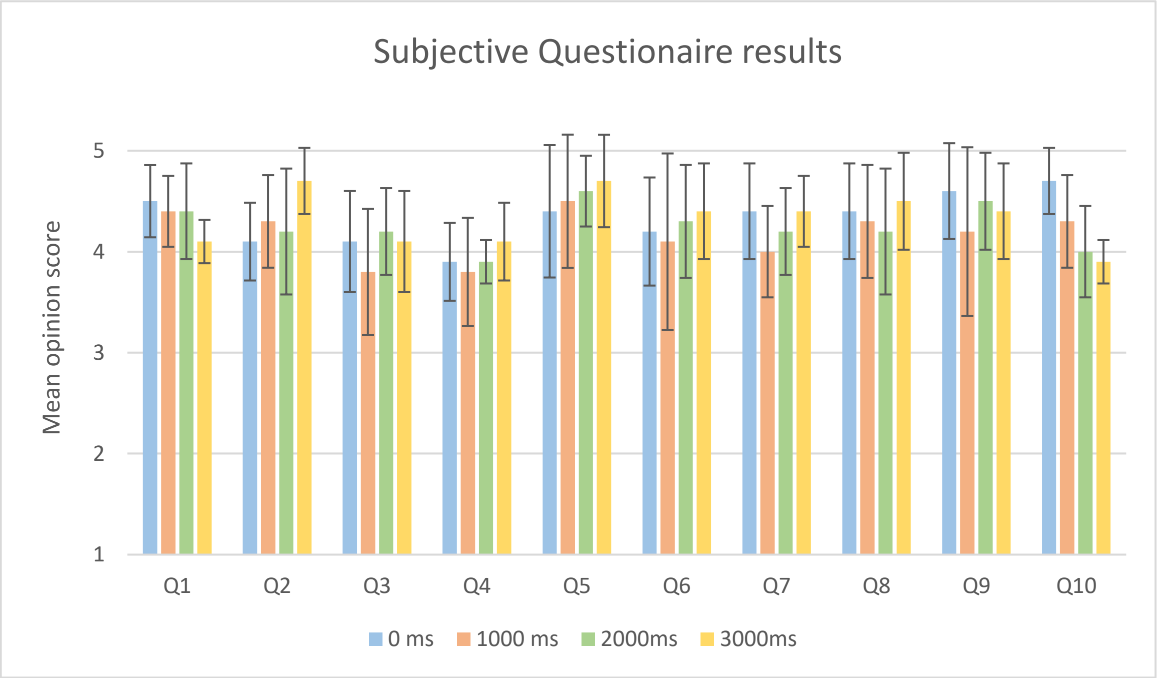


Figure 4.2: Visual Representation of user submitted answers in MOS on a 1-5 Likert Scale (Confidence level = p < 0.05)

Using tests for normality, the data was considered to be non-parametric or non-normal and as such, a Kruskal Wallis H test was performed to analyse this set of data. Question 10 ("There was a delay in the virtual environment") was found to be statistically significant, trending downwards as delay increased. No other questions showed a level of statistical significance when compared to the magnitude of delay. Closer examination revealed that participants answered lower on this question the more that delay increased, indicating being consciously aware of delay in environmental response to actions during test tasks. As indicated by Figure 4.3, mean dropped before tapering off. This trend of participants noticing a network delay as the magnitude increased was to be expected, however an unexpected outcome of the result was the fact that very few participants indicated that this delay was annoying and in fact may have felt that the delay did not inhibit their overall QoE as indicated by the high MOS in this question. (max 4.7 for 0s delay, min 3.9 for 3s delay).

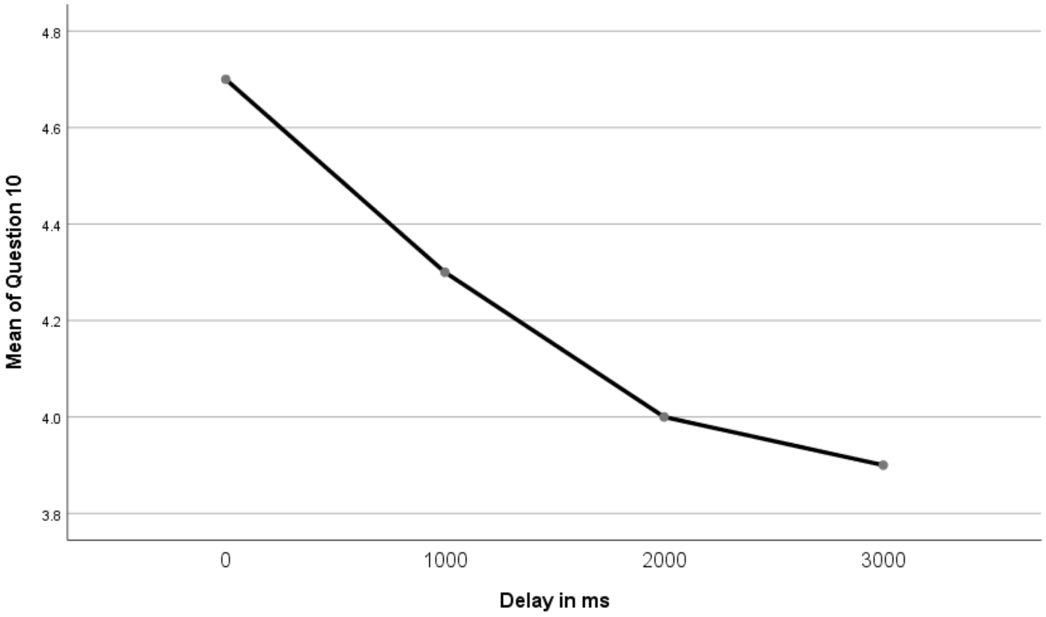


Figure 4.3: Means plot showing average answers for Q10

On average, the question with the lowest reported score by participants was Question 4 ("The environment felt natural"). This indicated that participants felt the system was somewhat unnatural and not correlative to real life, though no user explicitly indicated that the environment felt definitively unnatural (average MOS 3.925 for Question 4 across all groups). Question 5 ("I did not enjoy interacting with the virtual environment") had the most positively responded result. Even at the highest delays the response to this question was very good (average MOS 4.55 across all groups). Overall, when participants were asked to evaluate their overall experience of the environment, the response was very positive (average MOS 4.27 across all groups). Participants felt that the environment was satisfactory, with most positive responses occurring in the areas of immersion and interaction. Usability evaluations tended to fluctuate more as more delay was introduced to the environment. However even at the highest value of 3s, responses remained largely positive.

Overall, this study observed that even at higher measures of network delay within the virtual reality environment, participants did not feel that their overall experience of the environment was impacted to a noticeable degree. Though minor dips in QoE in certain aspects of the study were observed, for the most part participants were happy with their experience and responded positively in explicit evaluations of the system. This was unexpected as the initial hypothesis for this study stated that participants QoE would display much more noticeable drops at higher values of network delay. We also expected initially before this research, that lower values of delay (<1s) would result in a noticeable effect on user QoE. However, some participants even reported no perceived delay in the environment when responding to the questionnaire, even at network delay values of 2s.

Discussion of delay impact on subjective evaluation.

One of the major outcomes of this experiment was the observation that delay did not have a significant effect on a user's reported QoE.

In accordance to our research question, "*How does network delay impact a user's Quality of Experience of the application?*" it was hypothesized that network delay in the environment would have negative effect on overall reported explicit MOS. However as evidenced, this effect was not perceived and instead participants reported a strong enjoyment of the experience across all delay magnitudes, with no significant trend in positive or negative direction for MOS as indicated by high scores across the board in all questions. Question 10 did support this initial hypothesis on some level, with a trend of lowering MOS as delay increased. However, despite this, the magnitude of the effect was far less than anticipated with users remaining high in average enjoyment even at the most extreme levels of our tested delay (3000ms).

A proposed reason for the minimal impact of these delay magnitudes is the system architecture used for the virtual reality environment and the general design principles for virtual reality environments. Network delay affects a VR environment in a different way to input delay or motion delay. Input and motion delay affect the direct tracking of a user's movements on a local level in the virtual reality simulation, thus causing simulator sickness through loss of synchronization of the expected movements made by the user with the actual movement within the environment. However, network delay, especially in the context of a digital twin teleoperation design, does not affect user movement and tracking. Instead delay is localized to commands and operations performed on the machine within the VR environment. It is only when participants operate on the task at hand that delay becomes a factor. This is an intentional aspect of VR network design in third-party tools and development kits, intended to minimize simulation sickness by handling as much object motion tracking and user movement on the local client as possible. These design principles aim to reduce user discomfort as much as possible and so it can be suggested that participants found network delay more acceptable and were willing to overlook its impact. In addition, network delays in each group were constant throughout the experience. In real world terms, network delay would vary over time (jitter) and this may produce a stronger effect on the experience of a system like the one used in these experiments.

4.4 Implicit Results

Due to the limitations of purely subjective evaluations of QoE through surveys and questionnaires, especially in post-experience evaluations, it has become important to assess QoE from perspectives other than just an explicit manner. Thus, inspired by other examinations of QoE and developing methodologies [61] [62], this research examined how a user's basic biological emotion indicators looked when compared to the user-reported metrics.

As part of formulating a more robust model of QoE, biological signals were collected to assess a user's emotional response based on these signals. The physiological metrics considered during this test were heart rate and EDA, both of which were captured using an Empatica E4. The following sections describe the average heart rate and EDA experienced by each group over time, when transitioning from baseline resting measurements to the experimental phases of the study.

4.4.1 Heart Rate

From Table 4.2 and Figure 4.4, heart rate increased when transitioning from the baseline resting phase to the testing phase. All participants were within the range for average human heart rate, in beats per minute (bpm), when measured at baseline and no participants reported any health issues or previous history of heart problems. In addition, the increase in heart rate is displayed as a percentage by comparing the means of each individual as they transitioned between the baseline phase and the testing phase.

