The University of Birmingham

DETERMINING PRINCIPLES FOR THE DEVELOPMENT OF MIXED REALITY SYSTEMS FOR COMMAND AND CONTROL APPLICATIONS

by

CHRISTOPHER BIBB

A thesis submitted to

The University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

Department of Electronic, Electrical and System Engineering

School of Engineering

College of Engineering and Physical Sciences

The University of Birmingham

June 2019

ABSTRACT

The pace of advancement in emerging display and interface technologies supporting the development of mixed reality systems – those that exploit the existence of real-world objects to enhance the believability of virtual objects – is rapidly increasing. However, the availability of relevant human-system design standards underpinning the exploitation of interfaces is significantly lagging behind. To provide supporting principles to aid in the development and deployment of mixed reality systems, a series of studies was conducted to systematically investigate a range of design parameters relevant to mixed reality, and to determine the impact of those parameters on human-system performance, including cognitive and physical demands. An assessment of specific design standards was undertaken related to the performance of fundamental human-system interaction tasks in a mixed reality system. It was found that mixed reality is most suited to selection tasks, as opposed to more complex interaction tasks such as repositioning and rescaling virtual objects in 3D space. An evaluation was also made of the effects on *presence* of introducing physical "tangible" interface elements co-located with virtual content. The findings show that tangible interface objects have a significant positive effect on presence, in addition to usability and workload. Finally, an investigation was undertaken to assess the effects of vibration — a common, uncontrollable environmental condition — on human-system performance. Vibration is shown to have a significantly larger impact on accuracy for eye-based input than on head-based input when performing dwell-based interaction. The lowest frequencies have the greatest effect on accuracy, with higher frequencies producing similar effects to instances of zero vibration.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Professor Robert J. Stone for his expertise and guidance, as well as his continuous enthusiasm and encouragement throughout the entire process. I would also like to thank everyone in the Human Interface Technologies (HIT) Team, past and present, for the help and support throughout the project. Additionally, I would also like to thank the industrial sponsor, BAE Systems, with a special thanks to Jean Page and those at Warton who provided the guidance, knowledge and financial support that made the PhD research possible. I would like to thank my girlfriend Dana for her support and patience during the long field trips and many late nights throughout the PhD. Last but not least I would like to thank my family for the endless encouragement and unconditional support from the moment I started my undergraduate degree until the end of the PhD.

This thesis was copy edited for spelling and grammar by Scribbr.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
ANOVA	Analysis of Variance
AR	Augmented Reality
CAVE	Cave automatic virtual environment
C2	Command & Control
EEG	Electroencephalography
HCI	Human Computer Interaction
HF	Human Factors
HFI	Human Factors Integration
HIT	Human Interface Technologies
HMD	Head Mounted Display
HMI	Human Machine Interface
HOTAS	Hands On Throttle-And-Stick
HUD	Heads Up Display
IMU	Inertial Measurement Unit
IPQ	iGroup Presence Questionnaire
IR	Infrared
MBT	Main Battle Tank
MR	Mixed Reality
NASA-TLX	National Aeronautical and Space Administration Task Load Index
OST	Optical See-Through
PQ	Presence Questionnaire
RAF	Royal Air Force
SD	Standard Deviation

sUAV	Small Unmanned Ariel Vehicle
SSI	Short Subjective Instrument
SUS	System Usability Scale
UI	User Interface
VDU	Visual Display Unit
VE	Virtual Environment
VR	Virtual Reality
VST	Video See-Through
XR	Term describing “Virtual, Augmented and Mixed Reality”

Chapter 1: Introduction

1.1 Overview

1.1.1 Scope

The PhD research was undertaken as part of an iCASE (Industrial Collaborative Awards in Science and Engineering) studentship funded by the EPSRC (Engineering and Physical Sciences Research Council) and the industrial sponsor, BAE Systems (Air Sector based in Warton, UK). The research was conducted within the Human Interface Technologies Team, part of the School of Engineering at the University of Birmingham. The focus of the research is human factors (HF), human-computer interaction (HCI) and virtual reality (VR), augmented reality (AR) and mixed reality (MR). The research questions and experiments were developed in collaboration with BAE Systems subject-matter experts, including HF specialists and aircraft test pilots.

1.1.2 Problem Statement

The field of VR research has been steadily evolving since the 1980s. However, since the recent appearance of low-cost commercial VR technologies, interest by industry and researchers in VR and AR has grown rapidly (Cipresso, 2018). Consequently, the range of related technologies and systems being developed and released to market is increasing equally rapidly, thus providing many novel display and visualisation capabilities. While these new technologies may present novel capabilities and features, lessons learned over the past three decades (Stone, 2012) emphasise the fact that the suitability of such devices for specific tasks and users within a military environment must be investigated, preferably with a strong emphasis on the role of HF and human-centred design (Stone, 2016a).

However, while the research and technology associated with emerging MR devices have advanced rapidly, the formal standards and guidelines governing the design of such systems have not (Department of Defense, 2012). Compliance with formal design standards is often a legal requirement that must be met before industry and military systems can be qualified for use by the end user (ISO, 2019). However, many of the novel capabilities presented by emerging devices, such as MR systems, are outside the scope of existing standards and guidelines and no clear guidance for use is provided.

While the applicability of general design standards is well known when addressing the use of devices within a physical environment, devices may not be suitable when applied to a VR, AR or MR environment. The lack of defined design criteria could result in a situation in which there are no relevant design standards for the design of a system such that certification cannot be gained. Without certification, the device would not be used despite the benefits it may provide. Alternatively, lack of understanding of the appropriate use cases and tasks that a system may be used for could also lead to a specific device being used for a task or in a context that is not suitable and may have a negative impact on key factors such as performance, safety or usability.

Thus, to assess the suitability of any system, a formal investigation must be undertaken to ensure that the system is able to beneficially support the user in completing required tasks. The present research aims to address such issues by undertaking studies and performing experiments to assess the suitability of MR systems under a range of representative tasks and environmental conditions which are commonly presented to the military user.

1.2 Background

1.2.1 Definitions

The definition and common use of the term *mixed reality* is often used interchangeably with the terms *augmented reality* and *virtual reality*. Despite the similarities in software and hardware components, there is a subtle difference between VR, AR and MR regarding the relationship between the interaction with the virtual content and the real world. A summary of MR technologies by Stone (2016b) describes this relationship as a process of blending the "best of the real world with the best of the virtual". In addition, Stone (2016b) defines MR as a form of AR in which, instead of superimposing computer-generated material onto real-time images of the world, an attempt is made to exploit the existence of real-world objects to enhance the believability and usability of computer-generated or virtual elements. To differentiate between VR, AR and MR, Milgram and Colquhoun (1999) provide a spectrum to define the level of "reality" between the real world and virtual environments; this spectrum is referred to as the reality-virtuality continuum (Figure 1.1).

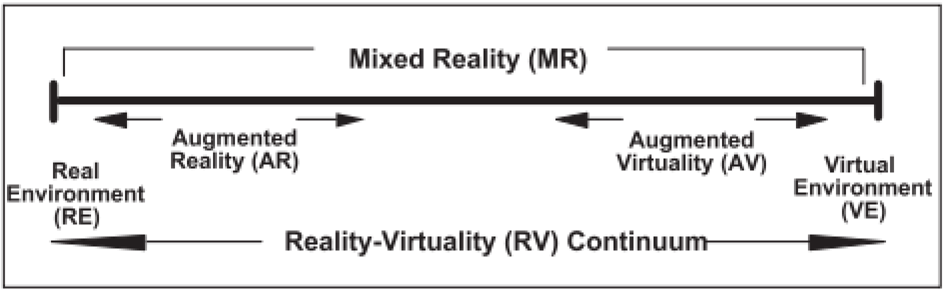

Figure 1.1: Reality-Virtuality Continuum (Milgram and Colquhoun, 1999)

Schnabel et al. (2007) further classify the subtle differences between "reality" concepts into six categories. These are mapped along two dimensions according to the level of interaction with real-world objects and the correlation between perception and action. While most of the industry and many academic publications do not differentiate beyond VR, AR and MR categories, it is beneficial to understand the level of interaction with real-world objects when comparing studies of the same classification, as shown in Figure 1.2.

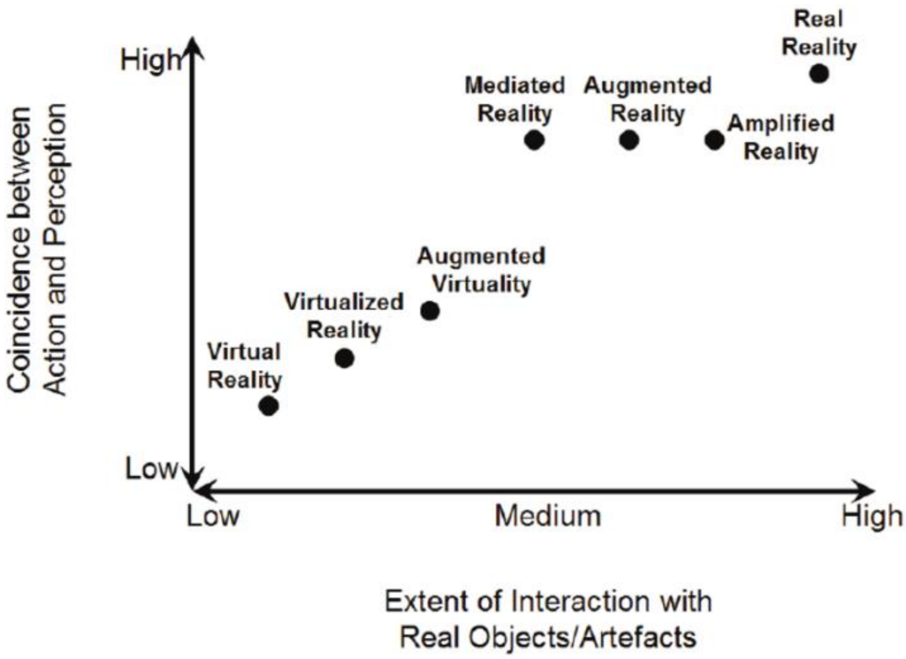

Figure 1.2: Classification of reality concepts according to the correlation between perception and action and level of interaction (Schnabel et al., 2007)

Figure 1.3 visually illustrates the conceptual differences between the different levels of "virtuality". The left image illustrates a VR system in which the operator's entire visual field is "immersed" within a virtual environment. A VR system will only display content that is inserted into the virtual environment and will have no visual or physical interaction with the real-world environment.

The centre image illustrates an AR system in which the real world can be seen in the user's normal field of view and the augmented, computer-generated content is superimposed onto the real-world scene. Whilst the virtual content may have a connection with physical objects within a scene, (e.g., a person's name and details are displayed above his or her head), there is no interaction between the person and the content except a virtual display which is registered to the person's location. Finally, the right image illustrates an MR system in which virtual content has a relationship with the physical environment such that a physical object can have a direct effect on the behaviour of the virtual object.

While older papers (such as Milgram and Colquhoun, 1999) refer to MR as a broad spectrum that encompasses "realities" that range between VR and the real environment, many researchers now refer to MR as an entity of its own which resides between AR and the real environment. Furthermore, Milgram and Kishino (1994) state that, within MR systems, the properties used to create spatial environments allow users to interact with both physical and digital information in an integrated way. For the purposes of this present research, the *working definition* of MR is defined as a system in which there is a substantial understanding of and relationship between the real-world environment and the virtual content being presented. An MR system requires that the computer-generated content is not simply presented as an overlay on the real world but that it is embedded in the physical environment.

When referring to all three of the "reality" concepts—VR, AR and MR—the term *XR* is often used. Because the technologies and interaction methods used for all three concepts are so similar, use of the term *XR* is of use when making a statement which applies to all three equally. The term *XR* may apply to all three reality types, as shown in Figure 1.3, with the differentiation being made with respect to the different levels of real-world inclusion compared to virtual content.

Figure 1.3: Visual representation of VR (left), AR (centre), MR (right) (Wired, 2017)

1.2.2 Applications

AR has many current and near-future applications in the military, including in the cockpit, for training, and in the battlefield (Livingstone et al., 2011). A common current use of AR in the military is within combat-aircraft cockpits. AR has been being used for over a decade in the form of a head-mounted displays (HMD) housed within a pilot's helmet, and head-up displays (HUD) mounted above an aircraft's instrument panel. These human-machine interfaces (HMI) can provide a vast amount of information to a user, including data from instrument panels, navigation and terrain data, and advanced features such as superimposed terrain and guidance information. Figure 1.4 illustrates a view from within a military-pilot HMD which displays the exterior real-world view, as seen from a night-vision camera, with instrument data streams superimposed. Figure 1.5 illustrates an HUD that is statically positioned within a cockpit and in full view of the pilot's central forward vision. These can be defined as AR instead of MR systems, as they provide an overlay of information only, with no further virtual and real-world interactivity.

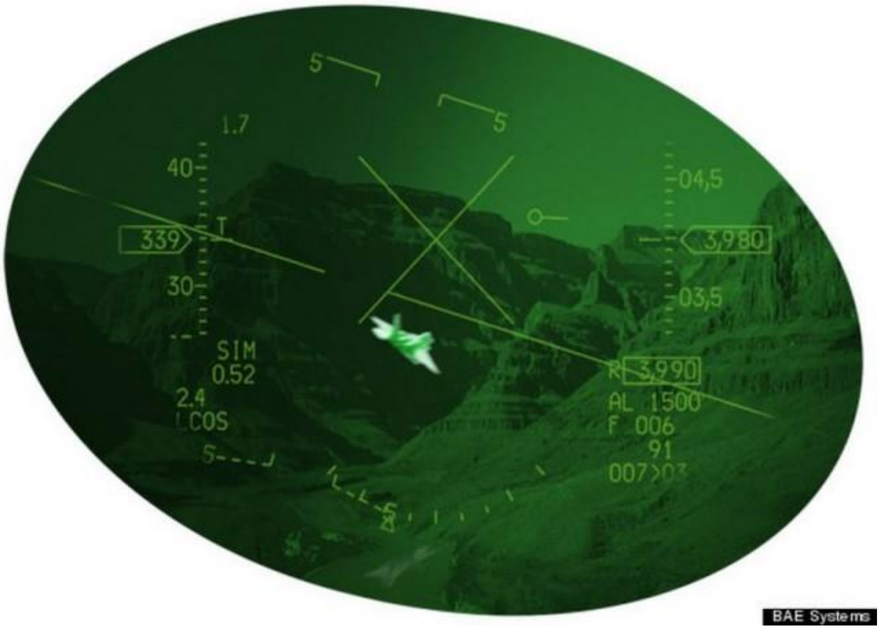

Figure 1.4: Striker II Digital Helmet-Mounted Display (BAE Systems, 2018a)

Figure 1.5: LiteHUD head-up display (BAE Systems, 2018b)

Applications of AR and MR are becoming increasingly numerous, both within the commercial sector and in the academic literature. De Crescenzio et al. (2011) claim that, by using advancing and emerging AR technologies, it is possible to ease the execution of complex operations and ensure efficiency in the transfer of knowledge. They further claim that AR can augment a user's existing skills and training by presenting additional, context-relevant materials without burdening the user with additional manuals or materials.

With the above in mind, AR has proven to be a valuable tool in complex environments in which the user often requires assistance in performing complex tasks. Many examples of AR applications are present within the literature, including the use for AR for many complex industrial applications, such as assembly (Siltanen et al., 2007), manufacturing (Sarwal et al., 2005), maintenance (Borsci et al., 2015; De Crescenzio et al., 2011; Henderson and Feiner, 2009; Macchiarella and Vincenzi, 2004) and safety-critical applications. Furthermore, AR has been utilised for the gamification of training exercises (Shooter, 2017), for spatial-ability training in understanding navigation, for orientation in dangerous areas (such as those encountered by firefighters) (Dunser et al., 2006), and for many more applications.

Other studies of AR in industrial applications have found that it can reduce the time taken and errors made when performing maintenance tasks and can reduce mental workload and the physical exertion experienced by the user (Henderson and Feiner, 2009). AR has also been shown to increase learning performance, to decrease training time, to improve recall and retention (Macchiarella and Vincenzi, 2004) and to reduce errors caused by insufficient training or misinterpretation of facts (De Crescenzio et al., 2011).

To evaluate the possibility of using AR to replace existing methods of displaying information, a prominent study by Henderson and Feiner (2009) compared an AR display against both a small monitor and an HUD. The study presented five forms of augmented content to assist the operator in maintenance task sequences, including the following: (i) attention-directing arrows (ii) text instructions and warnings (iii) registered labels showing location and surroundings (iv) a 3D virtual scene registered on the target, and (v) 3D models of tools and components registered at their projected location (Figure 1.6). Compared to the monitor and HUD conditions, the study found a significant improvement in task completion time and accuracy when using the AR HMDs. The authors also reported that users described greater levels of satisfaction and offered similar ratings of ease-of-use and intuitiveness.

Figure 1.6: Operator using the system (left), AR interface (right) (Henderson and Feiner, 2009)

Many of the studies presented above report a benefit to the use of AR. However, most studies are completed under laboratory conditions in which no external environmental factors are present. Environmental conditions such as vibration, lighting and sound may be present within the real operational environment but are not considered by the experimental method of the studies reported. The lack of consideration given to replicating conditions that apply to real environment limits the applicability of the findings of such studies to outdoor operational environments.

While the potential military uses of AR and MR are vast, many existing systems are used within static and well-defined areas which can be studied and mapped in advance. For example, a pilot within an aircraft or a maintainer in a factory. A challenging use case is within a dynamically changing, unknown environment when the system is under prolonged and intensive use. For example, consider a soldier on patrol who travels through new areas about which the system has no stored knowledge. Roberts et al. (2013) present a study to test and evaluate a military-oriented wearable AR system used in an outdoor environment. The study evaluated key metrics of an AR system worn by a soldier, including performance, usability and technology-related challenges of using a head-mounted system in an outdoor operational setting.

Based upon the study, Roberts et al. (2013) concluded that the requirements of a system worn by a patrolling soldier consist of three main elements: (i) to track the position and orientation of the soldier's head quickly and precisely (ii) to do so at a relatively low cost within a ruggedised package, and (iii) to operate in any arbitrary outdoor environment without requiring specific preparation or instrumentation of the environment. These requirements present many challenges and represent an active and growing field within the research and development communities.

Figure 1.7 illustrates a rendered demonstration of the same system overviewed by Roberts et al. (2013). The system presents elements which may be considered as both AR and MR solutions. First, there are text overlays (shown in green and white) which are in a fixed position of the display, which do not change with the users' head movements and which may be considered as AR display elements. Second, there are display elements (shown in red) which are fixed to the position and depth of the real world and which require a spatial understanding of the physical environment to operate. These display elements are fixed to a position in the real-world environment when users move their head or rotate their body. The figure represents the progression of AR technologies from systems which display only simple information overlays to spatially aware systems which can react to a dynamically changing environment.

Figure 1.7: Urban Leader Tactical Response, Awareness & Visualization (ULTRA-VIS) interface (Cheng, 2014)

1.2.3 Command and Control systems

Commonly, command and control (C2) systems comprise many screens and devices and are typically housed within a small enclosure, such as an ISO container or dedicated command room. C2 systems use a wide variety of HMI display and control technologies, including traditional workstations with fixed-location visual display units (VDU), large group displays (LGD), 3D surfaces and tables, volumetric displays, ambient displays and many others. The definition and use of the wide variety of displays, including those described above, is further detailed by Knight and Stone (2014).

An overview of the different HMI display and interaction devices used for C2 is summarised by Knight and Stone (2014), as are suggested use cases and the advantages and disadvantages of each display method. The existing display facilities listed above may be suitable within a formal command centre building that includes permanent infrastructure and support. However, they may not be suitable or applicable when used in a field setting where no support or permanent infrastructure is available. The different support requirements between display types is apparent when comparing display systems such as “CAVE” (cave automatic virtual environment) displays, which have extensive setup time and space requirements, with a wearable or mobile device for which no external infrastructure or devices are required.

The use of MR for C2 systems is becoming more prevalent due to the additional features and capabilities which the advanced visualisation and control facilities can provide. MR displays make it possible for individual operators to visualise many aspects of the myriad of display technologies available for military C2 applications but without the physical infrastructure usually required to support such devices.

Figure 1.8 illustrates an MR system which provides a high degree of interactivity with the virtual content and physical environment, in contrast to an AR system in which virtual content is simply superimposed. As can be seen from the figure, the virtual content is superimposed over a physical table in the centre of the room, thus providing a clear indication of the area of interaction and restricting users from walking through the area used for visualisation. Replacing physical displays with virtual displays reduces the weight and “cost of upgrade” of platforms and mitigates the need for physical modifications when changing display types.

Figure 1.8: Concept illustration (left), MR Interface (right) (Odom, 2017)

Many defence and associated industrial companies present concept illustrations of “command table” applications. These commonly involve a commander-level map view with various embedded vehicle and objective assets. Many of these are only concept illustrations with no known active development. Several companies have presented research and development platforms which have undergone HF and usability testing of the use of MR systems and virtual displays.

One such research platform is presented by BAE Systems (2015), as seen in Figure 1.9. The system illustrates the concept of virtually displaying multiple display concepts within a single MR interface. The system includes the displays commonly associated with C2 systems, including (i) 3D surfaces and tables (ii) large group displays (iii) ambient displays, and (iv) virtual avatars. Each of these display methods would usually occupy a large area and require extensive setup before use. Because virtual displays shown in the MR system are solely computer-generated, however, they do not occupy any physical space.

The system includes a physical circular table of the same size, shape and location as the virtual content; in this case, the virtual content consisted of terrain, virtual displays and associated mission assets. The table provides tactile feedback to the user regarding the correct position for the hands, and it prevents the user from placing his or her hands through the virtual content. This system is a further example of an MR system in which a clear differentiation from an AR system can be seen in the close relationship between the physical world and the virtual content. By virtually replicating the multiple display methods, this system allows for an infrastructure-free environment which can offer a wide variety of display facilities without the associated setup or cost.

Figure 1.9: Concept illustration (left), MR Interface (right) (BAE Systems, 2015)

1.3 Present Research: Aims and Objectives

The aim of the research described within this thesis was to determine principles for the development and deployment of MR technologies for data visualisation in a defence context. The main outcome of the research is to provide a set of evidence-based recommendations regarding the suitability of MR technologies for a selection of tasks and contexts. Five specific objectives were defined at the beginning of the research.

The first objective is to review the literature relating to the design and use of MR systems and to identify the most relevant design criteria available. There are many documents available for the design of human-machine interfaces. The contents of these documents range from formal design standards and guidelines to studies assessing performance across a wide range of design criteria. The design criteria recommendations vary greatly depending on the technologies used, the tasks to be undertaken, and the context and environment in which the device will be used in. Therefore, the aim is to compare differing and contrasting sources and present the design criteria which are most relevant to MR systems.

The second objective is to evaluate the suitability and readiness of a range of MR display and interaction devices when used for real-time visualisation and interaction with virtual displays. This objective was achieved through the development of a modular experimental software system which supported the rapid integration of a range of devices for a technology assessment and later experiments.

The third objective is to determine which tasks an MR system is best suited for concerning both performance and a range of subjective human-centred measures. Simple single-gesture selection tasks may require only gross rapid input, whereas more complex interaction tasks may require prolonged, precise and multi-gesture input. To determine which tasks are most suitable, two experiments were performed to assess the effect that different types of tasks can have on human performance metrics.

The fourth objective is to investigate the use of physical (tangible) objects within the environment as a tool for interaction with virtual content. Tangible objects provide tactile feedback to users and affordances and cues (such as shape) of the real-world physical environment.

The fifth objective is to investigate the effect of vibration when using eye- and head-based input for interaction with virtual displays. Vibration is a commonly occurring environmental condition within many ground, sea and air platforms, and the impact this vibration has on the user during common interaction methods must be understood.

1.4 Thesis Structure

Chapter 2 provides a review of the literature relating to the design and use of MR systems and is presented across three sections. The first section details studies and standards relating to interface design and their relevance to MR systems. The second section provides details on the topic of *presence*—which is a commonly measured metric within VR—and its relevance to MR systems. Tangibility is a key element of an MR system, and the effect of tangibility on presence within MR systems is accordingly investigated. The third section considers environmental factors that can affect eye- and head-based input, both of which are established and commonly used interaction methods for MR systems. Specific vibration frequencies are known to affect humans in specific ways, and this fact is discussed in detail.

Chapter 3 describes experimental software development and its features and capabilities. The system utilises data from a wide variety of sources, and the different methods of data authoring are detailed. A wide range of display and interaction devices are available for MR systems, and a range of the commonly used devices was assessed through a technology assessment. The technology assessment provided a selection of usability considerations of each device type and informed the technologies to be used in later experiments.

Chapter 4 presents two experiments which investigate the optimal user interface design criteria and the type of tasks that are best suited to MR systems. The experiments compared an MR system that uses virtual displays with one which is based on established physical technologies, including a touchscreen and joystick controller.

Chapter 5 presents an experiment which assesses the effects of using tangible interface objects on *presence* when using an MR system. As detailed previously (Section 1.2.1), interaction with the physical environment, which is often referred to as tangibility, is a key element of an MR system, and the effect of tangibility on presence and other related metrics is accordingly assessed.

Chapter 6 presents an experiment which investigates the effect of vibration frequencies on accuracy when interacting with a virtual display using head- and eye-based interaction. Vibration is a commonly occurring environmental condition that a user can experience when on a vehicle, and the performance implications must be identified if a system is ever to be used in an operational setting.

Chapter 7 summarises the previous chapters and presents the principles for the design and deployment of MR technologies based on the findings of the experiments. A further discussion details the limitations of the research and experiments. Finally, the implications for future work are discussed.

Figure 1.10 illustrates the structure of the thesis and the contents of each chapter.

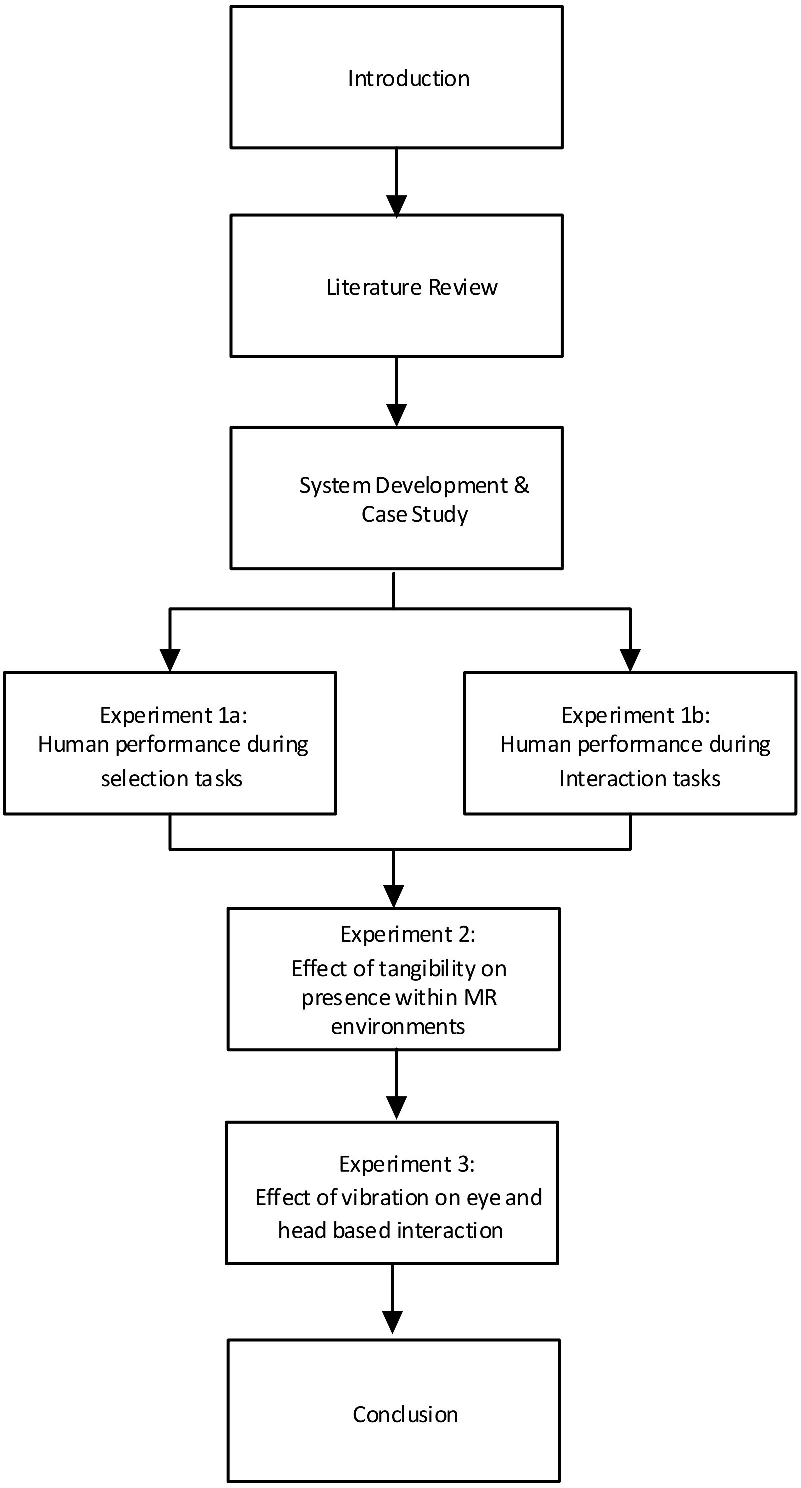

Figure 1.10: Thesis structure diagram

Chapter 2: Literature Review

2.1 Introduction

Many factors can affect human and system performance within MR systems. These can include the interface design choices made, the effect the system has on the user's cognitive state, and the environment in which the system operates in. While all of these factors are well studied for established technologies such as touchscreens and joysticks, few investigators have performed in-depth HF-based assessments on emerging technologies such as MR systems.

This chapter begins with a review of the methods available for evaluating MR systems and details the primary issues and limitations of each assessment method. The first part of Section 2.2 reviews the design standards and guidelines used within defence and their appropriateness relative to emerging technologies such as MR. Section 2.2 also reviews related studies which address the optimal design criteria for the user interfaces of existing technologies, which include touchscreens and eye-based input. Further details are provided regarding the reliability and validity of the experimental-design approaches used within the literature.

Section 2.3 provides details on the topic of *presence*—which is a key measure of virtual environments—and its relevance to MR systems. Furthermore, the section details the definitions of and relationship between presence and human performance. The section further details the main factors which contribute to presence and the various subjective and objective methods of measuring presence. Section 2.4 details the appropriateness of using MR systems under a variety of contexts, including the physical and environmental conditions in which they can operate. As detailed in Section 2.4, vibration is widely recognised across the literature as a key player in HMD-based interaction. It is accordingly investigated in detail.

2.2 Design of Mixed Reality Systems

2.2.1 Design and Evaluation Methods

Many HF-based assessment methods are available for evaluating the performance and usability of software and hardware systems. The primary methods available include standard compliance reviews, interviews, observations, human-performance modelling and experimental testing (Charlton and O'Brien, 2002). Each method has its own benefits, disadvantages and recommended contexts of use.

2.2.2 Standards and Guidelines

A common system design evaluation method examines compliance with official standards and guidelines. Many guidelines, standards, and HF guidance documents can be consulted during the design of a system and are available across many sectors, including commercial, government, industry and defence. Compliance with standards is mandatory in many sectors, including the military and various industrial sectors, and is often required in order to meet legal or contractual requirements (ISO, 2019). If compliance is required, until a system has been certified to be compliant with relevant standards, it cannot be sold or deployed to the end user. Design standards often provide detailed technical information regarding how a device or interface should look and behave, and this can aid in ensuring consistent design and behaviour between systems. Additionally, standardisation of design criteria across an industry provides familiarity across similar systems and can thereby reduce the training needs.

While the HF community provides very in-depth technical specifications for designing and evaluating software and hardware systems, most are aimed around established and commonly used technologies such as touchscreens and keypad input. New technologies, such as XR systems, provide novel capabilities compared to established and prolific technologies such as touchscreens and joysticks. In the case of MR technologies, the new capabilities provided include new methods for system interaction and data visualisation. However, the operational benefits of these technologies may be limited due to out-of-date design criteria and an absence of relevant HF-informed guidelines and standards (Department of Defense, 2012).

As the pace of introduction of emerging technologies increases, relevant standards may not follow for many years after introduction. The gap between the introduction of technologies and the creation of standards can result in no certification being issued. For safety-critical applications such as those used in aerospace and defence, the need is paramount for a rigorously tested and compliant product. The use of emerging technologies that have not been formally evaluated may lead to devices being used for inappropriate tasks with potential for serious problems and consequences.

The Department of Defense (2012) summarises the situation in the foreword to the defence design standards for the U.S. military:

> "Tomorrow's systems will depend on greater cognitive processing on the part of the human operator, maintainer, and support personnel. Portable or wearable computers are likely to be commonplace. New display concepts such as virtual reality, haptic (touch sensing), and three-dimensional [displays] are receiving a great deal of interest, as are voice, pointing, gesture, and eye-blink control systems. Technology, if misapplied, will impose human performance requirements that cannot be satisfied. Many technologies are evolving rapidly; the human is not. The benefits of new technologies may not be realized if one fails to consider human capabilities and limitations."

Many of the design and usability factors that may be encountered by a user wearing an MR device can be found within published HMD-specific design guidelines (e.g., Melzer and Moffitt, 1997; Velger, 1997), HF standards (e.g., Department of Defense, 2012; International Organisation for Standardization, 2010; Ministry of Defence, 2008a, 2008b) and display-specific documents (e.g. HFI, 2017; Tuason, 2012).

No known document addresses all aspects of the design, evaluation, and technical aspects of for the development of MR systems. Instead, a designer must assess a wide range of documents and attempt to understand the performance and usability issues that may be encountered based on similar technologies. To illustrate the limited information available for emerging technologies (such as for eye- and head-based controls), MIL-STD 1472G (Department of Defense, 2012) provides the following guidance:

- Eye and head-based controls may be used for a variety of tasks including teleoperations, instrument selection on a panel, and visual search tasks.
- Eye and head-based controls shall not be used in vibrating environments.
- Head-based controls shall not be used if the task requires frequent, precise head movements.
- Line-of-sight dwell times shall be minimized when using eye-based controls. Line-of-sight dwell times shall be not greater than 300 milliseconds.
- System response time shall be minimized. System response time shall be not greater than 100 milliseconds.

No further technical details are provided for the exact interface selection methods to use (such as dwell-time activation or additional selection-confirmation options) or the best methods of object manipulation (including selection, manipulation, and release methods). In addition, no details are provided regarding interface design options, such as the type, shape or size of cursors to use. To understand the effect of such design decisions (for example, if the cursor shape has an effect on performance), further investigation would be required.

Many of the large consumer-device manufacturers of XR systems have provided software toolkits for developers which include their own human interface guidelines, such as that provided by Apple (2018). However, in reality, the commercial guidelines are rudimentary and provide little guidance beyond generic statements such as “be mindful of the user’s safety”. This contrasts with the guidance provided by academia and industry, which provide detailed HF-informed, HMD-specific guidelines (e.g., Stone, 2012). HFI (2007) provides a list of many of the standards and guidelines that govern the design of displays and controls, including UK legislation, international standards, European standards, and military standards and guidelines both for the UK and U.S.

In 2017, the long-standing Defence Standard 00-250 (Ministry of Defence, 2008a) was made obsolete and replaced by a series of updated technical guides addressing HF integration. One such guide by HFI (2017), the *Controls and Displays Technical Guide*, provides guidance for several areas which are not covered in previous documents. This includes many aspects of selecting display and control facilities, including the key principles relating to visual displays, such as performance effects, portability, wearability and the effects of some environmental conditions. In addition, the document details related relevant standards, requirements and health and safety considerations.

The document also provides basic guidance regarding several interaction methods which are relevant to MR, including the following: auditory displays, tactile displays, olfactory displays, motor controls, gestural interfaces, haptic feedback, physiological controls, and control interactions in 3D. However, the HFI document presents guidance in the form of short textual descriptions of considerations, with no detail given on exact design parameters such as object sizes, shape, colours and other fundamental criteria.

Table 2.1 and Table 2.2 illustrate the general, high-level guidance provided within the technical guide. Many of the factors, such as field of view, require further in-depth analysis to determine the correct specification needed for the particular task and end-user requirements. However, the document does not address the key concepts of MR systems, such as interaction with objects in 3D space and the relationship between virtual objects and the physical external world.

Table 2.1: General considerations for HMDs (HFI, 2017)

Issue	Consideration
FOV	The FOV should be large enough to provide acceptable visual search performance.
Symbology registration	Symbology and information relevant to the task at hand should be viewed by the user in their instantaneous FOV, regardless of their head position.
Symbology location	Symbology should be displayed in the central 25 degree of the users' visual arc to minimize required eye movements.
Symbology clarity	Symbology and graphics must be large enough and clear enough to allow object recognition, and spatial orientation.
Visual restriction	HMDs shall not obscure task essential vision and should allow unrestricted view of all displays and controls.
Visual inhibition	If a HMD presents information by overlaying it onto the users' normal FOV, the imagery presented should not obscure the background and should be visually distinctive from the background.
Visual distraction	HMDs should not distract from the user's attention or increase their cognitive load by providing non task-oriented information with memory requirements. HMDs should only direct the user's attention to critical information.
Monochromatic colour shades.	Monochromatic HMDs should provide at least six shades of grey for alphanumeric and simple graphic information and at least nine shades of grey for complex graphic or sensor data. Depending on the image and information being presented more may be required.
Latency / Refresh Rate	Sluggish refresh rates can lead to discomfort or disorientation in wearers and can create problems for interpreting the display content.

Table 2.2: Wearability considerations for HMDs (HFI, 2017)

Issue	Consideration
Weight	The weight of the HMD should be minimised to reduce musculoskeletal loading and maximise comfort. Consideration must also be given to the potential effects of accelerations on the users head and neck e.g. the ejecting from an aircraft.
Weight distribution	The mass moment of inertia of the HMD system about the Centre of Mass (CoM) of the head should be minimised. Weight distribution of HMD-helmet mounted items should be balanced as much as possible to avoid or minimize neck strain and fatigue and improve comfort.
Movement	Head mounted systems should not restrict the user's head or shoulder motion required for the performance of any tasks
Posture – Augmented Reality alignment	Alignment of HMD virtual images with an external visual target should not require the wearer to adopt extreme head and neck postures.
Posture – visual restriction	Reduced FOV caused by the HMD housing should not require the wearer to adopt extreme head and neck postures to attend to external visual cues.
Adjustability	The user must be able to adjust the display to view the imagery.
Fit	HMDs should be adjustable to ensure a secure fit for the full range of users so that they do not move relative to the wearers head and to maximise comfort, including inter pupil distance and dioptre adjustment. HMD and helmet systems must not slip under vibration, acceleration, buffeting or due to sweating.
Comfort	The head mounted system must be comfortable for prolonged use and resist heat build-up.
Attenuation of Head motion	A HMD system should be designed so as to attenuate head motion in the 4 Hertz (Hz) range. Amplification of any head motion could result in significant problems in viewing the information display systems used for other tasks or the users normal FOV and could result in muscular strain or injury.
Compatibility/integration with other equipment	Power and data connections, connector types, physical integration with PPE such as glasses and helmet.

Roberts et al. (2013) present a paper based on a soldier-worn AR system which used a U.S. military design standard for the interface layout and symbology/icons used, as shown in Figure 2.1. The authors performed an HF assessment of the system by measuring human and system performance and usability. Furthermore, they detail the devices and components used, the tracking and registration methods applied, and the design standards used.

Abiding to standards may be required for the certification of a system, and as detailed previously, in some industries it is a legal and contractual requirement. In addition, applying defence standards to interface design allows cross-compatibility between systems, thereby providing familiarity and facilitating training and knowledge transfer with other systems complying with the same standards. The system presented in Figure 2.1 utilises the recommended symbology designs provided by the design standard selected and would therefore be familiar to military users that have previously used systems that applied the same standard.

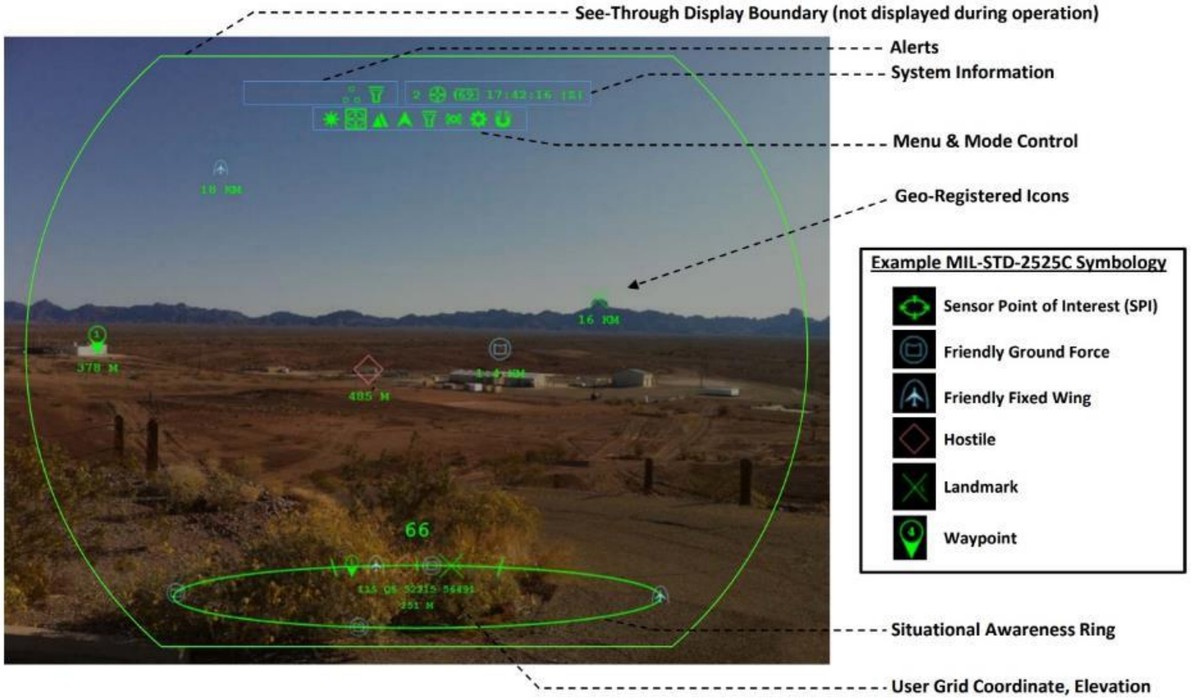

Figure 2.1: ULTRA-Vis user interface (Roberts et al., 2013)

2.2.3 Interface Design Criteria

Given the lack of technical detail provided by formal industry and military guidelines and standards, further investigation is required to assess the suitability of MR technologies for a range of users and tasks. This can be achieved via predictive mathematical modelling and human-performance testing.

It is possible to mathematically predict the point-select time of a target. Fitts' Law (Fitts, 1954) is a predictive model of human movement which is used within the human-computer interaction field. The law predicts the time it takes to rapidly move a cursor to a target based on the ratio between the distance of the cursor to the target and the width of the target, as shown in Figure 2.2. Intuitively, this law may be summarised in the following words: the larger the target is and the closer is to the pointer, the faster the pointing time will be. Fitts' law has been shown to apply across a variety of conditions, from hand inputs (Hoffman, 1991), head inputs (So and Griffin, 2000) and eye-gaze inputs (Zhang and MacKenzie, 2007), to mobile AR displays (Rohs et al., 2011). However, the predictive model does not extend to further design criteria, such as spacing, colour and shape. Therefore, further experimental testing is required to assess these further design choices.

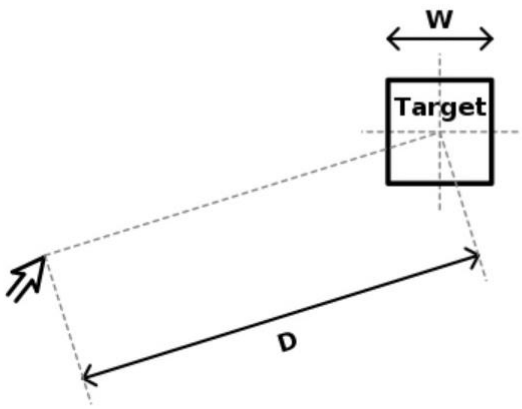

Figure 2.2: Fitts' Law draft of target size and distance to target (Fitts, 1954)

When predictive modelling is insufficient, human performance testing can be used in conjunction with design guidelines and standards as a method for validating system design and analysing the effectiveness of human-system interaction. Experimental testing can provide a much deeper insight into performance and usability than compliance with standards alone.

