LOUGHBOROUGH UNIVERSITY

Augmenting Low-Fidelity Flight Simulation Training Devices via Amplified Head Rotations

by

Luan Le Ngoc

A Doctoral Thesis

Submitted in partial fulfillment of the requirements for the award of

Doctor of Philosophy of Loughborough University

December 2013

ABSTRACT

Due to economic and operational constraints, there is an increasing demand from aviation operators and training manufacturers to extract maximum training usage from the lower fidelity suite of flight simulators. It is possible to augment low-fidelity flight simulators to achieve equivalent performance compared to high-fidelity setups but at reduced cost and greater mobility. In particular for visual manoeuvres, the virtual reality technique of head-tracking amplification for virtual view control enables full field-of-regard access even with limited field-of-view displays. This research quantified the effects of this technique on piloting performance, workload and simulator sickness by applying it to a fixed-base, low-fidelity, low-cost flight simulator. In two separate simulator trials, participants had to land a simulated aircraft from a visual traffic circuit pattern whilst scanning for airborne traffic.

Initially, a single augmented display was compared to the common triple display setup in front of the pilot. Starting from the base leg, pilots exhibited tighter turns closer to the desired ground track and were more actively conducting visual scans using the augmented display. This was followed up by a second experiment to quantify the scalability of augmentation towards larger displays and field of views. Task complexity was increased by starting the traffic pattern from the downwind leg. Triple displays in front of the pilot yielded the best compromise delivering flight performance and traffic detection scores just below the triple projectors but without an increase in track deviations and the pilots were also less prone to simulator sickness symptoms.

This research demonstrated that head augmentation yields clear benefits of quick user adaptation, low-cost, ease of systems integration, together with the capability to negate the impact of display sizes yet without incurring significant penalties in workload and incurring simulator sickness. The impact of this research is that it facilitates future flight training solutions using this augmentation technique to meet budgetary and mobility requirements. This enables deployment of simulators in large numbers to deliver expanded mission rehearsal previously unattainable within this class of low-fidelity simulators, and with no restrictions for transfer to other training media.

Acknowledgements

During about three years of research at the Advanced VR Research Centre, there are many people I would like to acknowledge and express gratitude in making this journey a swift and memorable experience.

First and foremost, I'd like to thank my parents and my sister for their unwavering support. This provided inspiration and motivation to see this project to its completion.

I'd like to thank Professor Roy Kalawsky, for his knowledge and wisdom have been inspirational, and his patience, kindness, understanding and enthusiasm a catalyst for this research right from the start. Giving me huge blocks of airspace to freely roam the skies really enabled me to perform to my maximum ability. I'd also like to acknowledge the unsung miracle worker of the department, Sharon Want for expediting 'the system' and keeping the gorilla's off my back.

The experiments were the most fun part of this research, so I would like to sincerely thank all the participants in this study. Their enthusiasm and feedback was a very rewarding experience after all the hard setup work. I'm in particular hugely indebted to Johan Kjölhede for his help debugging simulation code through the abyss of syntax errors in the twilight hours.

Sincere gratitude to Mike Southworth for being the industry champion and the Royal Aeronautical Society for supporting my work.

This enjoyable journey could not have been without my honorable colleagues, who also helped me build the simulator, Craig Wright and George Bedford, the latter solely responsible for turning me into a turbo petrolhead tanker. Special mention goes to Dr Petri Vitiello, the stats guru, on our many ramblings about life, the universe and the perils of doing a PhD.

A shout goes out to my bluegrass band: thanks for lighting my creative fire and giving me a spot in the limelight.

Finally, having never heard of this town before crossing the Channel, I appreciate Loughborough University for its generosity in granting me the Departmental PhD Scholarship Award in 2010 to fund this research.

This thesis is dedicated in spirit to the memories of my grandparents.

Contents

List of Figures

List of Tables

List of Abbreviations

AGL	Above Ground Level
AIAA	American Institute of Aeronautics and Astronautics
ANOVA	Analysis of Variance
ATC	Air Traffic Control
ATD	Average Turn Distance
AVRRC	Advanced VR Research Centre
BFM	Basic Fighter Manoeuvres
CAA	Civil Aviation Authority
CPU	Central Processing Unit
CTER	Cumulative Transfer Effectiveness Ratio
DISP	Display
DME	Distance Measuring Equipment
DMO	Distributed Mission Operations
DoD	Department of Defense
DOF	Degrees of Freedom
EFVS	Enhanced Flight Visual System
EGNX	East Midlands Airport ICAO code
EGPWS	Enhanced Ground Proximity Warning System
FAA	Federal Aviation Authority
FBS	Functional Breakdown Structure
FFD	Functional Flow Diagram
FOR	Field-of-Regard
FOV	Field-of-View
FSTD	Flight Simulation Training Device
GFOV	Geometric Field-of-View
GPS	Global Positioning System
GPU	Graphics Processing Unit
HDD	Hard Disk Drive
HFOV	Horizontal Field-of-View

HMD	Head-Mounted Display
HOTAS	Hands on Throttle and Stick
HUD	Heads-Up Display
ICAO	International Civil Aviation Organization
IEEE	Institute of Electrical and Electronic Engineers
ILS	Instrument Landing System
INCOSE	International Council on Systems Engineering
ITER	Incremental Transfer Effectiveness Ratio
IWG	International Working Group
LCD	Liquid Crystal Display
MANOVA	Multivariate Analysis of Variance
MoD	Ministry of Defence
NAA	National Aviation Authority
NASA	National Aeronautics and Space Administration
NDB	Non-Directional Beacon
OKR	Opto-Kinetic Reflex
PhD	Doctor of Philosophy
PPL	Private Pilot Licence
PRO	Triple Projectors
PSU	Power Supply Unit
QTG	Qualification Test Guide
RAeS	Royal Aeronautical Society
RMS	Root-Mean-Square
SEP	Systems Engineering Process
SGL	Single Monitor
SSD	Solid State Drive
SSQ	Simulator Sickness Questionnaire
SVS	Synthetic Vision System
SWAT	Subjective Workload Assessment Technique
TEE	Transfer Effectiveness Evaluation
TER	Transfer Effectiveness Evaluation
TLX	Task Load Index
TRL	Technology Readiness Level
TRP	Triple Monitor
UAS	Unmanned Aerial System
USD	United States Dollar
VE	Virtual Environments
VEC	Virtual Engineering Centre

VFOV	Vertical Field-of-View
VFR	Visual Flight Rules
VHF	Very High Frequency
VOR	VHF Omnidirectional Radio Range
VOR	Vestibular-Ocular Reflex
VR	Virtual Reality
XHSI	eXternal Highfidelity Simulator Instruments

Chapter 1

Introduction

This chapter describes the background motivation, defines the problem and transforms it into research goals and objectives composing the framework of this thesis. Finally, a thesis outline reveals the structure of the thesis and highlights the iterative process required to tackle complex systems problems in this study.

1.1 Background and Motivation

FLIGHT simulation has been an essential part of aerospace research and development since the dawn of flight. It also stands synonymous today with the important task of training aircrew. With progress in technology, simulators now encompass part-task trainers to high-fidelity full-mission simulators offering a wide range of training options in curriculum and fidelity levels. For training purposes, major advantages of using flight simulators are: increased efficiency, increased safety, lower overall training costs, practise rare real-world situations and the reduction in operational and environmental disturbance [1, 2].

In flight operations, visual cues are paramount to certain flying tasks and to maintain situational awareness. In 2010, the majority of general aviation accidents in the United States involved personal flying in fixed-wing aircraft [3]. The majority of these personal flying accidents occurred during the landing phase [3], despite its relative short duration as part of the total flight. This may be caused by the increased workload on flight crew and aircraft during takeoff and landing. The flight crew has many simultaneous tasks during this critical flight phase: control the aircraft, change altitude and speed, communicate with air traffic control (ATC) and/or other aircraft, and maintain separation from obstacles and other aircraft. This includes the carrying out of many manual operations such as changes to engine power settings, the possible operation of retractable landing gear, flaps, etc.

While the aircraft is at low altitude during takeoff and landing, it is also most susceptible to hazards caused by wind and weather conditions. In Australia from 1961 to 2003, mid-air collisions accounted for about 3% of fatal accidents involving general aviation aircraft. Most of these (78%) occurred in or near the traffic circuit area around the airport [4]. This reflects the higher traffic density in this area. Statistics further showed that a high proportion of the collisions (35%) occurred on final approach or the base-to-final turn [4]. These data indicate the importance of training in these crucial flight phases and is one of the reasons why this topic was chosen for study.

1.1.1 Flight Training

Simulation technology allows military pilots to participate in a continuous training cycle and maintain a high state of combat readiness by using cost-effective simulation alternatives in conjunction with live-fly operations and training missions [5, 6]. The current goal for air forces is to transition from frequency-based training systems using particular simulators for each training phase to a competency-based training system [7, 8]. This will eventually lead to a mixed usage of both simulation training solutions and real aircraft at the same time.

For air combat, many of the competency gaps revolve around higher order tasks or skills that can be gained from more complex experiences (e.g. team work, multi-team operations, complex tactical manoeuvres, etc.) [9–12]. Distributed Mission Operations (DMO) training, especially those using networked simulators, is often mentioned as a viable training medium for fulfilling many skill and experiential deficiencies [13, 14]. The usage of commercial-off-the-shelf technologies through innovative integration can therefore potentially provide affordable training solutions [15].

Until the late 1990s, pilots received training on complex tactical missions almost exclusively during infrequent, larger-scale, live-range exercises. Technological advances have made virtual environments (VE) more and more realistic and flexible enough to support the relatively new concept of DMO. In contrast to dedicated simulators built only to train for emergency procedures or routine tasks, DMOs networked environments enable frequent training of higher order individual and team-oriented skills which require more flexibility in simulator functionality [9].

DMO combines live (i.e. aircraft flying on a range), virtual (human-in-the-loop), and constructive (computer-generated models) assets to form a synthetic battlespace, shown in Figure 1.1. Table 1.1 further shows the training levels designed to enhance skills ranging from individual to full mission rehearsal as members of a team [16]. Effective application of multi-player simulation has been demonstrated for F-15 pilots [17], F-16 pilots [18], Tornado pilots and navigators [19], forward air controllers, and ground forces executing close air support [20], and Air Force Special Operations teams [21].

F-16 pilots who have flown in a distributed environment have rated it as particularly effective for training a four-ship formation against multiple enemy aircraft [23]. F-16 pilots have identified individual skills including radar mechanization (i.e., using the various modes and capabilities of the air-to-air radar to detect,

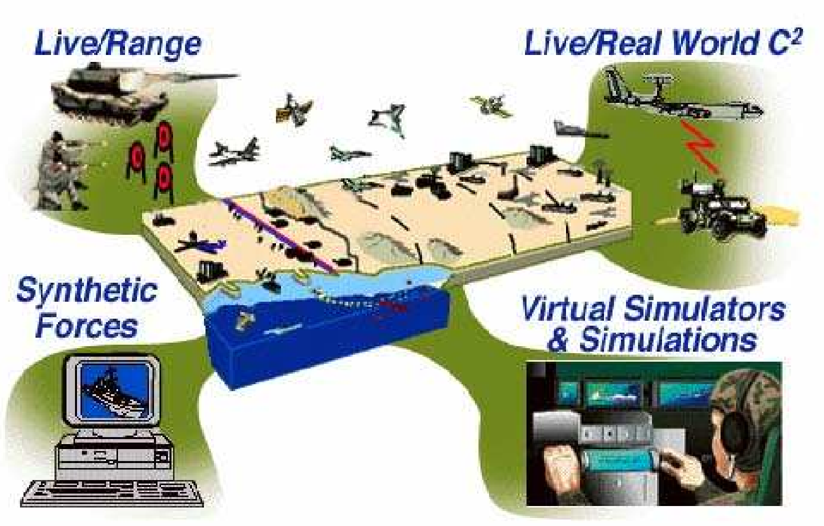


FIGURE 1.1: Synthetic battlespace [22]

TABLE 1.1: Distributed Mission Operations training levels and player types

Training Levels	**Player Types**
Individual	Live
Team	Virtual
Inter-Team	Constructive
Full Mission Rehearsal	

track, and target multiple aircraft), communication, and building situation awareness as being enhanced by such training. Further team skills which were promoted included mutual support, tactical execution, and flight leadership. The instructor pilots amongst the participants in the research reported it was a valuable complement to regular aircraft training.

Research on training effectiveness of multi-ship simulation systems has been ongoing for almost ten years primarily at the Air Force Research Laboratory. In addition to research on networked simulation technologies, activities have focused on application of DMO for continuation training of fighter pilots and air weapons controllers. Human factors issues and solutions lie at the heart of the DMO concept. Human factors improvement is necessary to optimize training effectiveness via any DMO activity. These human factors considerations include learning acquisition and retention, performance measurement, visual perception for simulator visual displays, brief and debrief capabilities, team interactions, mission rehearsal requirements, and advanced distributed learning [24]. The rapidly improving visual realism and fidelity of advanced virtual simulators has already provided significant improvement in training combat skills. In certain key aspects related to the accurate depiction of target, threat and natural environments, the training received in DMO mission simulators exceeds the realism and value of a live training mission [22].

1.1.2 Desktop Solutions

By providing pilots with effective (simulation) training, future accidents can potentially be prevented from happening. However, flying schools and clubs do not currently have access to high-fidelity simulators with large, visual systems that provide a wide field-of-view (FOV) compared to commercial operations. On typical low-fidelity flight training devices, FOV restrictions limit the utility for training visual flying. Even experienced pilots find flying a precise visual pattern on a desktop simulator difficult, resulting in over- and under-shoots during aircraft turns [25].

One such technique originating from virtual reality (VR) research is the implementation of amplified head rotations [26] in order to overcome FOV and field-of-regard (FOR) restrictions. By amplifying or exaggerating the head rotations, the user can be provided with a viewing scene that has turned further than it actually is. This technique relies on the visual dominance effect due to the mismatch between what the eyes see and what head movement the body senses [27]. Amplified head rotations have also been implemented in a variety of VR settings other than helmet-mounted displays (HMD) such as, desktop systems [28], fishtank VR [29], and surround-screen displays [30].

Studies have shown that this novel technique is both acceptable, useful [31] and natural [32]. A recent study by Kopper, Stinson and Bowman [27] investigated the effects of amplification with varying FOV and the detectability of amplification by users. This study used constant amplification factors up to a maximum of three times the unity scale. Participants had a visual scanning and a counting task to perform. For visual scanning, FOV changes were immediately noticed whereas amplification was only detected by half of the participants. It was also suggested that visual scanning can be performed without a performance penalty. In the counting task, however, none of the participants were able to detect amplification. Furthermore, a negative performance was found in counting performance when higher amplification levels were used with larger FOVs. This was due to a lack of sufficient visual cues in the synthetic environment to aid users with view orientation.

On desktop systems, there is a further constraint posed by the requirement to still physically view the display when rotating the head. Unlike a HMD where the display is automatically slaved, users can not make full 360° rotations without losing sight of the display. Amplified head rotations is a candidate technique allowing users to view their virtual surroundings by adopting head rotations with high amplification factors [27]. Although commercial head tracking devices do provide full control of each head axis for unity scaling, amplification requires manual fine-tuning of parameters such as dead zones and gain to achieve a satisfactory result for each individual user. This was a particular challenge to address in the later practical stages of this research as discussed in Chapter 4 and Section 5.5.4.

1.2 Problem Definition

With hardware head-tracking technology now becoming both affordable and available to everyday consumers, amplified head rotations have allowed users to increase utility and immersion in visual simulations [27]. Low-fidelity simulators based on desktop systems will not completely replace the functionality of full mission rehearsal simulators, but will enable pilots to participate in the operational environment and gain training experience at remote locations. Publications of amplified head rotations in the flight simulator domain and multi-monitor/projection system integration have been sparse [27, 33]. Potential cybersickness due to this technique in an applied piloting task is also unknown, as previous control studies were of short duration and did not cover any active control tasks [34]. Presence, commonly linked to virtual environment studies, is beyond the scope of this thesis as in flight training. This is due to pilots being accustomed to train in a wide variety devices regardless of fidelity levels (Section 3.2) due to the task oriented approach. The work in this thesis can therefore be categorized within the Technology Readiness Level (TRL) 2-4 scale range as defined in Table 1.2, with TRL being explained in Appendix I.

Addressing the lower-fidelity spectrum of flight simulation training devices, this thesis therefore covers monoscopic, computer displays or large screen projections and the augmentation of human interaction with the virtual world by means of amplified head rotations. HMDs are briefly mentioned to discuss similarities but again is not part of the research covered in this thesis. Since the research is primarily on visual cues, motion cues to stimulate the vestibular system were not

Table 1.2: Technology Readiness Levels [35]

TRL 1	Basic principles observed and reported.
TRL 2	Technology concept and/or application formulated.
TRL 3	Analytical and experimental critical function and/or characteristic proof-of-concept.
TRL 4	Technology basic validation in a laboratory environment.
TRL 5	Technology basic validation in a relevant environment.
TRL 6	Technology model or prototype demonstration in a relevant environment.
TRL 7	Technology prototype demonstration in an operational environment.
TRL 8	Actual Technology completed and qualified through test and demonstration.
TRL 9	Actual Technology qualified through successful mission operations.

considered. Following these considerations, the problem definition in this research study is stated below.

Problem definition

To investigate virtual reality augmentation on current low-fidelity, fixed-based flight simulation training devices to expand visual training capability.

1.3 Research Goal and Objectives

With the introduction of any new technology into a training solution, the benefits and drawbacks need to be carefully explored in order to qualify it for testing before development into operational use [36]. The usage of low-cost, commercially available technology needs to demonstrate financial benefits in order to serve the lower end of the aviation market [37].

1.3.1 Goal

With the problem defined in the aviation context of Section 1.2 above, the goal of this thesis can be formulated as follows:

Thesis goal

To determine and evaluate the effects of augmenting low-fidelity, fixed-based flight simulators with amplified head rotations to expand their mission rehearsal/training capability and the impact on pilot performance, workload and simulator sickness.

1.3.2 Objectives

With the aforementioned thesis goal shown above, research objectives were composed to steer the direction of research to obtain quantifiable results.

> **Objective One**
>
> To investigate the effects of amplified head rotations on fixed-based flight simulators in a basic flying task.

Establishing a baseline comparison of novel technology versus the most commonly used solution serves as a reference to validate enhancements and benefits before enabling more in-depth studies with advanced tasks. To comprehend the virtual reality technique behind amplified head rotations, a fundamental understanding of select human visual cue perception will be drawn from the literature. The selection of low-cost, low-fidelity, fixed-based flight simulators to research will be drawn from surveying the current simulator market, its training applications and device characteristics. What constitutes the chosen basic flying task will then be determined from the task/training needs analysis. Fitting with the established evaluation methods and practices from literature, the following research questions facilitate achieving the first research objective.

> **Research Question 1:** How does amplified head rotations affect visual flying technical performance compared to a legacy, non-augmented flight simulator?
> **Research Question 2:** How does head augmentation impact workload in a piloting task?
> **Research Question 3:** What simulator sickness side effects occur when amplified head rotations are introduced?

Expanding on establishing a baseline reference comparison as stated by the first research objective leads to the next step of studying how the proposed augmentation technique scales to higher-fidelity configurations to map further training benefits. The topic of fidelity and its relation to training transfer/effectiveness will also be part of the discussion.

> **Objective Two**
>
> To investigate the scalability of amplified head rotations to higher levels of visual display fidelity.

Since scalability by modifying FOV is relevant to flying tasks and has an impact on simulator design, potential workload and simulator sickness will also need to be studied again. Hence, the following research questions were composed to support the second research objective.

Research Question 4: How are head mapping profiles tuned for displays with larger field-of-view?
Research Question 5: How does amplified head rotations on larger displays affect visual flying technical performance in a more complex flying task?
Research Question 6: How does display size in conjunction with head augmentation affect pilot workload?
Research Question 7: Does simulator sickness scale with display size when using head augmented viewing?

1.4 Thesis Outline

Although the eight chapters of this thesis are laid out in a linear order describing the conducted research over a period of three years, the actual thesis structure is much more interwoven as illustrated in Figure 1.2. This roadmap is available again at the start of each chapter to guide the reader through this thesis. In contrast to the basic Systems V model [38], there are numerous loops reflecting the iterative process undertaken. This repeated application of definition decomposition followed by the repeated application of integration and verification ensures a robust approach to perform the research [39].

Chapter 2: Literature Review. This chapter provides a review of related work in support of the research problem and the identified knowledge gap. It starts by discussing related topics in human visual perception to better understand the context of this thesis. This is followed by explaining how flight simulation training devices are currently classified with inherent shortcomings. The chapter then focuses on virtual environments in simulation and highlights display technology parameters of importance and their issues. Finally, suitable view augmentation techniques are reviewed with a candidate solution indicated to solve

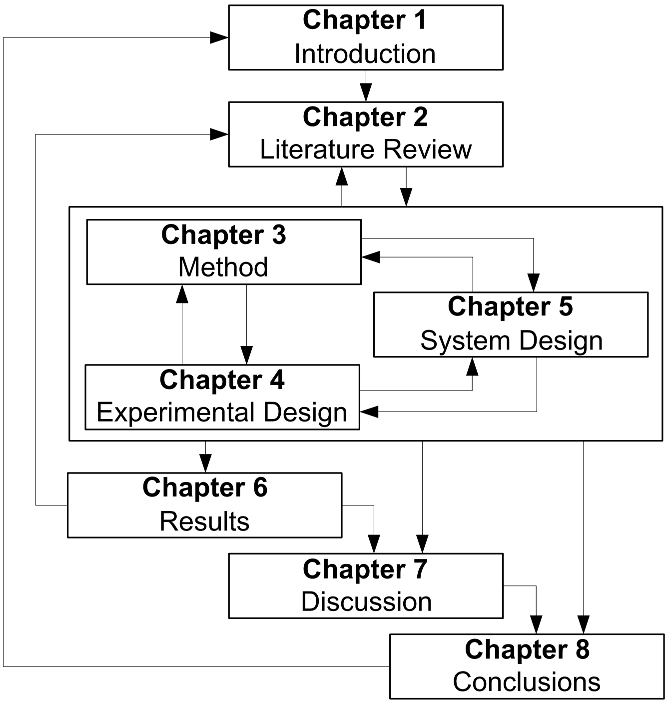


FIGURE 1.2: Thesis roadmap

the research problem. This chapter effectively initiates the iterative loop of experimental methodology, design and system realization. Chapters 3 to 6 are cross-referenced back to this chapter with each loop in order to obtain explanations and lessons learned from literature when similar problems and challenges occur during the research.

Chapter 3: Method. This chapter starts by explaining the relevance of training transfer and simulation fidelity in the context of flight simulation. To achieve the research objectives in quantifying the benefits of amplified head rotations, it discusses the utility evaluation method used to structure the experimental design. This starts by conducting a task analysis based on reviewing literature to define the mission scenario suitable for experimentation. Appropriate performance metrics and assessment tools are then identified. Requirements and constraints influencing the experimental design are also considered. Finally, the output of this chapter drives the experimental design in Chapter 4.

Chapter 4: Experimental Design. With the research objectives stated in Section 1.3.2 and input from Chapter 3, two experiments were designed to collect results to address the research questions. Section 4.1 provides an initial overview of how the experiments were prepared and the common procedures used. Section 4.2

describes the first experiment, which served as a baseline reference corresponding to stage one of the utility evaluation (Section 3.3). This establishes the practical usability of amplified head rotations in its most basic form compared to a legacy non-augmented triple-screen flight simulator. To obtain a fair comparison, the emphasis is on selecting an experimental task that does not impede the non-augmented setup.

The second experiment described in Section 4.3 covers the second research objective of investigating scalability by comparing three augmented displays to study the compounding effects when larger displays and FOVs are used. This was stage two (Section 3.3) of the evaluation process to determine performance improvement of the system itself. By not having any task restrictions posed on the flying task by a non-augmented display, this second experiment can fully utilize the freedom the head augmentation provides in supporting a more complex visual flying task, a feat undocumented before. Finally, Section 4.4 concludes by providing the design overview of the experiments and the tools and statistical techniques used to verify the results.

Chapter 5: System Design. This chapter documents the system design of the research flight simulator built to support the experiments proposed in Chapter 4. Knowledge gained from Chapter 2 and Chapter 3 highlighted the importance of deriving simulator features based on requirements following the ICAO 9625 process explained in Section 3.2.2. Hence, obtaining experimental results from conducting trials in a simulator qualified to these standards would add more credibility and ease of verification on other qualified devices. To successfully implement this process in this research, Section 5.1 explains why systems engineering is fundamental in the simulator system design.

The requirements analysis in Section 5.2 and functional analysis in Section 5.3 highlight the iterative nature of the systems engineering process. Since the simulator also supported two other projects within the AVRRC research group, input requirements from those projects resulted in the collaborative final functional requirements listed in Section 5.2.2. With common resource limitations and stringent timing constraints, risk management was also a collaborative effort as outlined in Section 5.4. This chapter culminates with Section 5.5 presenting the final simulator design after completing the design synthesis loop and verification/compliance.

Chapter 6: Results. This chapter presents the results of the two experiments after performing the statistical analysis. The experiments and obtained results provide insight into benefits obtained with augmentation compared to a standard setup, and serves as a launch platform for further experiments with augmentation on larger displays. This forms the basis leading to a full discussion of the results and its implications in Chapter 7.

Chapter 7: Discussion. The results obtained in Chapter 6 are discussed here in further detail. Interpretations are presented within the context of answering the research objectives and the link to the existing body of research. In addition to experimental findings, a critical evaluation of the simulator design process of Chapter 5 is also delivered.

Chapter 8: Conclusions. This chapter discusses the main research findings shown in Chapter 7 and provides appropriate conclusions for the whole study. It also highlights the key contributions of the study to the aerospace and virtual reality research domains. Finally, current research limitations are discussed with recommendations made for future research.

Chapter 2

Literature Review

This chapter provides a review of related work in support of the research problem and the identified knowledge gap. It starts by discussing related topics in human visual perception to better understand the context of this thesis. This is followed by explaining how flight simulation training devices are currently classified with inherent shortcomings. The chapter then focuses on virtual environments in simulation and highlights display technology parameters of importance and their issues. Finally, suitable view augmentation techniques are reviewed with a candidate solution indicated to solve the research problem. Figure 2.1 visualizes how this chapter fits in the overall thesis structure with its output initiating the iterative loop of experimental methodology, design and system realization. Chapters 3 to 6 are cross-referenced back to this chapter with each loop in order to obtain explanations and lessons learned from the literature when similar problems and challenges occurred during the research.

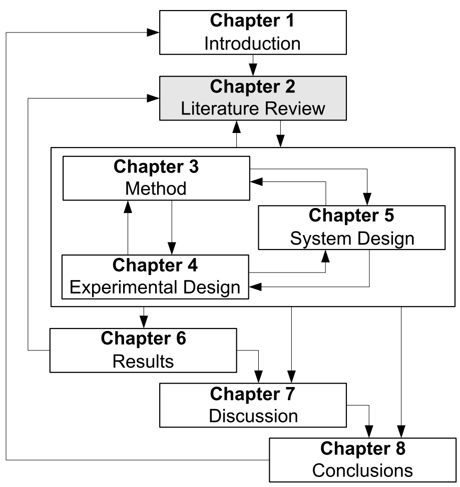


FIGURE 2.1: Chapter 2 in thesis roadmap

2.1 Visual Perception

In most critical phases of flight operations, pilots rely on visual cues to perform their tasks and to maintain situational awareness. The visual systems of a flight simulator is therefore essential in simulating this outside world environment. Before this can be discussed, there is a need to understand how human visual perception works. As the human visual system is a very complex and heavily researched sensory system, it is beyond the scope of this thesis to describe the full biological anatomy and processing involved. However, key characteristics vital to the purpose of this research will be discussed briefly in the following subsections.

2.1.1 Central/Peripheral Vision

In order to comprehend why central and peripheral vision is important in this research, the following excerpt of vision was summarized from Goldstein's textbook [40]. As light enters the eye via the pupil, it becomes focused on the retina. The retina is in essence the sub-system that converts light detection into brain stimuli. It is composed of an array of receptors with two different types of structures: rods and cones. The cones are densely packed into a small region of the retina called the fovea with the primary task of perceiving colour and fine detail. Table 2.1 compares the retinal properties of the foveal region with the peripheral region. The rods, for low light conditions, are sparsely distributed over the remainder of the retina. While our eyes move to centre the view on a target, the focus is to put this target image within the fovea area to get a clear picture. Hence this centralized area of perception is designated 'central vision'. The blurry, surrounding area is called 'peripheral vision', not good for seeing specific details but good at detection of moving objects and self-motion perception. Figure 2.2 illustrates this visual acuity effect against varying locations on the retina while Table 2.2 summarizes the key features of central and peripheral vision.

Table 2.1: Comparison of peripheral and foveal retina [41]

Property	**Foveal Retina**	**Peripheral Retina**
Threshold	Relatively high	Very low
Receptor distribution	Cones only	Rods and cones
Convergence	Limited or none	Extensive
Illumination	Photopic (Daylight)	Scotopic (Night)
Functions	Central/colour/detail vision	Peripheral/achromatic/blur vision

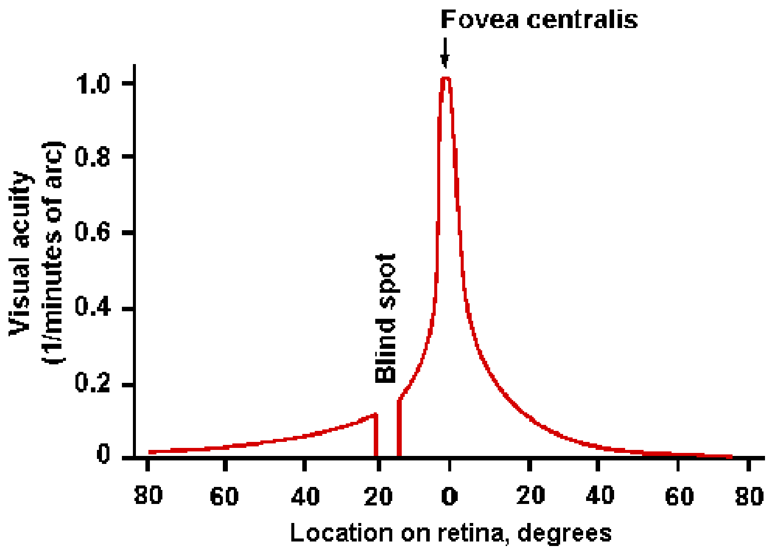


FIGURE 2.2: Visual acuity as a function of position on the retina [22]

TABLE 2.2: Summary of Peripheral and Central Vision Features [42]

Central Vision	**Peripheral Vision**
Serves to answer the question "what"	Serves to answer the question "where"
Small stimulus patterns, fine detail	Large stimulus patterns
Image quality and intensity important	Image quality and intensity not important
Central retinal area only	Peripheral and central retinal areas
Well represented in consciousness	Not well represented in consciousness
Object recognition and identification	Spatial localization and orientation

2.1.2 Optic Flow

Changes in the location of objects in peripheral vision over time provide information about how an observer is moving through their environment [40]. One of the key effects associated with this phenomenon is called optic flow, which according to Gibson's ecological foundations, as used in cybernetics and ecological interface design, is a vital part of perception [43]. He further argued that perception is a bottom-up process, which means that sensory information is analysed in one direction: from simple analysis of raw sensory data to ever increasing complexity of analysis through the visual system [44]. Gibson's theory was first developed during the Second World War when he was given the task of preparing training films for pilots. He attempted to provide training for pilots in depth perception and this work led him to the view that perception of surfaces was more important than depth/space perception. Surfaces contain features sufficient to distinguish different objects from each other. From this he developed a theory of optic flow patterns, for example when pilots approach a runway the point towards which the pilot is moving appears motionless, with the rest of the visual environment apparently moving away from that point, shown in Figure 2.3. Such optic flow patterns provide pilots with information about their direction, speed and altitude. This movement was previously often overlooked in the psychology and visual perception experimentation in the published literature.

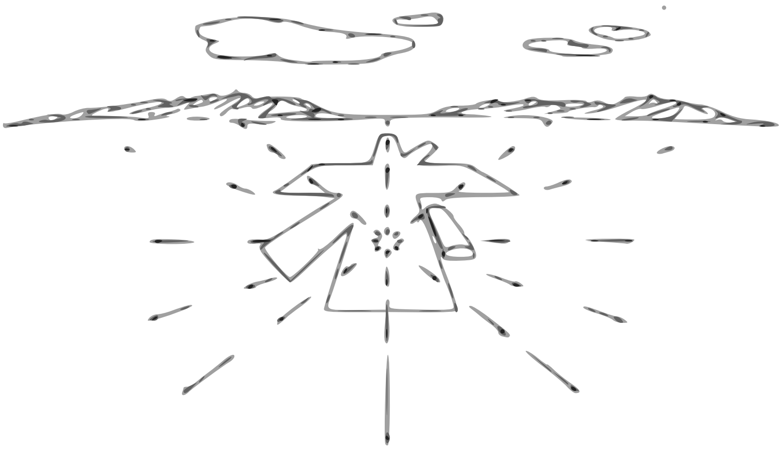


FIGURE 2.3: Outflow of the optic array in during runway approach [43]

Gibson then developed his theory into a general theory of visual perception which has three key components:

1. Optic Flow Patterns
2. Invariant Features
3. Affordances

A brief discussion of these components has been summarized as follows [45].

Optic Flow Patterns. Changes in the flow of the optic array contain important information about what type of movement is taking place. Examples are as follows.

1. Any flow in the optic array means that the perceiver is moving, if there is no flow the perceiver is static.
2. The flow of the optic array will either be coming from a particular point or moving towards one. The centre of that movement indicates the direction in which the perceiver is moving. If a flow seems to be coming out from a particular point, this means the perceiver is moving towards that point;

but if the flow seems to be moving towards that point, then the perceiver is moving away. Figure 2.4 illustrates how the an observer views the optic flow pattern from a moving train.

FIGURE 2.4: Optic flow pattern while looking back out of a moving train [43]

Invariant Features. Texture gradients provide another source of environmental information based on patterns or structures in objects. Because this flow of texture is invariant, occurring the same way every time in the environment, it provides an important direct depth cue for distance, speed, etc. This perception also involves almost little or no information processing by the cognitive system. Two examples of invariants are linear perspective as offered by the convergence of parallel lines of the railway track in Figure 2.4 and texture gradients shown in Figure 2.5.

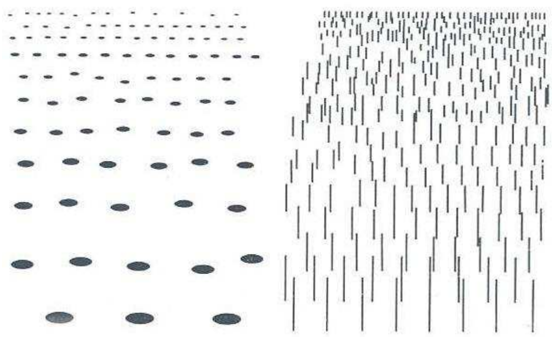


FIGURE 2.5: Examples of texture gradient [43]

Affordances. Affordances are in short, cues in the environment that aid perception as described in Table 2.3.

Table 2.3: List of affordance cues

Affordance Cue	Description
Optical Array	The patterns of light that reach the eye from the environment.
Relative Brightness	Objects with brighter, clearer images are perceived as closer.
Texture Gradient	The grain of texture gets smaller as the object recedes. Gives the impression of surfaces receding into the distance.
Relative Size	When an object moves further away from the eye the image gets smaller. Objects with smaller images are seen as more distant.
Superimposition	If the image of one object blocks the image of another behind it, the first object is seen as closer.
Visual Field Height	Objects further away are generally higher in the visual field.

2.1.3 Target Acquisition

In order to track and identify visual targets, human physiology possesses two reflexes, the opto-kinetic reflex (OKR) and the vestibulo-ocular reflex (VOR). Both are involuntary responses that work synergistically to produce a stable retinal image under a variety of dynamic viewing and motion conditions.

Opto-Kinetic Reflex. OKR is one of several eye movements that function to identify a target in the visual scene, to position the target on the fovea, and to keep it positioned there [46]. The OKR works by evaluating information from the entire retina to determine if image slip is occurring. In case of image slip, a corresponding movement is made in the eye position to eliminate it and to achieve image stabilization. For instance, when peering out the window of a moving vehicle, the reflex detects image slippage and applies a compensating movement to the eye with a gain equal to the motion and direction of the optic flow.

Vestibulo-Ocular Reflex. While motion is an important cue in high-fidelity flight simulation, this thesis focuses on fixed-based, low-fidelity devices. Yet how humans detect motion through the vestibular system is vital to head movements and the relationship with visual perception. The vestibular system is designed to detect and react to the position and motion of the head in space [47, 48]. Its function is critical to the ability to coordinate motor function, ensure correct eye movements and maintain posture. Since it is an unconscious system, its functionality is only experienced when disruptions occur due to certain diseases or acute conditions such as motion sickness.

The physiology of the vestibular system can be summarized as follows [46]. The inner ear is divided into two sections that include a complex set of mechanisms that allow us to hear and sense motion. The cochlea is associated with the sense of hearing and the peripheral vestibular system is associated with balance and sense of motion. The peripheral vestibular system rests in an area of the inner ear called the labyrinth. It is made of up a series of tubes (semicircular canals) and sacs (utricle and saccule). The semicircular canals are primarily responsible for detecting angular acceleration while the utricle and saccule are responsible for linear acceleration.

The three semicircular canals are oriented to detect motion in each of the three orthogonal planes in which motion can occur. Each canal detects motion in a single plane. These mechanisms are suspended in a fluid called perilymph and are filled with a fluid called endolymph. As the head moves, the endolymph within the tube flows causing tiny hairs to bend which generate nerve impulses. The nerve impulses are then transmitted to the brain through the vestibular nerve. The semicircular canals in humans are quite sensitive and can measure angular accelerations as low as 0.1 deg/s^2.

The utricle and saccule work through similar processes. The hair-like cilia of these organs are embedded in a gelatinous mass. This mass has clumps of crystals within it called the otolith. When linear acceleration occurs, the otolith provide enough inertia to flex and stimulate the hair cells. The stimulation results in the generation of nerve impulses that are then transmitted to the brain. The utricle is oriented to be able to detect motion in the horizontal plane and the saccule is oriented to detect motion in the vertical plane and fore-aft plane. These receptors are primarily responsible for our perception of vertical orientation with respect to gravity.

Once the brain receives the impulses from the entire vestibular system, it uses the information for perception of motion and also transmits information to the human visual system. There is a clear relationship between the vestibular and visual systems where angular acceleration information about head movement is supplied to the visual system [46]. The VOR together with the previously discussed OKR cooperate to maintain a stable retinal image regardless of the type of motion being experienced [49]. The VOR is a very fast-acting reflex, compensating head movements in the 1-7 Hz range but is much less accurate at lower frequencies with inferior gain. The OKR is the opposite with a longer latency due to the required evaluation of visual information to determine a response with near unity gain at low (< 0.1 Hz) frequencies. Between 0.1 and 1 Hz frequencies, the OKR begins to lose gain and also develops a phase lag due to inherent response latency. With the two reflexes working in unison, the human body is still able to provide stable retinal images through a wide range of frequencies. This ability enables adaptation to accommodate different sensory arrangements such as when looking through optics such as diving goggles. In simulation applications, subtle artifacts of poor simulator engineering might delay the VOR adaptation process either through inconsistent feedback or by altering the performance of the OKR through visual anomalies and potentially cause adverse effects.

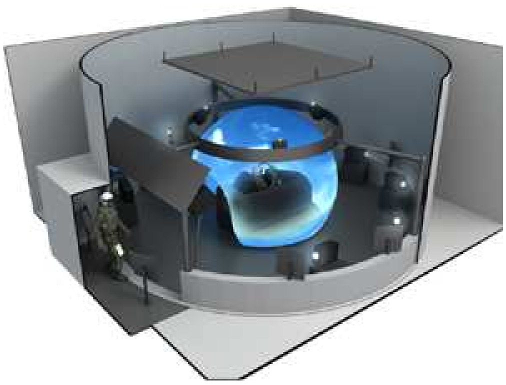
2.1.4 Perceived Self-Motion

Vection is the effect of experiencing the illusion of self-motion (or eigenmotion) while a person is actually static with respect to their environment [48]. In daily life, this occurs for example when looking out of the window and seeing another car pull away from a stoplight while your own vehicle remains static. In virtual environments and immersive simulations with fewer cues to a static reference frame, this is prone to causing vection. In the particular case of fixed-base simulators with no motion cueing, this is only attributed to optic flow changes. In the real-world, motion cues would provide corresponding vestibular information to corroborate the visual cues. Without this, the adverse effect of motion sickness is a common occurring symptom.