An ANOVA analysis was undertaken to examine the effect of delay on heart rate over time as participants progressed through the experiment. Heart rate increased as participants progressed through the experiment and this change was generally an increase of 9.31%. All measured heart rates came within the average range of human heart rates (60-100 bpm) and no one group displayed a noticeably higher rate of heart rate increase than any other. This indicates that, even though a certain level of arousal and excitement was experienced as participants progressed through the test, delay values had no major impact on this increase. Instead, increase in heart rate was more likely to be attributed to other factors, such as experiencing an immersive environment for the first time, stresses related to task difficulty or other environmental influences.

When heart rate was examined over time, certain patterns could be observed. The experiment was split into 10 slides (normalised measures of time frames) to gain a view of physiological changes over the course of the experiment. This was based on how data was presented in [62]. Slides 1-3 represent the baseline phase where participants were asked to wait to allow for their physiological metrics to return to their average state. Overall, heart rate began somewhat high and declined to a resting level within these slides. This is attributed to subject apprehension during introduction then returning to normal values during the baseline phase. Slides 4-10 then represent the remaining portion of the experiment. Due to the varying duration of the experiment, which was dependent on each individual, there is no definitive slide to pinpoint the transition between the training and testing phases. However, on average, training phases lasted from Slide 4 to Slide 5 or 6, with the testing phases lasting from Slide 5 or 6 onward. Within these slides a spike in heart rate was observed, which remained high throughout the experiment and began a decline towards the end. However, heart rate deviation showed no significant spike when related to delay. All groups showed a roughly equal increase in heart rate so while it can be concluded that subjects demonstrated a state of arousal while engaging with the environment, the effect of which cannot clearly be attributed to variability of delay. When considered alongside the high reported MOS scores, it is more likely to infer that subjects experienced more excitement or enjoyment during the test which lead to higher heart rates.

Table 4.2: User heart rate and percentage change in each phase of the experiment with significance ($p<0.05$) of network delay impact found during analysis. F is the level of variability between groups according to the F-test, whereas Sig is the p value (or statistical significance) of the variable at hand.

	No delay		1s delay		2s delay		3s delay			
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	F	Sig
Base (bpm)	84.278	11.3434	73.404	10.0015	81.029	15.63377	76.353	11.99671	1.513	0.228
Test (bpm)	90.527	10.7244	80.603	11.4041	88.602	18.73671	83.404	12.92444	1.284	0.295
Deviation (%)	8.075	11.06912	10.514	13.03287	9.161	7.258142	9.499	6.650969	0.104	0.957

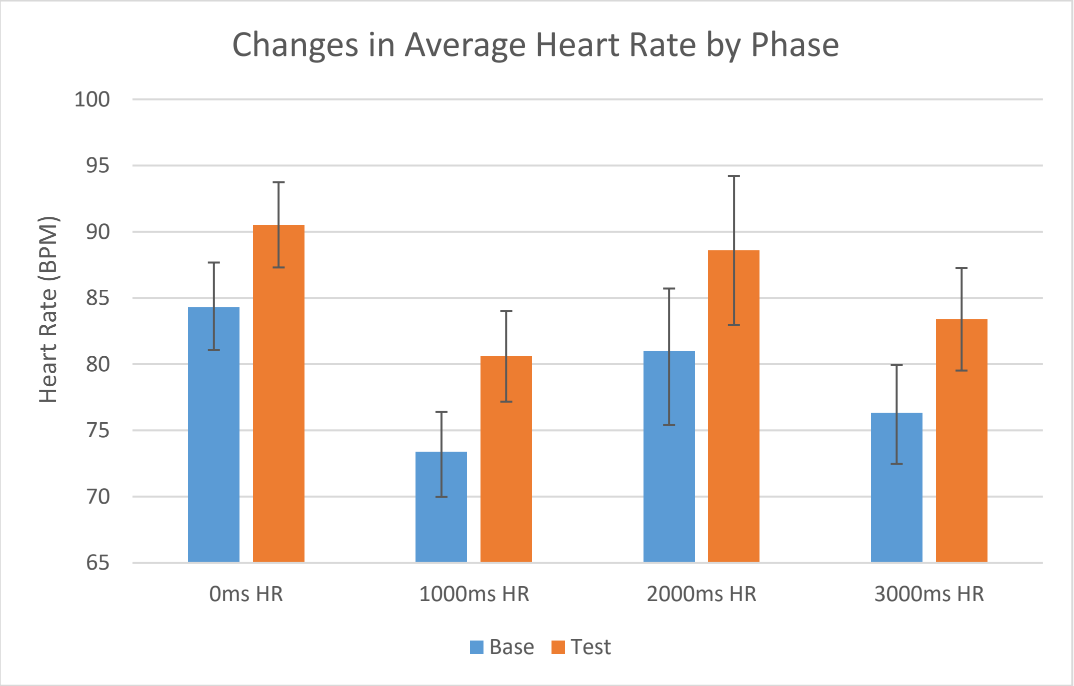


Figure 4.4: Change in average heart rate when comparing baseline average measure to the average measure in the testing phase. Lighter colours represent baseline, darker colours represent testing phase. (Confidence level = $p < 0.05$)

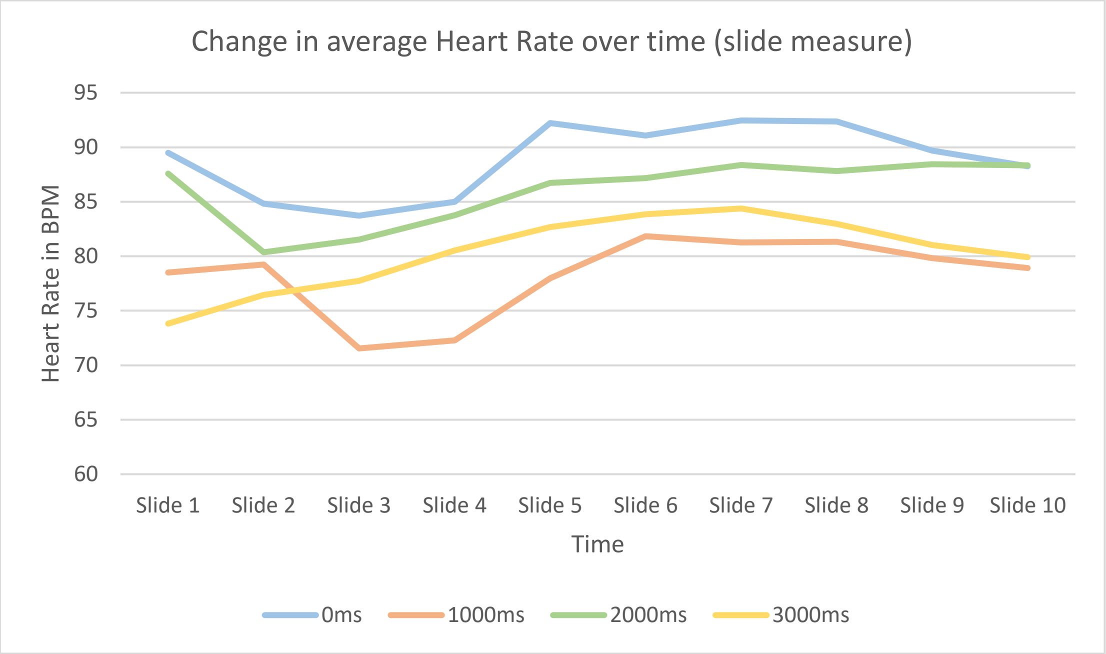


Figure 4.5: Change in heart rate over the course of the experiment. Baseline phase occurred during Slides 1-3, with training and testing occurring across slides 4-10.

4.4.2 Electrodermal Activity

Average user electrodermal activity (EDA) was measured and displayed in micro Siemens (μS) as it represents a way to indicate change in mood or emotion over extended amounts of time. The average user deviation of EDA is displayed as a percentage factor in Table 4.3 and visually in Figure 4.6 and Figure 4.7. Upon tests for normality, EDA data was found to be non-normal and so a Kruskal Wallis H test was used for statistical analysis

No statistical significances (Sig. = $p < 0.05$) could be found when comparing groups by the level of delay. Participants EDA tended to rise as they moved from the baseline measurement to the testing phase, similar to heart rate. As indicated by previous work on EDA as a measurement of emotional response, this rise in EDA possibly resulted from stress or arousal as participants proceeded through the experiment [61] [67].

When examining the change in EDA as a percentage factor, larger increases can be seen when compared to heart rate and the average values tends to vary to a much higher degree. For the most part, individuals tended to show an increase of around 80-120% in EDA. In addition, the 2s group contained two outlier participants who showed a larger than average spike in EDA with an increase of up to 3 times the recorded resting level, possibly due to physical health condition or environmental factors (some tests were conducted during the summer months). This can be observed in a much greater confidence interval at a significance level of $p < 0.05$, than other recorded groups.

Table 4.3: User EDA and percentage change in each phase of the experiment with the significance ($p<0.05$) of network delay as an impact. Standard Deviation calculated from Population.