While experimental testing is expensive, time-consuming and requires a lot of effort, its key benefit is that it removes the disassociation of subjective user preference from objective human performance (Charlton and O'Brien, 2002). In addition, in contrast to subjective measures such as questionnaires and interviews, which provide only qualitative feedback, experimental testing provides an assessor with accurate and quantitative data on the range of conditions tested.

Many studies assess the optimal design criteria of objects by testing a selection of design variations in performance assessments. Many studies are available for mature technologies such as touchscreens (see Section 2.2.4.1). For less-prevalent technologies, such as those involving eye-based interaction, far fewer studies are available (see Section 2.2.4.2). However, for emerging technologies such as MR systems, no academic study was found during the literature search.

2.2.4 Assessment of Design Criteria

A common task to be assessed for performance is a point-and-select task, as described by *ISO 9241-9: Requirements for Non-keyboard Input Devices* (International Organisation for Standardization, 2002). This standard describes performance testing considerations for evaluating human performance on non-keyboard input devices. This standard has since been updated by *ISO 9241-411: Evaluation Methods for the Design of Physical Input Devices* (International Organisation for Standardization, 2012). Defining the point at which an object is activated will also affect the selection time. Activation can be based on when an object is pressed or released. ISO 9241-9 defines the two touch strategies as either first-contact or last-contact. First-contact touch is based on actuation of the display area upon touching the display surface, and last-contact is based on the actuation of the display area upon withdrawing touch from the display surface. Tauson (2012) recommends activation based on last contact for selection tasks when using touchscreens.

2.2.4.1 Touchscreen Input

For the design of touchscreen interfaces, a large number of studies have assessed the effect of object sizes on human performance when using touchscreens under a variety of conditions. Many of these studies assess performance on a range of object sizes between $10mm^2$ and $20mm^2$. Throughout the literature, the recommended sizes based on findings varies widely, with similar-designed studies suggesting sizes of $9.6mm^2$ (Parhi et al., 2006), $12mm^2$ (Schedlbauer, 2007), $15mm^2$ (Tao et al., 2017), $17.5mm^2$ (Kim et al., 2014) $19mm^2$ (Wang et al., 2015) and $20mm^2$ (Colle and Hiszem, 2004).

All of the above studies found a significant main effect of object size on performance. Furthermore, all studies report that the largest object size tested produces the fastest response time and the lowest error rate. The one exception is reported in a study by Scott and Conzola (1997) in which the sizes assessed were 16, 18 and $20mm^2$. All three sizes are so similar that the effect would be expected to be minor. Even the smallest size, 16mm, is above the recommended sizes detailed by a range of the studies presented above and several standards and guidelines, including the main defence standard in the U.S., Military Standard 1472G (Department of Defense, 2012), which specifies a minimum of $15mm^2$.

Most previous studies that assess optimal object sizes do so by testing between two and five variations. In the literature, the majority of studies can be categorised into two types: those testing single object input, and those testing groups of objects such as keyboard and number-pad input. For grouped objects, the object to select is often surrounded by other closely located objects with varying spacing between them. The spacing of the surrounding objects is either fixed for the whole study (e.g., 1mm between objects) throughout all size conditions, or multiple spacing distances are used with multiple testing conditions (such as 1mm, 3mm and 5mm).

Wang et al. (2015) present a study in which five size conditions were assessed between 7 and $19mm^2$, with 3mm spacing between objects. Parhi et al. (2006) assessed five target size conditions between 5.8 and $13.4mm^2$, with 0mm spacing between objects. Studies assessing multiple spacing distances include that of Schedlbauer (2007), who tested two spacing conditions of 1.5 and $4.5mm^2$, and that of Colle and Hoszem (2006), who tested spacing of 1 and $3mm^2$. Both studies report no significant effect of spacing distances on selection time. This indicates that spacing does not affect performance when objects are spaced closely together. The effect of wider spacing on performance is not so widely reported.

Of the studies presented within this section, all used a within-subject experimental design in which all participants completed all object size and spacing conditions. Sample sizes were commonly reported between n=14 (Tao et al., 2017) and n=20 (Colle and Hoszem, 2006; Pahri, 2006). Other studies have used larger sample sizes when assessing group differences. A study by Sesto et al. (2012) used a large sample size (n=52) when assessing a range of object sizes. However, the study used a between-subject design in which multiple groups were assigned based on disability. Each group used a size similar to that reported by the within-subject studies, including the following groups: fine motor control (n=23), gross motor control (n=14), and a control group with no disabilities (n=15). As can be seen, a sample size of around n=20 is the most common sample size found in within-subject studies that assess a range of object size conditions.

All of these studies assess human performance using time-based measures. The terminology they use varies. For example, timing is sometimes referred to as “response time”, “completion time”, “pointing time” and “task time”. However, all of the terms are used to measure the time to complete a single input gesture. Similarly, across the studies, the objects to be selected are referred to variously, even given the same design (colour, ratio, size). The terms used vary between *object*, *button*, *target*, *square* and others. The lack of consistency in terminology between studies makes direct comparison of the studies difficult—especially given the wide variety of size and object design variations assessed.

Most of the studies were completed under university laboratory conditions in which there were no external environmental conditions which could have affected the participants. The academic studies, most of which where participants were seated and desk based, differ in object size recommendations compared to those conducted for defence applications. Cockburn et al. (2017) present one study in which a touchscreen and various object sizes were assessed within a vehicle subjected to varying levels of vibration to simulate turbulence. Others, such as Tauson (2012), account for further conditions that may be found to be present outside of laboratory conditions and recommend object sizes to use in military ground vehicles. No study could be found that attempts to assess military standards with the same methodology used in academic, laboratory-based studies.

Military standards provide further recommendations. As illustrated in Table 2.3, Tauson (2012) recommends that, when using a touchscreen interface, button sizes should be set to 38mm wide and 25mm high with 3mm spacing. However, as illustrated in Table 2.4, Military Standard

1472G (Department of Defense, 2012) states that the minimum button size should be 15mm^2 with 3mm spacing and that the maximum size should be 38mm^2 with 5mm spacing.

Table 2.3: Touchscreen button dimensions and separations (mm) in a military ground vehicle (Tauson, 2012)

	Actuation	Separation
Preferred	38mm x 24mm	3mm

Table 2.4: Military Standard 1472G Touchscreen button dimensions and separations (Department of Defense, 2012)

	Actuation	Separation
Minimum	15mm x 15mm	3mm
Maximum	38mm x 38mm	5mm

Some studies and standards differentiate between buttons and targets and provide different design criteria. For target sizes and spacing, Military Standard 1472G (Department of Defense, 2012), recommends smaller criteria, as shown in Table 2.5.

Table 2.5: Military Standard 1472G Touchscreen target dimensions and separations

	Actuation	Separation
Preferred	13mm x 13mm	-
Maximum	-	6mm

2.2.4.2 Eye-based Input

While not as prevalent as touchscreen systems, further studies have assessed object design criteria based on eye-based interaction. Murata et al. (2004) assessed the performance of a range of target shapes and sizes. They assessed pointing time, or selection time, for four target shapes (a square, diamond, a circle and a rectangle) and three sizes (small, medium and large). As shown in Figure 2.3, the authors found that all target shapes produced a very similar pointing time except for the rectangular target, which produced a significantly slower pointing time for all target sizes.

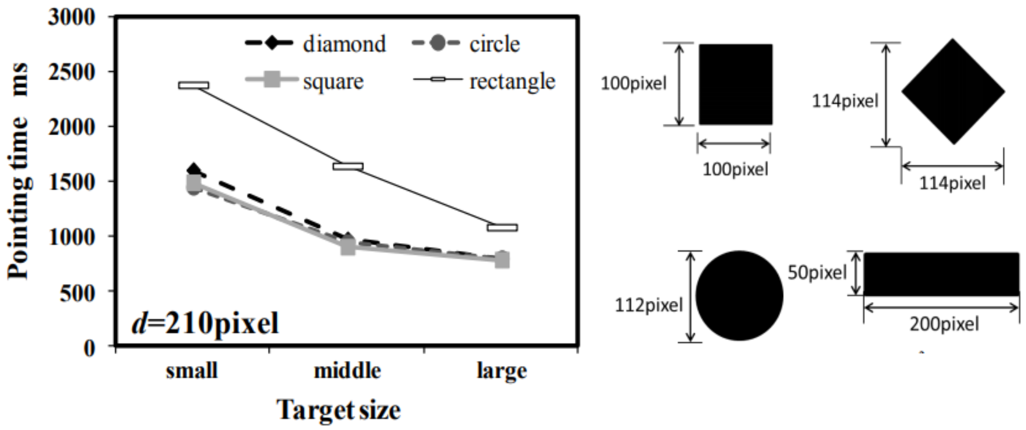

Figure 2.3: Pointing time as a function of target shape and target size (left), target size and shape experiment conditions (right) (Murata et al., 2004)

A second study by Murata et al. (2012) assesses the performance of a range of cursor shapes, including circles, crosses and dots. The authors assessed completion time and error rate when using the various cursor shapes to complete the tasks of selecting 12 icons, 32 icons and 64 icons. As shown in Figure 2.4, only a minor impact was found depending on the shape of the cursor used and the number of icons selected. The authors also report that completion time increased as the number of icons to select increased; however, this may be simply due to fatigue brought about by prolonged and continuous input.

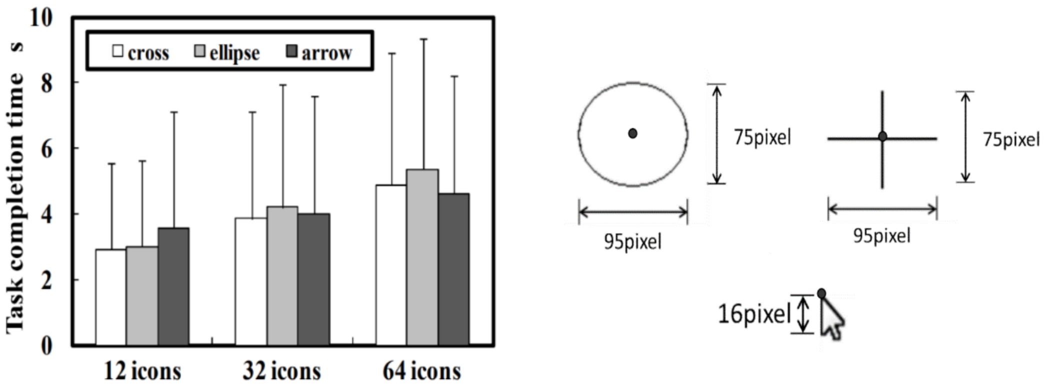

Figure 2.4: Task completion time across cursor shapes and number of input tasks (left), cursor shapes (right) (Murata et al., 2012)

As with most of the touchscreen studies detailed previously, both of the eye-tracking studies presented above used a within-subject, repeated-measures design and required participants to complete all shape and sizing conditions. While they used only a small sample size for the cursor and target experiments (N=16 and N=18 respectively), a statistically significant effect was found across all design criteria conditions. Alternatively, a between-subject study would ensure that no learning effect can occur when participants repeat conditions. However, this would have required a significantly larger sample size if a sample size similar to above was used (such as N=20 for each group with 5 conditions tested, resulting in 100 participants being needed). Therefore, it can be seen that, for studies which assess a large range of conditions, and when the availability of potential participants is limited, a within-subject design is most suitable.

2.3 Presence in Mixed Reality Systems

Stanney et al. (1998) claim that, due to the complex nature of XR systems, it is essential to develop effective multi-criteria measures to evaluate performance. These measures must account for the navigational complexity of interacting with virtual content in 3D environments. This includes measures outside of traditional HF and ergonomic testing. Two of the main factors found to affect human performance within immersive environments are presence and usability (Burdea and Coiffet, 1994).

2.3.1 Presence and Immersion

Presence is one of the most widely measured metrics for evaluating immersive environments, as it is a key concept in understanding and evaluating the effectiveness of virtual environments (MacIntyre et al., 2004). Within an immersive environment, the concept of presence is broadly defined as the sensation of "being there", or more precisely, as the user's physical response to sensory stimuli that result in the sensation that he or she is within the computer-generated environment (Steptoe, 2015). Alternatively, Lombard and Ditton (2006) describe presence as the perceptual illusion of non-mediation. In this context, non-mediation means that the "machinery" used to create the experience (the XR devices) is not evident to the user.

Witmer and Singer (1998) state that both immersion in and involvement with an immersive environment are required to experience presence. Witmer and Singer define immersion as a psychological state that is characterised by perceiving oneself to be enveloped by, included in and interacting with an environment that provides a continuous stream of stimuli and experiences. Furthermore, Witmer and Singer define involvement as a psychological state that is experienced as a consequence of focusing attention on a set of stimuli or meaningfully related activities and events.

Stanney et al. (1998) described several key factors of virtual environments that can impact human performance. These include the level of presence and the navigational complexity of the environment. Stanney et al. (1998) go on to suggest that affordances can reduce the navigational complexity of an environment. Affordances of an environment can be provided by physical objects within the virtual environment, which are often referred to as tangible interface objects. Reducing the navigational complexity of the environment and increasing the user's level of presence within an environment can provide the highest levels of human performance (Stanney et al., 1998). In addition to presence having a direct effect on performance, it is also significantly associated with other factors, including usability (Busch et al., 2014) and enjoyment or satisfaction (Sylaiou et al., 2010).

2.3.2 Perceptual Issues Affecting Human Performance

Three perceptual issues directly affect human sensory and motor capabilities within XR environments: visual perception, auditory perception, and the physiology of haptic and kinaesthetic perception (Stanney et al., 1998) or, in summary, "seeing, hearing and feeling." A literature review by Stanney et al. (1998) provides an overview of the many aspects of a VR system that can result in visual perception issues, many of which result from using an HMD. These include motion blur, field of view, stereo vision and depth perception. In addition, haptic feedback (i.e., touch and force/torque perception) is an important factor in providing a high level of presence, and it has been demonstrated to substantially enhance user performance in VR environments (Burdea and Coiffet, 1994).

Stanney et al. (1998) further explain that, due to the complex nature of interacting with a virtual environment, many factors can affect human performance. Of these, three of the main factors are (i) the navigational complexity of the virtual environment (ii) the degree of presence provided by the virtual environment and (iii) the user's performance as measured by benchmark testing. They go on to describe the relationship between affordances in the real world and human performance levels within a virtual environment, stating that well-designed affordances—which can be provided by tangible interfaces—may reduce the perceived navigational complexity of a virtual environment and thereby increase the user's benchmark task performance. Figure 2.5 illustrates the relationship between presence, performance and three-dimensional navigational complexity (the understanding of the physical world) and human performance.

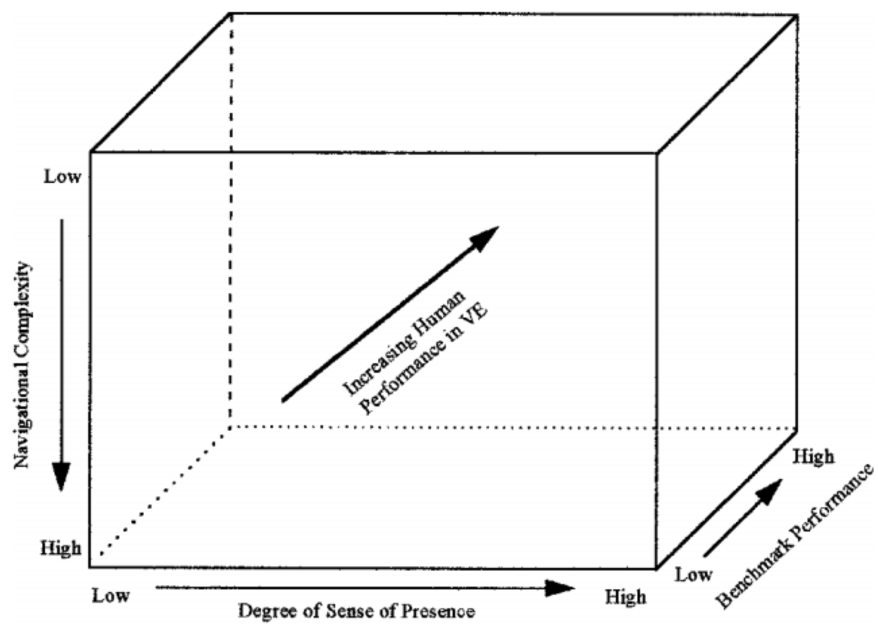

Figure 2.5: Components of human performance in virtual environments (Stanney et al., 1998).

Witmer and Singer (1998) also claim that, to achieve an acceptable level of immersion and presence, a set of criteria must be met. These criteria include (i) the immediacy of controls (provided by selecting input methods and devices with minimal latency) (ii) multimodal richness (provided by allowing proprioceptive feedback and normal movement with kinaesthetic motion, defined as self-movement and body position) and (iii) the scene realism (provided by incorporating many of the environmental elements present within the real scene, such as sound and lighting).

2.3.3 Factors Contributing to Presence

The factor that is often stated to primarily affect presence is *immersion*. The level of immersion within virtual environments has been shown to have a substantial impact on a user's sense of presence (Slater and Wilbur, 1997). However, many other factors are hypothesised to contribute to a sense of presence. As shown in Table 2.6, Witmer and Singer (2008) categorise these factors into the following groups: control factors, sensory factors, distraction factors and realism factors.

Table 2.6: Factors hypothesised to contribute to a sense of presence (Witmer and Singer, 1998).

Control Factors	Sensory Factors	Distraction Factors	Realism Factors
Degree of control	Sensory modality	Isolation	Scene realism
Immediacy of control	Environmental richness	Selective attention	Information consistent with objective world
Anticipation of events	Multimodal presentation	Interface awareness	Meaningfulness of experience
Mode of control	Consistency of multimodal information		Separation anxiety/ disorientation
Physical environment modifiability	Degree of movement perception Active search		

Furthermore, as illustrated in Table 2.7, Kalawsky (2000) shows that the range of variables that can positively and negatively affect presence is vast and is influenced by three main categories. The categories include the sensory outputs utilised, the virtual content used within the system and a selection of user characteristics.

The sensory factors, as defined by Witmer and Singer (1998) and Kalawsky (2000), are based primarily on human sensory modalities (e.g., visual, auditory and olfactory), the method of system interaction used and the technical specifications of the devices used. More recent studies have shown that characteristics of the technologies used can trigger a sense of presence, including technical specifications such as frame rate, latency (Meehan et al., 2003) and visual fidelity (Toczek, 2016). User characterises have also been shown to affect presence, including experience and physiological factors like susceptibility to simulator sickness (Sanchez-Vives and Slater, 2005). In addition, tangibility, the use of physical objects within an environment, has been shown to affect presence and is a key indicator of user performance (Hoshi and Waterworth, 2009; Stanney et al., 1998).

Table 2.7: Variables that can positively or negatively contribute to presence (Kalawsky, 2000)

Variable			Contribution
Form Variables – This group includes the more objective parameters			
Sensory outputs			
	Number of sensory outputs		Positive (for higher numbers)
	Consistency of sensory outputs		Positive when consistent)
	Visual outputs – have various dimensions		Strong – see dimensions below
		Display size	Positive (for larger proportion)
		Viewing distance	Positive (for larger proportion)
		Quality of image	Positive (for high quality)
		Depth cues	Positive
		Camera techniques	Positive
	Audible outputs – also has different dimensions		Strong
	Other sensory outputs (smells, touch etc)		Can be influential but usually less strong than audio or visual
	Body movement and force feedback		Positive when done well
Interactivity of medium			Positive
Visibility/obtrusiveness of medium			Negative
Interference from real world			Negative
Human contact			Positive
Content Variables – Can be both objective and subjective			
Characters and storylines			Positive and negative
Media conventions			Usually negative
Nature of representation			Positive and negative
Media user variables – These are highly subjective and depend directly on the individual			
Willingness to suspend disbelief			Positive
Previous experience			Positive or negative

2.3.4 Measurements of Presence

Klatt et al. (2012) claim that, due to the large number of variables in play, there are various challenges to developing valid and reliable measures of presence. Furthermore, Kalawsky (2000) claims that the complexities of a VR or AR system make it difficult to apply traditional objective and subjective testing tools during the evaluation process. Of the many methods of measuring presence which have been presented within the literature, most can be categorised as either subjective or objective measurements. At present, the most common method of measuring presence is via subjective ratings, which are primarily measured using *post-hoc* questionnaires.

2.3.4.1 Subjective Measures of Presence

The most commonly cited method of measuring presence found within the literature is the presence questionnaire (PQ) developed by Witmer and Singer (1998). The PQ categorises presence into three types, including involvement/control, naturalness and interface quality. Another questionnaire—developed by Usoh et al. (2000) and referred to as the Slater, Usoh and Steed questionnaire (SUS, not to be confused with the Systems Usability Scale questionnaire of Brooke (1996))—measures three themes, including the sense of being there, the extent to which the virtual environment feels like "reality" to participants and the extent to which it feels more like a "place". Both questionnaires used seven-point rating-scale questions and report a high level of reliability and internal consistency.

Another commonly used presence questionnaire is the iGroup Presence Questionnaire (IPQ) (iGroup, 2006) developed by Schubert et al. (1999). The questionnaire is based on a theoretical model of presence which is defined through three sub-elements: spatial presence, involvement and experienced realism. The IPQ is constructed based on a selection of questions taken from the Whitmer and Singer PQ (Witmer and Singer, 1998), from the Slater-Usoh-Steed questionnaire (Usoh et al., 2000) and from a study by Regenbrecht et al. (1998), with additional questions developed for the IPQ.

The IPQ has been validated across a range of technologies and studies (N = 246 and N = 296) (iGroup, 2006). It consists of 14 questions that are grouped into three elements as described above, with an additional single "general presence" scale. As with Usoh et al. (2000) and Witmer and Singer (1998), the IPQ questions contain seven-point ordinal scale ratings ranging from -3 to +3, with a neutral response category of 0. The questions have anchored responses between two extremes, mostly ranging from "fully disagree" to "fully agree".

The use of questionnaires may be inappropriate under certain testing conditions. For example, Usoh et al. (2000) used the Slater-Usoh-Steed questionnaire to measure presence across two conditions: a real office and a virtual office. The study used a between-subject design in which each group completed a task under one of the two conditions. No statistically significant differences were found between the two groups. Usoh et al. (2000) conclude that the questionnaire is valid only when participants compare the same types of virtual environments.

Questionnaires may also be an inappropriate method of measuring presence if they are very long and require completion by participants many times. Some questionnaires are very short and are not burdensome if completed several times in an experimental session. However, if the questionnaires are too long or are repeated too often, the reliability of the participants' responses can begin to deteriorate. This is due to a well-documented phenomenon referred to as "questionnaire fatigue" in which participants' motivation and attention levels drop due to repetitive or continuous questioning (Lavrakas, 2008).

2.3.4.2 Objective Measures of Presence

In addition, many objective methods have been developed for the analysis and measurement of presence. The objective measures include behavioural measures (Regenbrecht and Schubert, 2002), the measurement of reflexive motor acts and sensorimotor control (Mestre, 2005), and a range of physiological measures and techniques. Previous studies that have utilised physiological measurements include the use of skin conductance (galvanic skin response or electro-dermal activity), heart rate (Macedonio et al., 2007) and electroencephalography (EEG) (Clemente et al., 2014).

Because physiological sensors rely on physical changes in the user's body, the use of each measure requires that the environment provides a stimulus that is strong enough to produce observable physiological responses from each user (Meehan et al., 2002). A simple and uneventful environment or task may not produce a meaningful physiological response compared to environments or tasks that are stimulating, stressful or thought-provoking (Qian, 2015). In addition, environmental factors will impact the suitability of the type of measures that can be used. For example, sensors such as EEG devices are strongly susceptible to interference from movement and vibration. Many also require additional calibration and other set-up procedures. While many of these objective measures are available, no study found has validated the use of objective measurements of presence within AR and MR environments.

2.3.5 Presence within Mixed Reality Environments

The majority of studies measure presence with a VR environment, with only several previous studies using subjective and objective methods to measure presence specifically within an AR environment (Klatt et al., 2012). Although presence can be experienced in both VR and AR, Lombard and Ditton (2006) differentiate between the two. They claim that the feeling of "being there" may exist within VR but that AR elicits a different sense of presence: that "it is here". The feeling that "it is here" means that a person can interact and manipulate a virtual object in the same way that they would a physical object without physical or perceptual barriers to interaction. Compared to VR systems, different factors should be considered when measuring presence within AR or MR environments. Steptoe (2015) proposes that a theory of presence, when applied to an immersive AR system, should primarily consider both consistency and content.

Additionally, Steptoe (2015) hypothesises that a higher degree of presence might occur when both the virtual and real content are integrated effectively enough to form a consistent and perceptually unified environment. However, MR presents an additional barrier to interaction and immersion compared to AR: i.e., the need to account for the external, local, physical environment. As such, MR requires the seamless integration of augmented content in the physical environment.

Hoshi and Waterworth (2009) present a study of a "blended reality" game in which participants complete a task—playing tennis—which tests the effects of tangibility of objects on presence. The study was a within-subject study (n=16) in which participants completed two conditions. One condition tested the participants' sense of presence while playing tennis using a physical tool (a controller shaped like a tennis racket). The other tested the participants while using their hands without a physical tool. The method of assessing presence was a modified version of the *post-hoc* subjective questionnaire developed by Witmer and Singer (1998). The study found that participants who used the tangible tool reported a statistically significant higher level of presence than those who used their hands. The study concluded that tangibility (the use of tangible objects as interaction tools) increases users' perceived levels of presence when interacting with the game.

Perhaps the most relevant study used to assess presence within an MR game is one by Schaik et al. (2004). While Schaik et al. (2004) specifically state that they used a level of physical interactivity in their study, they do not state whether tangible interface objects were used. However, the study was a software evaluation and did not follow a formal experimental design. Within the paper, the authors recognise the limitations of the study in that the contributing factors that affect presence were not identified or manipulated. The study used the theoretical model of presence developed by Schubert et al. (2001) and a *post-hoc* questionnaire to measure presence based on a modified version of the iGroup presence questionnaire (IPQ). In addition, subjective interview questions were collected to measure factors such as confidence and motivation to play. No measures were made of performance or usability.

As the study did not manipulate presence through the use of testable independent variables, only the differences between the sub-elements of the presence model used can be assessed. The results indicate that the general presence element was strongly positively associated with spatial presence and involvement elements (Schaik et al., 2004). This is consistent with the statement by iGroup (2006) that the general presence element has high loadings on all three sub-elements, and especially on spatial presence.

In summary, in addition to presence being widely measured across VR systems, several studies have assessed presence within AR environments (Hoshi and Waterworth, 2009; Regenbrecht and Schubert, 2002) and MR environments (Schaik et al., 2004). Tangibility has been shown to affect presence (Hoshi and Waterworth, 2009); however, no experiment could be found that formally assessed the effect of tangibility on presence within an MR environment. By completing an experiment in which a task or procedure is repeated for two conditions (one without tangible interface elements and one with tangible interface elements), the effect of tangibility on the user's perceived level of presence can be measured directly. Further investigation to measure other subjective factors known to affect performance, such as usability and workload, would provide greater insight into the effects of tangibility on the user in MR.

2.4 Context of Use

A major factor that can affect both user and system performance is the environment in which the system is operating in. Most studies reported in the literature are conducted in tightly controlled laboratory conditions where no uncontrollable environmental conditions can occur. Lighting, movement, noise and many other factors are constant and precisely controlled. However, Gawron (2008) states that the environmental conditions of an experiment can have an even greater effect on performance than the independent variables in that experiment.

2.4.1 Interaction with the Physical Environment

The Ministry of Defence (2008a) provides an example of restrictive physical environments within the context of a fast-jet pilot. The document states that upward and rearward head movement can be severely restricted by a tight harness and by interference between the ejection seat headbox and the pilot's protective helmet. The Ministry of Defence (2008a) further states that the movement boundaries of the head or eye have little correlation with the pilot's anthropometric dimensions, such as sitting eye height, but are affected by the user's postural strategy (defined as the muscle pattern and positions of an upright human body). Furthermore, the postural strategy can be affected by the environment the user is in (e.g., a small, physically restrictive vehicle) or the equipment and clothing being worn.

In addition, head movement can be further limited by safety equipment such as the lifejacket, oxygen mask, helmet assembly and other bulky items. Harness tightness has been found to greatly affect the boundaries in which the head and eyes can operate to the extent that an unlocked or slack harness can allow for an additional 10^{o} to 20^{0} degrees of head movement. In addition, if the pilot's eyes are used in addition to head movement for aiming tasks, an additional 50^0 of pointing movement is available when used with a tightened and locked harness.

Figure 2.6 and Figure 2.7 illustrate the head and eye movement boundaries of a pilot with a slack harness and one with a tight harness, respectively. This example of physical restrictions having an impact on eye and head movements can be applied to other contexts, such as seatbelts within a ground vehicle or lifejackets on a ship. This specific limitation of head and eye movement when physically restricted due to environment or equipment should form part of the requirements when designing a user interface. This will ensure that no interactive or display elements are outside of the interaction boundary, and, therefore, that no input buttons or display panels are placed outside the accessible interactive space.

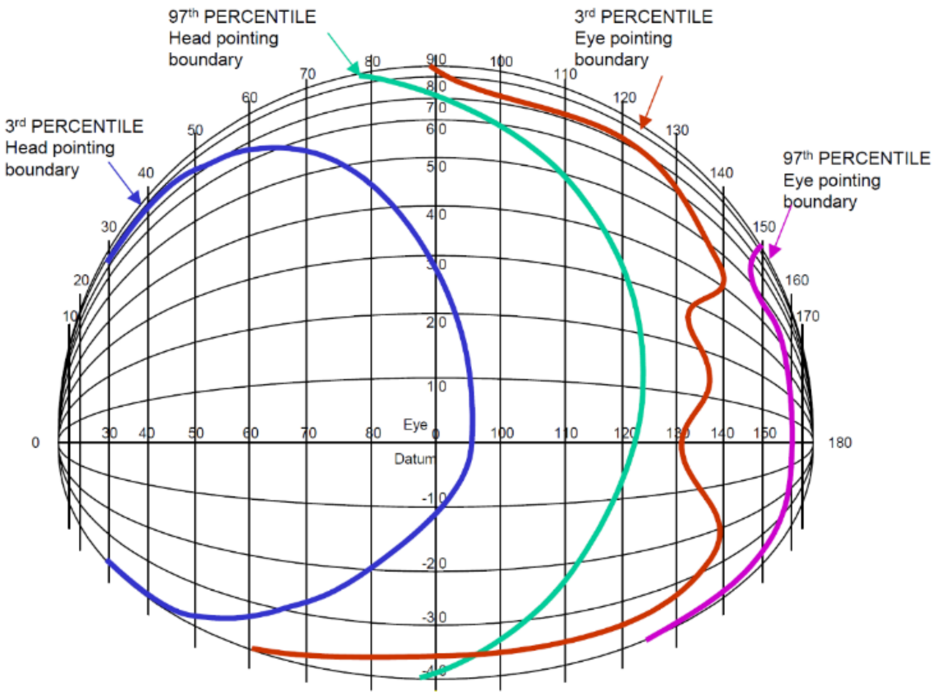

Figure 2.6: Head and eye pointing boundaries with a slack harness (Ministry of Defence, 2008a)

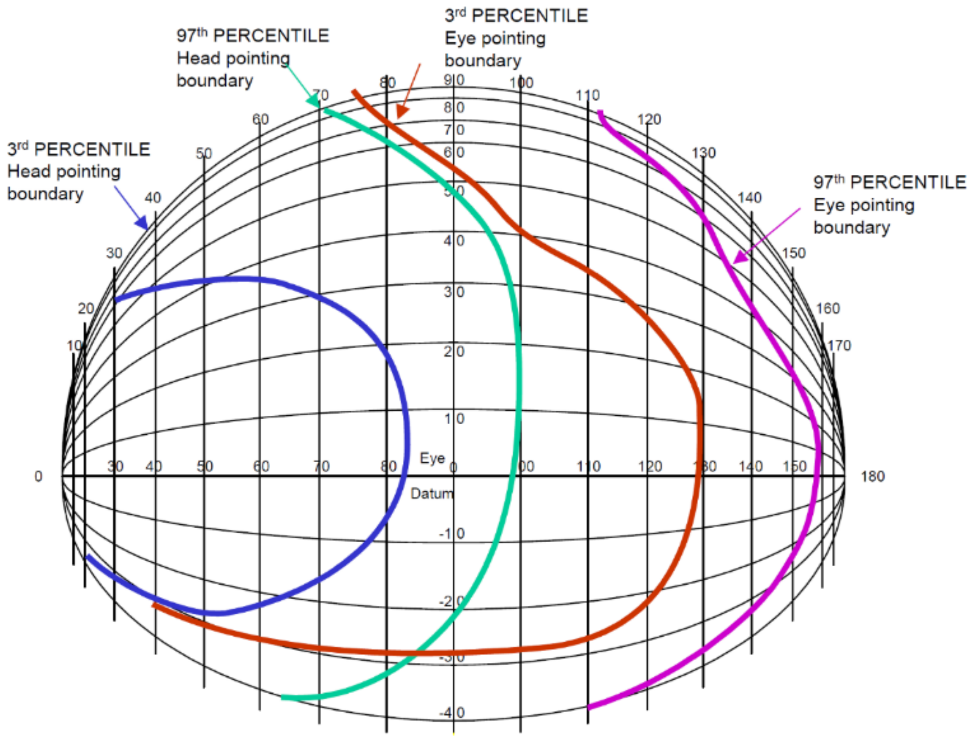

Figure 2.7: Head and eye pointing boundaries with a tight harness (Ministry of Defence, 2008a)

Any equipment or clothing which restricts a user's visual field, such as goggles or a protective/oxygen mask, must also be considered within the interface design process—especially if the task requires the user to view the whole unrestricted visual scene. If a system requires a user to freely rotate his or her head and body, then the system would be unsuitable in physically restrictive environments, such as a tight-harness which restricts a pilot from rotating his or her head beyond a static seating position. Figure 2.8 illustrates the restrictive working environment of a military jet pilot. The body movement is restricted due to the harness, and other movements, such as functional reach distance, are limited by the canopy enclosure.

Figure 2.8: Restrictive working environment of a pilot (Fingas, 2014)

2.4.2 Effect of Environmental Conditions

Many variable and uncontrollable environmental conditions can affect a user who is wearing an HMD. Environments such as aircraft can introduce many problematic factors, including vibration, acceleration, shock and high G-forces. Such factors can cause many issues on HMD systems, primarily causing registration issues.

2.4.2.1 Registration Issues

Using MR systems in an outdoor environment can cause many problems for HMD use. Many of these issues are a result of dynamic registration errors which are introduced into the HMD tracking system. Common issues associated with registration include hardware issues caused by physical aspects of the device design, and software issues associated with the system capabilities and limitations. Azuma (1997) states that registration errors can be divided into static and dynamic errors. Static errors are those which cause registration errors even when the user's viewpoint and the objects in the environment remain completely still. Dynamic errors have no effect until either the viewpoint or the objects begin to move. While dynamic tracking errors can occur with even minor movements, the problems are magnified when used in outdoor environments where lighting and movement are uncontrolled.

Robert et al. (2003) present a study of an HMD being used by a soldier in an outdoor setting. Figure 2.9 depicts the dynamic registration errors that occur, including *jitter* (high-frequency fluctuating motion, evident when staring into the environment), *wander* (low-frequency icon motion, observable over minutes to hours), *lag* (time lag between real-world motion and icon motion during dynamic movements), *bounce* (vertical icon motion with respect to the real-world scene during walking, starting and stopping), and *accuracy* (angular deviation of icon from real-world feature). Roberts et al. (2013) further state that many of the metrics are tempered by classic engineering trade-off situations such as stability vs. manoeuvrability, power vs. efficiency and sensitivity vs. specificity. The most important factor is accuracy, which is a key performance metric. The figure illustrates accuracy when the system is being used by a dismounted soldier in outdoor conditions. Further environmental conditions, such as vibration on a moving platform, may cause further degradation in accuracy.

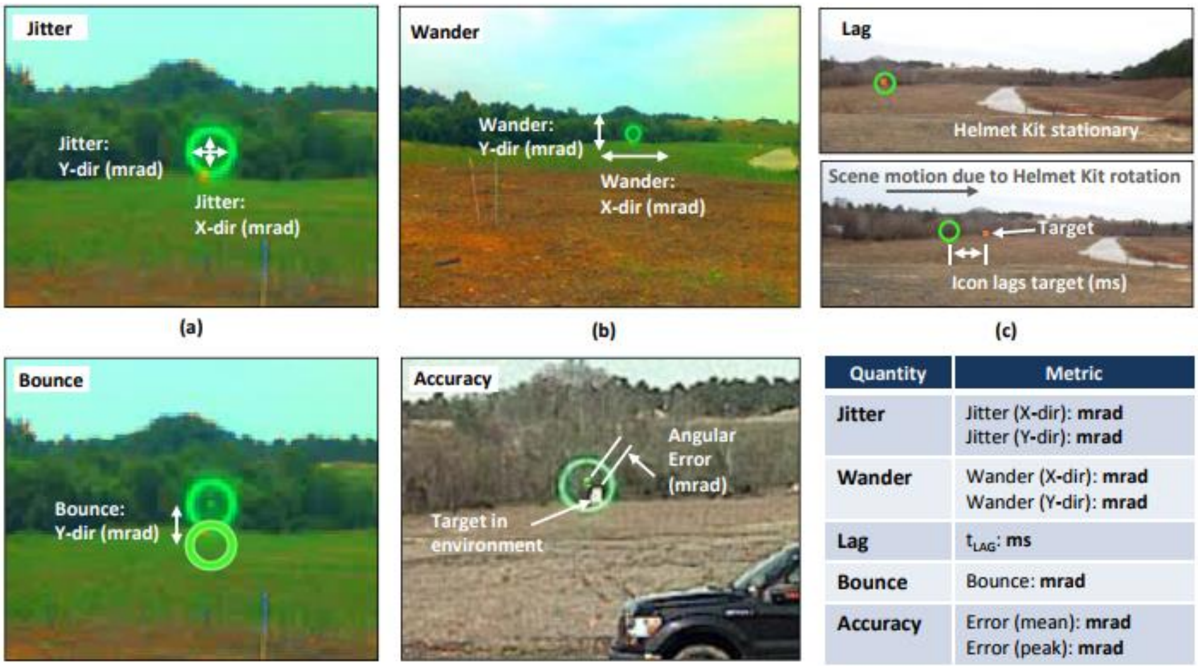

Figure 2.9: Illustration of dynamic error registration performance metrics (Roberts et al., 2003)

2.4.2.2 Vibration on Head and Eye Input

Vibration is an often unavoidable and persistent issue within an aircraft and land/sea vehicles. Vibration can be caused by a range of factors, such as turbulence (e.g., within an aircraft), uneven terrain (e.g., beneath fast-moving ground vehicles) or heavy seas (e.g., with high-speed boats). The intensity and frequency of vibrations can be static if one is moving at a constant speed or dynamic if one is moving over varying conditions, through changing environments or at different speeds.

While many ergonomic and human-centred design solutions can be used to minimise the effect of vibration within an HMD system—including weight distribution, fit, mounting methods, and posture (Knight and Baber, 2007)—these may only account for low levels of vibration. Asare (2015) explains that HMD systems are particularly susceptible to the effects of whole-body vibration, particularly on the tracking and optical sensors used.

Defence Standard 00-250 Part 3 Section 14 (Ministry of Defence, 2008c) provides a detailed overview of the effect of vibration on the human user, the variables influencing the human response and the symptoms and methods of protection. This specific standard details the effects of vibration on hand-arm interaction and the effects of vibration on whole-body motion. The standard further states that the motions of military vehicles and platforms can have adverse effects on the comfort, well-being and task performance of personnel and may compromise health and safety. The Ministry of Defence (2008c) continues to define four primary phenomena which result from induced body motion. Failure to attend to the above issues can result in an increased visual workload and may cause an increase in stress levels.

The four phenomena identified by Ministry of Defence (2008c) are as follows:

a) Motion sickness: low-frequency motion occurring with both short- and long-term exposure.
b) Motion-induced task interruptions: low-frequency, high-amplitude motion, and specific short-term events. Abrupt changes in acceleration are frequently present, e.g., in slamming, jolting or turbulence.
c) Motion-induced fatigue: low-frequency, large-amplitude motion resulting from long-term exposure.
d) Vibration: medium- to high-frequency, with exposure time depending on tolerance to motion severity.

An aircraft (specifically rotary, but not limited to) can produce a tri-axial (X, Y and Z axis) vibrating frequency ranging from 0.5-100Hz in all axes (Rash et al., 1999). However, the ability to use the eye as an effective means of system interaction can be impaired with low-frequency vibration environments. Asare (2015) states that compensatory eye movements become ineffective in stabilising images moving with the head at low frequencies (below 20Hz). In addition, low frequencies also cause visual blurring when a helmet-mounted display is used (Smith, 2004). The Ministry of Defence (2008c) states that vibrations in the 20-30Hz range can affect the head, neck and shoulders, while the eyeball resonates at a frequency in the 60-90 Hz range. It is shown that the range of frequencies experienced within an aircraft, ranging from 0.5 to 100Hz, can cause negative effects on the head (at 20-30Hz) and eyes (at 60-90Hz). Therefore, the effect of vibration on the head and eyes when used as an interaction method for the selection of virtual display elements across a large range of frequencies must be understood.

In addition, it has been shown that many other frequency ranges produce unwanted and troublesome effects, many of which occur at the lowest band of frequencies. 2-5Hz can cause the diaphragm in the chest to vibrate, thus creating a feeling of nausea (Ministry of Defence, 2008c); 4-6Hz causes severe visual performance degradation (Rash et al., 2009) and issues with display legibility (Tung et al., 2014); 2-10Hz produces significant unintended vertical eye movements (Uribe and Miller, 2013); 5-11Hz degrades reading performance by up to 20% (Lewis and Griffin, 1980); 2-20Hz increases target-tracking errors in both the head and eyes (Shoenberger, 1972), and 20-70Hz produces eye resonance which causes significant blurring (Griffin, 1990).

The effect of vibration on the hands has also been examined, with studies suggesting a strong effect on tracking and posture. Martin et al. (1991) report on the effect of high-frequency hand vibration at 150Hz on simultaneous ocular (eyes) and manual (hand) tracking performance. They found a decrement in eye- and hand-tracking performance which strongly correlates with vibration. In addition, Fletcher et al. (1992) examined vibration-induced perturbation of hand feedback on standing equilibrium and found that vibration of the hand can significantly increase sway and instability of posture.

Many simulator systems, including those used for defence training, include the ability to provide whole-body vibration, whether from direct controller feedback or based on the sound output. To experimentally test a specific mode of user interaction which is highly contingent on accurate head- or eye-based control, a method which can precisely and repeatably produce vibration frequencies is required.

One such simulator was developed by Compos and Menegaldo (2018), who present a simulator designed to provide horizontal whole-body vibration. The vibration frequencies were taken from readings within an operational vehicle, in this case, a main battle tank (MBT). In the experiment, the vibration platform provided the same vibration characterises and related tracking errors as that of the user within an MBT travelling over rough terrain. The ability to use a synthetic environment to simulate a real-world environment and platform will, they claim, reduce the time, cost and complexity of testing.

In summary, many studies and standards state that a range of vibration frequencies can negatively impact the head, eyes and hands when used as input devices. Despite this, all three methods of system interaction for MR systems are widely used both within academia and industry.

2.5 Discussion

The literature review presented here is comprised of three main sections. The first section has addressed the main considerations put forward by emerging technologies, such as MR. It has also addressed limitations of the existing design standards when applied to such technologies. This survey of the literature shows that many studies present novel applications of MR technologies across a wide range of industries. Many report higher performance and usability than more traditional display methods, such as touchscreens. Many of these studies are based on evaluations of software concept demonstrators and do not employ traditional and rigorous HF testing as seen for more mature technologies, such as touchscreens or joystick controllers.

Some of the most common methods of assessing system design and performance have been discussed in this chapter, including compliance with standards and guidelines, predictive performance modelling and experimental testing. A wide body of research assesses performance when using traditional and current generation display and interface methods and assesses the optimal design of user interface objects, including object shapes, sizes, colours, panel/workspace placement and many more elements. No such formal experimentation was found for MR systems. To address this gap, this thesis presents two experiments to assess the effect of a selection of design criteria on performance and various subjective measures when using an MR system. These experiments are considered in Chapter 4.

The second section of this chapter addresses the issue of exploiting real-world environments and objects within MR systems and the resulting effect on presence. Presence is a subjective response to a user's experience within a virtual environment which is defined as a person's feeling of "being there". Presence is also reported within AR environments, but as a different feeling of perceiving virtual content to "be there". Many studies state that presence has a positive effect on key factors such as performance and usability. However, presence has also been shown to be affected by a large number of variables. One of the key concepts of an MR system is the relationship with and interaction between the virtual content and the real-world environment.