The severity of vection is primarily due to two factors: FOV and optic flow rate [46]. Larger FOVs induce greater perception of motion due to the increased stimulus of peripheral vision whereas greater optic flow rates generate a sensation of greater speed. In flight simulation, greater optic flow rates aren't a factor in high altitude flying operations but are a factor when flying closer to the ground where ground objects, terrain features and scene complexity are more prevalent. Visual anomalies must therefore be avoided to prevent giving false cues than one would expect to get in the real-world.

2.2 Flight Simulation Training Devices

High-fidelity simulators, replicating full scale cockpits and with realistic visual systems surrounding the pilot FOV, discussed in Section 2.3.2, are normally used to provide DMO training [50], which was previously introduced in Section 1.1.1. However, such high-fidelity simulators (Figure 2.6(a)) are restricted by their cost, size and infrastructure to deploy on either a large scale or to field locations [51]. Low-fidelity simulators (Figure 2.6(b)) can also provide equivalent training benefits without the extra cost if used appropriately [52]. The primary difference between full-mission simulators and low-fidelity trainers is the significant reduction in the visual scene FOV. This requires fundamental research into methods as presented later in Section 2.4 of overcoming the challenges a narrow FOV presents in order to achieve the same level of training effectiveness without compromising normal task behavior. To fully comprehend the differences and commonalities between these simulator types, a census overview is given next for both civilian and military flight simulators in use today.



(a) High-fidelity dome simulator

(b) Low-fidelity trainer

FIGURE 2.6: Comparison between a) high-fidelity and b) low-fidelity simulators

2.2.1 Civil Flight Simulators

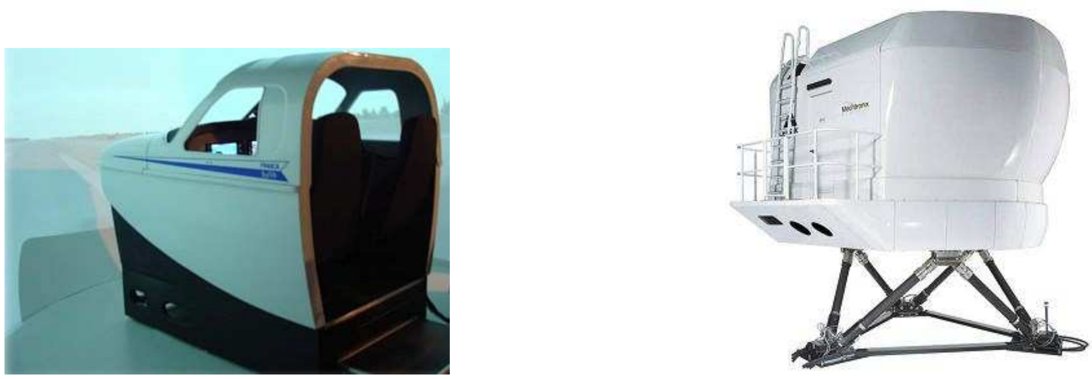
In the civil aviation world, airline operators, training organizations, and flight training centers use flight simulators, known as Flight Simulation Training Devices (FSTD), as a highly effective and economical method of training, testing and checking aircrew. These FSTDs are regulated by the National Aviation Authorities (NAA). That is, the FAA in the US and the CAA in the UK. Other NAAs have developed their own standards for the complete range of FSTDs, for both aeroplane and helicopters [36]. FSTDs range from instrument trainers with no visual displays, PC based desktop flight training devices to large motion-based full flight simulators commonly known as Level D simulators for airline checkrides.

Figure 2.7 shows two examples of desktop FSTDs which are FAA approved. An illustration of a static FSTD and full motion simulator is shown in Figure 2.8. With a variety of names and capabilities assigned to simulators by individual aviation authorities, it is difficult to correlate their attributes at a worldwide level. This may cause inefficiencies for pilot licensing, ratings and checks for all but the top-level, highest-fidelity simulators (Level D). This lack of harmonization occurs even between the two largest aviation bases of North America and Europe [53].

(a) PC/Basic

(b) Advanced

FIGURE 2.7: Examples of FAA approved desktop training devices [54]



(a) Static flight training device

(b) Level D full motion flight simulator

FIGURE 2.8: Flight training devices

Morrison [55] named five reasons behind the need to update the FSTD standards: 1) regulatory changes, 2) lack of harmonization, 3) new aircraft types, 4) new training types and 5) new technologies. The aviation industry's frustration due to the above mentioned reasons led to to an International Working Group (IWG) by the Royal Aeronautical Society (RAeS) in 2006 to review FSTD technical criteria. The IWS decided that a fundamental review was necessary to establish the simulation fidelity levels required to support each of the required training tasks for each type of pilot licence, qualification, rating or training type [36]. This task analysis process to support the updated flight simulator qualification standard is covered in Section 3.2.2.

2.2.2 Military Flight Simulators

The number of military flight simulators has been increasing yearly due to high cost of aircraft programmes and budget constraints. There are around 1,912 known military training devices worldwide according to a 2011 census [56]. The characteristics of these simulators and FSTDs are related to the role of the aircraft for which they are used as trainers. These devices have either an outside-the-window (OTW) visual system or a motion system, or both, full-size cockpit controls and mostly full-size replica cockpits. They also range from Unit Level Trainers with only one visual channel to full mission rehearsal simulators which are similar to a civilian Level D simulator. With over 1,050 simulators for fighter aircraft types, g-force cueing is an additional concern when compared to civilian simulators. Nearly 580 military simulators can be networked and only 60 out of the total amount are designed to be transportable as the survey showed [56]. Despite common terminology and designations across operators, there is no documentation of simulator standards because usually each simulator is a bespoke piece of equipment.

2.2.3 Research Flight Simulators

Regarding research simulators, Rehmann states:

> "The essential feature of simulator experimental investigations is to introduce the pilots into a closed-loop control situation, so that account is taken of their capabilities and limitations regarding the performance measure being evaluated. The expectation is that within the bounds of the experimental conditions, behaviour in the simulator will match their behaviour in the flight situation. Hence, the primary goal for a flight simulation researcher is to produce experimental conditions that elicit behaviour that would occur under similar circumstances in the real world." [57]

In research, other factors are of importance in addition to physical realism, such as the level of realism perceived by the pilot (perceptual fidelity) discussed in Section 3.2.1. Since non-research FSTDs do not have to consider this aspect, their classification system is not applicable to research simulators. This lack of classification for research simulators has lead to confusion when specifying what type of simulation is necessary for a particular research task, and is the major reason for performing analysis of fidelity requirements for simulation research.

2.3 Virtual Environments

The first immediate aspect of a desktop simulator a pilot notices is the restricted visual field due to the limited dimensions of the computer display. To overcome this shortcoming, desktop simulators offer the user alternative viewports or view panning by pressing appropriate keys or an axis on a controller. But this increases workload (discussed in Section 3.5) and is not always intuitive. Even experienced pilots find flying a precise visual pattern on a desktop simulator difficult, resulting in over- and under-shoots during turns [25]. To overcome this limitation, there are various emerging techniques in the gaming industry which have not been adopted by the modelling and simulation industry. By building upon this foundation of novel gaming technology, it will be possible to provide simulation solutions boosting low-fidelity simulators to higher levels of training capabilities [15].

2.3.1 Display Technology

To address the issue of not having a full FOV encompassing dome projection system, various display technologies in the field of virtual reality (VR) are available to choose from [58]. Kalawsky provides extensive technical descriptions and issues of VR systems based on the degree of immersion [59]. For mobile simulations, the two most common solutions are single or multi-monitor displays and head-mounted displays (HMD).

Head-Mounted Displays. The HMD is a popular device in the VR research community, but the narrow FOV and unwieldiness requires more user effort to operate compared to the real world. Besides research into advanced hardware technology, augmentation techniques have been proposed as a solution to overcome these limitations [32]. HMD technology is often associated with fully immersive VR systems and is a popular solution when portability is an important factor. HMDs can be classified into two categories: non see-through and see-through [59]. The ongoing drive in HMD technology is the ability to provide a wide enough FOV display that can fully encompass human vision [60]. Latency and visual resolution are also paramount to the functioning of the HMD system for training effectiveness since it affects human perception of visual (Section 2.1.1) and vestibular cues (Section 2.1.3). Existing visual perceptual issues with HMDs have been comprehensively documented [61, 62]. Reviewed FOV studies indicated that larger FOVs with HMDs led to better task performance for flying tasks as well as a reduction in subjective workload. Despite being popular in VR research, narrow FOV and inherent cumbersomeness (weight, physical restrictions, system parameters) requires more effort for a user to operate when compared to the real-world. Besides research into advanced hardware technology, augmentation techniques such as those discussed in Section 2.4 have been proposed as a solution to overcome these limitations [32].

Winterbottom, Patterson and Pierce [63] recommended that a FOV of 40° may be sufficient for tasks that occupy primarily the central vision. The importance of tasks that require peripheral cues (Section 2.1.1) or relative movement to other objects/vehicles (Section 2.1.2) would need a FOV much greater than 60°. Technical limitations also impose constraints on the design of HMDs, so a trade-off must be made between either a wide FOV or high resolution. Large FOVs also induce simulator sickness, which is discussed in further detail in Section 3.6.

Fixed Displays. The desktop flight simulator domain has traditionally used single or multiple computer displays to represent the external environment. In a study [51], a low-fidelity desktop training simulator equipped with triple 30″ monitors was compared to a full 360° high fidelity dome simulator. Although pilots had a negative opinion prior to conducting the experiment, the results showed that pilots performed at the same level in the desktop simulator. The task was to fly four-ship-air-combat sorties which included formation flying, a task that requires a large FOV. Pilots in the desktop simulator maintained better formation though at cost of a higher mental workload. An explanation given for this better result was due to the heightened state of awareness pilots were in because they were trying to compensate for the reduced FOV.

2.3.2 Field-of-View/Field-of-Regard

FOV here is defined as the momentary subtending visual angle of the scene at the pilots eye [64]. This section discusses this from a cognitive task perspective. Figure 2.9 shows the FOV of a pair of eyes looking straight ahead. The centre of the diagram represents the centre of vision of each eye and the grey portions are the regions seen by each individual eye only. The central white area depicts the overlap region seen by both eyes where stereovision is possible. The black areas resemble the cut off due to the brows, cheeks and nose. The field-of-regard (FOR), which incorporates eye and head movements, covers most of the sphere around the observer [65]. Operators of vehicles will encounter FOR obstructions due to the vehicle structure or aircrew equipment. Thus, the FOV at any one moment is a subset of the overall possible FOR.

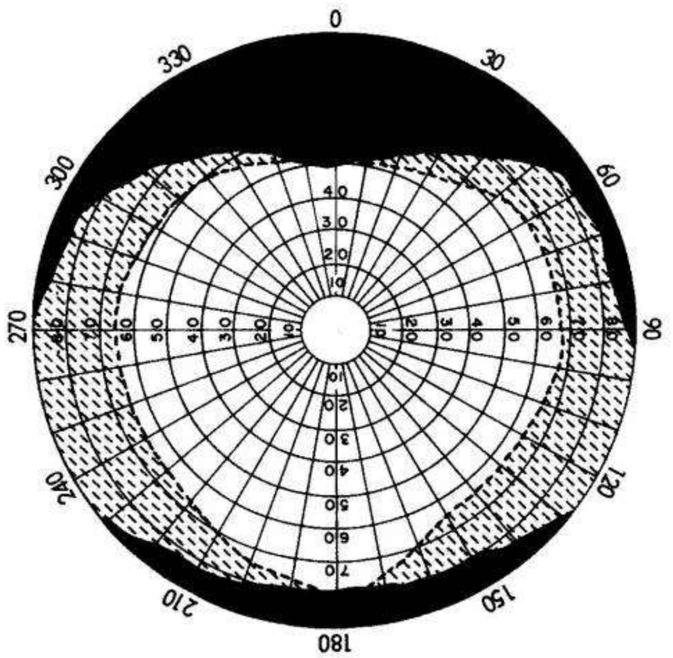


FIGURE 2.9: Field-of-view of a pair of eyes [66]

In the real-world, a pilot using head and eye movements can look around rapidly and each glance covers a very wide FOV. Similar to a car driver checking left and right at an intersection for traffic, the head movement that changes the gaze and FOR is traditionally limited by the available FOV of the fixed simulator displays. Translating this capability into a simulator FOV requirement is very dependent on the training task. Table 2.4 gives a general visual requirement on the FOV posed by varying manoeuvres for fixed-wing aircraft. Takeoff and landing is a critical visual manoeuvre in both civilian and military aviation, as Adams [53] states: “The visual display for lower-level devices is not necessarily elaborate. It is required for a few key tasks and transitioning from looking at the instruments to looking outside. So there needs to be some representation of the runway.” The seven International Civil Aviation Organization (ICAO) Flight Simulation Training Devices levels also have a visual requirement on the FOV, this is shown in Table 2.5.

TABLE 2.4: Typical fixed-wing aircraft simulation requirements [66]

Manoeuvre	**Total FOV**
Takeoff and landing	75° H ×30° V
Air-to-air, high altitude	Large
Air-to-air, low altitude	Large
Formation flight, high altitude	80° max
Inflight refueling	60° – 120°
Air-to-ground: weapon delivery, navigation	Large

TABLE 2.5: ICAO FSTD levels and (horizontal/vertical)FOV requirements [36]

ICAO FSTD Level	**HFOV**	**VFOV**	**Display Type**
1, 3, 5	200°	40°	direct-view
2, 4	45°	30°	flat screen
6, 7	200°	40°	collimated

The most common configuration, shown in Figure 2.10, representing the external environment for outside visuals is a three-channel display, typically giving a FOV between 150×40 and 180×45 degrees [56]. There are 320 simulators with a five-channel display, many giving a 240×45 FOV. There are 200 six-channel display systems, some having facets where back-projection is used with flat screens that surround the pilot and cockpit. There are about 60 seven-channel display systems including helicopter simulators with a five-channel display system spread horizontally and two extra so-called chin windows that use collimated monitor units to enhance downward view for hovering. Dome and wrap-around faceted systems have even more channels, about 160 simulators being listed with eight-channel displays and over.

An experiment examining the effect of reduced FOV on the control of a roll disturbance [67] indicated that there was little performance change for FOV larger than 40°. This suggests that FOV may be reduced in deployable systems without adversely affecting training effectiveness.

FIGURE 2.10: Three-channel visual display system configuration

HMD FOVs can affect the frequency and velocity of head movements made by a user (Section 2.1.3), and may consequently affect the extent of image displacements on the display caused by lags. A study by Wells reported significant interaction between FOV (between 20° and 120°), the number of displayed targets, the number of hits and the subjects' response time in a target acquisition task [68]. There was no effect of FOV on performance with three targets, but performance was significantly degraded by FOV of 60° or less when nine targets were displayed. Subjects were observed to move their heads less with the larger FOV, since more targets were simultaneously visible on the display, allowing them to be monitored more effectively using only eye movements. Head movements made with the larger FOV were observed to be faster than those with smaller FOV. This corresponds with intrinsic human properties of visual target acquisition as outlined in Section 2.1.3.

In further simulated flying tasks between wide conventional displays and helmet-mounted displays (slaved to the line of sight) with narrower FOV, flying accuracy down a narrow canyon was reported to be greater with a 23° HMD than with a 102° fixed display, unless the pilots were forced to make head movements by a secondary target capture task [69]. However, smaller horizontal position errors were found in a simulated aerial refuelling task with a 300° projection display than with a 40° FOV HMD [70]. Wider FOVs reduce the requirement for users to make head movements to keep a task in view, and may therefore reduce image position errors caused by system lags.

2.4 View Augmentation

A display device, being either a HMD or a computer display, when placed at a static distance from the observer will have an inherent display FOV. The geometric field-of-view (GFOV) is defined as the virtual viewing volume as an input to the display device. In most VR applications a perspective projection is chosen such that depth cues are consistent with a users real-world view, so the GFOV will match the HMD FOV resulting in a one-to-one (unity) mapping. Altering FOV parameters to non-unity levels would result in distorting the scene: users are presented with a mini- or magnification of the actual image. An experiment revealed that subjects preferred a virtual scene with a GFOV that was 14.9% larger than the HMD FOV [71]. This optical distortion has an impact on the users' scene perception such as distance estimation [72], which stems from the fact that a visual cue such as affordance (Section 2.1.2) has been affected.

Minification and FOV issues in aviation has been predominantly researched regarding synthetic vision (display) systems (SVS) and remotely-operated vehicles where the operator only has as small display available [73–75]. For flight training, this technique is not applicable because the optical distortions hinder transfer to the real world where negative training transfer as discussed in Section 3.1 is unacceptable.

2.4.1 Amplified Head Movements/Rotations

HMDs with head-tracking offer immersion to users by enabling them to look around the virtual scene by natural head rotations/movements. But the narrow FOV means more frequent and larger head movements are required to see parts of the environment where normal short eye movement in the real world would suffice [76]. As introduced in Section 1.1.2, desktop systems pose severe viewing angle constraints for head rotations unlike view slaved HMD systems. There are physical viewing angle limits for the user: turning ones head beyond these angles would mean the displayed image is no longer visible!

Amplifying physical head movements/rotations to the virtual camera effectively solves this by allowing head movements to view their virtual surroundings on a low FOV display instead of using a one-to-one mapping [27, 32]. Figure 2.11 illustrates the example of mapping a head movement to the right by an amount 'd' to a virtual camera movement of a gain factor 'E times d'. Applying gains to all three head-axis rotations and three movement/translatations allow for a full six degrees-of-freedom. While rotating the head in the real world, sensory information (Section 2.1) such as vestibular, proprioceptive as well as visual information create consistent sensory cues that indicate self-motion (Section 2.1.4). Amplification deceives the user into seeing a (virtual) scene that has turned or moved further than the actual physical head, but by exploiting the dominance of the visual cue above the vestibular confirmatory cue and the adaptability of the vestibulo-ocular reflex (VOR) as explained in Section 2.1.3.

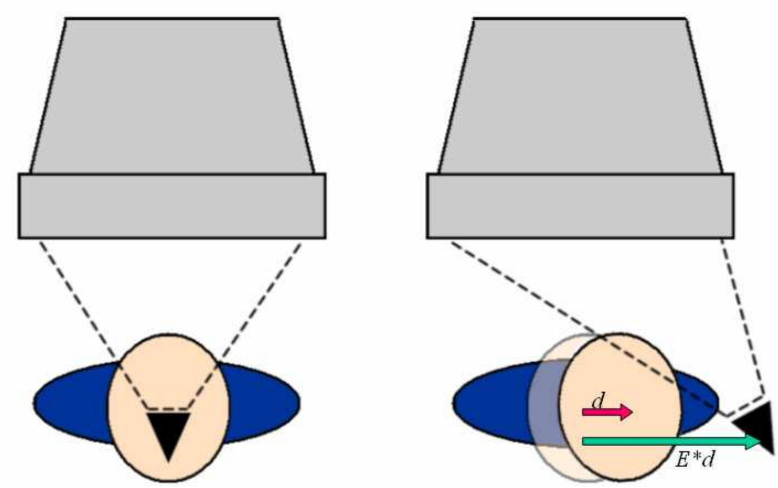


FIGURE 2.11: Example of amplified head movement [77]

This technique has been shown both to be acceptable and useful: users preferred an average amplification of 1.26 while using a HMD with 60° FOV [31]. Subjects also found amplification natural: they noticed head movement attenuations significantly faster than amplifications [32, 78]. Scene movement in the opposite direction of head rotations (i.e. amplifications) should be avoided [79].

The inherent benefit of controlling the viewpoint with one's head is intuitive [80], as humans already use their head in daily real life to look around. Furthermore, offloading this control from other traditional input devices such as a mouse, remote control, game-pad or hat switch enables the use of input devices for their originally intended function [32]. Looking around an aircraft whilst securely strapped in the cockpit means that head rotations rather than head-translational movements are of primary interest. As such, amplified head rotations have also been implemented in a variety of VR settings other than HMDs for visual search tasks: desktop systems [28], fishtank VR [29], and surround-screen displays [30].

Conversely, the majority of HMD systems had tracking errors and latency issues, resulting in severe vestibular and proprioceptive cue mismatch. Investigating the effect of latency on perceptual stability, researchers found that subjects were not able to detect small inconsistencies between real and virtual yaw rotations [31, 79]. Results showed that users may judge the virtual world as stable when the virtual rotation is slightly amplified compared to the real yaw rotation [79]. Furthermore, users tend to unwittingly compensate for small inconsistencies between vision and vestibular sensation [71]. Again, this is attributed to the adaptability of the human VOR (Section 2.1.3).

A current knowledge gap in the literature exists, with few studies addressing this emerging capability of amplification, and in particular to flight simulation applications [26, 81, 82]. With the many different visual display configurations available in terms of size and viewing angles, there has been scarce documentation on integration and usability [27]. Potential user sickness symptoms due to this technique in an applied piloting task are also unknown, with the majority of control studies being of short duration and not covering any active control tasks. After effects have also not been studied, an example of a yet to be researched topic is being how visual scanning patterns learned with amplification might inadvertently yield adverse effects. This is of interest for transfer of training to the actual operational environment.

2.4.2 Head-Coupled Factors

In order to document the evaluation process of integrating amplified head rotations into a flight simulator, lessons learned from prior research in head-coupled systems with HMDs is useful to systematically address common system design factors. A series of experimental studies have investigated the effects of lags in head-coupled systems on tasks involving tracking virtual targets with the head, tracking and manipulating virtual targets with the hands, simulated vehicle control, and target search and recognition [83]. It was shown that the target capture-time was significantly increased by imposing an additional lag of 67 ms on a head-coupled system which has a basic system lag of 40 ms for static targets. In an experiment, subjects were required to place the aiming reticle inside a target circle as quickly and as accurately as possible, and to keep it inside the circle for at least 350 ms. In another experiment involving the tracking of a continuously moving target, tracking errors were significantly increased by imposing an additional display lag of only 40 ms on a 40 ms basic system lag. The correlation between the head motion and the target motion also decreased with increasing lag, particularly at higher motion frequencies.

Another study on character search and recognition task in the HMD yaw axis reported consistent increases in search time with greater exponential lags in excess of 40 ms [84]. The frequency of head movements is also a factor influencing the performance of a manual control task with a head-slaved display. In HMD experiments with a fixed FOV, participants were forced to turn their heads to keep the task in view but as system delay was increased, subjects increasingly inhibited their head movements because of the de-coupling between head position and the position of the displayed image [85].

These results stress the importance of system latency in virtual reality systems with head movement. It is imperative to keep system latency below 40 ms for head-tracking applications as the task difficulty increases substantially and may force users to adopt different strategies to cope with this factor. Assessment of the likely effects of image lags in a particular head-coupled system therefore requires a knowledge of the pattern of head movements required by users.

2.5 Summary

The visual displays of a flight simulator provide the vital cues of the synthetic outside world that a pilot needs to perform visual flying manoeuvres. The central and peripheral vision features of the human visual system serve as the main detection mechanism of perceiving light. To detect visual changes over time, the phenomenon of optic flow is used comprising of patterns, invariant features and affordance cues. For tracking and identification of visual targets, human physiology utilizes two reflexes that can operate either independently or together. The first reflex is the opto-kinetic reflex (OKR) which uses eye positioning for image stabilization. The second reflex is the vestibulo-ocular reflex (VOR) which compensates for head movement in conjunction with cues from the vestibular system.

The current state of the flight simulation training device market was surveyed for civilian, military and research segments. It was found that there were many different types and classification of devices in the civilian domain, especially on the lower-fidelity solutions. Yet, there was a lack of harmonization between the national aviation authorities which led to inefficiencies in pilot licensing, rating and checks worldwide. The military and research flight simulators were even more diverse since each was often custom built for a specific platform or application.

The survey found that the most common visual configuration in the low-fidelity segment offered a triple-channel image using fixed computer displays. The limited FOV and fixed FOR of these systems severely limited their training application for visual flying. To overcome this, the virtual reality technique of amplified head rotations was proposed as a technical solution. Its non-unity mapping of the physical user head movement into the virtual scene change enabled full FOR coverage. Head movement for view control was not only an intuitive method of interaction, it also allows pilots to manually operate the flight controls and aircraft systems unimpeded.

Chapter 3

Method

This chapter starts by explaining the relevance of training transfer (Section 3.1) and simulation fidelity (Section 3.2) in the context of flight simulation. To achieve the research objectives in quantifying the benefits of amplified head rotations, Section 3.3 discusses the utility evaluation method used to structure the experimental design. This starts by conducting a task analysis based on reviewing literature in Section 3.4 to define the mission scenario suitable for experimentation. Appropriate performance metrics and assessment tools are then identified in Section 3.5 and Section 3.6. Requirements and constraints influencing the experimental design are discussed in Section 3.7. Finally, the output of this chapter drives the experimental design in Chapter 4, as the roadmap in Figure 3.1 illustrates.

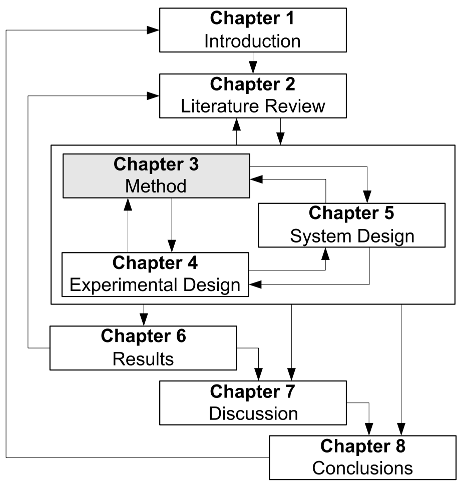


FIGURE 3.1: Chapter 3 in thesis roadmap

3.1 Training Transfer

Training transfer is defined as the indication of the effectiveness of simulator training [86]. One of the key benefits of flight simulator development is to reduce or replace training time in the actual aircraft for cost benefits, practical and safety considerations [11]. Allerton [87] emphasized three important aspects of training transfer: 1) measuring performance training transfer, 2) cost benefits of simulator versus real aircraft and 3) impact of simulator fidelity.

Training transfer has several directional outcomes. The first direction is forward transfer with the obviously desirable positive transfer for example when a trainee has learned something that can be transferred to the real world. Negative transfer, however, must be avoided where learning on a simulator interferes with operation on the real aircraft. In this research topic for instance, a pilot flying the simulator may learn a systems exploit by tilting his head in a certain, unrealistic direction to obtain better views, but this would impede his performance in the real aircraft. Secondly, the case where simulator training has no impact on actual performance obviously questions the value of simulator training. Finally, reverse transfer of training is also possible where experience with the real flying actually improves performance in the simulator. This can occur when expert pilots with vast experience and knowledge of advanced flying techniques perform much better in a simulator especially when facing an environment void of real danger or risk of life [88, 89].

Skill retention is another aspect crucial to aviation where pilots have to maintain proficiency for safety and licensing as a training task. In other fields, training remedies have only been administered via post-reported or observed decrements in performance [90]. Prior studies have also indicated that the likelihood for decay of skill depends greatly on the degree to which the skill was actually learned [91], arguing the case of extensive simulator training yielding a high acquisition environment. Therefore it is just as important to measure skill acquisition (degree to which the skill has been effectively learned) as it is to measure post-training transfer performance [92].

Common practice is to have a subject matter expert (often an instructor pilot) assess the performance of the simulator trainee and determine a positive or negative transfer of training [93]. More objective evaluations utilize a Transfer Effectiveness Evaluation (TEE) or Transfer Effectiveness Ratio (TER) [11, 94]. This metric compares the time difference on the real aircraft needed by a control group and a trainee group divided by the amount of hours needed on the simulator. An alternate metric exists in the form of Incremental or Cumulative Transfer Effectiveness Ratio (ITER/CTER) [11]. This was developed due to objections that the TEE/TER alone was not a sufficient indication of skill learning efficiency. CTER measures the average number of trials needed to reach standard proficiency on the actual aircraft. A CTER of 1.0 indicates that training on the simulator is identical to training on the real aircraft, while values above and below 1.0 indicates that the simulator is more or less efficient.

Due to the unavailability of wide-angle visual systems in past military research, only a modest amount of transfer has been demonstrated so far for aerobatics and air combat skills [95]. Due its classified nature, there is little public domain information regarding the value of simulation for air combat training. The last publicly available extensive literature review [96] found only limited supporting data, primarily subjective appreciation for the training potential and extremely limited actual transfer data. For air-to-air combat, two out of three transfer studies produced positive results but lacked statistical backing due to small effect sizes. Five out of six studies for surface-attack training also yielded positive transfer for conventional weapons delivery. Two studies demonstrating transfer to live exercises again yielded fairly small effect sizes. A twofold explanation for these findings can be made: there are not a lot of publications available in the first place due to the classified nature of work, and that these studies were done before the shift to competency-based training. The weakness of data from transfer evaluations to date is that they tend to be very task and weapons system specific [96].

Studies that compared a control group with no simulator training versus a simulation taught group handicap the overall training of the control group [95–97]. In order to log objective measurements of pilot performance to show benefits or drawbacks of simulator training, transfer studies to actual aircraft must be conducted, but in reality the complexity and practical costs are very high [87]. The same resource limitations were also applicable to this thesis: with no access to high-fidelity simulators or real aircraft, transfer effectiveness could not be measured except for performing an intrinsic comparison between simulator sub-variants.

3.2 Simulation Fidelity

Historically, simulators have been designed and produced under the concept that the effectiveness can be equated to its realism [57, 98]. However, simulation fidelity today is an extremely topical item in light of political and financial constraints. The lack of a clear-cut understanding of fidelity definitions by the supplier, procurer and end user in the past has led to over-specification, retro-modifications and unsatisfactory acquisition processes [99].

3.2.1 Definitions of Simulation Fidelity

Simulation fidelity is formally defined by the Fidelity Implementation Study Group [100] as:

> "1. The degree to which a model or simulation reproduces the state and behavior of a real-world object or the perception of a real-world object, feature, condition, or chosen standard in a measurable or perceivable manner; a measure of the realism of a model or simulation; faithfulness. Fidelity should generally be described with respect to the measures, standards or perceptions used in assessing or stating it. See accuracy, sensitivity, precision, resolution, repeatability, model/simulation validation.
>
> 2. The methods, metrics, and descriptions of models or simulations used to compare those models or simulations to their real-world referents or to other simulations in such terms as accuracy, scope, resolution, level of detail, level of abstraction and repeatability. Fidelity can characterize the representations of a model, a simulation, the data used by a simulation (e.g. input, characteristic or parametric), or an exercise. Each of these fidelity types has different implications for the applications that employ these representations."

Fidelity is then further divided into two descriptions: qualitative and quantitative fidelity. Qualitative fidelity is often used as a short description of fidelity for marketing purposes although subjective evaluations done by expert pilots are useful [100]. Quantitative fidelity is defined as the more common simulator requirement, dictating critical system parameters (accuracy, precision, resolution) and form an integral part of the iterative systems engineering process for modelling and simulation, as shown in Figure 3.2.

The research and development of simulators has focused on reaching technical merits replicating sensory cues available to humans such as those listed in Section 2.1. However, replicating the real-world to the maximum extent is not always necessary

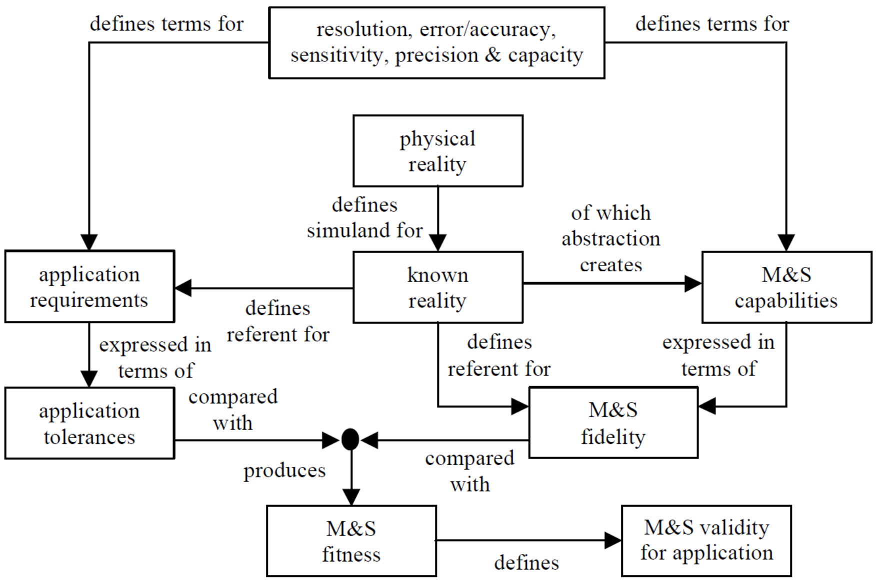


FIGURE 3.2: Fidelity-related conceptual relationships [100]

to achieve the training need. Hence, another term called operational fidelity can be defined as ensuring the simulation is an authentic representation of the complex operational environment of an organization [101]. Methods currently used for understanding the operation of complex systems frequently adopt an abstracted approach to data acquisition. For instance, Cognitive Task Analysis and Work Domain Analysis is used to inform systems design through the decomposition and analysis of individual work elements.

Further human factors pertaining to fidelity lies in psychological fidelity. This refers to the perceived realism and task fidelity of users [102] and can either enhance or cause detrimental performance effects. Task fidelity is commonly achieved by addressing vital physical and functional fidelity to recreate operational realism, but this is not a guarantee of achieving authentic scenarios. Lower fidelity simulators have proven to be able to provide training benefits without the extra expense and complexity of high fidelity simulators [2, 52, 103].

The development of new purposely designed lower-fidelity simulators does not represent a threat to existing high fidelity solutions, rather they offer optimal value for training and can be integrated as such to complement an overall curriculum of simulation solutions to form an optimal training package. This ideal situation in which low-fidelity desktop devices and high-fidelity simulators are employed as complementary elements of an integrated curriculum can assist in maximizing learning potential by delivering training in the early events of instruction that are typically underemphasized in simulation-based training [104], shown in Figure 3.3.

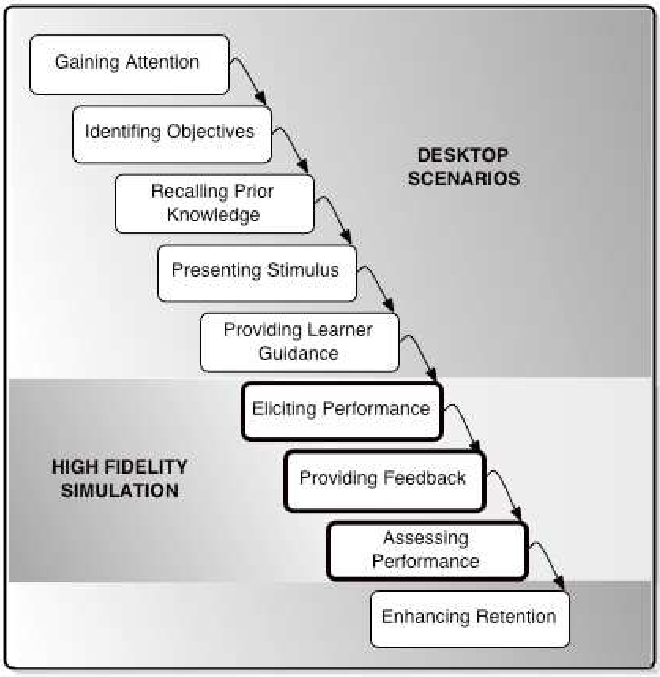


Figure 3.3: Proposed optimal use of mixed-fidelity training devices [104]

3.2.2 Simulator Specification and Qualification

To overcome the growing confusion of flight simulator classification by national aviation authorities (Section 2.2.1), it was necessary to update the fidelity levels to support specific training tasks depending on the exact licence, qualification, rating, or training type intended [105]. The output of this process in form of the ICAO 9625 publication [36] provides a step-by-step process map, illustrated in Figure 3.4 to determine the fidelity levels and qualification criteria for the simulation features according to such training task considerations. The end result is to produce a specific Qualification Test Guide (QTG) for the candidate flight simulation training device (FSTD) solution.

Fifteen training types were compiled from various global NAA definitions, which culminated in approximately 200 training tasks. Private pilot, instrument rating, commercial pilot and airline transport pilot licences are the most common examples of these training types. Then depending on these training tasks, the level of training or training to proficiency is selected in accordance with requirements and standards that need to be achieved (based upon existing NAA). Using a building block of 12 simulation features based upon the FSTD standards, one can then derive the level of fidelity needed in order to satisfy these training requirements.

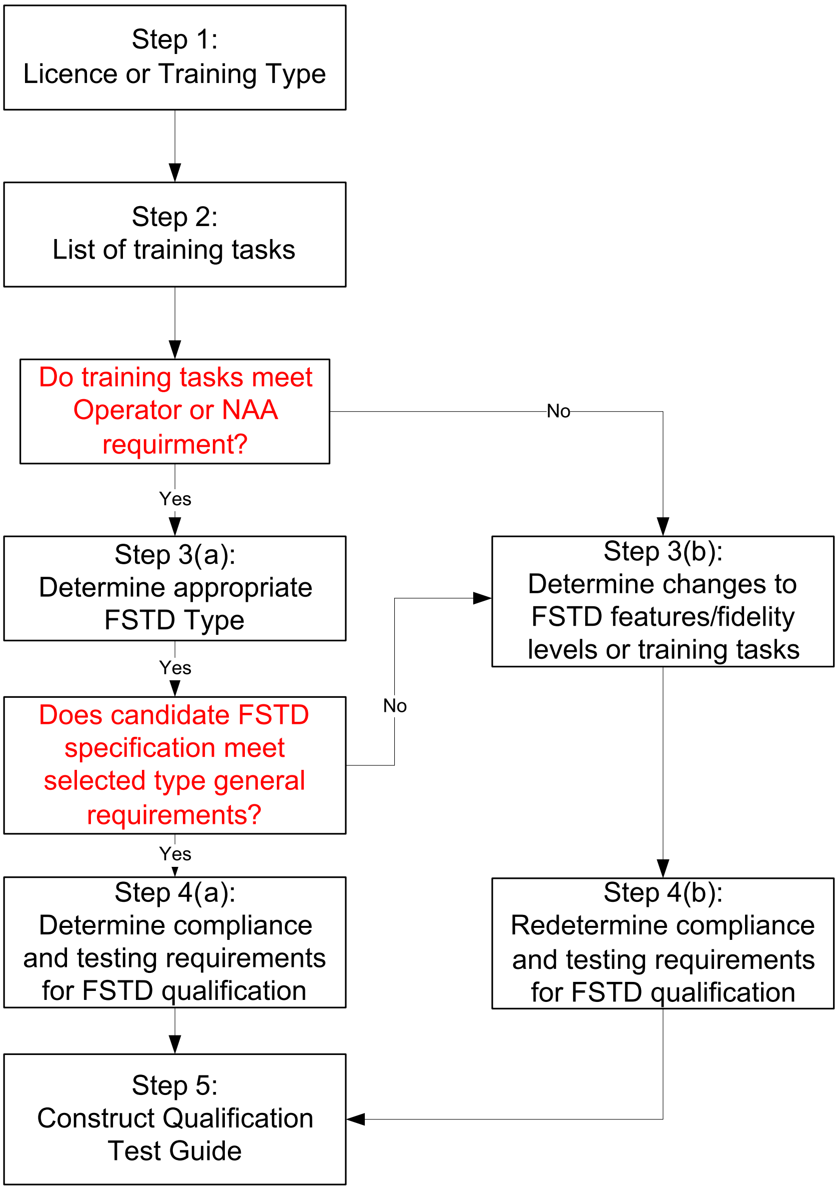


FIGURE 3.4: Simulator fidelity qualification specification process condensed from [36]

The training task analysis, shown in Figure 3.5, enables the true definition of any simulator device using these fidelity building blocks. This reduced the technical categories of FSTD down to a proposed standardized seven. Yet even with this harmonization, it is still possible to build a new FSTD tailored to specific training requirements which still can not be categorized by any of the seven proposed levels.