	No delay		1s delay		2s delay		3s delay				
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	df	Sig	Kruskal-Wallis H
Base (μS)	1.5	1.7393677	3.052	2.4256032	1.892	3.0769061	1.488	1.3270502	3	0.259	4.026
Test (μS)	3.398	4.1339806	5.098	4.3710753	4.564	6.3346336	2.925	2.5373882	3	0.797	1.017
Deviation (% Increase)	111.626	129.09957	49.225	75.51547	340.832	658.29449	130.718	197.24755	3	0.919	.498

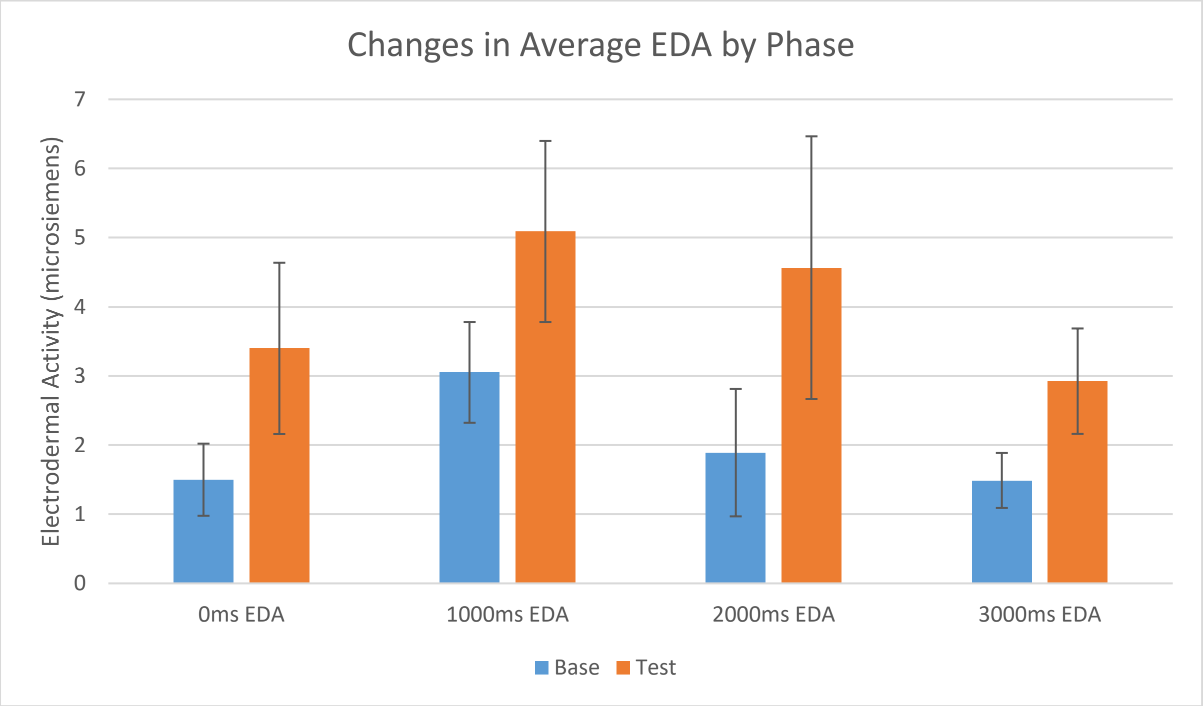


Figure 4.7: Change in average EDA when comparing baseline average measure to the average measure in the testing phase. Lighter colours represent Baseline, darker colours represent Testing phase. (Confidence level = $p < 0.05$).

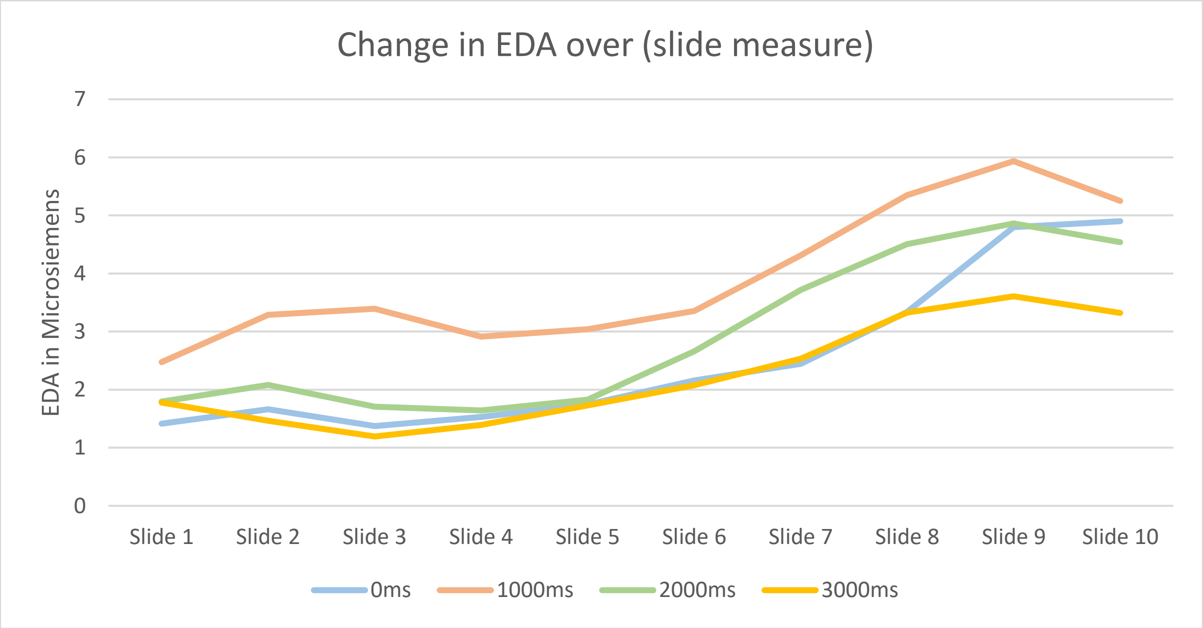


Figure 4.6: Changes in Average EDA over the course of the experiment by slide measure. Slides 1-3 represent the baseline phase where metrics could drop to resting state. Sides 4-10 represent the rest of the experiment where participants completed the task.

During the study, the most regularly observed trend in physiological values was that participants EDA and heart rate both rose noticeably from their measured resting state. Previous studies [67] [69] into the evaluation of physiological signals as emotional indicators, such as heart rate and EDA, suggest that heightened emotional state comes with an increase in activity from these metrics . Heightened heart rate indicates excitement or arousal, as blood vessels expand to allow more oxygen to enter the bloodstream. Heightened EDA can indicate a certain level of stress and frustration, as well as overall arousal, in any emotional situation where increased sweat production might lead to a more conductive galvanic skin response. This author is confident in claiming that participants who participated in the experiments for this thesis experienced a heightened emotional state in some regard. By examining this alongside the explicit reported values, there is a strong indication that many participants who participated in this study experienced some level of increase in excitement or enjoyment. This is generally consistent with the observation that participants who had never used VR before would be more immersed and engaged with a novel technology. However, this increase in heart rate was also observed among participants who had used a virtual reality system before and so the effect cannot be purely explained by novel interaction.

4.4.3 Interactions and simulation variables

As part of the evaluation of the groups, participants interactions were measured and recorded. These interactions include rate of trigger clicks and controller events and the time to task completion. Table 4.4 and Figure 4.8 the difference in aggregate controller interactions between groups can be seen. Normality tests showed that the data was non-parametric and so a Kruskal Wallis H test was used to examine the data. When examined at a confidence level of $p < 0.05$, a significance was found in the effect of delay on total interactions with a stronger effect on the use of the touchpad.

Table 4.4: User controller interactions while in testing phase. Clicks and touchpad presses measured in total number of recorded interactions averaged across the entire group. Sig = a<0.05.

	No delay		1s delay		2s delay		3s delay				
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	df	Sig	Kruskal-Wallis H
Trigger	23	11.19524	27.9	13.48621	30.4	12.63329	19.5	7.677529	3	0.093	6.411
Touchpad	14.2	5.329165	22.2	10.96256	29	16.77962	17.8	9.461031	3	**0.011**	11.132
Total	37.2	12.35404	50.1	12.35404	59.4	22.63822	37.3	16.77333	3	**0.055**	7.617

As illustrated in Figure 4.8, though the touchpad was more influenced by the network delay, trigger presses followed a very similar pattern of change. On average, participants who were subjected to a delay tended to click more often. In addition, those who were subjected to the delay showed a higher variance in their clicks and touchpad presses. This indicates that participants tended to click more often and more rapidly. This could be attributed to higher frustration as the environment was slow to respond to interactions as well as participants re-clicking as they did not receive immediate visual or audio feedback. A trend from 0s delay to 2s delay is exhibited in the number of interactions performed by a user during the test. Within this interval participants tended to click buttons more and press the touchpad more. This could indicate a level of erroneous pressing of a button or impatience with the irresponsiveness of an action.

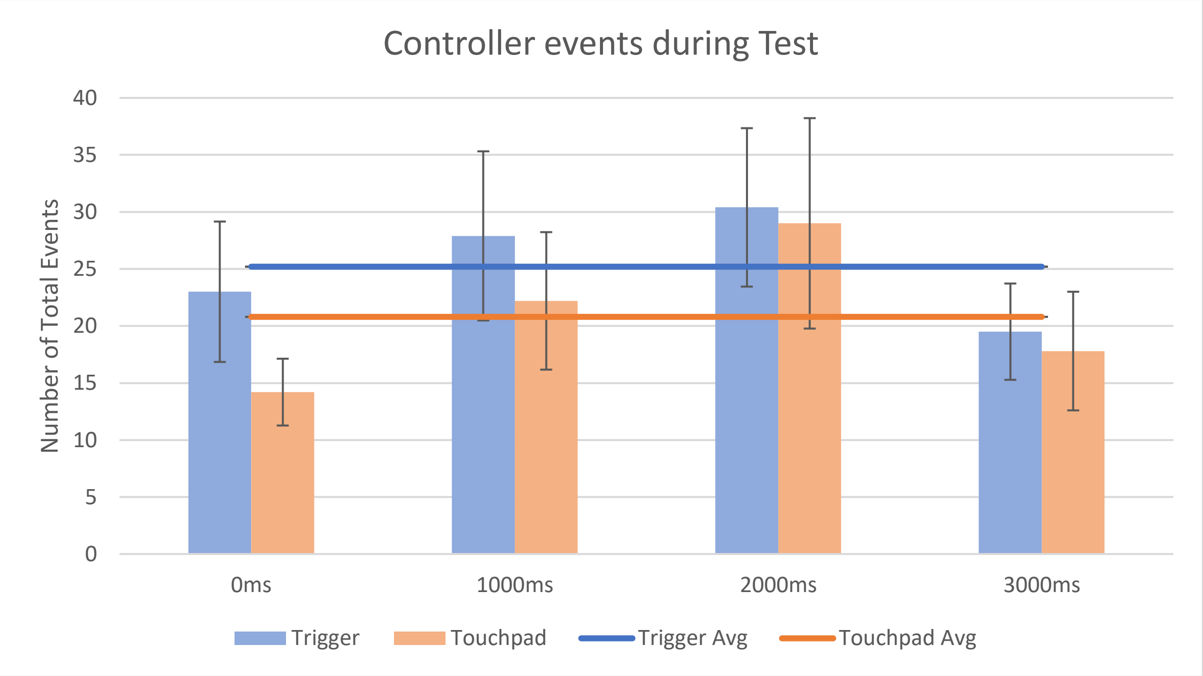


Figure 4.8: Visual Representation of user interactions. (Confidence level = $p < 0.05$). Trigger Avg. and Touchpad Avg. are the average value of each respective variable when averaged across all subjects in all groups