The use of real, familiar objects within the local physical environment provided by an MR system is referred to as a tangible user interface and has been shown to affect presence within VR systems. However, no study has to date assessed the impact of such interaction on MR users in a formal experimental study. There are many objective and subjective methods for measuring presence within VR. However, the only validated method of measuring presence within MR is the iGroup Presence Questionnaire. The questionnaire is based on the theoretical model of presence developed by Schubert et al. (2001), and is used in a later experiment within the present study. An experiment to assess the effect of tangibility on presence within an MR environment is reported on in Chapter 5.

The third section of this chapter focusses on the effect of environmental conditions when using MR systems in operational settings. It found that vibration is the most commonly uncontrollable environmental condition a military user may experience within a vehicle. Many frequencies are known to affect the user when interacting with displays, and several frequencies are known to affect the user when using eye- and head-based input methods. Many studies assess the effect of vibration on physical display usage, but no study specifically mentions interaction with fully virtual displays within an HMD. This section also detailed the frequency ranges known to occur within an aircraft and the specific frequencies within this range that are known to affect specific user motor functions. An experiment used to assess the effect of vibration frequencies on accuracy when using eye- and head-based interaction is discussed in Chapter 6.

Chapter 3: System Development

3.1 Introduction

This chapter details the experimental software development process together with the key features and functionality of the system. During the design process, multiple iterations of physical mock-ups and design elements were developed. A technology assessment, as reported later in this chapter (Section 3.4), was used to test and evaluate several interaction devices and to highlight their strengths and drawbacks. To integrate the wide range of displays and devices, a modular system design was used. A modular system design allows rapid changes to be made to the interface layout and assets used, and it facilitates easy integration and rapid change-in/change-out of a large number of display and tracking devices. To facilitate the rapid testing and evaluation of design concepts and emerging devices, the system allows for the addition and/or change of content and assets (including text datasets, 3D models and animation) and hardware (including display and interface devices) without the need to modify any code. It also provides the ability to "drag and drop" objects into lists to specify which devices and datasets to use for the experiments.

3.2 System Design

The system (see Figure 3.1) consists of four main elements. First, an interface manager is responsible for the external display interface and user interface layout, which changes depending on the display method being used. Second, the system-monitoring controller automatically collects performance data (such as task completion time and error rates) and continuously and automatically stores data for later analysis. Third, the tracking/registration controller is responsible for the integration of display and interface devices and for configuring the system to use a specific device (such as changing between an HMD and a monitor). Finally, the data-management controller defines the task-related datasets to be used, such as the text and images used in the tasks.

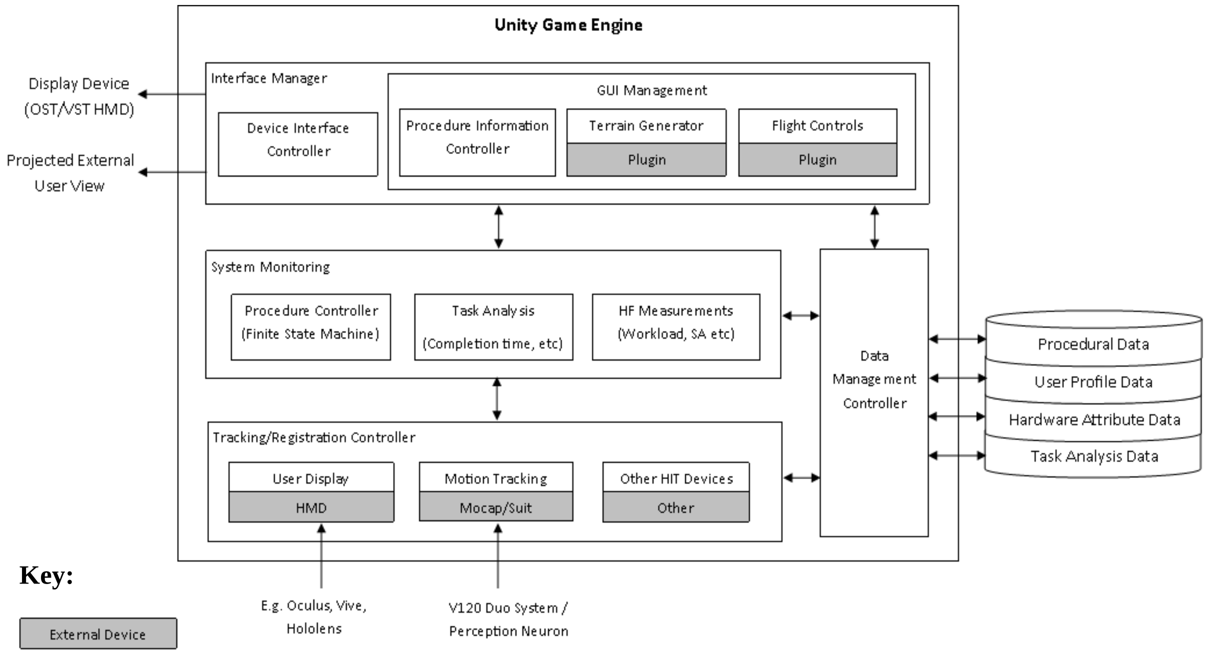

Figure 3.1: System design block diagram

When developing the system, a wide range of devices and techniques were integrated to review the current technologies available for use in MR systems. Whilst the list of potential interactive device candidates is vast, the following technologies and techniques were assessed within the scope of this thesis. These technologies are commonly used methods of system interaction for XR systems and were all available within the Human Interface Technologies Team research group laboratory:

- Tracking technologies
    - Finger/hand tracking
    - Head and body tracking
- Display technologies
    - Video see-through (VST) HMDs
    - Optical see-through (OST) HMDs
- Interaction techniques
    - Head-tracked-based input
    - Eye-gaze-based input
    - Hand- and finger-based input
    - Touchscreen input
    - Various interface devices (e.g., 3D mouse)

3.3 Data Acquisition

A wide variety of authoring methods was used to create the assets in the technology assessment and experiments. The authoring methods used include 3D modelling, 3D reconstruction based on multiple methods of scanning and mapping, and off-the-shelf plugins and 3D assets sourced from online repositories. An overview is presented below of the three main methods of data acquisition and sourcing used and the range of technologies and methods available.

3.3.1 3D Modelling

Developers are able to use a variety of methods when sourcing assets for a project. These may include tools used to support the manual 3D modelling of an object, the sourcing of objects from external suppliers through free and paid databases and repositories or advanced 3D reconstruction methods. Vajak and Livada (2017) describe a system in which multiple acquisition methods are used together to reproduce a real-world location within a VR simulation. The examples shown combine photogrammetry, 3D modelling and real-time information gathering for a highly immersive VR experience, thus illustrating the ability to combine multiple different data types in one system. Manual 3D modelling is the most time-consuming method of authoring and requires knowledge and expertise in modelling software to create models that are optimised for use in real-time systems. Yu et al. (2011) provide an overview of current research and the technical aspects of 3D modelling and describe issues associated with the use of 3D model data, including methods of distribution, compression and protection (copyrighting and watermarking).

3.3.2 Repositories

Data repositories are available in many forms: as online stores, as parts of game engines or as open source or freely available files. Using externally sourced data provides time- and cost-savings, allows expertise to be delegated and provides the most up-to-date and accurate data. A wide range of 3D model repositories exist online, both paid (TurboSquid, 2018) and free (Sketchup, 2018). In general, free models are of basic geometry and low-resolution textures, whilst paid models can be very high fidelity and may be “rigged” (animations applied) and optimised for game engines. In addition, there are repositories of 3D models for the purpose of 3D printing (Thingiverse, 2018), which can be used to develop tangible interface objects, as described earlier. Other alternative methods of sourcing traditional and non-traditional data and information include datasets from social media, location-based data and crowdsourced data. However, these present their own challenges to analysis, interpretation and integration with particular difficulties in data validation, “cleansing” (identifying incomplete, inaccurate or irrelevant parts of a dataset and replacing, modifying or deleting the coarse data; Wu., 2013) and maintenance (Grabowski et al., 2018).

The UK Ministry of Defence provides a shared repository of data sources for simulation-based software systems. The datasets available range from terrain to 3D models (Defence Academy of the United Kingdom, 2018). The 3D-model catalogue (Defence Simulation Centre, 2018) provides a comprehensive list of objects, including land, marine and air vehicles; structures; weapons; avatars (humans) and a wide range of military and civilian objects. In addition, the UK government has surveyed the majority of the UK mainland geography and topography and has made a range of terrain elevation data (including laser-based 3D reconstructions) freely available to the public. The available data formats include Digital Elevation Model (DEM), Digital Terrain Model (DTM) and Digital Surface Model (DSM) (DEFRA, 2018).

These can be processed and converted into 3D mesh objects and textured with aerial imagery to create textured terrain reconstructions. Further terrain elevation data is available from Ordnance Survey (2018). These datasets contain additional data including geo-located addresses and locations, roadways, points of interest and more.

3.3.3 3D Reconstruction

A range of 3D-reconstruction methods was used to capture assets for this study, both for the technology assessment and experimental set-ups. This section describes the main methods used for 3D reconstruction and the projects untaken within this study to capture data for subsequent experiments.

When a required 3D model of a real-world object is not available via a repository—or if the object is too large, too detailed or too complex to be manually modelled—it can be reconstructed using a variety of geometry and textural-capture techniques, including laser scanning, depth cameras and vision algorithms. 3D scanning offers an accurate method of scanning from small objects to very large terrain areas. Unlike vision-based systems, it can scan rough, shiny and dark surfaces, and it is less sensitive to changing light conditions and ambient lighting (EMS, 2018). A large array of 3D scanning technologies is available, ranging from integrated micro-sensors to larger mechanical machines, as reviewed by Ebrahim (2013).

Bures and Polcar (2016) conducted a study comparing 3D scanning and manual 3D modelling of a workplace environment. Their results showed that, whilst the manually modelled 3D model was more visually pleasing (in that it was a rigid and uniform representation of the objects), it was much less precise and took far longer to produce than the 3D scanning technique. As shown in Figure 3.2, the 3D scanning technique produced a reconstruction of the real environment, but was of far lower quality with many errors and artefacts.

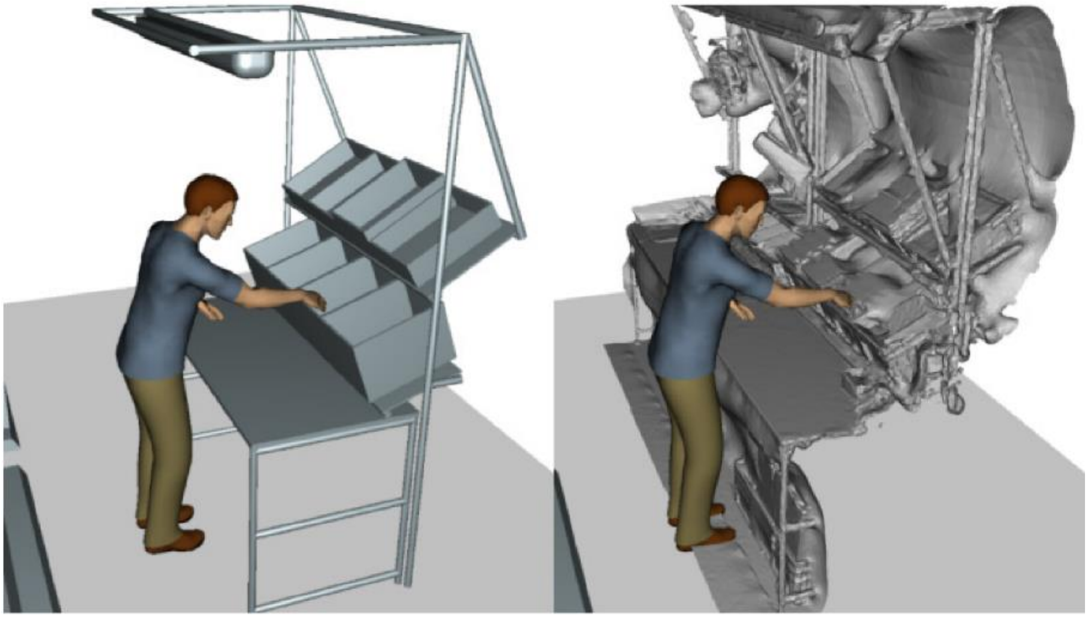

Figure 3.2: 3D model (left), 3D scanning surface reconstruction (right) (Bures and Polcar, 2016)

Photogrammetry provides a method for automatic reconstruction of the geometry of an object based on photographic images alone. Photogrammetry is a technique that uses vision processing to extract 3-dimensional coordinate information from two-dimensional images, thereby creating photorealistic, textured 3D models of the physical geometry. The main benefit of using photogrammetry techniques is that it provides a cost-effective method for rapidly capturing and producing a photorealistic, textured, 3D mesh of large and complex objects such as buildings and terrains. Photogrammetry has also been used extensively to create photorealistic environmental assets for video games, thus allowing developers to increase visual fidelity and reduce the time and cost of creating assets (Valve, 2017). Studies have shown that visual fidelity has an effect on various measures such as performance and presence.

One study, for example, had participants walk around a mountain scape with several conditions adjusting visual fidelity. The study reported that increasing visual realism leads to a statistically significant increase in presence, as measured using the IPQ (Toczek, 2016). The results show that the level of visual realism provided can affect the perceived level of presence, and that realism that can be controlled by the authoring method chosen.

Several photogrammetry projects were undertaken during the course of this study, as detailed below, to provide 3D models for use in later experiments. Figures 3.3 and 3.4 illustrate a photogrammetry survey conducted by a small, unmanned aerial vehicle (sUAV) or “drone” to remotely reconstruct a simulated air-crash investigation site. Such data can be converted into a 3D model through a point-cloud surface-reconstruction process and imported directly into a game engine or other 3D environment. To manually model a crash with the aim of producing an accurate and true model of the vehicle would not have been possible given the complexity of the scene. The ability to rapidly and accurately model complex objects illustrates the usefulness of alternative surveying techniques for 3D model production.

Figure 3.3: sUAV performing photogrammetry scan

Figure 3.4: Point cloud reconstruction from sUAV scan

Additionally, aerial photogrammetry 3D reconstruction has been used for large sites with many complex objects. Another photogrammetric scan completed for this study included the Goonhilly Earth Station site, which consists of many buildings, several satellite dish complexes and varied terrain (Figure 3.5). Second, a quarry site was reconstructed as a very high-fidelity large-area terrain model (Figure 3.6). It would not be possible to reconstruct these two sites from DTM data alone, as the resolution is so low that only general terrain topology is captured. By using photogrammetry, a large-area can be geo-referenced and reconstructed in very high detail, which would otherwise require intensive and prolonged surveying and manual modelling. The building complex which was reconstructed within this study, as shown in Figure 3.5, was used within further experiments (see Chapter 5).

Figure 3.5: Multiple building complex photogrammetry 3D reconstruction

Figure 3.6: Large terrain area with environmental complexity (such as water)

3.4 Technology assessment: “Wearable” Cockpit

A technology assessment was undertaken to assess the suitability and readiness of a range of technologies for use in the subsequent experiments presented in this thesis. The assessment focused on a “wearable” cockpit concept, as described by BAE Systems (2017), in which all physical display and control facilities are replaced by virtual alternatives. To capture usability and qualitative feedback of the devices assessed, the assessment consisted of informal observation and *post-hoc* interview sessions with a range of expert users. The users included subject-matter experts from the industrial sponsor, including system engineers, HF specialists and test pilots.

The devices assessed included display and input methods which may be used by a pilot, including an HMD, touchscreens, joysticks and other emerging technologies such as head-, eye- and hand-based input. By assessing the various technologies, the study aimed to evaluate key performance and usability factors, including robustness, accuracy, wearability and suitability to applications and environments.

The technology assessment aimed to develop and assess the following objectives:

- to integrate a variety of MR display and interaction devices for both physical and virtual cockpit designs;
- to investigate simulated external-world viewing methods (such as terrain and HUD displays) to replicate a flight simulator;
- to investigate methods of interaction with and manipulation of virtual displays;
- to investigate input techniques (hand tracking, head tracking and eye tracking) for interacting with the virtual displays; and
- to investigate the use of a third-person instructor view so that an experiment assessor can monitor users in real-time and to provide instructions.

3.4.1 Physical Testbed

To provide a representative working space in which to house the display and interface devices, a physical mock-up was required. During testing of the initial implementation of the experimental cockpit testbed (Figure 3.7, top image), it became clear that, without restrictions of lower and upper body movement and functional reach, the experiments would not impose the same physical restrictions on system interaction as is experienced by a pilot. This was determined through engagement with BAE Systems subject-matter experts and the chief test pilot. When interacting with virtual-display elements, it was important that the user's movements did not exceed the perimeter of the "cockpit" (e.g., by moving the hand through the area where a canopy would usually be).

Figure 3.7 illustrates the three design iterations of the physical experimental testbed. The first iteration (Figure 3.7, top image) housed the interface devices in an approximate location but allowed free and unrestricted movement of the body and arms. It became clear that by not enclosing the users arms and legs it would not impose the same physical restrictions on movement and interaction as experienced within a cockpit environment. The second iteration (Figure 3.7, middle image) progressed to a testbed consisting of a basic shell which restricted the lower and upper body movement envelope and functional reach of the user. The testbed contained accurately-placed display and interface devices. In order to further increase the physical realism of the testbed and limit hand movements to those capable within manned combat aircraft, further additions were required. The third and final iteration (Figure 3.7, bottom image) provided a further enclosed environment and included part of an actual aircraft cockpit canopy and further enclosed the user's arms and functional reach.

In current combat-aircraft cockpits, the main instrument panel and canopy enclosure partially obstruct the user's visual field of the external world. The same viewing restrictions were included in the system presented within the third testbed iteration. To provide a basic level of visual realism, the design mimics the look and device layout of a current-generation cockpit and utilises real aircraft components such as the canopy windscreen from a *Tornado* military aircraft.

Figure 3.7: Physical testbed design iterations

3.4.2 Simulating External Environments

Multiple methods of viewing the external environment were assessed during the technology assessment. The system provides an interface to enable the user to select between the different display methods using a menu screen. When an option is selected, the layout and all other elements of the system change to become compatible with the new display option. The three options include the following:

i. HMD only (virtual background terrain),
ii. HMD + panoramic imaging (3 large LCD screens), and
iii. HMD + large projector screen.

Two methods were evaluated to visualise the external-world terrain as viewed within an HMD. The first method included the terrain within the virtual environment, and the second method viewed the terrain through a camera as it was displayed on externally mounted screens. The external screen included a multiple-screen setup and a projector-based display.

In both methods, the cockpit display interface is placed directly in front of the seated user and an HUD overlay of the “artificial horizon” is shown overlaid on the terrain. The cockpit display interface is presented virtually in both methods. The terrain used in the technology assessment used a third-party plugin for the Unity game engine that processed digital terrain model (DTM) data and overlaid satellite photography to provide a photorealistic large-area terrain of >100km^2. The source of the data and the methods whereby they were acquired is detailed in Section 3.3.2.

In the first method, the virtual terrain was included within a virtual environment along with the display interface and the HUD. When the terrain is viewed in the virtual environment, both the display interface and terrain are of high visual fidelity, with all elements of the terrain being clear and distinguishable, as illustrated in Figure 3.8 (top).

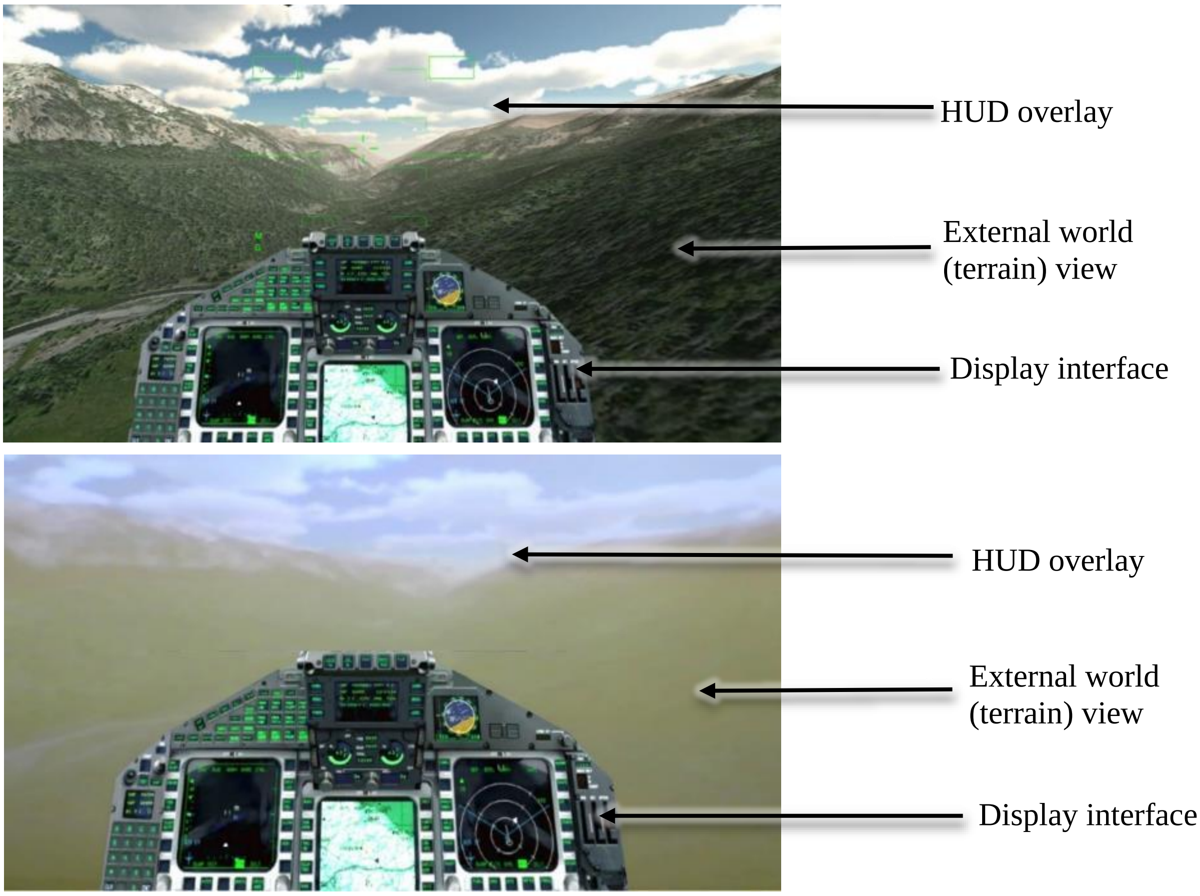

Figure 3.8: Terrain within VR environment (top), terrain displayed on external screens (bottom)

In the second method, the virtual terrain was displayed on externally mounted displays (both a projector screen and large LCD-panel screen) and relayed to the user's view through a monoscopic "pass-through" camera mounted on an HMD. A pass-through camera is a camera mounted on the front of VST HMD which allows a user to view the outside world, as viewed through the camera, beyond the opaque display. As can be seen in Figure 3.8 (bottom), viewing the terrain on a projector screen through the pass-through camera provided very low visual fidelity, and details of the terrain appeared blurry. Terrain imagery projected onto a screen can appear washed out and unclear when the ambient lighting of the room is high. The issues of conflicting luminance and contrast levels when viewing the external world are only present when using a video see-through HMD system. When using a VST HMD, the external world is seen through a pass-through camera displayed on an opaque display. When using an OST HMD, the external world can be seen directly through the transparent lenses without obstruction, delay or reliance on additional sensors that present technological and perceptual issues. However, at the time of development, no commercially available OST HMD with a sufficiently large field of view was available. In addition, current optical systems are limited to indoor low-light conditions because the luminance capability of the display and optics result in glare, high reflectivity and image wash-out.

When viewing the terrain on three large LCD screens via the pass-through camera, the display is less prone to being affected by environmental conditions such as the high ambient lighting of the room. However, the use of monitors introduces additional problems, such as flicker, which is an artefact produced when the camera and the monitor utilize different refresh rates. The difference in refresh rate between the pass-through camera and the monitors can cause synchronisation issues and a potentially distracting perceptual flicker. Consequently, when the user glances away from the monitor (such as to the control devices or hands), the difference in luminance levels causes a delay with the camera system as it adjusts the brightness levels.

4.4.3 Tasks

The technology assessment included a range of interaction methods to complete two complete two types of tasks: selection and interaction with virtual displays in 3D space. The interaction methods assessed included head- and eye-based interaction, and hand- and finger-based interaction.

The first set of tasks required the user to select buttons that were designed to toggle functions on and off. The buttons performed the following actions:

- adding additional new virtual displays;
- toggling display elements on and off, including a 3D map of an area; and
- resetting the layout of the virtual displays.

In the second set of tasks, objects were manipulated in 3D space by performing positioning and resizing gestures. One method used eye- and head-based movements. When the user's head-point or eye-gaze was positioned over the object, he or she held down a button to "slave" it to their movement, then released the button to deselect the object. The other method was hand tracking, in which the user performed a "grabbing" function to select, move and deselect the virtual display panels.

Figure 3.9 illustrates the layout of the system interface. Four buttons are positioned above the three virtual display panels and perform the actions detailed above. The three virtual display panels can be interacted with in 3D space (including repositioning and resizing). Figure 3.10 displays a user in the physical testbed using the head-slaved cursor input. In the upper-right portion of the monitor, a button that has been selected by the user appears in a green colour to show that it is active.

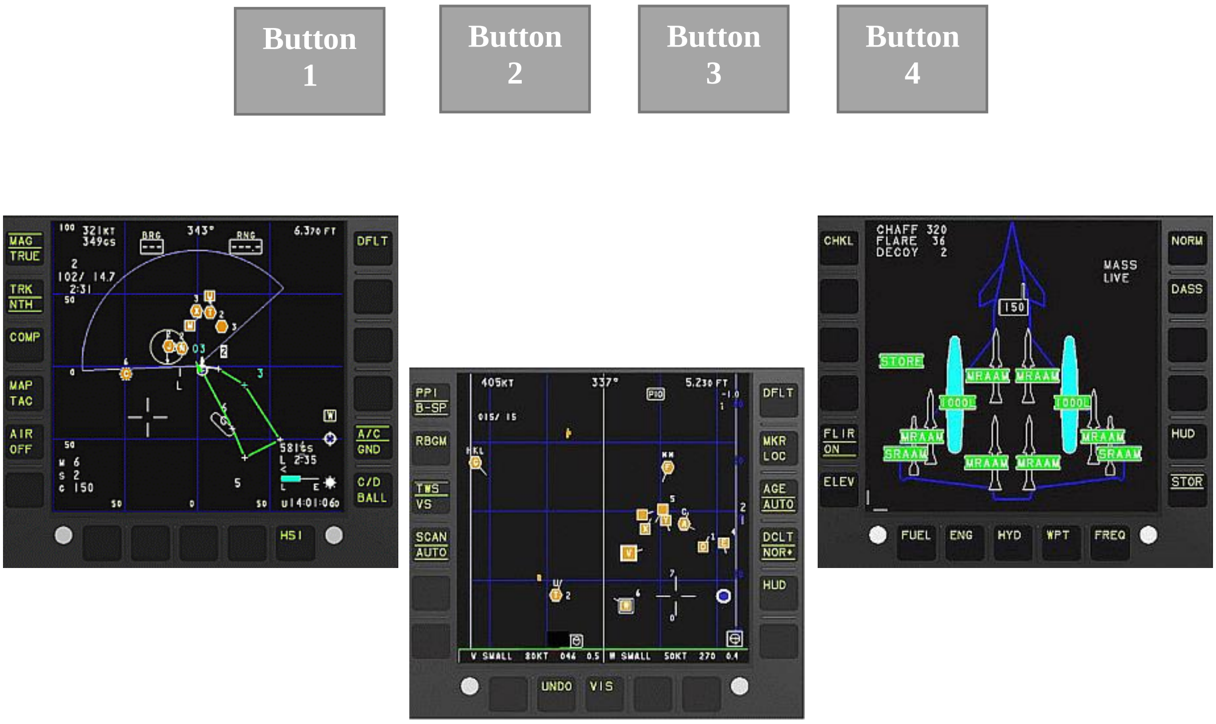

Figure 3.9: Technology assessment system interface

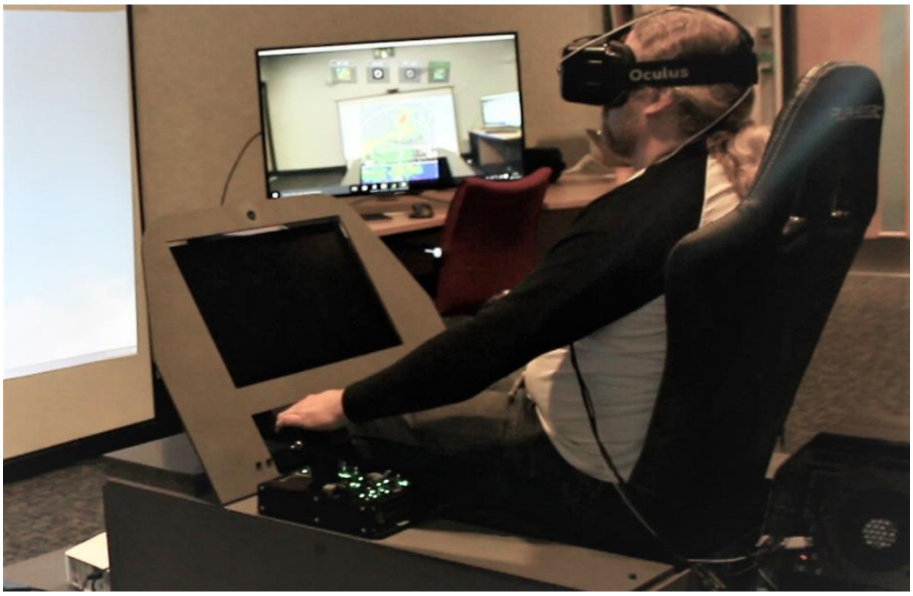

Figure 3.10: Head-slaved interaction with dwell and confirmatory button input

3.4.4 Head and Eye Input

This section assessed two input methods; eye- and head-based input. Eye-based input was achieved by the user consciously looking at an object to interact with it. The point on the display in which the eye is looking at is referred to as the gaze-point and is displayed to the user as a "reticle" or "cursor" on the display. Head-based input was achieved through the user adjusting their head orientation to control a reticle or cursor on the display. Figure 3.11 illustrates the head-based input method. As is shown, as the user yaws, pitches, and to a lesser extent, rolls the head, the head-slaved cursor (shown as a crosshairs) moves to match current orientation of the head.

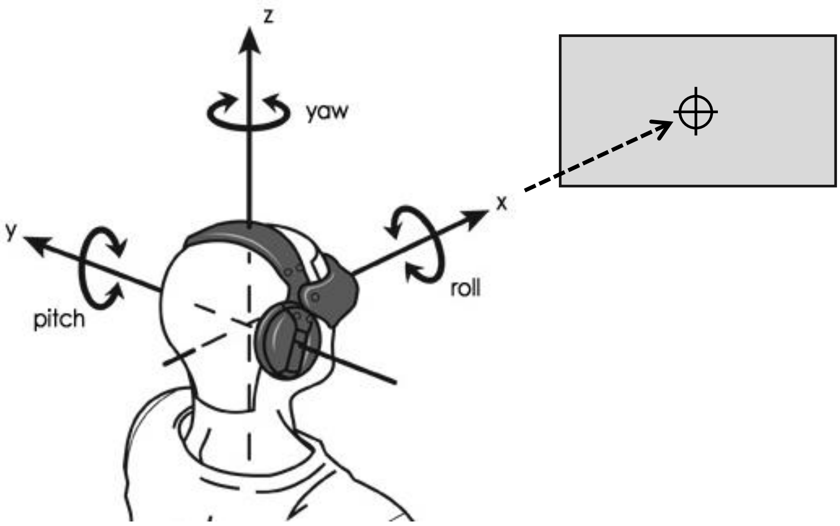

Figure 3.11: Illustration of the head-based input method during input
(modified from Strickland, 2007)

When performing the tasks in the previous section, two methods of interaction were used to assess head- and eye-based input. These include dwell-time activation alone and dwell point with an additional activation confirmation. Dwell time alone required continuously focusing the eye gaze-point or head-slaved cursor on an object for a period of time. A continuous period of dwell time is required to ensure that activation is not unintentionally caused by a passing glance. A drawback to using dwell time as a unimodal input method is the “Midas touch” effect, a situation in which a user’s passing glance over an interactivity area within an interface can unintentionally activate an object or trigger an event (Jacob and Karn, 2003). A sufficiently long dwell period reduces the likelihood of this happening. Dwell-time activation used an activation period of 500ms for both eye-gaze input and head-slaved cursor input. When dwelling on a button, the virtual button started as a light-green colour when the cursor first intersected it and then gradually intensified to a dark-green over a period of 500ms, at which point it became activated.

Dwell point with additional activation confirmation requires placing the eye-gaze or head-slaved cursor over an object and using an additional confirmation method, in this case, a keyboard-key press, to confirm the input. The additional confirmation action ensures that only intentional activation would occur. When using the additional confirmation method, the user moves the head/eye cursor and, when the cursor intersects a virtual button, it changes from light to dark green instantly to indicate the object selected to the user.

The cursor icon used for both input methods is a crosshair reticle. With both methods, the object reverts to its original state (no colour) when the cursor moves away from it. The eye tracker involves the proprietary “smoothing” technique provided by the manufacturer to smooth the rapid saccade eye movement and present a stable cursor.

The main elements and considerations of using head- and eye-based input for simple UI button selection are summarised in Table 3.1.

Table 3.1: Eye- and head-based input elements

	Eye gaze-based selection		Head-based selection	
	Dwell Time	Dwell + Button	Dwell Time	Dwell + Button
Device/System	Tobii EyeX (externally mounted)		Oculus CV1 HMD	
Modules and mounting	One externally screen-mounted tracker		HMD with integrated sensor and one external positional tracker	
Cost	£80		£500	
Interference susceptibility	External IR light can cause minor instability		Little to no effect of external IR light	
Calibration requirements	Calibration once during setup		Initial calibration procedure each use	
Scalability	Limited scalability: Limited by viewing arc of the eye-tracking sensor, seated position		Highly scalable: Tracking area can be extended with additional sensors	
Stability	Inherently unstable rapid eye movements. Requires precise focus for a short period	Inherently unstable rapid eye movements. Minimised due to additional confirmation action required	Very stable, head movement can be precisely and consistently positioned	
Issues	Accidental activation of buttons when glancing around	Use of additional action (button press) causes distraction from the intuitive interaction	Not suitable for environments with vibration	

The second set of tasks required manipulation of objects in 3D space. The input methods used yielded different usability feedback than the set of simple input tasks. For manipulation tasks which require the user to precisely select, manipulate and deselect an object, eye-gaze interaction is found to be unsuitable. During normal vision, the eye performs continuous small rapid movements, known as saccades. Due to the rapid eye movements made during a saccade, when an object is selected and “slaved” to the movement of the eye, it moves rapidly and uncontrollably. However, head orientation was found to be very stable and appears to allow precise positioning by the user with controlled minor head movements.

3.4.5 Hand and Finger Input

Hand and finger input were tested using three technologies: (i) a “motion capture” suit with integrated inertial measurement unit (IMU) based sensors (ii) an externally mounted infrared (IR) based motion-capture system, and (iii) a depth-based sensor. The IMU based motion capture suit required no external devices to work and was a “wearable” device worn by the user. The IR based motion capture suit required an external camera looking at the user for operation. The depth sensor was integrated in to the HMD and also required no external tracking devices.

To assess the different methods of hand and finger tracking, users were asked to select, manipulate and release virtual displays. As with eye- and head-based tracking, three virtual display elements were presented to the user, who would perform a “grabbing” gesture to select the object, move the hand to the desired position, then unclasp the hand to release the object. In addition, a finger-press gesture could be used to select a button. This was performed by reaching the hand out and intersecting the tip of the index finger with a virtual button for 500ms. Removing the fingertip from the virtual button deselected the button.

The first technology assessed—the motion-capture suit—used an array of IMUs distributed throughout a “suit” consisting of gloves and arm straps (see Figure 3.12). The motion-capture suit was unsuitable for prolonged or precise use, as the nature of inertial measurement sensors resulted in inaccuracy (positional drift) and jitter due to the lack of an external device to provide a relative-world position. In addition, this method required a time-consuming procedure for attaching the sensors in the correct location of the body. It also required a multi-step calibration procedure in which the user had to assume multiple body positions.

The externally located IR camera-tracking system used three or more IR cameras to monitor the location of IR retro-reflective spheres, referred to as markers. The markers are worn on the user to triangulate his or her position and orientation. IR motion-capture systems are very accurate; however, the system is also expensive, takes a long time to set up and requires a large workspace to capture user motions. Motion-capture systems are suitable for applications that can be set up in a static location or fixed facility and do not need to be transported regularly.

Finally, the depth-based tracking sensor provided a method of tracking the user’s hand and finger positions and orientation. Tracking each finger accurately allowed the user to perform gestures and complex input commands such as “pinch-to-zoom” as opposed to gross movements such as zooming by clenching both hands and moving them apart. Depth sensors offer the lowest cost and least involved setup procedure of the three methods presented. As such, they are, at the time of writing, the most commonly used and most rapidly evolving tracking and gesture recognition method within consumer AR/MR HMD systems. Using depth cameras also allows for environmental scanning, mapping for tracking and registration, and a method for hand occlusion.

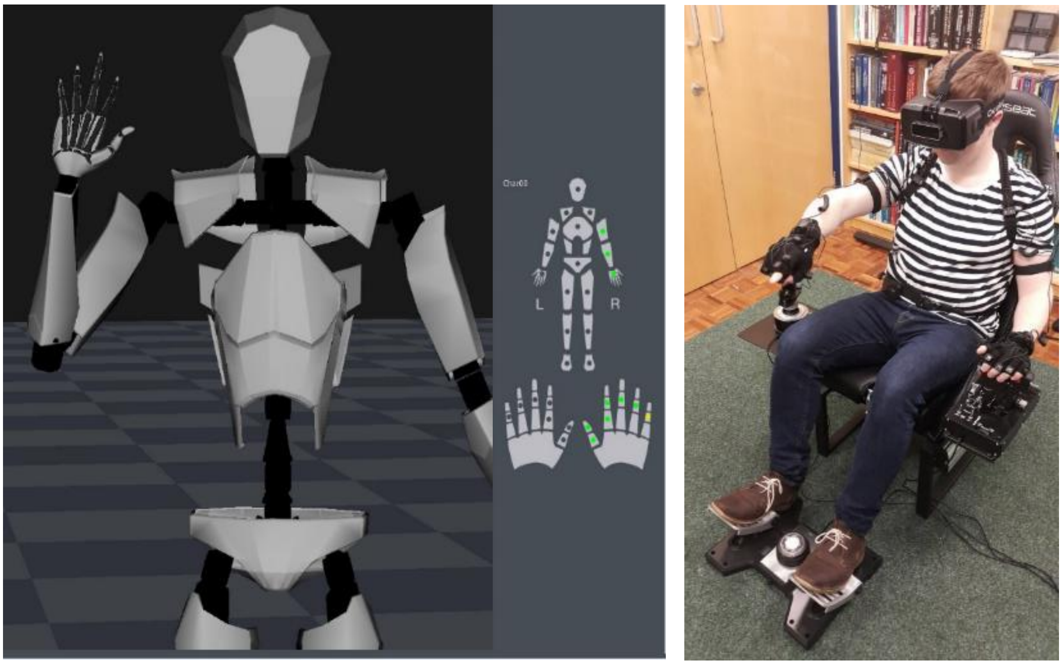

Figure 3.12: IMU-based motion-capture suit

All three hand-input methods were observed to be stable and accurate for simple gestures, such as pressing a large virtual button. However, for complex gestures such as “pinch to zoom”, all were observed to be unstable and often required very slow, exaggerated gestures for the system to recognise the action being performed. In addition, gestures such as grabbing would often require repeating the gesture multiple times; otherwise the system would lose track of the gesture during the interaction. Table 3.2 provides an overview of hand- and finger-based tracking and interaction by comparing the main components of each device type and the technical considerations. These include setup time, interference of environment conditions, calibration time, scalability and accuracy of tracking and the stability and robustness of tracking and input.

Table 3.2: Hand-tracking device comparison

	Motion-capture suit	Motion-capture Cameras	Depth Camera
Device/System	Perception Neuron	Optitrack Flex 13	Leap Motion
Modules and mounting	One suit, 11 sensors on gloves and arm straps	12 IR cameras, gloves with IR markers required	One device, no glove required
Cost	£500	£17,000	£50
Setup	15 minutes	Several hours	No setup required
Interference susceptibility	Highly susceptible to electronic interference	External IR light can cause major instability	External IR light can cause minor instability
Calibration requirements	Long calibration process, calibration needed when interference introduced	One short process at the beginning of the session	One short process at the beginning of the session
Scalability	Highly scalable: Can track basic motions to full sub-finger tracking	Highly scalable: No limits, can add as many tracking objects as needed	Not scalable: Can currently add only one to HMD
Stability	Drift and jitter over time	Robust tracking	Jitter and loss of tracking frequently

3.5 Discussion

This chapter has provided an overview of the design and features of the experimental software and physical testbeds used for subsequent experiments. To assess the most applicable technologies to use when conducting experiments, a review was performed of the wide range of tracking, display and interaction technologies and devices. As part of the review, a technology assessment was conducted to assess the technological and usability impacts of a range of display and interface devices when users performed specific tasks in specific environments. A technical overview of the considerations and drawbacks was further detailed.

The technology assessment reveals a range of limitations with many of the technology devices available and their suitability to MR systems. A physical testbed mock-up was required for the technology assessment, and an overview of the multiple design iterations was provided. Multiple methods of interacting with and manipulating virtual displays were investigated using hand-, eye- and head-based input. Whilst all three appear to allow for the accurate and intuitive selection of simple UI objects, only head tracking allows the robustness and precision required for the precise manipulation of objects in 3D space (such as resizing and repositioning).

Chapter 4: Performance and Usability

4.1 Introduction

A wide body of research studies, standards and guidelines recommend specific design criteria when using current display and interface technologies, such as a touchscreen and joystick controller, within a cockpit environment. However, no study found extends the formal experimental approach to AR/MR systems in which both head-based input and virtual display panels are used instead of hand-based input and physical displays. Virtual displays have numerous advantages over physical displays: They reduce weight, they cost less to change, they allow for rapid upgrades, and they allow the display interface to be customised for each pilot and task without replacing any physical equipment.

This chapter assesses human performance and associated metrics associated with using various HMI devices while conducting a series of interface selection and interaction tasks. To understand the performance and usability of MR systems, studies must be performed to determine the user interface design parameters that provide the highest level of performance and usability. The design criteria commonly assessed in studies such as this include the shape, colour, size and separation distances of interface objects. Therefore, the present experiments build upon standards, guidelines and studies which inform design criteria for currently used technologies such as the touchscreen, and then applying the same methodology to emerging devices such as MR systems in which displays are replaced by fully virtual alternatives.

This chapter describes two experiments: 1a and 1b. Experiment 1a was used to investigate the effect of object design criteria during interface selection on objective performance and subjective human-centred metrics. The experiment assessed a range of HMI technologies and a selection of object types, object sizes and object spacing distances. Experiment 1b was conducted to investigate the effect of HMI technologies used to perform various interaction tasks and their effect on human-performance measured objectively and subjectively.

4.2 Experiment 1a: Selection Tasks

4.2.1 Introduction

This experiment was conducted to assess the effect of various HMI technologies and object design criteria on human performance when performing input tasks. The experiment used an assessment framework which integrated a range of design criteria, including the shape, colour, size and spacing of objects based of a number of conflicting studies and HF standards and guidelines, as detailed within the literature review (Section 2.2.4). The experiment did not aim to assess the standards and guidelines directly but rather aimed to use them to assess the HMI technologies and specific design criteria presented within this experiment. The experiment focused on a simple multi-directional point-and-select task, as described by ISO 9241-9 (International Organisation for Standardization, 2010). ISO 9241-9 is a widely cited and applied standard for human-centred design. A point-and-select task is a generic input task that is common across most interactive systems and it forms the basis of human-machine interaction. Selection tasks are used across most interactive technologies and are used in a wide range of contexts, including ground, air and sea platforms for both manned and unmanned systems.

4.2.2 Aims

The aims of the experiment were rated as either primary (P) or secondary (S). The primary aim of the experiment was to investigate the effect of HMI devices and object design criteria on objective performance metrics. The secondary aim was to investigate the effect of HMI devices on participants based on subjective measures.