Four fidelity levels are categorized: None, Generic, Representative and Specific. By determining the minimum level of fidelity required for each training task, a fidelity/feature matrix can be produced, shown in Table 3.1. A worked example of this process as used in this research study is available in the form of a statement of compliance in Appendix L.

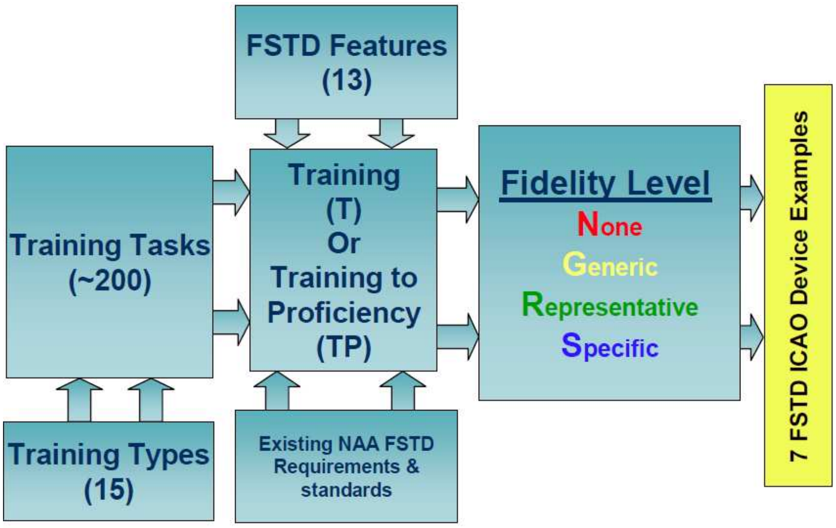


FIGURE 3.5: Training analysis process [36]

TABLE 3.1: Fidelity level and simulation feature matrix [36]

Level	Aircraft	Cueing	Environment
None	Not required.	Not required.	Not required.
Generic	Not specific to model, type or variant.	Generic to an aeroplane of its class. Simple modeling of key basic cueing features.	Simple modeling of key basic environment features.
Representative	Representative of an aeroplane of its class, e.g. four engine turbo-fan aeroplane. Does not have to be type specific.	For sound and motion cueing only: replicates specific aeroplane to maximum extent possible. However, physical limitations currently only provide representative, not specific, cues. For visual cueing only: representative of the real world visual environment and perspective.	Representative of the real world environment.
Specific	Replicates the specific aeroplane.	For visual cueing only: replicates the real world visual environment and perspective.	Replicates the real world environment as far as possible for any specific location.

The main simulation features, with more details available in Appendix B, are then grouped into three categories as follows.

Aircraft simulation

1. Cockpit layout & structure
2. Flight model (aero & engine)
3. Ground handling
4. Aircraft systems
5. Flight controls and forces

Cueing simulation

1. Sound cue
2. Visual cue
3. Motion cue

Environment simulation

1. ATC
2. Navigation
3. Weather
4. Airports & terrain

3.3 Utility Evaluation

Bell *et al.* [96] grouped the different approaches for the training effectiveness of flight simulation from literature into three broad categories, namely: utility evaluations, in-simulator learning and transfer of training. Utility evaluations for user acceptance were found to be too subjective since it involved subject matter experts rating the effectiveness of the system. Despite this, positive user opinion is a prerequisite for system acceptance and clears the way for more extensive evaluations. Simulator learning is a category that requires proof that users improve performance as they practise more in the simulator. If this does not occur, then it is believed that there will be no benefits in the actual aircraft. Transfer of training is the final category where performance benefits in the real aircraft are noticeable after training in the simulator. Bell et al. [96] then recommended a new multistage systematic evaluation approach in order to quantify the value in their case of simulator-based air combat training, as listed below.

Stage 1 Utility Evaluation

Stage 2 Performance Improvement

Stage 3 Transfer to Alternative Simulation Environment

Stage 4 Transfer to Flight Environment

Stage 5 Extrapolation to Combat Environment

By applying this technique to actual aviation tasks, this research study will complete Stage 1 of the utility evaluation mentioned above, bringing the research objectives to a similar level as operational test and evaluation objectives for simulator acquisitions. The objectives of this stage are twofold:

Utility Evaluation objectives

1. Evaluate the accuracy/fidelity of the technique applied to the flight simulation environment
2. Gather data concerning the potential value of this application within a training environment

To support these objectives a series of piloted experiments is required to gather objective data as well as subjective feedback. Objective performance measures can only be established once suitable scenarios/tasks have been selected. User opinion and simulator sickness ratings can be administered in a straightforward manner. Mulder [106] distinguishes four phases in this experimental method: design, implementation, analysis and synthesis. Following this approach ensures a robust experiment and transparency for replication.

For the initial design phase, there are three actions. First, the scenarios/tasks need to be defined and performance measures identified. Secondly, the scenario is designed and developed in the simulation of choice. Third and last, a data collection system is designed to provide the raw data for follow-up statistical analysis.

The actual experiment is then referred to as the implementation phase. This consists of the experiment briefing where subjects receive information pertaining to the task to be performed and the schedule. Subjects then proceed to the learning phase, where they adapt to the experimental facility and experiment itself. Once they are accustomed and sufficient learning to perform the experiment is shown, then the actual experiment is conducted when objective performance data are recorded. After the experiment, subjective data are measured by administering questionnaires to the subject.

The analysis phase consists of analysing the recorded quantitative and qualitative data using statistical analysis. The results of these findings are then used to design future experiments and/or form a basis for utility recommendations which take place in the synthesis phase.

3.4 Task Analysis

To evaluate the contribution that amplified head rotations add to the flight simulator, a suitable flying scenario was needed to carry out the experiments. The challenge was to define a scenario that would be representative of potential training applications yet would enable an objective comparison without favouring any particular system. Since Section 1.1 identified military mission rehearsal and general aviation flight training to be the two main areas of application, the following literature review of published studies formed the basis for selecting an appropriate flying scenario for experimentation.

3.4.1 Military Training

Military training scenarios are quite prevalent throughout the literature with many air combat tasks being studied for simulator usage as mentioned in Section 1.1.1. Air-to-air fighter combat has been investigated in its most simple form as one versus one Basic Fighter Manoeuvres (BFM) [107] to test for its training applicability on low-fidelity training devices. In this study, participants flew a jet aircraft positioned initially 15 nautical miles head-on from an opposing computer controlled aircraft. The objective was for the subject to reacquire the target aircraft within a 20° frontal cone as quickly as possible. The computer controlled target aircraft flew a random profile out of nine available presets after the merge. If the target aircraft was not acquired after 75 seconds, then this was considered a failure. This experiment task was the initial candidate scenario for experimentation due to its direct air combat origins. Unfortunately, it was unfeasible to replicate due to the lack of access to qualified combat pilots for experimentation.

More elaborate studies involving team training have been conducted with a four-ship F-16 formation flying an air defence scenario [51]. The formation's integrity was measured by recording the average ranges between the wingman and his flight lead, the number of times the wingman strayed out of formation and pilot opinions on the visual fidelity of the simulator. This task is extremely focused on a particular military training scenario and again not enough qualified subjects could be recruited nor were there adequate simulation facilities available to support such an experiment.

Brickner [65] studied how pilots performed flying a helicopter around a slalom course based on established rotorcraft handling requirements criteria [108]. Performance measures were the number of pylon hits, misses and average turn distance (ATD), shown in Figure 3.6. This scenario was initially selected and modified for fixed-wing aircraft with an entry and exit corridor/window marking the start and end of the timed course. The aircraft has to fly alternating between the pylons, set at random distances to prevent pilots from memorizing the course but distances are within the aircraft's turning performance capabilities. This course simulates an evasion training exercise and also tests the handling quality of the aircraft.

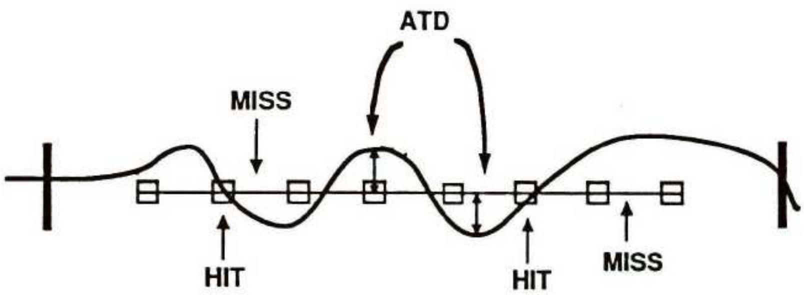


FIGURE 3.6: Slalom course based on [65]

This slalom scenario was then dropped for two reasons: 1) the task was predominantly in the frontal forward view hence FOR could be fixed and the head augmentation would not have added any value and 2) there was no access to participants with rotary wing experience to perform the experiment. Although the slalom course could be adopted and modified for fixed-wing aircraft, it was simply not a typical, representative flight phase for fixed-wing aircraft. Performing low-level slaloms again belonged to the military aircrew domain where pilot familiarity with the aircraft is essential before attempting such manoeuvring, and again lack of qualified subjects for this flying task eliminated it from being selected.

3.4.2 Student Pilot Training

The use of simulation in general aviation to provide training for gaining of private and commercial licenses is currently mostly restricted to instrument training, although studies have shown that training quality improved by using part-task simulator training [109–111]. Again, financial costs are currently preventing widespread adoption of simulator based training despite the clear advantages offered [37]. A complex visual scenario widely used in training is the basic airport traffic pattern, an established procedure path to be flown designed to let air traffic flow in and out of an airport in an organized manner. Patterns may vary depending on airports but most are based on the basic rectangular ground course that all pilots are familiar with as part of their flight training [112, 113].

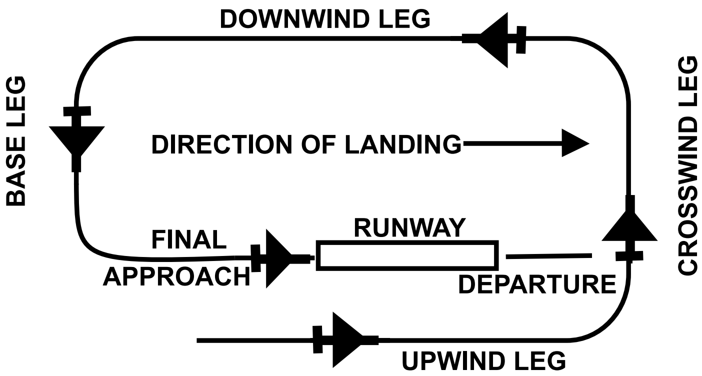


Figure 3.7: Airport traffic pattern example [114]

A standard rectangular traffic pattern is illustrated in Figure 3.7. The pattern altitude is usually 1000 feet above ground elevation of the airport surface. Using a common briefed altitude at a particular airport enables collision avoidance especially at uncontrolled airfields. The Airplane Flying Handbook [112] further states: "When entering the traffic pattern at an airport without an operating control tower, inbound pilots are expected to observe other aircraft already in the pattern and to conform to the traffic pattern in use."

The last three phases prior to landing: downwind, base and final legs as mentioned in Section 1.1 are the most critical with the highest percentage of accidents. Hence, these flight segments are of particular interest to this research aiming to improve pilot performance.

The traffic pattern can be broken down into flight segments with the following description for each leg.

Downwind leg. A course flown parallel to the runway, but in the opposite landing direction. The lateral distance of this leg from the runway is approximately 0.5 to 1.0 mile out at the specified traffic pattern altitude. During this leg, pilots complete the before landing checklist and extend the landing gear. Pattern altitude should be maintained until abeam the approach end of the landing runway where power is reduced and a slight descent begun. The downwind leg continues past a point abeam the runway threshold to a point approximately 45° past it, where pilots perform a medium bank turn onto the base leg.

Base leg. The transitional part of the traffic pattern between the downwind leg and the final approach leg. Depending on wind, it is flown at a sufficient distance from the approach end of the landing runway to permit a gradual descent to the intended touchdown point. The ground track of the airplane while on the base leg should be perpendicular to the extended centerline of the landing runway, although the longitudinal axis of the airplane may not be aligned with the ground track when it is necessary to turn into the wind to counteract drift. While on the base leg, the pilot must ensure, before turning onto the final approach, that there is no danger of colliding with another aircraft that may be already on the final approach.

Final approach. The last and most important leg of the pattern with a descending flight path starting from the completion of the base-to-final turn and extending to the point of touchdown. Here, the pilots judgment and procedures must be the sharpest to accurately control the airspeed and descent angle while approaching the intended touchdown point.

Covelli [64] had subjects fly a visual flight rules (VFR) traffic pattern in a helicopter simulator. Subjects were positioned initially on the base leg of the pattern, had to make a coordinated turn to line up with the runway and subsequently land at the runway intersection. Time to land, ground track and vertical path were recorded. Head and eye movements were also monitored. Flying a visual pattern is a prerequisite during basic flight training, so this task modified for a fixed-wing aircraft would be highly suitable.

Another variation of the generic traffic pattern is its military equivalent: the overhead pattern. This was designed to get both individual and formations of (high-performance) aircraft on the ground in a fast organized manner, minimizing the time the aircraft is low and slow being vulnerable to threats. This pattern is shown in Figure 3.8. The pilot initially lines up with the runway, overflying until the break point, then performs a break turn up to 60° bank (dependent on aircraft type, airspeed) in transition to the downwind leg. The aircraft is then slowed down and configured for landing. Upon reaching the perch point, of which a visual cue example is shown in Figure 3.9, the pilot then makes a final turn to line up on the runway and make a full-stop landing.

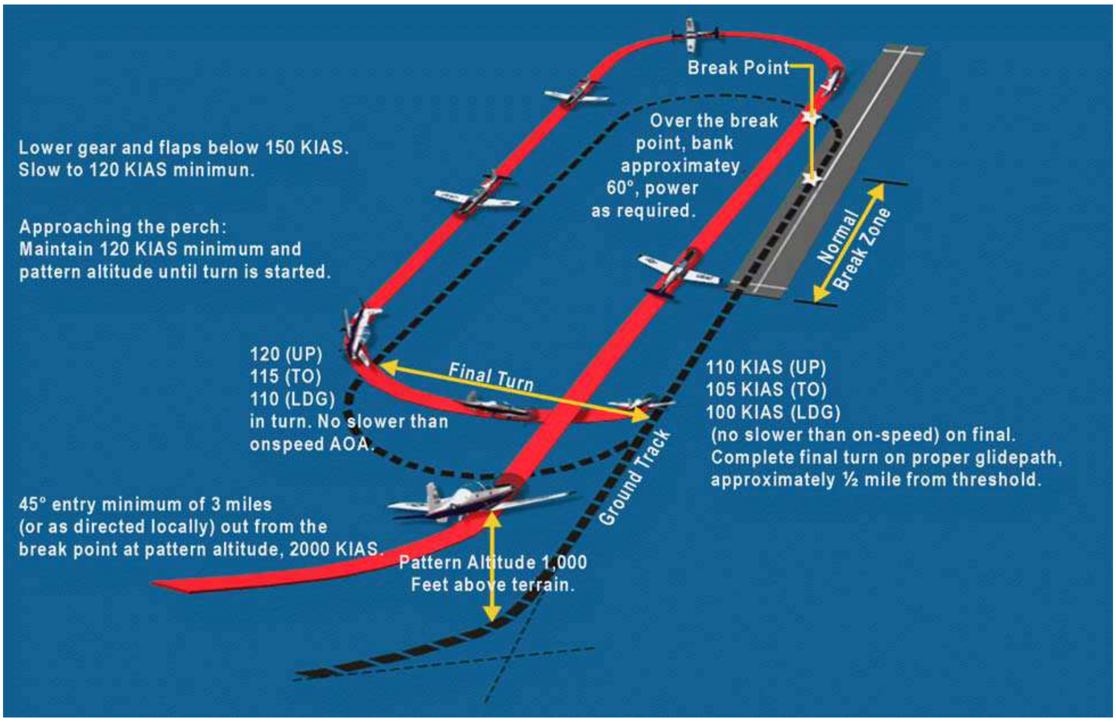


FIGURE 3.8: Military overhead pattern [115]

3.4.3 Task Selection

On the basis of aforementioned task discussion, a basic, civilian visual flying circuit was selected as the test scenario. The military overhead break of Figure 3.8 was deemed too complicated and unusual for civilians and would also have required a high performance aircraft to execute. A regular visual circuit is familiar to all pilots as it is part of basic flight training and can also be easily divided into sections for clarity and measurement. Another task in vicinity of the airfield is to maintain a lookout for other (airborne) traffic, this secondary visual search task forms an excellent complement to the selected flying task since it already is a integral part of the pilot duties. Kopper [27] researched visual scanning and counting in a

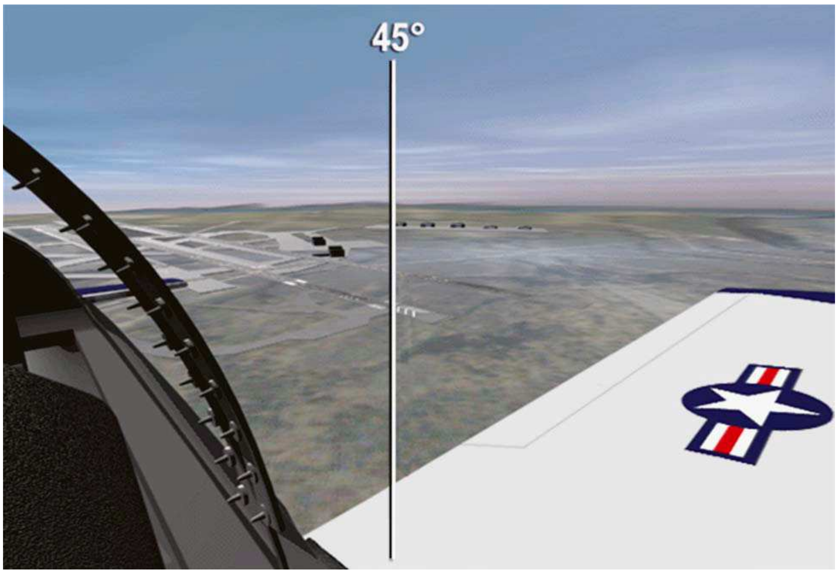


FIGURE 3.9: Perch point visual reference in the T-6 Texan II [115]

virtual urban environment from a moving vehicle with head amplifications. The percentage of identified threats, number of people counted as well as detectability of head amplification were measured. Extending this work towards its aviation equivalent of visual scanning for airborne traffic out of a moving aircraft would extend this knowledge area.

3.5 Workload Assessment

Workload is an important factor when introducing new technologies. Assessments are therefore commonly used in human factors research and the field of human-machine interaction. In this research, workload assessments provide a key insight into the utility validation of the augmented system and how it scales across the various display configurations.

The measurement of workload is classified into three main categories: physiological, subjective and performance-based measures [116, 117]. Physiological measurements consist of: 1) eye related measures; 2) brain related measures; 3) heart related measures and 4) other measures such as skin and muscle activity. Although heart rate monitoring is relatively cheap, the remainder of the physiological measures are more expensive to implement and all are quite intrusive to the participant, hence these were all discarded for this research.

Table 3.2: Sources of load [119]

Choose between each pair of loads	
◯ Mental demand	◯ Physical demand
◯ Mental demand	◯ Temporal demand
◯ Mental demand	◯ Performance
◯ Mental demand	◯ Effort
◯ Mental demand	◯ Frustration level
◯ Physical demand	◯ Temporal demand
◯ Physical demand	◯ Performance
◯ Physical demand	◯ Effort
◯ Physical demand	◯ Frustration level
◯ Temporal demand	◯ Performance
◯ Temporal demand	◯ Effort
◯ Temporal demand	◯ Frustration level
◯ Performance	◯ Effort
◯ Performance	◯ Frustration level
◯ Effort	◯ Frustration level

The subjective methods are the most popular assessment method with subjective workload rating scales being the de facto standard in aerospace [116]. The Subjective Workload Assessment Technique (SWAT) and the NASA Task Load Index (NASA-TLX) are the two most often used, reliable and validated methods. In an evaluation by Rubio *et al.* [118], which directly compared these two methods based on five factors (intrusiveness, sensitivity, convergent validity, concurrent validity and diagnosticity) NASA-TLX was recommended. Hence, NASA-TLX [119] was the method used in this research for workload assessment. It is administered by having respondents fill in a subjective, off-line self-assessment form in two steps. The first step is to rate their experience of the experiment on six metrics (mental, physical, temporal demands, performance, effort, frustration) on a scale, see Figure 3.10, then finally to pick the most appropriately felt source of load when being offered pair-wise comparisons as weighting factors, Table 3.2.

In short, the weights for each individual participant and experiment condition are calculated first [88]. Weighting factor for each workload source is determined by counting the number of times it was selected in the pair-wise comparison. The weighting factor has a maximum value of 5 and a minimum of 0. The sum of all weights is always equal to 15. Magnitudes range from 0 to 100 depending on how participants marked it on the scale rating. Finally, the overall workload score is calculated by multiplying the weights together with the appropriate magnitudes and dividing the compound score by 15.

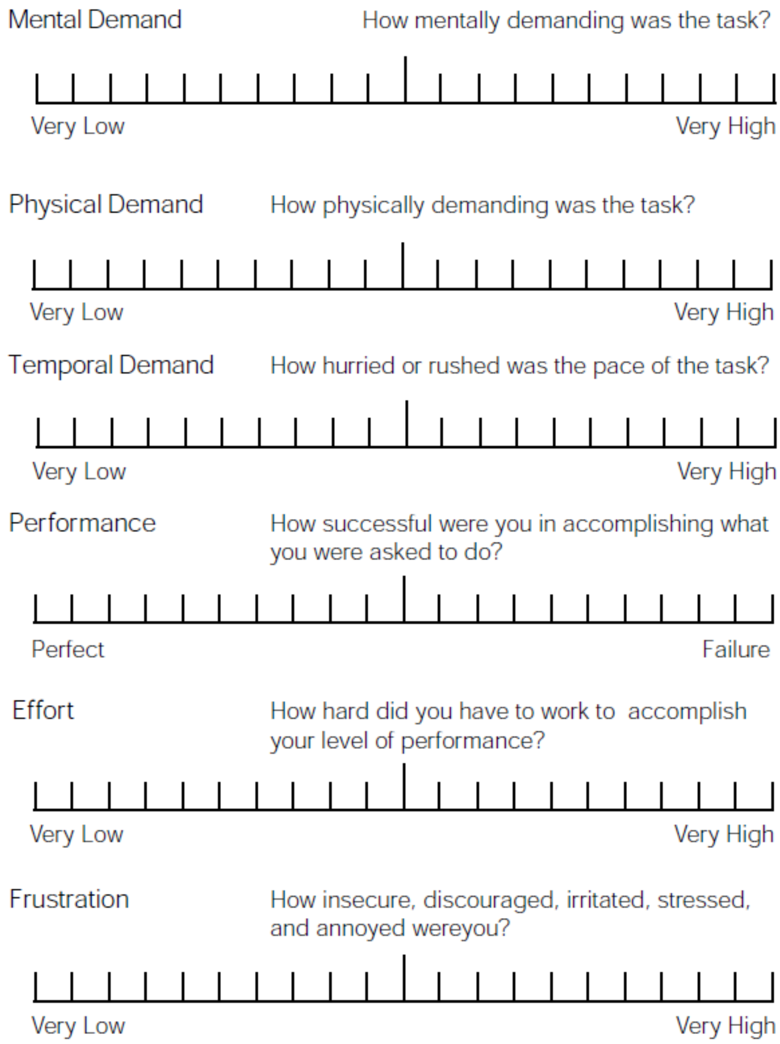


FIGURE 3.10: NASA-TLX magnitude of load selection [119]

Besides these subjective measures, it is crucial to have another source of data to cross-check in order to validate experimental results. Performance-based measures examine part of the operator's capability to perform tasks in order to objectively assess the workload. These consist of two categories: primary task and secondary task measures. Primary task measures in flight simulation evaluations typically record measures such as flight control input and vehicle state parameters. Secondary task measures are applied in conjunction with a primary task, and indicate the spare capacity of an operator.

Secondary tasks in flight simulation literature are often critiqued for being intrusive when operators are assigned an artificial secondary task that bears little relevance to the scenario, such as finger tapping, which can be detrimental to the primary flying task [116]. It was therefore recommended to utilize embedded tasks such as radio control panel usage, monitoring of cockpit system alerts to generate measurements that are of high sensitivity and realism in the experimental task. In this thesis, objective flight technical performance is used as a direct performance comparison criteria, hence the secondary task method was selected to provide an additional workload metric to correlate with the subjective, self-report NASA-TLX results.

3.6 Simulator Sickness

Immersion in simulation and virtual environments (VE) can induce side effects collectively referred to as simulator sickness, also popularly known as cybersickness [120]. Although similar to motion sickness, simulator sickness is less severe and is caused by a mismatch of sensory information from two or more sensory systems (mostly visual and vestibular systems). This often results from exposure to simulated environments on the ground, such as fixed-based simulators and virtual environments, where strong visual cues are present without any confirmatory vestibular backup information (motion) [121].

3.6.1 Symptoms of Simulator Sickness

Simulator sickness is categorized into three classes of symptoms [122]:

Simulator Sickness Symptoms

1. Disorientation
2. Oculo-motor
3. Nausea

After-effects, such as visual flashbacks and balance disturbances have occasionally occurred up to 12 hours after exposure and symptoms of motion sickness and postural disturbances have been reported to be higher in simulators employing space-stabilized helmet-mounted displays (HMD) (i.e. immersive virtual reality systems) than in simulators with dome-based projection systems or fixed display monitors [122, 123].

3.6.2 Contributing Factors

There are three prominent theories extensively postulated in literature behind the occurrence of simulator sickness, namely: cue conflict theory, poison theory and postural instability. Cue conflict theory is simply sickness occurring due to a sensory mismatch between what a person expected versus what the simulator presents. Poison theory has the premise that sensory artifacts such as visual instability is similar to being intoxicated so the body reacts to this due to evolution by vomiting. Postural instability is when an individual is trying to maintain stability under an (new virtual) environment where they have not yet learned strategies to do so. All three theories remain controversial and do not fully explain the phenomenon of simulator sickness [46].

Extensive compilations of simulator sickness factors have also been published [46, 124]. These range from individual factors such as age, gender, illness to technology factors such as display technology, simulated task and duration and are divided into three main categories: user, system and task characteristics. The factors associated with each characteristic are listed in Table 3.3, Table 3.4 and Table 3.5 respectively.

Table 3.3: User characteristics contributing to simulator sickness [122, 125, 126]

Physical	Experience	Perceptual
Age	With VR system	Flicker fusion frequency
Gender	With corresponding	Mental rotation ability
Ethnic origin	real-world task	Perceptual style
Postural stability		
Health state		

Table 3.4: System characteristics contributing to simulator sickness [122, 125, 126]

Display	System lags
Contrast	Time lag
Flicker	Update/refresh rate
Luminance level	
Phosphor lag	
Refresh rate	
Resolution	

With such a large number of factors, it is beyond the scope of this thesis to discuss them all. Because there is yet a method to completely eliminate simulator sickness

Table 3.5: Task characteristics contributing to simulator sickness [122, 125, 126]

Movement in VE	**Visual image**	**Task interaction**
Control of movement	Field-of-view	Duration
Speed of movement	Scene content	Head movements
	Vection	Sitting vs. standing
	Viewing region	
	Optic flow	

[48], the implication towards engineering new technologies is to at least try and minimize the likelihood with proper control of technical factors based on reported findings [127]. For this thesis, system and task characteristics are most relevant as they are directly linked to the engineered application and appropriate design considerations can be made during the requirements specification stage already. Hence, these factors need to be taken into account during the system design stage of Chapter 5 and iteratively checked during the verification before experimental trials are conducted.

3.6.3 Assessment of Simulator Sickness

Johnson [128] discussed literature reports of difficulties with measuring simulator sickness. Most relevant to this thesis are: simulator sickness is polysymptomatic, hence one cannot simply measure one dependent variable, individuals vary in susceptibility to simulator sickness with trials commonly reporting half of the participants not experiencing any symptoms and weak effects that disappear quickly upon exiting the simulator [129].

There have been various ways of measuring simulator sickness in the literature [128]: direct observations of participants for signs such as facial pallor/sweating, instrumented measurements of physiological measures such as respiration rate/stomach activity, postural equilibrium tests to measure simulator-induced disorientation/ataxia and self-report assessment on types/severity of symptoms in the form of a questionnaire. Of all these tests the Simulator Sickness Questionnaire (SSQ) [123] is the most often used and validated method for measurements [130]. The advantages of the SSQ test is that it provides not only an overall metric value but also diagnostic information on the symptom categories, its ease of administration and quick measurement process. It also enables comparisons across simulators, populations and temporal use within the same simulator even.

Table 3.6: Simulator Sickness Questionnaire [122]

Circle how much each symptom below is affecting you right now:	
01. General discomfort	None - Slight - Moderate - Severe
02. Fatigue	None - Slight - Moderate - Severe
03. Headache	None - Slight - Moderate - Severe
04. Eye strain	None - Slight - Moderate - Severe
05. Difficulty focusing	None - Slight - Moderate - Severe
06. Salivation increasing	None - Slight - Moderate - Severe
07. Sweating	None - Slight - Moderate - Severe
08. Nausea	None - Slight - Moderate - Severe
09. Difficulty concentrating	None - Slight - Moderate - Severe
10. Fullness of the head	None - Slight - Moderate - Severe
11. Blurred vision	None - Slight - Moderate - Severe
12. Dizziness with eyes open	None - Slight - Moderate - Severe
13. Dizziness with eyes closed	None - Slight - Moderate - Severe
14. Vertigo	None - Slight - Moderate - Severe
15. Stomach awareness	None - Slight - Moderate - Severe
16. Burping	None - Slight - Moderate - Severe

The SSQ is a self-report symptom checklist [123] as shown in Table 3.6; respondents indicate for 16 symptoms the severity in four levels (none, slight, moderate, severe) of what they are experiencing currently. These four severity levels are assigned a numerical value as follows.

None = 0
Slight = 1
Moderate = 2
Severe = 3

The particular count of symptoms in Table 3.6 is then tallied.

Nausea = items (1+6+7+8+9+15+16)
Oculo-Motor = items (1+2+3+4+5+9+11)
Disorientation = items (5+8+10+12+13+14)

The next step is then used to compute the three sub-scale (factor analysis based) scores [123, 130].

Nausea subscore = Nausea $\times$ 9.54
Oculo-Motor subscore = Oculo-Motor $\times$ 7.58
Disorientation subscore = Disorientation $\times$ 13.92

Finally, the total severity score for SSQ is calculated. This is achieved by using the sum of the three previous three sub-scales multiplied by another (factor analysis derived) constant [123].

$$\text{Total score} = (\text{Nausea} + \text{Oculo-Motor} + \text{Disorientation}) \times 3.74$$

3.7 Requirements and Constraints

Before designing and performing simulator trials, there were pre-experiment requirements and constraints that impacted the design and methodology. This section describes the requirements and constraints which were applicable in the experiments.

Top-level requirements. The top-level requirements are basically the test objectives, these can be formally restated as: (i) compare amplified head rotations versus a baseline non-augmented setup and its scalability across various display fidelity levels; (ii) evaluate pilot performance; (iii) quantify workload differences between configurations and (iv) report occurrences of simulator sickness.

Experiment requirements. Experimental observations in a controlled manner require two important factors: (i) an equal number of runs for each condition to be able to statistically compare them; (ii) different conditions and runs in a randomized order to ensure that systematic similarities having strong influence on the observations can be minimized.

Experiment constraints. These are practical constraints which comprise of time and resources. The timing aspect was that the experiments had to be completed in time so that the results could be analysed and reported in order to meet publication deadlines. Also the sequential order of the experiments dictated that the first experiment had to be successfully completed in time before systems integration work could be finished and the systems prepared for the next experiment.

A further constraint was the availability of participants in both quantity and quality (flying experience). Novice and low-hour private pilots could be recruited from the local student population and flying club. Although it is better for the scientific validity of the experiments that professional pilots from commercial airlines or service personnel participate in the experiment since their opinions and feedback are of more value, access to such experienced pilots was very limited such that any sufficient numbers were practically unobtainable.

3.8 Summary

The key benefit of augmenting low-fidelity flight simulators is to increase their training effectiveness with cost benefits, practical and safety advantages. Training transfer and skill retention are two main indicators of training effectiveness, with positive training transfer being the most desirable metric. Measuring these, however, require access to high-fidelity simulators or real aircraft. Since such assets were unavailable to this research study, the thesis instead focuses on a direct comparison between simulator sub-variants.

Training effectiveness and simulation fidelity are often linked together, with research focusing on replicating the real-world as accurately as possible. However, low-fidelity simulators have proven to deliver more cost-effective training than their higher-fidelity counterparts. The current issue lies in the mismatch and confusion of simulator types across aviation authorities worldwide. This is overcome by adopting the redefined ICAO 9625 simulator qualification specification process, which maps simulation features/fidelity levels driven by the intended training task.

To enable the simulator sub-variant comparison, this thesis evaluates the utility of amplified head rotations validate its potential value. This is done by conducting a series of experiments that gathers both objective and subjective performance measurements. A literature review of published studies and task analysis resulted in the selecting a basic, visual flying circuit as the mission scenario. Besides using aircraft telemetry as an objective performance measure, workload and simulator sickness were also taken into account. NASA-TLX was found to be the most popular and validated workload assessment tool and therefore used. The objective secondary task method, however, was also selected to provide an additional workload metric to correlate against the subjective NASA-TLX results.

Immersive virtual environments such as fixed-based simulators without motion cueing can induce simulator sickness. Symptoms are categorized into three classes:

disorientation, oculo-motor and nausea. There are many contributing factors ranging from user, system and task characteristics that cause simulator sickness. The best current practice to minimize symptoms therefore is to carefully consider these factors during the simulator system design and verify it before operational use. Many methods for measuring simulator sickness exist, from physiological measurements to self-report assessments. This thesis uses the subjective SSQ test as it is the most popular and validated tool by researchers.

Before commencing the experimental method, there are requirements and constraints to be considered. In order to achieve a controlled experiment, sufficient data needs to be generated to enable statistical analysis and the experimental design has to minimize biases. Practical constraints were mostly time and resource limits. Limited access to large numbers of pilots meant that only predominantly novice and low-hour private pilots from the local flying club were available for experiments.

Chapter 4

Experimental Design

With the research objectives stated in Section 1.3.2 and input from Chapter 3, two experiments were designed to collect results to address the research questions. Section 4.1 provides an initial overview of how the experiments were prepared and the common procedures used. Section 4.2 describes the first experiment, which served as a baseline reference corresponding to stage one of the utility evaluation (Section 3.3). This establishes the practical usability of amplified head rotations in its most basic form compared to a legacy non-augmented triple-screen flight simulator. To obtain a fair comparison, the emphasis is on selecting an experimental task that does not impede the non-augmented setup.

The second experiment described in Section 4.3 covers the second research objective of investigating scalability by comparing three augmented displays to study the compounding effects when larger displays and FOVs are used. This was stage two (Section 3.3) of the evaluation process to determine performance improvement of the system itself. By not having any task restrictions posed on the flying task by a non-augmented display, this second experiment can fully utilize the freedom the head augmentation provides in supporting a more complex visual flying task, a feat undocumented before. Finally, Section 4.4 concludes by providing the design overview of the experiments and the tools and statistical techniques used to verify the results.

To aid navigating through the thesis, Figure 4.1 shows the thesis structure once more and in particular how this chapter is part of the iterative loop of method, experimental design and system design/realization.

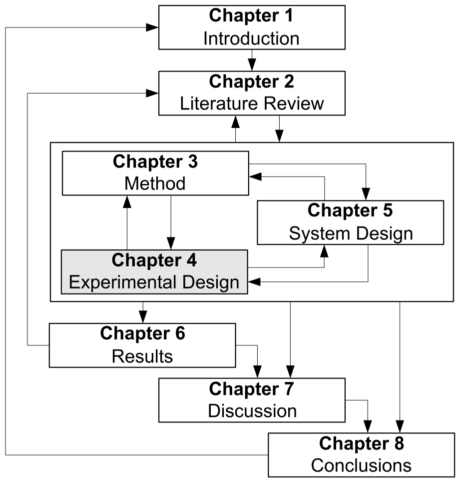


FIGURE 4.1: Chapter 4 in thesis roadmap

4.1 Overview

THIS section provides an overview of the common experimental design/procedure for both experimental scenarios together with the participant selection and screening method.

4.1.1 Participants

Since the experiment requires familiarity with flying and landing a simulated airplane, it is vital that participants either have affinity or actual flying experience to ensure that the trials are not advertently biased due to poor piloting skills nor advanced piloting skills, such bias will cause unwanted reverse training transfer as discussed in Section 3.1. Participants should also not have had any prior virtual reality experience with the head augmentation configuration utilized in these experiments. Participants were thus recruited from the general university population (in particular members of Loughborough Students Flying Club/Loughborough Students Union Gliding Club), pilots flying from East Midlands Airport (Donair flying club and airline crew) and acquaintances/contacts, with prior relevant experience.

With small sample sizes being very common in aviation studies [131], no other restrictions were imposed on the selection of participants except for the age bracket, which was between 18-65 years old. To further ensure participants had the required minimum piloting skills to produce valid results, a short checkride based on the private pilot practical test standards [113] was administered in the simulator prior to commencing the actual experiments. Checkride details are provided in Section 4.1.3.

Participants, upon expressing their interest in the simulator trials, received the participant information sheet (Appendix C) through their email, explaining the experiment, what was expected of them and their available course of actions. Voluntary consent to experiment participation was further collected by having participants sign the ethical consent form (Appendix D) on the day of the experiment. Participants could withdraw from the experiment at any stage without hindrance.

4.1.2 Design Approach

The survey (Section 2.2) of current non-motion flight simulators identified the three-channel display monitor (Figure 4.2(a)) or large projector (Figure 4.2(b)) setup to be most prevalent (Section 2.3.2). A progressive two-stage approach based on the utility evaluation process described in Section 3.3 was used in order to gather data. First, a baseline controlled experiment was designed to compare the three-display simulator versus the most basic form of an augmented setup with just a single display. This also served as a verification trial to further improve augmentation integration in the follow-up experiment.

Once completed, the next step was to integrate head augmentation to higher visual display fidelity levels in order to understand its scalability. By solely comparing between the augmented systems, this capability unlocked more complex scenarios and representative flying tasks that would fully test the extended mission training scenarios that the new technology brings. This second experiment would also give insight into what display fidelity is required to fully benefit the use of head augmentation and gather evidence if larger displays automatically scale in terms of usefulness.

(a) Triple display configuration [51]

(b) Generic 3-channel large display simulator

FIGURE 4.2: Representative simulator display configurations

Figure 4.3 shows the experimental design overview which is applicable to both experiments. Each of the experiments was independent, and all display configurations (main independent variable) were performed by the same participants in a single session. In case of the first experiment there are only two levels so the flowchart points out that the debrief follows the second (last) display and terminates. The scalability experiment contained three displays.