However, when delay was increased to 3s there was a decrease in the click values generated in the test. The 3s group also showed a clear drop in standard deviation of clicks and touchpad presses. In addition, higher deviation was seen in the 1s and 2s delay groups, illustrated in Figure 4.9. Here we can see that the 1s and 2s delay groups had several higher outlying data points where participants exhibited a higher variance of controller interactions.

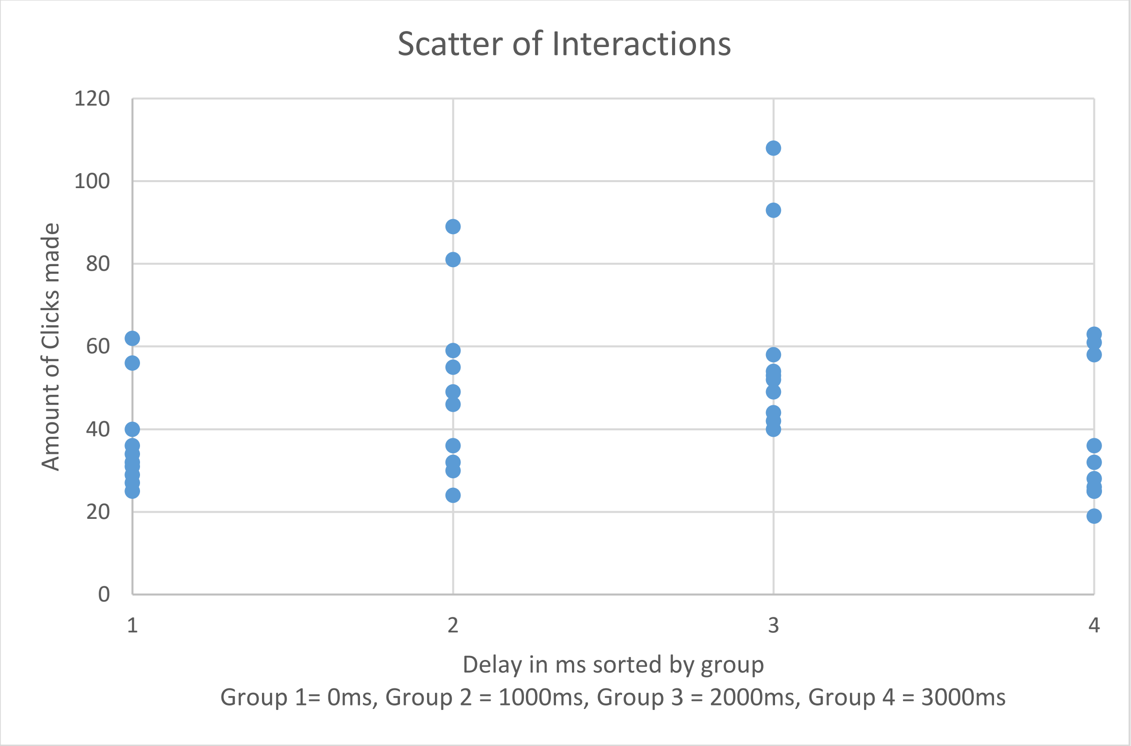


Figure 4.9: Scatter plot of total number of interactions made by participants during the testing portion of the experiment.

From the scatter in Figure 4.9, a more indicative trend of increasing controller click can be seen with tighter groupings and less variance in both the no delay group and the 3s group. Possible interpretations of this trend could suggest that there is a threshold beyond which participants begin to become aware of the more obvious delay and adapt easier to more noticeable delay patterns, while a lesser delay is enough to cause misclicks and re-clicks from possible frustration and not immediately noticing the delay. The 3s delay group did not have a higher ratio of participants who had previously used VR when compared to other groups, with only 3 out of the 7 participants in this group having had experienced a virtual environment before this experiment.

Table 4.5 illustrates the difference between results when accounting for previous user experience in VR. For the most part, participants who had used VR before and were familiar with a HMD tended to click less frequently and make more deliberate actions. However, the total population taking part in the study was heavily biased towards those who had not used VR before, as mentioned before in the population overview in Section 4.1, so any inferences to be made from this must take that into account. Acknowledging this, the trend does indicate that those who have used VR before were less likely to input controller interaction events multiple times in quick succession.

Table 4.5: Comparison of VR experience with number of clicks made. (Sig. = p <0.05).

Group Statistics					
	Has Used VR?	Mean	Std. Deviation	F	Sig.
Trigger	No	25.68	13.419	4.216	0.047
	Yes	24.08	7.166		
Touch	No	20.64	14.038	1.105	0.300
	Yes	21.17	7.209		
Total	No	46.29	23.188	3.426	0.072
	Yes	45.25	12.396		

One trend noticed was that participants tended to have difficulty with certain aspects of the test, with the steps of the experiment involving moving the injection unit into the correct place proving the most challenging. This trend peaked with the 2s user group, which also contained the highest number of outlying click and press instances. However, when compared to questions in the questionnaire designed to test the user's perception on controller ease of use, most participants responded that the system was easy to use and interact with. It was intended that the task presented to participants was largely simplistic and not designed to introduce overly complex mechanics to the environment. It is suggested that the simplistic nature of the point-and-click remote control design may have overshadowed any explicit annoyances that delay might have introduced.

Discussion of delay impact on interaction with the simulation.

For the most part, it was identified that as network delay was increased, participants would make controller interactions more frequently. However, when the delay of the highest value used (3s), a noticeable drop in controller interaction was observed, showing interaction values similar to the no delay group. Additionally, it was shown that the variance and deviation of clicks also tightened, indicating that re-clicks happened less often. Examination of this trend has led to the conclusion that when delay increases past the point where it becomes noticeable, even by those who had never used VR before, participants would begin to compensate behaviourally for the delay in their actions. In future, it would be recommended to perform a closer examination of this phenomenon, as the results of this study do indicate some relationship between network delay and user interaction with a VR environment.

When observing the relationship between increases in biometrics versus controller interactions no clear trends were observed. Though the 2s group experienced both the most extreme increase in biological signals, as well as the most frequent number of interactions, closer examination of the results indicated that no correlations could be found between these two parameters. While this could show that physiological signal increases are not a good indicator of a user making mistakes or making multiple clicks, which can be inferred to show frustration, the interaction metric also tended to be the metric most prone to outlying data.

4.5 Eye Tracking

Eye tracking metrics were analysed using a few methods. Firstly, gaze data was filtered and organised by fixations and saccades according to a set universal velocity threshold based on the context of our virtual environment (40 deg/s). Then any grouped fixations were ordered by fixation group indices and the area of interests those points lay in. This allowed for a broader examination of the available eye gaze data. Areas of interest were defined as the instructions, the face, and the console. The instructions refer to the instruction panel that guides the participants through the experiment. The face refers to the face of the Fanuc where most of the