The experiment aimed to address the following:

i) (P) Determine whether the HMI device used has an effect on performance (measured by reaction time and error rate) when selecting interface objects within a cockpit-like environment. The null hypothesis is that the HMI conditions have no effect on performance metrics.

ii) (P) Determine whether the object types, sizes and spacing affect performance (measured by reaction time and error rate) when selecting interface objects. The null hypothesis is that the object type, sizing and spacing of objects have no impact on reaction times and error rates.

iii) (S) Determine whether the HMI device used has an effect on subjectively measured parameters, including physical demand, as measured by exertion and discomfort, and difficulty in selecting a range of objects of varying types, sizes and spacing. The null hypothesis is that the HMI device used has no effect on physical exertion, discomfort and difficulty.

iv) (S) Determine if users prefer some HMI devices over others for selecting a range of objects of varying types, sizes and spacing. The null hypothesis is that participants equally prefer all HMI conditions.

4.2.3 Method

4.2.3.1 Participants

A total of 19 participants were recruited, including 15 male and 4 female undergraduate and postgraduate students with a mean age of 26.0 and a standard deviation of 8.6. Participants were recruited using an opportunistic (availability) approach via a recruitment email sent out within the School of Engineering at the University of Birmingham. Ethical approval was granted from the Science, Technology, Engineering and Mathematics Ethical Review Committee within the University of Birmingham.

4.2.3.2 Independent Variables

4.2.3.2.1 Primary Variable

The primary independent variable (IV) within the experiment is the HMI technology used, and consists of three conditions, as illustrated in Figure 4.1:

i. Touchscreen: 23-inch touchscreen capacitive LCD monitor
ii. MR: Oculus CV1 HMD with pass-through camera
iii. HOTAS: “Warthog HOTAS” throttle and stick controller

Figure 4.1: Touchscreen (left), HMD (centre), HOTAS (right)

The first HMI condition is the touchscreen. Touchscreens are common in the latest generation of aircraft and are likely to become more common in future generations. They provide many benefits not afforded by the HOTAS input method, such as allowing for rapid, gross input selection (Tauson, 2012). However, during vehicle operation, touchscreens can be affected by several issues, such as vibration, turbulence (Cockburn et al., 2017), the effect of G-force, and fatigue caused by prolonged arm extension (Savage-Knepshield et al., 2012). While these factors present issues with the use of touchscreens in vehicles, they are nonetheless currently utilised within existing-generation aircraft and other ground and sea vehicles and therefore were included within this experiment. The participants interacted with the touchscreen by using a finger press for selection.

The second HMI condition is an MR system. The MR condition utilizes a VR HMD, as shown in Figure 4.1 (centre), equipped with a pass-through camera and an external tracking system. The latter allows the tracking of participants' head movements in six degrees of freedom, including positional movement and rotation. The camera allows participants to see the real-world environment with the virtual displays superimposed over their field of view. The MR condition contained the same display imagery as the touchscreen and HOTAS conditions, but in this case, they were virtually superimposed as a virtual display in the same position as the touchscreen. The participants navigated and interacted with the display using a cursor controlled by head-tracked movement and a button on the HOTAS throttle stick for selection confirmation. During this assessment method, the touchscreen was deactivated and appeared black so as not to distract the user.

The third HMI condition was the hands-on throttle and stick controller, otherwise known as HOTAS. The HOTAS device is commonly used within current-generation combat aircraft and includes a mouse, buttons and switches placed on the throttle lever and control stick, as shown in Figure 4.1 (right). The control stick allows participants to interact with cockpit functions without removing their hands from the throttle and stick. This method of interaction is most commonly used amongst existing and past generation combat aircraft. The HOTAS condition navigated and interacted with the physical display using a cursor controlled by a two-axis joypad, in addition to a button on the throttle stick for the selection command.

For all three HMI conditions, input selection was based upon “last contact”, as recommended by Tauson (2012). In this method, first contact is made when an object is first selected, and the last contact is made when an object is deselected and thus activated. To achieve this, the participant first presses the object using a finger or cursor, and the object is highlighted. At this point, the object is not yet activated. The participant then releases the press on the highlighted object, thereby activating it.

Table 4.1 provides an overview of the interaction gestures the participants were required to perform for each HMI condition.

Table 4.1: Experiment 1a technology/task interaction table

HMI Technology	Input Method	Selection Task
Touchscreen	Finger press	Touch the object on the touchscreen
Mixed Reality	Head-based cursor	Position cursor over the object, press HOTAS “select” button
HOTAS	Mouse-based cursor	Position cursor over the object, press HOTAS “select” button

4.2.3.2.2 Secondary Variables

The secondary IVs in Experiment 1a consisted of multiple design options for UI objects, including the object type, size and spacing. The first-layer IV was the object type, which contained two conditions (buttons and targets). The buttons were square and white, and the targets were rectangular and red. The second-layer IV was size, which contained three conditions (small, medium and large). The third-layer IV was the spacing of distractors, which contained four conditions (no distractors, near, medium and far). To summarise, each of the two object types contained three sizes, and each size contained four distractor spacing distances, for a total of 24 unique conditions. The specific object types, sizes and spacing used are further detailed in Section 4.2.3.4.1.

4.2.3.3 Apparatus

A physical cockpit-like testbed was built to replicate certain aspects of the working environment of a pilot or military ground vehicle, which involves imposing restrictions on the participant’s head, hands and body movements. The testbed integrated the three HMI conditions: a touchscreen, HOTAS controller and an MR HMD. The participant's view of the touchscreen and HMD was mirrored to a secondary screen so the assessor could monitor the experiment and be informed of any issues or provide instructions to the participant. Figure 4.2 (top) shows the assessor display that mirrors the MR HMD view. Figure 4.2 (middle) shows the MR configuration in which both the HMD and HOTAS devices are used. Figure 4.2 (bottom) shows the touchscreen condition.

As the procedure was repeated for the three HMI conditions, a menu allowed the assessor to select the HMI device to be used and to reset the procedure. Once the assessor selected the “n” key to begin the experiment, the system began recording objective data, and the 24 object conditions were sequentially completed by the participant without any further input from the assessor. Upon completion of the procedure, the menu re-appeared on the screen, and the participant completed questionnaires about the subjective measures. All objective performance data (reaction time and error rate) was recorded automatically from the system and was output as a log file which could be input into a spreadsheet for analysis. Objective data points were recorded for each trial in all 24 object conditions.

Limiting the user’s ability to move freely created restrictions on both input devices and interaction techniques that are similar to those encountered in a cockpit. Additional environmental effects such as vibration, movement and G-force are also present within a cockpit’s working environment, but they were not integrated into this system due to cost and complexity. Gawron (2008) states that the environmental conditions of an experiment can have a greater effect on performance than the independent variables included in the experiment. All input devices, including the throttle, stick and monitors are positioned in the same layout as is found in current-generation combat aircraft.

Figure 4.2: Assessor screen (top), MR condition (middle), touchscreen condition (bottom)

4.2.3.4 Procedure

Prior to the experiment, participants read an information sheet (Appendix A) detailing the experimental procedure. They also reviewed a health and safety information sheet listing exclusion criteria, such as severe eye conditions and a range of health conditions (Appendix B). Next, the participants completed a consent form (Appendix C) that detailed their right to withdraw and affirmed that the data would be anonymised. Finally, the participants filled in a participant data questionnaire (Appendix D), which asked 21 questions designed to determine the participants' knowledge of and skill and experience with the different technologies used.

The experiment employed a within-subject repeated-measures design consisting of three HMI conditions (HMI devices) and 24 object conditions (object type, size and spacing). Participants completed tasks under all 24 object conditions for each of the three HMI conditions. The experiment used a Latin-square method (as shown in Appendix E) for assigning the order in which the participants completed each HMI condition. The order of the 24 object conditions was randomised. The experiment lasted approximately 45 minutes with a short practice session for each HMI condition prior to beginning each procedure.

The tasks assessed a range of object design criteria which varied in the object type, the size and the spacing between surrounding objects. Two object groups (buttons and targets) contained 12 tasks each. The 12 subtasks consisted of three different object sizes (small, medium and large), each containing four different spacing distances between surrounding objects. The surrounding objects acted as "distractor" objects and varied in the proximity to the object to select (no distractors, near, medium and far).

The participants completed 10 trials for each of the 24 tasks, for a total of 240 individual trials per HMI condition. For each of the 10 trials, a single mean variable was calculated for reaction time and error rate and used in the subsequent analysis. To begin the experiment, the assessor selected the HMI condition from the menu. The assessor then pressed the “n” key to begin the experiment, and the first object (“target”) appeared in a random location on the screen. After the participant selected the object, it was then “destroyed” and appeared on a new random location on the screen. In total, the individual tasks were repeated 240 times across all object conditions and trials. Following the completion of all 24 tasks, the display was cleared and the participant answered the subjective rating questionnaires.

The reason for repeated trials of each size and spacing condition is that the position of objects on the display was randomised; thus, the distance of the randomly positioned object from the previous object may have affected the participant's reaction time. If the object was randomly positioned close to the cursor or hand, a decreased reaction time may have resulted; conversely, if an object was randomly positioned far from the cursor or hand, an increased reaction time may have resulted. Therefore, by repeating each task across 10 trials, the potential impact of randomised positioning was reduced.

Distraction objects were used because there are many interface elements on displays which can cause distractions. For example, more than one object can be present on the display at once, as with a radar display in which several detected objects may be clustered closely together. Alternatively, several buttons may be located together with minimal spacing between each one, as with a keyboard or number pad. Therefore, the effect of distractor objects in close proximity to other selectable objects was investigated.

4.2.3.4.1 Tasks

As discussed in Chapter 2, a wide variety of recommended design criteria are evident within the literature. The most common design criteria assessed include the size, shape, colour, placement and a selection of object types. The majority of studies and standards are based around physical displays, predominantly touchscreen displays, with few based around virtual displays or MR systems. Many studies have assessed the effect of object sizes on human performance when using touchscreens, mainly measuring the error rate. As detailed within the literature review (see Section 2.2.4.1), many studies have assessed the performance of a range of object sizes, all of which suggest object sizes ranging between 10 and 20mm^2, from 9.6mm^2 to 20mm^2 (Colle and Hiszem, 2004). No single study found during the literature review process describes research which assesses sizes in addition to further important design criteria including shape, colour and spacing.

While a wide variety of studies have assessed object sizes for touchscreens, they have done so under laboratory conditions without external environmental conditions which may have affected the user. Cockburn et al. (2017) present one study in which a touchscreen and various object sizes were assessed within a cockpit context in which varying levels of vibration were used to simulate turbulence. Others, such as Tauson (2012), account for conditions that may be present outside of laboratory conditions and recommend object sizes with specific reference to use within military ground vehicles.

Object Sizes and Spacing

The experiment employed existing defence-related guidelines and standards as a means for selecting design criteria—in this instance, the colour, shape, size and spacing between objects. The selected criteria included recommendations by Tauson (2012) and Military Standard 1472G (Department of Defense, 2012), as detailed in Section 2.2.4.1. By using military standards, it ensures that the experiment has the greatest degree of relevance to the end-user demographic of this research – military personnel.

For the two object types, buttons and targets, three object sizes were selected. For each of the size conditions, a further four spacing conditions were selected. For object sizes, a total of 12 object criteria variations were assessed for both buttons and targets, as shown in Tables 4.2 and 4.3.

One spacing condition included no surrounding distractor objects, while the other three included surrounding distractor objects of varying spacing. For tasks with surrounding distractor objects present, participants selected the object (button or target) labelled "A" from a block of eight other closely co-located distractors with letters designated B to I. After each object was selected, the block of objects moved to a different randomised position on the screen. The object to select was always placed in the centre of the block of surrounding distractor buttons.

Figure 4.3 and 4.4 lists the button and target size and spacing conditions.

Table 4.2: Button sizes and distractor separation spacing based on Military Standard 1472G (Department of Defence, 2012)

Button Size (Height x Width)	**Distractor buttons separation distance (mm)**	**Button ID**
Small (10mm x 15mm)	No Distractors	1
	1	2
	3	3
	5	4
Medium (18mm x 25mm)	No Distractors	5
	1	6
	3	7
	5	8
Large (25mm x 35mm)	No Distractors	9
	1	10
	3	11
	5	12

Table 4.3: Target sizes and distractor separation spacing based on Military Standard 1472G (Department of Defence, 2012)

Target Size (Height x Width)	**Distractor button separation distance (mm)**	**Target ID**
Small ($5mm^2$)	No Distractors	1
	5	2
	10	3
	15	4
Medium ($10mm^2$)	No Distractors	5
	5	6
	10	7
	15	8
Large ($15mm^2$)	No Distractors	9
	5	10
	10	11
	15	12

Figure 4.3: 10mm x 15mm with no distractors (top)

10mm x 15mm with 1mm spacing distractors (centre)

10mm x 15mm with 3mm spacing distractors (bottom)

Figure 4.4: 5mm x 5mm with no distractors (top)

5mm x 5mm with 15mm spacing distractors (centre)

15mm x 15mm with 5mm spacing distractors (bottom)

4.2.3.5 Dependent Variables

Six dependent variables were measured in this experiment, and they were classified as objective and subjective measures. Performance was assessed using two objective measurements: reaction time and error rate. The two objective measurements were collected for all 24 tasks. The measurements were collected automatically within the system, and logging began once the assessor instructed the participant to begin and pressed the “n” key to start the experiment.

- Reaction Time: Reaction time was defined as the time which passed between the moment the object was displayed to the moment the participant selected it. Because selection was based upon “last contact”, it was considered to have occurred when the participant pressed and then released to activate the selection.
- Error Rate: An error was defined as the selection of an incorrect object or as the selection of a location on the display other than the object.

The second type of dependent variable was subjective in nature and based on the participant’s perceptions of the HMI conditions. Four subjective measurements were taken immediately following each HMI condition (detailed below). All subjective measurements used existing and validated rating questionnaires which are used widely in the literature. The questionnaires use non-technical language throughout and can be deployed to participants with no prior training.

Wearing an HMD or extending an arm for prolonged periods can become uncomfortable. Accordingly, the body postures adopted when wearing an HMD can affect biomechanical loading which may result in the sensation of musculoskeletal discomfort and localised muscle fatigue (Knight and Baber, 2007). To measure the physical demand on the participants, exertion and discomfort were measured following each HMI condition. The participants were required to complete a large number of tasks; ten trials were completed for each of the 24 tasks, for a total of 240 actions by the user. Therefore, it was expected that a degree of exertion and discomfort would occur during the prolonged input periods and would vary depending on the HMI device used. The four subjective measures used are as follows:

i. Exertion Rating: Exertion was measured using the Borg Rating of Physical Exertion (RPE) scale (Borg, 1982) in which participants rated exertion between 6 ("No exertion at all") and 20 ("Maximum exertion"). See Appendix F.

ii. Discomfort Rating: Discomfort was measured using a rating scale (Harich, 2002) that measures discomfort on a range from 1–10 ("very mild" to "unspeakable"). See Appendix G.

An additional two measurements were taken upon completion of the experiment. These included subjective ratings of difficulty and preference across the three HMI conditions.

iii. Difficulty Rating: Difficulty was measured using a scale of 1-10 ("very easy" to "very hard") at the end of the experiment. The question asked is the following: "How difficult would you rate selecting objects for this HMI device? Rate between 1 (very easy) and 10 (very difficult)".

iv. Preference Rating: At the end of the experiment, participants ranked the three HMI technologies in the assessment from 1-3 according to their preference. The following question was asked: "Which of the three technologies in the experiment did you most prefer for selecting objects? Rate the three technologies between 1 and 3, 1 being most preferred and 3 being least preferred".

4.2.3.6 Statistical Analysis

Means are presented in a standardised format displaying the mean value followed by the standard deviation (SD) value, presented as ($M = x \pm SD$). The mean is presented with two decimal places.

Objective data, including reaction time and error rate, was analysed using a repeated-measures analysis-of-variance (ANOVA) test to check for statistical significance between the means of all measures. All assumptions required for using an ANOVA were met, primarily that three or more conditions were assessed which consisted of interval-level data. Across all ANOVA analyses, a p-value of $p = 0.05$ was used as a criterion for statistical significance. Because there were four independent variables—HMI type, object type, object size and separation distances—a four-way repeated-measures ANOVA was used. Mauchly's test of sphericity was used to test for sphericity violation. For repeated-measures ANOVAs, sphericity violation was tested to see if the variances of differences between the condition combinations of related groups were not equal. As suggested by Field (2013), when sphericity was violated, the Greenhouse-Geisser correction is used for the correction of F-values when $\hat{\varepsilon} < 0.75$, and the Huynh-Feldt correction is used when $\hat{\varepsilon} > 0.75$. Subsequent pairwise comparisons were used to test condition combinations for statistical significance.

The subjective measures within the experiment include exertion, discomfort, difficulty and preference. All the rating scales collected ordinal-level data for three conditions and met all the assumptions needed for use of a Friedman test. The Friedman test is used instead of a one-way repeated-measures ANOVA when the data is non-parametric. All subjective ratings met all four assumptions for the Friedman test: namely, that (i) three or more different conditions were used (ii) the participant groups were a random sample of the population (iii) the dependent variable was ordinal (rank order of 1 to 3), and (iv) the data was not normally distributed.

For the subjective ratings, to further examine where differences occurred between the three HMI conditions, a post-hoc analysis using the Wilcoxon signed-rank test of the different combinations of HMI conditions was completed. Because there were more than two HMI conditions, a Bonferroni correction was used to provide multiple comparisons of the three HMI conditions. To apply the Bonferroni adjustment, the significance level used ($p < 0.05$) is divided by the number of tests used (in this case, three) and rounded. Therefore, the adjusted significance level is $p < 0.017$.

4.2.4 Results

4.2.4.1 Reaction Time

First, reaction time was assessed for the primary IV, HMI devices. As shown in Figure 4.5, the touchscreen condition produced the fastest overall reaction time across all 24 object conditions ($M = 92.82, \pm 2.30$); the MR condition produced the second fastest ($M = 172.52, \pm 4.98$); and the HOTAS condition produced the slowest reaction time ($M = 253.89 \pm 12.01$).

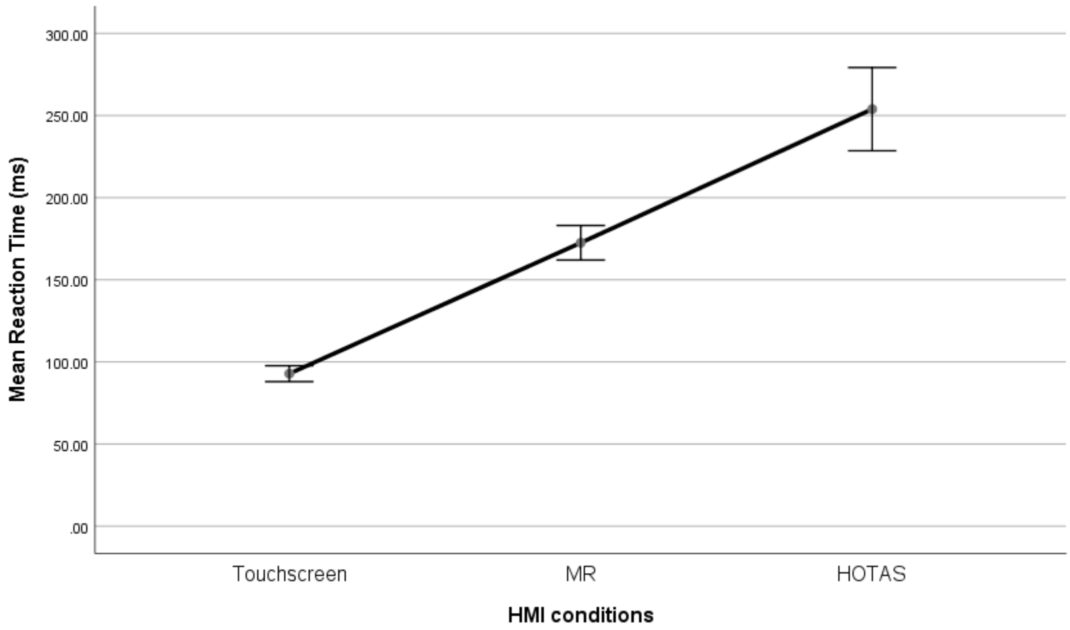

Figure 4.5: Overall Mean reaction times across HMI conditions

Reaction time data was analysed using a four-way (3x2x3x4) repeated-measures ANOVA to compare the overall effect of HMI conditions on the participants' reaction time. The four factors included HMI type (touchscreen, MR, HOTAS), object type (buttons and targets), object size (small, medium and large) and spacing distances (no distractors, near, medium and far). Sphericity was assumed with a significance of $p > 0.05$. However, Mauchly's test of sphericity indicated that the assumption of sphericity was violated. A Greenhouse-Geisser correction was used and the adjusted results demonstrated a significant main effect of HMI conditions on reaction time [$F(1.29,21.22) = 146.59$, $p < 0.001$]. Subsequent pairwise comparisons of the three HMI conditions showed that all conditions had a statistically significant difference ($p < 0.001$). Therefore, the HMI devices used were found to significantly affect reaction times.

Figure 4.6 illustrates the mean reaction times for both button and target objects for the three sizing conditions. As can be seen, the trend described above is repeated across all object types and sizes. The touchscreen produces the fastest reaction, the MR second, and the HOTAS produces the slowest reaction time.

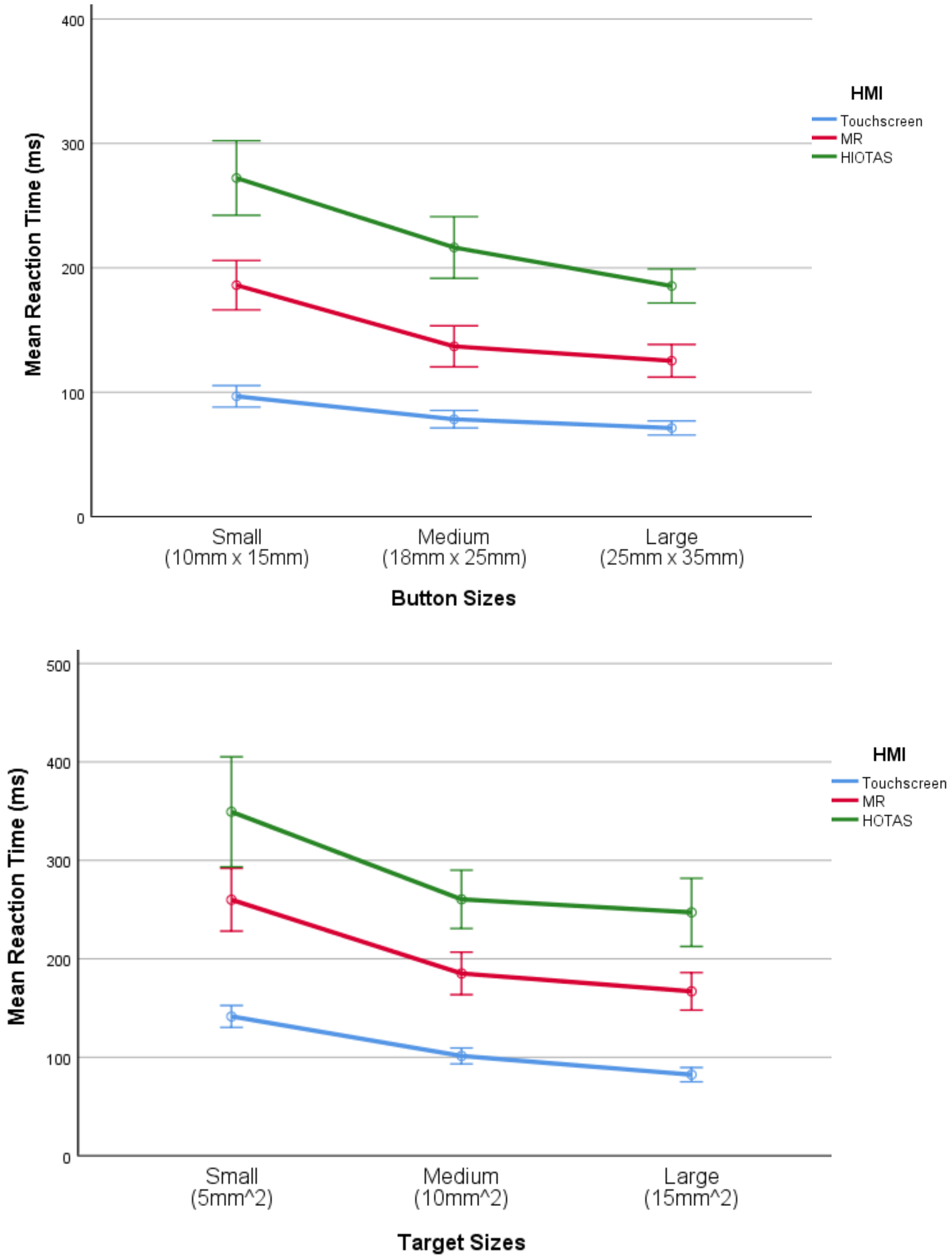

Figure 4.6: Mean reaction time for button sizes (top) and target sizes (bottom) across HMI conditions Vertical error bars indicate the standard deviation from the average values

The secondary IVs—the object types, sizes and spacing distances—were analysed further. A 27% difference in mean reaction time was found across the object Type IV, comprising button and target conditions. Across all three HMI conditions, the mean reaction time was shorter for the button objects (M = 149.55 ± 4.21) than for target objects (M = 196.60 ± 7.07). The mean reaction time for object size IVs varied widely across the three HMI conditions. Overall, it was found that, as the object sizes increased, the reaction time decreased. The small buttons and targets produced the highest reaction time (M = 216.19 ± 8.46). The medium-sized buttons and targets produced the second highest (M=163.06 ± 5.05), and the large buttons produced the lowest (M = 141.64 ± 4.06). A one-way ANOVA showed a significant main effect of object sizes on reaction times [$F(2,34) = 107.85$, $p < 0.001$]. Further pairwise comparisons showed statistical significance differences ($p<0.001$) across all conditions.

The mean reaction time for the spacing IV varied little across the four conditions (no distractors, near, medium and far). The condition with no distractor objects present produced the slowest reaction time (M=176.18 ± 5.79), the “near” (M=173.02 ± 0.24) and “medium” (M= 173.90 ± 5.15) conditions produced very similar results, and the “far” condition produced the fastest reaction time (M=169.21 ± 5.35). A one-way ANOVA found a significant main effect of spacing distances on reaction time [$F(3,15)=4.51$, $p=0.007$]. However, subsequent pairwise comparisons of spacing conditions did not show statistical significance for any condition combinations except for a selection of the “far” condition combinations, as is shown in Table 4.4. Because the mean reaction times between spacing conditions were minor and there was no statistical significance between most condition comparisons, the effect of distractor spacing on reaction time was shown not to have a significant main effect except for the furthest spacing condition.

Table 4.4: Pairwise Comparison of the four conditions within the spacing IV

Key: 1 = No distractors; 2 = "near", 3 = "medium", 4 = "far".

Spacing ID (I)	Spacing ID (J)	Mean Difference (I-J)	df error	Significance
1	2	3.156	1.974	**0.128**
	3	2.282	1.847	**0.234**
	4	6.973	2.083	0.004
2	1	-3156	1.974	**0.128**
	3	-0.874	1.923	**0.655**
	4	3.817	1.711	**0.399**
3	1	-2.282	1.847	**0.234**
	2	0.874	1.923	**0.655**
	4	4.691	2.033	0.034
4	1	-6.973	2.083	0.004
	2	-3.817	1.711	0.039
	3	-4.691	2.033	0.034

4.2.4.2 Error Rate

Twenty-four individual object conditions were tested for a total of 240 individual trials per HMI condition across object types, sizes and spacing. While many trials were completed, few errors occurred overall. Figure 4.7 illustrates the mean error rate per object size condition and HMI condition. A four-way (3x2x3x4) repeated-measures ANOVA was used to assess error rate across four factors. The four factors included HMI type (touchscreen, MR and HOTAS), object type (buttons and targets), size (small, medium and large) and spacing distances (no distractors, near, medium and far). Figure 4.7 (top) and 4.7 (bottom) illustrate the mean error rate for each HMI condition across buttons and targets, respectively.

Overall, across all object conditions, the HOTAS produced the fewest errors (M=0.35 ± 0.06), with the MR condition producing significantly more (M=0.542 ± 0.12) and the touchscreen producing the highest error rate by a significant margin (M=0.90 ± 0.15). A significant main effect of HMI condition on error rate was found [$F(2,34) = 11.14$, $p<0.001$]. The Huynh-Feldt correction was applied for sphericity violation. A further pairwise comparison found significant differences between all size condition combinations ($p<0.05$).

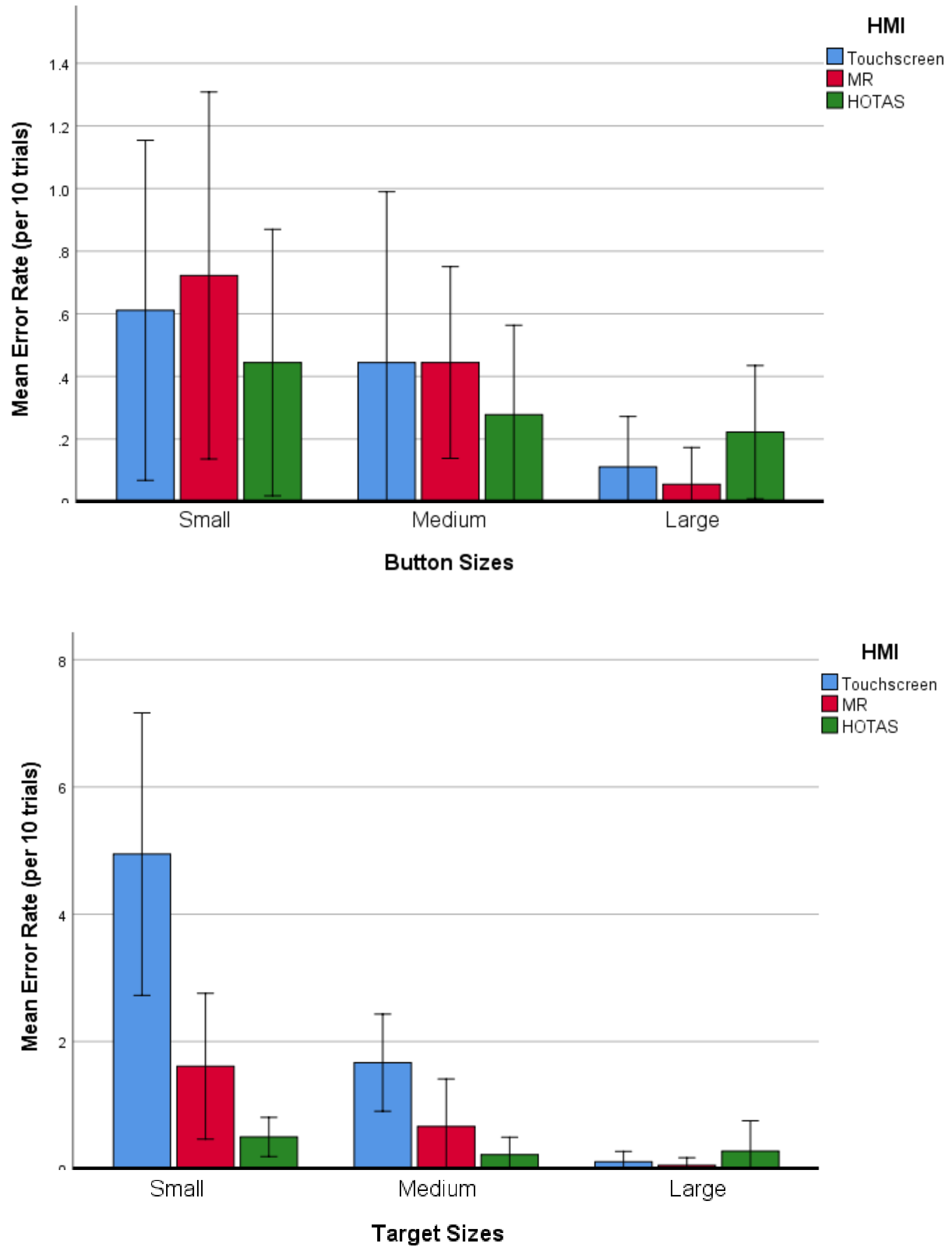

Figure 4.7: Mean error rate for button sizes (top) and target sizes (bottom) across HMI conditions Vertical error bars indicate the standard deviation from the average values

As shown in Figure 4.7, the error rate differed between HMI conditions and between object types and sizes. All three HMI conditions produced similar error results for button objects of all sizes. However, for targets, the touchscreen produced a far greater mean error rate than the other two HMI conditions for both the small and medium target sizes, but not for the large target size. The ANOVA demonstrated a significant main effect of object type on error rate [$F(1.248,21.223) = 146.55$, $p < 0.001$]. Sphericity was assumed. Subsequent pairwise comparisons found significant differences across all condition combinations ($p < 0.001$). A significant main effect of object size on error rate was also found [$F(1.29,21.223) = 146.59$, $p < 0.001$]. Sphericity was met. Subsequent pairwise comparisons found significant differences across all condition combinations ($p < 0.001$).

The spacing of surrounding objects was found to have a minor effect on error rates across all three HMI conditions. The condition with no distractor objects present produced a relatively high error rate ($M=0.74 \pm 0.13$), with the error rate rising across the "near" ($M=0.48 \pm 0.08$) "medium" ($M= 0.59 \pm 0.80$), "far" conditions ($M=0.81 \pm 0.18$). This may be explained by the "near" condition displaying surrounding objects very close to the object to select and could mean participants took extra precaution to ensure that no unintentional input errors occurred. The ANOVA with a Greenhouse-Geisser correction applied found a significant main effect of spacing distances on error rate [$F(1.29,21.22) = 146.548$, $p < 0.001$]. Subsequent pairwise comparisons revealed a statistically significant difference between the "no distractors" and "near" conditions ($p < 0.001$). For "medium" and "far" spacing conditions, no significance was found. Therefore, it can be concluded that distractor targets have a significant, albeit small, effect on error only when positioned "near" other objects.

4.2.4.3 Exertion

Participants completed the RPE rating scale (Borg, 1982) following each HMI condition. The scale rates exertion between 6 (no exertion at all) and 20 (maximum exertion). The results found a minor difference between conditions, all of which were rated within the "very light" bracket.

The exertion rating results found little difference in the participants' perceived level of physical exertion when completing the high volume of input tasks, with all scoring within the rating scale bracket "very light". The MR condition reported the least exertion ($M = 10.25 \pm 3.25$). The HOTAS condition reported the most exertion ($M = 10.86 \pm 3.42$), and the touchscreen condition was rated in between ($M = 10.40 \pm 3.16$). A Friedman test found no significant difference in exertion between the HMI conditions [$\chi2(2) = .646$, $p = 0.724$]. Therefore, no further statistical analysis was required.

The low rating for exertion across all three HMI conditions may be due to the fact that participants were seated and required only minimal movement to complete the tasks. Although the exertion rating shows that all HMI methods required low overall effort, the rating was generalised to whole-body exertion and did not specifically measure aches or fatigue, which is why discomfort was also measured.

4.2.4.4 Discomfort

A discomfort rating scale (modified from Harich, 2002) was filled in following the end of each HMI condition. It used a scale ranging from 0 to 10 (“Nothing at all” to “Extremely strong” respectively). Participants rated all HMI conditions below 3 out of 10, within the bracket denoting “minor—able to adapt to” on the scale.

The MR condition produced the lowest level of discomfort ($M = 1.65 \pm 0.67$), with the HOTAS producing a marginally higher level of discomfort ($M = 2.22 \pm 0.95$) and the touchscreen the highest ($M = 2.45 \pm 1.05$). The touchscreen condition required participants to extend their arms for a prolonged period of continuous movement, whereas the HOTAS condition required only that the thumb be moved, and the MR condition required only minor head movements.

A Friedman test reported a significant difference in discomfort between theHMI device used [$\chi^2(2) = 10.107$, $p = 0.006$]. *Post-hoc* analyses with the Wilcoxon signed-rank test was conducted with the Bonferroni correction applied ($p < 0.05$ becoming $p < 0.017$). A significant difference was found between the touchscreen and MR conditions [$Z = -2.84$, $p = 0.005$]. However, no statistically significant difference was found between the MR and HOTAS conditions [$Z = -1.61$, $p = 0.107$], nor the touchscreen and HOTAS conditions [$Z = -0.80$, $p = 0.422$].

4.2.4.5 Difficulty

Upon completion of all three HMI conditions, participants gave a subjective rating of difficulty for each HMI condition. The following question was asked: “How difficult would you rate selecting objects for this HMI device? Rate between 1 (very easy) and 10 (very difficult)”

Participants rated the touchscreen the least difficult to use ($M = 2.35 \pm 1.42$), the MR condition second ($M = 3.40 \pm 2.06$), and HOTAS the most difficult ($M = 4.17 \pm 3.16$). Participants commented that, when using the HOTAS controller, the smaller precise movements by the thumb-controlled HOTAS mouse were harder to achieve than larger gestures requiring less precision, such as pressing a touchscreen with a finger or moving the head-slaved cursor by looking at the desired location.

A Friedman test revealed a significant main effect of HMI condition on difficulty [$\chi^2(2) = 10.586$, $p = 0.005$]. *Post-hoc* analyses with the Wilcoxon signed-rank test was conducted with the Bonferroni correction applied ($p < 0.05$ becoming $p < 0.017$). A significant difference was found between the touchscreen and MR conditions [$Z = -2.67$, $p = 0.008$] and the touchscreen and HOTAS conditions [$Z = -2.68$, $p = 0.007$]. However, no statistically significant difference was found between the MR and HOTAS conditions [$Z = -1.48$, $p = 0.139$].

4.2.4.6 Preference

Upon completion of all three HMI conditions, participants provided a subjective preference rating of HMI devices. The following question was asked: "Which of the three technologies in the experiment did you most prefer for selecting objects? Rate the three technologies between 1 and 3, 1 being most preferred and 3 being least preferred".

Figure 4.8 shows the total for each preference rank for each HMI condition. Across all 24 object selection tasks, nine participants rated the MR condition as their first preference, compared to eight for the touchscreen. Eight participants also ranked the touchscreen as their second preference. The HOTAS condition was rated the least preferred by twelve participants.

A Friedman test demonstrated a significant main effect of HMI conditions on preference ratings for selection tasks [$\chi^2(2) = 9.380$, $p = 0.009$]. *Post-hoc* analyses with the Wilcoxon signed-rank test was conducted with the Bonferroni correction applied ($p < 0.05$ becoming $p < 0.017$). The test found no statistically significant difference between the touchscreen and MR conditions [$Z = -0.49$, $p = 0.683$] or the MR and HOTAS conditions [$Z = 2.34$, $p = 0.019$]. However, there was a statistically significant difference between the touchscreen and HOTAS condition [$Z = -2.946$, $p = 0.003$].

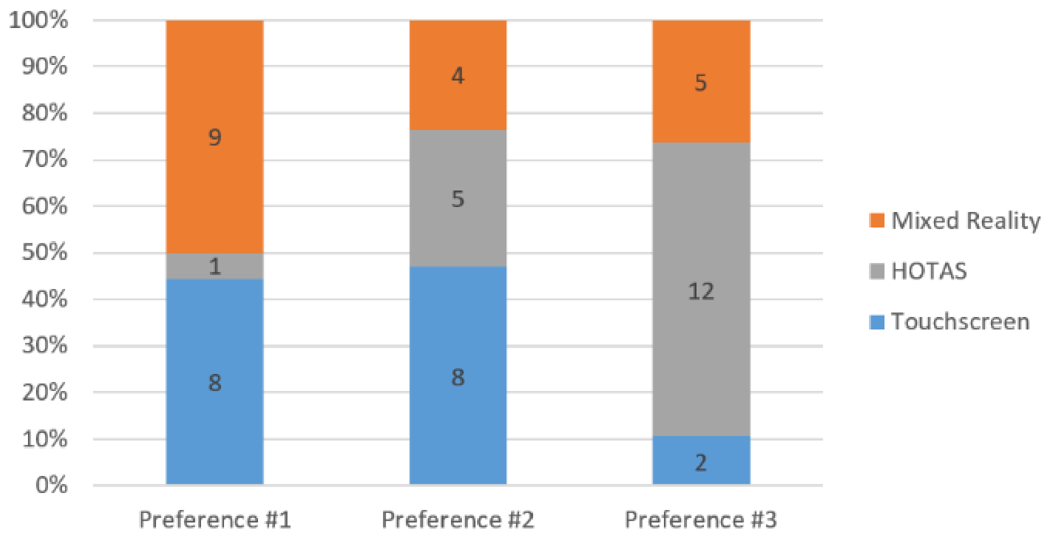

Figure 4.8: Results of preference ratings for HMI conditions

4.2.5 Discussion

The experiment addressed four aims. The first was to determine whether the HMI device used had an effect on performance (measured by reaction time and error rate) when selecting interface objects within a cockpit-like environment. The results found that the HMI device used had a statistically significant effect on reaction times. The touchscreen produced a very fast mean reaction time of 92ms, whereas the MR condition produced a significantly slower mean reaction time of 172ms. The HOTAS was slowest by a large margin at 253ms. To summarise, Experiment 1A showed that the touchscreen is best suited for prolonged and continuous input tasks that requires rapid responses.

While few errors occurred across all three HMI conditions, a statistically significant difference was found. The HOTAS produced the least errors by a relatively large margin, with the touchscreen producing the most. The MR system scored in-between the touchscreen and HOTAS across both performance metrics. Therefore, the null hypothesis that the HMI conditions have no effect on performance metrics is rejected.

The second aim was to determine whether the object types, sizes and spacing affect performance (measured by reaction time and error rate) when selecting interface objects. It was found that overall, all object variables, including object type, size and spacing, have a statistically significant effect on reaction times. The object type has an impact on reaction times in that buttons produce a slightly faster response time than targets, and this was seen for all button sizes. The size of objects also has an effect on reaction times, in that the reaction times decrease as the object size increases. The results indicate that, as the object size increases, the participants are able to select the object with less precision required and thereby increase the speed with which they interact.

The spacing of surrounding objects had very little impact on reaction time, with less than 7ms difference between object conditions with and without distractor objects present. Furthermore, only the objects which were spaced “far” from the object to select had a statistically significant, though very minor, effect on reaction times. The findings are consistent with previous studies which found the spacing distances of surrounding objects to have a very minor effect (Schedlbauer, 2007; Colle and Hoszem, 2006). A statistically significant difference in error rate was found for the object types, sizes and spacing, but only for specific combinations of conditions. Few errors occurred across all object conditions except for the small target sizes, which produced the most errors for the touchscreen condition. In addition, only objects with spacing classified as “far” from each other had a significant effect on error compared to those without any objects spaced around at all. Therefore, the null hypothesis that the type, size and spacing of objects would have no effect on reaction times and error rates is rejected.

The third aim was to determine whether the HMI device used has an effect on participants, as reported using a number of subjective measures (including physical demand, as measured by exertion and discomfort, and difficulty) when selecting a range of objects of varying types, sizes and spacing. The same pattern as was found previously for performance was found for exertion, discomfort and difficulty, with the touchscreen producing the lowest levels, the MR scoring marginally higher and the HOTAS marginally higher again. For the exertion rating, all conditions were scored between “very light” and “fairly light”, and no statistically significant difference was found. Similarly, for the discomfort rating, all conditions were scored within the “minor” category. A statistically significant difference was found only between the touchscreen and MR conditions.

For the difficulty rating, the scores exhibited a wider variance than was found for the exertion rating; however, all scoring was low. The analysis found a statistically significant difference in difficulty between HMI conditions, but not between the MR and HOTAS condition combination. While the results lead to rejection of the null hypothesis (i.e., that the HMI device used has no effect on the multiple subjective measures), the effect found was very minor and occurred only for a selection of conditions. Overall, participants reported minimal exertion, discomfort and difficulty for all three HMI conditions.

The final aim was to determine whether participants prefer some HMI devices over others when selecting a range of objects of varying types, sizes and spacing. The results demonstrate that participants marginally prefer the MR condition over the touchscreen, and the HOTAS condition was firmly rated as least preferred. While the initial analysis found a statistically significant difference, further analysis found significance only between the touchscreen and HOTAS conditions. The null hypothesis (that participants would prefer all HMI conditions equally) is rejected.

4.3 Experiment 1b: Interaction Tasks

.4.3.1 Introduction

This experiment was conducted to assess the effects of various HMI technologies and interaction tasks on objective performance measures and subjective metrics. The experiment tested several types of interaction tasks that are commonly performed on physical and virtual displays.