Prior to the experiment, participants had the opportunity to read the information sheet and sign the consent form. They then received an oral briefing on the experiment (this was to ensure they had no prior detailed knowledge of the experiment

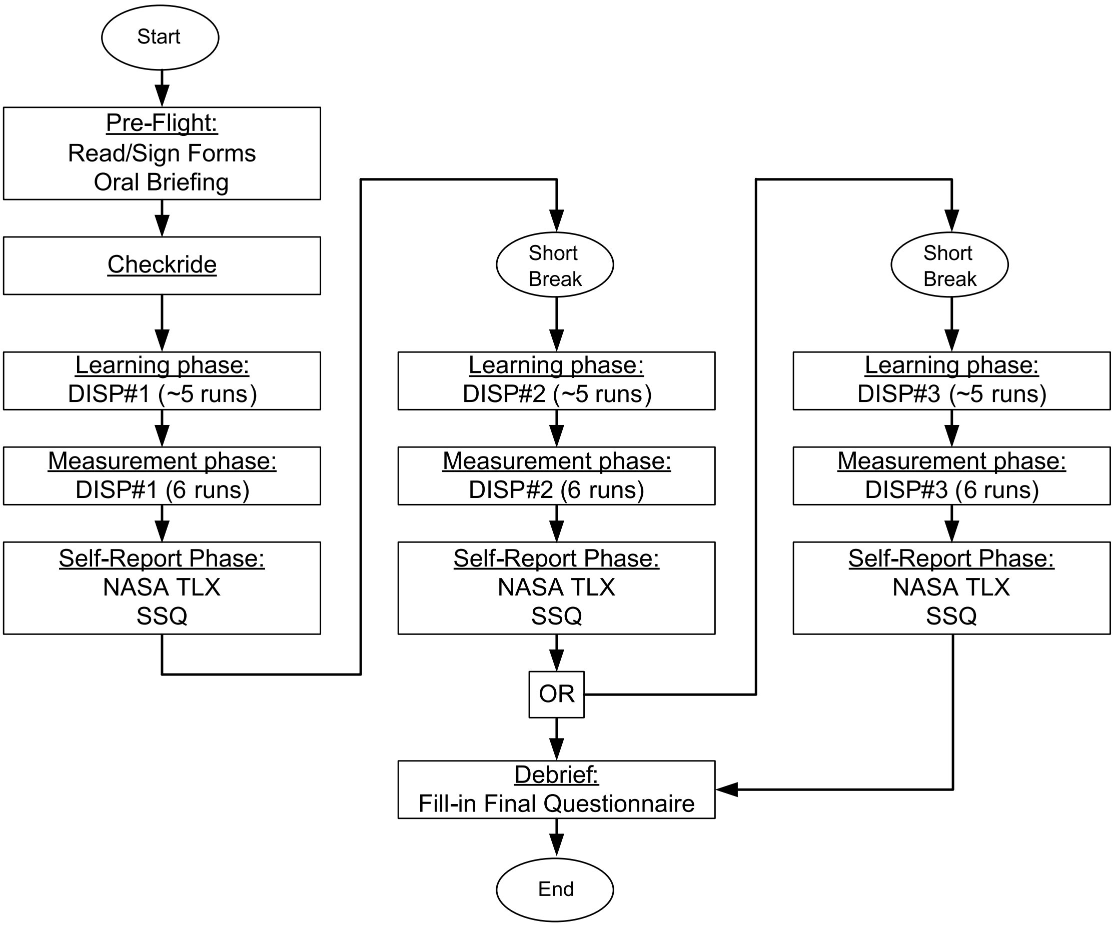


FIGURE 4.3: Experimental design flowchart

to prevent any self-administered training) and undertook the checkride. Upon passing the checkride, participants started the actual experiment.

Display order was randomized to ensure no learning biases would occur. Upon loading up the first assigned display configuration, participants underwent the learning phase where they familiarized themselves with the scenario and task; this usually took about five runs where participants performance reached a consistent result.

In the measurement phase, a total of six runs were flown and data logged. This amount of runs was deliberately chosen to prevent pilot boredom and fatigue. After completing the measurement runs, participants filled out a NASA-TLX rating in digital form (Appendix E) and the simulator sickness form (Appendix F). A short break of five minutes was given until the next display configuration was imposed to alleviate pilot fatigue and allow full recovery from any sickness symptoms. After the last display configuration, participants completed a post-experiment questionnaire (Appendix G) which then ended the experiment. Table 4.1 lists a

TABLE 4.1: Experimental procedure

Step	Description	Time duration (minutes)
	Pre-Experiment	
1	Email Experiment Briefing	-
2	Email Consent Forms	-
	Experiments	
1	Sign/Collect Consent Forms	0.5
2	Display order assignment	0.5
3	Listen Oral Briefing	5
4	Perform Familiarization Checkride	5
5.a	Perform DISPLAY 1 experiment	25
5.b	Perform DISPLAY 2 experiment	25
5.c	Perform DISPLAY 3 experiment	25
6	Complete NASA-TLX form	5
7	Complete SSQ form	1
8	**SHORT BREAK**	5
9.a	Perform DISPLAY 1 experiment	25
9.b	Perform DISPLAY 2 experiment	25
9.c	Perform DISPLAY 3 experiment	25
10	Complete NASA-TLX form	5
11	Complete SSQ form	1
12	**SHORT BREAK**	5
13.a	Perform DISPLAY 1 experiment	25
13.b	Perform DISPLAY 2 experiment	25
13.c	Perform DISPLAY 3 experiment	25
14	Complete NASA-TLX form	5
15	Complete SSQ form	1
16	Complete final questionnaire	5

more detailed step-by-step procedure of the experiment together with indications of time duration.

4.1.3 Checkride Details

To ensure participants in the experiments had a common, minimum piloting skill set to perform the experiment without detrimental results for fair comparison of the simulator systems, a short checkride based on the Private Pilot Practical Test Standards [113] was developed and administered. Due to the anticipated limited availability of licensed pilots, the checkride delivers a means to qualify unlicensed participants to a similar standard. Appendix H provides excerpts from the practical standards which formed the basis for the following checkride procedure.

In the checkride, participants started out on the active runway with the aircraft preset in the same weight and configuration which they would perform the experiment. They were then asked to perform a take-off with no gear/flap configuration changes once airborne, then to establish a straight and level flight at a pattern altitude of 1000 feet as indicated on their altimeter and at 85 knots indicated airspeed with the following criteria:

Checkride: straight-and-level flight

Maintain altitude, ±100 feet; heading, ±20°; and airspeed, ±10 knots.

They then had to demonstrate their capability of performing turns using up to 30° of bank to change their heading to any new heading requested by the researcher within the following parametres:

Checkride: turns

Maintain turn entry altitude, ±100 feet, airspeed, ±10 knots, bank, ±5°; and roll out on the exit heading, ±10°.

To check whether participants were capable of flying a traffic pattern around a ground reference, they were tasked to fly a rectangular course around a designated runway at pattern altitude. This was evaluated against the following set of requirements:

Checkride: ground reference manoeuvre

- Exhibit satisfactory knowledge of the elements related to a rectangular course.
- Plan the manoeuvre so as to enter a left or right pattern, at pattern altitude at an appropriate distance from the selected runway.
- Divide attention between airplane control and the ground track while maintaining coordinated flight.
- Maintain pattern altitude, ±100 feet, maintain airspeed, ±10 knots.

Upon successful completion of these flight manoeuvres, participants were tasked to land the aircraft at the nearest runway available to demonstrate their landing competency. Due to the complicated nature of landing technique and skills, participants passed the landing test if the aircraft was put on the ground on the runway without overstressing the airframe or triggering a crash in the simulator.

If participants failed to perform more than two manoeuvres in the checkride within the criteria, they were excluded from further participation in the experiment, all successfully screened participants proceeded immediately on to the actual experiment phase.

4.2 Experiment 1: Baseline Comparison

In this exploratory experiment, the goal is to obtain data to enable a quantifiable comparison between the common, triple monitor fixed-based flight simulator versus the most basic implementation of amplified head rotations on a single monitor. This allows for an exploratory validation of how the novel virtual reality implementation stacks up in terms of pilot performance, workload and simulator sickness by searching answers to the first research objective as stated in Section 1.3.2. The results gathered in this experiment then forms the basis to extrapolate the augmentation system onto larger displays in the next experiment.

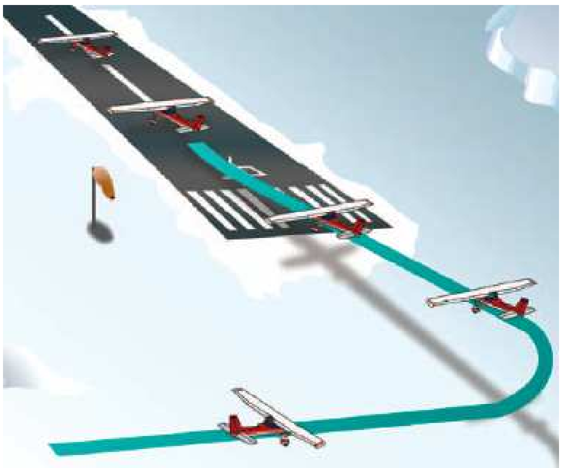


FIGURE 4.4: Basel leg and final approach adapted from [114]

Participants were tasked to land an aircraft airborne from the base leg in a visual traffic circuit pattern, as shown in Figure 4.4. With the runway being the primary visual reference cue, this starting point was specifically chosen to avoid handicapping and biasing the triple static monitor configuration which served as the control condition. Setup with a static 120° FOV on the triple displays, placing the aircraft on the base leg ensures that the runway is visible during the whole manoeuvre. Participants took control of the aircraft once the experiment started, then performed a left hand descending turn, lined-up on the runway and performed a full stop landing.

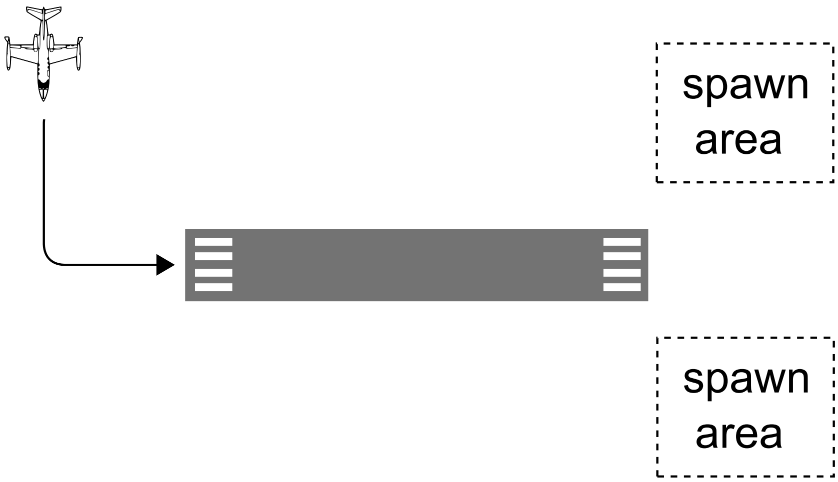


FIGURE 4.5: Popup blimp spawn areas

During real flight, pilots are supposed to maintain a visual lookout for other traffic in the area to avoid mid-air collisions. This lookout is the secondary task during the experiments which participants must do. A popup blimp in the vicinity of the runway was chosen to simulate this aspect rather than having other aircraft flying around adding to the complexity of the scenario. When the aircraft arrives at a preset distance from the runway threshold, a timer is triggered which then after a random amount of time spawns a single blimp randomly in either of the two spawn areas as seen in Figure 4.5. The participant presses a button on the stick when a blimp has been spotted. Having this simulated airborne traffic ensures that pilots maintain their visual lookout without slacking since they do not know when and where the blimp might popup.

4.2.1 Independent Variables

As mentioned before, the experiment had one within-subject independent variable: simulator display configuration (DISP) with two levels: single augmented and triple non-augmented, shown in Table 4.2.

TABLE 4.2: Experiment #1 cases

Case	Display	Head augmentation
1	Triple	No
2	Single	Yes

4.2.2 Dependent Variables

The design of this experiment enabled collection of both quantitative and qualitative data. As discussed in the assessment methods review, objective measures collected such as primary task performance, workload together with subjective self-reporting measures provide a comprehensive insight for assessing human performance parametres across the two displays.

Performance measures. To obtain flight technical performance data, the entire flight was divided into separate flight segments. Each flight segment then had its own performance measures as follows:

1. *Base-to-final turn*
   (a) *Maximum bank angle (degrees):* the maximum bank angle the pilot used to perform this turn was captured as a metric to gauge the tightness/aggressiveness of the turn.
   (b) *Turn start centerline distance (metres):* by capturing when the turn was initiated as a distance from the (extended) runway centerline, a consistent metric is obtained to compare across displays.
2. *Approach*
   (a) *Glideslope deviation (degrees):* the root-mean-square (RMS) of the vertical deviation from the 3 degree nominated glideslope allows to quantify how well participants were capable of tracking the ideal vertical descent flightpath.
   (b) *Cross-track error (metres):* this RMS metric tracks a participant's lateral deviation from (extended) runway centerline, a measure of how well the aircraft is lined up with the runway.
3. *Touchdown location*
   (a) *Longitudinal (metres):* touchdown point measured along runway length from threshold, this provides a metric if a participant landed too far or too short.
   (b) *Lateral (metres):* touchdown point distance from runway centerline, measures how far the aircraft landed left or right of the centerline.

Workload. Although the aforementioned flight technical performance measures are a good primary workload indication as stated in Section 3.5, further workload validation data were achieved by sampling the secondary task performance of blimp spotting using two measures: blimp detection rate and detection time. Detection rate is the percentage amount of blimps spotted whereas detection time is the elapsed time in seconds between blimp spawn and button press. Additional workload data were acquired through the subjective workload measures provided by the results of the administered NASA-TLX forms.

Simulator sickness. Simulator sickness data were obtained via participants self-reporting of the SSQ forms, the best measurement method as discussed in Section 3.6.

4.2.3 Experiment Setting

Apparatus. The experiment was conducted in the Advanced Cockpit Facility, Figure 4.6, at the Advanced VR Research Centre (AVRRC) of Loughborough University. The laboratory consisted of an experiment area with a fixed-base simulator and control desk. Detailed information on the system development and construction of this research simulator is provided in Chapter 5 documenting the technical specifications and systems engineering process based on Section 3.2.2.

FIGURE 4.6: AVRRC Advanced Cockpit Facility

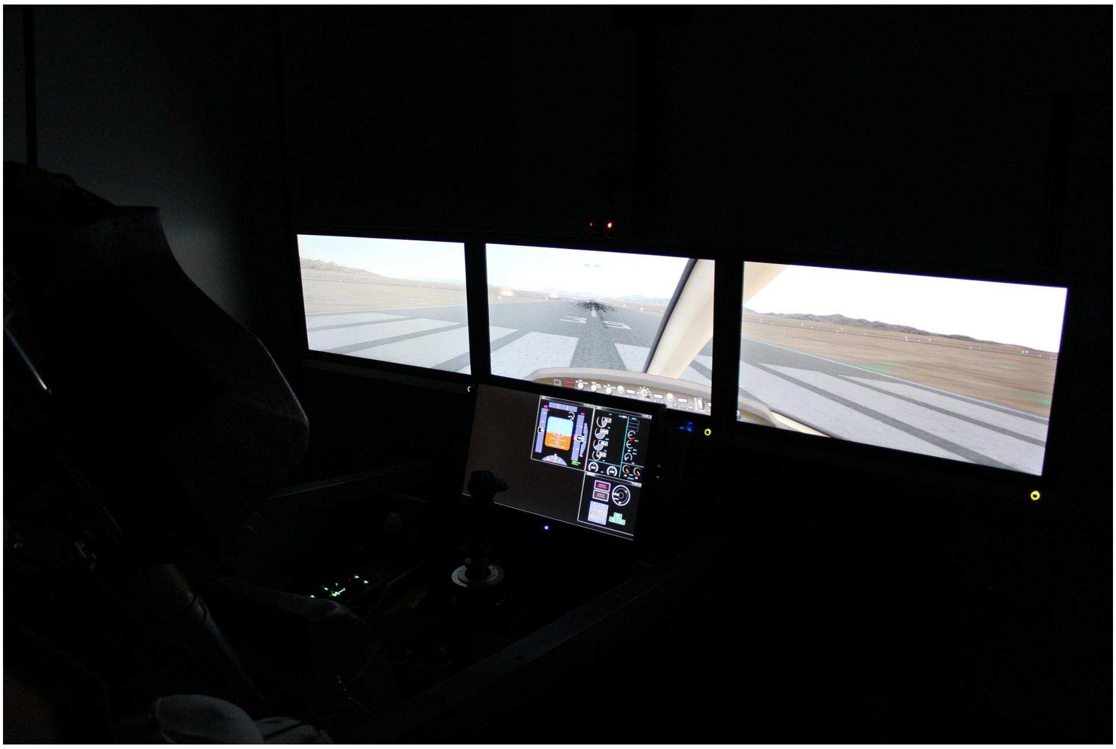


FIGURE 4.7: Triple display simulator mode

Displays. In this experiment, the participant was seated in a (fighter type) cockpit in a darkened environment, shown in Figure 4.7 and Figure 4.8. Directly in front of the pilot were three 27″ 1080p LCD screens in landscape mode. Together, these formed a single, large display area of (bezel-corrected) 6160×1080 pixels providing a virtual 120° horizontal FOV. The left and right displays were disabled for the single augmented configuration, with approximately 40° remaining horizontal FOV. A single 22″ 1080p LCD touchscreen provided flight (instrument) displays (so in case of the single augmented display, this was redundant as the virtual cockpit instruments could also be viewed on the single monitor by physically looking down).

The instrument displays provided by the bottom screen are shown in Figure 4.9. These were driven by the XHSI [132] avionics suite and offered three sub-displays: an electronic Primary Flight Display (centre top) showing aircraft attitude, airspeed and heading information representing a generic layout found in many modern aircraft. To its right is the engine display offering information on engine turbine speed, fuel flow and temperature. Right below is the aircraft configuration/status display, showing the status of the flaps, landing gear and wheelbrakes. Although the pilot did not have any control over this, showing this screen enables confirmation of the correct aircraft state in normal operations to support realism and

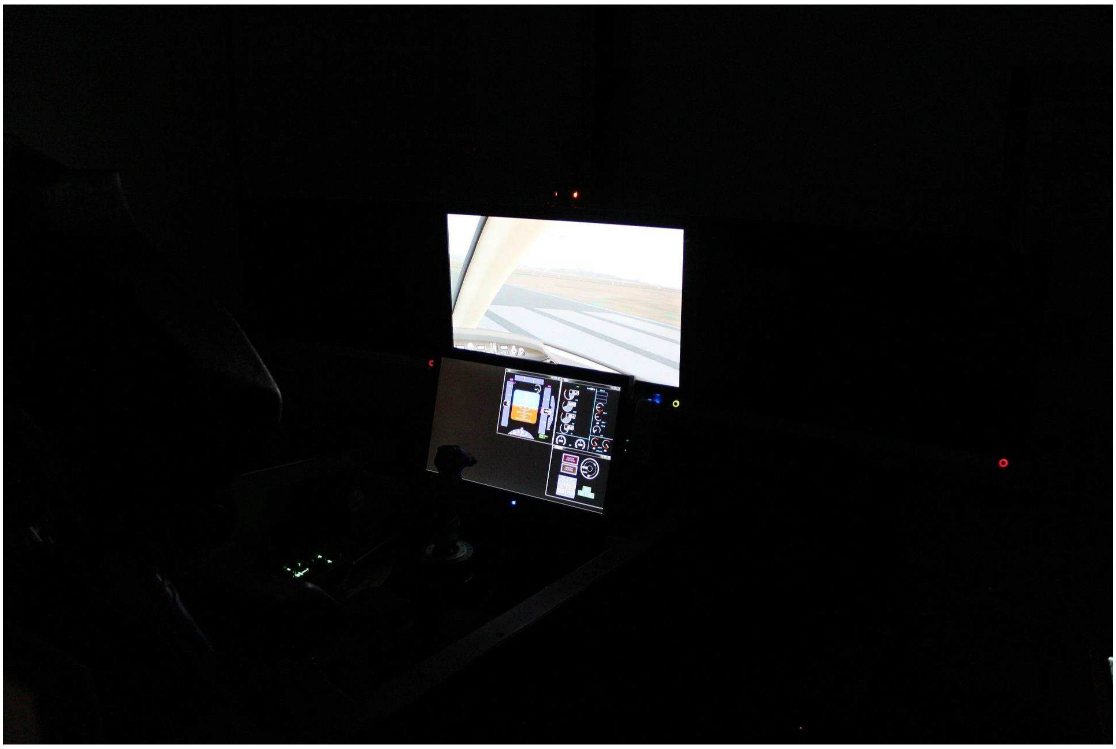


Figure 4.8: Augmented single display simulator mode (note virtual view panned to the right)

confirm the pre-landing checks.

Experiment conditions. Aircraft control was available via a right-handed, mechanically, spring loaded, passive centre stick (elevator and aileron). The throttle quadrant was available on the left-side panel with only the throttle levers being used. The flight simulation software used was X-Plane 9.70 [133], with the default CirrusJet aircraft model selected. The aircraft was preset in the landing configuration with flaps and gear down, trimmed for 85 knots indicated airspeed. Wheelbrakes automatically activated upon touchdown, therefore participants did not have to operate the trim, rudders, brakes, flaps and gear to reduce the amount of pilot input variables to a bare minimum required to perform the task. The unavailability of rudder control also negated any advanced piloting techniques that participants might employ in order to salvage last-minute poor performance.

An infrared head-tracker (TrackIR 5) [134] measured participant head position/orientation at 120 Hz which controls the virtual pilot head camera in the simulation. The darkened environment was to ensure that no stray light or reflections could

Figure 4.9: Heads-down instrument displays [132]

impede the operations of the head-tracker and also facilitated participant immersion into the virtual environment by letting them focus on the monitors without distractions from the real world adjacent to the simulator.

The amplified head rotations were mapped to this single-display configuration through a custom-designed profile, shown in (Figure 4.10). Although the device manufacturer supplied generic templates, each profile must be adapted to the existing simulator displays as these can vary in size. With no guidelines or reference works on tweaking this, this profile was developed through in-house empirical testing with 20 staff and students to produce as an initial reference which could be further improved as more experiments were completed.

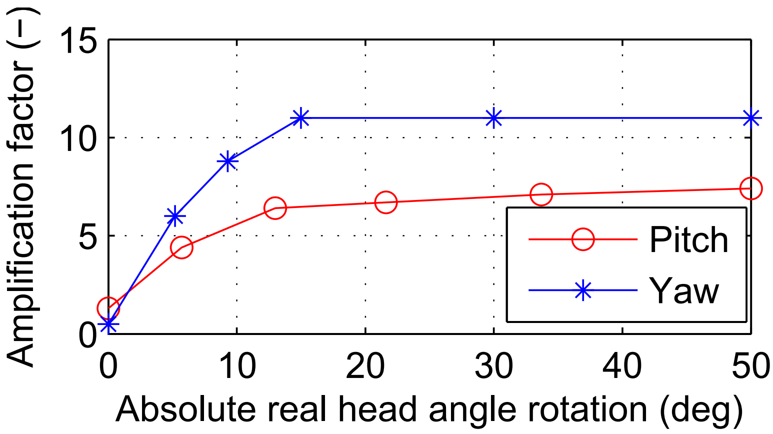


FIGURE 4.10: Head amplification single display profile

4.2.4 Procedure

Each experiment took about two hours and consisted of two phases. The first phase was considered to be a *training phase* in which the participants received an oral briefing and became familiar with the simulator to get accustomed to landing with the various simulator configurations. Participants took three to five runs before setting a consistent ground track. This was to eliminate learning curve effects so the actual measurements would be able to record consistent tracks. The second phase was the *measurement phase* consisting of 6 runs to be flown per display type. A single experiment run lasted around three minutes. A screenshot of the virtual world environment is shown in Figure 4.11. After completing the runs on a display, the participant filled out the self-report NASA-TLX workload sheet, digital format shown in Appendix E and a SSQ form, with the online form of Appendix F. There was a short break in between, this allowed participants to stretch their legs and prevent fatigue/boredom as well as making sure any simulator sickness symptoms expired prior to the next display trial. Finally, participants filled out a general questionnaire at the end of the experiment.

4.3 Experiment 2: Scalability

The first experiment described in Section 4.2 was of exploratory nature to document head augmentation as an alternative solution to the common, static flight simulator visual displays. In this second experiment, the second and last research objective of Section 1.3.2 is answered by evaluating how this augmentation scales with (higher) levels of display fidelity, i.e. when more or larger size displays are

FIGURE 4.11: Screenshot of virtual simulation world

used in the simulator to provide larger FOVs. This was done by enabling augmentation on all displays, there were no more non-augmented setups in this evaluation. The reasons for this were two-fold: 1) this allowed for a fair comparison between display fidelity levels with no mismatched biasing if a non-augmented system was also tested in between and 2) this eliminates (visual) scenarios restrictions that a non-augmented setup even with larger FOVs would still impose so that the expanded capability of the head augmentation can be properly demonstrated. The end goal is to provide answers and recommendations on the questions of how effective head augmentation is as a retrofit to existing static simulator displays and does lower or higher-display fidelity levels together with augmentation make a significant impact on operator performance.

Participants were tasked to take control of an airborne aircraft starting downwind in a visual traffic circuit pattern, as shown in Figure 4.12. The primary task was therefore to complete the circuit pattern by flying downwind and performing two turns (indicated by Turn 1 for downwind-to-base and Turn 2 for base-to-final) prior to a full-stop landing. A secondary visual search task for airborne traffic like in the first experiment was retained, however, with the longer flight duration and ground path covered the amount of simulated traffic was also increased from just a single blimp to three blimps.

The blimp sequence was as follows: on downwind at the start of the run, the aircraft would trigger a timer after which the first blimp would spawn in area 1 at a randomly selected time interval. Once further downwind, the first blimp would

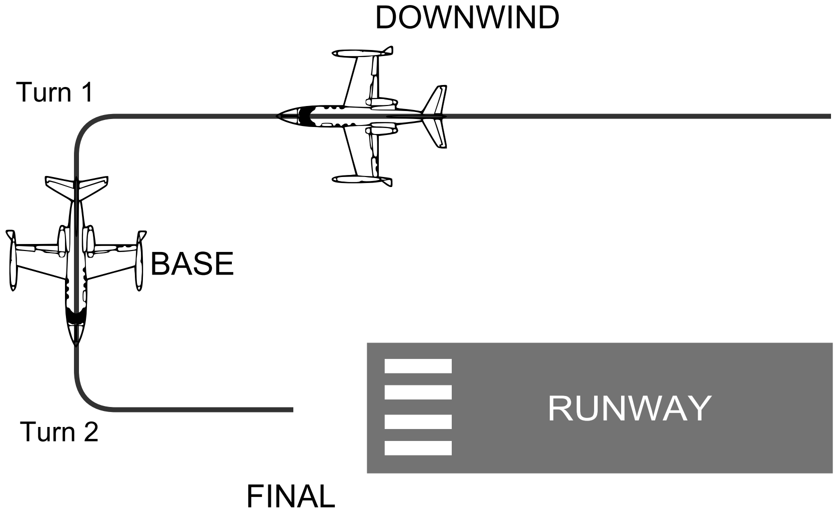


Figure 4.12: Downwind-base-final traffic pattern

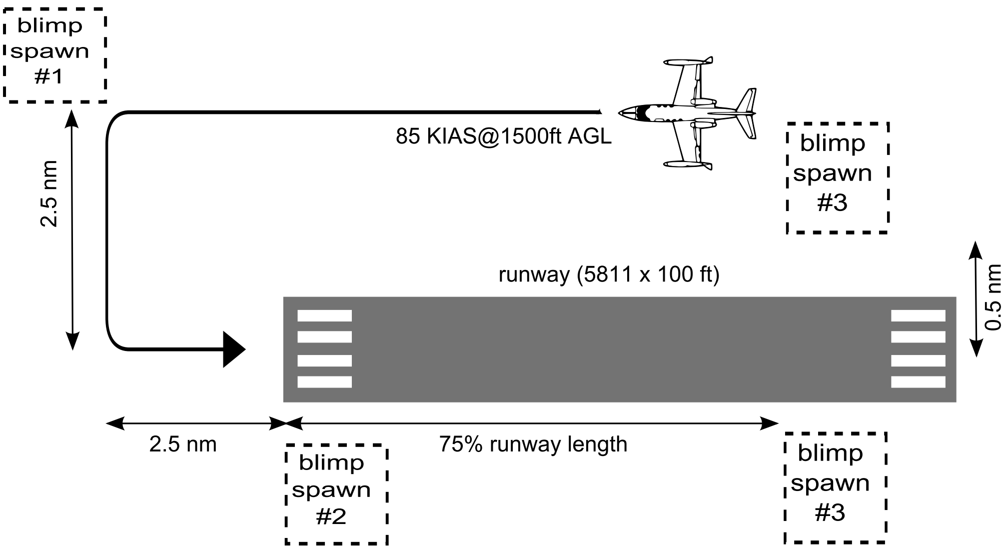


Figure 4.13: Downwind traffic pattern with three blimp spawn areas

disappear and another timer triggered for when Blimp 2 would spawn in its box. A similar procedure applied to Blimp 3, however the final blimp had two box areas to randomly choose to spawn in. Timings for the blimps were selected such that Blimp 1 would be detectable on the downwind leg only, Blimp 2 near the end of the downwind left through to base leg. Blimp 3 would only spawn whilst transiting from base leg to touchdown. Even though participants might have remembered the spawn sequence, the timings in combination with the spawn boxes still made it sufficiently challenging so participants had to actively visually search in order to detect a blimp.

4.3.1 Independent Variables

This experiment again had only one within-subject independent variable: simulator display configuration (DISP) but now with three levels: single augmented, triple augmented and projector augmented, with characteristics listed in Table 4.3. Figure 4.14, Figure 4.15 and Figure 4.16 show pictures of the corresponding configurations actually used in the experiment.

TABLE 4.3: Display configuration screen characteristics

Display	HFOV × VFOV (deg)	Physical size
Single monitor (SGL)	33×23	27″ single
Triple monitor (TRP)	100×23	27″ triple landscape
Triple projector (PRO)	112×29	3.2m curved radius, 2m height

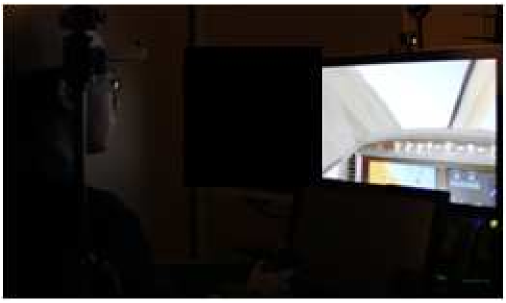


FIGURE 4.14: Single augmented display

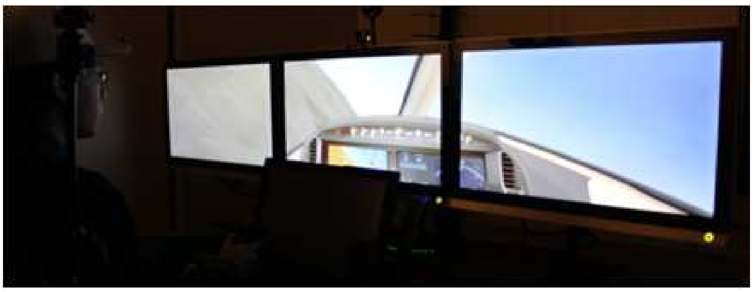


FIGURE 4.15: Triple augmented display

FIGURE 4.16: Augmented projectors display

4.3.2 Dependent Variables

Extending the first experiment, data were similarly collected for both quantitative and qualitative measures. Having increased the scenario and task complexity, the objective performance measures were expanded to cover the additional downwind flight phase and extra dependent variables associated with them.

Performance measures. Flight technical performance data were again captured and separated per flight segment. Expanding the traffic pattern to start from downwind meant the two additional segments (downwind and the downwind-to-base turn) were incorporated. Also with lessons learned from the first experiment and the anticipated higher task workload, an additional dependent variable was added into most flight segments: deviation from reference indicated airspeed. Furthermore, runway alignment error was added as a stand-alone dependent variable after the base-to-final turn to gauge how well participants had finished the turn prior to final approach. The following lists all performance variables per flight segment:

1. *Downwind*
   (a) *Altitude deviation (metres):* root-mean-square (RMS) of altitude deviation from pattern altitude of 1500 feet during downwind leg.
   (b) *Cross-track error (metres):* RMS lateral deviation from designated ground track 2.5 nautical miles parallel to runway centerline.
   (c) V_{REF} *error (knots):* RMS deviation from approach reference indicated airspeed of 85 knots.
2. *Turn 1 (Downwind-to-base)*
   (a) *Maximum bank angle (degrees):* root-mean-square (RMS) of maximum bank pilot used to perform the turn to base.
   (b) *Turn starting distance from runway threshold (metres):* provides comparison of how far downwind pilots started the turn to base.
   (c) V_{REF} *error (knots):* RMS deviation from approach reference indicated airspeed during Turn 1.

3. *Base leg*
    (a) *Altitude deviation (metres):* root-mean-square (RMS) of altitude deviation from 1500 feet during base leg, indicates pilot descent strategy to manage energy.
    (b) *V_{REF} error (knots):* RMS deviation from approach reference indicated airspeed.
    (c) *Heading error (degrees):* RMS of base leg magnetic heading deviation from ideal orthogonal runway heading.
4. *Turn 2 (Base-to-final)*
    (a) *Maximum bank angle (degrees):* root-mean-square (RMS) of maximum bank pilot used to perform the turn to final.
    (b) *Turn starting distance from runway threshold (metres):* provides comparison of how far downwind pilots started the turn to final.
    (c) *V_{REF} error (knots):* RMS deviation from approach reference indicated airspeed during Turn 2.
5. Initial line-up
    (a) *Runway alignment error (degrees):* the angle formed between the extended runway centerline and the aircraft track [135] on upon roll-out from base-to-final turn. A higher value indicates more error, requiring larger corrections for line-up which is less desirable.
6. *Final approach*
    (a) *V_{REF} error (knots):* RMS deviation from approach reference indicated airspeed.
    (b) *Glideslope deviation (deg):* RMS of vertical deviation from 3 degree glideslope.
    (c) *Cross-track error (metres):* RMS lateral deviation from (extended) runway centerline.
7. *Touchdown location*
    (a) *Longitudinal (metres):* touchdown point measured along runway length from threshold.
    (b) *Lateral (metres):* touchdown point distance from runway centerline.

Workload. Primary task workloads were already inherently measured by the performance measures. In this experiment, additional dependent variables were added in an attempt to cover any performance differences across the displays. The secondary task of blimp spotting was now extended to three blimps with the same variables: detection rate and detection time. Further subjective workload data were gathered by means of the self-report NASA-TLX questionnaires.

Simulator sickness. Simulator sickness data were obtained from participants self-reporting of the usual SSQ forms.

4.3.3 Experiment Setting

Apparatus. The experiment was conducted at the AVRRC of Loughborough University. The single and triple display cases used the Advanced Cockpit Facility setup, previously shown in Figure 4.6. To facilitate the projector display case, the Virtual Engineering Centre (VEC) with its triple immersive projection system was reconfigured to support the experiment. This facility is directly adjacent to the Cockpit Facility and is on the same computer network. The same mission control station is used to supervise the experiments using the projectors.

Displays. In this experiment, the displays exclusively provide the outside visual world and the virtual cockpit representation with all avionics displays inherent on the aircraft in the virtual space shown in Figure 4.17. Because augmentation was available on all three display cases, participants could use their head to simply glance down to view the instruments when needed. The central lower screen was therefore disabled and not used in this experiment. Having full view interaction within the virtual cockpit also enhanced the realism and familiar environment for pilots, resulting in more presence experienced in the virtual aircraft. The usage of (virtual) vehicle references has been proven to improve operator task performance and reduce task difficulty [136–138]. Users performed both tasks on a large curved screen wall projected display and on the desktop monitors, as such in order to negate the effects of screen size and user distance to the screen, the virtual cockpits for all displays was kept at a constant visual angle [33, 139].

Experiment conditions. Aircraft control was available via a right-handed, mechanically, spring-loaded, passive centre stick (elevator and aileron). The throttle

Figure 4.17: Screenshot of the virtual cockpit

quadrant was available on the left side panel with only the throttle levers being used. The flight simulation software used was X-Plane 9.70 [133], with the default CirrusJet aircraft model selected. The aircraft was preset in the landing configuration with flaps and gear down, trimmed for 85 knots indicated airspeed. Wheel brakes automatically activated upon touchdown, therefore participants did not have to operate the trim, rudders, brakes, flaps and gear to reduce the amount of pilot input variables to a bare minimum required to perform the task. The unavailability of rudder control also negated any advanced piloting techniques that participants might employ in order to salvage last minute (poor) performance.

An infrared head-tracker (TrackIR 5) [134] measured participant head position/orientation at 120 Hz which controls the virtual pilot head camera in simulation. The darkened environment was to ensure that no stray lights or reflections would impede the operations of the head-tracker and also facilitated participant immersion into the virtual environment by letting them focus on the monitors without distractions from the real world adjacent to the simulator.

The amplification profiles for all three displays are shown in Figure 4.18. Pitch mapping for single monitor and triple monitors (TRP) was initially identical because the vertical viewing angles are the same on both configurations (TRP simply had 2 extra monitors to the side compared to SGL). Later, it was found out that there was a slight tweak necessary required at the first node-to-node gradient for increased comfort. The yaw profile for the triple monitor and triple projector

(PRO) was not surprisingly similar as their respective horizontal FOVs were very close. What Figure 4.18 also theoretically refers to is that on the extreme edge of (50,7) for pitch (SGL/TRP) this would mean looking 50° to the right physically would translate into 350° of virtual angle. A simple limiter on maximum virtual angles simulating maximum human head rotation angles in place obviously eliminated this issue.

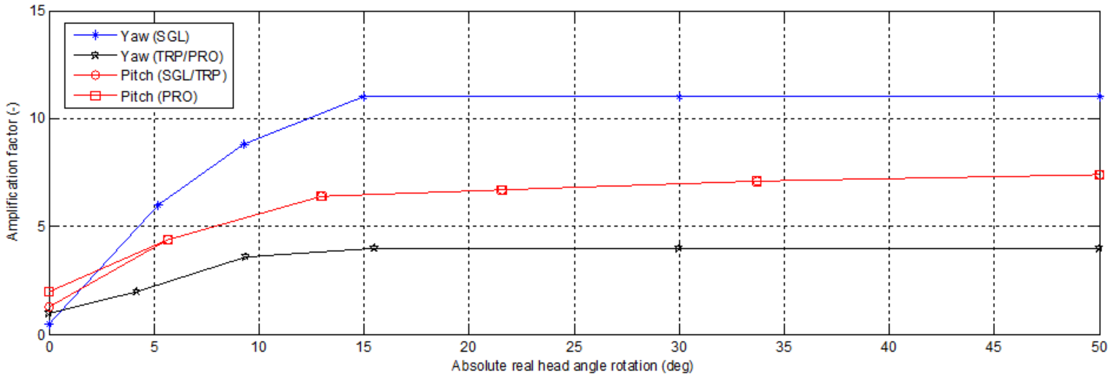


FIGURE 4.18: Amplification profile for single, triple monitor and triple projector displays

4.3.4 Procedure

Each experiment took about two hours and consisted of two phases. The first phase was considered to be a *training phase* in which the participants received an oral briefing and became familiar with the simulator to get accustomed to landing with the various simulator configurations. Participants took three to five runs before setting a consistent ground track. This was to eliminate learning curve effects so the actual measurements would be able to record consistent tracks. The second phase was the *measurement phase* consisting of six runs to be flown per display type. A single experiment run lasted around three minutes. A screenshot of the virtual world environment is shown in Figure 4.11. After completing the runs on a display, the participant filled out the self-report NASA-TLX workload sheet, digital format shown in Appendix E and a SSQ form, with the online form of Appendix F. There was a short break in between, this allowed participants to relax and rest as well as making sure any simulator sickness symptoms expired prior to the next display trial. Finally, participants filled out a general questionnaire at the end of the experiment.

4.4 Statistical Analysis

After collecting all the generated data from the experiments, a suite of software tools was used to perform the statistical analysis. IBM SPSS Statistics Version 19.0.0 was the primary software used for the analysis; with JMP 10.0.0 from SAS for visual interaction with the datasets and the generation of charts. Microsoft Excel 2007 was used for light calculations and pre-processing of the questionnaires before analysis. The NASA-TLX digital forms were online based whereas the SSQ and post-experiment questionnaires were custom made in Google Forms.

4.4.1 Statistical Method

With both experiments having only one independent variable, the appropriate statistical analysis method was determined using the flowchart in Figure 4.19. As stated in reference statistical textbooks [140, 141]: factor design checks whether the dependent variables were collected for different participants, if so then it is unrelated and if the same participants were used then the related case applies.