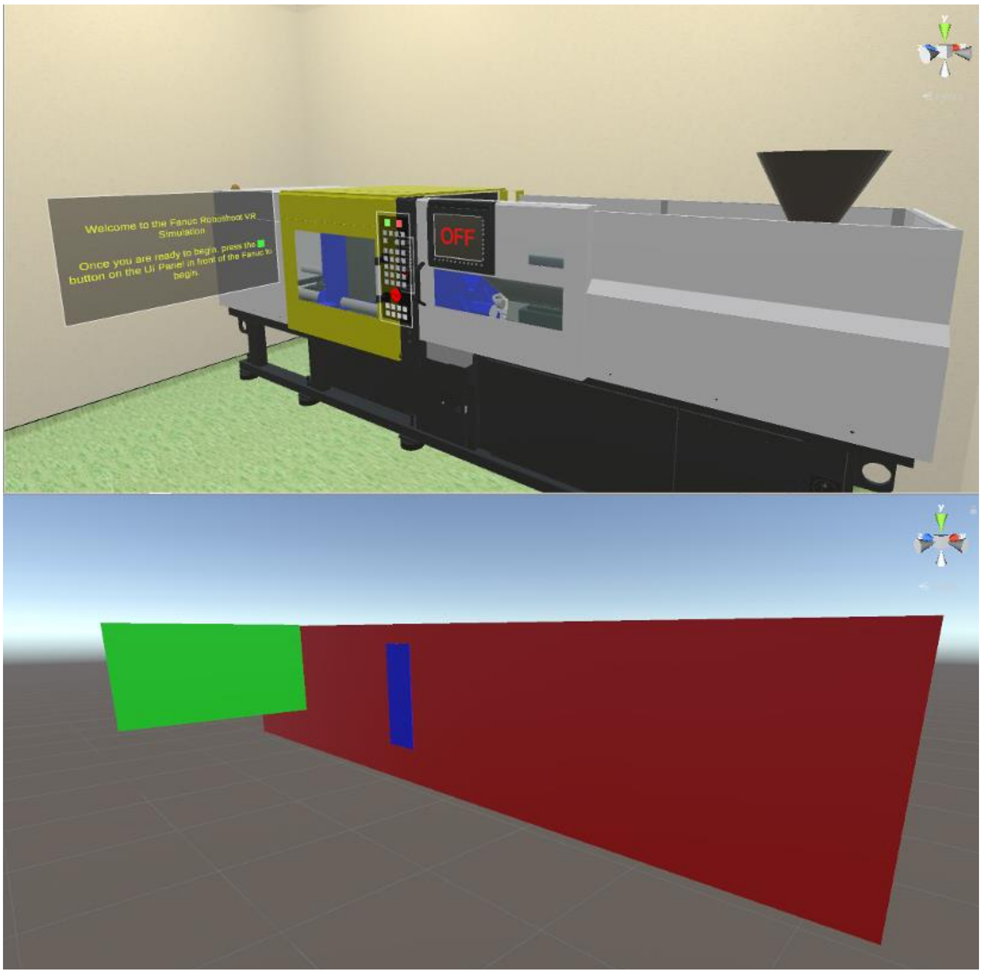


Figure 4.10: (Top) View of what participants can see in virtual environment and (Bottom) Area of Interest colliders showing the areas of interest selected as important: Instructions (Green), Face (Red) and Console (Blue)

moving parts of the machine could be observed. Finally, the console refers to the interface panel where the buttons to control the operation of the Fanuc were located. A visual representation of this can be seen in Figure 4.10.

Table 4.6: Statistical Results from Eye Gaze Points data. Eye gaze measured in total recorded gaze raycast hits averaged across all groups. Fixations and fixation groups measured in number gaze raycasts that fell within the velocity threshold.

	No delay		1s delay		2s delay		3s delay			
	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	Mean	Std. Dev	F	Sig.
Total Gaze	19421.70	9537.431	22975.30	9919.181	22296.60	8490.020	20141.20	2629.31	0.429	0.733
Fixations	13013.50	7089.157	14203	8243.507	14097.3	7645.239	12521.4	1656.984	0.152	0.928
Saccades	6408.20	3044.099	8772.30	2105.317	8199.30	2330.995	7619.80	1773.877	1.834	0.158
Total Fixation Groups	965.00	551.6225	1139.9	370.1851	1030.1	370.7606	993.7	129.5111	0.394	0.758

From results presented in Table 4.6, network delay in these experiments had no statistically significant influence on eye gaze behaviour. However, there was a weak influence on the total number of saccades. In addition, the no delay group tended to display more variance in results between individuals. This is also reflected in the variance and standard deviation of the 3s group. Delay groups also showed higher number of fixation groups while individual gaze points remained largely the same. This meant that fixations in groups with more delay tended to be smaller and of less duration. From this it can be suggested that the delay caused participants to demonstrate more erratic gaze behaviour.

Table 4.7: Correlation between area most fixated on and delay

		Delay in Ms	Area most fixated on (Point)
Delay in Ms	Pearson Correlation	1	.389*
	Sig. (2-tailed)		0.013
Area most fixated on (Point)	Pearson Correlation	.389*	1
	Sig. (2-tailed)	0.013	

As seen in Table 4.7, a correlation was found between delay and what area of interest participants were most fixated on over the duration of the experiment. Participants in the no delay group tended to focus much more on the instruction panel. As network delay was introduced, most fixations were observed on the console and face areas, with less focus on the instruction panel. This could indicate that as delay was introduced to the environment and participants actions were impeded, participants tended to focus more on the buttons they were pressing and waited for the visual and auditory feedback before breaking focus.

Table 4.8: Area most fixated on by each subject ordered by group.

Subject Number	No delay	1s delay	2s delay	3s delay
Number 1	Instructions	Instructions	Instructions	Instructions
Number 2	Instructions	Console	Console	Instructions
Number 3	Instructions	Instructions	Console	Console
Number 4	Instructions	Console	Console	Console
Number 5	Instructions	Instructions	Face	Instructions
Number 6	Instructions	Console	Console	Console
Number 7	Face	Face	Face	Console
Number 8	Instructions	Face	Instructions	Face
Number 9	Instructions	Instructions	Instructions	Instructions
Number 10	Instructions	Console	Face	Console

From the analysis of the eye-tracking data, while no significant results (Sig. = $p < 0.05$) were found as part of the velocity threshold analysis, some trends in the data were observed as an influence of delay. Participants showed greater saccadic behaviours and fixation groups tended to be smaller and more frequent, indicating shorter fixations with more saccadic actions between them. From established behavioural research on eye tracking [72] [73], this can be inferred to mean that participants were possibly more frustrated or stressed, or possibly exhibited a heightened awareness state or excitedness. Inversely, the longer gaze durations and reduced saccadic behaviour of the no delay group would suggest a more relaxed and calm response. In addition, area-of-interest analysis showed that the higher the delay value, the less

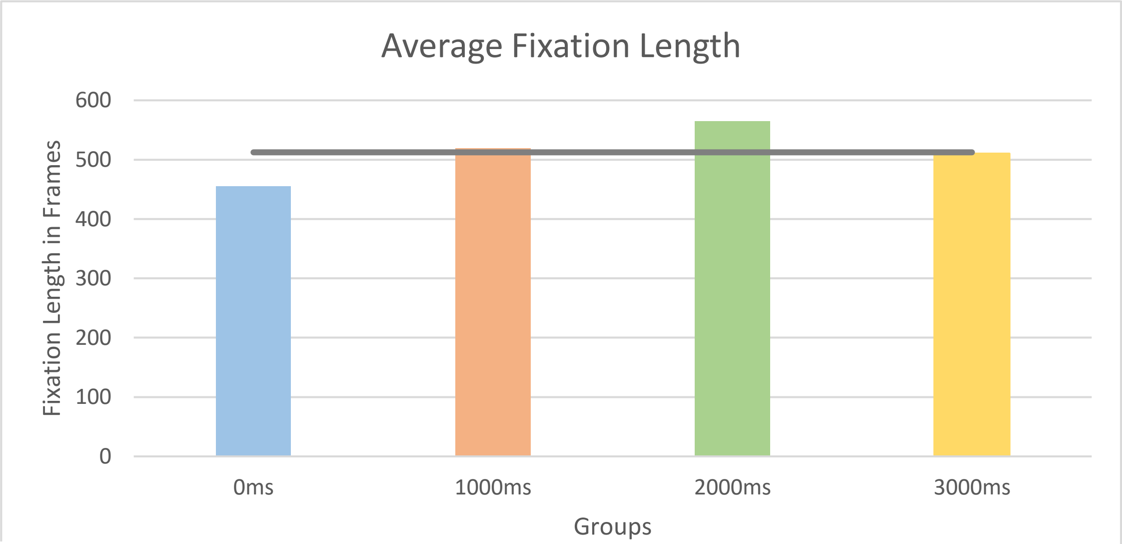


Figure 4.11: Average duration of a fixation measured in frames per second (recorded at 60fps). (Confidence level = $p < 0.05$)

focused on one area the participants would be. The no delay group showed a near unanimous fixation on the instruction panel, however the more network delay increased, less agreement between participants was observed. Participants in the 3s delay group were much more likely to focus on the control panel during the experiment, suggesting a focus on the lack of responsive input or a tendency to wait for auditory or visual feedback from the virtual environment.

Figures 4.11 shows the average duration of a fixation in a test. On average, small fixations tended to last for around 27 frames of measurement, which at 60 frames a second, results in fixations that lasted approximately half a second. The average duration of long fixations was around 512 frames of measurement, or around 8.5 seconds. During statistical analysis (ANOVA) of the impact of network delay on these values, no significant value (Sig. = $p < 0.05$) could be found. When comparing the values of long fixation duration and the number of interactions made during a test, a weak correlation was found, as illustrated by the similar patterns of increase and decrease seen across delay values. However, the effect seen is much less pronounced in the case of fixation duration. A possible explanation for this pattern is that participants fixated on a particular element and made more than a single click or interaction on that element. One interesting outcome is that the margin of error in gaze patterns for the 3s delay group is much smaller than other groups. This suggests that participants exposed to extreme measures of delay showed similar patterns of gaze. Like the interactions, it could be an indication that participants may have adapted to the extreme value of delay or that participants fixated more on individual elements in the environment while waiting for visual or auditory feedback.

Eye-Tracking as a metric to capture QoE has been explored in other areas [66], however in immersive technology research, especially when considering virtual reality, evaluations using eye-tracking are virtually non-existent. This is largely due to a number of factors, including the difficulty of including eye-tracking as a separate metric in any study, the absence of concise and standardized eye-tracking design principles for VR examinations, as well as technological factors such as HMDs covering a user’s eyes completely and the unavailability of tools for use in VR environments. Using a Tobii Pro and the SDK supplied within, in the developed system a user’s gaze could be captured by the custom HMD, the gaze direction translated to Unity co-ordinates and these co-ordinates stored for use in various evaluations and interactions. One such use for this was the creation of heatmaps in a 3D virtual space. To date, to the author’s knowledge, no study implementing 3D virtual reality eye tracking as an evaluation metric has been published. Further development of this framework could serve as a highly valuable methodology for capturing QoE for VR.