4.3.2 Aims

The aims of the experiment are rated as either primary (P) or secondary (S) aims. The primary aim of the experiment was to investigate the effect of HMI devices and interaction on completion time and error rate. The secondary aim was to investigate the effect of HMI devices and interaction tasks as rated by participants using subjective measures. The experiment aimed to address the following:

i) (P) Determine whether the HMI devices used affect performance (measured by completion time and error rate) when performing three interaction tasks (selection, rescaling and repositioning). The null hypothesis is that the HMI device used has no effect on performance across the three tasks.

ii) (P) Determine whether the individual interaction tasks affect performance (measured by completion time and error rate) when using the HMI devices. The null hypothesis is that the interaction tasks have no effect on performance.

iii) (S) Determine whether the HMI devices used affect various subjectively measured parameters, including workload, usability and difficulty. The null hypothesis is that the HMI device used has no effect on the subjective ratings.

iv) (S) Determine if participants prefer some HMI devices over others when completing a range of interaction tasks. The null hypothesis is that participants prefer all HMI conditions equally.

4.3.3 Method

4.3.3.1 Participants

The same participants who completed Experiment 1a were recruited. The group consisted of 15 male and 4 female undergraduate and postgraduate students with a mean age of 26.0 and a standard deviation of 8.6.

4.3.3.2 Independent Variables

The independent variables consist of primary and secondary variables. As with Experiment 1a, the primary independent variable was the HMI device used, which was characterized by three conditions:

i. Touchscreen: 23-inch touchscreen capacitive LCD monitor
ii. MR: Oculus CV1 HMD with pass-through camera
iii. HOTAS: “Warthog HOTAS” throttle and stick controllers

The secondary independent variable was the interaction task used in the procedure. This included three conditions: selection, resizing and repositioning.

Table 4.5 provides an overview of the interaction methods for each HMI condition and task. The touchscreen used a finger press for selection and a dragging gesture for interaction tasks. The MR system used both head movement and an additional button on the HOTAS for confirmation. The HOTAS used a thumb-controller mouse to move a cursor with an additional button press for selection confirmation.

Table 4.5: Experiment 1b Task and Technologies Interaction Table

HMI Technology	Input Method	Task 1: Selecting	Task 2: Resizing	Task 3: Repositioning
Touchscreen	Finger press	Touch the target with a finger	Press to activate, drag the corner to resize, release once the desired size is reached	Tap to activate, drag to reposition
MR	Head-based cursor	Move head to target, press HOTAS “select” button	Move head to a resize icon, select icon with HOTAS “select” button, move head to the desired position, deselect icon with HOTAS “select” button	Move head to centre of object, select object with HOTAS “select” button, move head to the desired position, deselect object with HOTAS “select” button
HOTAS	Mouse-based cursor	Position cursor over the target, press HOTAS “select” button	Position cursor over a resize icon, select icon with HOTAS “select” button, move the cursor to the desired position, deselect icon with HOTAS “select” button	Position cursor over the centre of object, select object with HOTAS “select” button, move the cursor to the desired position, deselect object with HOTAS “select” button

4.3.3.3 Apparatus

As with Experiment 1a, the physical cockpit testbed shown in Figure 4.2 was used, which integrated the three HMI conditions. Objective performance measurements, including completion time and error rate, were automatically recorded by the system for all HMI conditions, tasks and trials.

4.3.3.4 Procedure

The experiment employed a within-subject repeated-measures design which included three conditions (HMI devices) and three interaction tasks. The participants completed all three tasks across all three HMI conditions. The experiment used a Latin-square method (as shown in Appendix E) for assigning the order in which participants completed each HMI condition. The order of the three tasks was randomised. For each of the tasks, participants completed 10 trials, for a total of 40 individual trials. The experiment lasted approximately 30 minutes with a short practice session for each HMI device prior to beginning each condition.

The three tasks used for the experiment included selection, repositioning, and resizing. Each interaction task represents a common generic task within a manned cockpit interface, but the tasks might also apply to other manned and unmanned systems on ground, in the air and at sea. The tasks and interface design were defined via discussion with test pilots and HF specialists employed by the industrial sponsor of the study: BAE Systems Military Air and Information (Warton). The participant progressed to the next task once the assessor deemed the current task to have been correctly completed by pressing the “n” key on the assessor's console, as illustrated in Figure 4.2. The system then “destroyed” the objects (“targets”) on the display and created new objects in a randomised position. Following the completion of each task, the display was cleared, and the participant answered the subjective rating questionnaires.

As recommended by Tauson (2012), input selection was based upon last contact. The user selected the target on the first contact; the selected object was highlighted when the operator touched the screen but was not yet activated. The object only became activated when the participant removed the finger or cursor from the display, thereby breaking the last contact. Completion time and errors were measured using this method.

4.3.3.4.1 Tasks

Task 1—Selection

Task 1 required the user to select a circular target which represented a radar object (the "target"). The selection task was repeated 10 times. The completion time for selection of the target (the time lapsed from when the object appeared until it was successfully selected and activated) and the error rate (when the user selected an area of the screen which was not the target, such as the background terrain image) were measured.

The target had a diameter of 15mm. Figure 4.9 illustrates the display and the objects displayed to the participants.

Figure 4.9: Task 1 Interface and Target

Task 2—Resizing

Task 2 required the user to resize an object until it reached a defined size, as dictated by a white outline. The user was required to drag the corner icon (the blue triangle in any of the four corners of the object) to either increase or decrease the rectangle size until it matched the white outline. Once the object was resized to the correct scale so that it overlaid the white outline, the assessor pressed the “n” key to indicate that the task had been completed. The object and outline were “destroyed” and a new object appeared in a random position on the display. The resizing task was repeated 10 times. The completion time (the time it took to reposition the object to the correct location) and the error rate (when the user selected an area of the screen which was not the object or incorrectly resized the object) were measured for each trial.

The resize corner area was 15mm x 15mm. The active area for selecting the corner comprised the whole 15mm^2; however, the displayed icon was a triangle that covered only half of the area. The triangle served as an arrow to indicate the direction in which the rectangle would be resized when the user interacted with and dragged the icon. Figure 4.10 illustrates the object and interface design.

Figure 4.10: Interface before task (top), interface after task (bottom)

<u>Task 3—Repositioning</u>

Task 3 required the user to select an object and move it to a defined position, as dictated by a white outline of the same size as the object. Once the object was placed in the correct position so that it overlaid the white outline, the assessor pressed the “n” key to indicate that the task was complete. The object and outline were “destroyed”, and a new object appeared in a randomly positioned point on the display. The repositioning task was repeated 10 times. The user's completion time (the time taken to reposition the element to the correct location) and the error rate (when the user selected an area of the screen which was not the object or incorrectly positioned the element) were measured.

Any part of the object could be used to select and drag the object, except for the resizing icons in the corners. Figure 4.11 illustrates the object and interface design.

Figure 4.11: Interface before task (top), interface after task (bottom)

4.3.3.5 Dependent Variables

The dependent variables in this experiment consisted of objective and subjective measurements. Objective measures included the completion time and error rate when completing tasks. Subjective measures included the user’s perceived level of workload, usability, and difficulty when completing tasks and preference rankings of HMI devices.

Performance was measured using two objective measurements: completion time and error rate. These measurements were collected for all HMI conditions and all tasks. The objective measurements were collected automatically by the system and were logged starting when the assessor instructed the participant to begin and pressed the “n” key to start the experiment.

- Completion time: Completion time was defined as the time elapsed between the time when an object was displayed the time when the participant selected and activated the object based on last contact. The timer began when the assessor pressed the “n” key to begin.
- Error rate: An error was defined as the participant either selecting a location on the display that was not the object or incorrectly performing the gestures required to complete a task.

Subjective usability and workload ratings were collected for each HMI condition following the completion of each of the four task groups. Information regarding workload and usability was collecting using rating-scale questionnaires.

i. Workload: Workload was measured using the National Aeronautics and Space Administration Task Load Index (NASA-TLX) questionnaire (Hart and Staveland, 1988). See Appendix I.

ii. Usability: Usability was measured using the System Usability Scale (SUS) Questionnaire by Brooke (1996). See Appendix H.

In addition, participants rated the difficulty of each of the three HMI conditions and identified their preference amongst them at the end of the experiment.

i. Difficulty Rating: Difficulty was measured at the end of the experiment using a scale of 1-10 (“very easy” to “very hard”). The following question was asked: “How difficult did you find it to complete the interaction tasks for this HMI device? Rate between 1 (very easy) and 10 (very difficult)”.
ii. Preference Rating: Participants ranked the four HMI technologies in order of preference at the end of the experiment. The following question was asked: “Which of the three technologies in the experiment did you most prefer for interaction tasks, including repositioning, resizing and zooming? Rate the three technologies between 1 and 3, 1 being most preferred and 3 being least preferred”.

4.2.3.6 Statistical Analysis

Mean values are presented in a standardised format displaying mean value followed by standard deviation value, presented as (M = x ± SD). The mean is presented with two decimal places.

Objective data, including completion-time and error rate data, was analysed using an ANOVA test. Because there were two independent variables (HMIs, tasks), a two-way, 3 (HMI) x 3 (tasks) repeated-measures ANOVA was used. A p-value of $p = 0.05$ was used as a criterion for statistical significance. Mauchly's Test of Sphericity was used to determine whether sphericity was violated. Because more than two conditions were used, subsequent pairwise comparisons were used to test condition combinations for statistical significance.

Subjective data was analysed using multiple methods. The usability and workload questionnaire employed multiple Likert-scale ratings, which are ordinal variables; however, under certain circumstances, the data can be treated as interval data and analysed using an ANOVA. While the SUS and NASA-TLX questionnaires involved multiple Likert questions, when represented in a single variable output of overall workload and usability, the data can represent interval data. The use of an ANOVA for summarised scores of ordinal questionnaires can be seen widely across the literature, such as for SUS (Trujillo, 2011) and NASA-TLX (Bowers, 2014; Qian, 2015; Kitchin and Baber, 2018). Therefore, usability and workload were analysed using a one-way (HMI), repeated-measures ANOVA with testing for sphericity violation and subsequent pairwise comparison to test condition combinations.

Difficulty and preference data was of ordinal scale and were analysed using non-parametric tests. In this case, a non-parametric version of an ANOVA, the Friedman test, was used. The Friedman test was chosen because all the assumptions for use were met, including that it requires at least three conditions, random sampling, and ordinal-level data. Subsequent pairwise comparisons were used to compare the combinations of conditions. The pairwise comparisons used a Wilcoxon signed-rank test. Because there were three conditions, a Bonferroni adjustment was used with an adjusted level of $p < 0.017$.

4.3.4 Results

4.3.4.1 Completion Time Analysis

Completion time was analysed for all HMI conditions and tasks. As illustrated in Figure 4.12, overall, the touchscreen condition produced the lowest completion time (M = 247.89 ± 11.98), the MR condition produced the second lowest completion time (M = 353.77 ± 18.75) and the HOTAS condition produced the highest completion time (M = 556.83 ± 27.20).

A two-way, 3 (HMI conditions) x 3 (tasks), repeated-measures ANOVA was used to analyse the effect of HMI conditions on completion time. The assumption of sphericity was met. The ANOVA revealed that the HMI condition had a significant main effect on task completion time [$F(2,32) = 100.49$, $p < 0.001$]. Subsequent pairwise comparisons found statistical significance across all HMI condition combinations ($p < 0.001$).

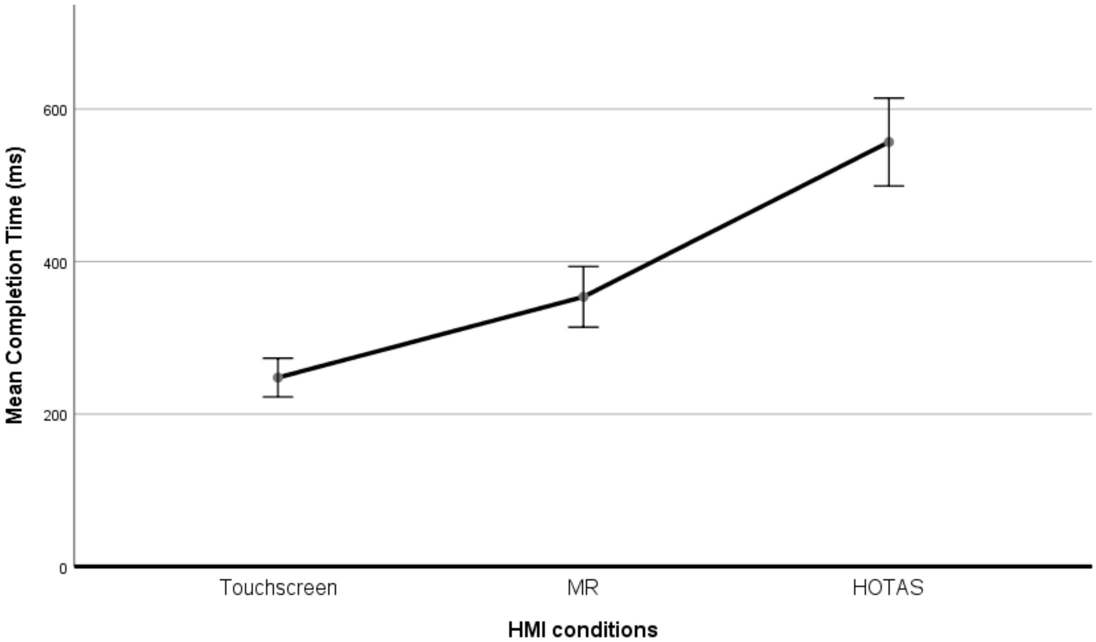

Figure 4.12: Mean completion time (ms) for HMI conditions

Vertical error bars indicate the standard deviation from the average values

The large difference in completion time may be explained by the varying number of steps and the amount of effort required by each HMI condition to complete each task. Table 4.5 details the interaction gestures that the participants were required to complete per HMI condition. These varied from the touchscreen, which required a single reach-and-press gesture for selection; to the MR system, which required head movement and a button; to the HOTAS condition, where participant had to use an X-Y plotter to move the cursor across the display and then press a button to select.

Figure 4.13 further details the completion times for each combination of HMI and task conditions. It was found that the touchscreen produced a lower completion time across all three tasks, with MR second and HOTAS third. Across the three HMI conditions, the selection task reported a lower mean completion time compared to the resizing and repositioning tasks, which scored similarly. The exception was the resizing task in the HOTAS condition, for which the mean completion time was greater than the repositioning task.

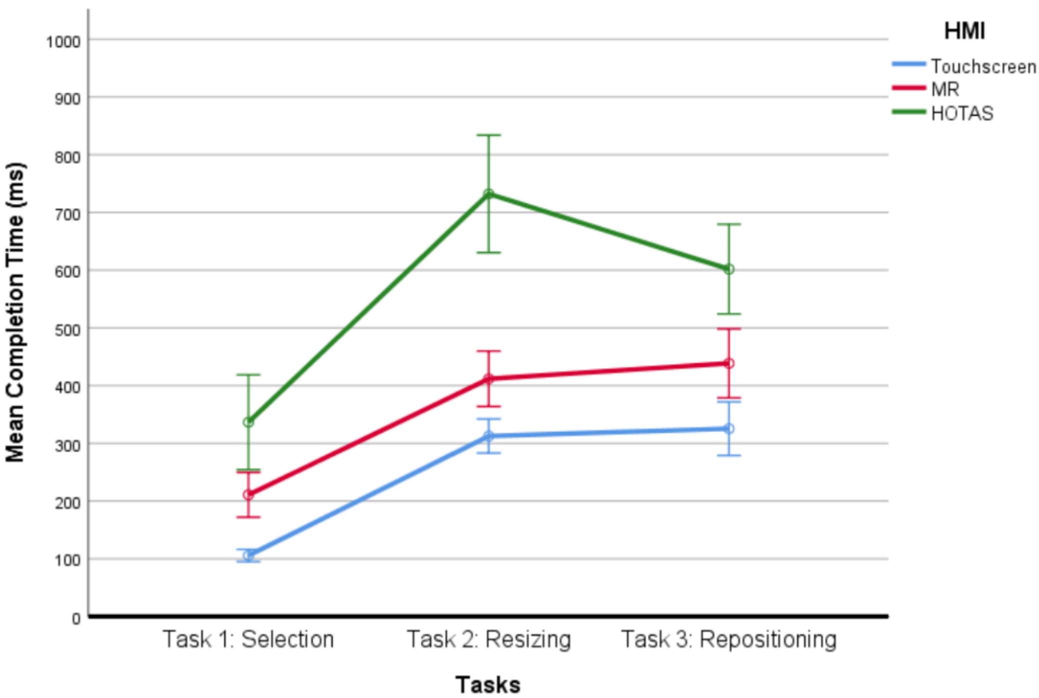

Figure 4.13: Mean completion time (ms) for tasks and HMI conditions

Vertical error bars indicate the standard deviation from the average values

The effect of tasks on completion time was further analysed using the ANOVA. Mauchly's test of sphericity found that the assumption of sphericity was violated and Greenhouse-Geisser correction applied. Using the correction, the three tasks had a significant main effect on completion time [$F(1.47, 23.5) = 79.94$, $p < 0.001$]. A pairwise comparison reported statistical significance across all task comparisons except for the resizing and repositioning tasks, for which no statistically significant difference was found [$Z = 30.286$, $p = 0.65$].

4.3.4.2 Error Analysis

Overall, participants made few errors when completing the interaction tasks. Figure 4.14 illustrates the total number of errors for each HMI condition and interaction task condition. Ten trials were repeated for each task, and the mean error rate shown is the total across all 10 trials. Therefore, the average per trial is the stated figure divided by 10. A two-way, 3 (HMI conditions) x 3 (tasks), repeated-measures ANOVA was performed. It found a significant main effect of HMI conditions on error rate [$F(2,34) = 5.586$, $p = 0.08$]. Sphericity was assumed. A further pairwise comparison found a significant difference only between the touchscreen and HOTAS conditions [$Z = 0.37$, $p = 0.14$].

Several errors were recorded for all three HMI conditions for the selection task. No errors were recorded for the repositioning task for any of the three HMI conditions. The small number of recorded errors was expected, as the participants were able to select anywhere on the object and drag it to reposition it. Because the area for selection was so large, little chance was left for an incorrect selection that resulted in an error. For the resizing task, few errors were recorded for the MR and HOTAS conditions; however, a significant number were recorded for the touchscreen condition. The resizing task required the participants to select and drag a small rectangular icon on the corner to drag and resize. As the selection target icon was small, it explains the high rate of error recorded for the touchscreen.

A further comparison of tasks found a significant main effect of tasks on error rate [$F(2,34) = 5.05$, $p = 0.012$]. The Greenhouse-Geisser correction was applied for sphericity violation. A further pairwise comparison found a significant difference between all task combinations except between the selection and resizing tasks [$Z = 0.74$, $p = 0.625$].

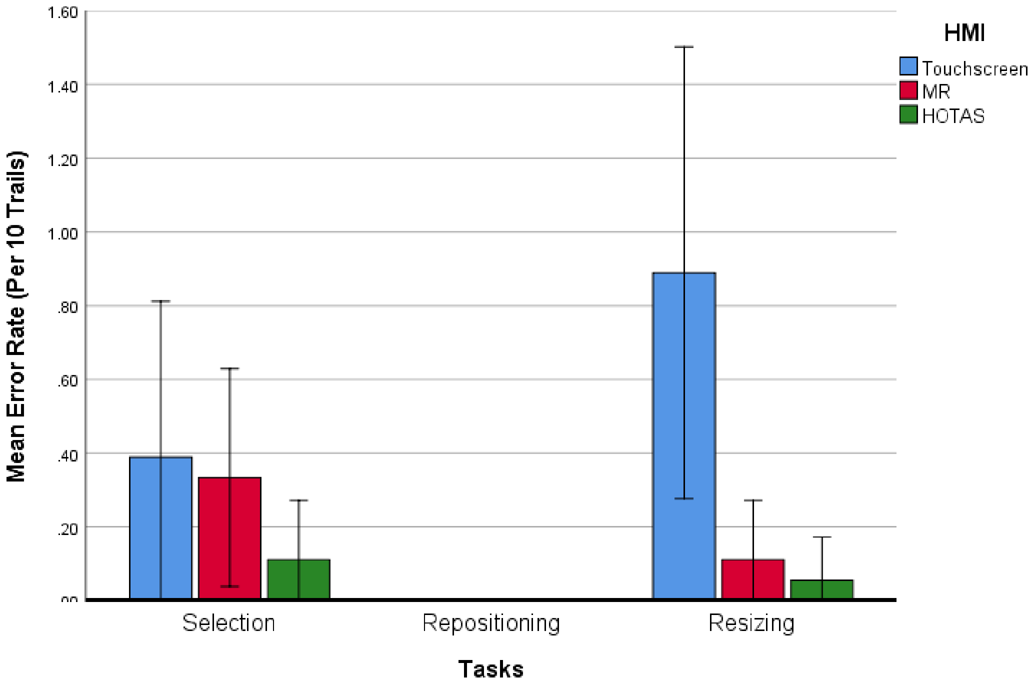

Figure 4.14: Mean error rate for tasks and HMI conditions

Vertical error bars indicate the standard deviation from the average values

4.3.4.3 Usability

Following the completion of trials for each HMI condition, participants filled out an SUS questionnaire. Brooke (1996), the author of SUS, states that any conditions with a score above the universal average score of 68 are deemed acceptable. The score of 68 is an industry benchmark based on the 50th percentile score of over 500 studies that tested systems and applications using the SUS questionnaire (Sauro, 2011). Additional studies confirm the average score of 68 with a standard deviation of 12.5 (Sauro and Lewis, 2016). Furthermore, Tullis and Albert (2008) state that an average score of less than 60 indicates relatively poor usability, while an average score of more than 80 indicates relatively good usability. Figure 4.15 illustrates the SUS scores and corresponding "acceptability ranges" and "grading scale".

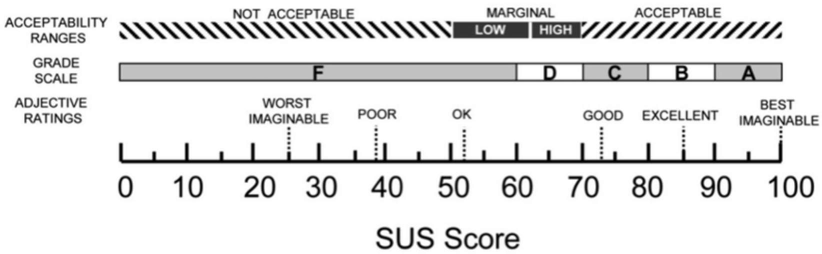

Figure 4.15: A comparison of the acceptability scores, grading scales and adjective ratings in relation to the universal average SUS score (Brooke, 1996)

As illustrated in Figure 4.16, all conditions scored above the universal average of 68. The green dotted line indicates an "acceptable" system. The touchscreen condition received the highest SUS score (M = 92 ± 2.15), rated "excellent". The MR condition scored second highest (M = 79 ± 3.94), rated "good"; and HOTAS scored the lowest (M = 71 ± 3.65), rated "OK".

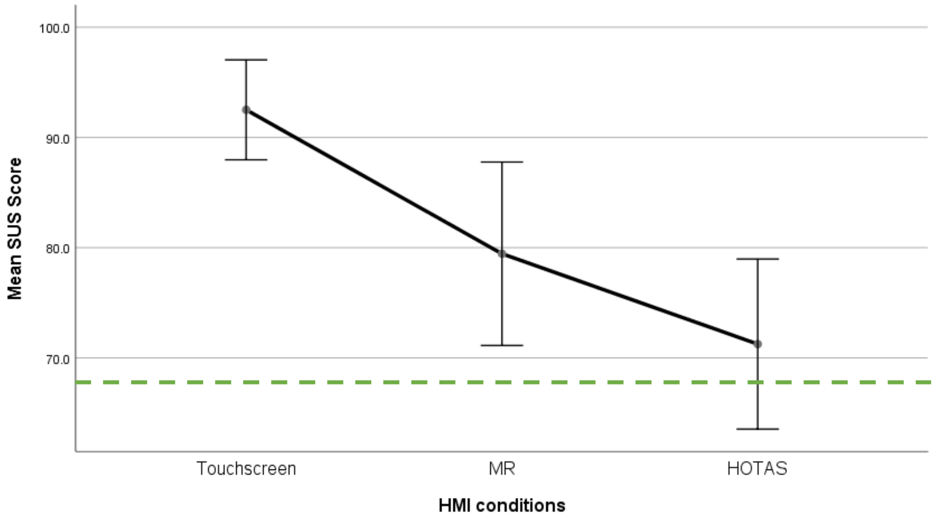

Figure 4.16: Usability score per HMI condition.

Error bars indicate the standard deviation from the average values.

Green bar indicates an "acceptable" system based on the universal score of 68.

A one-way, repeated-measures ANOVA was used to analyse the SUS scores. The assumption of sphericity was met. The results show a significant main effect of HMI condition on usability [$F(2,34) = 13.96$, $p < 0.001$]. Subsequent pairwise comparisons reveal a statistically significant difference between all technology combinations ($p < 0.001$) except between the MR and HOTAS conditions, for which no statistical significance was found ($p = 0.105$).

4.3.4.4 Workload

Participants completed a NASA-TLX workload questionnaire after each HMI condition. Workload was manipulated by the HMI condition used. Depending on the HMI condition, the number of steps required to complete the tasks varied. For example, during the repositioning task, participants in the touchscreen condition could move the object with a single gesture. However, for the HOTAS condition, the same task required multiple gestures, as the cursor moved on an X and Y plot. Therefore, the HMI conditions used manipulated the mental and physical demand on the participant.

The NASA-TLX provides a single overall score for workload and a score for all the six sub-scales, including mental demand, physical demand, temporal demand, performance, effort and frustration. While a score can be given for each sub-scale, many researchers report only a single "raw TLX" score, which is an overall score between 0 and 100 (Hart, 2006). A single overall score has been shown to improve the experimental validity of the workload rating (Bustamante and Spain, 2008). Endsley (1988) states that a workload score below 50 is considered acceptable.

Figure 4.17 illustrates the raw TLX scores across each HMI condition. All three conditions received an overall workload score below 50 and thus were regarded as "acceptable". The MR condition scored the lowest ($M = 32.88 \pm 5.33$), whereas both the touchscreen ($M = 45.13 \pm 5.33$) and MR ($M = 43.19 \pm 5.29$) conditions produced a similar but higher score.

A one-way (HMI) repeated-measures ANOVA was used to analyse the effect of HMI conditions on workload. The assumption of sphericity was met. The results show a significant main effect of HMI condition on workload [$F(2,30) = 3.54$, $p = 0.04$]. However, subsequent pairwise comparisons found a statistically significant difference only between the MR and HOTAS conditions [$Z = 10.31$, $p = 0.64$].

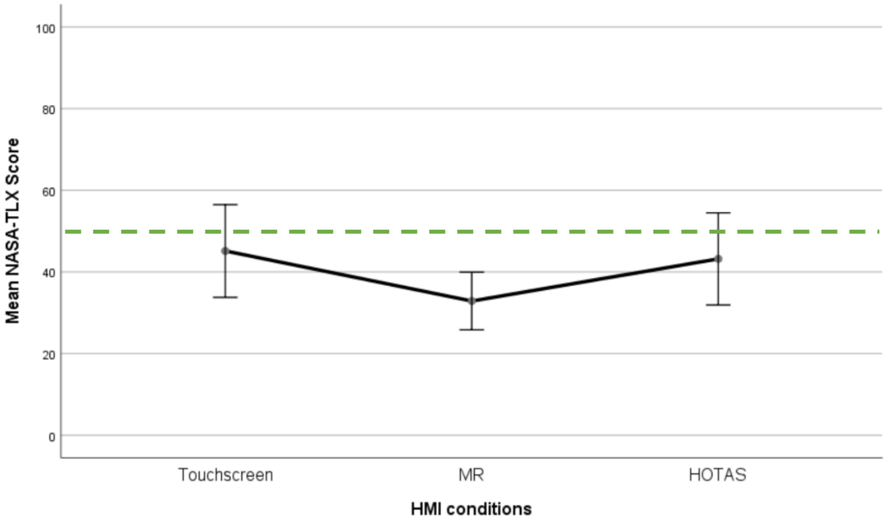

Figure 4.17: Workload score per HMI Condition

Vertical error bars indicate the standard deviation from the average values.

Green bar indicates an "acceptable" workload score.

4.3.4.5 Difficulty

Difficulty was rated upon completion of all three HMI conditions. The following question was asked: "How difficult did you find it to complete the interaction tasks for this HMI device? Rate between 1 (very easy) and 10 (very difficult)". The rating was given audibly by the participants. As shown in Figure 4.18, the touchscreen produced the lowest mean rating of difficulty (M = 2.50 ± 0.46), the MR condition producing a marginally higher rating (M = 3.67 ± 0.52), and the HOTAS condition producing the highest rating (M = 5.61 ± 0.53). A Friedman test found a significant difference in difficulty depending on which HMI condition was used [$\chi^2(18) = 18.98$, $p < 0.001$]. *Post-hoc* analyses with the Wilcoxon signed-rank test was conducted with the Bonferroni correction applied ($p < 0.05$ becoming $p < 0.017$). The result demonstrated a statistically significant difference across all three HMI combinations ($p < 0.01$).

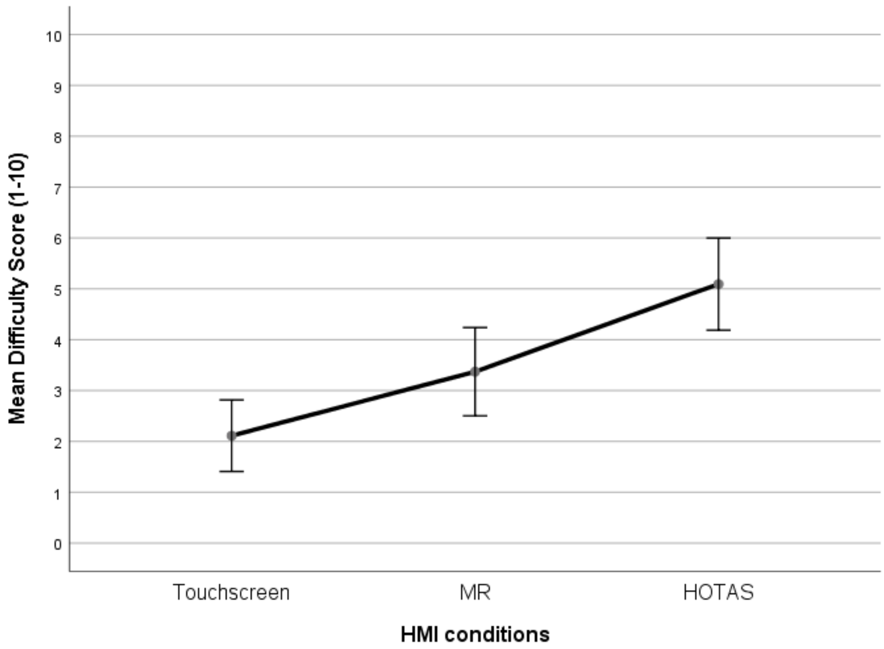

Figure 4.18: Mean difficulty score HMI conditions

Vertical error bars indicate the standard deviation from the average values.

4.3.4.6 Preference

Upon completion of all three HMI conditions, participants provided a subjective preference rating of HMI devices. The following question was asked: "Which of the three technologies in the experiment did you most prefer for interaction tasks, including repositioning, resizing and zooming? Rate the three technologies between 1 and 3, 1 being most preferred and 3 being least preferred". The data of two participants was excluded because they incorrectly filled in the form.

As shown in Figure 4.19, 15 of 18 participants ranked the touchscreen as their first choice, thus indicating a clear preference. Thirteen of 18 participants rated the MR system as their second choice, again indicating a clear preference. Finally, 13 of 18 rated the HOTAS as their least-preferred HMI device. The results contrast with the preference rankings of Experiment 1a in that the three HMI conditions have a clear ranking order. In Part 1a, no clear preference for first emerged, with 9 out of 18 participants preferring the MR system and 8 out of 18 preferring the touchscreen. Both Experiments, 1a and 1b, found the HOTAS condition to be the least-preferred HMI condition by a clear majority.

A Friedman test reported a significant main effect of HMI conditions on preference ratings for the interaction tasks [$\chi^2(2) = 22.40$, $p < 0.001$]. *Post-hoc* analyses with the Wilcoxon signed-rank test was conducted with the Bonferroni correction applied ($p < 0.05$ becoming $p < 0.017$). Subsequent Wilcoxon signed-rank tests found a significant difference between all three condition combinations ($p < 0.017$).

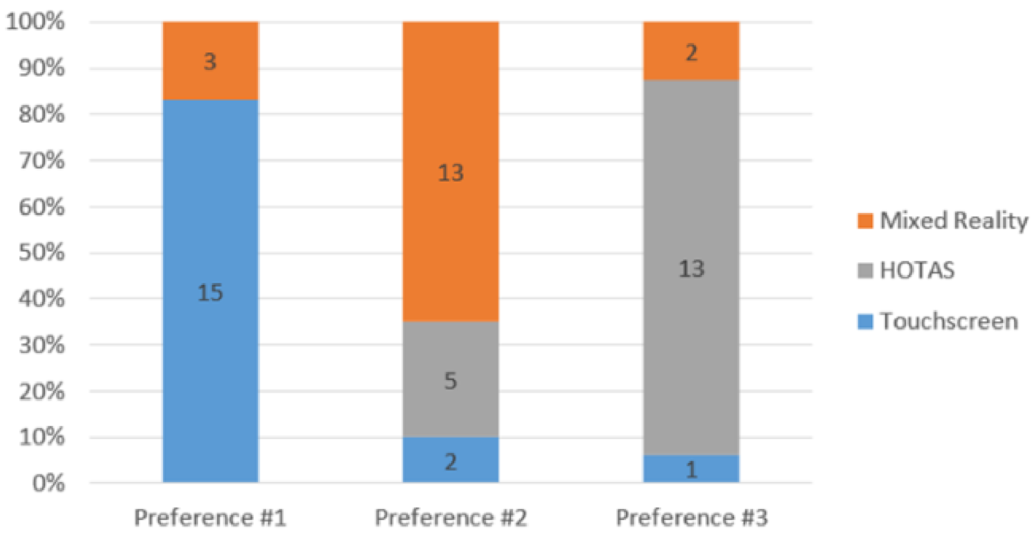

Figure 4.19: Preference ratings for HMI conditions

4.3.5 Discussion

The experiment had three aims. The first was to determine whether the HMI devices used had an effect on performance (measured by completion time and error rate) when completing three interaction tasks (selecting, rescaling, repositioning). A statistically significant difference was found between the HMI devices. The touchscreen produced the slowest completion time and the HOTAS produced the fastest completion time. As with Experiment 1a, few errors occurred overall, but a statistically significant difference was reported. Therefore, the null hypothesis that the HMI device used had no effect on performance is rejected.

The second aim was to determine whether the individual interaction tasks had an effect on performance (measured by completion time and error rate) when using the HMI devices. For the three tasks, the same trend as above was found. The touchscreen yielded the fastest completion time across the three tasks, followed by the MR, and then the HOTAS condition. One outlier was found: The resizing task during the HOTAS condition required a much longer completion time than it did during the other HMI conditions. Few errors were reported, with the majority occurring during the resizing task on the touchscreen condition. No errors were recorded for the repositioning task across the three HMI conditions. Based on these results, the null hypothesis that the interaction tasks would have no effect on performance is rejected.

The third aim was to determine whether the HMI devices used had an effect on various subjectively measured parameters, including workload, usability, and difficulty. As with previous results, the touchscreen produced a significantly higher usability rating and a significantly lower difficulty rating than the other two devices, with the MR and HOTAS performing second and third. All three HMI conditions were rated as “acceptable” for usability, as described by Brooke (1996). Similarly, all three HMI conditions had a low difficulty score.

For the workload rating, the MR had the lowest score, with the touchscreen and HOTAS conditions scoring close to each other. However, all scored below 50 and were deemed "acceptable", as described by Endsley (1988). The null hypothesis that the HMI device used had no effect on the subjective ratings is rejected.

The final aim was to determine if users would prefer some HMI devices over others when completing a range of interaction tasks. Participants indicated a statistically significant clear first preference for the touchscreen condition and a clear second preference for MR, with HOTAS being the least preferred. The null hypothesis that participants would prefer all HMI conditions equally is therefore rejected.

4.4 Conclusion

This chapter reports on two experiments, each assessing the same three HMI devices, but using different tasks and measures. Experiment 1a assessed performance and a range of subjective metrics when participants performed simple object selection gestures across three HMI conditions: A touchscreen, a HOTAS controller, and an MR system. A range of interface design options were tested, including the shape, size and spacing of objects. Experiment 1b assessed performance and a range of subjective metrics when participants performed more complex interaction gestures across the same three HMI conditions.

The two experiments were performed with the overall aim of assessing the use of virtual displays to replace physical fixed-function, fixed-location displays. Virtual displays were delivered using an MR system which utilised head-tracking for head-based input and a video see-through ("pass-through") HMD. The MR system was assessed against two input types that used a physical display: a touchscreen for finger-press input and a HOTAS controller for mouse- and button-based input. Overall, the touchscreen produced the highest performance in terms of reaction times and completion times but not for error rate. The HOTAS produced the fewest errors, but it also produced the slowest reaction times and completion times by a significant margin. The MR condition scored between the other two conditions for all measures. The same result was observed across all selection and interaction tasks.

For Experiment 1a, a framework for assessing the suitability of a range of interface design criteria for various HMI devices was presented. The experiment measured a range of objective and subjective measures in order to determine the suitability of devices and design choices to specific tasks. Suitability may be defined by performance alone (e.g., fastest reaction time and lowest error rate), but it may also be defined relative to human-centred metrics, such as by providing the highest level of usability and putting the least physical and cognitive demand and on users. When selecting the design parameters to include within the experiment, a wide range of academic and industrial studies, standards and guidelines were analysed. Based on the above, a range of object design criteria was selected, including the shape, size and spacing of objects.

The experiment found that, overall, the selected range of object design criteria had a significant effect on performance as measured by both reaction times and error rate. During the selection tasks, a range of subjective human-centred measurements were collected, including exertion, discomfort, difficulty and preference. No significant effect on exertion was found, and very little discomfort was found. The only significant result was some minor discomfort felt while using the touchscreen. This latter result was expected because the touchscreen requires users to extend their arms for a long period of time. During the interaction tasks, a further range of subjective human-centred measures were collected, including workload, usability, difficulty and preference. All three HMI conditions were scored at low levels of workload, usability and difficulty, with all scores considered acceptable as per the authors of the rating methods used.

Experiment 1a found no significant differences in preference ranking for first and second preference. However, during Experiment 1b, a clear preference was seen across all rankings. During selection tasks, participants did not exhibit a significant preference for either the touchscreen or MR devices. However, for interaction tasks, the participants' clear first preference was the touchscreen. Across both experiments, participants least preferred the HOTAS by a significant margin. The result may be explained by the fact that the HOTAS controller was not operated by a single, simple gesture of pointing or looking but instead required precise and continuous input via a mouse. This resulted in significantly slower reaction times and completion times than the other conditions.

The subjective ratings for both Experiment 1a and 1b indicate that, apart from the effect on performance, the three HMI devices used had little impact on the user, with all rankings producing very similar results and low levels of workload, difficulty and usability issues reported across all conditions. The main difference found was in participants' preference, with the MR device being the most preferred for selection tasks by an insignificant margin, whereas the touchscreen was most preferred for interaction tasks.

In the academic and industrial literature, various standards suggest a number of conditions under which head-based input should not be used. For instance, it is not suitable for precise input or under vibration (Department of Defense, 2012). However, the same conditions may equally affect touchscreen performance, and this is still a currently used technology within defence platforms (Tauson, 2012). One example is the Military Standard 1472G (Department of Defense, 2012), which states that head-based controls shall not be used if the task requires frequent, precise head movements. However, the MR system provided reaction times and completion times which were better than those of the HOTAS controllers, and error rates which were better than the touchscreen condition. Both the touchscreen and HOTAS devices are used within current operational-vehicle platforms. This experiment has therefore shown that an MR system in which virtual displays replace physical displays can provide performance levels comparable to existing interaction and display systems that are currently used within cockpits (including touchscreen- and joystick-based controllers), and could be considered a viable compromise between speed and accuracy.

4.4.1 Future Work

While there is a wide variety of standards for mature and widespread technologies such as the touchscreen, few to none are available for emerging technologies such as MR systems. These often contain varied and even conflicting recommendations and emphasise different design elements, such as the shape, colour, size and spacing of objects. Therefore, a range of criteria was chosen based upon similarities between a selection of standards, and an assessment framework was used to define the optimal design criteria.

As the experiment assessed a range of object variables including shape, size and spacing, the number of variations of each was limited so as not have an overly complex and long experimental procedure. Future work could build upon this study by introducing a further range of sizes with greater resolution, such as from 5mm–25mm in 2mm increments. By not assessing a further range of nested conditions, such as spacing, additional conditions could be assessed without the experiment becoming very long and complex.

An additional experiment could include assessment of the effect of user demographics on performance. The effect of a user’s experience with the HMI devices used may also be further assessed. For devices such as a HOTAS controller, combat aircraft pilots will have substantial experience with the device and will have used it regularly for a minimum of several years both during training and operationally. The thumb-controlled cursor used on the HOTAS is not common across other consumer-input devices and non-pilot users may have far less experience with this input scheme. A further experiment could be undertaken in which a between-subject experiment assesses HOTAS input on virtual displays within an HMD, one group of pilots and one group of non-pilots with no thumb-controller input experience.

During the experiments, a physical, modular, cockpit-like testbed was developed to produce a working environment representative of a manned cockpit wherein the arms, head and body movements are restricted to the physical dimensions of the seated, enclosed vehicle. Using a modular physical testbed made it possible to integrate and place the HMI devices in the locations they ordinarily occupy within a cockpit.

However, due to budget and time limits, many additional factors were not included and tested within the testbed, including the wearing of gloves, helmets and seatbelts. In addition, there are multiple environmental factors such as movement (vibration, G-force, jolts) and lighting (such as bright sunlight or low-lighting during night-time) that might affect the results. Future work could aim to include the factors detailed above to more closely represent the environment of specific use cases, such as the cockpits of fast jets and manned, land-based combat vehicles.

Finally, the experiment collected data of the participants' age and sex, but this was not analysed in reference to the results found within this study. Investigation of the effect of age could be a consideration for further experimentation by other researchers. For comparison, the participants within the two studies presented in this chapter had a mean age of 26.0 and a standard deviation of 8.6. This differs from the average age of all Officers within the Royal Air Force (RAF) which was 37, while the average age of all other ranks was 30 (House of Commons, 2019). Data for the average age of operational RAF pilots is not readily published, however, within the RAF, pilots are of officer rank. The entry requirements for pilot training within the UK RAF is 17.5 to 25 years (Royal Air Force, 2019). RAF pilots serve an initial commission of 12 years, therefore dictating the maximum age of a RAF pilot to ordinarily be 40 years old.

Chapter 5: Tangibility and Presence

5.1 Introduction

Many potential applications of MR for C2 have employed the use of augmented virtual content in the form of fixed-location interactive command tables. However, many of these take the form of “floating” scenarios or vistas in 3D space without any physical barriers that serve to intuitively inform users about—or limit them to—the area of interactivity. The absence of such barriers also means that users are not provided with any cues or affordances related to the external world or to the actions they have undertaken when interacting with the command table. One method of overcoming the lack of affordances and cues to the physical environment is to “anchor” virtual objects to physical objects in the external world. Employing the affordances of the local physical environment can provide interaction cues and haptic feedback to the user.

In addition to passively utilising the affordances of the surrounding environment, a system may also incorporate the inclusion of physical objects that are related to the virtual content presented to the user in the MR environment. When these physical objects have a relationship to virtual content and are used during interaction, they can be referred to as *tangible interface objects*. Tangible interfaces have emerged as a novel interface type over the last two decades; however, the field is still in its infancy and still requires extensive research to fully understand the implications of using tangible interfaces to bridge the physical and virtual (Orit and Eva, 2009). Tangible interfaces have further been shown to affect human performance, usability and cognitive states such as presence, as detailed in Section 2.3.

5.2 Aims

The aims of the experiment were rated as either primary (P) or secondary (S). The primary aim of this experiment was to understand the effect of tangibility within an MR environment on perceived levels of presence (as reviewed in Section 2.3). A tangible interface is achieved by using co-located physical objects that closely match virtual objects in terms of shape, dimensions, position and movement effects. The secondary aim was to assess the effect of tangibility on further general HF metrics known to affect performance, including workload and usability.

The experiment aimed to address the following:

i) (P) Determine whether the inclusion or absence of tangible interface elements within an MR environment (the inclusion of physical objects that are co-located with virtual objects) have an effect on participants' rated levels of presence. The null hypothesis is that tangible interface objects have no effect on presence.

ii) (S) Determine whether the inclusion or absence of tangible interface elements within an MR environment (the inclusion of physical objects that are co-located with virtual objects) has an effect on participants' rated levels of workload and usability. The null hypothesis is that tangible interface objects have no effect on workload usability.