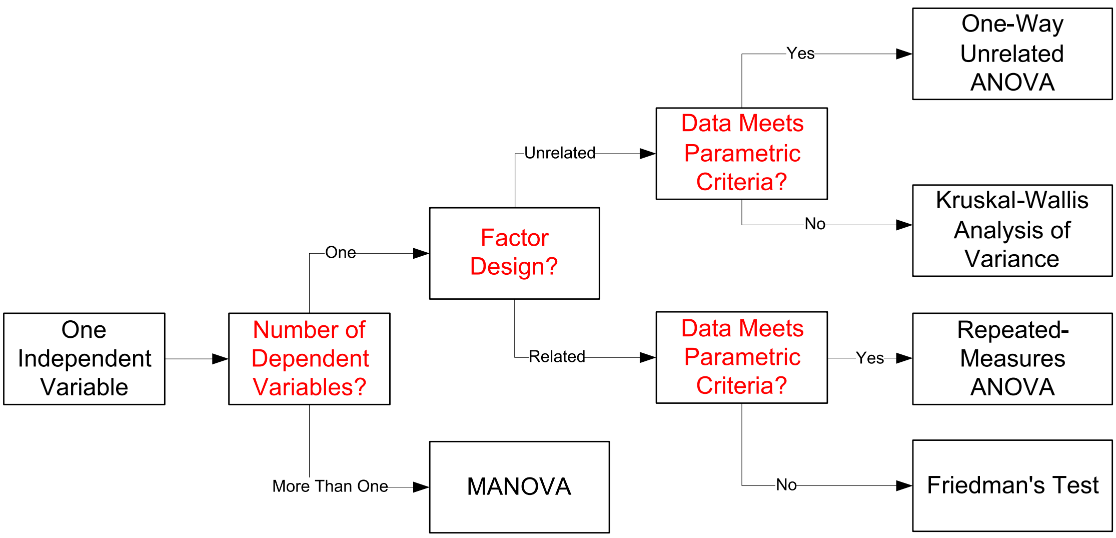


FIGURE 4.19: Statistical method flowchart for one independent variable [142]

As Vitiello [142] cited from textbooks [140, 141], the parametric criteria is valid if the collected data satisfies the following tests:

Normality. To test if the data were normally distributed, the Shapiro-Wilks test was used instead of Kolmogorov-Smirnov and the outcome considered valid if it yielded non-significant ($p \leq 0.05$) results. This test was selected because it is the most powerful normality test [143] and best in rejecting the null hypothesis at the smallest sample size (< 2000) compared to the other tests, for all levels of skewness and kurtosis of these distributions [144].

Homogeneity of variance. Levene's test was used to determine if the dependent variables had homogenous variances [141], this was true if the test yielded non-significant results ($p \leq 0.05$).

Interval data. Data with meaningful intermediate values.

Independence. All dependent variables are independently collected.

Homogeneity of covariance. Only for the MANOVA method, an additional test in form of Box's M test of equality of covariance matrices was used to determine the outcome of these criteria, and yielded true if the results were non-significant ($p \leq 0.001$).

Although the ANOVA and MANOVA statistical methods are robust to violation of normality [145], contingency courses of actions are still available when normality is violated. However, ANOVAs are still robust even for data skewness if the sample sizes are equal and large enough (> 20) and outliers removed from data [146–148]. Pillai's trace test was also found to be robust for normality violations [141, 147]. When normality is violated, an alternative is to transform the data in an attempt to satisfy the parametric tests, but this is only applicable if homogeneity of variance was also violated. If the parametric tests still failed even with the transformed data, then non-parametric tests (Kruskal-Wallis or Friedman's ANOVA) were performed instead on the original, untransformed data without correcting the significant levels due to the exploratory nature of the experiments.

With a lack of extensive reporting of complete statistical results, especially power level and effect sizes, in (aviation) research [131, 149], this thesis provides a full account of obtained statistical results by reporting effects sizes using Pearson's correlation coefficient r. Cohen's guidelines [150] was used to define r: when $r > 0.10$ is a small effect, $r > .030$ is a moderate effect and $r > 0.50$ is a large effect. Effect sizes are only reported whenever statistical significance was obtained for the ANOVA and Kruskal-Wallis non-parametric tests (with the latter calculated following a *post hoc* Mann-Whitney test). Effect sizes for Pearson's chi-squared tests are also reported but designated with the equivalent Kramer's ϕ symbol instead of r.

4.4.2 Sample Size

Before performing the actual experiments to collect data for statistical analysis in answering the research questions, suitable theoretical sample sizes can be calculated using a priori power analysis. These power calculations were performed using the software G*Power version 3.1.4 [151, 152] with the following results:

One-way ANOVA. In the first experiment, two groups of displays (augmented versus non-augmented) were compared. If initially values were selected for power = 0.80 and a significance level of $\alpha = 0.05$, then to detect a large size effect $f = 0.40$ required a sample size of 52 whereas a medium size effect $f = 0.25$ required 128. Table 4.4 shows further sample size calculations with reducing power and significance levels that could have been selected due to the exploratory nature of the experiment. Even with the most modest selection, a sample size of 24 was required.

TABLE 4.4: One-way ANOVA power analysis for two groups

Power	α level	Effect size	Sample size
0.8	0.05	0.25	128
0.8	0.05	0.40	52
0.8	0.10	0.40	42
0.7	0.10	0.40	32
0.6	0.10	0.40	24

In the second experiment, with three two groups of displays, a similar power analysis can be calculated whenever a one-way ANOVA was performed. Table 4.5 gives the calculated sample sizes with varying input values when there are three groups. Here, a sample size of 33 was required to achieve modest power and α levels *a priori* estimates.

MANOVA. The second experiment had a maximum of three dependent variables per flight segment where these were expected to be correlated, hence the choice for MANOVA. With three groups, the power analysis results to achieve a

Table 4.5: One-way ANOVA power analysis for three groups

Power	α level	Effect size	Sample size
0.8	0.05	0.25	159
0.8	0.05	0.40	66
0.8	0.10	0.40	51
0.7	0.10	0.40	42
0.6	0.10	0.40	33

power level of 0.80 with $\alpha = 0.05$ and large effect size $f^2 = 0.35$ required a sample size of 24. The large and medium effect size values for MANOVA differs from the one-way ANOVA [150]. Table 4.6 illustrates that selecting the lowest input values still required a minimum sample size of 15.

Table 4.6: MANOVA power analysis for three groups and dependent variables

Power	α level	Effect size	Sample size
0.8	0.05	0.15	51
0.8	0.05	0.35	24
0.8	0.10	0.35	21
0.7	0.10	0.35	18
0.6	0.10	0.35	15

4.4.3 Limitations and Mitigation

Considering the experimental constraints mentioned in Section 3.7, namely the scarcity of qualified pilots to serve as participants, this posed a serious risk to the statistical validity of the results with low sample sizes. An analysis of statistical power in aviation research by Ison [131] confirmed that small sample sizes were indeed common in aviation studies with a significant majority of publications being underpowered and more than half missing critical data to validate power post hoc.

Despite the theoretical objections against such studies, aerospace research is often carried out with limited resources and a set timescale to produce results. Exploratory research, especially in the development of novel technology and prototyping typically has access to even fewer resources. In this research, even with a hypothetical ample pool of qualified pilots, fiscal and practical limits prevented achieving desired sample sizes calculated in Section 4.4.2.

In order to mitigate this, the maximum available laboratory timeslots were booked to offer participants a wide range of dates to confirm their attendance. Also all participants performed all display configurations in each of the individual experiments to maximize their useful time in the simulator. Multiple measurement runs were also logged to build up the sample size keeping in mind the aforementioned learning curve effects and avoiding fatigue and/or boredom biasing the results.

Another method is to accept a larger alpha level in order to increase power, recommended for studies that have no immediate safety or large financial implications [131]. Diverting from normal practice by selecting a higher 0.10 significance level than the standard 0.05 has been suggested to be taken into consideration in this regard [153, 154], with some cases going even further by taking a 0.20 alpha level [155] as appropriate. In light of aerospace publications with the common 0.05 alpha level and an awareness of potentially being underpowered, this thesis retained this normal significance level to adhere to the limited available literature. To strengthen the findings learned from the design, implementation and analysis of the experiments, extra measure was taken to report results that could be considered borderline significant (very close to 0.05 alpha level) and trends in the data.

4.5 Summary

The experimental design used a two-stage approach to gather data for utility evaluation. The first experiment served as a baseline controlled test to compare the three-display, non-augmented simulator with the single, augmented configuration. The selected scenario was a visual traffic circuit starting from the base leg. Participants had the primary task of performing a left-handed turn on to final approach and subsequent full stop landing. They also had a secondary task to keep a visual lookout for a blimp simulating airborne traffic. Performance measures were flight technical error, subjective workload assessment (NASA-TLX), objective workload (secondary task of blimp spotting) and simulator sickness assessment via SSQ self-reporting forms.

The second experiment integrated amplified head rotations to three different display levels: single screen, triple screen and triple projectors. This scalability test also enabled a more complex flying scenario. The same visual traffic circuit pattern was used but it was extended to start earlier from the downwind leg. The number of spawning blimps increased from one to three. Collected performance metrics were similar to the first experiment.

Due to access and availability, participant selection was limited to pilots from the local flying club. To avoid bias in pilot skills, a checkride based on the PPL standards was administered to ensure everyone had the basic skills to successfully accomplish the experimental task. Prior to the measurement phase during the experiments, participants had the opportunity to familiarize themselves with the simulator and head-tracker in the training phase. This took about 10-15 minutes per individual to achieve a consistent performance level. A total of eleven pilots participated in the first experiment and nine in the second.

Chapter 5

System Design

This chapter documents the system design of the research flight simulator built to support the experiments proposed in Chapter 4. Knowledge gained from Chapter 2 and Chapter 3 highlighted the importance of deriving simulator features based on requirements following the ICAO 9625 process explained in Section 3.2.2. Hence, obtaining experimental results from conducting trials in a simulator qualified to these standards would add more credibility and ease of verification on other qualified devices. To successfully implement this process in this research, Section 5.1 explains why systems engineering is fundamental in the simulator system design.

The requirements analysis in Section 5.2 and functional analysis in Section 5.3 highlight the iterative nature of the systems engineering process. Since the simulator also supported two other projects within the AVRRC research group, input requirements from those projects resulted in the collaborative final functional requirements listed in Section 5.2.2. With common resource limitations and stringent timing constraints, risk management was also a collaborative effort as outlined in Section 5.4. This chapter culminates with Section 5.5 presenting the final simulator design after completing the design synthesis loop and verification/compliance. Figure 5.1 restates the role of this chapter in the overall thesis.

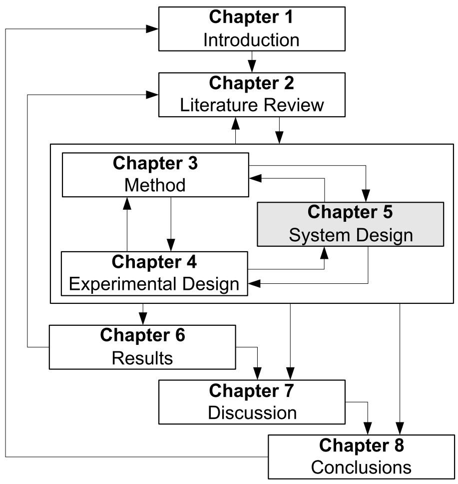


FIGURE 5.1: Chapter 5 in thesis roadmap

5.1 Simulator Engineering Approach

In order to facilitate the experimental trials drawn up in Chapter 4, a research flight simulator had to be designed and constructed. Since a research flight simulator has to support numerous, varying research projects over time, it had to have a robust architecture to cope with uncertainties, flexibility and sustainability. To deal with changes and modifications/addition of new components over its operational lifespan, one needs to approach its design from a holistic point of view in engineering systems [156]. Systems engineering has been successfully applied to products as varied as ships, aircraft, environmental control, urban infrastructure, automobiles, computer hardware and software [157, 158]. It has been shown that investing into systems engineering during projects improves development quality [159].

In general, systems engineering approaches have four characteristics: holistic, multidisciplinary, integrated/value-driven, and long-term/life cycle oriented [158]. The holistic approach considers the full technical system, including technical performance criteria as well as potentially non-technical concerns (human factors or societal impacts). The multidisciplinary nature reflects the complex nature of the topic, in the case of flight simulation involves engineering, computational, and social sciences to cover the human operator. It is also integrated and value-driven by considering the needs and interests of all customers and stakeholders. Finally, systems engineering is focused on the long-term or life cycle of the system and takes into account the cradle-to-grave life of the system for economic and sustainable concerns of today's world.

INCOSE [157] sum up the systems engineering approach in the following way: "the benefits of systems engineering include not being caught out by omissions and invalid assumptions, managing real-world changing issues, and producing the most efficient, economic and robust solutions to the need being addressed. By using the Systems Engineering approach, project costs and timescales are managed and controlled more effectively by having greater control and awareness of the project requirements, interfaces and issues and the consequences of any changes."

5.1.1 Systems Engineering Process Overview

In a document produced by DoD Systems Management College in the USA [160], it states: "The Systems Engineering Process (SEP) is therefore a comprehensive, iterative and recursive problem solving process by integrated teams. It transforms needs and requirements into a set of system product and process descriptions, generate information for decision makers, and provides input for the next level of development. The process is applied sequentially, one level at a time, adding additional detail and definition with each level of development. The process includes: inputs and outputs; requirements analysis; functional analysis and allocation; requirements loop; synthesis; design loop; verification; and system analysis and control."

The fundamental systems engineering activities consist of Requirements Analysis, Functional Analysis and Allocation, and Design Synthesis. These are all balanced by techniques and tools collectively called System Analysis and Control. Systems engineering controls are used to track decisions and requirements, maintain technical baselines, manage interfaces, manage risks, track cost and schedule, track technical performance, verify requirements are met, and review/audit the progress.

5.1.2 Implementation

To ensure the simulator system design incorporated all requirements necessary to support the project experiments, a top-down approach using SEP was chosen. This allows for transparency in the design process and verification of the final design to the ICAO 9625 qualification standards. The SEP illustrated in Figure 5.2 can be applied as follows.

Process Inputs. Inputs in the context of this thesis consist primarily of the researchs needs, objectives (Section 1.3), requirements and project constraints (Section 3.7). They include the intended research missions (in this case the experimental trials briefed in Chapter 4), measures of effectiveness (Section 3.2.2), environments, available technology base, output requirements from prior application of the systems engineering process, academic department requirements, etc.

Requirements Analysis. By analyzing the process inputs, the requirements analysis is then used to develop functional and performance requirements. This will also highlight any design constraints that may pop-up. Section 5.2 documents this step.

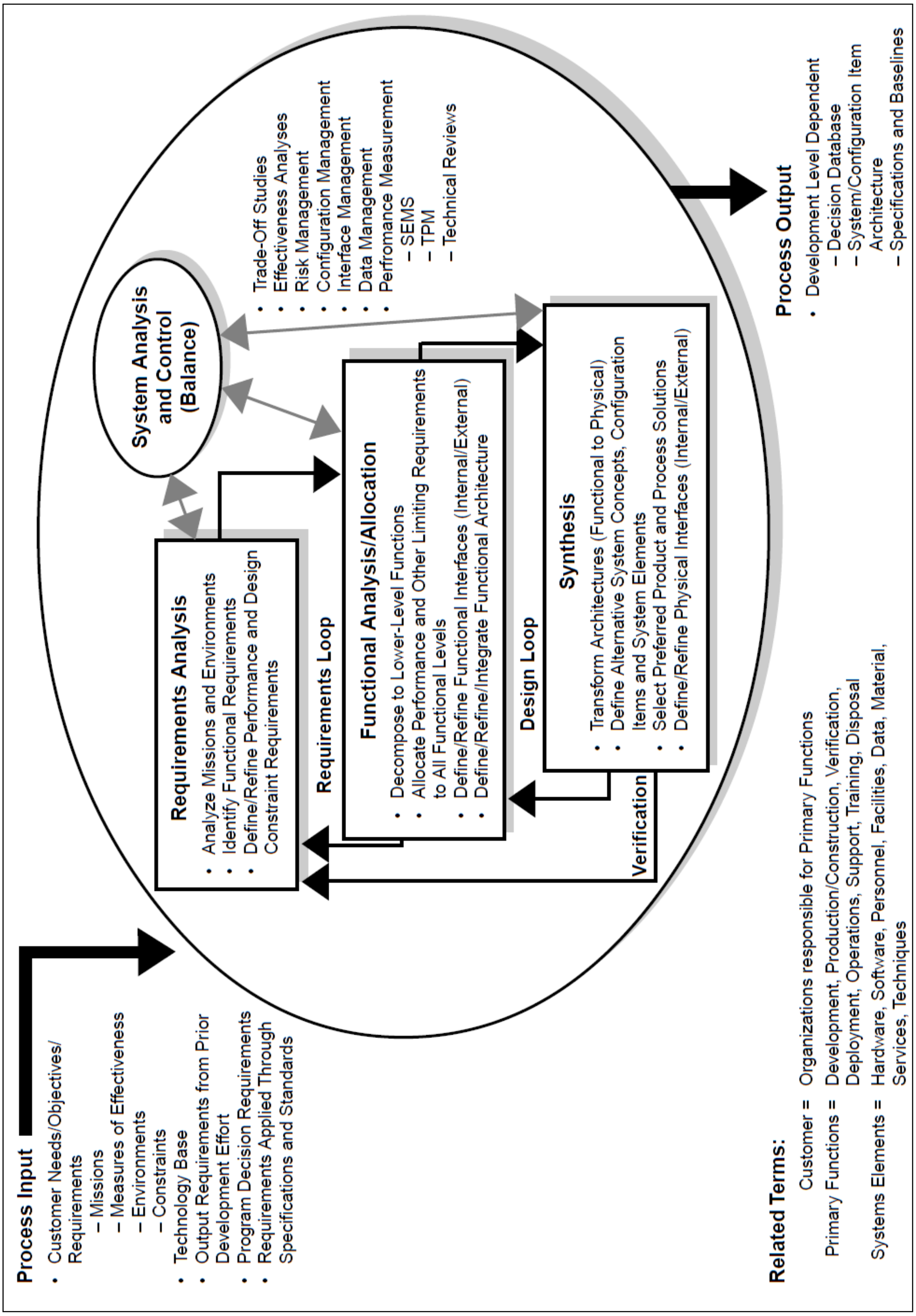


FIGURE 5.2: The Systems Engineering Process extracted from [160]

Functional Analysis/Allocation. Functions are analyzed by decomposing higher level functions identified through the requirements analysis into lower-level functions. The performance requirements associated with the higher level are allocated to lower functions. The result is a description of the research simulator in terms of what it does logically and what the performance is that is required. Two visualization results stemming form this step are the Functional Flow Diagram (FFD) and Functional Breakdown Structure (FBS) of the simulator in Section 5.3.

Requirements Loop. In this iterative step the functional analysis and requirements analysis are matched for traceability. Also new requirements are determined from low-level translated functions as this process loops in order to fully specify the simulator systems.

Design Synthesis. This is the process of defining the simulator system and subsystems in physical hardware and software components that together form the working solution. Also known as the physical architecture, the result is that each component meets at least one functional requirement, and any part of components may support many functions. The physical architecture is the basic structure for generating the specifications and baselines shown earlier in Section 3.2.2. The end result of the design iterations is presented in Section 5.5.

Design Loop. The creative aspect of the design synthesis in producing conceptual designs needs to be iterated in order to verify the physical design can satisfy the requirements and to yield an optimized design. It is also evident that some requirements may impact design features so only through iterations can the whole design be fully comprehended.

Verification. During each step of the systems engineering process, solutions are continuously checked against requirements. Due to the nature of this thesis and limited resources (uncertainties in evolving requirements and limited technical data of some system components) this also meant that formal testing and evaluation had to be done ad hoc and post hoc as best as possible within the time constraints. As such, extensive verification methods such as modelling and simulation, testing of the simulator itself was not always possible. Furthermore, being a research simulator, in the scope of this thesis the goal was to prove that the design was based on the latest task-driven simulator qualification process (Section 3.2.2) but not to go as far as the validation steps [36] since the simulator was never intended to be certified for flight training licence purposes.

Systems Analysis and Control. With limited resources, the technical tools used in specifying the simulator were documentation of conceptual designs, trade-off studies, risk management leading to the final design and construction. Post-construction activities did include acceptance testing prior to experimentation of critical systems.

5.2 Requirements Analysis

Requirements are the primary focus in SEP because the processs primary purpose is to transform the requirements into designs. As discussed in Section 5.1.2, the requirements analysis loop is iterated with functional analysis to converge to requirements for identified functions, and to verify that synthesized solutions can satisfy requirements.

To start with, there is a higher level drive behind the design and construction of the new research flight simulator facility at the AVRRC. Prior research simulators on-site were typically one-off devices custom-built by previous researchers to support their individual projects. To prevent such resource fragmentation and waste, a clean slate was envisioned to modernize the laboratory with a new simulation facility to not only support just this thesis but also two other concurrent aerospace projects listed in Appendix J as well as future projects. Therefore, a mission need statement for the new research simulator was formulated together with an accompanying project objective statement:

Mission Need Statement
Provide an academic research flight simulator to facilitate departmental research in manned and unmanned aviation.

Project Objective Statement
Design, build and operate a fixed-based, research flight simulator using low-cost, commercial off-the-shelf PC technology to support experiments by three PhD researchers in time for start of the academic year 2011-2012 with flexibility for future projects.

5.2.1 Operational View

The Operational View addresses how the system will serve its users. The main user is the PhD researcher, so from an individual project point of view (i.e. this thesis), there are basically four main stages (theoretical analysis, setup experiment, conduct experiment and post analysis) throughout the research process illustrated in Figure 5.3.

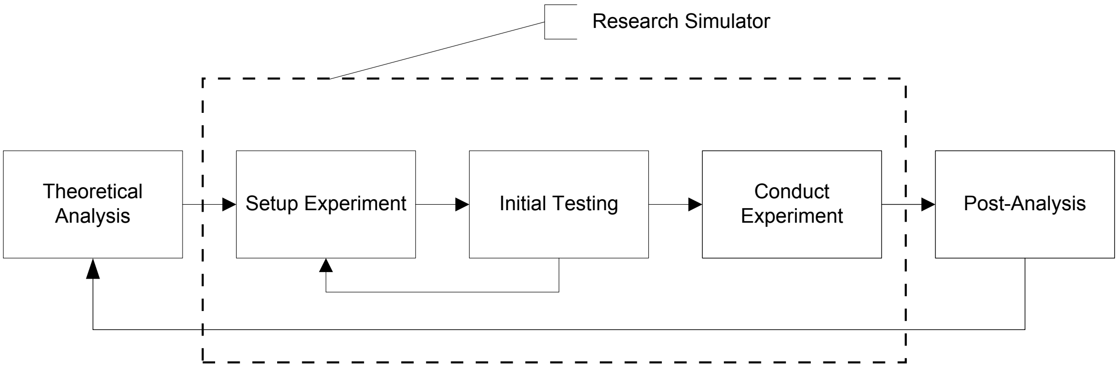


FIGURE 5.3: The role of the research simulator in this thesis

Theoretical Analysis. This phase of operation consists of each researcher developing and understanding their needs and requirements for their respective projects; typically includes gaining a fundamental understanding of the research topic, direction of research, theory, hypothesis and concepts of experimentation backed up with literary reviews and reports. In engineering projects yielding experimentation to verify theoretical concepts, the simulator is part of the final chain of simulation tools [87].

Setup Experiment. This is where the proposed experimental plan will be put into action; requires setting up the simulator to the required researcher's specifications and includes hardware/software testing to fine-tune system with respect to each experiment needs. Testing is carried out for each planned experiment to make sure that the experiment will produce usable data and results. Viability of the finalized experiment will also be fully understood during this testing phase. If initial testing does not meet the research needs then experimental setup will be changed or repeated until the desired outcome is reached.

Conduct Experiment. At this point the planned experiment is carried out on the simulator with real participants (if required), with the formalized settings and procedures determined. Data is logged for both post-analysis and real-time monitoring during the experiment.

Post-Analysis. The data collected during the Conduct Experiment phase is analyzed by using methods proposed during the experimental design (like in Section 4.4); these methods could include the use of a tool set available within the simulator or on an external machine. Network capabilities could help with the distribution of the relevant data to external points of access. The dry run before the actual experimentation serves as a verification step to ensure the collected data is sufficient to yield results, if not the experimental design in the theoretical analysis or a simulator setup design needs to be revisited.

Besides the direct individual operation of the simulator in support of research experiments, it is also fundamental to understand other stakeholders and organizations with an interest in the simulator. This marketing potential has a longer-term cascading effect with implications not only to the current research projects like this thesis for which the simulator is immediately intended but it can also to future research projects not yet in effect or even devised. It is beyond the scope of this thesis to fully cover this aspect, but from a requirements perspective a market analysis was performed covering the simulator associations with academia, industry and government. The market requirement analysis results are available in Appendix K.

5.2.2 Functional View

This subsection provides the outcome of the requirements loop explained in Section 5.1.2 which iterated both steps of requirements analysis (Section 5.2) and functional analysis (Section 5.3). The experimental design as presented in Chapter 4 was based on the primary flying task of performing a visual traffic circuit to land the aircraft as selected in Section 3.4.3. Breaking down this task into competency-level tasks meant that participants needed basic prerequisite flying skills as outlined by the checkride of Section 4.1.3 and Appendix H. The simulator fidelity specification was determined by going through the fidelity matching process of Figure 3.4 as discussed in the simulator qualifications section of Section 3.2.2.

5.2.3 Feature Requirements

Table 5.1 shows the results of deriving the final simulator feature fidelity levels from the master matrices stated in ICAO 9625 [36]. With the seven ICAO FSTD devices being the overriding master plan, Type I indicates the proposed simulator to be at the lowest fidelity level suitable for the training of private pilots. Details of each simulator device feature is available in Appendix B with the fidelity levels of: None (N), Generic (G), Representative (R) and Specific (S) previously explained in Table 3.1.

The required fidelity levels are very similar for these tasks matching the experimental design, with the Type I device and the landing task being the most demanding. The difference between the two lies in the airport and terrain environment feature, this needs to be representative or specific in case of the PPL trainer because when used for licence training the pilots need to be in the exact same simulated environment as they would be flying. With a research simulator and dependent on the flight simulation (software) suite, faithful modelling of this environment can be made possible but is not the minimum level required for the experiments

TABLE 5.1: Simulator fidelity master matrix derived from [36]

Licence	**ICAO FSTD**	**Cockpit Layout & Structure**	**Flight Model (Aero & engine)**	**Ground Handling**	**A/C Systems**	**Flight controls and forces**	**Sound Cue**	**Visual Cue**	**Motion Cue**	**Environment - ATC**	**Environment - Navigation**	**Environment - Weather**	**Environment - Airports & Terrain**
PPL	Type I	R	R	R	R	R	G	R	N	N	S	G	R(S)
Source	**Training Task**												
ICAO	8.1 Land the aircraft	R	R	R	R	R	G	R	N	N	S	G	G
FAA	12.4.3 Landing transition from a visual approach	R	R	N	R	R	G	G	N	N	N	G	G
MISC	Climbing and descending turns with $10° - 30°$ bank	R	R	N	R	R	G	G	N	N	N	G	G
AVRRC	Research simulator	R	R	N	R	R	G	G	N	N	N	G	G(R)

in this thesis. Hence, the landing task fidelity levels are used for this simulator design. Note that in particular the requirement for no motion cueing as stated in the research goal (Section 1.3.1) and project objective statement (Section 5.2) was determined by prior knowledge before the analysis but is also confirmed to be valid as such by completing this process. The navigation environment requirement was dropped due to the research not needing long distance flying, but this capability still exists in reserve.

5.2.4 Functional Requirements

Appendix L lists the compiled simulator feature fidelity requirements per technical category as well as the statement of compliance once the final design was completed. The distillation of these requirements together with physical and practical factors culminate in the Requirements Discovery Tree, illustrated in Figure 5.4.

Overall, the research simulator needs to provide two main functions: 1) provide a pilot station for the participant to perform the experiment and 2) provide a control station from where the experiment is controlled. Each research project with its

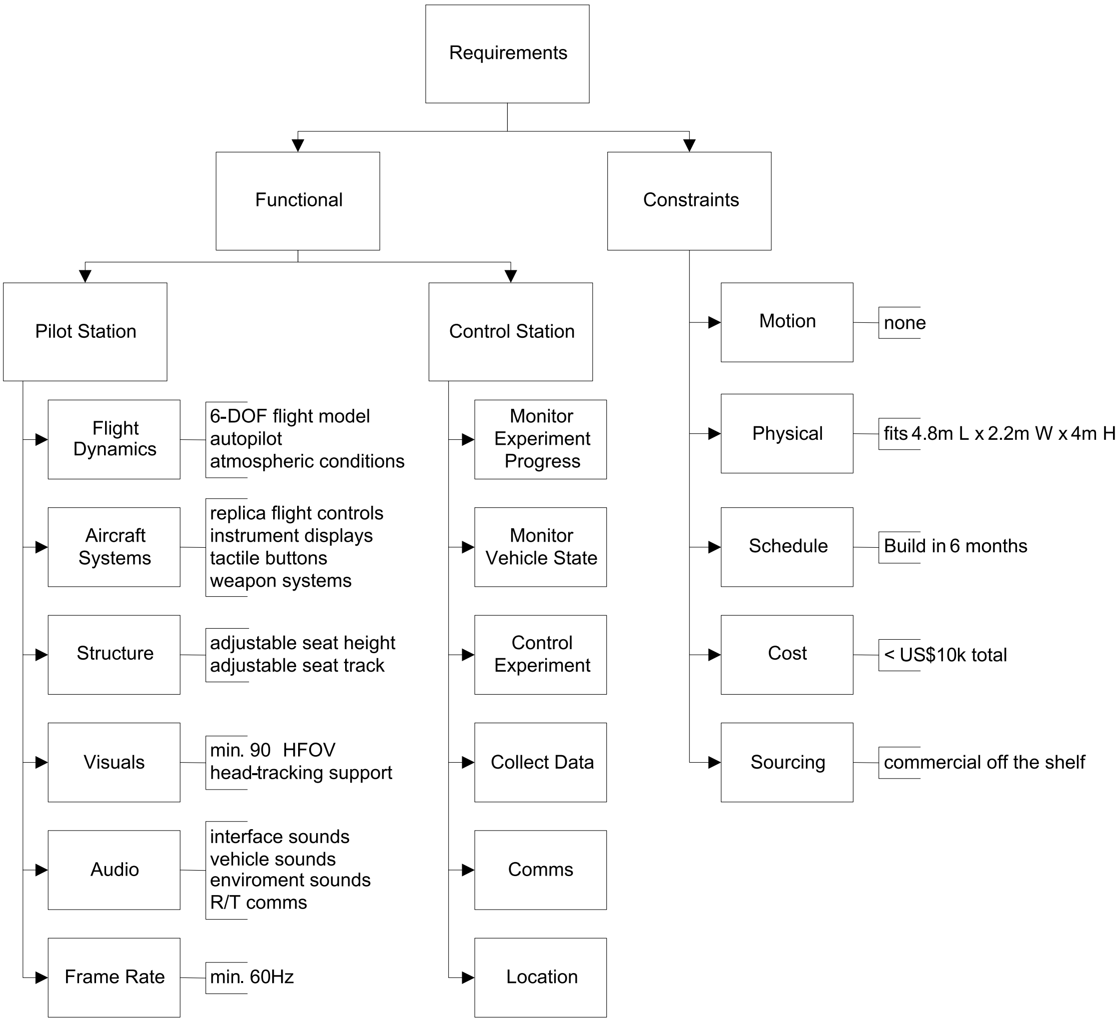


FIGURE 5.4: Requirements Discovery Tree

own experiments will require certain functionality out of the simulator but with a lot of common functionality and smart modular design, the simulator can be quickly reconfigured. The following summary recaps the main requirements.

A. Pilot Station

A.1 Flight Dynamics Based on 2.R and 11.G of Appendix L, an accurate 6-DOF flight model is needed in order to prove the validity for manned flight experiments. Although extremely detailed flight models obtained from flight test data might not be available, the aircraft flight model must ensure the pilot is able to perform normal control behavior. Atmospheric simulation of wind, gust and turbulence options to enable the researcher to conduct tests simulating weather conditions as well as particular flight manoeuvres i.e. cross-wind landings.

A.2 Aircraft Systems Flight controls must resemble those used in real life to provide a baseline reference and familiarity. Instruments displays must be reconfigurable to provide displays of legacy aircraft instruments and future displays, this complies with 4.R of Appendix L. For UAS purposes, secondary map display and video feeds are also required to replicate the analyst workstation.

A.4 Visual systems Provide forward horizontal field of view of at least 90 degrees with high-resolution, brightness, contrast, 7.R of Appendix L. Be able to integrate VR head/pose tracking technology into the simulator.

A.6 Frame rate The simulator needs to run in (quasi) real-time and at a minimum frequency of 60 Hz to ensure smooth synchronization with the visual display outputs at their native resolution, per 13.8.R Appendix L.

B. Control Station

B.1 Monitor experiment progress Means for the instructor to keep track of how the experiment run progresses as 13.2.R Appendix L. Provide overview of experiment phase and plan.

B.2 Monitor vehicle state Information in graphical or textual form presenting data and outside visuals for the instructor to assess the aircraft state, per 13.1.R Appendix L.

B.3 Control experiment Means to control the simulated environment, aircraft configuration and run scenarios.

B.4 Collect data Select, record telemetry for debrief and post analysis.

B.5 Communications Ability to communicate between instructor and pilot stations.

B.6 Location Isolated from the pilot station to minimize interference/distractions according to 1.2.2.R in Appendix L.

5.2.5 Constraint requirements

There were certain constraints posed on the simulator, with physical space limitations and budget the most demanding.

C.1 Motion The simulator must be fixed-based.

C.2 Physical The simulator must fit in the allocated laboratory space of 4.5m length by 4.5m width by 4m height.

C.3 Schedule The simulator shall be built by three PhD researchers within 6 months elapsed time.

C.4 Cost There shall be a total budget of USD 10k (currency selected for ease of comparison between systems worldwide) to procure the whole facility, including recycling of existing equipment.

C.5 Sourcing The simulator shall make use of commercial off-the-shelf equipment as much as possible.

5.2.6 Special Requirements

Based on these sets of requirements, *critical requirements* were identified to be motion and frame rate. Without motion, this reduces the complexity and costs substantially. Frame rate has implications for having sufficient computing power to deliver this performance. *Driving requirements* are representative flight controls, reconfigurable instrument displays, with schedule, cost and sourcing meaning the project uses off-the-shelf components and can be hand built without specialist equipment. The single *key requirement* is to be representative of the kind of simulators requiring a small physical footprint and mobility so that they can be quickly transported and deployed.

5.3 Functional Analysis

This top-down process describes the simulator functions in logical sequences after translating the requirements identified in Section 5.2. The Functional Flow Diagram in Figure M.1 clarifies the functional terms and the corresponding Functional Breakdown Structure of Figure M.2 are two visualization tools to aid traceability from requirements to solutions.

5.4 Risk Assessment

A risk assessment of the simulator design and construction process was performed in collaboration with Wright [161]. Potential risks were identified in Table 5.2 with descriptions of the issue, mitigation plans and potential solutions. Each issue is then weighted on a scale of 1-10 on the probability of it happening (10 being most likely). Each issue is also rated on its impact to the project on a scale of 1-10 (10 be highest impact). This was then mapped to a risk map as illustrated in Figure 5.5.

Table 5.2: Risk classification, mitigation and solutions

#	Cause	Mitigation	Solution
1	Data loss	Backup work to network storage	Increase work to recreate lost documents
2	Project requirement conflictions	Communication between project groups during requirements analysis	Compromise between the two or alternative interchangeable solutions
3	Unavailable parts	Source components from reliable sources	Reorder of alternative component
4	Delayed parts	Source parts from reliable suppliers	If overly long delay, cancel order and reorder with alternative company
5	Lack of space	Ensure that required space is available before hand	Negotiate alternative space or location
6	Over budget	Ensure that prices are sourced and all costs have been accounted for before any purchase is to be made	Cancelling orders and finding alternatives
7	Component failure	Get 3 year warranty on all parts and support from manufactures	Use warranties to replace parts and get support to have minimal down time
8	Loss of team members	Work to documented and plan to be laid out	Other members use documentation to take over role
9	Final solution fails to meet needs	Checks while working to ensure that what is needed is being sought	Make changes as needed, alter what needs to be changed
10	Failure of safety testing	Ensure that parts are sourced from quality suppliers and meet EU and UK standards	Return items and replace with safe items

5.5 Design Synthesis

This section provides a synoptic overview of the final results of the design loop process introduced in Section 5.1.2 leading to the eventual simulator build. Hence, the requirements and functional architecture have also been revisited many times

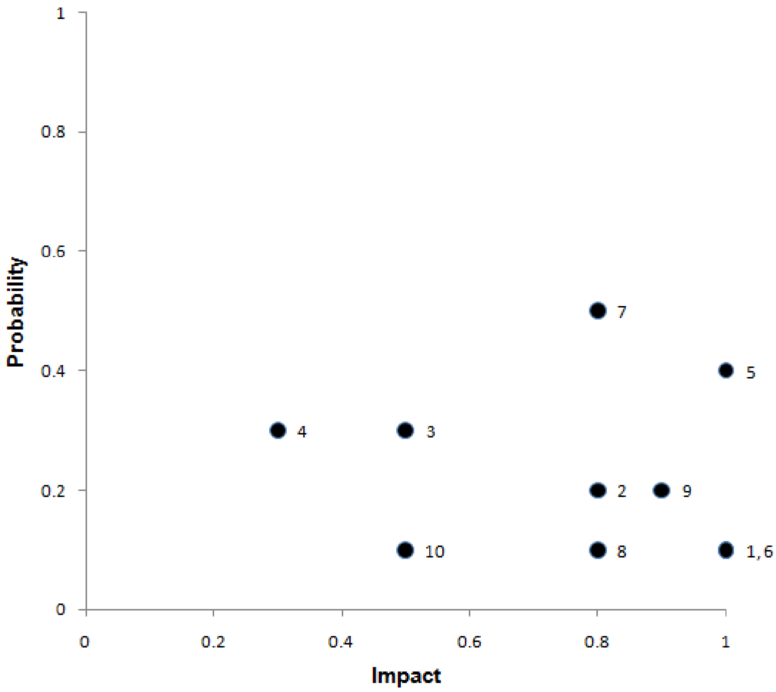


FIGURE 5.5: Simulator Project Risk Map [161]

before synthesizing the physical design. Also the verification process was continuous throughout the iterations to verify solutions were matching requirements and to capture the evolving requirements with time and project news updates.

5.5.1 Simulation Software

With the symbiosis between simulator hardware architecture and software options, a few iterations quickly showed that synthesizing the simulator based on the software choice first was the most logical approach considering experimentation requirements and cost. Cost constraints also meant that the hardware options were narrowed down to PC systems running Microsoft Windows 7 operating system as the university already had licenses and support for this. A survey of simulator software options boiled down to three categories: industry-grade, commercial/-consumer titles and the open source/homebrew custom programs, as visualized in Figure 5.6.

After a trade study was performed using fidelity, potential risk (knowledgebase, accessibility), interoperability (esp. with sponsors reviewed in Appendix K), adaptability, reliability and maintainability as criteria. The industry-grade simulations ([162, 163] were dropped due to expensive licensing, lack of access to source code/-modifications and knowledgebase. Open source[164]/homebrew was also not viable solution due immaturity, lack of extensive documentation and uncertain development lead times. Of the remaining three commercial candidates [133, 165, 166], X-Plane 9.70 was selected as the winner due to its maturity, advanced features in compliance with the requirements (Appendix L), ease of integration and established used for a wide range of aerospace research [167–170]. Last but not least, the researchers had the most experience with this software and were most confident in its ability to produce experiment scenarios for testing in the least amount of time.

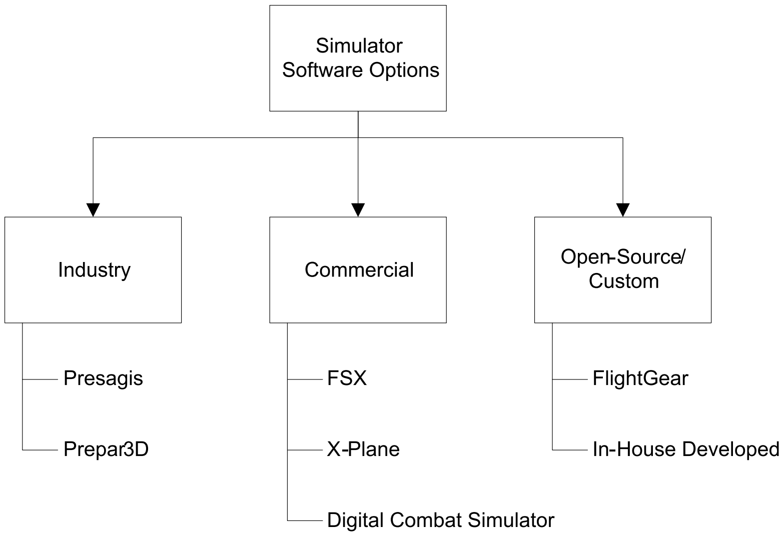


Figure 5.6: Software Options Tree

5.5.2 System Architecture

The end result of iterations using inputs from the requirements and functional analysis yielded the research simulator architecture shown in Figure 5.7. A division is made between the pilot/experiment station where participants do the experiment and the control station where the investigator monitors and controls the experiment. The experiment station is powered by the host PC, running the main flight simulation with visuals and flight controls, whereas a secondary slave PC and a laptop power the control station which job is to monitor and control the experiment. All computer systems are networked through a 1 Gbit local area network.