4.6 Summary

In this section we presented and analysed our results that we recorded over the course of this work. Overall, participants were found to have enjoyed their experience with the environment, they did not feel that network delay had a major negative impact on their experience of the environment, and they did not feel that the usability of the environment was impacted greatly by the presence of delay. In the explicit results we identified that participants did notice a delay at the most extreme ends of our delay spectrum, but they did not report it as annoying or an impediment in the questionnaire evaluation. Though participants did report that the environment felt somewhat unnatural, there was a high reported immersion and enjoyment. Implicitly, from our physiological results we identified that participants did experience increases in both heart rate and electrodermal activity which indicates an increase in mood and arousal. By comparing this to user reported statistics, it can be inferred that this increase in mood was a positive one and users experienced increases in enjoyment and excitement while using the virtual reality environment. Also measured implicitly was number and types interactions with the environment, which showed increasingly erratic patterns the higher the delay experienced, this then dropped off dramatically at the highest levels of delay. This was identified as participants becoming used to the extreme delay supported by participants reporting noticeable delay only at that extreme level. Eye gaze showed no strong patterns of behaviour other that could indicate a change in behaviour based on delay increase. There was a decrease in variance at the most extreme levels of delay mirroring the interaction metrics, however this is still within reasonable difference from other groups. Overall little significance could be found on the impact of network delay on participant QoE on the teleoperation VR environment.

Section 5: CONCLUSION

5.1 Discussion

In conclusion, over the course of this study, certain gaps in the knowledge base of quality of experience evaluations relating to immersive multimedia and its application in Industry 4.0 were identified. To investigate these, a virtual reality simulation of a digital twin teleoperation application was created and developed. A method of applying networking to a VR simulation and controlling network traffic on a client-server architecture was developed. We developed an experimental framework to allow the capture of user QoE metrics was implemented in this context in order to examine the impact that network delay would have on participants.

From the analysis of the results obtained from this QoE capture methodology, participants QoE was largely unaffected by the presence of network delay when performing a simple teleoperation task. The author speculates that this was influenced by factors mentioned in discussion, primarily with system architecture. Other factors such as novelty, experience and task simplicity were also suggested as influences on the overall QoE results. Eye-gaze analysis showed that participants had more erratic patterns of gaze when subjected to higher values of delay, indicating some level of minor increase in cognitive load. participants showed a universal emotional arousal when experiencing the virtual reality environment, and when corroborated with the subjective questionnaire results it can be inferred that participants generally had a high QoE overall. In relation to the research objectives, the goal was to locate a threshold of network delay intensity beyond which a user would consider the VR application was unacceptable for use. The value at which the network delay became generally noticeable to participants was around 2-3 seconds of delay, however, a delay value of 3s was still within the realm of acceptability. The QoE in a virtual reality environment network delay is largely dominated by factors such as exposure to novel technologies, ease of task completion and the level of participant experience with VR. Due to these, even at high network delay values participants will accept latency in virtual reality environments.

Eye gaze evaluations in VR to gauge human emotion, particularly from a QoE perspective, are sparse and so the eye-tracking framework developed as part of this research aimed to help broaden the knowledge in this area, as well as to demonstrate how eye-tracking allows for more robust QoE assessments in virtual reality environments. We also hope that this framework will help contribute towards understanding better practices for implementing eye-tracking into immersive technology research.

5.2 Limitations and Recommendations for Future Work

Limitations

- The experimental results are influenced by an experience bias due to the much higher population of participants who had never experienced VR before (28 no experience vs 12 experienced). There is an indication that technological novelty had a positive impact on QoE throughout the test and that having no point of reference to an environment not impacted by delay would lead to the effect of said delay being diminished among those who were experiencing VR for the first time. To account for this in the future, a degradation category rating QoE design could be considered, whereby participants are presented with an environment where no delay is present to allow for a baseline to be achieved, and then presented with a second environment with network delay.
- During statistical analysis, this relationship was not explored in depth. However, from examination of overall results the impact of this relationship is apparent. Our results show that users experienced very high QoE despite the delay and it was suggested that this was due to excitement from participants experiencing a novel technology for the first time, as well as not being experienced in a teleoperation task of this sort. As such we would propose an examination focusing on this relationship would be worth undertaking.
- Sample size is a limitation to consider for this test. Each group only contained 10 participants in the results examination. This increased the possibility of outside influences on the results, such as extreme outliers in physiological signals, or users rating subjective experience in extremes (All 1's or all 5's). As such, a future study of this type should include a larger subject population.
- Virtual reality eye tracking frameworks are still largely proprietary and third-party tools for eye-tracking are sparse. In addition, it is difficult to implement many of the established research principles for eye-tracking in the context of VR, as most eye-tracking research is conducted in a 2D context (the face of a computer screen), which translates poorly to a 3D context.
- While a simple velocity threshold analysis is easy to implement with a clear output of fixation/saccades, it is susceptible to noise, especially if the velocity threshold selected for measurement is too strict. In addition, while an area-of-interest analysis can give a good indication of where participants were looking in the environment, much

information is lost in terms of individual fixation groups without a concise framework to account for them.

- Initially, experiment design was simplified to minimise the possible points of external frustration for the test, so that network delay could be isolated as an impact factor. This included things like the controller scheme, task complexity and feedback from the digital twin. The simplicity of the task may have made the participants less prone to stress and frustration as perceived ease-of-use overpowered the impact of delay.

Recommendations for Future Work

Immersive technology applications in Industry 4.0 are still a growing area, especially when considering academic research. Many best practices and standards have yet to be defined. It is hoped that this work shall serve to inform future researchers on how QoE evaluations for immersive technologies can be developed. Some potential recommendations for similar work in the future are:

- The relation between user experience with virtual reality and the acceptance of an environment affected by network delay merits further examination.
- There is great potential for eye-tracking as a QoE metric as researchers have a great deal of freedom in transformation and representation of captured eye-tracking data, particularly in the 3D representation of eye tracking data.
- One area this work touched upon that has high potential for future research is that of networked virtual reality environments. For this study, the framework developed was only used to measure static delay in communication between client and server. Investigation of other aspects of poor network connections, such as jitter, are also possible and could be touched upon in future work.
- In future an expanded examination of system interaction metrics should be included. This could include the time between steps of instructions as well as out-of-order pressing (examining optimal map of button presses vs user deviation).
- In future, we would consider increasing the complexity of the task, or adding a timed element to the environment.
- On reflection of the study, further examination, and improvements on the implementation of the eye-tracking framework should be considered, based on the limitations discussed in the limitation section above. In addition, other eye related metrics deserve examination, such as pupillometry and blink rate

REFERENCES

[1] K. Henning, “Recommendations for implementing the strategic initiative INDUSTRIE 4.0.,” 2013.

[2] *Recommendation ITU-T Y.2060, Overview of the Internet of Things. ITU-T Study Group 13,* 2012.

[3] S. Khaitan and J. McCalley, “Design techniques and applications of cyberphysical systems: A survey.,” *IEEE Systems Journal,* vol. 9, no. 2, pp. 350-365, 2014.

[4] R. Baheti and H. Gill, “Cyber-physical systems,” *The impact of control technology,* vol. 12, no. 1, pp. 161-166, 2011.

[5] D. Gorecky, M. Schmitt, M. Loskyll and D. Zühlke, “Human-machine-interaction in the industry 4.0 era,” in *2014 12th IEEE international conference on industrial informatics (INDIN),* 2014.

[6] P. F. Hokayem and M. W. Spong, “Bilateral teleoperation: An historical survey,” *Elsevier Automatica,* vol. 42, no. 12, pp. 2035-2057, 2006.

[7] J. Yan, X. Yang, X. Luo, C. Chen and X. Guan, “Consensus of teleoperating cyber-physical system via centralized and decentralized controllers,” *IEEE Access,* vol. 5, pp. 17271-17287, 2017.

[8] A. Langley, G. Lawson, S. Hermawati, M. D'Cruz, J. Apold, F. Arlt and K. Mura, “Establishing the usability of a virtual training system for assembly operations within the automotive industry,” *Human Factors and Ergonomics in Manufacturing & Service Industries,* vol. 26, no. 6, pp. 667-679, 2016.

[9] J. I. Lipton, A. J. Fay and D. Rus, “Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing,” *IEEE Robotics and Automation Letters,* vol. 3, no. 1, pp. 179-186, 2017.

[10] “VIVE™ | Discover Virtual Reality Beyond Imagination,” HTC Corporation, 29 August 2019. [Online]. Available: https://www.vive.com/eu/. [Accessed 29 August 2019].

[11] "Microsoft HoloLens | Mixed Reality Technology for Business," Microsoft, 25 July 2019. [Online]. Available: https://www.microsoft.com/en-us/hololens. [Accessed 29 August 2019].

[12] J. Kurose and K. Ross, Computer Networking: A Top-Down Approach, 7th Edition, Pearson Education Limited, 2017.

[13] C. Cavaro-Ménard, Z. G. Lu and P. Le Callet, "QoE for telemedicine: Challenges and trends," *Applications of Digital Image Processing XXXVI,* vol. 8856, p. 88561A, 2013.

[14] D. Zhang, Y. Xu and C. Cheng, "A QoE assessment system in distance education," *Engineering - Scientific Research Publishing,* vol. 3, no. 1, p. 90, 2011.

[15] J. E. Otto and J. B. Ritchie, "The service experience in tourism," *Elsevier Tourism management,* vol. 17, no. 3, pp. 165-174, 1996.

[16] T. Ebrahimi, "Quality of multimedia experience: past, present and future," in *MM'09: Proceedings of the seventeen ACM international conference on Multimedia*, ACM, 2009, pp. 3-4.

[17] R. C. Streijl, S. Winkler and D. S. Hands, "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives," *Springer Multimedia Systems,* vol. 22, no. 2, pp. 213-227, 2016.

[18] R. Rajkumar, I. Lee, L. Sha and J. Stankovic, "Cyber-physical systems: the next computing revolution," in *Design Automation Conference,* IEEE, 2010, pp. 731-736.