5.3 Method

5.3.1 Participants

A total of 22 undergraduate and postgraduate students were selected on an opportunistic-availability basis from within the School of Engineering at the University of Birmingham. The participants consisted of 18 males and 4 females, with a mean age of 27.2 years and a standard deviation of 7.4 years. All participants had prior experience with VR HMDs and had no prior experience with hand-tracking based system interaction.

5.3.2 Independent Variable

The experiment assessed tangibility across two conditions: one that included physical objects that were co-located with virtual objects (MR), and one in which no physical objects were present (AR). Therefore, the independent variable within this experiment was the “reality” concept used. The conditions included the following:

i. AR: No tangible objects were included in the experiment and thus no affordances were provided by physical objects that were co-located with the virtual content.

ii. MR: Tangible objects were included in the experiment. Physical objects of the same shape, dimensions and position as virtual objects were included. The tangible objects provided tightly-coupled affordances and tactile cues for interaction.

5.3.3 Apparatus

The system used within this experiment was a modified version of the experimental testbed described by BAE Systems (2015), which provided a reconfigurable, tabletop-based MR system. The testbed was developed for the purpose of experimenting with visualisation methods and interaction techniques for displaying large amounts of data to the user in one central location. The system allowed the user to visualise and interact with objects and data in real time, in 3D space and viewed stereoscopically. The system used an OptiTrack motion-capture tracking system and an Oculus Rift DK2 HMD. The motion-capture system provided precise tracking for the HMD, hand-tracked gloves, and tracked physical objects. This allowed the user to navigate, interact with and manipulate the virtual content using tablet-computer-like gestures. Figure 5.1 illustrates the table that was used within the experiment; the left panel shows the user wearing the apparatus and interacting with the system while the right panel depicts the user's view as seen through the HMD. For the MR condition, the augmented content is precisely overlaid onto a physical circular table of the same shape, dimensions and position. For the AR condition, no table was used. Participants were able to see real-world objects, such as the hands and external world, through the HMD provided by the pass-through camera.

Figure 5.1: External view of the table (left) and HMD view of the table with augmented content (right)

The system consisted of several main elements, the first of which was the superimposed terrain, with which participants interacted with by placing their hands flat on the virtual terrain and performing gestures, including repositioning, rotating and rescaling. As illustrated in Figure 5.2, movement (repositioning) was achieved by placing one hand flat on the terrain and moving the hand in the desired direction to perform a “dragging” motion. Rotation was achieved by placing one hand on the edge of the table and moving around the edge. The virtual content rotated to match the location of the hand along the perimeter of the table. Rescaling was achieved by placing both hands flat on the table next to each other and moving them toward or away from each other to “zoom” in or out.

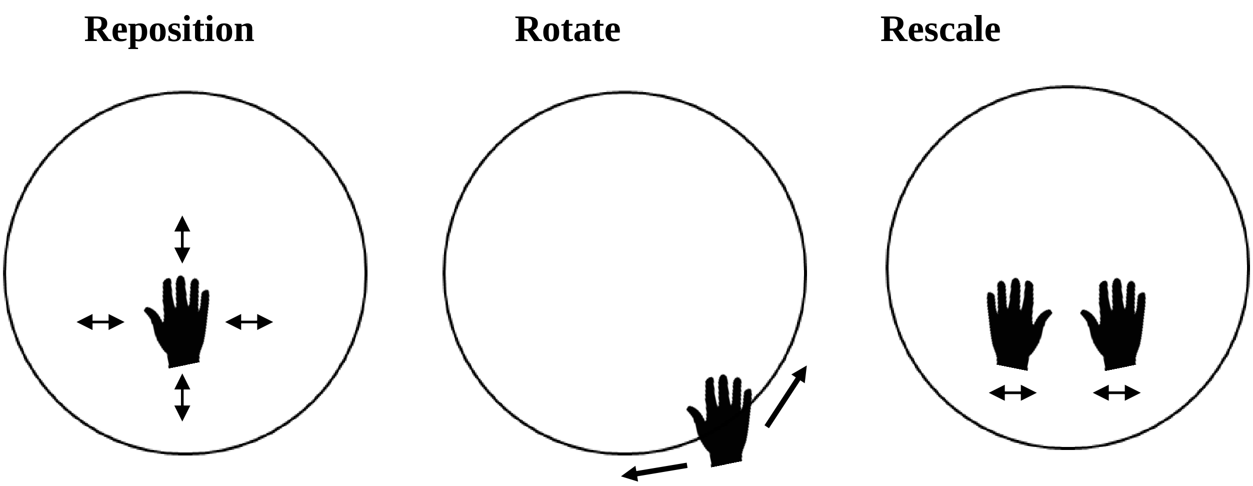

Figure 5.2: Repositioning gesture (left), rotating gesture (centre) and rescaling gesture (right)

The second element was the virtual display panels. Participants interacted with them in the same way as with the table interface, including repositioning, rescaling and selecting. As the virtual panels were positioned freely within 3D space, there was no element of haptic feedback or co-location of any physical object. Figure 5.3 illustrates these virtual display panels. The panel graphics represented mission objectives and included related generic text and image information. The user was not required to read the text, as the displays were used as a placeholder to represent virtual display panels. The panels illustrated a mission in which a small unmanned aerial vehicle (sUAV) based 3D mapping task (the process of which is detailed in Chapter 3) has been completed, and the video (Figure 5.3, top left panel) was used to produce a 3D model output of that map (Figure 5.3, top middle panel). The 3D model, as represented within the virtual panel, was also embedded within the virtual terrain.

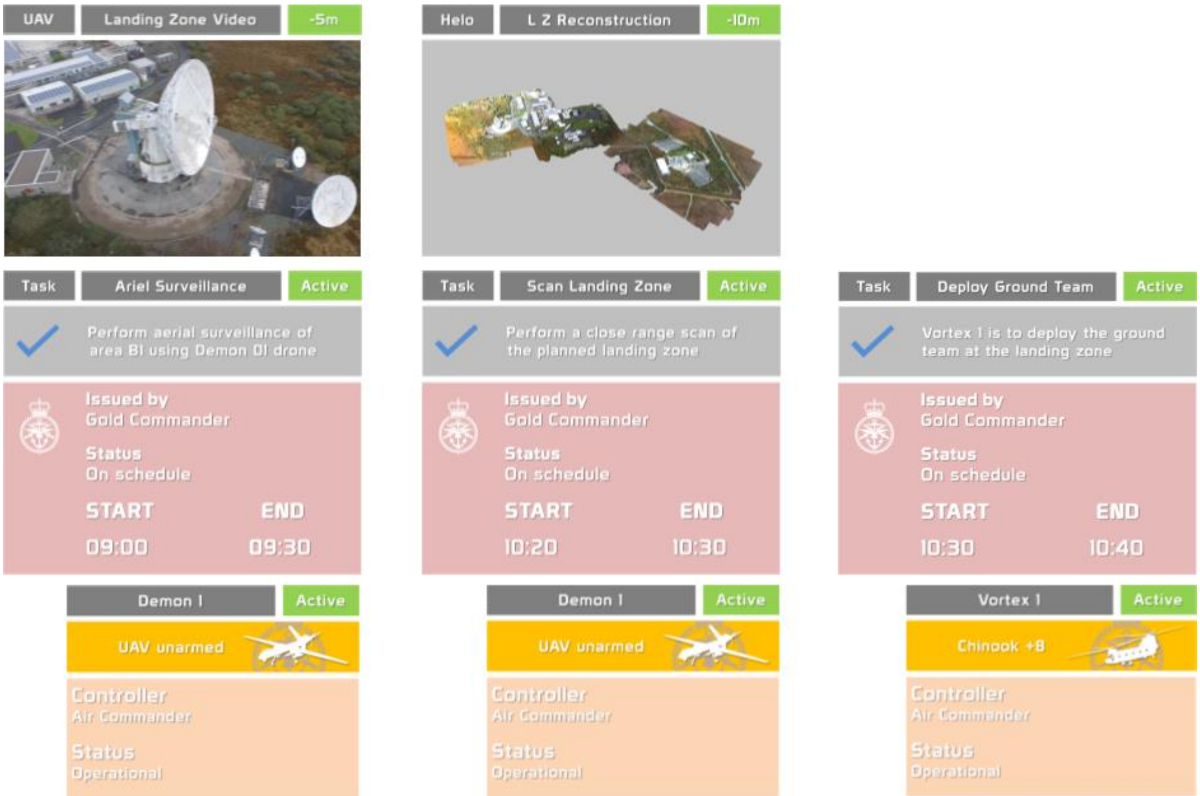

Figure 5.3: Virtual-display panels with SUAV video (top left) and 3D reconstruction (top centre)

As demonstrated in Figure 5.4, the gestures used to interact with the 2D virtual display were similar to those used to interact with the virtual terrain, including selecting, repositioning and rescaling. Repositioning the displays on a 2D plane (left, right, up and down) was achieved by placing one hand over the panel and moving the hand in the desired direction to perform a "dragging" motion. Rescaling was achieved by placing both hands next to each other in front of the panel and moving them toward or away from each other to "zoom" in or out. Selection was achieved by intersecting the hand with the panel to select it and then withdrawing the hand to deselect it.

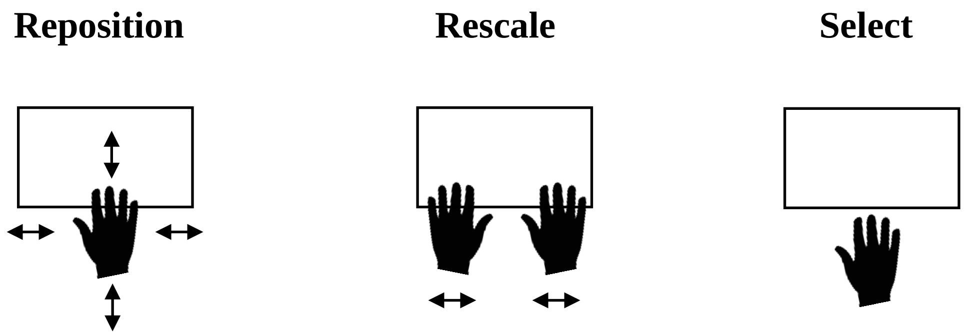

Figure 5.4: Repositioning gesture (left), rescaling gesture (centre) and selection gesture (right)

The third element was a feature that enables the user to “draw” in 3D space and was achieved by two means: one by using a tracked physical “puck” object (during the MR condition), and another using the hand alone (during the AR condition). During the AR condition, the participants selected one of three virtual buttons (“green”, “red” and “off”) from a virtual display panel by touching the virtual button with the hand. The button that had been selected was highlighted to provide visual feedback regarding the active selection. The hand drawing process was activated when the user touched the virtual terrain and began to draw in 3D. Drawing was then deactivated when the user selected the “off” button on the virtual display panel. The AR condition provided no haptic feedback or interaction with physical objects.

During the AR condition, participants used their hands only to draw in 3D space, with no associated physical objects. During the MR condition, participants used a physical object to draw in 3D space, as illustrated in Figure 5.5. The participants selected the colour to draw with by manipulating a physical object, in this case a 3D-printed puck shaped object of the same shape, size and location as a virtual puck object. Participants selected the drawing option by rotating the physical puck, which rotated the co-located virtual object, until a directional arrow at the base of the puck pointed at the option required (a letter *G* for green, *R* for red, and *M* for “movement” to turn off and move the puck without drawing). When the participant rotated the physical puck and pointed the directional arrow at the required letter, the option was selected and displayed above the puck to give visual feedback regarding the active selection, as shown in Figure 5.5 (bottom image). The puck-shaped object was tracked by the same motion-capture system as the head and hands, but beyond the tracking of the object, the object was inanimate. The use of a tangible user interface object was included because it provided the ability to deliver credible haptic feedback to the user when drawing in 3D space and selecting a colour.

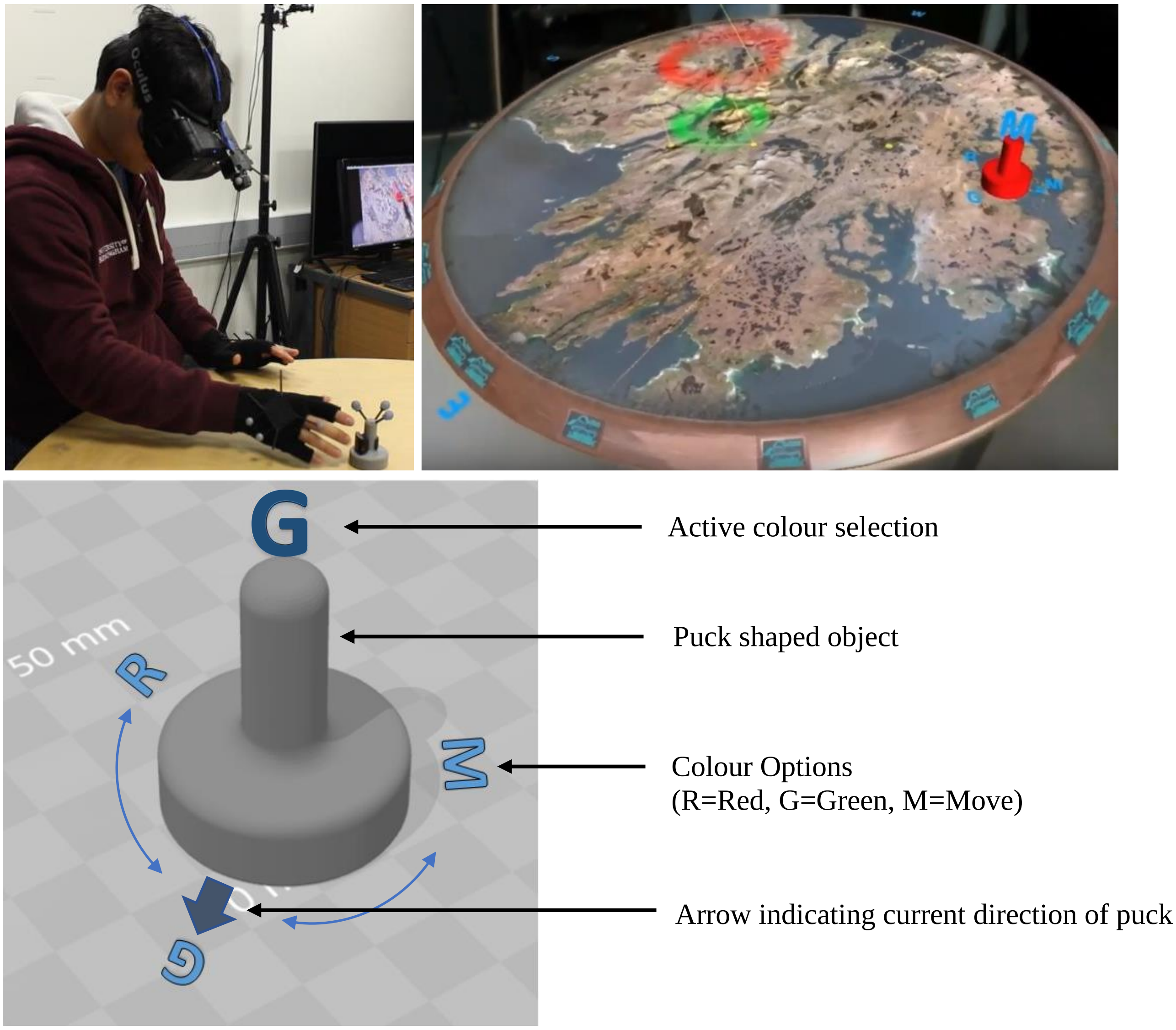

Figure 5.5: Participant interacting with the physical puck (top left), participants view within the HMD illustrating the puck and colours drawn in 3D space (top right), illustration of the puck features (bottom)

A puck-shaped object was selected for several reasons. Knight and Stone (2014) recommend the use of a puck in a paper in which they evaluate the use of tangible user interfaces on interactive tabletop systems in particular. In addition to providing affordances and haptic feedback, Knight and Stone claim that a major advantage of the use of a puck is that dynamically changing virtual controls can be assigned to it, which can be changed depending on the task, context and situation. Across two studies which assess puck-based interaction, Paelke et al. (2012) and Zuckerman and Gal-Oz (2013) report that, although participants rated puck-based tangible interfaces as inferior in usability compared to traditional graphical user interfaces (GUI), puck-based tangible interfaces were still rated as a highly effective interaction method across both studies. In addition, Zuckerman and Gal-Oz claim that the properties of a puck that were identified to be preferred by users included include the physical interaction, rich feedback, and high levels of realism.

The table used in the present experiment, during the MR condition only, provided an additional intuitive haptic feedback response, together with limitations to hand movement when interacting with the virtual content. The virtual content, in the form of a circular overlay, was precisely calibrated in a 1:1 position and scale with the physical table. The physical table allowed the participants to lay their hands flat on the table or touch the edges of the table, and the system precisely matched the boundaries of the virtual content. In addition, by utilising a pass-through camera to allow the user to see the outside world, the experiment aimed to eliminate the perceptual barriers of the user feeling removed from the environment, thus allowing a greater sense of “being there”.

5.3.4 Procedure

Prior to beginning the experience, participants read an experiment information sheet (Appendix J), completed a consent form (Appendix C) and read a health and safety document detailing exclusion criteria (Appendix B).

The experiment employed a between-subject design that consisted of two conditions (AR and MR). Participants were assigned to two groups: one group completing the AR condition without tangible interface objects, the other group completing the MR condition with tangible interface objects present. Participants were randomly assigned to the groups, with 11 participants per group. A procedure was completed by both groups which consisted of four interaction tasks. The tasks used were designed to ensure that each interactive element of the system was interacted with. The experiment lasted approximately 20 minutes, with a short familiarisation period prior to beginning the procedure.

The participants completed similar tasks for both conditions, which differed only in the inclusion or exclusion of tangible objects within the environment. The AR condition used no physical objects within the environment, and the virtual content floated freely in 3D space. No haptic feedback or affordances were provided, and the condition required users to position their hands precisely over the virtual content by vision alone. The MR condition included co-located physical objects of the same shape, dimensions and position as the virtual content. A physical table of the same dimensions as the virtual table provided affordances of the environment as well as giving haptic feedback cues to the hands to indicate the limit of the area of interactivity.

A between-subject experimental design was utilised to counteract well-known demand characteristics, as described by Klatt et al. (2012), whereby participants may attempt to guess what the experiment is examining and what outcomes the investigator aims to find. In such situations, the participant may answer according to, or contrary to, these predictions and thus invalidate the results. As the AR and MR conditions vary only in the inclusion or absence of physical objects within the environment, the purpose of the experiment may have become apparent. By utilising a between-subject experimental design, the purpose is made less apparent, and thus the demand-characteristics problem is less likely to occur. Other similar studies have utilised a between-subject design to measure presence with a sample size of n=10 for each group (Usoh et al., 2000).

5.3.4.1 Tasks

The four experimental tasks were selected to ensure that all elements of the system were interacted with, including 2D objects (such as menus) that move on a 2D plane, and 3D objects (such as the map) that move in 3D space. The four tasks were as follows:

i. Task 1: Familiarisation with the 3D interface and 2D virtual displays.
ii. Task 2: Interaction with the 3D map (by performing repositioning, rotating and rescaling gestures).
iii. Task 3: Customisation of the interface (by selecting, resizing and repositioning displays).
iv. Task 4: Interaction with content within 3D space.

The tasks were based on a representative mission scenario in which participants first inspected a 3D map to locate and identify an object. After this, the participants customised a 2D display to show imagery taken from an aerial vehicle. Finally, the user marked, or “drew” on the map with colours to indicate safe (green) and dangerous (red) areas. The participants had to interrogate the terrain to locate the embedded 3D model, as illustrated in Figure 5.4 (top left panel). Each of the tasks and sub-tasks within this scenario were individually identified and categorised into one of four single repeatable gestures, consisting of (i) selecting, (ii) repositioning, (iii) rotating and (iv) rescaling. List 1 details the procedure, which consists of the four tasks and multiple sub-tasks.

List 1: Task and sub-task breakdown by fundamental input type.

Task 1 System Interface Familiarisation

1.1 **Rescale:** Zoom in and out on the map.

1.2 **Reposition:** Move the map location on the system.

1.3 **Rotate:** Rotate the map on the system.

1.4 **Rescale:** Resize the virtual displays by increasing and decreasing the scale.

1.5 **Reposition:** Move the virtual displays up, down, left and right.

Task 2 Interact with 3D Map

2.1 **Reposition:** Move the map location to locate the area.

2.2 **Rescale:** Zoom in and out of the map to locate the area.

(Repeat steps 2.1 and 2.2 until the SUAV 3D scan model is located and placed in the centre of view).

Task 3 Customise Interface by Selecting, Resizing and Repositioning Displays

3.1 **Selection:** Select the virtual display to interact with.

3.2 **Rescale:** Zoom in on the display to the defined size.

3.3 **Reposition:** Move the display to a defined area.

(Repeat steps 3.1 to 3.3 until the virtual display with the SUAV video is located above the 3D scan model, as can be seen in Figure 5.7).

3.4 **Rescale:** Zoom in on the display until it reaches the desired size.

3.5 **Reposition:** Move the display to a defined area.

(Repeat steps 3.4 and 3.5 until the virtual display with the 3D scan model is located above the 3D scan model).

Task 4 Draw Zones in 3D Space with a Hand/Puck

Draw low-altitude safe zone:

4.1 **Selection:** Select the “green” colour option to draw in green.

4.2 **Reposition:** Move the puck/hand around at ground level to draw a green circle at the table level.

4.3 **Selection:** Select the “off” colour option to stop drawing.

Draw high-altitude danger zone:

4.4 **Selection:** Select the “red” colour option to draw in red.

4.5 **Reposition:** Move the puck/hand around at a higher level to draw a red circle in 3D space.

4.6 **Selection:** Select the “off” option to stop drawing.

Figure 5.6 illustrates the default layout of the system before the experiment, and Figure 5.7 illustrates the system after the experiment. Task 3 required the participants to interact with the virtual displays by repositioning and resizing them until the top left panel had been enlarged to fill the whole area. Task 4 required the participants to draw a circle in 3D space by drawing a green circle at a low position and a red circle at a higher position. Both completed tasks can be seen below.

Figure 5.6: Pre-experiment interface layout

Figure 5.7: Post-experiment interface layout

5.3.5 Dependent Variables

Three subjective measures were included in the experiment in the form of *post-hoc* questionnaires. Following the procedure, the participants completed all three subjective measures. The three dependent variables in this experiment are as follows:

i. Presence: Measured using the iGroup Presence Questionnaire (IPQ) (iGroup, 2006).
ii. Usability: Measured using the System Usability Scale questionnaire (SUS, not to be confused with the Slater, Usoh and Steed presence questionnaire) (Brooke, 1996),
iii. Workload: Measured using the NASA-TLX questionnaire (Hart and Staveland, 1988).

The method chosen to measure presence is the subjective IPQ questionnaire (iGroup, 2006). Based on the model of presence by Schubert et al. (2001), the IPQ model of presence comprises three independent sub-elements and a single variable metric of general presence, as follows:

1) Spatial presence (SP): the sense of being physically present in the virtual environment.
2) Involvement (INV): measuring the attention devoted to the VE and the involvement experienced.
3) Experienced realism (REAL): measuring the subjective experience of realism in the virtual environment.
4) General presence (G): measuring the subjective sensation of “being there”.

Although the reliability and validity of a single metric to represent presence is unknown, iGroup has stated that the general presence element has high loadings on all three sub-elements, and especially on spatial presence.

The IPQ consists of 14 questions that were grouped into three sub-elements and a single question attributed to the single general-presence element. The questions were accompanied by anchored responses between two extremes (for example, between “fully disagree” to “fully agree”). The questions contained seven scale points ranging from -3 to +3, with a neutral response category of 0. Appendix L details the questions used, the corresponding categories, the anchors used and the sources that were used to generate the IPQ questions. The full list of questions and the format of the questionnaire can be found in Appendix M, and the calculations and process used to compute the results are detailed by iGroup (2006).

Further research by Regenbrecht and Schubert (2002) has validated the use of the IPQ and the applicability of the presence model presented by Schubert et al. (2001) in AR environments, while Schaik et al. (2004) have demonstrated the use of the model in MR environments. As AR and MR environments purposefully include elements of the real world, whereas VR environments purposefully exclude the real world, several question anchor responses were inverted. Questions INV2 and INV3 measure the inclusion of real-world environments as a negative response. This is correct for VR environments, but for AR/MR environments it would be a positive response. The two previous studies referenced above (Regenbrechy and Schubert, 2002; Schaik et al., 2004) also inverted anchor responses.

The three elements within the framework presence model were based upon principal-component analysis and used questions both from a variety of academic sources and others that were composed specially for the questionnaire. The questionnaire framework was validated across two studies (N = 246 and N = 296), and it is widely cited in the literature. While the applicability of the IPQ to MR is not explicitly stated by the authors, AR and MR environments may be considered to be the same for the purpose of the present experiment, with the exception of additional affordances that are provided by the tangible interface objects within the MR condition.

5.3.6 Statistical Analysis

Mean values were presented in a standardised format that displayed the mean value followed by the standard deviation value (M = x ± SD). The mean was presented with two decimal places.

Each of the four sub-elements of presence within the presence model used were analysed separately for each of the two “reality” conditions: AR and MR. As the experiment was a between-subject study with two independent groups (AR and MR), and because the data collected was of ordinal scale, a Mann-Whitney U test was performed. The Mann-Whitney U test is a rank-based nonparametric test that can be used to determine whether there are differences between two groups of ordinal data types. It is an alternative to the independent sample t-test, which tests for significance between groups of interval level data.

To analyse the statistical significance of the difference between the means of the two “reality” conditions, an independent sample t-test was used for the workload and usability measures. Though the data collected by the questionnaires was of ordinal scale, it met all the assumptions for interval data when the multiple variables were calculated to produce a single overall mean score (using the calculations defined by the questionnaire authors). The assumptions required for the use of interval-scale analysis methods for ordinal data are detailed in Experiment 1a.

As the six sub-scales of the workload measure were also of interest, further analysis was performed to determine the individual effects. The sub-scales included mental demand, physical demand, temporal demand, own performance, effort and frustration. Each sub-scale was measured using ordinal scale questions and analysed using the Mann-Whitney U test, as above.

5.4 Results

5.4.1 Presence

Following the experimental procedure, participants completed the IPQ questionnaire (iGroup, 2006) to measure presence for both the AR and MR groups. As recommended by iGroup, the results are presented on a plot diagram, or "radar chart" which represents a "presence profile", as illustrated in Figure 5.8. The three elements of presence (spatial presence, involvement, and experienced realism) are displayed on three axes within a range of 0 (lowest) to 7 (highest). In addition, as recommended by iGroup, the general presence element is displayed as a "bow" on the left of the diagram. The bow consists of an individual curved line for each measure of general presence, with the length of the line denoting the rated value of general presence. A short line represents a low rating of general presence and a long line represents a high rating of general presence.

As all the sub-elements of presence were of ordinal level data, had the same scale points (0-7), and the axes begin at the same point, the radar chart allows for a graphical method of displaying a single observation of the multivariate sub-elements of the presence model on a two-dimensional chart. While the use of a radar chart is disputed, by using the chart recommended by the questionnaires authors, it allows direct comparison with other studies that also used the IPQ questionnaire and default recommended chart.

As shown in Table 5.1, the experiment found that the MR condition, in which tangible user-interface objects were present, resulted in a higher rated level of presence across all four elements of the IPQ compared to the AR condition in which tangible objects were absent. The experienced realism sub-element produced a very small difference between the MR and AR conditions, an element which was not expected to be strongly affected by the use of tangible interfaces as the content remained the same. However, for the spatial presence sub-element, which was expected to be strongly affected by the use of tangible interfaces, a large difference was found. The high levels of tangibility within the MR system provided affordances and tactile feedback of the real world. Thus, the MR system would be anticipated to provide a greater level of spatial understanding of the environment.

A Mann-Whitney U test was used to assess the statistical significance of the difference between the AR and MR groups for general presence and the three sub-elements. The analysis reported that the MR condition was statistically significantly higher than the AR group for general presence [$U = 13$, $p < 0.001$], spatial presence [$U = 16$, $p = 0.003$] and involvement [$U = 78$, $p < 0.001$]. However, no statistically significant difference was found between the experienced realism sub-element in the AR and MR conditions [$U = 54$, $p = 0.66$].

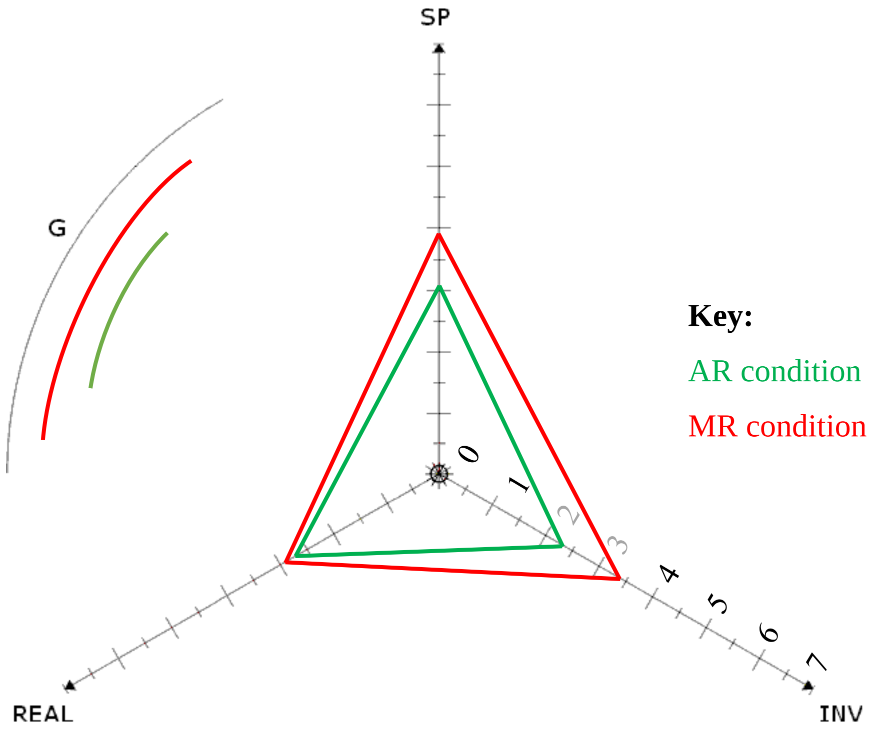

Figure 5.8: Presence profile of IPQ results for AR and MR conditions

Key: G = general presence; SP = spatial presence; INV = involvement; REAL = experienced realism

Table 5.1: Descriptive Statistics of IPQ Results

Dependent Variable	Condition	Mean Score
G	AR	2.36 ± 1.02
	MR	4.09 ± 0.83
SP	AR	3.04 ± 0.82
	MR	3.95 ± 0.69
INV	AR	2.30 ± 0.67
	MR	3.47 ± 0.62
REAL	AR	2.68 ± 0.66
	MR	2.93 ± 1.57

5.4.2 Workload

Following the procedure, participants completed a NASA-TLX (Hart and Staveland, 1988). workload questionnaire for the "reality" condition in which they were grouped. The NASA-TLX provides a single overall score for workload in addition to six sub-scales: mental demand, physical demand, temporal demand, own performance, effort and frustration. As mentioned in Section 4.2.4.4, Endsley (1988) states that a workload score of less than 50 is considered acceptable.

Overall, the AR condition scored far higher than the acceptable score of 50 (M = 68.82 ± 19.18), which indicates that the condition requires an unacceptable level of workload during use. The MR condition scored lower than 50, thereby indicating an acceptable system (M = 40.27 ± 19.00). Physical workload for the MR condition would be expected to be lower given that the tangible interface objects within the MR condition provide physical support such that participants can rest their hands and lean on the table. The MR system also provides tactile cues for the correct positioning of the hands. An independent-samples t-test was used to assess the mean workload scores between the AR and MR groups. The results of this t-test suggest that the MR condition had a statistically significantly lower rating of workload than the AR condition [$t(20) = 2.48$, $p = 0.02$].

While previous experiments undertaken within this thesis considered workload only as a single variable, this experiment assessed the differences of the individual sub-scales of the TLX between the AR and MR conditions. Figure 5.9 illustrates the scores between conditions for all six sub-scales. The results correlate with the overall TLX scores in which the MR condition produced a lower level of workload across all six sub-scales. The MR condition provided an indication of the correct point in 3D space to position the hands by providing tactile cues, by limiting the user’s hand movement, and by leaving little room for inaccurate hand placement. Therefore, it was hypothesised that the MR condition would place less mental and physical demand on the participants than the AR condition.

The “physical” sub-scale produced the largest difference between the AR ($M = 12.0 \pm 5.4$) and MR ($M = 6.1 \pm 3.6$) conditions. The results may be attributed to the inclusion of the table within the MR condition which provided physical support for the hands and ensured correct posture and hand position. In addition, the frustration metric displayed a large difference between the AR ($M = 10.5 \pm 5.8$) and MR ($M = 6.7 \pm 4.9$) conditions. Visible frustration was observed during the AR condition in which participants incorrectly placed their hands in 3D space when attempting to complete the interaction gestures. Furthermore, participants noted their inability to consistently perform the required actions, which caused them to become frustrated.

A Mann-Whitney U test of each sub-scale pairing revealed that the MR condition was statistically significantly higher than the AR condition ($p < 0.05$) across all six sub-scales, except for frustration, which exhibited no significant difference [$U = 36$, $p = 0.11$].

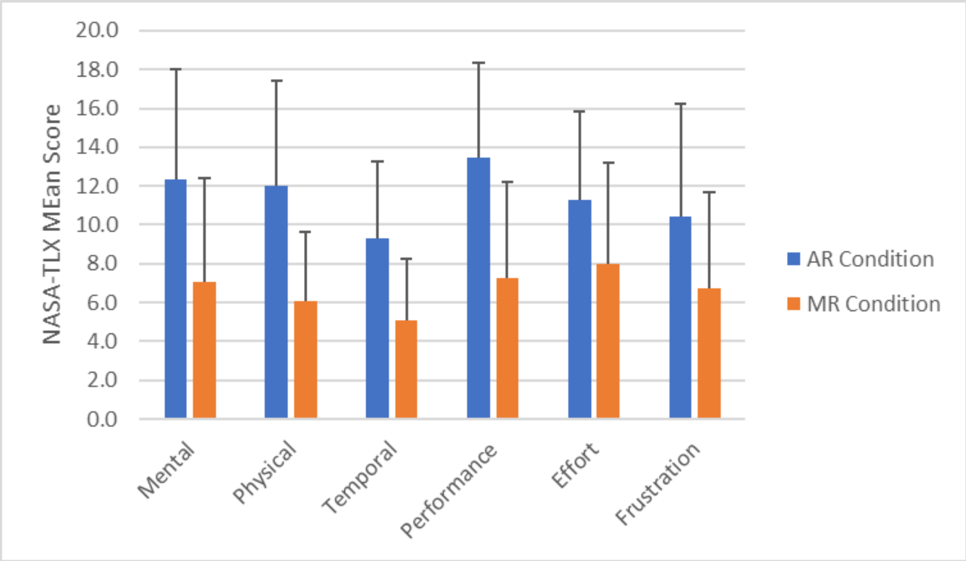

Figure 5.9: Mean NASA-TLX ratings across the two conditions (AR and MR). Error bars indicate the standard deviation from the average values.

5.4.3 Usability

Following the procedure, participants completed an SUS questionnaire (Brooke, 1996) for the condition in which they had been grouped. As illustrated in Figure 5.10, the results demonstrate that the AR condition was rated to have "poor" usability (M = 51.14 ± 11.58) while the MR condition was rated at an "acceptable" usability level (M = 74.77 ± 12.77). As previously stated, the universal average SUS score is 68 (Brooke, 1996; Sauro, 2011), and a score of less than 60 indicates relatively poor usability, while an average score of more than 80 indicates relatively good usability (Albert, 2008).

An independent samples t-test revealed that the MR condition featuring tangible interface objects was rated statistically significantly higher in usability than the AR condition in which such objects did not feature [$t(20) = 4.55$, $p<0.001$].

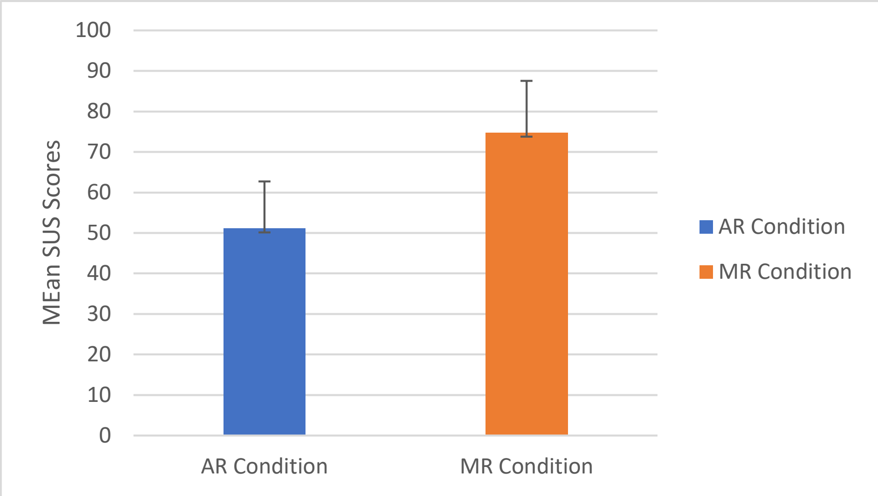

Figure 5.10: SUS scores for AR and MR conditions.

Error bars indicate standard deviation from the average values.

5.5 Conclusion

The experiment described in this chapter assessed the effect of tangibility on presence within MR systems. The first aim of the experiment was to determine if tangible interface elements within an MR environment had an effect on rated levels of presence. For the MR condition, tangible objects of the same size, shape and position as the virtual content were used. These provided affordances of the environment and tactile cues relating to the area of interactivity. They also restricted the participants' hand movements to the area of interactivity. The model of presence used within the experiment was the theoretical framework presented by Schubert et al. (2001).

The results demonstrate that the inclusion of tangible interface objects within an MR system produces a statistically significant positive effect on the multiple elements of presence, including general presence, spatial presence and involvement. However, no significant difference was found for the sense of realism sub-element. Therefore, the null hypothesis, that tangible interface objects have no effect on presence, is rejected, except for the sense of realism.

This result was expected, as the tangible objects within the MR condition provide physical spatial cues and tactile feedback relevant to the environment and virtual content.

The second aim was to determine whether the use of tangible interface elements within an MR environment affect rated levels of workload and usability. For both usability and workload, the MR condition scored significantly better for all measures. The null hypothesis, that tangible interface objects would have no effect on workload and usability, is therefore rejected. The results align with previous research which states that tangibility—the inclusion of tangible interface elements within a system—has a positive correlation with presence (Hoshi and Waterworth, 2009) and exhibits a significant association between presence and usability (Busch et al., 2014). The results also match those of previous studies which report that, when interacting in 3D space, tangible interfaces provide many benefits over traditional input devices for both performance and usability (Besancon et al, 2017).

Throughout the literature, many studies use the terms *VR*, *AR* and *MR* interchangeably without providing a precise working definition of each term. Without a precise working definition of the "reality" type used, it becomes unclear whether a study assessed an AR system in which information was simply superimposed over the world or an MR system in which there is a relationship between virtual and physical objects.

Many of the numerous presence questionnaires available are designed for use in VR systems, and few are relevant to AR or MR systems. The IPQ is one of the few questionnaires which had been previously tested and validated with AR environments (Regenbrecht and Schubert, 2002) and MR environments (Schaik et al., 2004). It should be noted that the study by Regenbrecht and Schubert (2002) uses a questionnaire based on the same theoretical model of presence as presented by Schubert et al. (2001), but it does not use the IPQ directly.

In addition, the IPQ is based on a theoretical model of workload that was consistent with the aims of this experiment: namely, the measurement of spatial presence. The IPQ provides a multi-dimensional method for measuring a user's subjectively rated sense of presence. Importantly, the model measured an aspect of presence that may be the most affected by the use of tangible objects: the degree of spatial presence the user perceives. This experiment demonstrates the application of the IPQ for use in MR systems under the working definition presented in this thesis.

5.5.1 Future work

Future work could be undertaken to build upon this study to assess a further range of tangible objects: not only using pre-defined objects, but also by using the immediate environment to opportunistically integrate virtual controls and displays with physical objects within the user's functional workspace in real-time. Early work by Henderson and Feiner (2010) assesses the use of "opportunistic" tangible interfaces that can leverage the affordances of the real world to embed virtual content and controls to real-world surfaces as and when they are found.

When comparing two conditions – one in which virtual content was integrated on to surfaces within the physical environment, and one in which the content was floating freely in 3D space – the study found that opportunistically placing virtual content onto appropriate tangible (physical) surfaces produces a faster completion time and error rate than was obtained for the baseline condition which displays the virtual content floating freely in 3D space.

However, the study consisted only of simple virtual content (labels and images) and limited interactivity (selecting buttons). The experimental testbed of the present experiment required intensive interactivity and used a variety of hand gestures to interact on 2D planes and in 3D space. Therefore, an experiment that utilises the concept presented by Henderson and Feiner (2010), but which assesses tasks which require a far greater degree of interactivity, would validate the use of opportunistic tangible interfaces for complex interactive tasks in addition to simple selection tasks.

Chapter 6: Environmental Conditions

6.1 Introduction

Previous chapters have addressed a number of assessment methods and evaluation criteria used to measure human performance with MR technologies in support of a variety of tasks and applications. This chapter continues the thread of investigation by assessing the use of MR technologies whilst under an uncontrollable environment that can affect performance. In this case, vibration is considered—which is one condition often encountered in military operations.

As detailed in Section 2.4.2.2, the literature states that an aircraft can produce vibration frequencies from 0.5Hz to 100Hz in all axes (Rash et al., 1999). However, there are specific frequencies within this range that are known to negatively affect the head (at 20-30Hz) and eyes (at 60-90Hz) when used to make system control inputs. A wide body of research has assessed the performance of the head and eye under specific frequency ranges, with most focusing on a narrow range of frequencies. This experiment assessed the performance of head and eye interaction over a wider range of frequencies: from 0 to100Hz.

In addition, numerous frequencies have been identified which are known to affect the human user in various ways:

- 2-5Hz causes a feeling of nausea (Ministry of Defence, 2008c)
- 4-6Hz causes severe visual performance degradation (Rash 2009) and display legibility (Tung et al., 2014)
- 2-10Hz produces significant unintended vertical eye movements (Uribe and Miller, 2013)

- 5-11Hz causes a degradation in reading performance (Lewis and Griffin, 1980)
- 2-20Hz causes an increase in target tracking errors with both the head and eyes (Shoenberger, 1972)
- 20-70Hz produces eye resonance which causes significant blurring (Griffin, 1990).

6.2 Aims

The aims within this particular experiment were, as before, rated as either primary (P) or secondary (S). The experiment aimed to do the following:

i) (P) Determine whether the interaction method used has an effect on accuracy under multiple whole-body vibration frequencies. The null hypothesis is that the interaction method used have no effect on accuracy.

ii) (P) Determine whether the vibration frequency used has an effect on accuracy. The null hypothesis is that the vibration frequencies used have no effect on accuracy either for eye- or head-based interaction.

iii) (S) Determine whether the frequency of whole-body vibration and the interaction method used has an effect on human mental workload. The null hypothesis is that the vibration of whole-body vibration has no effect on mental workload.

6.3 Method

6.3.1 Participants

A total of 11 participants were recruited on an opportunistic-availability basis. The sample consisted of 10 male undergraduate and postgraduate students and 1 female undergraduate student, with a mean age of 28.1 years and standard deviation of 8.0 years. All participants had prior experience with VR HMDs and had no prior experience with HMD-integrated eye-based interaction.

6.3.2 Independent Variables

The primary independent variable was the interaction method used. This consisted of two conditions: eye- and head-based system interaction. Both conditions utilised an HMD.

i) Eye-based interaction was achieved by having the participants move the eyes and the gaze-point to control the position of a cursor.
ii) Head-based interaction was achieved by having the participants move the head and the head-pose to control a head-slaved cursor.

The secondary independent variable is vibration frequencies. Five frequencies were considered: 0Hz (no vibration), 25Hz, 50Hz, 75Hz and 100Hz.

6.3.3 Apparatus

The device used for providing vibration, the *Buttkicker Gamer2* (Figure 6.1, left), is a commercial off-the-shelf device primarily used for gaming applications. The device was chosen due to the availability of the device within the research group laboratory and the simple fact that it provides for the precise control of vibration frequencies and vibration intensity. The device includes a controllable, high-power vibration motor which provides the ability to induce whole body vibration at frequency level ranges of 5-200 Hz. Because an aircraft can experience vibration frequencies between 0.5-100Hz, the lowest range of frequencies (0.5-5Hz) could not be provided by the device.