The system was initially built with two identical PC hardware configurations: one was mounted in a reused aircraft cockpit shell (for manned flying experiments), as shown in Figure 5.8. The other system was installed in a more generic simulator

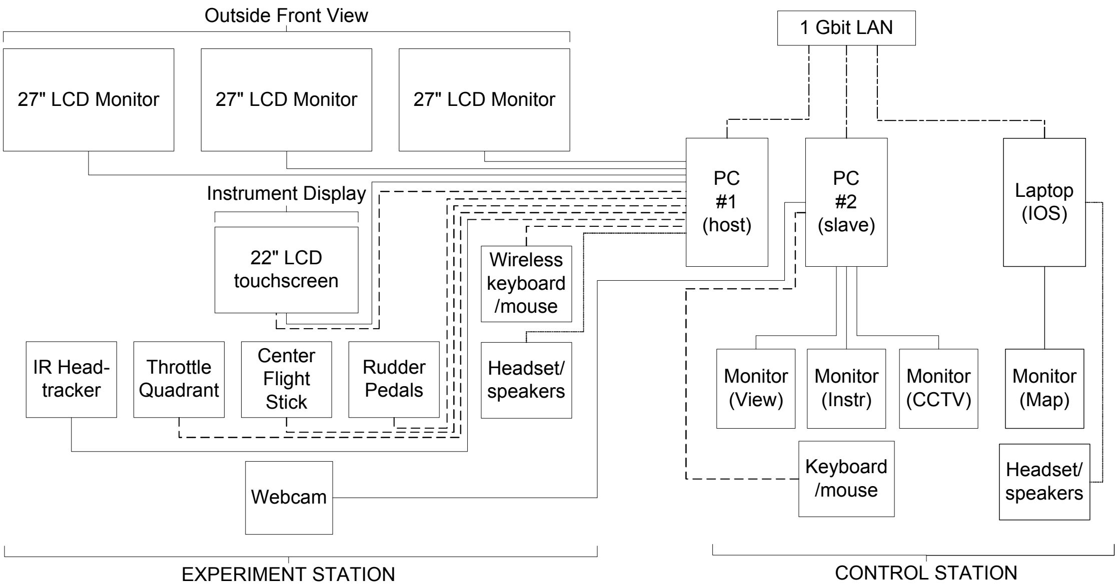


FIGURE 5.7: Research Flight Simulator configuration diagram

frame [171], which demonstrated system flexibility for quick reconfiguration into an unmanned aerial system (UAS) ground control station as shown in Figure 5.9. This way, if an experiment was in progress on either of the systems, the other would serve as the control station. Details of the PC system configurations per station with a cost breakdown is provided in Appendix N. With the LCD displays running at a native 60 Hz refresh rate, X-Plane is set to a vertical synchronization of 60 frames per second in order to match the displays. This was also well within the performance capability of the computer systems to avoid latency issues. The flight model was set update twice per frame, resulting in a 120 Hz rate. The external instrument displays are rendered by XHSI [132], a JAVA software package that reads flight parameters through X-Plane's data output. The experiment scenario and data logging was done by a custom built 'fat' plugin accessing X-Plane's data references to the simulator state variables in Appendix O.

After completion of both pilot stations, additional equipment as listed in Table N.3, was utilized to expand the research simulator facility by building a third station to serve as a dedicated control station shown in Figure 5.10. This enabled the simultaneous operation of the two existing pilot stations. With the simulation software suite chosen in Section 5.5.1, it was very easy and quick to add this extra functionality to maximize the newly available display real estate. As Figure 5.11 shows, the control station duplicates the visuals and the instrument displays. Via a flight-logging software [172] tool together with flight path visualization in Google Earth 7.1.1.1888, this provides all the required monitoring and control functionality. In the bottom right corner, the software for the webcam serves as a closed-circuit

FIGURE 5.8: Cockpit station with refitted displays and flight controls

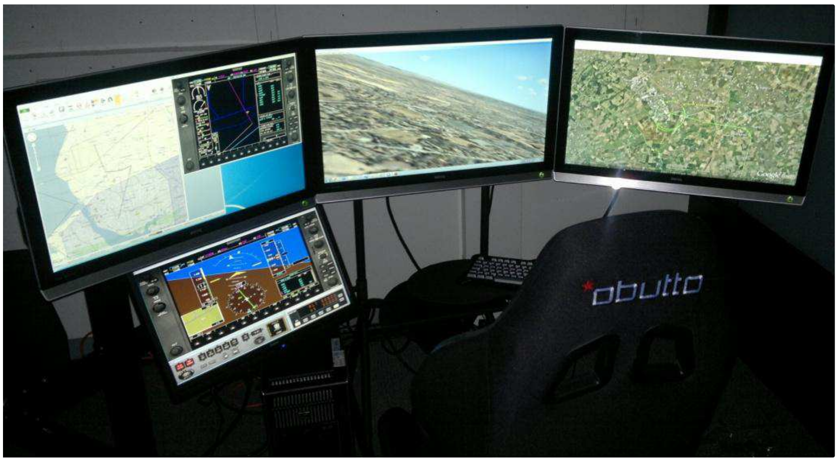


FIGURE 5.9: UAS ground control station simulator

television for the pilot station occupant's safety during experimentation. Also shown is the laptop which was used for development work and which can also be added into the network to serve as additional control station. The laptop then utilizes X-Plane's built-in Instructor Operator Station as a means to control the experiment station.

5.5.3 Projector Geometrical Correction

There were three existing LCD projectors (1400×1024 resolution) available in the laboratory but they were not configured to deliver a seamless outside visual image

Figure 5.10: Control station

Figure 5.11: Control station multi-display configuration

for flight simulation. To solve this problem, there were several consumer level software packages available [173–175] listed in Appendix N (Table N.4). After extensive testing, Immersive Display PRO [173] which despite being the cheapest solution was also evaluated to be the best. It provided all the required geometric correction features with the easiest, transparent calibration process and the additional bonus of displaying the native operating desktop system and automatic camera calibration capability.

The big issue found with the majority of these software solutions was the lack of clear documentation to aid the end-user to apply a successful blend/warp. Automatic camera calibrated solutions were also not able to perform as advertised in restricted physical spaces, either requiring an expensive, very high resolution camera or adequate field of view. Manual intervention defining the physical warp boundaries within the camera's viewpoint was also imprecise compared to manual calibrations. After extensive practice and experience, a full manual calibration can be completed under 10 minutes. The most time consuming task was setting the overlap blend.

Therefore, it is recommended to combine both methods by performing an initial manual calibration to correct geometry and then and then let a camera finish the process with an automatic blend of the overlap region. Another issue end-users need to be aware of is black level correction; unfortunately no consumer software package is now capable of delivering this performance without artifacts.

5.5.4 Head-tracking Integration

With the main simulator specifications finalized, the remaining task is to integrate head-tracking to facilitate the view control augmentation (Section 2.4.1) used in the experiments described in Chapter 4. Specifically, motion tracking systems most often derive pose estimates from electrical measurements of mechanical inertial, acoustic, magnetic, optical, and radio frequency sensors [176]. Because there is no universal single technology solution for all applications, each of their advantages and limitations need to be evaluated before adopting a particular solution [177].

For flight simulation, optical and inertial technology is most often used [178]. This is due to magnetic systems having too much latency and distortion from the metallic structures in the environment. Acoustic and radio trackers also require extensive calibration of the environment and are prone to external noise which is not practical for deployable solutions. With the low cost, commercial-off-the-shelf requirement in mind, an optical solution was selected and three candidate solutions found as shown in Figure 5.12.

The Kinect system [179] uses a combined 640×480 pixel camera system in the infrared and RGB color spectrum providing a 30 Hz update rate. Regular camera solutions like FaceTrackNoIR [180] utilizes face-tracking technology [181] to work with a wide variety of webcamera's built-in laptops to stand-alone high-definition external cameras. The drawback with webcamera's is that they require adequate

FIGURE 5.12: Optical tracking candidates (infrared, combination infrared/color camera and webcam based solutions)

lighting to function, which conflicts with the darkened environment requirement (1.R Appendix L) and during test struggled to maintain a steady 30 Hz update rate at the same resolution as the Kinect. The remaining and chosen solution was the TrackIR 5 [134] system, which delivered 120 Hz using passive infrared camera and position markers, shown in Figure 5.13. With a latency of just 9 ms, this was well within the 40 ms desired limits as found in Section 2.4.2. Furthermore, TrackIR had an established development track record with flight simulation software titles and provided the quickest integration route in this thesis. Prototype integration on actual military trainers as shown in Figure 5.14 was further proof of the maturity and acceptance of this solution for flight simulators.

FIGURE 5.13: TrackIR system with infrared emitter and ballcap marker [182]

Since the first TrackIR system retailed in 2001, it has been used by consumers and researchers on a wide variety of configurations. Nevertheless, there was no documentation regarding fine-tuning the amplification parameters. The manufacturer did supply default generic templates which formed the basis for in-house

FIGURE 5.14: Head-tracking demo on a part task trainer [183]

testing with the simulator. This required a learning curve and largely through trial-and-error feedback runs led to the used profiles as described in Section 4.2.3 and Section 4.3.3. Although the simulation software X-Plane, had built-in support for TrackIR, a custom view plugin [184] was used instead to replicate the view control while providing more access to the view control data.

5.6 Summary

In order to host the two-stage experiments designed to generate data for utility evaluation, a new research flight simulator had to be designed. The goal was to have a simulator system representative of its low-fidelity class so that experimental results could be validated against other systems qualified to a similar standard. To ensure a transparent and accountable system design, a top-down systems engineering process approach was used. This process meant that several iterations of requirements and functional analysis steps were completed to synthesize the final design.

The system requirements were also influenced by the fact that the simulator also had to support other projects. The inputs gathered to determine the total set of requirements were obtained from collaborating with AVRRC colleagues. This resulted in the identification of two critical requirements: fixed-based simulator and minimum frame rate of 60 Hz. Driving requirements were the use of commercial-off-the-shelf components to reduce cost and representative, reconfigurable flight controls and displays. Matching the low-fidelity class of simulators, the key requirement was a small physical footprint and mobility for quick transport.

The final design of the simulator system architecture is based on a twin-station configuration running X-Plane: an experiment station where the participant is actively controlling the aircraft and a control station where the observer supervises the experiment. The main hardware features are a fighter-type cockpit shell housing three computer displays for the experiment station powered by a single PC. The control station consisted of two PCs driving the replicate cockpit displays for monitoring the experiment and the instructor operator console for manipulating the experiment settings. The triple projectors uses manual, software calibration to warp and blend a single, large display. Head-tracking for amplified head rotations was done using a single, infrared TrackIR camera and markers.

Chapter 6

Results

As Figure 6.1 illustrates: after carrying out the experiments designed in Chapter 4 with the simulator system built in Chapter 5, this chapter presents the results of both experimental evaluations after performing the statistical analysis methods described in Section 4.4. The experiments and the obtained results provide insight into benefits obtained with augmentation compared to a standard setup and serves as a launch platform for further experiments with augmentation on larger displays. This forms the basis leading to a full discussion of the results and its implications in Chapter 7.

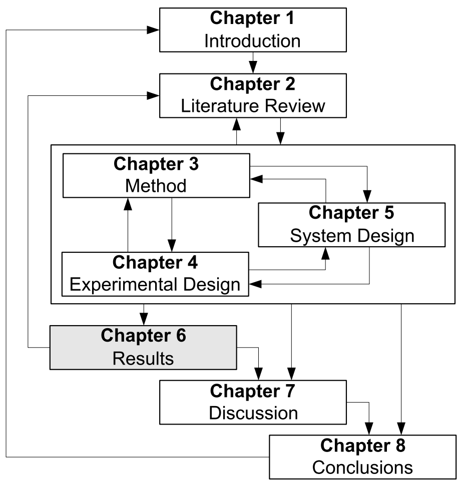


FIGURE 6.1: Chapter 6 in thesis roadmap

6.1 Experiment 1: Base-to-Final Circuit

Eleven male participants were recruited from the university for this experiment. All were current engineering students with a background or keen interest in aviation. The mean age was 28 years, with a standard deviation of seven years. Although some participants had logged flight time as a pilot, all were briefly assessed prior to the experiment on basic flying manoeuvres such as turns, descents and landings so that they could be considered equally capable of carrying out the experiment task. None of the participants had prior experience with amplified head rotation applications.

The participants generated a total of $(11 \times 12 =)132$ measurement runs. With a single group of participants and one independent variable (DISP), a one-way ANOVA was utilized for statistical analysis where the assumptions of normality and homogeneity of variance was met. Some dependent variables were not normally distributed and had heterogeneity of variance, of which skewness could not be corrected through transformation, therefore a non-parametric Kruskal-Wallis test was applied instead.

6.1.1 Flight Technical Performance Results

Base-to-final turn. The MANOVA test (Pillai's trace) reported a highly significant difference on base-to-final turn performance between DISP, $F(2, 129) = 21.938, p < 0.01$. Follow-up univariate ANOVAs, with Bonferroni correction accepting a reduced significant level of ($p = 0.025$), found that both the maximum bank angle ($F(1, 130) = 43.024, p < 0.01, r = 0.249$) and turning point distance from centerline ($F(1, 130) = 15.297, p < 0.01, r = 0.105$) were contributing measures to this difference as well as being both significantly different between DISP. Figure 6.2 shows that participants applied larger maximum bank angles during the turn on the single augmented display. Furthermore, Figure 6.3 reveals that the base-to-final turn was started later and closer to the extended runway centerline on the single augmented display. Comments from participants on this matter were that although a part of the runway was still visible on the edge of the triple monitor when approaching their desired turning point, they preferred to keep a larger part of the runway in sight at all times rather than flying a steeper turn with the risk of temporarily losing sight of the runway.

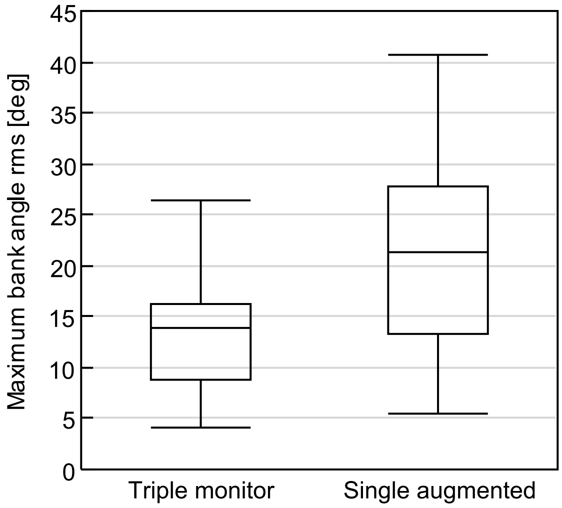


FIGURE 6.2: Boxplot of maximum bank angle during base-to-final turn

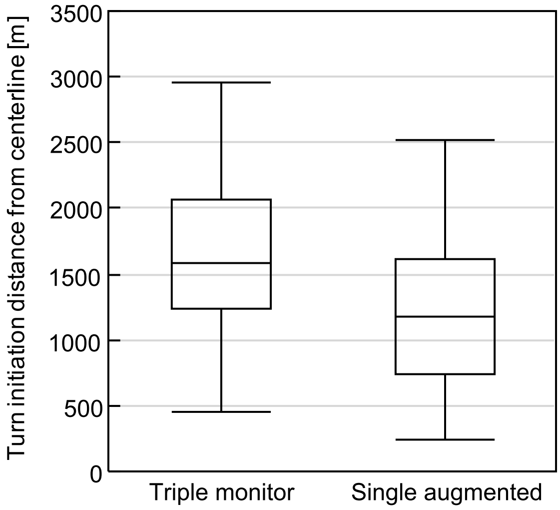


FIGURE 6.3: Boxplot of turn initiation distance to runway centerline

Final approach. MANOVA (Pillai's trace) results indicated significant difference between DISP during final approach ($F(2, 129) = 4.589, p = 0.012$). Following-up again with univariate ANOVAs yielded no significant difference for glideslope deviation ($F(1, 130) = 3.690, p = 0.057$). However, cross-track error was found to be significant ($F(1, 130) = 5.416, p = 0.021, r = 0.0419$) (with the corrected ($p = 0.025$) level of significance value). The single augmented display caused larger lateral deviations during lining up on the runway on final approach, shown in Figure 6.6. This can be explained by the smaller FOV and lack of peripheral view for better stabilizing the aircraft.

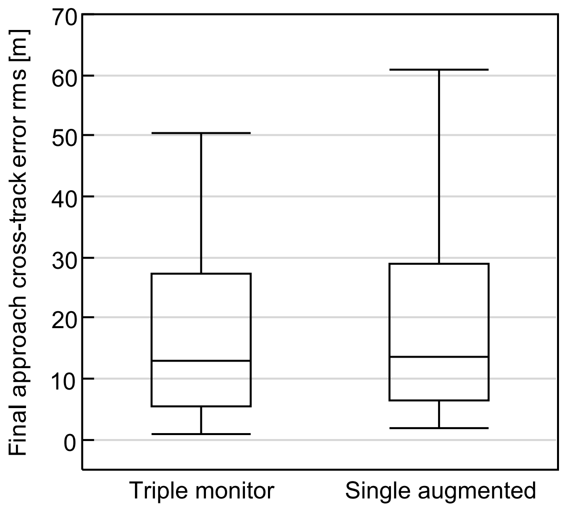


FIGURE 6.4: Boxplot of cross-track error during final approach

Touchdown. Wilks' lambda (MANOVA) results for touchdown location found no significant difference ($F(2, 129) = 1.313, p = 0.273$). All displays yielded similar results upon landing. Just the task of a straight-in approach and landing is in itself complex and challenging, with past studies favouring larger (horizontal) FOVs in particular for peripheral vision. Recent studies have shown that it is possible to achieve equivalent results across limited FOV displays for landing if the pilots are aware of the strategies available to them and use them accordingly [185].

6.1.2 Workload Results

NASA-TLX. One-way ANOVA of the NASA-TLX workload scores found a significant difference between DISP, $F(1, 20) = 4.851, p = 0.040, r = 0.195$. Figure 6.5 shows that the single augmented display had a higher workload than the triple monitor. Multiple ANOVAs for the individual workload scales then revealed (taking into account a Bonferroni correction of ($p = 0.0083$)) that mental demand ($F(1, 20) = 1.246, p = 0.278$), temporal demand ($F(1, 20) = 2.448, p = 0.133$), performance ($F(1, 20) = 2.902, p = 0.104$) and effort ($F(1, 20) = 2.530, p = 0.127$) were all not significantly different. Only physical demand ($F(1, 20) = 11.072, p = 0.003, r = 0.356$) was found to be significantly higher on the single augmented display. Frustration level was considered borderline significant ($F(1, 20) = 8.212, p = 0.010, r = 0.291$). This was not unexpected since participants required more active head movement in order to scan the whole visual world when confronted with a smaller FOV.

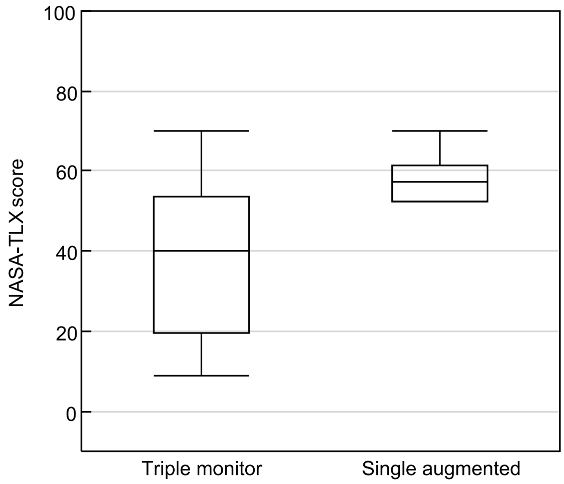


FIGURE 6.5: Boxplot of NASA-TLX workload rating

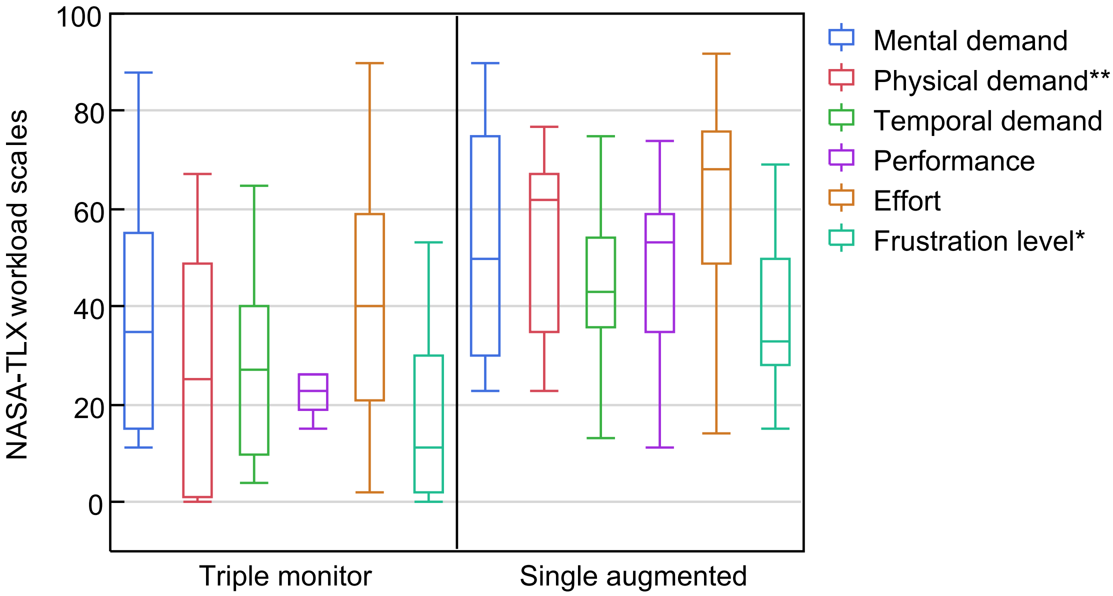


FIGURE 6.6: Boxplot of TLX workload scales (** for $p < 0.01$, ** for borderline)

Secondary task. Participants did occasionally fail to spot the single blimp during a measurement run. Table 6.4 shows the number of times the blimp was spotted or missed per display as well as the resulting detection rate. With negligible difference between the two displays, they can be considered equivalent in detection rate. The recorded detection times for when the blimp was spotted was also analyzed using ANOVA. The results indicated that detection time did not vary per display, $F(1, 126) = 1.292, p = 0.285$, which meant the single augmented display again provided equivalent performance in this regard to the triple monitor.

Table 6.1: Blimp detection statistics

DISP	Spotted	Missed	Detection Rate
Triple monitor	63	3	95%
Single augmented	64	2	97%

6.1.3 Simulator Sickness Results

The results of the administered SSQ showed that some participants reported simulator sickness symptoms during this experiment. Table 6.2 provides the number and corresponding percentage of participants out of the total of nine that got symptoms. The single augmented display had overwhelmingly more reports of nausea and oculomotor symptoms among participants. Disorientation was evenly matched across both displays.

Table 6.2: Number and percentage of participants reporting symptoms

DISP	Nausea	Oculomotor	Disorientation
Triple monitor	3 (14%)	2 (9%)	9 (41%)
Single augmented	19 (86%)	20 (91%)	13 (59%)

A non-parametric test (Kruskal-Wallis) of SSQ total score confirmed a significant difference between displays, $H(1) = 4.055, p = 0.044, r = -0.429$. Shown separately in Figure 6.7 for the SSQ total score and in Figure 6.8 for the individual factor scores, there was a greater reported simulator sickness on the single augmented display. This is backed up by the fact that all three SSQ factor scores are also pronouncedly higher for the single augmented display.

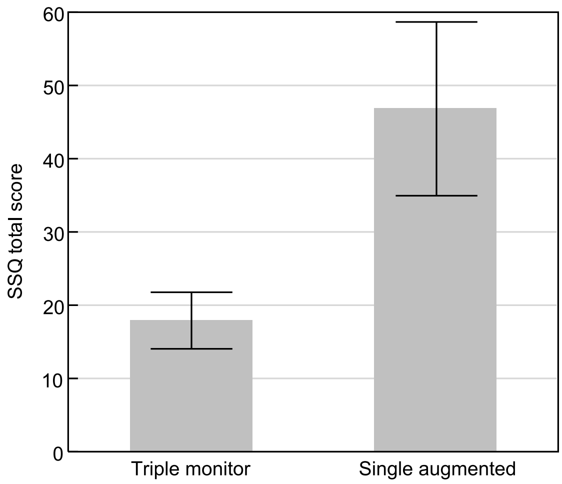


Figure 6.7: Mean SSQ total scores with standard error bars

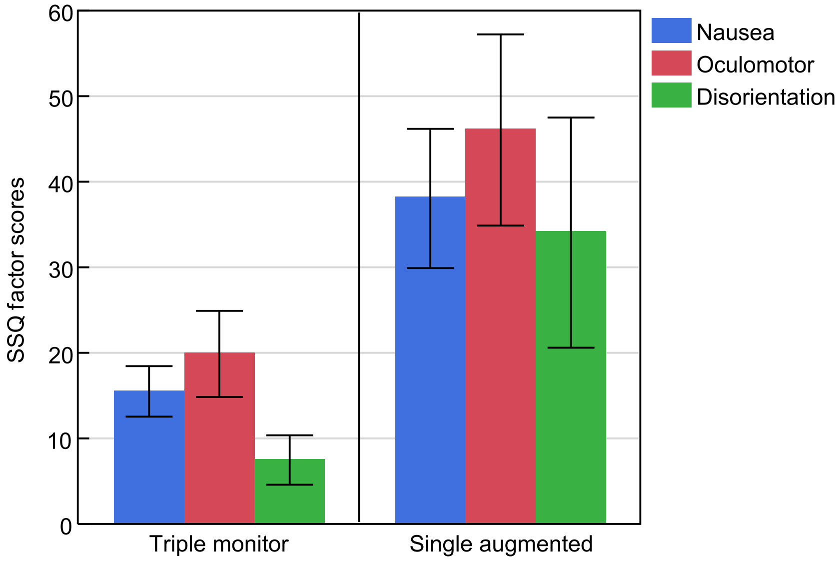


FIGURE 6.8: Mean SSQ factor scores with standard error bars

6.1.4 Correlation Analysis of Performance Measures

Because the dependent performance measures were not normally distributed, a non-parametric Spearman test was applied. The test showed that maximum bank angle was highly correlated to turn initiation point ($r = -0.559, p < 0.01$), this was found earlier in the flight performance measure plots as larger bank angles occurred when turns were performed closer to the runway centerline. Maximum bank was also further correlated to cross-track error ($r = -0.195, p < 0.025$) and longitudinal touchdown location ($r = 0.215, p < 0.013$): larger bank angles led to larger deviation during runway line-up and subsequent long landings. Turn initiation point was found to be highly correlated to cross-track error ($r = 0.244, p < 0.01$), meaning that whenever participants started their turns earlier, this led to a larger intercept angle with the runway centerline resulting in more deviations while trying to line-up with it. During approach, glide slope deviation had a major correlation to longitudinal touchdown point ($r = 0.810, p < 0.01$): larger deviations resulted in landings further down the runway, a logical consequence of non-stabilized approaches. Cross-track error in a similar regard was linked to both touchdown points: longitudinal ($r = -0.208, p < 0.016$) resulting in short landings and lateral ($r = 0.198, p < 0.023$) meant larger deviations during line-up resulted in off-centerline touchdowns. Finally, the number of blimps detected was found to be highly correlated to cross-track error ($r = -0.244, p < 0.01$): the more deviations from runway centerline during final approach meant participants were too busy with the primary flying task and would detect fewer blimps.

6.1.5 Final Questionnaire Results

To verify that the aircraft dynamics and simulator apparatus were not detrimental to the experimental task, participants scored the aircraft handling via the Cooper-Harper rating scale. This scale is a set of criteria used during flight test to evaluate the handling qualities of aircraft [186, 187]. The scale ranges from 1 to 10, with 1 indicating the best handling characteristics and 10 the worst. The criteria are evaluative and thus the scale is considered subjective. The pilot ratings of the aircraft dynamics and outside visuals are shown in frequency plots of Figure 6.15, ranging between 1-3 for all but one respondent. These scores fall under the Level 1 US Military Specifications for Handling Qualities, confirming that the flying qualities in this setup were adequate for the mission flight phase at hand [188]. Participants also reported positive feedback regarding the outside visuals, with the majority scoring it as realistic or very realistic. Finally, seven out of a total of eleven pilots (64%) preferred the single augmented display.

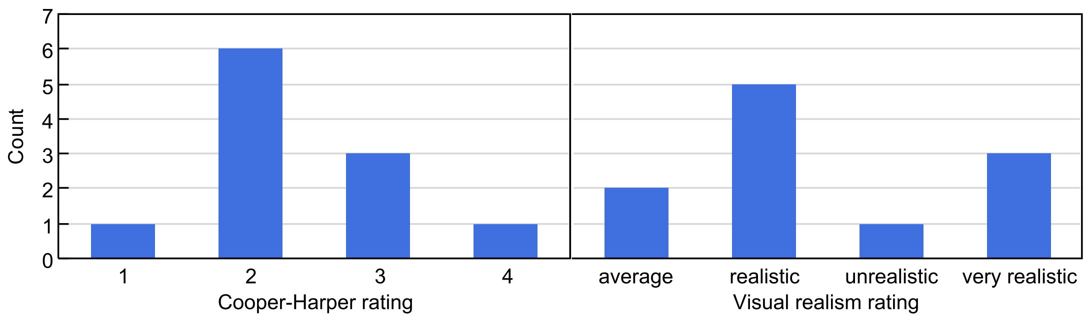


FIGURE 6.9: Aircraft dynamics and visual realism ratings

6.2 Experiment 2: Downwind-Base-Final Circuit

Nine male participants were recruited from the local flying club at East Midlands Airport (EGNX) for this experiment. All were licensed pilots or currently in flight training, with their individual age and flying experience listed in Table 6.3. None of the participants had any prior experience flying any of the experimental simulator configurations.

The participants generated a total of $(9 \times 18 =)162$ measurement runs. The performance dependent variables were analyzed using univariate Analysis of Variance

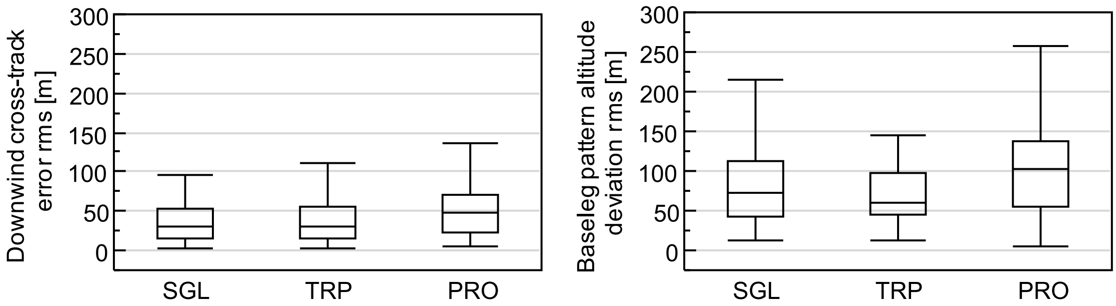
TABLE 6.3: Participant characteristics

Pilot	Age	Hours	Aircraft types
1	22	365	C152/172, PA28/38
2	25	36	G115
3	20	65	C172, AT01, SR20
4	19	20	G115
5	22	90	C152/172, PA28/38, T67
6	58	575	PA28
7	19	51	C150, PA28/38
8	26	60	C152, PA28
9	25	12	Glider, C152, G115

(ANOVA) per flight segment. With the expectancy that multiple dependent variables per segment would be correlated, a Multivariate ANOVA (MANOVA) was conducted to protect against the Type I error rate [140, 141]. The assumptions required for ANOVA/MANOVA tests regarding normally distributed data could not always be satisfied. However, with homogeneity of variance assured and equal sample sizes, ANOVA is still robust enough and has been reported/used as such in numerous studies [189]. Results using Wilks' lambda was taken in case both Box's M test for homogeneity of variance-covariance and Levene's test for homogeneity of error variances were satisfied, else Pillai's trace test was used instead [141].

6.2.1 Flight Technical Performance Results

Downwind leg. The MANOVA test (Pillai's trace) reported a significant difference on downwind performance between DISP, $F(6, 316) = 2.483, p = 0.023$. Follow-up univariate ANOVAs with Bonferroni correction accepting a reduced significant level ($p = 0.0167$) found that the downwind cross-track error was significant between DISP, $F(2, 159) = 4.297, p = 0.015, r = 0.0513$. Both pattern altitude, $F(2, 159) = 2.634, p = 0.023$, and reference airspeed deviations, $F(2, 159) = 0.840, p = 0.434$, were not significant. Figure 6.10(a) shows the boxplot for the downwind cross-track error across the three DISP levels, indicating the means and confidence level were higher for the triple projector compared to both monitor setups. Although the scenario is identical across all DISP with the aircraft trimmed for level flight, participants tended to drift to the right more on the triple projectors, they were not aware of this themselves even when they were 5° off the original heading. This can be explained due to the physical commonality of SGL and TRP where for the downwind segment, the middle screen sufficed to provide all the required visual cues. Flying on the triple projector would therefore be more different and the larger physical size providing stronger compelling visuals coupled with the head movement to cause this drift. However, a *post hoc* Tukey test (with alpha correction $p = 0.0056$) revealed that the only significant difference yet was between SGL and PRO ($p = 0.017$). SGL versus TRP ($p = 0.833$) and TRP versus PRO ($p = 0.074$) were insignificant.



(a) Downwind cross-track error plot

(b) Baseleg altitude deviation from 1500 feet pattern altitude

FIGURE 6.10: Boxplots of cross-track and altitude deviations for two significant flight phases

Turn 1 (downwind-to-base). Wilks' lambda reported no significant difference on Turn 1 between DISP, $F(6, 316) = 1.610, p = 0.144$. Although the ground tracks seemed to indicate differences in Turn 1, the scatter and variance was sufficient to cause statistical insignificant conclusions.

Base leg. Wilks' lambda found a significant difference between DISP, $F(6, 314) = 2.683, p = 0.015, r = 0.0539$. This was followed-up with ANOVAs (Bonferroni correction) at a reduced significant level ($p = 0.0167$). Baseleg heading, $F(2, 159) = 3.229, p = 0.042$, and Vref deviations, $F(2, 159) = 1.244, p = 0.291$, were not found to be significant. Base leg altitude deviation was found to be significant, $F(2, 159) = 4.528, p = 0.012$, shown in Figure 6.10(b). Despite this finding, *post hoc* Tukey test (alpha corrected $p = 0.0056$) did not detect significant pairwise differences: SGL-TRP ($p = 0.490$), SGL-PRO ($p = 0.160$) and TRP-PRO ($p = 0.009$), probably again due to limited sample size. This can also be explained by individual piloting styles, as some participants liked to lose more altitude during the base leg to avoid having excess potential energy left to have to bleed off during final approach.

Turn 2 (base-to-final). Wilks' lambda reported no significant difference on Turn 2 between DISP, $F(6, 314) = 0.799, p = 0.572$.

Initial line-up. One-way ANOVA showed a borderline significant difference for the runway alignment error, $F(2, 159) = 2.825, p = 0.062, r = 0.0343$. This is visualized in Figure 6.11, suggesting a trend of increasing positive angular errors from SGL to TRP to PRO. This meant that participants tended to overshoot the runway extended centerline, being most prevalent on PRO.

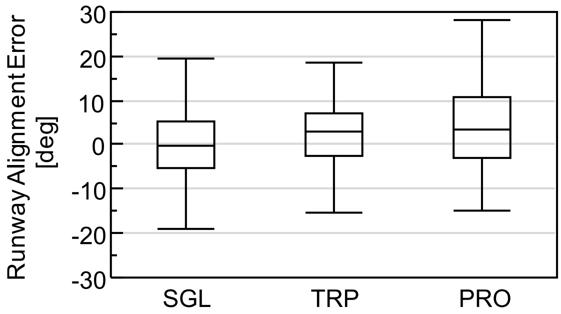


FIGURE 6.11: Boxplot of the runway alignment error during final approach

Final approach. MANOVA (Pillai) indicated a significant difference between DISP during final approach, $F(6, 316) = 2.820, p = 0.011$. Following-up with univariate ANOVAs (Bonferroni corrected) did not yield further explanation of this fact. The results showed no significant differences between DISP for Glideslope deviation ($F(2, 159) = 1.302, p = 0.275$), approach cross-track error ($F(2, 159) = 2.983, p = 0.053$) and Vref deviation ($F(2, 159) = 2.896, p = 0.058$).

Touchdown. MANOVA (Pillai) for touchdown location found no significant difference, $F(6, 318) = 1.829, p = 0.123$, all DISP yielded similar results upon landing. This matches the previous results found in the first experiment for this same flight segment.

6.2.2 Workload Results

NASA-TLX. To compensate for different subjective rating ranges per participant, all resulting TLX workloads were converted into z-scores per participant [88, 190] to allow proper comparisons between them. One-way ANOVA of the

TLX workload z-scores did not result in any significant differences between DISP, $F(2, 24) = 1.322, p = 0.285$. ANOVA for any of the TLX workload demand subscales did not find any significant differences either between DISP. Thus subjectively according to participants, head-tracking augmentation could be implemented across these three DISP levels with no impact on the TLX workload nor individual workload demands.

Secondary task. Participants did occasionally fail to spot blimp(s) during a measurement run. Table 6.4 shows the count of blimps that were missed per DISP for each of the three blimps that spawned. The percentages in brackets is the corresponding detection rate overall for each blimp per DISP. Pearson's chi-squared test was then applied to check if each of the blimp spot counts were significantly different due to DISP. This was not the case with Blimp 1 ($\chi^2 = 2.971, p = 0.226$) nor Blimp 2 ($\chi^2 = 3.419, p = 0.181$) but Blimp 3 ($\chi^2 = 18.459, p < 0.01, \phi = 0.338$) proved to be significant. The lowest detection rate occurred on SGL whereas PRO resulted in the highest. This could be expected beforehand since the triple monitor and projector offered increasing larger viewing areas to perform the visual scan hence an easier time to detect the blimp.

Table 6.4: Number of missed blimp spottings and corresponding detection rates in percentages

DISP	Blimp 1	Blimp 2	Blimp 3
Single monitor	7 (96%)	2 (99%)	26 (84%)
Triple monitor	5 (97%)	5 (97%)	14 (91%)
Triple projector	2 (99%)	1 (99%)	6 (96%)

Statistical tests were conducted separately on each blimp detection time to see if DISP influenced the time it took to detect the blimps. Figure 6.12 plots the detection times for each blimp per DISP. Overall mean times considered, Blimp 1 was the quickest to be spotted followed by Blimp 3. Blimp 2 proved to take the longest time for participants to detect.

Although attributed to spawning within the central viewing area during the low workload downwind leg (so it was relatively easy to spot across all DISP), ANOVA showed no significant difference in detection times for Blimp 1, $F(1, 126) = 1.292, p = 0.258$.