[19] D. C. Sol, A. R. Devidas, M. V. Ramesh and e. al., "Design and Implementation of Context Aware Cyber Physical System for Sustainable Smart Building," in *2018 International Conference on Smart Grid and Clean Energy Technologies (ICSGCE),* IEEE, 2018, pp. 162-167.

[20] E. A. Lee, "Cyber physical systems: Design challenges," in *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC),* IEEE, 2008, pp. 363-369.

[21] J. Lee, B. Bagheri and H.-A. Kao, "A cyber-physical systems architecture for industry 4.0-based manufacturing systems," *Elsevier Manufacturing letters,* vol. 3, pp. 18-23, 2015.

[22] M. Stilman, K. Nishiwaki and S. Kagami, “Humanoid teleoperation for whole body manipulation,” in *2008 IEEE International Conference on Robotics and Automation*, IEEE, 2008, pp. 3175--3180.

[23] D. Whitney, E. Rosen, E. Phillips, G. Konidaris and S. Tellex, “Comparing robot grasping teleoperation across desktop and virtual reality with ros reality,” in *Proceedings of the International Symposium on Robotics Research*, 2017.

[24] T. H.-J. Uhlemann, C. Lehmann and R. Steinhilper, “The digital twin: Realizing the cyber-physical production system for industry 4.0,” *Elsevier Procedia Cirp*, vol. 61, pp. 335-340, 2017.

[25] “Google Earth,” Google LLC, 4 September 2019. [Online]. Available: https://www.google.com/earth/. [Accessed 4 September 2019].

[26] D. Egan, C. Keighrey, J. Barrett, Y. Qiao, S. Brennan, C. Timmerer and N. Murray, “Subjective Evaluation of an Olfaction Enhanced Immersive Virtual Reality Environment,” in *Proceedings of the 2nd International Workshop on Multimedia Alternate Realities*, ACM, 2017, pp. 15-18.

[27] “Oculus Rift: VR Headset for VR Ready PCs | Oculus,” Facebook Technologies, LLC., [Online]. Available: https://www.oculus.com/rift/. [Accessed 29 August 2019].

[28] “PlayStation VR | The VR gaming system for PS4 | PlayStation,” Sony Interactive Entertainment Limited, 2019 August 2019. [Online]. Available: https://www.playstation.com/en-ie/explore/playstation-vr/. [Accessed 25 August 2019].

[29] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. O’Brien, V. K. Bansal, D. K. Andersen and R. M. Satava, “Virtual reality training improves operating room performance: results of a randomized, double-blinded study,” *Annals of surgery - Lippincott, Williams, and Wilkins*, vol. 236, no. 4, p. 458, 2002.

[30] D. A. Guttentag, “Virtual reality: Applications and implications for tourism,” *Elsevier Tourism Management*, vol. 31, no. 5, pp. 637-651, 2010.

[31] J.-R. Li, L. P. Khoo and S. B. Tor, “Desktop virtual reality for maintenance training: an object oriented prototype system (V-REALISM),” *Elsevier Computers in Industry*, vol. 52, no. 2, pp. 109-125, 2003.

[32] “Unity Real-Time Development Platform | 3D, 2D VR & AR Visualizations,” Unity Technologies, 28 August 2019. [Online]. Available: https://unity3d.com. [Accessed 29 August 2019].

[33] “What is Unreal Engine 4,” Epic Games, [Online]. Available: https://unrealengine.com/en-US/. [Accessed 11 October 2019].

[34] “Godot Engine - Free and open source 2D and 3D game engine,” Godot Engine (Open Source), [Online]. Available: https://godotengine.org/. [Accessed 11 October 2019].

[35] M. Koutek, Scientific Visualization in Virtual Reality: Interaction Techniques and Application Development, 2003.

[36] F. Argelaguet, “Adaptive navigation for virtual environments,” in *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, IEEE, 2014, pp. 123-126.

[37] E. M. Kolasinski, “Simulator Sickness in Virtual Environments.,” Army Research Inst for the Behavioral and Social Sciences Alexandria Va, Alexandria VA, 1995.

[38] G. Schirner, D. Erdogmus, K. Chowdhury and T. Padir, “The future of human-in-the-loop cyber-physical systems,” *IEEE Computer*, vol. 46, no. 1, pp. 36-45, 2013.

[39] C.-B. Zamfirescu, B.-C. Pirvu, J. Schlick and D. Zuehlke, “Preliminary insides for an anthropocentric cyber-physical reference architecture of the smart factory,” *Studies in Informatics and Control*, vol. 22, no. 3, pp. 269-278, 2013.

[40] D. Zuehlke, “SmartFactory—Towards a factory-of-things,” *Annual reviews in control*, vol. 34, no. 1, pp. 129-138, 2010.

[41] A. Monferrer and D. Bonyuet, “Cooperative robot teleoperation through virtual reality interfaces,” in *Proceedings Sixth International Conference on Information Visualisation*, IEEE, 2002, pp. 243-248.

[42] L. Oliveira and A. Alvares, “Axiomatic Design Applied to the Development of a System for Monitoring and Teleoperation of a CNC Machine through the Internet,” *Elsevier Procedia Cirp,* vol. 53, pp. 198-205, 2016.

[43] *ITU-T Recommendation E.800, Definitions of terms related to quality of service,* 2008.

[44] “Latency [Def. 2] in Webopedia.,” [Online]. Available: https://www.webopedia.com/TERM/L/latency.html. [Accessed 22 July 2020].

[45] M. Ries, P. Svoboda and M. Rupp, “Empirical Study of Subjective Quality for Massive Multiplayer Games,” in *2008 15th International Conference on Systems, Signals and Image Processing*, IEEE, 2008, pp. 181-184.

[46] M. Fiedler, T. Hossfeld and P. Tran-Gia, “A generic quantitative relationship between quality of experience and quality of service,” *IEEE Network,* vol. 24, no. 2, pp. 36-41, 2010.

[47] M. Sithu, Y. Ishibashi, P. Huang and N. Fukushima, “Influences of network delay on quality of experience for soft objects in networked real-time game with haptic sense,” *International Journal of Communications, Network and System Sciences,* vol. 8, no. 11, p. 440, 2015.

[48] Y. Maeda, Y. Ishibashi and N. Fukushima, “QoE assessment of sense of presence in networked virtual environment with haptic and auditory senses: Influence of network delay,” in *2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE)*, IEEE, 2014, pp. 679-683.

[49] M. Fujimoto and Y. Ishibashi, “The effect of stereoscopic viewing of a virtual space on a networked game using haptic media,” in *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology,* ACM, 2004, pp. 317-320.

[50] A. Tatematsu, Y. Ishibashi, N. Fukushima and S. Sugawara, “QoE assessment in tele-operation with 3D video and haptic media,” in *2011 IEEE International Conference on Multimedia and Expo,* 2011.

[51] *Recommendation ITU-T G1010. End-user multimedia QoS categories. ITU-T Study Group 12,* 2001.

[52] *Recommendation ITU-T G1032. Influence factors on gaming Quality of Experience. ITU-T Study Group 12,* 2014.

[53] *Recommendation ITU-T P10/G.100: Vocabulary for performance, quality of service and quality of experience..*

[54] S. Möller and A. Raake, Quality of experience: advanced concepts, applications and methods, Springer, 2014.

[55] U. Reiter, K. Brunnström, K. De Moor, M.-C. Larabi, M. Pereira, A. Pinheiro, J. You and A. Zgank, "Factors Influencing Quality of Experience," in *Quality of Experience,* Springer, 2014, pp. 55-72.

[56] *Recommendation ITU-T P.910, Subjective video quality assessment methods for multimedia applications. ITU-T Study Group 12,,* 2008.

[57] T. Hoßfeld, P. E. Heegaard, M. Varela and S. Möller, "QoE beyond the MOS: an in-depth look at QoE via better metrics and their relation to MOS," *Springer Quality and User Experience,* vol. 1, no. 1, p. 2, 2016.

[58] "Real-time physiological signals | E4 EDA/GSR sensor," Empatica Inc., [Online]. Available: https://www.empatica.com/research/e4/. [Accessed 29 August 2019].

[59] *Recommendation ITU-T P.913, Methods for the subjective assessment of video quality audio quality and audio-visual quality of Internet video and distribution quality television in any environment. ITU-T Study Group 12,,* 2016.

[60] M. Waltl, C. Timmerer and H. Hellwagner, "Improving the quality of multimedia experience through sensory effects," in *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX),* IEEE, 2010, pp. 124-129.

[61] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer and N. Murray, "An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method," in *2016*

Eighth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, 2016, pp. 1-6.

[62] C. Keighrey, R. Flynn, S. Murray, S. Brennan and N. Murray, “Comparing User QoE via Physiological and Interaction Measurements of Immersive AR and VR Speech and Language Therapy Applications,” in *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, ACM, 2017, pp. 485-492.

[63] S. R. Gulliver and G. Ghinea, “Stars in their eyes: What eye-tracking reveals about multimedia perceptual quality,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 34, no. 4, pp. 472-482, 2004.

[64] “Eye tracking technology for research - Tobii Pro,” Tobii Pro AB, 26 August 2019. [Online]. Available: https://www.tobiipro.com. [Accessed 29 August 2019].

[65] “Physiology [Def 1.] In Oxford Dictionary Online.,” [Online]. Available: https://www.lexico.com/en/definition/physiology. [Accessed 29 August 2019].

[66] U. Engelke, D. P. Darcy, G. H. Mulliken, S. Bosse, M. G. Martini, S. Arndt, J.-N. Antons, K. Y. Chan, N. Ramzan and K. Brunnstöm, “Psychophysiology-based QoE assessment: A survey,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 1, pp. 6-21, 2016.