The vibration frequency was controlled by an analogue audio output source, and the vibration intensity was controlled by a control box and remote control (Figure 6.1, right). The frequency produced by the vibration device was tested for accuracy using a vibration monitoring mobile application which was placed on the chair in which participants sat. To simulate the varying frequency conditions within the experiment, an audio playlist with audio files of the five vibration frequencies were included, each lasting a period of 20 seconds each. The assessor could select each frequency audio file individually and in any order. The vibration intensity was set to maximum and static across all conditions. Using the integrated clamp, the device was attached to the frame of a chair and thus aimed to provide full-body vibration to the person sitting. The vibration frequencies provided could be precisely adjusted for the conditions within the experiment.

Figure 6.1: *Buttkicker Gamer2* Vibration device mounted on chair stem (Buttkicker, 2017)

The experiment used a FOVE HMD, a commercially available headset featuring integrated eye-tracking sensors (Figure 6.2), which was integrated with the Unity3D game engine using an external plugin. The plugin provided gaze-point data for the left and right eyes using the internal eye-tracking sensor module. It provided such data for the head orientation through a combination of internal IMU sensors and an external IR-based position tracking device. In order for participants to precisely position the head and eye within the interface, a cursor was provided in the form a reticule for both head and eye-based interaction conditions. The reticule was displayed continuously during the head-slaved conditions. However, during eye-based interaction, the update frequency of the eye-slaved cursor was once per 0.5 seconds.

The update frequency of the eye-slaved cursor must be carefully considered. During normal vision, the eye continuously performs small rapid movements, known as saccades. To account for this, some eye-tracking systems provide software methods for smoothing the gaze-point cursor. However, as the integrated eye-tracking system within the FOVE did not provide a built-in feature for eye-movement smoothing or update frequency, a custom cursor and display method was used. The eye gaze-point was tracked at a high refresh rate, but to compensate for the lack of cursor smoothing, the cursor displayed to the user was updated every 0.5 seconds.

Figure 6.2: FOVE HMD with integrated eye-tracking sensors (FOVE, 2017)

6.3.4 Procedure

Prior to beginning, participants read an experiment information sheet (Appendix K), completed a consent form (Appendix C) and read a health and safety document detailing exclusion criteria (Appendix B). The health and safety document detailed the risks and side effects of prolonged use of head-mounted displays was provided, and specifically, due to the use of an HMD and vibrations, the health and safety document specified that the participants must not wear glasses within the HMD, have any severe eye conditions or inner ear infections, any pain symptoms associated with the musculoskeletal system or suffer from severe motion sickness.

The experiment employed a within-subject design consisting of two interaction-type conditions (eye-based and head-based interaction) and five vibration-frequency conditions. The order in which participants completed each condition was dictated by a Latin-square ordering method. The vibration frequency order was randomised, with each of the five frequency conditions being tested once per interaction method. The experiment lasted approximately 20 minutes.

During the experiment, the participants were required to continuously look precisely at a point in the centre of the virtual display for the duration of the experiment: the “focus point”, as shown in Figure 6.3. For the two interaction-type conditions, the user completed the task of fixating on a point on the virtual display under five vibratory frequency conditions. Specific frequencies are known to produce specific issues for both eye and head-based interaction. The frequencies were as follows: no vibration (0Hz, a baseline comparison), low-frequency vibration (25Hz, known to cause problems with head movement), medium-frequency vibration, high-frequency vibration (75Hz, known to have a negative effect on the eye) and very-high-frequency vibration (100Hz). Each of the five vibration-frequency conditions lasted 20 seconds, and each was followed by a break of approximately 30 seconds for the participant to subjectively rate workload. As illustrated in Figure 6.4, the workload question was displayed within the HMD, and the participants verbally stated their rating between 0 and 10.

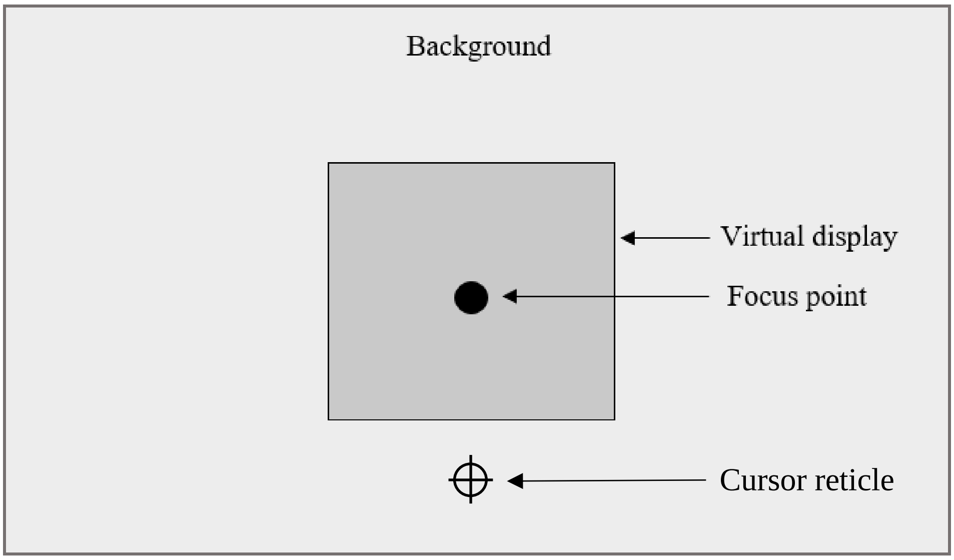

Figure 6.3: Illustration of interface layout and elements as seen through HMD

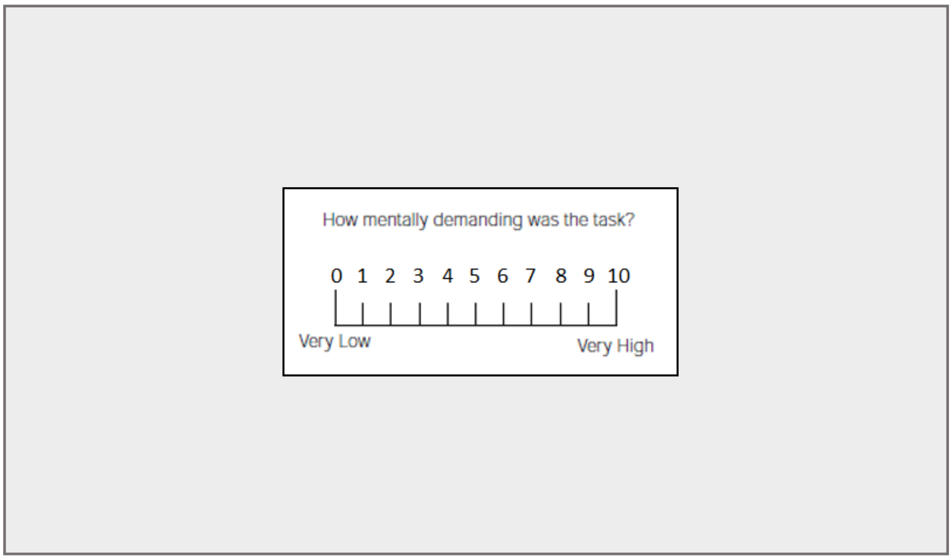

Figure 6.4: Illustration of workload question as seen through HMD

Prior to beginning the experiment, the HMD was adjusted to provide maximum comfort and to ensure a tight and secure mounting, thereby minimising incidences of spurious movements from vibration. During the subjective rating process, the user did not remove the HMD to help ensure that the mounting and fit did not change between conditions.

As the head and eyes were expected to move involuntarily during vibration, a variable to measure the deviation from the focus point was used. For this, a variable referred to as *divergence* was used. Divergence is defined as the mean radial distance of each individual measurement of the eye- or head-slaved cursor from the focus point in the centre of the virtual display. The divergence data for head and eye positions were collected automatically at a rate of 20Hz (one sample every 50ms) for the 20 seconds duration of each of the five vibration-frequency conditions A total of 400 samples was acquired over the 20 seconds duration. At the end of each condition, the system calculated a mean divergence variable for the X and Y coordinates. From this, a single mean result was recorded to an accuracy of two decimal places.

When fixating on the focus point, the system provided a position-tracking resolution in the X and Y axes of up to five decimal places. However, the level of accuracy and precision of both the head- and eye-tracking technologies are far below that of the position resolution recorded within the Unity game engine; thus, a lower-resolution divergence dataset was collected and rounded to two decimal places. Within the experiment, the divergence value was measured on a scale between 0.00 and 100.00. Because the square virtual display tracking area provided values for both X and Y, ranging from -100 to +100 across the radius of the central focus point, the mean distance for the negative figures was inverted (e.g. -5.00 becoming 5.00) so that, once averaged, the mean distance was a single variable of distance from the centre focus point, and the positive and negatives numbers did not negate each other. In addition, the X and Y values were averaged to provide a single divergence variable.

6.3.5 Dependent Variables

Two measures were included in the experiment, including accuracy and workload. These were the dependent variables within the experiment. Accuracy measurements were recorded continually and automatically throughout the experiment by the system.

i) Accuracy—Accuracy was measured using a divergence variable. Divergence was defined as the radial distance of the eye or head-controlled cursor reticle to the focus point in the centre of the virtual display. The divergence variable described the accuracy of the two conditions under the varying levels of vibration. A large divergence result signified a high level of inaccuracy.
ii) Workload rating—Following each condition, and without removing the HMD, the participants were asked to rate workload using the following question based on the mental workload question from NASA TLX (Hart and Staveland, 1988): "How mentally demanding was the task, between 0 (very low) and 10 (very high)".

The full NASA TLX questionnaire asks multiple questions, each addressing one dimension of the perceived workload: mental demand, physical demand, time pressure, perceived success, overall effort level and frustration level. Because there were 10 conditions in the experiment, 10 sets of TLX questionnaires would have been very intensive, so only mental workload was selected. Other questionnaires assess workload through a single question for overall workload, such as the short subjective instrument (SSI) questionnaire. One study compared the TLX and SSI questionnaires and found that both exhibit significant differences in scores relative to a range of tasks designed to vary in workload (Windell et al., 2006). This study also found that the single question questionnaire (SSI) was more sensitive than the multi-question questionnaire (TLX). Participants completed the workload questionnaire following each condition.

6.3.6 Statistical Analysis

Means are presented in a standardised format by displaying the mean value followed by the standard deviation value, presented as (M = x ± SD). The mean is presented with an accuracy of two decimal places.

As with previous experiments, objective accuracy readings was collected of interval-level data for more than three conditions, thereby enabling an ANOVA to be used. Overall, a 2 (interaction type) x 5 (vibration frequencies) repeated-measures ANOVA was used to assess the overall interaction effect of interaction types and vibration frequencies on accuracy. A further one-way, 5 (vibration frequencies) repeated-measures ANOVA was completed to individually assess the effect of vibration frequencies on eye- and head-based interaction. The assumption of sphericity was tested (as described in Section 4.2.3.6) along with subsequent pairwise comparisons for each vibration frequency.

Subjective workload reports were collected in the form of ordinal-level data for all vibration frequencies and both interaction methods. Unlike previous experiments within this thesis where workload was measured through a multi-question questionnaire, this experiment used a single question. Therefore, the ordinal-level data cannot be treated as interval data. Thus, a Friedman test was completed individually for eye- and head-interaction conditions to assess the effect of vibration frequencies on accuracy. Subsequent Wilcoxon signed-rank tests was further used to assess the differences between individual frequency combinations.

6.4 Results

6.4.1 Accuracy

Accuracy was defined as the mean radial distance of the eye- and head-controlled cursor from the centre of the screen in which a focus-point was placed and was described using a "divergence" variable. Figure 6.5 illustrates the mean divergence of both head- and eye-based interaction method conditions under the varying vibration frequency conditions.

The 0Hz condition was one in which the participants were not under any whole-body vibration, and it provided a baseline for comparison with the other vibration frequencies (25Hz, 50Hz, 75Hz, 100Hz). As shown in Figure 6.5 and Table 6.1, it was found that the overall effect of vibration frequencies on head-based interaction was minimal. The largest impact was observed during the 25Hz condition, and it reduced linearly across the higher frequency conditions. However, the effect on accuracy of eye-based interaction was very high across all frequencies, and also reduced across the higher frequency conditions. The 25Hz condition produced the largest negative effect by a large margin. Under the 25Hz condition, the eye-based interaction method would no longer a suitable method of system interaction where accuracy is required. Accuracy increased progressively across the 50Hz, 75Hz and 100Hz conditions, and only a small difference could be seen between the 0Hz and 100Hz conditions.

The effects of interaction method and vibration frequencies on accuracy was analysed using a two-way, 2 (interaction method) x 5 (vibration frequencies), repeated-measures ANOVA. Mauchly's Test of Sphericity reported that the assumption of sphericity was met. The ANOVA reported an overall significant main effect of vibration on accuracy [$F(4,36) = 32.320$, $p < 0.001$].

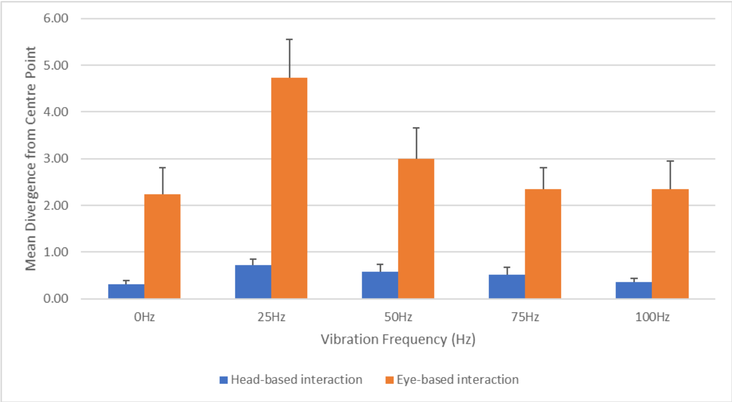

Figure 6.5: Effects of vibration frequencies on the head- and eye-interaction accuracy

Table 6.1: Results table for the mean accuracy of head- and eye-based interaction under varying vibration frequencies

	Head and Eye-Based Interaction Accuracy				
Mean Divergence from Centre	**Condition 1 (0 Hz)**	**Condition 2 (25 Hz)**	**Condition 3 (50 Hz)**	**Condition 4 (75 Hz)**	**Condition 5 (100 Hz)**
Head-based interaction	0.33 SD=0.07	0.74 SD=0.13	0.61 SD=0.16	0.55 SD=0.15	0.37 SD=0.09
Eye-based interaction	2.23 SD=0.57	4.80 SD=0.83	2.90 SD=0.67	2.32 SD=0.45	2.31 SD=0.61

A further analysis was performed to independently assess the significance of vibration frequencies on accuracy for eye- and head-based interaction conditions. First, the effect of vibration frequencies on accuracy during head-based interaction was assessed using a one-way, 5 (vibration frequencies) repeated-measures ANOVA. Mauchly's Test of Sphericity reported that the assumption of sphericity was violated. An ANOVA with Greenhouse-Geisser reported a significant main effect of vibration frequencies on accuracy during head-based interaction [$F(2.67,24.07) = 41.19$, $p < 0.001$].

A subsequent pairwise comparison found statistically significant differences between a range of frequency combinations, as seen in Table 6.2. When compared against the baseline condition with no vibration (0Hz), all frequencies were significantly different except for the highest frequency condition (100Hz) which shown no significant difference. This mirrors the accuracy findings (Figure 6.5) which reports that the 100Hz condition had a very minor difference to the 0Hz condition. Further insignificant differences were found across the mid-range frequencies combinations (50Hz and 75Hz).

Table 6.2: Pairwise comparison significance table for head-based interaction

(I) Vibration Frequencies	(I) Vibration Frequencies	Mean Difference (I-J)	Std. Error	Sig.[b]
0Hz	25Hz	-.404	.033	.000
	50Hz	-.260	.033	.000
	75Hz	-.207	.032	.000
	100Hz	-.038	.034	**.289**
25Hz	0Hz	.404	.033	.000
	50Hz	.144	.029	.001
	75Hz	.197	.034	.000
	100Hz	.366	.042	.000
50Hz	0Hz	.260	.033	.000
	25Hz	-.144	.029	.001
	75Hz	.053	.026	**.073**
	100Hz	.222	.048	.001
75Hz	0Hz	.207	.032	.000
	25Hz	-.197	.034	.000
	50Hz	-.053	.026	**.073**
	100Hz	.169	.048	.006
100Hz	0Hz	.038	.034	**.289**
	25Hz	-.366	.042	.000
	50Hz	-.222	.048	.001
	75Hz	-.169	.048	.006

Similarly, the effect of varying vibration frequencies on accuracy when using the eye-based interaction condition was assessed. To assess this, a one-way, 5 (vibration frequencies) repeated-measures ANOVA was taken. The assumption of sphericity was violated and a Greenhouse-Geisser correction was applied. The corrected ANOVA reported a significant main effect of vibration frequencies on accuracy for the eye-based interaction condition [$F(2.41,21.72) = 46.18$, $p < 0.001$]. As with the head-based interaction condition, a pairwise comparison reported significance differences for several frequency comparisons, as shown in Figure 6.3. When compared against the baseline condition with no vibration (0Hz), no significant differences was found with the higher-range frequency conditions (including 75Hz 100Hz). However, a statistically significant difference was found between the 0Hz condition and the lower-range frequencies (25Hz and 50Hz).

Table 6.3: Pairwise Comparison significance table for eye-based interaction

(I) Vibration Frequencies	(I) Vibration Frequencies	Mean Difference (I-J)	Std. Error	Sig.b
0Hz	25Hz	-2.491	.233	.000
	50Hz	-.750	.152	.001
	75Hz	-.105	.107	**.352**
	100Hz	-.100	.240	**.686**
25Hz	0Hz	2.491	.233	.000
	50Hz	1.741	.305	.000
	75Hz	2.386	.193	.000
	100Hz	2.391	.276	.000
50Hz	0Hz	.750	.152	.001
	25Hz	-1.741	.305	.000
	75Hz	.645	.134	.001
	100Hz	.650	.250	.029
75Hz	0Hz	.105	.107	**.352**
	25Hz	-2.386	.193	.000
	50Hz	-.645	.134	.001
	100Hz	.005	.211	**.982**
100Hz	0Hz	.100	.240	**.686**
	25Hz	-2.391	.276	.000
	50Hz	-.650	.250	.029
	75Hz	-.005	.211	**.982**

6.4.2 Workload

Following each frequency condition, the participants were asked to rate their responses to the following question, which was constructed based on the mental workload question from NASA TLX (Hart and Staveland, 1988): “How mentally demanding was the task, between 0 (very low) and 10 (very high)”. Figure 6.4 illustrates the subjective workload results for both head- and eye-based interaction.

Overall, the head-based interaction condition produced a low level of workload for all vibration frequencies, whereas the eye-based interaction condition produced a higher level of workload across all vibration frequencies. The results closely match the accuracy results, as seen in Figure 6.6 and Table 6.4, where the 25Hz condition produces the highest level of workload, with the workload rating decreasing steadily until 100Hz, where it produces a response similar to that of the 0Hz condition (with no vibration present). The results for the eye-based interaction produced a higher workload rating across all measures relative to the head-based interaction.

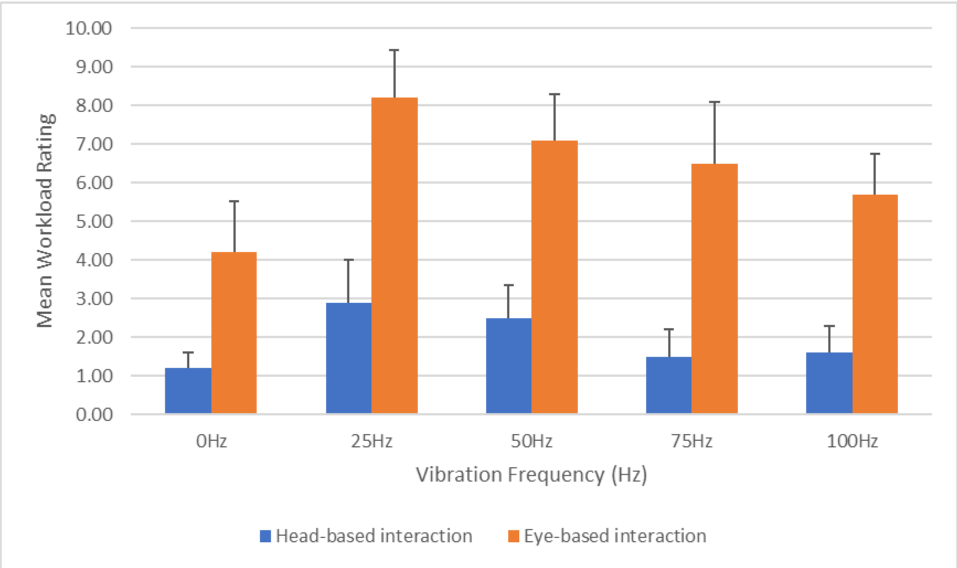

Figure 6.3: Effects of vibration frequencies on mental-workload ratings for head and eye interaction

Table 6.4: Results table for mean-workload rating for head- and eye-based interaction under varying vibration frequencies

	Workload Rating Scores				
Mean Workload Rating	**Condition 1 (0 Hz)**	**Condition 2 (25 Hz)**	**Condition 3 (50 Hz)**	**Condition 4 (75 Hz)**	**Condition 5 (100 Hz)**
Head-based interaction	1.20	2.90	2.50	1.50	1.60
Eye-based interaction	4.20	8.20	7.10	6.50	5.70

For the eye-based interaction condition, a Friedman test found a significant difference in accuracy across the five vibration frequencies [$\chi^2(4) = 23.45$, $p < 0.001$]. *Post-hoc* analyses with the Wilcoxon signed-rank test was conducted with a Bonferroni correction applied ($p < 0.05$ becoming $p < 0.01$). As detailed in Table 6.5, the analyses reveals statistically significant differences between a range of frequency conditions. When compared against the baseline condition with no vibration (0Hz), significant differences was found across all vibration conditions except for the highest range frequency condition (100Hz). Additional significant differences was found for the remaining condition combinations, as shown below. Table 6.5 shows all the frequency combinations and correlate results, with only unique results included. Duplicates (e.g. 0Hz and 25Hz, and 25Hz and 0Hz) were removed as they produce the same results.

Table 6.5: Wilcoxon signed-rank test results for eye-based interaction

(I) Vibration Frequencies	**(I) Vibration Frequencies**	**Z value**	**Sig.**
0Hz	25Hz	-2.952	0.003
	50Hz	-2.956	0.003
	75Hz	-2.687	0.007
	100Hz	-2.214	**0.027**
25Hz	50Hz	-2.176	**0.030**
	75Hz	-2.687	0.007
	100Hz	-2.823	0.005
50Hz	75Hz	-1.723	**0.085**
	100Hz	-2.754	0.006
75Hz	100Hz	-1.994	**0.046**

For head-based interaction, a Friedman test found a significant difference in accuracy across the five vibration frequencies [$\chi^2(4) = 32.07$, $p < 0.001$]. *Post-hoc* analyses with the Wilcoxon signed-rank test were conducted with a Bonferroni correction applied ($p < 0.05$ becoming $p < 0.01$). As detailed in Table 6.6, the analyses reveals statistically significant differences between a range of frequency conditions. When compared against the baseline condition with no vibration (0Hz), no significant differences was found with the higher-range frequency conditions (including 75Hz and 100Hz). However, a statistically significant difference was found between the 0Hz condition and the lower-range frequencies (25Hz and 50Hz). Further differences between conditions can be seen below.

Table 6.6: Wilcoxon signed-rank test results for head-based interaction

(I) Vibration Frequencies	(I) Vibration Frequencies	Z value	Sig.
0Hz	25Hz	-2.850	0.004
	50Hz	-2.739	0.006
	75Hz	-1.342	**0.180**
	100Hz	-1.635	**0.096**
25Hz	50Hz	-1.667	**0.096**
	75Hz	-2.877	0.004
	100Hz	-2.724	0.006
50Hz	75Hz	-2.640	0.008
	100Hz	-2.260	**0.014**
75Hz	100Hz	-1.000	**0.317**

6.5 Conclusion

The experiment assessed a commonly occurring uncontrollable environment condition, vibration, and its effect on human performance (measured by accuracy) with respect to both eye- and head-based interaction within an MR system. Furthermore, the effect of vibration frequencies on mental workload was assessed for each vibration frequency across both eye and head interaction conditions.

The first aim of the experiment was to determine whether the interaction method used had an effect on accuracy under multiple whole-body vibration frequencies. Overall, the head-based interaction method produced a far greater degree of accuracy than the eye-based interaction, measured by the divergence of the eye- and head-controlled cursor from the centre of the display. Statistical analysis revealed a significant overall main effect between the interaction conditions. Therefore, the null hypothesis, that no effect can be found between interaction methods and accuracy, is rejected.

The second aim was to determine whether particular vibration frequencies affect accuracy for both interaction methods. Accuracy was individually measured for all five vibration frequencies and for both interaction conditions. It was found that, when using the head-based interaction condition, the vibration frequencies tested produced a very minor effect on accuracy. The level of accuracy was high, and it allowed participants to focus on a point on a screen accurately for a prolonged period, presenting no issues for dwell-based interaction. At no point during the experiment was it observed that the participants jolted or moved their heads to the extent that a dwelling period of 500ms would be interrupted or that it caused any involuntary movements.

The literature states that frequencies between 20-30Hz can cause negative effects on the head, neck and shoulders (Ministry of Defence, 2008c). However, this experiment shows that, whilst 25Hz produced the lowest accuracy, it has only a minor impact and thus does not invalidate the use of head-tracking for interaction purposes. However, the minor effect of vibration on accuracy may be due to the relatively low intensity of vibration provided by the *Buttkicker* vibration device (compared to operational vehicles which can produce intensive whole-platform turbulence).

However, during eye-based interaction, the effect was very large across all vibration conditions and was most prominent during the lowest-frequency vibration condition tested of 25Hz. During the condition in which no vibration was present (0Hz), the eye remained stable such that it could successfully dwell on a point for a period of 500ms. However, multiple instances were observed in which the gaze-point would involuntarily move momentarily away from the focus point. Under the 25Hz condition, the eye gaze-point would often sporadically and unintentionally move around the screen and would therefore be unsuitable for dwell-based interaction. The literature states that the eyeball resonates at a frequency range of 60-90Hz (Ministry of Defence, 2008c) and would, therefore, be unusable for intentional input. While the 75Hz condition did provide a high level of inaccuracy, it was did not produce a significantly large level of accuracy degradation compared to the other frequencies tested.

Statistical analysis revealed a significant overall main effect of accuracy due to vibration frequencies for both eye- and head-based interaction conditions. However, further analysis revealed this was not found between all combinations of vibration-frequency conditions. Therefore, the null hypothesis, that vibration frequency has no effect on accuracy for both interaction conditions, is rejected.

The third aim of the experiment was to determine whether the frequency of whole-body vibration and the interaction method used had an effect on mental workload. Statistical analysis revealed a significant main effect of vibration frequencies on mental workload for both interaction methods. A further analysis revealed a significant effect of various vibration frequencies for both eye- and head-based interaction, but not for all individual frequency combinations. Across both interaction conditions, no significant differences was found between the baseline condition with no vibration (0Hz) and the high-frequency condition (100hz). The mean accuracy data mirrors this findings where the 0Hz and 100Hz frequency conditions produced very similar results. Therefore, the null hypothesis, that that vibration frequency has no effect on mental workload, is rejected.

In conclusion, the experiment found that vibration has a minor impact on head-based interaction and a major impact on eye-based interaction when interacting with virtual displays within an MR HMD. The results indicate that head-based interaction is suitable for dwell-based interaction during vibration, whereas eye-based interaction is unsuitable during vibration. All vibration conditions had a negative impact on accuracy and mental workload, with the lower-frequency vibrations (25Hz) providing the largest negative impact on accuracy. The effect decreases as the frequencies increased across 50Hz, 75Hz, and 100Hz.

6.5.1 Future Work

The vibration frequencies tested were intended to investigate a broad range known to affect specific areas of user or technology performance. However, as they covered such a broad range, with large intervals in between (25Hz increments), the intermediate intervals may provide significant effects but were not tested. To build upon the experiment presented within this chapter, future work could aim to increase the frequency-range conditions from very broad (25Hz to 100Hz in increments of 25Hz) to a greater resolution (such as 25Hz to 50Hz in increments of 5Hz). In addition, further work could assess a narrow range of frequencies of increased resolution on a lower frequency range, such as those presented within other research, including 0.5-5Hz in increments of 0.25Hz.

In addition, the experiment tested the eye-based interaction condition by instructing the participant to keep his or her head pose stationary on the focus point in the centre of the display. This was selected for technical reasons, as the stability of the eye tracker was known to reduce as the users' gaze point reached the extremities of the tracking area. To provide a stable and reliable gaze-point reading, the participants were required to look at the centre of the display, which provided the most stable eye tracking. However, requiring the participants to keep a static head-pose is not representative of a real-world display in which the point of focus may be anywhere on the display (and, indeed elsewhere in the user's immediate workspace). Therefore, the experiment could be repeated with multiple targets that are distributed throughout the display by using non-static targets that move around the virtual display.

The vibration device used in this case was a low-cost commercial off-the-shelf device attached to the stem of a chair to provide whole-body vibration of equal intensity. However, because this was such a low-cost device as opposed to a full motion platform, it is not certain whether the device provided vibration equally throughout the whole body or was localised to the seated torso area, reducing in intensity as the effects travelled up the body towards the head and eyes. A reduction in vibration upwards from the seat may have reduced the impact on head and eye stability and accuracy. However, even if a reduction in intensity did occur, the experiment still measured vibration over specific frequencies but only with low intensity. A further experiment could be undertaken to test the different vibration frequencies with an additional independent variable of intensity.

Chapter 7: Discussion & Conclusions

7.1 Discussion

The aim of the research described within this thesis was to determine a set of principles for the development and deployment of MR technologies in the context of future defence human-system interfaces. The main outcome of the studies undertaken was the provision of evidence-based recommendations regarding the suitability and appropriateness of MR technologies for a selection of tasks and contexts. The experiments were performed primarily to measure human performance when interacting with virtual displays in an MR system, specifically addressing the use of different variations in interface designs under varying environmental conditions. This was achieved through a human-centred technology assessment, supplemented by a series of experiments, each addressing different areas of the design and use of MR systems in a simulated operational setting. Five specific objectives were defined at the beginning of the study.

The first objective was to review the literature relating to the design and use of MR systems and to identify the most relevant design criteria available. A review of a range of standards, guidelines and studies found many detailed design recommendations for mature technologies such as touchscreens, but few relevant to virtual displays within AR or MR systems. Similarly, a wide body of research was found which presents human-performance assessments for mature technologies, but none were found to assess MR systems. Many standards and studies detail the effect of using head- and eye-based input on performance while under vibration. Both head and eye input methods are commonly used with MR systems, but both are known to be susceptible to environmental conditions such as vibration. A range of frequency vibrations were identified that might affect the user and were presented with relevance to head- and eye-based input.

The second objective was to evaluate the suitability and readiness of a range of MR display and interaction devices when used for real-time interaction with virtual displays. To achieve this, a technology assessment was undertaken based around an MR "wearable cockpit" (a particular topic of interest on the part of the co-sponsor of the research, BAE Systems). This included the development of a modular software and hardware testbed which integrated a wide range of devices, including tracking technologies, input devices and display methods. The testbed supported the rapid integration of the wide range of MR devices available to the author for further HF experimentation. Observations and informal interview sessions were performed with experienced systems engineers using the system, with qualitative usability feedback being recorded. The sessions found that eye- and hand-based inputs were unsuitable for complex system interactions requiring multiple or multi-step gestures, with only head-based input being stable and reliable.

The third objective was to determine which tasks an MR system would be most suitable for, measured by objective human performance metrics and a range of subjective human-centred measures. To address this objective, Chapter 4 presented two experiments assessing human performance when using virtual displays in an MR system: the first when performing simple, single-gesture input tasks; the second when performing more complex, multi-gesture interaction tasks. Two further HMI devices were assessed that utilised a physical display, including a touchscreen and a HOTAS joystick, both of which were representative of those used in current-generation aircraft.

Individual findings across tasks and subjective measures can be found in Chapter 4. In summary, the results revealed that, when completing both selection tasks and interaction tasks, the MR system performed below the touchscreen condition but above the HOTAS condition for both reaction times and error rate. Both the touchscreen and HOTAS are utilised within existing-generation aircraft. If an application requires that users exercise both fast reaction times and low error rates, then a MR system can considered for use. Subjectively, the results also indicated that, for selection tasks, participants marginally preferred the MR system over a touchscreen. However, for interaction tasks, participants preferred the touchscreen by a significant margin. The HOTAS was rated as least preferable for both tasks. This is in line with the recommendations from military design standards which state that head-based input is most suitable for simple input tasks and not complex or prolonged input activities.

The fourth objective was to investigate the use of physical (tangible) objects within the environment as a tool for interaction with virtual content. A key factor of an MR system is its relationship to the real-world physical environment. In the context of XR, tangibility refers to the interaction between physical and virtual objects in an environment, and has been shown to affect presence, usability and other factors. While the effect of tangibility on presence has been demonstrated in multiple studies for VR, no formal experiments have assessed the effect of tangibility within MR.

To assess the effect, Chapter 5 presented an experiment that assessed the effect of tangibility on presence within an MR environment when interacting with a C2 system featuring a high level of interactivity. The group of participants that completed tasks with virtual objects co-located with physical objects reported significantly higher ratings presence, usability and workload than the group without physical objects. This result indicated that tangibility is positively correlated with presence in MR, as found by previous studies for VR. The experiment also further utilised the use of the IPQ presence questionnaire and demonstrated its use for MR systems in addition to VR and AR systems as used in previous studies.

The fifth objective was to investigate the effect of vibration on the use of eye- and head-based interaction with virtual displays. To address this, Chapter 6 assessed accuracy levels when using head- and eye-based interaction methods when interacting with virtual displays under varying vibration frequencies (0Hz, 25Hz, 50Hz, 75Hz, 100Hz). The frequencies tested broadly covered the frequencies present within rotary aircraft operating environments. It was found that head-based interaction provided higher levels of accuracy than eye-based interaction when dwelling on a point under vibration. Dwell-based input is a common method of interaction and simply requires a user to dwell on an object for a period of time. A duration of 500ms has been shown to minimise unintentional input. In addition, for both interaction type conditions, further analysis revealed that lower frequencies (25Hz) had a statistically significant effect on accuracy and workload, and that higher frequencies (100Hz) did not produce a significant effect. However, for eye-based interaction, vibration produced a statistically significant degradation in accuracy, with the most prominent effect caused by lower frequencies (25Hz). Using a dwell-time activation period of 500ms, the eye-based input was unsuitable as an interaction method, as the eyes would jitter uncontrollably around the screen, meaning it was not possible for participants to maintain a stable, intentional gaze-point for a 500ms period.

7.2 Principles for the Deployment of Mixed Reality Systems

This study has contributed to the academic field of applied HF and human-machine interaction by determining a number of early principles supporting the development and deployment of MR technologies in a defence context. During the literature search conducted in support of the research, a significant absence was identified of formal studies and in-depth assessments of MR systems. At the time of the literature search, no single study had performed a rigorous and formal ergonomic assessment of MR systems in the same way that more established technologies such as touchscreens had been assessed. While there were studies assessing human performance of MR systems, they were often based on software evaluations with no consideration for important HF such as the physical and cognitive demands imposed on the user by the devices being utilised.

A series of reviews and experiments were conducted to systematically investigate aspects of the design and use of MR systems including the relevance of design criteria used for MR, the types of tasks most suitable to MR systems, and the context in which MR may operate. The principles of development of MR systems are as follows:

- Observations and interviews with the end-user demographic of a system will provide invaluable qualitative feedback regarding the key challenges and considerations faced by end users that may not be readily apparent within existing design guidelines. The environment in which the user operates can restrict the types of devices that can be used and can affect human and system performance (e.g. by imposing body movement restrictions due to an enclosed seated position or vibration caused by a moving platform).
- The types of display and input devices used have a significant impact on performance and usability. When selecting MR display and interaction devices to use, a human-centred

design evaluation of appropriateness, as described by Stone (2008, 2012), should be undertaken to identify the suitability of devices for a range of tasks and contexts.

- For emerging technologies such as MR systems, there are no established or formal design standards that define the interface design criteria. It has been shown that the shape, colour and size of objects affect performance (measured by reaction time and error rates) and influence subjective measures such as workload and usability. Therefore, a performance evaluation should be undertaken to ensure that the design criteria chosen are the most optimal.
- When performing selection and interaction tasks, the display and interaction devices used have a significant impact on performance and usability. While MR systems are suitable for simple selection gestures, MR may not be suitable (compared to more mature technologies such as touchscreens) for complex interaction gestures in performance-critical applications.
- The use of tangible user interface objects (virtual objects co-located with physical objects), can improve the reported presence, usability and workload experienced by users. By including physical objects of the same shape, size, and position as virtual objects, the user can employ the affordances of the real world to provide haptic feedback and spatial cues to the user, thereby indicating the area of interactivity and the correct position to place the hands when interacting with virtual objects.
- Environmental conditions that are imposed on a user or system can significantly affect human performance and workload. It was found that vibration has a minor impact on performance in relation to head-based input. However, vibration has a significant and negative impact on eye-based input. In addition, outside environments may render an optically see-through HMD unusable due to technical elements of the display and optics. A human-centred evaluation should be undertaken prior to selecting a device – preferably in the environment in which the device is to operate.

7.3 Industry and Public Engagement

As the PhD is an industrially sponsored iCase studentship, this study involved collaboration with the industrial sponsor, BAE Systems, during development and testing of the experimental testbeds. As stated by the funding council, EPSRC, an iCase studentship aims to provide an industrially relevant, applied experimental approach to provide the researcher with facilities and expertise not available in academic settings alone. During the early stages of the system-development process, the concept testbed iterations were demonstrated to the industrial sponsor, including systems engineers, HF engineers and test pilots. Seeking qualitative general usability and appropriateness feedback from experts and end-user demographics gave insight into the potential future adoption of MR technologies. Informal observations and interviewing provided a unique insight into the needs and limitations imposed upon end users that are not commonly addressed within the literature. In addition, the software and hardware testbeds developed over the course of this study were demonstrated to a wide range of audiences at industry and academic conferences (Bibb, 2017), exhibitions and workshops. Informal usability feedback was sought from end users, from military personnel (including test pilots, soldiers and commanders), internal personnel from within BAE Systems and from the general public.

7.4 Limitations and Future Work

The technology assessment presented in Chapter 3 considered the usability and suitability of a range of display and interaction devices used for MR. The technology assessment collected only informal observations and interview data relating to usability, with no objective performance data collected. A future extension of this work would seek to perform a formal technology evaluation and introduce appropriate objective human and system performance metrics to support the findings and recommendations presented.

The first experiment, 1a, investigated the effect of interface design criteria on human performance and a range of subjective factors. The experiment assessed three devices as participants completed 24 input tasks with varying object sizing and spacing (six sizes and four spacing conditions). One limitation was in the limited number of conditions that could be tested, due to the layered experiment design (three device types x six object sizes x four spacing distances). Due to the large number of conditions tested, additional design criteria were not tested. This was done to stop the experiment from becoming very long and complex. Because many spacing distances were found to have no statistically significant effect, the spacing conditions could be removed and replaced with other object design criteria variations. This would allow an experiment to be conducted to assess other design options, such as increasing the number of object size conditions from 6 to 24 without increasing the experiment procedures time or complexity.

A further limitation of the study is in the limited availability of high-end (enterprise or research-level) devices and related software packages. As detailed in Chapter 2, the technical specifications of HMD systems, such as the field of view and resolution, have been shown to have an effect on several factors, including human performance and presence (Qian, 2015;

Toczek, 2016). The specifications of display technologies are rapidly increasing, resulting in VR HMD systems increasing in display resolution from relatively low in 2010 (10 pixels per degree; Oculus Rift DK1 HMD) to “human-eye” resolution displays in 2019 (60 pixels per degree; Varjo VR-1 HMD).

Further experimentation could be undertaken to repeat the experiments within this study using new, higher specification devices as they are made available. One such experiment might, for example, investigate the effect of specific device specifications, such as resolution, to see if they produce different results. This would demonstrate whether the objective performance and subjective metrics (such as presence and usability) found within this thesis can be generalised to account for other HMD systems, or if the effects differ depending on external factors, such as the resolution of the HMD used. As the software testbed presented within this thesis used a modular system design, further devices could be easily and quickly integrated for further experimentation.

REFERENCES

Apple (2018) Human Interface Guidelines (Online). Available at: https://developer.apple.com/ios/human-interface-guidelines/technologies/augmented-reality/ (Accessed 10 January 2018)

Asare, C. (2015) Human Performance Issues With Helmet Mounted Displays In Rotary Aircraft Cockpits (Online). Available at: http://www.perceptionenhancement.com/docs/course /Case%20History%202.pdf (Accessed 08 March 2018)

BAE Systems (2015) Augmented reality systems set to revolutionise battlefield operations (Online). Available at: https://www.baesystems.com/en-uk/article/augmented-reality-systems-set-to-revolutionise-battlefield-operations (Accessed 13 April 2018)

BAE Systems (2017) BAE Systems mixed reality cockpit technology aims to revolutionise future aircraft (Online). Available at: https://www.baesystems.com/en/article/bae-systems-mixed-reality-cockpit-technology-aims-to-revolutionise-future-aircraft----- (Accessed 21 July 2018)

BAE Systems (2018a) Striker® II Digital Helmet-Mounted Display (Online). Available at: https://www.baesystems.com/en-uk/product/striker-ii-digital-helmet-mounted-display (Accessed 11 January 2018)

BAE Systems (2018b) LiteHUD® Head-Up Display (Online). Available at: https://www.baesystems.com/en-us/product/litehud-headup-display (Accessed 11 January 2018)

Besancon, L., Issartel, P., Ammi, M. and Isenber, T. (2017) ‘Mouse, Touch and Tangible Inputs for 3D Data Manipulation’, *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. Colorado, USA: CHI, pp. 4727-4740.

Bibb, C. (2017) ‘A Human Factors Evaluation of Mixed Reality Technologies for Command and Control Applications’, *Defence and Security Doctoral Symposium (DSDS17).* Available at https://cord.cranfield.ac.uk/articles/A_Human_Factors_Evaluation_of_Mixed_Reality_Technologies_for_Command_and_Control_Applications/5752803/1 (Accessed 9 September 2019)

Borg, GA. (1982) ‘Psychophysical bases of perceived exertion’, *Medical Science Sports Exercise*, 14(5), pp. 377–81.

Borsci, S., Lawson, G. and Broome, S. (2015) ‘Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance’, *Computers in Industry*, 67, pp. 17-26.

Brooke, J. (1996) ‘SUS-A quick and dirty usability scale’, *Usability evaluation in industry*, 189(194), pp. 4-7.

Burdea, G. and Coiffet, P. (1994) *Virtual reality technology*. New York, USA: Wiley

Bures, M. and Polcar, J. (2016) ‘Comparison of 3D scanning and 3D modelling of a workplace from various aspects’, *2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM).* Bali, Indonesia: IEEE, pp. 306-310.

Busch, M., Lorenz, M. and Tscheligi, M. Hochleitner, C. and Schulz, T. (2014) Being there for real: presence in real and virtual environments and its relation to usability. Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. Helsinski, Findland: ACM, 117-126.

Bustamante, E. and Spain, R. (2008) "Measurement Invariance of the NASA TLX", *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52(19), pp. 1522–1526.

Buttkicker (2017) Buttkicker Gamer 2 (Online). Available at: https://thebuttkicker.com/buttkicker-gamer2/ (Accessed 4 June 2017)

Carlin, A.S., Hoffman, H.G. and Weghorst, S. (1997) ‘Virtual reality and tactile augmentation in the treatment of spider phobia: a case report’, *Behaviour Research and Therapy*, 35(2), pp. 53-158.

Charlton, S. and O’Brien, T. (2002) *Handbook of Human Factors Testing and Evaluation.* 2nd edn. Mahwah, New Jersey, USA: Lawrence Erlbaum Associates

Cheng, Y (2014) DARPA-funded augmented reality technology moves to commercial sector -- Defense Systems (Online). Available at: https://defensesystems.com/articles/2014/05/29/ara-augmented-reality.aspx (Accessed 11 January 2018)

Cipresso, P., Giglioli, I., Raya, M. A. and Riva, G. (2018) ‘The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature’, *Frontiers in Psychology*, 9, pp. 1-20.

Clemente, M., Rodriguez, A., Rey, B. and Alcaniz, M. (2014) ‘Assessment of the influence of navigation control and screen size on the sense of presence in virtual reality using EEG’, *Expert Systems with Applications*, 41, pp. 1584-1592.