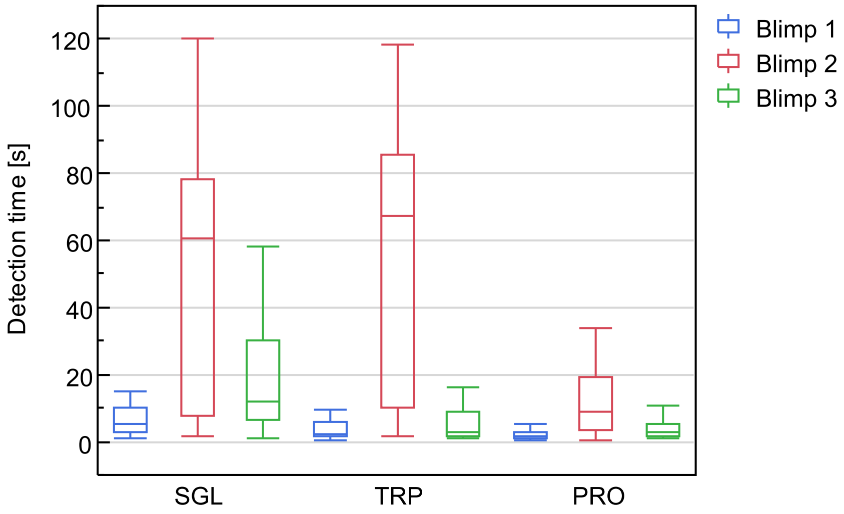


FIGURE 6.12: Blimp detection time

Since the data for Blimp 2 times was unsuitable for ANOVA and could not be transformed, a non-parametric Kruskal-Wallis was used instead. This gave a significant result for Blimp 2, $H(2) = 20.053, p < 0.01$. Due to unequal groups in sample size, further *post hoc* investigation was done via Tamhane's test (with significance levels again corrected per Bonferroni to $p = 0.0167$). This showed SGL and TRP to be inconclusive ($p = 0.553$), whereas SGL-PRO and TRP-PRO were both significantly different (both $p < 0.01$). This was confirmed by a *post hoc* Mann-Whitney test revealing large effect sizes for SGL-PRO at $r = -0.464$ and TRP-PRO at $r = -0.618$. Participants detected Blimp 2 the quickest on the triple projector DISP.

6.2.3 Simulator Sickness Results

The results of the administered SSQ showed that some participants reported simulator sickness symptoms during this experiment. Table 6.5 provides the number and corresponding percentage of participants ($n = 9$) who were symptomatic for simulator sickness. Oculomotor was the least reported symptom out of the three categories. Pearson's chi-squared test was then applied to check if each of the individual symptom categories were significantly different due to DISP. This was not the case with nausea $\chi^2 = 2.077, p = 0.354$, oculomotor $\chi^2 = 0.355, p = 0.837$, and disorientation $\chi^2 = 0.297, p = 0.862$.

TABLE 6.5: Number and percentage of participants getting symptoms

DISP	Nausea	Oculomotor	Disorientation
Single monitor	3 (33%)	4 (44%)	4 (44%)
Triple monitor	5 (56%)	3 (33%)	5 (56%)
Triple projector	6 (67%)	3 (33%)	5 (56%)

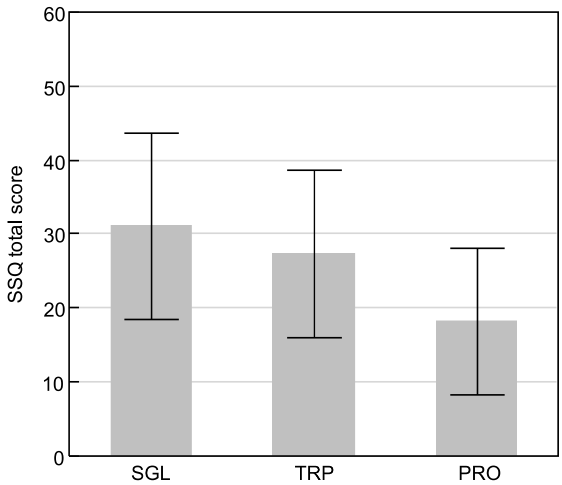


FIGURE 6.13: Mean SSQ total scores with standard error bars

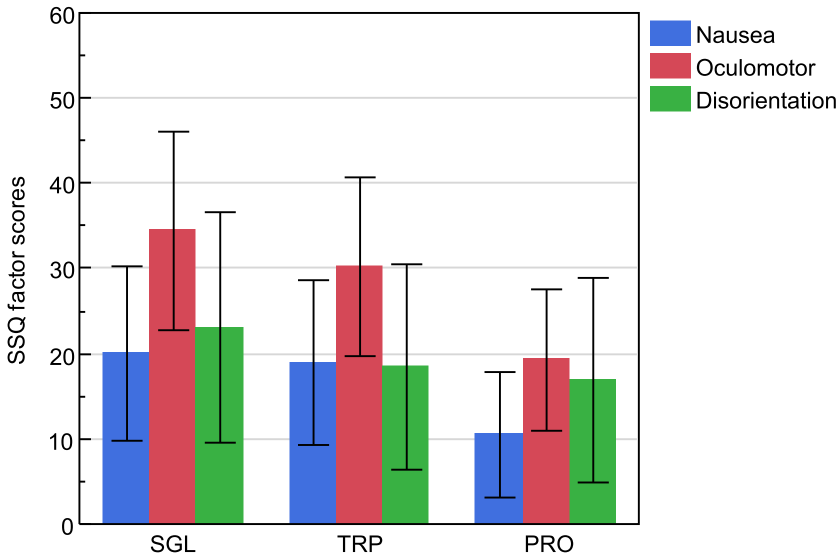


FIGURE 6.14: Mean SSQ factor scores with standard error bars

The SSQ total score as well as the factors score are both shown in Figure 6.13 and Figure 6.14 respectively. Statistical analysis (ANOVA) indicated no significant differences for the total SSQ score between DISP, $F(2, 24) = 0.342, p = 0.713$. However, the bar charts for the mean values suggested there was a trend of reduced simulator sickness going from single monitor on one end to triple projector on the other end. ANOVA of the SSQ factor scores did not reveal any significant differences between DISP. Again, the bar charts suggest a reduction in oculomotor factor from single monitor to triple projector. Furthermore, the indication of higher oculomotor factor means compared to nausea and disorientation is not surprising since this experiment focuses on the visual display system with head-tracking [120].

6.2.4 Correlation Analysis of Performance Measures

Since the dependent performance measures were not normally distributed, a non-parametric Spearman test was used for correlations, again for flight segments. However during applicable flight segments, the blimp detection times from the objective workload measures data were combined to cross-check workload indications.

The test showed that there was no correlation during the downwind leg between the performance dependent measures themselves and Blimp 1 or Blimp 2 detection time. For Turn 2, the turn distance from runway threshold had a significant positive correlation to the maximum bank angle ($r = 0.479, p < 0.01$) and reference airspeed deviation ($r = 0.173, p < 0.05$). This implied that if the downwind turn to base was started further away from runway threshold, pilots would use larger bank angles as well deviate more from the reference airspeed. There was no correlation between Turn 2 manoeuvring and detection time of Blimp 2.

On base leg, Blimp 2 and Blimp 3 detection times were added to the existing three performance measures. The test showed that base leg altitude deviation had a significant positive correlation to reference airspeed deviation ($r = 0.223, p < 0.01$) and Blimp 3 detection ($r = 0.383, p < 0.01$). This indicates that larger altitude deviations were coupled with larger airspeed deviations, which is fundamental in flight dynamics. Furthermore, the larger altitude deviations also resulted in a larger workload to correct for hence the longer it took to detect Blimp 3.

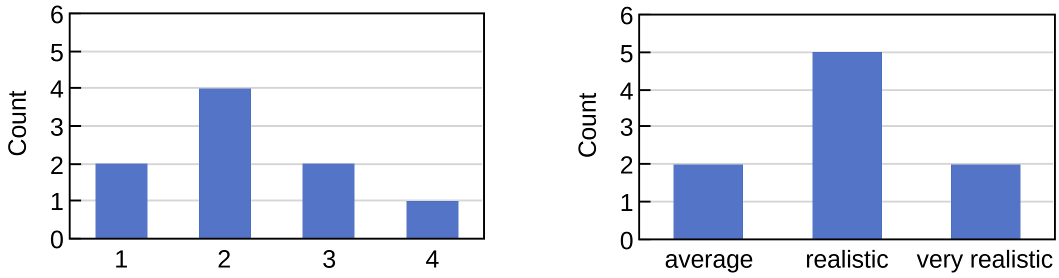
Turn 2, from base-to-final with Blimp 3 added, the test showed that maximum bank angle had a significant negative correlation to turn start distance from centerline ($r = -0.354, p < 0.01$), positive correlation to airspeed deviation ($r = 0.165, p < 0.01$) and Blimp 3 detection time ($r = 0.191, p < 0.01$). This is understandable as pilots who started their final line-up turn closer to the runway (extended) centre line had less turning room to do so, hence leading to larger bank angles used, meaning larger airspeed deviations had to be corrected, in short a higher amount of workload leading to a longer time before spotting Blimp 3.

For final approach and touchdown, the three approach variables were combined with initial runway alignment error, touchdown location and Blimp 3 detection time. The test found a significant positive correlation between glide slope deviation and cross-track error ($r = 0.309, p < 0.01$). This means that pilots who were struggling with maintaining the glide slope tended to also have more issues with lining up with the runway. Furthermore, glide slope deviation had a significant correlation to both touchdown locations, for longitude ($r = 0.732, p < 0.01$) and latitude ($r = 0.370, p < 0.01$). This meant pilots who strayed too far from the intended glide slope would land further down the runway with a larger chance of being off-centre. Further lateral touchdown location correlation was for cross-track error ($r = 0.562, p < 0.01$) and longitudinal touchdown location ($r = 0.155, p < 0.01$), meaning that the more misaligned the pilot was on the runway, the more off-centre the resulting touchdown would be as well as tending to be further down the runway.

Finally, the test between the TLX subjective workload and SSQ scores did not reveal any correlations. So any perceived workload was not linked to any experienced simulator sickness symptom(s).

6.2.5 Final Questionnaire

A Friedman's ANOVA test was conducted to determine whether participants had a differential rank ordered preference for the three levels of DISP. Results indicated that there was a significant rank preference, $\chi^2 = 9.250, p = 0.010$. The median for triple projector was being first choice, triple monitor as second choice and single monitor the least preferred. *Post hoc* comparison of the preference rankings was conducted using a Wilcoxon signed-rank test to determine if these medians



(a) Cooper-Harper Rating

(b) Visual realism rating

FIGURE 6.15: Pilot ratings of aircraft dynamics and outside visuals

could be substantiated by significant pair-wise differences. A Bonferroni correction was applied to account for three tests [141], thereby reducing the statistical significance level of acceptance to $p < 0.0167$. Results of this indicated that there were significantly more favourable rankings of triple monitor over single monitor ($Z = -2.640, p = 0.008$), borderline favouring of triple projector over single monitor ($Z = -2.309, p = 0.021$) but no significant preferences between triple monitor and triple projector ($Z = -302, p = 0.763$).

The pilot ratings of the aircraft dynamics and outside visuals were again sampled using the Cooper-Harper rating scale, with the results shown in Figure 6.15. The score results were between 1-3 for all but one respondent. This met the Level 1 US Military Specifications for Handling Qualities, verifying that the flying qualities were sufficient for this experiment [188].

Participants gave positive feedback regarding the outside visuals, with the majority scoring it as realistic or very realistic. Positive comments in particular were about the high resolution and detailed virtual airport environment such as individual modelling of runway lights. Half of the participants were initially skeptical regarding the usability of head-tracking augmentation but were pleasantly surprised at its utility and short learning time to adapt.

Chapter 7

Discussion

Following the thesis structure guide in Figure 7.1, this chapter discusses in detail the obtained results from the experiments done in Chapter 6. This interpretation is presented within the context of answering the research objectives and the link to the existing body of research. In addition to experimental findings, a critical evaluation of the simulator design process of Chapter 5 is also delivered.

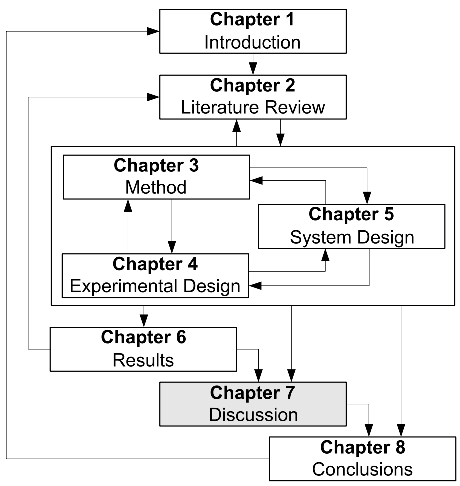


FIGURE 7.1: Chapter 7 in thesis roadmap

7.1 Retrospective Summary

THE objectives of this thesis were to research the effects of enhancing a low-fidelity, fixed-based flight simulator with amplified head rotations in various display configurations on a basic traffic pattern flying task. This was made possible by performing two human-in-the-loop simulator trials with the augmentation technique integrated into the simulator. The experiments were conducted using a within-subject study design since the independent variable was solely the display configuration.

In the first experiment, the focus was on establishing a baseline comparison study by evaluating the most limited implementation of the amplified head rotations on a single outside visual display against the common, non-augmented, triple-display configuration. The task was to complete a visual traffic pattern from the base leg to a full stop landing while keeping a visual lookout for a single blimp simulating airborne traffic. Quantification was done by selecting suitable performance metrics to capture a wide spectrum of human factors. Objective performance measures recorded were flight technical performance data (the scenario was split into three phases: base-to-final, approach and touchdown). Workload was assessed with the secondary task of blimp detection as well as subjective workload ratings for cross-validation. Since the experiment involved a fixed-based simulator, simulator sickness issues were also assessed by self-reporting of symptoms.

With the first experiment results providing a baseline/control reference to validate the novelty and gain more system development maturity with the augmentation, the second experiment was to extrapolate the technique to higher visual display fidelity levels (physical size/FOVs). The goal was to determine how scaling impacts upon the usefulness of augmentation in order to justify potential retrofitting of this technique to existing higher fidelity simulators. Comparing within augmented systems enabled a more complex flying task to fully demonstrate the expanded training task potential. The visual traffic circuit was therefore extended to encompass the downwind leg and the number of spawning blimps increased to three. The performance measures were similar to the first experiment with the exception of additional flight segments.

Although not explicitly stated as a research objective; a new research flight simulator had also to be designed and built to facilitate both experiments. By complying with the latest simulator fidelity level qualification philosophy [36], the constructed simulator can be verified to be a representative low-fidelity class device to validate the research. The selection of certain solutions and components to meet the requirements to produce the final simulator specification meant that these also had implications on the experiment and are discussed where relevant.

7.2 Research Objectives/Questions Findings

Section 1.3.2 defined the two research objectives and seven corresponding research questions, and each of the two experiments matched a research objective. This section addresses these in numerical order with explanations and cross-references to highlight the holistic nature of this research study. (N.B. where the term significant is used, this refers to a statistical significance at a p-value $p < 0.05$ as determined in the results unless a Bonferonni correction was applied).

7.2.1 Objective One

Investigate the effects of amplified head rotations on fixed-based flight simulators in a basic flying task.

This research objective was completed by carrying out the experimental plan in Section 4.2 with the results obtained in Section 6.1.

Research Question 1
How does amplified head rotations affect visual flying technical performance compared to a legacy, non-augmented flight simulator?

The results showed that for the base-to-leg turn segment, the amplified head rotations on a single display significantly generated larger maximum bank angles and turns were initiated closer to the runway extended centerline compared to the legacy, non-augmented triple display configuration. Considering a limited sample size, the size of the effect was small ($r = 0.25$) for maximum bank angle, with a 59.0% increase in mean values (Figure 6.2). The turn initiation distance also had a small effect size ($r = 0.105$), with a 26.1% decrease in mean distance from the runway centerline (Figure 6.3). This more desirable rectangular flight path result contributes to the compliance with the basic rectangular traffic pattern aim in reducing the possibility of conflicts at airports without an operating control tower [112].

From a human visual perspective as outlined in Section 2.1.1, the single desktop display primarily uses human central vision, whereas the wider FOV on the triple display also provided more peripheral vision. A further limitation due to the low-cost drive behind the simulator architecture presented in Section 5.5.2 is that the triple display was effectively outputting a single, large virtual display. This meant that only a single virtual camera projected this software FOV using a standard perspective projection, which was on the limits of what typical 3D graphical engines could output. At this wide limit, there is considerable distortion and pixel inefficiency i.e. fish-eye effect. One solution then is to render multiple perspective views that cover a superset of the required FOV and combine these together to form the required FOV, noting that some additional image warping may be required to deal with the geometry or other details of the projection geometry [191] but this would have required multiple image generators and more computer systems which conflicted with the low-cost driving requirement.

Stabilized approaches are key to good landings [192], as such a small effect size ($r = 0.04$) was found that showed the single, augmented display generated an 110.33% increase (significant) in mean cross-track error (Figure 6.6) while tracking the centerline. This was attributed to the reduced FOV and lack of peripheral cues offered by the single display [67]. It has been demonstrated that piloted approaches and landings can be successfully achieved with a reduced FOV in remote television controlled landings [193] or on complete synthetic vision systems [194, 195]. However, without an experienced pilot or additional symbology [185] to aid this approach task, pilots have to rely on environmental cues as provided by the real world [73, 196–199] and hence the reduction in peripheral vision caused detrimental effects as the results have shown. The non-significant results on glide slope was explained by the sufficient availability of visual cues such as runway size and geometry in the focal vision to facilitate vertical path control [200]. Overall, these results contribute to further the qualification process in flight simulator training devices [36] to validate the suitability of novel technological solutions to the desired training task and the further understanding of FOV effects on manoeuvring performance [30].

Research Question 2

How does head augmentation impact workload in a piloting task?

In terms of subjective workload, the NASA-TLX results showed that for a small effect size ($r = 0.2$) the single, augmented display was significantly more demanding with a 43.2% increase in mean workload. As the box plots in Figure 6.5 have shown, the mean and 95% confidence levels for the single, augmented display showed that respondents were in high agreement with this rating. A closer inspection into the workload scales revealed that physical demand was the significant contributing factor with a medium effect size ($r = 0.4$). This was not surprising since participants were adapting to a new virtual interaction control method they had no prior experience with (despite the learning phase put in place, introduced in Section 4.1.2). As this was the first simulator trial with the amplification profiles, this was not optimal with post-experiment feedback recommending more personalized profiles. The borderline significance for frustration level confirmed this aspect.

There was no significant difference found between the two displays for blimp detection with the detection rate for both displays 95% for the triple monitor and 97% for the single augmented respectively. This was surprising considering the visual scanning behaviour [201, 202] required to fly the aircraft and maintain a visual lookout despite having a smaller FOV on with the augmentation.

Research Question 3

What simulator sickness side effects occur when amplified head rotations are introduced?

As Figure 6.7 and Table 6.2 have shown, a significant difference was indeed found between the two displays with a medium size effect ($r = -0.4$). The single augmented display had almost double mean total score of the triple display. This was further supported by the high number of reported symptoms of nausea (86%) and oculomotor (91%) categories for the single augmented compared to the low reporting of 14% and 9% respectively in case of the triple display. This high occurrence of nausea and oculomotor symptoms were typical of prior studies with head-tracking backed by the cue conflict theory discussed in Section 3.6.2, though it must be kept in mind that participants were still getting used to the interaction technique. The fact that disorientation was not a factor between the two displays was explained by the relatively restrained anticipated self-motion (i.e. sitting in an aircraft, one can expect the travel motion the aircraft makes).

It is notable that despite these formally obtained results, the simulator sickness questionnaire (SSQ) is still a subjective rating method where participants in a laboratory testing environment are still more likely to be critical of their own responses (i.e. when in doubt they would report a symptom) [203]. Despite this, the sickness symptoms were of very short duration matching existing research [129] as the short breaks designed into the experiment procedure in Table 4.1 were effective. Participants did not experience any sickness symptoms prior to starting an experiment run with a new display configuration.

7.2.2 Objective Two

Investigate the scalability of amplified head rotations to higher levels of visual display fidelity.

This research objective was fulfilled by performing the simulator trials planned in Section 4.3 with the results obtained in Section 6.2.

Research Question 4
How are head mapping profiles tuned for larger displays with larger FOVs?

As Section 4.3.3 described, an empirical method had to be used due to the lack of literature or any available guidelines to tune the amplification profiles. Lessons had to be learned and user feedback after the first experiment was therefore used as guidance. As the experiments only used two degrees of freedom in pitch and yaw: the yaw profile was most heavily influenced with display configuration as the horizontal display area was extended in the experiments. As Figure 4.18 already illustrated, the yaw profile for the single display has a much steeper initial gradient than the triple displays/projectors. A further zoom in near the origin are as shown in Figure 7.2 shows that the larger (wider) displays offer a larger deadzone before the amplification takes effect, this offered a larger stabilized zone for looking straight ahead which reduced jittering and improved user comfort. The single display obviously had to trade-off between a larger deadzone which meant there was less room left to map the required virtual camera angles or vice versa which would decrease view stability near the centre view.

For the pitch head angle, the single and triple display used the same profile as they had identical vertical FOV. Because the triple projector screen was also very close the computer monitors in vertical FOV as listed in Table 4.3, it was not surprising that the original profile used for the single display could be simply extrapolated with only minor modifications for the initial angles. As Figure 7.2 illustrates, the triple projector offered more deadzone area and a reduced initial amplification gradient due to the larger physical display size. To sum up in general, larger displays require less steep initial gradients and offer a larger deadzone near the centre view for increased head stability and user comfort.

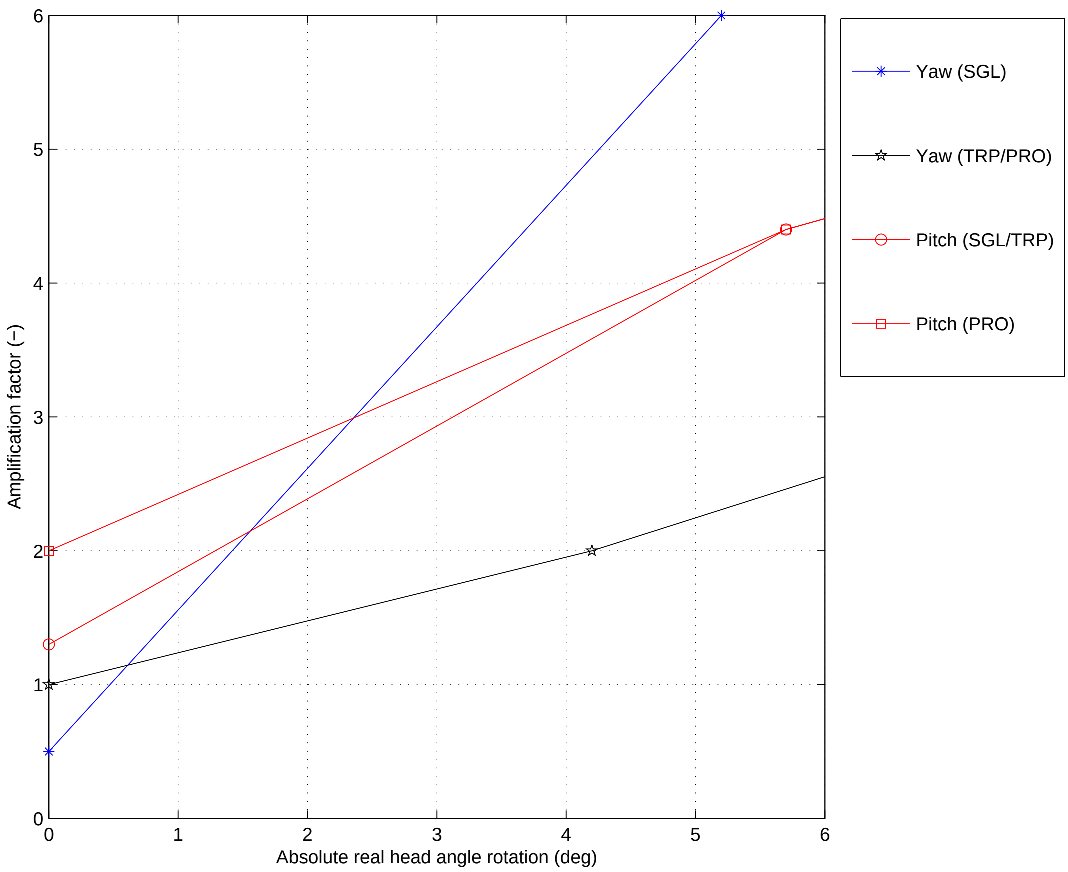


FIGURE 7.2: Amplification profile close-up of Figure 4.18 near origin

Research Question 5

How does amplified head rotations on larger displays affect visual flying technical performance in a more complex flying task?

In general, increased viewport size improves user performance, but it is task dependent [204]. By comparing between all augmented displays, the visual traffic circuit was extended to start from downwind. Again dividing the whole experiment scenario into flight segments, the downwind phase found a significant difference between the displays. However, this was only found for a very small effect size ($r = 0.05$) for the cross-track error with only significant differences found between the single display and the triple projector. A trend was observed suggesting larger errors with larger displays, but analysis of the pattern altitude and reference airspeed deviations did not find any differences. Upon starting the experiment, the aircraft was trimmed for level flight and airspeed, so if the participant did not touch the control the aircraft would continue dead straight ahead, so this explained the altitude and airspeed insignificant differences. The surprising drift to the right of track behaviour observed on the triple projectors can be explained by the fact that participants barely noticed slight visual scene changes on the larger display area since the overall big picture was consistent enough for them. In contrast, detecting central vision cues on the computer monitors was easier since the focus was already on the immediate central viewing area with less peripheral vision available.

The statistical analysis of Turn 1 (downwind-to-base) did not yield any differences between the displays. Keeping in mind the limited sample size ($n = 11$), this result showed that augmentation is an equalizing tool enabling participants to perform this turn across this variety of displays. Another weak effect size ($r = 0.05$) accounted for a significant difference during base leg, with only the altitude deviation found to be the main contribution. Since the pair-wise comparison test couldn't further discriminate any results, this was attributed to individual piloting technique and the limited sample size.

Turn 2 (base-to-final) was indifferent between the displays. Similar to Turn 1, this task was achievable on all displays due to the enabled augmentation participants could easily look left to spot the runway and orient themselves without any issues. The extra recorded measure of runway alignment error did spot a borderline difference, albeit with a weak effect size ($r = 0.03$): participants tended to overshoot the extended centerline the most on the projectors. An explanation for this was the different combined FOV and physical display area provided by the projectors compared to the computer monitors which just had the left and right screens turned off which meant participants might have reverted to using slightly altered visual cues provided by the virtual cockpit reference.

Although final approach yielded a significant difference and touchdown did not, the follow-up analysis for both segments could not determine a specific cause. This agrees with the findings of Section 7.2.2 wherein it was matched to literature that the approach and landing task could be sufficiently carried out using just the cues available in the central focal vision. Despite offering the largest physical display area, the horizontal FOV on the triple projectors was close to the triple monitors. Hence the peripheral vision differences were minimal to have made any impact on the landing task. With similar FOV specifications, curved displays have been shown to improve performance over flat displays [204] but the limited sample size and weak effect sizes previously found for the other dependent measures also reflect the limited exploratory nature of this experiment.

> **Research Question 6**
>
> How does display size in conjunction with head augmentation affect pilot workload?

The NASA-TLX results showed that there were no significant differences between the displays of the total workload as well as none for any of the subscales. This confirmed that the amplification profiles produced for the second experiment have achieved satisfactory user adaptation across such all three display systems. User frustration is significantly less on the larger displays than on single monitors, corresponding to the greater use of human visual capacity and more natural physical navigation (i.e. head movement) that reduces potential frustrations of virtual navigation [204] but the subjective results obtained here prove that the successful integration of this virtual reality technique can overcome this when profiles have been generated based on user feedback from a large group during development.

The secondary task measurement of blimp spotting resulted in a significant difference. This was solely caused by detection of the third and last blimp. A medium effect size ($r = 0.3$) showed that the single display had the lowest detection rate (Table 6.4), well over three times that of the lowest which was the triple projectors. This confirms that single-display users experienced the highest workload. Its visual limitations meant that the primary task of flying didn't leave spare capacity to perform the visual search to the same merit as the other displays. The detection rates for the first and second blimp confirms this as this was during the lower demanding flight segments where users had ample time and capacity left to search for visual traffic.

> **Research Question 7**
>
> Does simulator sickness scale with display size when using head augmented viewing?

The administered SSQ showed (in Table 6.5) that despite the statistics not revealing any significant differences (low sample size limitations), twice the number of participants reported this on the projected display compared to the single monitor. There was a trend indicating more nausea with increasing FOV. This was in line with fixed-based simulator research that high FOV results in higher levels of reported nausea [34]. This matches the explanation of increased stimulation of the peripheral retina in eliciting self-motion perception as discussed in Section 2.1.4. Some studies also report vection with small FOVs [205].

The results for the obtained SSQ total and factor scores were also yet unable to generate a statistical difference (due to limited sample size again). However, Figure 6.13 suggested a trend: a reduction in total and factors scores with increasing FOV. So despite the earlier higher occurrence of reported nausea with higher FOVs, the combined report and severity account by participants resulted in suggesting that head augmentation works better with larger FOVs. This results contradicts the established view that high FOV does not automatically mean more simulator sickness [206]. The user role is the distinguishing factor: passengers who are not in control of their own movement would experience sickness [34] whereas drivers (pilots in this case) are exempt from this view. Post-experiment feedback also confirmed this claim, as participants ranked the triple projector as their first choice, followed close second by the triple monitor, and single monitor as the least favored.

Chapter 8

Conclusions

This chapter discusses the main research findings shown in Chapter 7 and provides appropriate conclusions for the whole study. It also highlights the key contributions of the study to the aerospace and virtual reality research domains. Finally, current research limitations are discussed with recommendations made for future research.

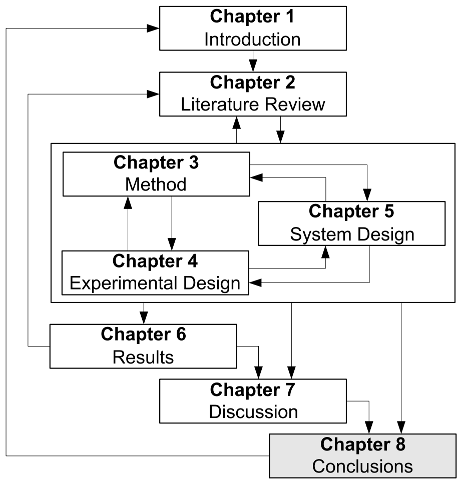


FIGURE 8.1: Chapter 8 in thesis roadmap

8.1 Conclusion

THE main goal of this research was to determine the effects of augmenting low-fidelity, fixed-based flight simulators with amplified head rotations to expand their mission rehearsal/training capability by evaluating the impact on pilot performance, workload and simulator sickness. This led to the derivation of two research objectives, each with a matching experiment to obtain responses to a total of seven supporting research questions.

The first objective, with three corresponding research questions, led to an initial control experiment. Its aim was to collect data as reference comparison between a basic augmented system and a common, non-augmented system. This helped find a response to the first research question of how the head interaction method influenced flight technical performance. The results showed that an augmented, single display with two degrees of head freedom in pitch and yaw had a significant, positive effect on how participants performed the base-to-final turn. The tighter turns closer to the runway generated a more rectangular pattern which more closely resembled the real life desired ground track [112]. The task of flying the final approach was found to be no different between the displays.

The second research question regarded the quantification of workload. The single augmented display was rated subjectively as being significantly more demanding in overall workload. This was primarily due to a higher physical demand, with the head amplification adding an extra interaction control. The secondary task of detecting the airborne blimp served as an objective workload measure to verify the subjective workload results. The statistical results, however, showed no significant differences between the two displays. This is an important finding, revealing that despite participants subjectively reporting that amplified head rotations increased workload, the objective data showed that it did not.

With regards to simulator sickness issues posed by the third and last research question, participants reported a higher number of nausea and oculomotor symptoms on the augmented single display with the total sickness score almost double that of the control case. All symptoms were of short duration though and had completely disappeared prior to the next test. This highlights the expanded flying task capability head augmentation enables, even on a single integrated display without an objective increase in workload demand. Furthermore, any occurrence of simulator sickness was primarily due to user adaptation and subsided after exiting the simulator.

With the first experiment having shown that amplified head rotations can be successfully applied to expand flying visual manoeuvres even on limited FOV displays, the second objective of this research was to extend the validity of this technique by scaling it to larger FOV displays. The goal now was to investigate if there is a point of diminishing returns. By comparing all three displays with augmentation applied, the visual flying circuit was extended to cover the downwind leg and the amount of airborne traffic was increased.

The first research question queried the profiling of amplification to larger displays. Based on user feedback from the first experiment, refinements were made to produce mapping profiles for the triple monitor and projectors. It was encouraging to see that the new technology introduced was accepted and users were able to adapt quickly. The results showed that this was well received as there were no subjective workload rating differences between all three configurations. This was further compounded by the flight technical performance: only a small difference was recorded, resulting in a drift to the right during downwind on larger FOVs. There were no significant differences between all displays with the remainder of the flight segments.

The subsequent three research questions were similar to that from the first objective, as flight performance, workload and simulator sickness effects were again evaluated. The biggest difference found was in the secondary task of visual scanning for traffic. The final approach segment resulted in the highest workload and participants had the lowest detection rate of blimps on the single display and the highest on the triple projectors. This clearly supports the benefit of having larger FOVs for demanding tasks requiring peripheral vision cues. The impact of simulator sickness was found to be the opposite to what the literature suggested regarding fixed-based simulators. Although there were more reports of nausea with larger displays/FOVs, the total Simulator Sickness Questionnaire (SSQ) score actually decreased with increasing FOV. This supports the effectiveness of scalability using amplified head rotations with increasing displays size and FOV. This view was supported by the finding that the triple projection was the favourite choice selected by participants.

In support of the two experiments, a new research flight simulator was designed and built that comprised two pilot stations with identical hardware systems using off-the-shelf technology. It used a systems engineering process in the design phase in order to identify technical requirements and match functional features to the ICAO 9625 qualification standard. This allowed the simulator to be constructed as a technology demonstrator representative of its intended fidelity class in current and future generations of devices following the ICAO standard. This added value to the experimental results by enabling their verification on other simulators qualified to the same standard.

In short, research and development in this knowledge area are very attractive for integration into low-cost flight simulators. These can either be retrofit or newly produced cost-effective training solutions that give performance gains. This can enable affordable, smart training that can be accessed by a large number of users. Hence, the increasing requirement to field networked training on a large scale and in conjunction with live/virtual/construct environments for further training benefits previously unattainable with lower-fidelity simulators can be met.

8.2 Research Contributions

This study has applications in both aerospace and virtual reality domains with regards to advancing novel technological advancements in physical human navigation interaction in immersive virtual environments [80]. It was achieved by systems integration into low-cost, low-fidelity flight simulation training devices [36] for future flight training benefits [5, 6, 15, 37]. The contributions as presented in this section to match the order of the research objectives stated in Section 1.3.2.

Using a utility-based approach, this research investigated the benefits of amplified head rotations specifically applied to the aviation domain via two pilot-in-the-loop simulator trials. With the first experiment providing a baseline reference for comparison against a common, non-augmented system, the second experiment expanded the understanding of head amplification research by evaluating scalability towards multi monitor and large projector displays. This thesis provides insight into optimal usage of this technique where the training task performance is enhanced in configurations without drawbacks such as extra workload or suffering more simulator sickness.

In support of this research, a new research flight simulator facility was also designed and built based on the latest flight simulation qualification standards [36]. The potential applications of this research are huge: this study offers the (aviation) simulation industry as well as consumers insight into optimal usage/systems integration of such head amplification techniques for upgrading their simulator equipment to maximize its (training) value.

8.2.1 Contribution One

Quantification of amplified head rotation effects in comparison with a non-augmented baseline reference.

1. The primary contribution was that this study supports the theory [207] that physical navigation methods such as natural head movement is indeed an efficient and valuable interaction method. It reduces dependency on less-efficient virtual navigation devices such as switches and hand controllers [204].

2. Introduced in Section 1.2, this fundamental research built upon the most recent virtual reality study [27] of amplified head rotations in a visual search task, but specifically applied this to aviation by relegating it as a secondary task together with a primary manual control task of flying a simulated aircraft.

3. By studying a visual-oriented task, this research contributed to the further understanding of FOV effects on manoeuvring performance [30], sharing commonality in synthetic vision display research discussed in Section 7.2.1.

4. The base-to-final turn scenario contributed to identifying pilot strategies/behaviour in their flight performance helping gain more insight into the human visual cues (Section 2.1) available to a pilot to judge and perform this turn [208, 209].

5. The recorded ground tracks flown on the augmented display yielded the best match to the desired rectangular traffic pattern (discussed in Section 3.4.2). Hence, this finding supported the virtual reality notion [210] that head-tracking interaction leads to improved user performance and spatial orientation.

6. The novelty of this experimental design compared to prior research in amplified head rotations rested in the comprehensive sampling of both objective and subjective workload for cross-verification, as well as being the first to explicitly measure simulator sickness in a fixed-base simulator.

7. The simulator sickness findings showed that simulator sickness symptoms quickly disappeared after use (Section 7.2.1). This is key for operational acceptance and potentially enables unimpeded training transfer to other training devices without time-based restrictions.

8.2.2 Contribution Two

Quantification of the scalability of amplified head rotations to larger displays and FOV.

1. The positive feedback from participants in the second experiment demonstrated that the empirical process of obtaining mapping profiles for larger displays by extrapolating them from smaller displays forms a viable subject for further research.

2. As such, the amplification profiling process revealed that large displays require less steep initial amplification gradients. The increase in centre dead zone offers more head stability and user comfort.

3. This research explicitly addressed the scientific agenda [204] of determining the breaking point where diminishing returns occur when amplified head rotations are scaled up to larger, more immersive virtual environments. In this case, moving from triple screens to the larger, triple-projector display did not yield significant improvements.

4. This research also demonstrated that novel technology proved to be an equalizer as the head augmentation led to very few flight technical performance differences between display sizes with equal user workload ratings.

5. User feedback contributed to the scientific basis that users exhibit greater engagement, immersion, and focus, with less distraction by tedious interface controls such as navigating through virtual environments [211].

6. The simulator sickness results contradicted existing research that automatically corroborates high FOV with increased sickness [206] but supported the exception when user are in control of their own view [34].

8.2.3 Supporting Contributions

Design and construction of a low-cost, flexible research flight simulator representative of current and future devices.

1. The final design of the research simulator produced a device representative of the low-cost, low-fidelity, fixed-based simulator for which amplified head

rotations were intended to augment. By following the ICAO 9625 qualification standards [36] throughout the systems design, results obtained from experiments in this simulator can be more easily be verified by other facilities qualified to a similar, transparent standard.

2. By matching features and fidelity levels to the ICAO 9625 qualification standards, the constructed simulator disproved the notion that higher display resolutions automatically allow flight and training simulators to be much more realistic and effective [211]. Instead, it has proven that this needs to be justified based on the training task.

3. The systems engineering approach to tackle the holistic issues during design ensured that the new research simulator facility was capable of hosting the current research project yet could be also be quickly adapted to support future projects.

4. The successful construction and operation of this flight simulator contributed to further widespread incorporation of commercial consumer technologies by demonstrating that low-cost solutions lead to system flexibility and quick turnaround times for research and development.

8.3 Future Research

With the experimental constraints and sample size limitations in this thesis, further work is required to fully understand the application and effects of amplified head rotations in flight simulators. Future research should strive to quantify user benefits in finer detail as well as providing more supporting data to validate the results of this research.

The key challenge faced in this research (common in aerospace studies) was the recruitment of sufficient participants to obtain desired sample sizes to generate statistical power. Hence, the primary focus of any future work should be the confirmation of these results using more pilots. Potential options could be to extend the run time of these experiments over a much longer period of time (up to a year). An alternative would be to build a portable facility and deploy it to flying clubs on location in order to provide access to a larger number of pilots. Considering the ability to recruit enough qualified participants, more advanced flying tasks would provide further evidence on the effects of head augmentation.

Additionally, this could be expanded to other domains such as driving and virtual navigation.

As the mapping of the amplification profiles was currently done using an empirical process, it would be beneficial to gain insight into mass user preferences in order to quantify optimal mapping profiles for a range of display configurations. This could be done by online surveying of profiles, perhaps in cooperation with the original hardware supplier. An analysis of raw physical head movement data in conjunction with frequency analysis is also suggested to further comprehend potential postural instabilities associated with simulator sickness.

The addition of the head roll axis into future studies is also highly recommended as this is particularly useful for stabilization at the extreme edges of yaw usage making it more natural for the user. This can then be followed by further studies that can expand the mapping into the full, final six degrees of freedom.

Finally, the key prerequisites allowing widespread implementation for flight training lies in fully documenting the transfer of training effects when using this augmentation and moving towards higher-fidelity simulators and/or the real aircraft. Access to such higher-fidelity aircraft or a flying laboratory would easily facilitate such an experiment. Long-term effects like skill retention should also be studied as there is sparse literature available whenever new technologies are introduced.

References

[1] Rolfe, J. and Staples, K. "Introduction to flight simulation." *Flight Simulation*, pp. 1–13. Cambridge University Press, Cambridge, UK, 1986.