[67] R. Blankendaal, T. Bosse, C. Gerritsen, T. De Jong and J. De Man, “Are Aggressive Agents as Scary as Aggressive Humans?,” in *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*, International Foundation for Autonomous Agents and Multiagent Systems, 2015, pp. 553-561.

[68] “Galvanic Skin Response [Def 1.] In Merriam-Webster Dictionary Online.,” [Online]. Available: https://www.merriam-webster.com/dictionary/galvanic%20skin%20response. [Accessed 29 August 2019].

[69] S. D. Kreibig, “Autonomic nervous system activity in emotion: A review,” *Elsevier Biological Psychology*, vol. 84, no. 3, pp. 394-421, 2010.

[70] M. Pantzar, M. Ruckenstein and V. Mustonen, “Social rhythms of the heart,” *Taylor & Francis Health Sociology Review*, vol. 26, no. 1, pp. 22-37, 2017.

[71] C. M. van Ravenswaaij-Arts, L. A. Kollee, J. C. Hopman, G. B. Stoelinga and H. P. van Geijn, “Heart rate variability,” *Annals of internal medicine [American College of Physicians],* vol. 118, no. 6, pp. 436-447, 1993.

[72] W. Y. Akamine and M. C. Farias, “Video quality assessment using visual attention computational models,” *Journal of Electronic Imaging,* vol. 23, no. 6, p. 061107, 2014.

[73] J. Beatty, “Task-evoked pupillary responses, processing load, and the structure of processing resources.,” *Psychological Bulletin,* vol. 91, no. 2, p. 276, 1982.

[74] D. D. Salvucci and J. H. Goldberg, “Identifying fixations and saccades in eye-tracking protocols,” in *Proceedings of the 2000 symposium on Eye tracking research & applications,* ACM, 2000, pp. 71-78.

[75] W. Wu, A. Arefin and R. Rivas, “October. Quality of experience in distributed interactive multimedia environments: toward a theoretical framework.,” in *Proceedings of the 17th ACM international conference on Multimedia*, ACM, 2009, pp. 481-490.

[76] “About the VIVE controllers,” HTC Corporation, [Online]. Available: https://www.vive.com/us/support/vive/category_howto/about-the-controllers.html. [Accessed 30 1 2020].

[77] “Blender Home,” Blender Foundation, 30 1 2020. [Online]. Available: https://www.blender.org/. [Accessed 30 1 2020].

[78] U. Technologies, “SteamVR Plugin | Integration | Unity Asset Store,” Valve Corporation, [Online]. Available: https://assetstore.unity.com/packages/tools/integration/steamvr-plugin-32647. [Accessed 2 6 2020].

[79] “VRTK - Virtual Reality Toolkit,” Extend Reality Ltd, [Online]. Available: https://vrtoolkit.readme.io/. [Accessed 18 11 2019].

[80] “ROBOSHOT electrical injection moulding machine,” Fanuc, [Online]. Available: https://www.fanuc.eu/dk/en/roboshot. [Accessed 18 11 2019].

[81] “Home | JL Goor,” JL Goor, 11 9 2020. [Online]. Available: https://www.jlgoor.ie/. [Accessed 11 9 2020].

[82] M. A. Srinivasan and C. Basdogan, “Haptics in virtual environments: Taxonomy, research status, and challenges,” *Elsevier Computers & Graphics,* vol. 21, no. 4, pp. 393-404, 1997.

[83] D. R. Begault and L. J. Trejo, 3-D sound for virtual reality and multimedia, Academic Press Professional, Inc, 2000.

[84] “Photon Unity 3D Networking Framework SDKs and Game Backend | Photon Engine,” Exit Games, 1 August 2019. [Online]. Available: https://www.photonengine.com/en/pun. [Accessed 29 August 2019].

[85] M. Pol, “Network Emulator Toolkit The virtual world of Marc O Polo,” 14 May 2013. [Online]. Available: https://blog.mrpol.nl/2010/01/14/network-emulator-toolkit/ . [Accessed 29 August 2019].

[86] V. Clay, P. König and S. U. Koenig, “Eye tracking in virtual reality,” *Journal of Eye Movement Research,* vol. 12, no. 1, 2019.

[87] P. Ugwitz, Z. Stachoň and Č. Šašinka, “Eye-Tracking in VR: From Heatmaps to Object AoIs,” in *14th International Conference on Spatial Information Theory*, Regensburg, Germany, 2019.

[88] C. H. Graham, “Vision and visual perception,” 1965.

[89] J. Pokorny, B. Collins, G. Howlett, R. Lakowski, M. Lewis, J. Moreland, H. Paulson, V. C. Smith and S. Shevell, Procedures for testing color vision: Report of working group 41, National Academies, 1981.

[90] K. Brunnström, M. Sjöström, M. Imran, M. Pettersson and M. Johanson, “Quality of Experience for a Virtual Reality simulator,” *Electronic Imaging,* vol. 2018, no. 14, pp. 1-9, 2018.

[91] H. Dingman, “Tobii's eye-tracking tech is niche on PCs, but makes virtual reality feel like magic | PCWorld,” PCWorld, 18 3 2018. [Online]. Available: https://www.pcworld.com/article/3266634/tobii-vr-eye-tracking-vive.html. [Accessed 11 9 2020].

APPENDIX 1: EXPERIMENT INFORMATION SHEET

Information Sheet

An evaluation of the QoE impact of network delay on a Virtual Reality industry training application.

I would like to invite you to take part in a research study conducted on the behalf of the Athlone Institute of Technology. Before you decide to take part in this research, you will be informed of why this research is being conducted and what it will involve. Take your time to read the following information carefully. You are free to ask questions or inquire about the procedures if there is anything that is unclear or if you require more information. Please read some of the frequently asked questions below.

Who am I and what is this study about?

My name is David Concannon. I am a postgraduate research student here in AIT. As part of my Masters research program I am investigating the factors that influence a user's Quality of Experience (QoE) when operating multimedia devices in the realm of Virtual Reality. Virtual Reality is a novel technology that allows users to experience virtual environments, and the technology has improved significantly in recent years. The improvements are such that users can now experience a sense of realism within these environments. This has rapidly changed the future potential for use cases in VR.

In this study, I will be aiming to evaluate to what degree, the effect network latency will have on a user's QoE of a Virtual Reality training simulation.

What will this study involve?

Before the investigation begins you will be asked to take an eyesight test and a colour blindness test which will inform the results of the experiment. You will be asked to experience a virtual training environment via a VR Head Mounted Display (HMD) wherein you will be asked to operate a machine based on a Fanuc Roboshot Injection Moulding machine. Within the environment you will find instructions guiding you on how to operate the machine itself. You will interact with the machine via a VR control device known as a 'wand' which the researcher will instruct you how to use before you enter the environment. It is anticipated that you will be experiencing the VR environment for **10-15 minutes**.

Before and during the investigation, the researcher will be taking a reading of your temperature, electro dermal activity (EDM) and Heart Rate from a wrist worn device. Once the test is completed you will be asked to rate your experience of the environment via a short survey.

Overall the duration of your participation will be **20-30 minutes.**

Do I have to take part?

Your participation in this experiment is completely voluntary and you have the right to refuse to participate if you do not wish to take part. If at any time you wish to stop the experiment simply inform the researcher present and you will be allowed to withdraw without any consequence or necessary reasoning.

What risks are associated with participating in this research?

The operation of the VR software and hardware is safe and harmless. However, you may experience mild nausea or motion sickness if you have not experienced a VR environment before. You may experience a sense of losing balance within the VR environment, so you will be asked to wear a safety harness for the duration of the test.

If at any point you feel a high level of nausea, please inform the supervising researcher.

Will taking part be confidential?

To ensure participant confidentiality, all personal information gathered during the test will be minimal and any information recorded will be discarded once the research for this Masters degree is completed. There are several situations where we may have to break confidentiality: For example, should the researcher believe any serious risk or harm come to the participant or another individual. Be aware that we will be taking certain non-anonymised data in the form of signed consent forms, however you will not be asked to provide any information outside of your name and age.

All data collected during this investigation will be stored in a secure manner and it will not be possible to identify you from the data collected during the experiment.

Under the freedom of information legislation, you are entitled to access any information you have provided to me during this investigation.

What will happen to the results of the study?

The result of this study will be used to inform the writing of an academic paper as part of my research.

If you have any further questions, please contact me at:

David Concannon
d.concannon@research.ait.ie

Thank you very much for your time!

APPENDIX 2: CONSENT FORM

An evaluation of the QoE impact of network delay on a Virtual Reality remote operation application.

Researcher: David Concannon

Please tick the following boxes:

- I, ____________________, voluntarily agree to participate in this study on the date of ___/___/2018. ☐
- I understand that my participation is voluntary and that I am free to withdraw at any time. ☐
- I confirm that I have read the information sheet accompanying this form and have had the opportunity to ask questions. ☐
- I do not suffer from photosensitive epilepsy or any other form of epilepsy ☐
- I am not pregnant or experiencing any signs or symptoms of pregnancy ☐
- I am satisfied that I understand all the information provided and have been given the time and opportunity to consider the given information. ☐
- I understand that any data collected in the course of this study will be used for research purposes only and in the strictest confidence. Any information related to me will be discarded at the completion of this research. ☐

______________________ ______________________

Name of Participant *Date* *Signature*

______________________ ______________________

Researcher *Date* *Signature*

APPENDIX 3: QUESTIONNAIRE

Post Test Questionnaire

I was immersed in the virtual environment

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

I felt a strong sense of presence within the virtual environment

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

The virtual environment did not feel realistic

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

Using the system was natural

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

I did not enjoy interacting with the virtual environment

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

I was not restricted in my movements

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

The system was slow and unresponsive

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

The task was not easy to complete

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

The instructions were not useful

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree

There was a delay in the virtual environment

Imperceptible	Perceptible but not annoying	Slightly Annoying	Annoying	Very Annoying