Cockburn, A., Gutwin, C., Palanque, P., Deleris, Y., Trask, C., Coveney, A., Yung, M. and MacLean, K. (2017) ‘Turbulent touch: Touch screen input for cockpit flight displays’, *Proceedings of the 2017 CHI Conference for Human Factors in Computing Systems*, pp. 6742-6752

Colle, H. and Hiszem, K. (2004) ‘Standing at a kiosk: Effects of key size and spacing on touch screen numberic keypad performance and user preference’, *Ergonomics*, 47(13), pp. 1406-1423.

Compos, L. and Menegaldo, L. (2018) ‘A battle tank simulator for eye and hand coordination tasks under horizontal whole-body vibration’, *Journal of Low Frequency Noise Vibration and Active Control*, 37(1), pp. 1-12.

De Crescenzio, F., Fantini, M., Persiani, F., Di Stefano, L., Azzari, P. and Salti, S. (2011) ‘Augmented Reality for Aircraft Maintenance Training and Operations Support’, *IEEE Computer Graphics and Applications*, 31(1), pp. 96-101.

Defence Academy of the United Kingdom (2018) Defence Simulation Centre (Online). Available at: https://www.da.mod.uk/colleges-schools/technology-school/defence-simulation-centre (Accessed 3 January 2018)

Defence Simulation Centre (2018) 3D Models: DTEC Catalogue (Online). Available at: https://www.da.mod.uk/Portals/0/Documents/DSC/DTEC_Catalogue_Part_1_Reuse_Catalogue_3D_Models.pdf?ver=2017-10-30-135745-253 (Accessed 4 January 2018)

DEFRA (2018) Survey Open Data (Online). Available at: http://environment.data.gov.uk/ds/survey/#/survey?grid=SP08 (Accessed 3 January 2018)

Dempsey, N. (2019) Briefing Paper CBP7930: UK Defence Personnel Statistics, House of Commons Library. UK: House of Commons Library

Department of Defense (2012) MIL-STD-1472G, Department of Defence Design Criteria Standard: Human Engineering, Department of Defence. USA: Department of Defense

Dunser, A., Steinbugl, K., Kaufmann, H. and Gluck, J. (2006) ‘Virtual and augmented reality as spatial ability training tools’, *Proceedings of 7th international conference on computer-human interaction: design centres HCI (SIGCH).* Christchurch, New Zealand: ACM, pp. 125-132.

Ebrahim, M (2013) '3D Laser Scanners' Techniques Overview', *International Journal of Science and Research*, 3(10), pp. 323-331.

EMS (2018) Types of 3D Scanners and 3D Scanning Technologies (Online). Available at: https://www.ems-usa.com/tech-papers/3D%20Scanning%20Technologies%20.pdf (Accessed 4 January 2018)

Endsley, V. J (1988) 'Situational Awareness Global Assessment Technique (SAGAT)', *Proceedings of the National Aerospace and Electronics Conference*. pp. 789-795.

Field, A. (2013) *Discovering statistics using IBM SPSS statistics*. 4th edn. London, UK: Sage Publications

Fingas, J (2014) New fighter pilot helmet delivers night vision without goggles (Online). Available at: https://www.engadget.com/2014/07/16/bae-striker-ii-helmet/ (Accessed 19 January 2018)

Fitts, P. (1954) 'The information capacity of the human motor system in controlling the amplitude of movement', *Journal of Experimental Psychology*, 47(6), pp. 381–391.

Fletcher, L., Park, H. and Martin, B. (1992) 'Effects of Hand Vibration on Postural Stability', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 36(10), pp. 765-769.

FOVE (2017) FOVE Eye Tracking Virtual Reality Headset (Online). Available at: https://www.getfove.com/ (Accessed 4 June 2017)

Gawron, V (2008) *Human Performance, Workload, and Situational Awareness Measures Handbook*. Mahwah, NJ, USA: CRC Press

Grabowski, M., Rowen, A. and Rancy, J. (2018) 'Evaluation of wearable immersive augmented reality technology in safety-critical systems', *Safety Science*, 103, pp 23-32.

Griffin, M (1990) *Handbook of Human Vibration*. London, UK: Academic Press

Harich, J (2002) Comparative Pain Scale (Online). Available at: https://www.thwink.org/personal/ComparativePainScale.html (Accessed 16 October 2017)

Hart, S. (2006) 'NASA-Task Load Index (NASA-TLX); 20 Years Later', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), pp. 904–908.

Hart, S. and Staveland, L. E (1988) ‘Development of NASA-TLX (Task Load Index): Results of impirical and theoretical research’, *Advances in Psychology,* 52, pp. 139–183.

Henderson, S. and Feiner, S. (2009) ‘Evaluating the Benefits of Augmented Reality for Task Localization in Maintenance of an Armoured Personnel Carrier Turret’, *IEEE International Symposium on Mixed and Augmented Reality,* 8, pp. 135-144.

Henderson, S. and Feiner, S. (2010) ‘Opportunistic Tangible User Interfaces for Augmented Reality’, *IEEE Transactions on Visualization and Computer Graphics*, 16(1), pp. 4-16.

Hendrix. C.M. (1994) *Exploratory Studies on the Sense of Presence in Virtual Environments as a Function of Visual and Auditory Display Parameters*, Master's thesis. Washington: University of Washington.

HFI (2017) *Human Factors Integration: Technical Guide for Display and Control Technologies*. UK: Defence Authority

Hoffmann, E. (1991) "A comparison of hand and foot movement times", *Ergonomics*, 34(4), pp. 397–406.

Hoshi, K. and Waterworth, J. (2009) ‘Tangible presence in blended reality space’, *Proceedings of the 12th Annual International Workshop on Presence*, pp. 1–10.

iGroup (2006) igroup presence questionnaire (IPQ) (Online). Available at: http://www.igroup.org/pq/ipq/download.php (Accessed 15 February 2018)

International Organisation for Standardization (2002) ISO 9241–9. 2000 Ergonomics requirements for once work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices. Geneva, Switzerland: International Organisation for Standardization.

International Organisation for Standardization (2010) ISO 9241-210:2010 Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems. Geneva, Switzerland: International Organisation for Standardization.

International Organisation for Standardization (2012) ISO 9241-411Ergonomics of human-system interaction – Part 411: Evaluation methods for the design of physical input devices. Geneva, Switzerland: International Organisation for Standardization.

ISO (2019). International Organisation for Standardization: Certification (Online). Available at: https://www.iso.org/certification.html. (Accessed 19 June 2019)

Jacob, R. and Karn, K. (2003) ‘Eye tracking in Human-Computer Interaction and usability research: Ready to deliver the promises’ in Hyona, J., Radach, R. and Deubel, H. (ed.) *The Mind’s Eye: Cognitive and applied aspects of eye movement research.* Oxford, England: Elsevier Science, pp 573-605.

Kalawsky, R (2000) The Validity of Presence as a Reliable Human Performance Metric in Immersive Environments. 3rd International Workshop on Presence. Delft, Netherlands

Kim, H., Kwon, S., Heo, J., Lee, H. and M. K. Chung, M.K. (2014) ‘The effect of touch-key size on the usability of In-Vehicle Information Systems and driving safety during simulated driving’, *Applied Ergonomics*, 45(3), pp. 379-388.

Kitchin, J. and Baber, C. (2018) ‘Exploring workload and performance through the use of visual analytics’, *Ergonomics and Human Factors Conference,* Birmingham, UK: University of Birmingham

Klatt, J., Pitten, A., Broeke, S., McCall, R., Kramer, C., Wetzel, R., Blum, L., Oppermann, L. and Klatt, J. (2012) ‘Subjective and behavioral presence measurement and interactivity in the collaborative augmented reality game TimeWarp’, *Interacting with Computers*, 24(4), pp. 317-325.

Knight, J. and Baber, C. (2007) ‘Effect of Head-Mounted Displays on Posture’, *The Journal of Human Factors and Ergonomics Society,* 49, pp. 797-804.

Knight, J. and Stone, R.J. (2014) *An Overview of Human-Machine Interface Research Relevant to Virtual and Mixed Reality Environments for Command and Control.* Unpublished Technical Report for BAE Systems. Future Mission Systems 2035 Project. Birmingham, University of Birmingham

Lavrakas, P. J. (2008*). Encyclopedia of survey research methods.* Thousand Oaks, CA, USA: Sage Publications, Inc

Lewis, C. and Griffin M. (1980) ‘Predicting the effects of vibration frequency, and axis and seating conditions on the reading of numeric displays’, *Journal of Sound and Vibration,* 72(3), pp. 355-377.

Livingston, M., Rosenblum, L., Brown, D., Schmidt, G., Julier, S., Baillot, Y., Edward, J., Ai, Z. and Maassel, P. (2011) ‘Military Applications of Augmented Reality’ in Furht, B. (ed.) *Handbook of Augmented Reality.* New York, USA: Springer

Lombard, M. and Ditton, T. (2006) ‘At the Heart of It All: The Concept of Presence’, *Journal of Computer-Mediated Communication,* 3(2).

Macchiarella, N. and Vincenzi, D. (2004) ‘Augmented Reality in a Learning Paradigm for Flight and Aerospace Maintenance Training’, *Digital Avionics Systems Conference*, 23, pp. 1-9.

Macedonio, M., Parsons, T, Digiuseppe, R., Weiderhold, B. and Rizzo, A. (2007) ‘Immersiveness and physiological arousal within panoramic video-based virtual reality’, *Cyberpsychology and Behavior*, 10(4), pp. 508-515.

MacIntyre, B., Bolter, J. and Gandy, M (2004) “Presence and the Aura of Meaningful Places”, *7th Annual International Workshop on Presence (PRESENCE 2004),* Valencia, Spain: Polytechnic University of Valencia

Martin, B., Roll, J. and Di Renzo, N. (1991) ‘The interaction of hand vibration with oculomanual coordination in pursuit tracking’, *Aviation, Space, and Environmental Medicine,* 62(2), pp. 145-152.

Meehan, M., Insko, B., Whitton, M. and Brooks, F. (2002) ‘Physiological measures of presence in stressful virtu'al environments’, *ACM Transactions on Graphics (TOG).* New York, USA: ACM, pp. 645-652.

Meehan, M., Razzaque, S. and Whitton, M. and Brooks, F. (2003) ‘Effect of latency on presence in stressful virtual environments’, *Proceedings of the IEEE Virtual Reality*. Washington, USA: IEEE, pp. 141-148.

Melzer, J and Moffitt, K (1997) *Head-Mounted Displays: Designing for the User*. New York, US: McGraw-Hill

Mestre, D (2005) ‘Immersion and Presence’, *Movement and Perception.* Marseille: France: University of the Mediterranean, pp. 1-10

Milgram, P. and Colquhoun, H. (1999) ‘A taxonomy of real and virtual world display integration’ in Ohta, Y. and Tamura, H. (eds) *Mixed Reality: Merging Real and Virtual Worlds.* Berlin, Germany: Springer-Verlag, pp 5-30.

Milgram, P. and Kishino, F. (1994) ‘A taxonomy of mixed reality visual displays’, *IEICE Transactions on Information Systems*, 12, pp. 119–127.

Ministry of Defence (2008a) Defence Standard 00-250 Part 3 Technical Guidance, Section 9 People Characteristics. UK: Ministry of Defence

Ministry of Defence (2008b) Defence Standard 00-250 Part 3 Technical Guidance, Section 11 The Workplace. UK: Ministry of Defence

Ministry of Defence (2008c) Defence Standard 00-250 Part 3 Technical Guidance, Section 14 HFI Method, Tools and Techniques. UK: Ministry of Defence

Murata, A., Uetsugi, R. and Fukunaga, D (2004) 'Effects of target shape and display location on pointing performance by eye-gaze input system', *SICE Annual Conference,* Sapporo, Japan: IEEE, pp. 955-962.

Murata, A., Uetsugi, R. and Hayami, T (2012) 'Study on Cursor Shape suitable for Eye-gaze input system', *SICE Annual Conference,* Akita, Japan: IEEE. pp. 926-931.

Odom, J (2017) Royal Australian Air Force Using HoloLens to Experiment with Augmented Reality (Online). Available at: https://hololens.reality.news/news/royal-australian-air-force-using-hololens-experiment-with-augmented-reality-0175955/ (Accessed 14 April 2018)

Ordnance Survey (2018) Find the right OS data product for your business need (Online). Available at: https://www.ordnancesurvey.co.uk/business-and-government/products/finder.html (Accessed 3 January 2018)

Orit, S. and Eva, H. (2009) 'Tangible User Interfaces: Past, Present, and Future Directions', *Foundations and Trends in Human-Computer Interaction*, 3(2), pp. 1-137.

Oyekoya, O. and Stentiford, F (2005) 'A performance comparison of eye tracking and mouse interfaces in a target image identification task', *Integration of Knowledge, Semantics and Digital Media Technology EWIMT*. London, UK: IET, pp. 139-144.

Paelke, V., Nebe, K., Geiger, C., Klompmaker, F. and Fischer, H. (2012) 'Multi-modal, multi-touch interaction with maps in disaster management applications', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science,* 39, pp. 55-60.

Parhi, P., Karlson, A. and Benderson, B. (2006) 'Target size study for one-handed thumb use on small touchscreen devices*', Proceedings of 8th International Conference on Human Computer Interaction with Mobile Devices and Services*. Espoo, Finland: ACM, pp. 203-210.

Qian, C (2015) *Determining principles for the development of virtual environments for future clinical applications,* PhD thesis. Birmingam: University of Birmingham.

Rash, C., McLean, W., Mozo, B., Licina, J. and Mcentire, B. (1999) *Human Factors and Performance Concerns for the Design of Helmet-Mounted Displays*. USA: U.S. Army Aeromedical Research Laboratory

Rash, C., Russo, M., Letowski, T. and Schmeisser, E. (2009) *Helmet Mounted Displays: Sensation, Perception and Cognition Issues*. Fort Rucker, AL, USA: U.S. Army Aeromedical Research Laboratory

Regenbrecht, H. and Schubert, T. (2002) 'Measuring presence in augmented reality environments: design and a first test of a questionnaire', *Proceedings of the 5th Annual International Workshop Presence.* Porto, Portugal.

Regenbrecht, H., Schubert, T. and Friedmann, F. (1998) 'Measuring the sense of presence and its relations to fear of heights in virtual environments', *International Journal of Human-Computer Interaction,* 10(3), pp 233-249.

Roberts, D., Menozzia, A., Cooka, J., Sherrilla, T., Snarskia, S., Russlera, Pat., Clippa, B., Karla, R., Wengera, E., Bennetta, M., Maurgera, J., Churcha, W., Towlesa, H., MacCabea, S., Webba, J., Lupoa, J., Frahmb, J., Dunnb, E. and Leslieb, C. (2013) 'Testing and evaluation of a wearable augmented reality system for natural outdoor environments', *Defense, Security, and Sensing,* pp. 1-16

Rohs, M., Oulasvirta, A. and Suomalainen, T. (2011) 'Interaction with Magic Lenses: Real-World Validation of a Fitts ' Law Model', *Conference on Human Factors in Computing Systems – Proceedings.* Vancouver, Canada: CHI, pp. 2725-2728.

Royal Air Force (2019) RAF Recruitment | Pilot | Royal Air Force (Online). Available at https://www.raf.mod.uk/recruitment/roles/roles-finder/aircrew/pilot (Accessed 11 September 2019)

Sanchez-Vives, M. and Slater, M. (2005) 'From presence to consciousness through virtual reality', *Nature Reviews Neuroscience*, 6, pp. 332-339.

Sarwal, A., Baker, C. and Filipovic, D. (2005) 'Head-Worn Display-Based Augmented Reality Systems for Manufacturing', *The International Society for Optical Engineering*. San Jose, CA, USA: Springer, pp. 115-123.

Sauro, J. (2011) 'SUStisfied? Little-known System Usability Scale facts', *User Experience: The Magazine of the User Experience Professionals Association,* 10(3).

Sauro, J. and Lewis, J. (2016) *Quantifying the user experience: Practical statistics for user research*. San Francisco, CA, USA: Morgan Kaufmann

Savage-Knepshield, J., Martin, J., Lockett, J. and Allender, L. (2012) *Designing Soldier Systems: Current Issues in Human Factors.* London, UK: CRC Press

Schaik, P., Turnbull, T., Wersch, A. and Drummond, S. (2004) ‘Presence Within a Mixed Reality Environment’, *Cyberpsychology and Behaviour*, 7(5), pp. 540-552.

Schedlbauer, M. (2007) ‘Effects of key size and spacing on the completion time and accuracy of input tasks on soft keypads using trackball and touch input’, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting. SAGE Publications*, 41(1), pp. 429-433.

Schnabel, M,. Wang, X., Seichter, H. and Kvan, T (2007) ‘From Virtuality to Reality and Back’, *International Assosiation of Societies of Design Research.* Hong Kong: The Hong Kong Polytechnic University Hung Hom, pp. 2-15

Schubert, T., Friedmann, F. and Regenbrecht, H. (2001) ‘The experience of presence: Factor analytic insights’, *Presence: Teleoperators and Virtual Environments,* 10(3), pp. 266-281.

Scott, B. and Conzola, V. (1997) ‘Designing touch screen numeric keypads: Effects of finger size, key size, and key spacing’, *Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting.* Santa Monica, CA, USA: Human Factors and Ergonomics Society, pp. 360-364.

Sesto, M., Irwin, C., Chen, K., Chourasia, A. and Wiegmann, D. (2012) ‘Effect of Touch Screen Button Size and Spacing on Touch Characteristics of Users With and Without Disabilities’, *The Journal of the Human Factors and Ergonomics Society,* 54(3), pp 425-36.

Shoenberger, R. (1972) ‘Human response to whole body vibration’, *Perceptual and Motor Skills*, 34(1), pp 127-160.

Shooter, C (2017) *Virtual and Augmented Reality and Gamification as Training Tools in Safety Critical Industries*. UK: Rail Safety and Standards Board

Siltanen, S., Hakkarainen, M., Korkalo., O,m Salonen, T., Saaski, J., Woodward, C., Kannetis, T., Perakakis, M. and Potamianos, A. (2007) ‘Multimodal User Interface for Augmented Assembly’, *IEEE Workshop on Multimedia Signal Processing*, 9, pp. 78-81.

Sketchup (2018) 3D Warehouse (Online). Available at: https://3dwarehouse.sketchup.com/?hl=en (Accessed 3 January 2018)

Slater, M. and Usoh, M. (1994) ‘Representations Systems, Perceptual Position, and Presence in Immersive Virtual Environments’, *Presence Teleoperators and Virtual Environments,* 2(3), pp 221-233.

Slater, M. and Wilber, S. (1997) ‘A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments’, *Presence: Teleoperators and Virtual Environments,* 6(6), pp 603-616.

Smith, S. (2004) ‘Cockpit seat and pilot helmet vibration during flight operations on aircraft carriers’, *Aviation Space Environment Medicine,* 75, pp. 247–254.

So, R. and Griffin, M. (2000) ‘Effects of target movement direction cue on head-tracking performance’, Ergonomics, 43(3), pp. 360–376.

Stanney, K., Mourant, R. and Kennedy, R. (1998) ‘Human Factors Issues in Virtual Environments: A Review of the Literature’, *Presence: Teleoperators and Virtual Environments,* 7(4), pp. 327-351.

Steptoe, W. (2015) What is Presence in Immersive Augmented Reality? (Online). Available at: http://willsteptoe.com/post/91843919913/what-is-presence-in-immersive-augmented-reality (Accessed 12 February 2018)

Strickland, J. (2007) How Virtual Reality Gear Works (Online). Available at: https://electronics.howstuffworks.com/gadgets/other-gadgets/VR-gear.html (Accessed 25 September 2019)

Stone, R.J. (2012) *Human Factors Guidance for Designers of Interactive 3D and Games-Based Training Systems.* 2nd ed. UK: Human Factors Integration Defence Technology Centre Publication

Stone, R.J. (2016a) 'Getting VR Right Then and Now … The Indispensable Role of Human Factors and Human Centred Design', *Presence: Virtual and Augmented Reality*, 25(2), pp. 151-160.

Stone, R.J. (2016b) ‘Blending the Best of the Real with the Best of the Virtual: Mixed Reality Case Studies in Healthcare and Defence’ in Jung T. and Dieck, C. (eds) *Augmented Reality and Virtual Reality*. Cham, Switzerland: Springer, pp. 277-293.

Sylaiou, S., Mania, K., Karloulis, A. and White, M. (2010) ‘Exploring the relationship between presence and enjoyment in a virtual museum’, *International journal of human-computer studies,* 68(5), pp. 243-253.

Tao, D., Chen, Q., Yuan, J., Liu, S., Zhang, X. and Qu, X. (2017) ‘Effects of Key Size, Gap and the Location of Key Characters on the Usability of Touchscreen Devices in Input Tasks’, *International Conference on Engineering Psychology and Cognitive Ergonomics*. Cham, Switzerland: Springer, pp. 133-144.

Tauson, R. (2012) ‘Operations on the move: Vehicle movement and soldier performance.’ In Savage-Knepshield, J., Martin, J., Lockett, J. and Allender, L. (2012) *Designing Soldier Systems: Current Issues in Human Factors.* London, UK: CRC Press, p 119-133.

Thingiverse (2018) Digital Designs for Physical Objects (Online). Available at: https://www.thingiverse.com/ (Accessed 3 January 2018)

Toczek, Y. (2016) *The influence of visual realism on the sense of presence in virtual environments,* PhD thesis. Eindhoven, German: Eindhoven University of Technology

Trujillo, A.C. (2011) Evaluation of Electronic Formats of the NASA Task Load Index. Virginia, USA: Langley Research Centre

Tullis, T. and Albert, B. (2008) *Measuring the user experience.* San Francisco, CA, USA: Morgan Kaif,amm Publishers, Inc

Tung, K., Miller, M. and Colombi, J., Urbide, D. and Smith, S. (2014) ‘Effect of vibration on eye, head and helmet movements while wearing a helmet-mounted display’ *Journal of the Society for Information Display*, 22(10), pp. 535-544.

TurboSquid (2018) 3D Models for Professionals (Online). Available at: https://www.turbosquid.com/ (Accessed 3 January 2018)

Uribe, D. and Miller, M. (2013) ‘Eye Movements When Viewing a HMD Under Vibration’, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), pp. 1139-1143.

Usoh, M., Catena, E., Arman, S. and Slater, M. (2000) ‘Using presence questionnaires in reality’, *Presence Teleoperators and Virtual Environments*, 9(5), pp. 497-503.

Vajak, D. and Livada, C. (2017) ‘Combining photogrammetry, 3D modeling and real time information gathering for highly immersive VR experience’, *Zooming Innovation in Consumer Electronics International Conference (ZINC).* Nova Sad, Serbia: IEEE, pp. 82-85.

Valve (2017) SteamVR/Environments/Photogrammetry - Valve Developer Community (Online). Available at: https://developer.valvesoftware.com/wiki/SteamVR/Environments/Photogrammetry (Accessed 5 January 2018)

Velger, M. (1997) Helmet-mounted displays and *sights*. Norwood, MA, USA: Artech House Publishers

Wang, L., Cao, Q., Chang, J. and Zhao, C. (2015) ‘The Effect of Touch-key Size and Shape on the Usability of Flight Deck MCDU’, *The Eighth International Conference on Advances in Computer-Human Interactions, 41(1),* pp. 234-238.

Wired (2017) The Untold Story of Magic Leap, the World’s Most Secretive Startup (Online). Available at: https://www.wired.com/2016/04/magic-leap-vr/ (Accessed 28 December 2017)

Witmer, B.G. and Singer, M.J. (1994) *Measuring Presence in Virtual Environments, ARI Technical Report.* Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences

Witmer, B.G. and Singer, M.J. (1998) ‘Measuring Presence in Virtual Environments: A Presence Questionnaire’, *Presence: Teleoperators and Virtual Environments*, 7(3), pp. 225-240.

Wu, S. (2013), ‘A review on coarse warranty data and analysis’, *Reliability Engineering and System,* 114, pp. 1–11.

Yu, F., Lu, Z., Lou, H. and Wang, P. (2011) *Three-dimensional model analysis and processing.* Berlin, Germany: Springer-Verlag

Zhang, X. and MacKenzie, I. (2007) ‘Evaluating eye tracking with ISO9241 - Part 9’, *Proceedings of HCII 2007. Lecture Notes in Computer Science*, 4552, pp. 779–788.

Zuckerman, O. and Gal-Oz, A. (2013) ‘To TUI or not to TUI: Evaluating Performance and Preference in Tangible vs. Graphical User Interfaces’, *International Journal of Human-Computer Studies*, 71(7), pp. 803-820.

APPENDIX

Appendix A Experiment 1 Information Sheet

Experiment Information Sheet

A Human Factors evaluation of a Mixed Reality display technologies

Overview: The aim of the experiment is to assess performance and general human factors issues arising when interacting with and manipulating targets using a variety of Human-Machine Interface (HMI) technologies within a constrained-space, cockpit-like environment. Participant will perform a series of simple interaction tasks in which performance data will be collected, including completion time, and error rate. Following each task, a series of questionnaires will be completed.

The 3 technologies being used include head-tracking (via a Virtual Reality Head Mounted Display), a touchscreen and a joystick. Interaction tasks include selecting, resizing, repositioning and zooming in on objects on displays.

Please complete the tasks as quickly and accurately as you can.

Risks: The experiment includes using a Head-Mounted Display (HMD) which can cause discomfort after prolonged use. If you feel nausea, headaches, or sick then stop using the system immediately. Health & Safety guidelines for the use of HMD's are provided.

Time: The experiment will take a total of 1-2 hours.

Data Collection: If you choose to withdraw before the end of the experiment all collected data will be destroyed. All data will be anonymised and stored on the University IT system and used solely for the purposes of this PhD research. The anonymised data may be published in academic journals or conferences. Further details and copies of published results will be provided on request.

Contact Details:

Name: Christopher Bibb

Email: cxb978@bham.ac.uk

Appendix B HMD Health & Safety Guidelines (Stone, 2012)

Health & Safety Guidelines
Virtual Reality Head-Mounted Displays (HMDs)
GUIDANCE NOTE 1

CHECK BEFORE USE

Participants reporting/showing any signs of the following should NOT take part in HMD-based trials

NB. The HMD shown above – the Rockwell Collins *ProView XL-50* is now an old and discontinued product (however, see Case Studies 6 and 20) – the HMD is presented here for general illustration only

Conjunctivitis	Pregnancy (Heavy or 5 months+)
Corneal Ulcers	Ear infections/ear disease
Corneal Infections	Influenza
"Dry Eye"	Head Colds
Iritis	Respiratory Ailments
Cataracts or Glaucoma	"Heavy" Hangover

Participants reporting/showing any signs of the following should be OBSERVED CLOSELY whilst taking part in HMD-based trials and should be debriefed after trials to ascertain their state of health

Extreme Fatigue	Digestive Problems
Significant Sleep Loss	Emotional Stress
Mild Hangover	Anxiety

ENSURE THAT THE HMD IS ALIGNED STRAIGHT ON THE USER'S HEAD (SEE PICTURE ABOVE), TO AVOID ANY VISUAL DISTURBANCES CAUSED BY THE TRACKING SYSTEM RECORDING ABNORMAL ELEVATION ANGLES AT START-UP

CHECK DURING USE

OBSERVE PARTICIPANTS AT REGULAR PERIODS DURING EACH TRIAL AND TAKE IMMEDIATE ACTION (HALT THE SIMULATION AND STAND EASY) IF THEY REPORT SYMPTOMS SUCH AS DISORIENTATION, NAUSEA, EYESTRAIN OR ANY FORM OF MALAISE

CHECK AFTER USE

- Does the participant show any signs of disorientation?
- Does the participant show any signs of nausea or malaise?
- Does the participant show any signs of eyestrain?
- Does the participant show any signs of unstable posture? If unsure, test participant - walking a straight line with eyes closed and arms folded

IF THE ANSWER TO ANY OF THESE IS "YES" THEN INSTRUCT THE PARTICIPANT TO STAND DOWN AND RELAX. DO NOT ALLOW THE PARTICIPANT TO OPERATE MACHINERY OR DRIVE FOR 60 MINUTES

IF IN DOUBT – ASK!

Appendix C Experiment Consent Form

ID #___

Consent Form

PhD Research Experiment

	Please tick to confirm
I am over 18	
I have read and understood the information sheet	
I have been given the opportunity to ask questions about the study	
I understand that I have the right to withdraw from the experiment at any time	
I agree to report any discomfort that might arise from using the system	
I agree to the anonymous data being collected from me to be used for research purposes	
I understand that all data recorded will be anonymised and will not be linked to me	
I agree to take part in the above-mentioned experiment	

Name………………………...

Signature…………………………………………… Date ………………………

Appendix D Experiment 1 Participant Data Questionnaire

ID #___

Experiment User Data Questionnaire

A Human Factors evaluation of

Mixed Reality Interface and Display Technologies

Age ____________

Gender ____________

Height ____________

Occupation ____

1. Do you play computer/video games? *(Please tick one option)*

❑ Yes ❑ No (if no go to question 9)

2. How long have you been playing computer games?

_________ years _________ months (if under 1 year)

3. How often do you play computer/video games?

❑ Every day ❑ 1 – 2 times per month

❑ 5 – 6 times per week ❑ Once every two months

❑ 3 – 4 times per week ❑ Once every six months

❑ 1 – 2 times per week ❑ Less than once per year

4. How many hours do you play computer/video games each time you play?

________ hours per session

5. On which of the following do you play computer games? *(Select all that apply)*

- ❑ PC
- ❑ Games console eg Xbox and Playstation
- ❑ Handheld e.g. PSP
- ❑ PDA
- ❑ Mobile phone
- ❑ Interactive TV
- ❑ Arcade
- ❑ Other (Please specify)

6. What type of computer games do you play? *(Select all that apply)*

Before you start, please turn to page 6 for game definitions

- ❑ First Person Shooter
- ❑ Fighting
- ❑ Role Playing Games
- ❑ Vehicle Simulations
- ❑ Other (Please specify)
- ❑ Real Time Strategy
- ❑ Sports
- ❑ Arcade
- ❑ Puzzles

7. What features are important to you when you play a video game?

(Select all that apply)

- ❑ Graphics
- ❑ Music/Soundtrack
- ❑ Storyline
- ❑ Realism
- ❑ Fantasy/make-believe
- ❑ Having clear rules (what you can/can't
- ❑ Having clear goals/objectives
- ❑ Immediate feedback
- ❑ Humour
- ❑ Other (Please specify)
- ❑ Rewards/penalties
- ❑ Simple to play (easy to pick up)
- ❑ Hard to play (requires practice)
- ❑ Variety of levels/progressive challenge
- ❑ Chance to win
- ❑ Control (being in control)
- ❑ Problem-solving activity
- ❑ Challenge
- ❑ Competition/contest

8. In which of the following environments do you play? (*Select all that apply)*

- ☐ Single player
- ☐ Dual player, on the same computer
- ☐ More than two on the same computer
- ☐ Other (Please specify)
- ☐ Multiplayer locally networked
- ☐ Multiplayer Internet games
- ☐ MMPGs (Massively Mulitplayer (Online)

9. What would be your preferred methods to choose to learn a new task? *(Select one option)*

- ☐ Listening to someone explain how to do it ("auditory")
- ☐ Watching someone do it ("visual")
- ☐ By trying it yourself ("kinaesthetic")

10. How confident do you feel using new technologies and devices*? (Please tick one option)*

Not confident	○	○	○	○	○	○	○	Very confident
	-3	-2	-1	0	1	2	3	
				Moderately confident				

11. How confident do you feel using new software systems*? (Please tick one option)*

Not confident	○	○	○	○	○	○	○	Very confident
	-3	-2	-1	0	1	2	3	
				Moderately confident				

12. How confident do you feel using new system interaction techniques (e.g. gestures and touch instead of keyboard input)*? (Please tick one option)*

Not confident	○	○	○	○	○	○	○	Very confident
	-3	-2	-1	0	1	2	3	
				Moderately confident				

13. How much experience do you have with Touchscreen displays (desktop, not mobile)? *(Please select one option)*

- ❑ Never used
- ❑ Used once (e.g. a demo)
- ❑ Used on several different occasions
- ❑ Used many times
- ❑ Use on a regular basis

14. How much experience do you have with Eye Tracking technologies? *(Please select one option)*

- ❑ Never used
- ❑ Used once (e.g. a demo)
- ❑ Used on several different occasions
- ❑ Used many times
- ❑ Use on a regular basis

15. How much experience do you have with VR/AR Head Mounted Displays? *(Please select one option)*

- ❑ Never used
- ❑ Used once (e.g. a demo)
- ❑ Used on several different occasions
- ❑ Used many times
- ❑ Use on a regular basis

16. How much experience do you have with non-VR/AR Head Tracking technologies? *(Please select one option)*

- ❑ Never used
- ❑ Used once (e.g. a demo)
- ❑ Used on several different occasions
- ❑ Used many times
- ❑ Use on a regular basis

17. How much experience do you have with HOTAS (Joystick and Throttle)? *(Please select one option)*

- ❑ Never used
- ❑ Used once (e.g. a demo)
- ❑ Used on several different occasions
- ❑ Used many times
- ❑ Use on a regular basis

18. How much experience do you have with Flight Simulators? *(Please select one option)*

- ❑ Never used
- ❑ Used once (e.g. a demo)
- ❑ Used on several different occasions
- ❑ Used many times
- ❑ Use on a regular basis

19. How much experience do you have with cockpit interfaces and layouts? *(Please select one option)*

- ❑ Never used
- ❑ Used once (e.g. a demo)
- ❑ Used on several different occasions
- ❑ Used many times
- ❑ Use on a regular basis

20. How knowledgeable are you about aircraft cockpit interfaces and layouts? *(Please select one option)*

- ❑ Novice - minimal or 'textbook' knowledge
- ❑ Beginner - working knowledge of key aspects of subject, limited experience using techniques or concepts
- ❑ Competent - good working and background knowledge of subject area, comfortable using techniques or concepts
- ❑ Proficient - depth of understanding subject area, broad experience using techniques or concepts
- ❑ Expert - authoritative knowledge and deep tacit understanding across subject area, advanced experience using techniques or concepts

21. How knowledgeable are you about User Interface design methods and concepts? *(Please select one option)*

- ❑ Novice - minimal or 'textbook' knowledge
- ❑ Beginner - working knowledge of key aspects of subject, limited experience using techniques or concepts
- ❑ Competent - good working and background knowledge of subject area, comfortable using techniques or concepts
- ❑ Proficient - depth of understanding subject area, broad experience using techniques or concepts
- ❑ Expert - authoritative knowledge and deep tacit understanding across subject area, advanced experience using techniques or concepts

Game Definitions

First Person Shooter	Three-dimensional shooter games with a first person perspective (i.e. looking down the barrel of a gun, as in *FarCry*, *Half-Life 2*, etc.)
Fighting	Games which simulate hand-to-hand combat, usually between pairs of fighters, modelled on Asian martial arts techniques (e.g. *Mortal Combat*).
Real Time Strategy	Games which allow the player to command some type of operation, typically a military operation, involving the player in planning a series of actions and managing resources to build or expand a community, army or empire (e.g. *Civilisation*).
Vehicle Simulators	Simulations which create the feeling of driving or flying a vehicle (either real or imaginary) in a realistic situation (e.g. Microsoft's *Flight Simulator*).
Sports	Games which simulate some aspect of a real or imaginary athletic sport (e.g. *Winning Eleven*).
Role Playing	With role-playing games, players manage either a person or a team through a series of quests, in a fantasy or science fiction setting, building the character's power, and abilities and inventories to meet increasing and evolving conflicts (e.g. *Dungeons and Dragons*, *EverQuest*).
Arcade	Often coin-operated entertainment machines, often installed in pubs and video arcade (e.g. *Space Invaders*).
Puzzles	Games in which puzzle solving is the primary activity (e.g. *Tetris*).

Appendix E Latin Square Ordering

A Touchscreen

B HOTAS

C Mixed Reality system + HOTAS

Participant #	Task Order		
1	A	C	B
2	B	C	A
3	A	B	C
4	C	B	A
5	C	A	B
6	B	C	A
7	B	A	A
8	A	C	B
9	A	C	B
10	C	B	A
11	B	A	C
12	A	C	B
13	B	A	C
14	B	C	A
15	A	C	B
16	B	C	A
17	A	B	C
18	C	B	A
19	A	C	B
20	B	C	A

Appendix F Exertion Rating

BORG Rating of Perceived Exertion (RPE) Scale (Borg, 1982)

#	Level of Exertion
6	No exertion at all
7	
7.5	Extremely light (7.5)
8	
9	Very light
10	
11	Light
12	
13	Somewhat hard
14	
15	Hard (heavy)
16	
17	Very hard
18	
19	Extremely hard
20	Maximal exertion

Appendix G Subjective Discomfort Rating

Subjective Discomfort Scale (modified from Harich, 2002)

Minor Able to adapt to pain	1 - Very Mild	Very light barely noticeable discomfort
	2 - Discomforting	Minor discomfort, like lightly pinching
	3 - Tolerable	Very noticeable discomfort, like an injection
Moderate Interferes with many activities	4 - Distressing	Strong, deep discomfort like a toothache
	5 - Very Distressing	Strong, deep, piercing discomfort, such as a sprained ankle
	6 - Intense	Strong, deep piercing discomfort like several bee stings
Severe Unable to function properly	7 - Very Intense	Comparable to an average migraine headache
	8 - Utterly Horrible	Comparable to very bad migraine headache
	9 - Unbearable	Intense pain you cannot tolerate
	10 - Unspeakable	Maximum pain threshold that you cannot carry on

Appendix H Usability Questionnaire

System Usability Scale from Brooke (1996)

	Strongly disagree				Strongly agree
1. I think that I would like to use this system frequently	1	2	3	4	5
2. I found the system unnecessarily complex	1	2	3	4	5
3. I thought the system was easy to use	1	2	3	4	5
4. I think that I would need the support of a technical person to be able to use this system	1	2	3	4	5
5. I found the various functions in this system were well integrated	1	2	3	4	5
6. I thought there was too much inconsistency in this system	1	2	3	4	5
7. I would imagine that most people would learn to use this system very quickly	1	2	3	4	5
8. I found the system very cumbersome to use	1	2	3	4	5
9. I felt very confident using the system	1	2	3	4	5
10. I needed to learn a lot of things before I could get going with this system	1	2	3	4	5

Appendix I Work Load Questionnaire

NASA TLX (Hart and Staveland, 1988)

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Appendix J Experiment 2 Information Sheet

Experiment Information Sheet

A Human Factors evaluation of a Mixed Reality display technologies

Overview: The aim of the experiment is to assess a variety of user centred metrics when using an interactive MR 'tabletop' system and wearing a HMD. A procedure is given which involves selecting, rotating, resizing and repositioning virtual objects. Following the procedure a series of questionnaires will be completed.

Please complete the tasks as accurately as you can.

Risks: The experiment includes using a Head-Mounted Display (HMD) which can cause discomfort after prolonged use. If you feel nausea, headaches, or sick then stop using the system immediately. Health & Safety guidelines for the use of HMD's are provided.

Time: The experiment will take a total of 1 hour.

Data Collection: If you choose to withdraw before the end of the experiment all collected data will be destroyed. All data will be anonymised and stored on the University IT system and used solely for the purposes of this PhD research experiment. The anonymised data may be published in academic journals or conferences. Further details and copies of published results will be provided on request.

Contact Details:

Name: Christopher Bibb

Email: cxb978@bham.ac.uk

Appendix K Experiment 3 Information Sheet

Experiment Information Sheet

A Human Factors evaluation of a Mixed Reality display technologies

Overview: The aim of the experiment is to assess performance and mental workload when interacting with a virtual display using head- and eye-based input within a HMD while subject to a variety of vibration frequencies. Each vibration frequency continues for a period of 20 seconds, followed by participants audibly rating how hard it is to accurately focus during that specific vibration frequency.

Please keep the cursor over the “focus point” as accurately as you can.

Risks: The experiment includes using a Head-Mounted Display (HMD) which can cause discomfort after prolonged use. If you feel nausea, headaches, or sick then stop using the system immediately. Health & Safety guidelines for the use of HMD’s are provided.

Time: The experiment will take a total of 1 hour.

Data Collection: If you choose to withdraw before the end of the experiment all collected data will be destroyed. All data will be anonymised and stored on the University IT system and used solely for the purposes of this PhD research experiment. The anonymised data may be published in academic journals or conferences. Further details and copies of published results will be provided on request.

Contact Details:

Name: Christopher Bibb

Email: cxb978@bham.ac.uk

Appendix L iGroup Presence Questionnaire (IPQ) items

Key: G = General Presence; SP = Spatial Presence; INV = Involvement; REAL = Experienced Realism.

IPQ Name	English Question	English Anchors	Copyright (Item Source)
G1	In the computer-generated world, I had a sense of 'being there'.	Not at all—very much.	Slater and Usoh (1994)
SP1	Somehow, I felt that the virtual world surrounded me.	Fully disagree—fully agree.	IPQ
SP2	I felt like I was just perceiving pictures.	Fully disagree—fully agree.	IPQ
SP3	I did not feel present in the virtual space.	Did not feel—felt present.	Not Stated
SP4	I had a sense of acting in the virtual space, rather than operating something from outside.	Fully disagree—fully agree.	IPQ
SP5	I felt present in the virtual space.	Fully disagree—fully agree.	IPQ
INV1	How aware were you of the real-world surroundings while navigating in the virtual world (e.g., sounds, room temperature, and other people)?	Extremely aware/moderately aware/not aware at all.	Witmer and Singer (1994)
INV2	I was not aware of my real environment.	Fully disagree—fully agree.	IPQ
INV3	I still paid attention to the real environment.	Fully disagree—fully agree.	IPQ
INV4	I was completely captivated by the virtual world.	Fully disagree—fully agree.	IPQ
REAL1	How real did the virtual world seem to you?	Completely real—not real at all.	Hendrix (1994)
REAL2	How much did your experience in the virtual environment seem consistent with your real-world experience?	Not consistent/moderately consistent/very consistent.	Witmer and Singer (1994)
REAL3	How real did the virtual world seem to you?	About as real as an imagined world—indistinguishable from the real world.	Carlin, Hoffman, and Weghorst (1997)
REAL4	The virtual world seemed more realistic than the real world.	Fully disagree—fully agree.	IPQ

Appendix M iGroup Presence Questionnaire (IPQ) Questions

How aware were you of the real world surrounding while navigating in the virtual world? (i.e. sounds, room temperature, other people, etc.)?

extremely aware	○	○	○	○	○	○	○	not aware at all
	-3	-2	-1	0	+1	+2	+3	
				moderately aware				

64/inv1/0

How real did the virtual world seem to you?

completely real	○	○	○	○	○	○	○	not real at all
	-3	-2	-1	0	+1	+2	+3	

48/real1/1

I had a sense of acting in the virtual space, rather than operating something from outside.

fully disagree	○	○	○	○	○	○	○	fully agree
	-3	-2	-1	0	+1	+2	+3	

31/sp4/2

How much did your experience in the virtual environment seem consistent with your real world experience?

not consistent	○	○	○	○	○	○	○	very consistent
	-3	-2	-1	0	+1	+2	+3	
				moderately consistent				

7/real2/3

How real did the virtual world seem to you?

about as real as an imagined world	○	○	○	○	○	○	○	indistinguishable from the real world
	-3	-2	-1	0	+1	+2	+3	

59/real3/4

I did not feel present in the virtual space.

did not feel	○	○	○	○	○	○	○	felt present
	-3	-2	-1	0	+1	+2	+3	28/sp3/5

I was not aware of my real environment.

fully disagree	○	○	○	○	○	○	○	fully agree
	-3	-2	-1	0	+1	+2	+3	37/inv2/6

In the computer generated world I had a sense of "being there"

not at all	○	○	○	○	○	○	○	very much
	-3	-2	-1	0	+1	+2	+3	62/g1/7

Somehow I felt that the virtual world surrounded me.

fully disagree	○	○	○	○	○	○	○	fully agree
	-3	-2	-1	0	+1	+2	+3	44/sp1/8

I felt present in the virtual space.

fully disagree	○	○	○	○	○	○	○	fully agree
	-3	-2	-1	0	+1	+2	+3	33/sp5/9

I still paid attention to the real environment.

fully disagree	○	○	○	○	○	○	○	fully agree
	-3	-2	-1	0	+1	+2	+3	40/inv3/10

The virtual world seemed more realistic than the real world.

fully disagree	○	○	○	○	○	○	○	fully agree
	-3	-2	-1	0	+1	+2	+3	47/real4/11

I felt like I was just perceiving pictures.

fully disagree	○	○	○	○	○	○	○	fully agree
	-3	-2	-1	0	+1	+2	+3	

30/sp2/12

I was completely captivated by the virtual world.

fully disagree	○	○	○	○	○	○	○	fully agree
	-3	-2	-1	0	+1	+2	+3	