[2] Carretta, T.R. and Dunlap, R.D. "Transfer of Training Effectiveness in Flight Simulation: 1986 to 1997." Tech. rep., Air Force Research Laboratory, Brooks AFB, TX, 1998.

[3] National Transportation Safety Board. "Review of U.S. Civil Aviation Accidents, Calendar Year 2010." Tech. rep., NTSB, Washington, DC, 2010.

[4] Australian Transport Safety Bureau. "Review of Midair Collisions Involving General Aviation Aircraft in Australia between 1961 and 2003." Tech. rep., Australian Government, Canberra, Australia, 2004.

[5] Zalcman, L. and Blacklock, J. "USAF Distributed Mission Operations, an ADF Synthetic Range Interoperability Model and an AOD Mission Training Centre Capability Concept Demonstrator - What Are They and Why Does the RAAF Need Them?" Tech. rep., Defence Science and Technology Organisation, Fishermans Bend, Australia, 2010.

[6] Schank, J.F., Thie, H.J., Graf II, C.M., Beel, J. and Sollinger, J.M. "Trade-offs between live and simulated training." *Finding the right balance: simulator and live training for Navy units*, chap. 3, pp. 43–66. RAND Corporation, Santa Monica, CA, 2002.

[7] Walker, D. and Geiss, K. "The PSI Handbook of Virtual Environments for Training and Education: Developments for the Military and Beyond, Vol 3: Integrated Systems, Training, Evaluations, and Future Directions." chap. Future Air Force Training, pp. 392–398. Praeger Security International, Westport, CT, 2009.

[8] Colegrove, C.M. and Bennett Jr., W. "Competency-Based Training: Adapting to Warfighter Needs." Tech. rep., Air Force Research Laboratory, Mesa, AZ, 2006.

[9] Schreiber, B.T., Schroeder, M. and Bennett Jr., W. "Distributed Mission Operations Within-Simulator Training Effectiveness." *The International Journal of Aviation Psychology*, vol. 21(3), pp. 254–268, Jul 2011.

[10] Schreiber, B.T., Stock, W.A. and Bennett Jr., W. "Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Metric Development and Objectively Quantifying the Degree of Learning." Tech. Rep. 3, Air Force Research Laboratory, Mesa, AZ, USA, 2006.

[11] Morrison, J.E. and Hammon, C. "On Measuring the Effectiveness of Large-Scale Training Simulations." Tech. Rep. 10, Institute for Defense Analyses, 2000.

[12] Krusmark, M., Schreiber, B.T. and Bennett Jr., W. "The Effectiveness of a Traditional Gradesheet for Measuring Air Combat Team Performance in Simulated Distributed Mission Operations." Tech. Rep. 5, Air Force Research Laboratory, Mesa, AZ, USA, 2004.

[13] Schreiber, B.T. and Bennett Jr., W. "Distributed Mission Operations Within-Simulator Training Effectiveness Baseline Study: Summary Report." Tech. Rep. 3, Air Force Research Laboratory, Mesa, AZ, USA, 2006.

[14] Crane, P., Robbins, R. and Bennett Jr., W. "Using Distributed Mission Training to Augment Flight Lead Upgrade Training." Tech. rep., Air Force Research Laboratory, Mesa, AZ, USA, 2001.

[15] Alexander, A.L., Brunye, T., Sidman, J. and Weil, S.A. "From Gaming to Training: A Review of Studies on Fidelity, Immersion, Presence, and Buy-in and Their Effects on Transfer in PC-Based Simulations and Games." *DARWARS Training Impact Group*, 11, pp. 1–14. Aptima, Inc., Woburn, MA, 2005.

[16] Hoog, T. "Distributed Mission Training overview." *99 Industry/Interservice Training Systems Conference*. National Security Industrial Association, 1999.

[17] Houck, M., Thomas, G. and Bell, H. "Training evaluation of the F-15 advanced air combat simulation." Tech. rep., Armstrong Laboratory, Williams, AZ, 1991.

[18] Berger, S. and Crane, P. "Multiplayer simulator-based training for air combat." *Proceedings of 15th Industry/Interservice Training Systems Conference*. National Security Industrial Association, 1993.

[19] Huddlestone, J., Harris, D. and Tinworth, M. "Air combat training - the effectiveness of multi-player simulation." *99 Industry/Interservice Training Systems Conference*. National Security Industrial Association, 1999.

[20] Bell, H., Dwyer, D., Love, J., Meliza, L., Mirabella, A. and Moses, F. "Recommendations for planning and conducting multi service tactical training with Distributed Interactive Simulation technology (A Four-Service Technical Report)." Tech. rep., U.S. Army Research Institute, Alexandria, VA, 1996.

[21] Nullmeyer, R.T. an Spiker, V. "The importance of crew resource management in MC-130P mission performance: Implications for training effectiveness evaluation." *Military Psychology*, vol. 15(1), pp. 77–96, 2003.

[22] Lollar, G. and Hambleton, O. "USAF Distributed Mission Operations (DMO) 2005 NATO M&S Group Conference." *The Effectiveness of Modelling and Simulation - From Anecdotal to Substantive Evidence. RTO-MP-MSG-035*. RTO, Neuilly-sur-Seine, France, 2005.

[23] Crane, P., Schiflett, S. and Oser, R. "RoadRunner 98: Training effectiveness in a Distributed Mission Training exercise." Tech. rep., Air Force Research Laboratory, Mesa, AZ, 1999.

[24] Andrews, D., Moses, F., Hawkins, H., Dunaway, M., Matthews, R. and Singer, T. "Recent Human Factors Contributions to Improve Military Operations." *Human Factors and Ergonomics Society Bulletin*, vol. 46(12), p. 1, 2003.

[25] Bradley, D.R. and Abelson, S.B. "Desktop flight simulators: Simulation fidelity and pilot performance." *Behavior Research Methods, Instruments, & Computers*, vol. 27(2), pp. 152–159, Jun 1995.

[26] Le Ngoc, L. and Kalawsky, R.S. "Evaluating Usability of Amplified Head Rotations on Base-to-Final Turn for Flight Simulation Training Devices." *Proceedings of IEEE Virtual Reality 2013*. IEEE, 2013.

[27] Kopper, R., Stinson, C. and Bowman, D.A. "Towards an Understanding of the Effects of Amplified Head Rotations." *3rd IEEE VR Workshop on Perceptual Illusions in Virtual Environments*. 2011.

[28] Poupyrev, I., Weghorst, S., Otsuka, T. and Ichikawa, T. "Amplifying spatial rotations in 3D interfaces." *CHI '99 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '99, pp. 256–257. ACM, New York, NY, 1999.

[29] Mulder, J. and van Liere, R. "Enhancing fish tank VR." *Proceedings IEEE Virtual Reality 2000*, pp. 91–98. IEEE Comput. Soc, 2000.

[30] Jansen, S.E.M., Toet, A. and Delleman, N.J. "Restricting the Vertical and Horizontal Extent of the Field-of-View: Effects on Manoeuvring Performance." *The Ergonomics Open Journal*, vol. 3, pp. 19–24, 2010.

[31] Jaekl, P., Allison, R., Harris, L., Jasiobedzka, U., Jenkin, H., Jenkin, M., Zacher, J. and Zikovitz, D. "Perceptual stability during head movement in virtual reality." *Proceedings IEEE Virtual Reality 2002*, pp. 149–155. IEEE Comput. Soc, 2002.

[32] Bolte, B., Bruder, G., Steinicke, F., Hinrichs, K. and Lappe, M. "Augmentation techniques for efficient exploration in head-mounted display environments." *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, pp. 11–18. ACM, New York, NY, 2010.

[33] Ni, T., Bowman, D.A. and Chen, J. "Increased display size and resolution improve task performance in Information-Rich Virtual Environments." *Proceedings of Graphics Interface 2006*, GI '06, pp. 139–146. Canadian Information Processing Society, 2006.

[34] Seay, A.F., Krum, D.M., Hodges, L. and Ribarsky, W. "Simulator Sickness and Presence in a High FOV Virtual Environment." *Proceedings of the Virtual Reality 2001 Conference*. 2001.

[35] Agency, D.P. "Findings of the 3rd Technology Maturation Study." *Managing Research under Output Ownership*. Joint MoD/Industry Workshop, 2004.

[36] ICAO. *Manual of Criteria for the Qualification of Flight Simulation Training Devices*, vol. I. International Civil Aviation Organization, 3rd edn., 2009.

[37] Allerton, D. "The case for flight simulation in general aviation." *Aeronautical Journal*, vol. 106(1065), pp. 607–618, 2002.

[38] Forsberg, K. and Mooz, H. "The Relationship of System Engineering to the Project Cycle." *Proceedings of the First Annual Symposium of NCOSE*. NCOSE, 1991.

[39] Dickerson, C. and Mavris, D. *Architecture and Principles of Systems Engineering*. Complex and Enterprise Systems Engineering. Taylor & Francis, 2010.

[40] Goldstein, E. *Sensation and Perception*. Thomson Wadsworth, Belmont, CA, 7th edn., 2007.

[41] Mann, M. "Vision." http://michaeldmann.net/mann7.html, 2007. (Accessed: 30-Aug-2013).

[42] Money, K. "Theory underlying the peripheral vision horizon device." *Proceedings of the NASA Peripheral Vision Display Conference*. NASA, Edwards, CA, 1983.

[43] Gibson, J. *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston, MA, 1966.

[44] Gibson, J. *The Pyschology of Knowing*. Gordon & Breach, New York, NY, 1972.

[45] McLeod, S. "SimplyPsychology." http://www.simplypsychology.org/perception-theories.html, 2007. (Accessed: 30-Aug-2013).

[46] Mollenhauer, M.A. "Simulator Adaptation Syndrome Literature Review." Tech. rep., Realtime Technologies Inc., Royal Oak, MI, 2004.

[47] Draper, M. *The Adaptive Effects of Virtual Interfaces: Vestibulo-Ocular Reflex and Simulator Sickness*. Ph.D. thesis, University of Washington, 1998.

[48] LaViola Jr., J.J. "A Discussion of Cybersickness in Virtual Environments." *SIGCHI Bulletin*, vol. 32(1), pp. 47–56, 2000.

[49] Zacharias, G. and Young, L. "Influence of combined visual and vestibular cues on human perception and control of horizontal rotation." *Experimental Brain Research*, vol. 41(2), Jan 1981.

[50] I/ITSEC. "Desktop to Dome: Air Force Integrates Global Training at the Speed of Real World Operations." *I/ ITSEC Show Daily*, 2010.

[51] Estock, J.L., Alexander, A.L., Stelzer, E.M. and Baughman, K. "Impact of Visual Scene Field of View on F-16 Pilot Performance." *Human Factors and Ergonomics Society Annual Meeting Proceedings*, pp. 75–79. 2007.

[52] Wickens, C.D. and Hollands, J.G. *Engineering Psychology and Human Performance*. Prentice Hall, Upper Saddle River, NJ, 2000.

[53] Adams, R. "27 Into 7 Equals 9625 - The New Regulatory Matrix." *Civil Aviation Training Magazine*, (1), 2009.

[54] Precision Flight Controls. "Desktop Systems." https://flypfc.com, 2011. (Accessed: 2-Dec-2011).

[55] Morrison, G. "Applying Simulation to First Officer Competencies." *2011 National Training Aircraft Symposium*. Daytona Beach, FL, 2011.

[56] Strachan, I. "Military Flight Simulator Census." *MS&T Magazine*, (4), pp. 51–70, 2011.

[57] Rehmann, A.J., Mitman, R.D. and Reynolds, M.C. "A Handbook of Flight Simulation Fidelity Requirements for Human Factors Research." Tech. Rep. 12, Federal Aviation Administration, Atlantic City International Airport, NJ, 1995.

[58] Bowman, D.A., Kruijff, E., LaViola, J.J. and Poupyrev, I. *3D User Interfaces: Theory and Practice*. Pearson Education. Addison-Wesley, Redwood City, CA, 2005.

[59] Kalawsky, R.S. "Keynote Address: What is Essential for Virtual Reality Systems to Meet Military Human Performance Goals?" *RTO MP-058*, vol. 44. 2000.

[60] Sisodia, A., Bayer, M., Townley-Smith, P., Nash, B., Little, J., Cassarly, W. and Gupta, A. "Advanced helmet mounted display (AHMD)." *Proceedings of SPIE*, vol. 6557, pp. 65570N–65570N–11, 2007.

[61] Patterson, R., Winterbottom, M.D. and Pierce, B.J. "Perceptual Issues in the Use of Head-Mounted Visual Displays." *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 48(3), pp. 555–573, Sep 2006.

[62] Rash, C.E., Russo, M.B., Letowski, T.R. and Schmeisser, E.T. *Helmet-Mounted Displays: Sensation, Perception and Cognition Issues*. U.S. Army Aeromedical Research Laboratory, Fort Rucker, AL, 2009.

[63] Winterbottom, M.D., Patterson, R. and Pierce, B.J. "Review and Analysis of Helmet-Mounted Displays for Use in Air Force Training and Simulation." Tech. Rep. March 2004, Air Force Research Laboratory, 2005.

[64] Covelli, J.M., Rolland, J.P., Proctor, M.D., Kincaid, J.P. and Hancock, P.A. "Field of View Effects on Pilot Performance in Flight." *The International Journal of Aviation Psychology*, vol. 20(2), pp. 197–219, Mar 2010.

[65] Brickner, M.S. and Foyle, D.C. "Field of View Effects on a Simulated Flight Task with Head-down and Head-up Sensor Imagery Displays." *Proceedings of the Human Factors Society 34th Annual Meeting*, pp. 1567–1571. HFES, Santa Monica, CA, 1990.

[66] Rolfe, J. and Staples, K. "Visual Systems." *Flight Simulation*, pp. 130–160. Cambridge University Press, Cambridge, UK, 1986.

[67] Winterbottom, M.D., Patterson, R., Gaska, J., Amann, R. and Prost, J. "Field of View Requirements for Deployable Training Systems: Effect of Field of View on Roll Control." *Human Factors and Ergonomics Society Annual Meeting Proceedings*, pp. 1560–1564. Human Factors and Ergonomics Society, 2008.

[68] Wells, M. and Venturino, M. "Performance and head movements using a helmet-mounted display with different sized fields-of-view." *Optical Engineering*, vol. 29(8), pp. 870–877, 1990.

[69] Grunwald, A., Kohn, S. and Merhav, S. "Visual field information in nap-of-the-earth flight by teleoperated helmet-mounted displays." vol. 1456, pp. 132–153. 1991.

[70] Woodruff, R., Hubbard, D. and Shaw, A. "Comparison of helmet-mounted visual displays for flight simulation." *Displays*, vol. 7(4), pp. 179–185, 1986.

[71] Steinicke, F., Bruder, G., Kuhl, S., Willemsen, P., Lappe, M. and Hinrichs, K.H. "Judgment of natural perspective projections in head-mounted display environments." *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pp. 35–42. ACM Press, New York, NY, 2009.

[72] Kuhl, S.A., Thompson, W.B. and Creem-Regehr, S.H. "HMD calibration and its effects on distance judgments." *ACM Transactions on Applied Perception*, vol. 6(3), pp. 1–20, Aug 2009.

[73] Schnell, T., Keller, M. and Etherington, T. "Trade-Offs in Synthetic Vision System Display Resolution, Field of Regard, Terrain Data Density, Texture, and Shading During Off-Path Operations." *International Journal of Aviation Psychology*, vol. 19(1), pp. 33–48, Jan 2009.

[74] Glaab, L., Kramer, L., Arthur, T., Parrish, R. and Barry, J. "Flight Test Comparison of Synthetic Vision Display Concepts at Dallas/Fort Worth International Airport." Technical Paper NASA/TP-2003-212177, Langley Research Center, National Aeronautics and Space Administration, Hampton, VA, 2003.

[75] Parrish, R.V., Bailey, R.E., Kramer, L.J., Jones, D.R., Young, S.D., Arthur III, J.J., Prinzel III, J., Harrah, S.D. and Glaab, L.J. "Aspects of Synthetic Vision Display Systems and the Best Practices of NASA's SVS Project." Tech. Rep. NASA/TP-2008-215130, NASA, May 2008.

[76] Jay, C. and Hubbold, R. "Amplifying Head Movements with Head-Mounted Displays." *Presence: Teleoperators and Virtual Environments*, vol. 12(3), pp. 268–276, Jun 2003.

[77] Teather, R.J. and Stuerzlinger, W. "Exaggerated head motions for game viewpoint control." *Proceedings of the 2008 Conference on Future Play Research, Play, Share - Future Play '08*, p. 240. ACM Press, New York, NY, 2008.

[78] Pausch, R., Proffitt, D. and Williams, G. "Quantifying Immersion in Virtual Reality." *Proceedings of SIGGRAPH 1997*. SIGGRAPH, 1997.

[79] Jerald, J., Peck, T., Steinicke, F. and Whitton, M. "Sensitivity to scene motion for phases of head yaws." *ACM Symposium on Applied Perception in Graphics and Visualization*, pp. 155–162. ACM Press, New York, NY, 2008.

[80] Dourish, P. *Where the Action Is: The Foundations of Embodied Interaction*. The MIT Press, 2001.

[81] Le Ngoc, L. and Kalawsky, R.S. "Visual Circuit Flying with Augmented Head-tracking on Limited Field of View Flight Training Devices." *AIAA Modeling and Simulation Technologies Conference 2013*. AIAA, 2013.

[82] Le Ngoc, L. and Kalawsky, R.S. "Expanding Virtual Mission Training via Low-cost Head-tracking Augmentation." *RAeS Spring Flight Simulation Conference 2013*. Royal Aeronautical Society, 2013.

[83] So, R. and Griffin, M. "Simulated and virtual realities." chap. Head-coupled virtual environment with display lag, pp. 103–111. Taylor & Francis, Inc., Bristol, PA, 1995.

[84] Lewis, C., Higgins, A., Wells, M. and Griffin, M. "Stabilisation of head-coupled systems for rotational head motion." Tech. rep., MacAuley-Brown, Inc., Canberra, Australia, 1987.

[85] Liu, A., Tharp, G., French, L., Lai, S. and Stark, L. "Some of what one needs to know about using head-mounted displays to improve teleoperator performance." *Robotics and Automation, IEEE Transactions on*, vol. 9(5), pp. 638–648, 1993.

[86] Rolfe, J.M. and Caro, P.W. "Determining the training effectiveness of flight simulators: some basic issues and practical developments." *Applied ergonomics*, vol. 13(4), pp. 243–50, Dec 1982.

[87] Allerton, D. *Principles of Flight Simulation*. Aerospace Series. Wiley, Chippenham, UK, 2009.

[88] Borst, C. *Ecological Approach to Pilot Terrain Awareness*. Ph.D. thesis, Delft University of Technology, 2009.

[89] Kowalski-Trakofler, K., Vaught, C. and Scharf, T. "Judgment and Decision Making Under Stress: an Overview for Emergency Managers." *International Jouornal of Emergency Management*, vol. 1(3), pp. 278–289, 2003.

[90] Rose, A., Radtke, P., Shettel, H. and Hagman, J. "Users manual for predicting military task retention (ARI Report No. 85-26)." Tech. rep., US Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA, 1985.

[91] Loftus, G. "Evaluating forgetting curves." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 11(2), pp. 397–406, 1985.

[92] Healy, A. and Sinclair, G. "Long term retention of trained skills (Technical Report No 94-001)." Tech. rep., Naval Air Warfare Center Training Systems Division, Orlando, FL, 1994.

[93] Martin, E. "Training effectiveness of platform motion: Review of motion research involving the advanced simulator for pilot training and the simulator from air-to-air combat (AFHRL Report No.TR-79-51)." Tech. rep., Air Force Human Resources Laboratory, Williams, AZ, 1981.

[94] Hays, R.T., Jacobs, J.W., Prince, C. and Salas, E. "Requirements for Future Research in Flight Simulation Training: Guidance Based on a Meta-Analytic

Review." *The International Journal of Aviation Psychology*, vol. 2(2), pp. 143–158, 1992.

[95] Waag, W.L. "Training Effectiveness of Visual and Motion Simulation." Tech. rep., Air Force Human Resources Laboratory, Williams AFB, AZ, 1981.

[96] Bell, H.H. and Waag, W.L. "Evaluating the Effectiveness of Flight Simulators for Training Combat Skills: A Review." *The International Journal of Aviation Psychology*, vol. 8(3), pp. 223–242, 1998.

[97] Norman, G., Dore, K. and Grierson, L. "The minimal relationship between simulation fidelity and transfer of learning." *Medical Education*, vol. 46(7), pp. 636–647, 2012.

[98] Research, A.G.F.A. and Development. "Fidelity of Simulation for Pilot Training (AGARD-AR-159)." Tech. rep., North Atlantic Treaty Organization, Orlando, FL, 1980.

[99] Northam, G. "Simulation fidelity - getting in touch with reality." *Proceedings of SimTecT 2000*. Simulation Industry Association of Australia, 2000.

[100] Gross, D. "Report from the Fidelity Implementation Study Group." *Fidelity Implementation Study Group Glossary 3.0*. ISG, 1998.

[101] Thomas, M. "Operational fidelity in simulation-based training: the use of data from threat and error management analysis in instructional systems design." *Proceedings of SimTecT2003: Simulation Conference*. SimTecT 2003 Organising and Technical Committees, 2003.

[102] Macfarlane, R. "Simulation as an instructional procedure." G. Hunt, ed., *Designing instruction for Human Factors training in aviation*. Ashgate, UK, 1997.

[103] Schneider, W. "Training high-performance skills: Fallacies and Guidelines." *Human Factors*, vol. 27(3), pp. 285–300, 1985.

[104] Thomas, M. "Integrating Low-Fidelity Desktop Scenarios into the High-Fidelity Simulation Curriculum in Medicine and Aviation." *Proceedings of SimTecT 2004*. Simulation Industry Association of Australia, 2004.

[105] Micheli, G.S. "Analysis of the Transfer of Training, Substitution, and Fidelity of Simulation of Transfer Equipment." Tech. rep., Naval Training Equipment Center, Orlando, FL, 1972.

[106] Mulder, M. *Cybernetics of Tunnel-in-the-Sky Displays.* Ph.D. thesis, Delft University of Technology, 1999.

[107] Chien Wei, C. "Low Cost Virtual Cockpits for Air Combat Experimentation." *Interservice/Industry Training, Simulation, and Education Conference*, 1596, pp. 1–11. 2004.

[108] United States Army Aviation and Missile Command. *ADS-33E-PRF Aeronautical Design Standard Performance Specification Handling Qualities Requirements for Military Rotorcraft.* Redstone, AL, 2000.

[109] Lintern, G. and Liu, Y.T. "Explicit and implicit horizons for simulated landing approaches." *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 33(4), pp. 401–417, 1991.

[110] Lintern, G., Taylor, H., Koonce, J., Kaiser, R. and Morrison, G. "Transfer and Quasi-Transfer Effects of Scene Detail and Visual Augmentation in Landing Training." *The International Journal of Aviation Psychology*, vol. 7(2), pp. 149–169, Apr 1997.

[111] Lintern, G. "An Analysis of Slant for Guidance of Landing Approaches." *The International Journal of Aviation Psychology*, vol. 10(4), pp. 363–376, Oct 2000.

[112] U.S. Federal Aviation Administration. *Airplane Flying Handbook.* U.S. Department of Transportation, Washington, DC, FAA-H-8083-3A edn., 2004.

[113] U.S. Federal Aviation Administration. *Private Pilot Practical Test Standards for Airplane (SEL, MEL, SES, MES).* U.S. Department of Transportation, Washington, DC, FAA-S-8081-14B edn., 2011.

[114] U.S. Federal Aviation Administration. *Aeronautical Information Manual.* U.S. Department of Transportation, Washington, DC, FAA AIM 4-3-2 edn., 2009.

[115] US Air Force. *AFMAN11-248 T-6 Primary Flying.* Bibliogov, 2006.

[116] Rehmann, A. "Handbook of Human Performance Measures and Crew Requirements for Flightdeck Research." Tech. rep., Federal Aviation Administration, Wright-Patterson, OH, 1995.

[117] Miller, S. "Workload Measures: Literature Review." Tech. rep., National Advanced Driving Simulator, Iowa City, IA, 2001.

[118] Rubio, S., Díaz, E., Martín, J. and Puente, J.M. "Evaluation of Subjective Mental Workload: A Comparison of SWAT , NASA-TLX , and Workload Profile Methods." *Applied Psychology: An International Review*, vol. 53(1), pp. 61–86, 2004.

[119] Hart, S. and Staveland, L. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." P. Hancock and N. Meshkati, eds., *Human Mental Workload*, *Advances in Psychology*, vol. 52, pp. 139–183. North-Holland, Amsterdam, the Netherlands, 1988.

[120] Kennedy, R., Drexler, J., Compton, D., Stanney, K., Lanham, D. and Harm, D. "Configural scoring of simulator sickness, cybersickness and space adaptation syndrome: similarities and differences." L. Hettinger and M. Haas, eds., *Virtual and Adaptive Environments Applications Implications and Human Performance Issues*, chap. 12, pp. 247–278. CRC Press, Boca Raton, FL, 2003.

[121] Draper, M.H., Viirre, E.S., Furness, T.A. and Gawron, V.J. "Effects of Image Scale and System Time Delay on Simulator Sickness within Head-Coupled Virtual Environments." *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 43(1), pp. 129–146, Jan 2001.

[122] Kennedy, R., Lanham, D., Drexler, J., Massey, C. and Lilienthal, M. "Cybersickness in several flight simulators and VR devices: a comparison of incidences, symptom profiles, measurement techniques and suggestions for research." *Proc. FIVE95 (Framework for Immersive Virtual Environments)*. 1995.

[123] Kennedy, R.S., Lane, N.E., Berbaum, K.S. and Lilienthal, M.G. "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness." *The International Journal of Aviation Psychology*, vol. 3(3), pp. 203–220, 1993.

[124] Straus, S.H. "SPR 559 - New, Improved, Comprehensive, and Automated Drivers License Test and Vision Screening System." Tech. Rep. 1, Arizona Department of Transportation, Phoenix, AZ, 2005.

[125] Kolasinski, E. "Simulator sickness in virtual environments." Tech. rep., U.S. Army Research Institute for the Behavioural and Social Sciences, Alexandria, VA, 1995.

[126] Pausch, R., Crea, T. and Conway, M. "A literature survey for virtual environments: military flight simulator visual systems and simulator sickness." *Presence: Teleoper Virtual Environ*, vol. 1(3), pp. 344–363, Jul 1992.

[127] Renkewitz, H. and Alexander, T. "Perceptual Issues of Augmented and Virtual Environments." Tech. rep., Defense Technical Information Center, Fort Belvoir, VA, 2007.

[128] Johnson, D.M. "Introduction to and Review of Simulator Sickness Research." Tech. rep., U.S. Army Research Institute for the Behavorial and Social Sciences, Arlington, VA, 2005.

[129] R.S., K. and Fowlkes, J. "Simulator sickness is polygenic and polysymptomatic: Implications for research." *International Journal of Aviation Psychology*, vol. 1(1), pp. 23–28, 1992.

[130] Balk, S.A., Bertola, A. and Inman, V.W. "Simulator Sickness Questionnaire: Twenty Years Later." *7th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*. 2013.

[131] Ison, D. "An Analysis of Statistical Power in Aviation Research." *International Journal of Applied Aviation Studies*, vol. 11(1), pp. 67–84, 2011.

[132] XHSI. "XHSI - eXternal High-fidelity Simulator Instruments." http://xhsi.sourceforge.net/, 2012. (Accessed: 30-Jan-2012).

[133] Laminar Research. "X-Plane.com." http://www.x-plane.com/store/desktop/x-plane_9_flight_simulator/, 2012. (Accessed: 12-Mar-2012).

[134] Naturalpoint, Inc. "TrackIR." http://www.naturalpoint.com/trackir/, 2012. (Accessed: 30-Nov-2011).

[135] Keller, M., Schnell, T., Lemos, K., Glaab, L. and Parrish, R. "Pilot performance as a function of display resolution and field of view in a simulated terrain following flight task using a synthetic vision system." *22nd Digital Avionics Systems Conference Proceedings*, pp. 9.E.5–91. IEEE, Indianapolis, IN, 2003.

[136] DeVries, S. and Padmos, P. "Steering a simulated unmanned aerial vehicle using a head-slaved camera and hmd: effect of hmd quality, visible vehicle references and extended stereo cueing." *SPIE Conference on Helmet and Head Mounted Displays III*. SPIE, 1998.

[137] Kappe, B., van Erp, J. and Kortelin, J. "Effects of head-slaved and peripheral displays on lane-keeping performance and spatial orientation." *Human Factors*, vol. 41(3), pp. 453–466, 1999.

[138] Kappe, B. and Padmos, P. "Headslaved displays in a driving simulator." *Proceedings of the 1st Human-Centered Transportation Simulation Conference*. The University of Iowa, 2001.

[139] Tan, D.S., Gergle, D., Scupelli, P.G. and Pausch, R. "With similar visual angles, larger displays improve spatial performance." *Proceedings of the conference on Human factors in computing systems - CHI '03*. ACM Press, New York, NY, 2003.

[140] Coolican, H. *Research Methods and Statistics in Psychology*. Hodder Education, London, UK, 5th edn., 2009.

[141] Field, A. *Discovering Statistics Using SPSS*. SAGE Publications, London, UK, 3rd edn., 2009.

[142] Vitiello, P. *Use of Performance Predictors in Visual Analytics*. Ph.D. thesis, Loughborough University, 2013.

[143] Razali, N. and Wah, Y. "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests." *Journal of Statistical Modeling and Analytics*, vol. 2(1), pp. 21–33, 2011.

[144] Ahad, N., Yin, T., Rahman, A. and Yaacob, C. "Sensitivity of Normality Tests to Non-normal Data." *Sains Malaysiana*, vol. 40(6), pp. 637–641, 2011.

[145] Glass, G., Peckham, P. and Sanders, J. "Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance." *Review of Educational Research*, vol. 42(3), pp. 237–288, 1972.

[146] Borst, C., Sjer, F.A., Mulder, M., Van Paassen, M.M. and Mulder, J.A. "Ecological Approach to Support Pilot Terrain Awareness After Total Engine Failure." *Journal of Aircraft*, vol. 45(1), pp. 159–171, Jan 2008.

[147] Olson, C. "Comparative Robustness of Six Tests in Multivariate Analysis of Variance." *Journal of the American Statistical Association*, vol. 69(348), pp. 894–908, 1974.

[148] Lunney, G. "Using Analysis of Variance with a Dichotomous Dependent Variable: an Empirical Study." *Journal of Educational Measurement*, vol. 7(4), pp. 263–269, 1974.

[149] McMillan, J. and Foley, J. "Reporting and Discussing Effect Size: Still the Road Less Traveled?" *Practical Assessment, Research & Evaluation*, vol. 16(14), pp. 1–11, 2011.

[150] Cohen, J. "A power primer." *Psychological Bulletin*, vol. 112(1), pp. 155–159, 1992.

[151] Prajapati, B., Dunne, M. and Armstrong, R. "Sample size estimation and statistical power analyses." *Optometry Today*, vol. 16(7), 2010.

[152] Faul, F., Erdfelder, E., Buchner, A. and Lang, A.G. "Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses." *Behavior Research Methods*, vol. 41(4), pp. 1149–1160, 2009.

[153] Leahey, E. "Alphas and asterisks: The development of statistical significance testing standards in Sociology." *Social Forces*, vol. 84(1), pp. 1–24, 2005.

[154] Noymer, A. "Alpha, significance level of test." P.J. Lavrakas, ed., *Encyclopedia of Survey Research Methods*. SAGE Publications, Inc., Thousand Oaks, CA, 2008.

[155] Hayes, A., Gee, T., Price, I., Cooksey, R. and Patrech, A. "What α-level?" http://www.une.edu.au/WebStat/unit_materials/c5_inferential_statistics/what_alpha_level.html, 2000. (Accessed: 1-Oct-2013).

[156] Moses, J. "Foundational issues in engineering systems: A framing paper." *MIT 2004 Engineering Systems Symposium*. MIT, Cambridge, MA, 2004.

[157] Systems Engineering Applications Technical Committee. "Systems Engineering Applications Profiles Version 3.0." Tech. rep., International Council On Systems Engineering, Seattle, WA, 2000.

[158] Dykes, K. and Meadows, R. "Applications of Systems Engineering to the Research, Design, and Development of Wind Energy Systems." Tech. rep., National Renewable Energy Laboratory, Golden, CO, 2011.

[159] Honour, E. "Understanding the Value of Systems Engineering." *INCOSE International Symposium*. INCOSE, Toulouse, France, 2004.

[160] Department of Defense Systems Management College. *Systems Engineering Fundamentals*. Defense Acquisition University Press, Fort Belvoir, VA, 2001.

[161] Wright, C. and Kalawsky, R. "Simulating overload of aircrew attention for optimal training benefit." http://tinyurl.com/avrrc-overload, 2013. (Accessed: 30-Oct-2013).

[162] Presagis USA, Inc. "Presagis Vega Prime." http://www.presagis.com/products_services/products/modeling-simulation/visualization/vega_prime/, 2011. (Accessed: 18-May-2011).

[163] Lockheed Martin. "Prepar3D." http://www.prepar3d.com, 2011. (Accessed: 11-Sep-2011).

[164] FlightGear. "FlightGear Flight Simulator." http://www.flightgear.org/, 2013. (Accessed: 30-Oct-2013).

[165] Wikipedia. "Microsoft Flight Simulator X." http://en.wikipedia.org/wiki/Microsoft_Flight_Simulator_X, 2013. (Accessed: 30-Oct-2013).

[166] Eagle Dynamics. "Digital Combat Simulator." http://www.digitalcombatsimulator.com/en/, 2013. (Accessed: 30-Oct-2013).

[167] Craighead, J., Murphy, R., Burke, J. and Goldiez, B. "A Survey of Commercial Open Source Unmanned Vehicle Simulators." *2007 IEEE International Conference on Robotics and Automation*, pp. 852–857. 2007.

[168] Ertem, M.C. "An airborne synthetic vision system with HITS symbology using X-Plane for a head up display." *The 24th Digital Avionics Systems Conference*. 2005.

[169] Ribeiro, L. and Oliveira, N.M.F. "UAV autopilot controllers test platform using Matlab/Simulink and X-Plane." *IEEE Frontiers in Education Conference*. 2010.

[170] Goh, J. and Wiegmann, D. "An Investigation of the Factors that Contribute to Pilots' Decisions to Continue Visual Flight Rules Flight into Adverse Weather." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 45(2), pp. 26–29, 2001.

[171] Obutto. "Obutto." http://www.obutto.com/home.html, 2012. (Accessed: 22-Feb-2012).

[172] Bussemaker, J. "BlackBox flightlogger." http://www.utr-online.com/Software.asp?page=4, 2012. (Accessed: 30-Jan-2012).

[173] Fly Elise-ng. "Immersive Display Pro." http://www.fly.elise-ng.net/index.php/immersive-display-pro, 2012. (Accessed: 17-May-2011).

[174] UniVisual. "Warpalizer Warp and Blend Software." http://www.warpalizer.com/, 2012. (Accessed: 4-Mar-2012).

[175] NTHUSIM. "NTHUSIM Plus Edition." http://www.nthusim.com/, 2011. (Accessed: 18-May-2011).

[176] Richards, J. "The Measurement of Human Motion: A Comparison of Commercially Available Systems." *Proceedings 5th International Symposium 3D Analysis of Human Movement*. 1998.

[177] Foxlin, E. "Handbook of Virtual Environment Technologies." chap. Motion Tracking Technologies and Requirements, pp. 163–210. Lawrence Erlbaum Publishers, Hillsdale, NJ, 2002.

[178] Welch, G. and Foxlin, E. "Motion Tracking: No Silver Bullet, But a Respectable Arsenal." *IEEE Computer Graphics and Applications*, vol. 22(6), pp. 24–38, 2002.

[179] Corporation, M. "Kinect for Windows." http://www.microsoft.com/en-us/kinectforwindows/, 2013. (Accessed: 31-Oct-2013).

[180] Vriend, W. "FaceTrackNoIR: FaceTracking with just your webcam!" http://facetracknoir.sourceforge.net, 2013. (Accessed: 31-Oct-2013).

[181] Seeing Machines. "faceAPI." http://www.seeingmachines.com/product/faceapi/, 2013. (Accessed: 31-Oct-2013).

[182] Plewe, J. "Head banging..." https://weblogs.java.net/blog/herkules/archive/2007/09/head_banging.html, 2007. (Accessed: 7-Apr-2011).

[183] Arenalogic. "Link 16 Part Task Trainer." http://www.arenalogic.com/, 2011. (Accessed: 7-Apr-2011).

[184] Barbour, S. "Pilot View." http://www.xpluginsdk.org/pilot_view.htm, 2012. (Accessed: 7-Jul-2012).

[185] Le Ngoc, L., Borst, C., Mulder, M. and Van Paassen, M. "The Effect of Synthetic Vision Enhancements on Landing Flare Performance." *AIAA Guidance, Navigation and Control Conference 2010*. AIAA, 2010.

[186] Payne, K.H. and Harris, D. "A Psychometric Approach to the Development of a Multidimensional Scale to Assess Aircraft Handling Qualities." *The International Journal of Aviation Psychology*, vol. 10(4), pp. 343–362, 2000.

[187] Harper, R.P. and Cooper, G.E. "Handling qualities and pilot evaluation." *Journal of Guidance, Control, and Dynamics*, vol. 9(5), pp. 515–529, Sep 1986.

[188] Hodgkinson, J. *Aircraft Handling Qualities*. AIAA Education Series. American Institute of Aeronautics and Astronautics, Oxford, UK, 1999.

[189] Borst, C., Mulder, M. and Van Paassen, M.M. "Design and Simulator Evaluation of an Ecological Synthetic Vision Display." *Journal of Guidance, Control, and Dynamics*, vol. 33(5), pp. 1577–1591, Sep 2010.

[190] Mueller, M., McNally, D., Rentas, T., Aweiss, A., Thipphavong, D., Gong, C., Cheng, J., Walton, J., Walker, J., Lee, C., Sahlman, S. and Carpenter, D. "Controller and Pilot Evaluation of a Datalink-Enabled Trajectory-Based Operations Concept." *9th USA/Europe Air Traffic Management Research and Development Seminar*. ATM2011, Berlin, Germany, 2011.

[191] Bourke, P. "iDome: Immersive gaming with the Unity game engine." *Proceedings of the Computer Games & Allied Technology 09*. Research Publishing Services, 2009.

[192] Flight Safety Foundation. "Stabilized Approach and Flare Are Keys to Avoiding Hard Landings." *Flight Safety Digest*, vol. 23(8), pp. 1–25, 2004.

[193] Gee, S.W., Carr, P.C., Winter, W.R. and Manke, J.A. "Development of Systems and Techniques for Landing an Aircraft Using Onboard Television." Tech. Rep. NASA TP-1171, National Aeronautics and Space Administration, Feb 1978.

[194] Goodman, A., Foyle, D.C., Hooey, B.L. and Wilson, J.R. "Characterizing visual performance during approach and landing with and without a synthetic vision display: a part study." *Proceedings of the 2003 NASA Aviation Safety Program Conference on Human Performance Modeling of Approach and Landing with Augmented Displays, Moffett Field (CA), March 6*, (NASA/CP-2003-212267), pp. 71–89, 2003.

[195] Snow, P., Alter, K. and Davis, R.C. "Challenges of reducing landing minima and enabling single-pilot operations through the use of synthetic

vision/highway-in-the-sky technologies." *Proceedings of the 24th Digital Avionics Systems Conference, Washington (DC), October 30 - November 3*, vol. 2, pp. 13.C5–1–13.C.5–11, 2005.

[196] Gibb, R., Schvaneveldt, R. and Gray, R. "Visual Misperception in Aviation: Glide Path Performance in a Black Hole Environment." *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50(4), pp. 699–711, Aug 2008.

[197] Entzinger, J.O. and Suzuki, S. "Visual Cues in Manual Landing of Airplanes." *Proceedings of {KSAS}–{JSASS} Joint International Symposium on Aerospace Engineering*, pp. 388–395. The Korean Society for Aeronautical & Space Sciences, Nov 2008.

[198] Gibb, R. and Gray, R. "Terrain Orientation Theory for a Visual Approach to Landing." *Human Factors and Ergonomics Society Annual Meeting Proceedings*, vol. 50, pp. 1661–1665(5), 2006.

[199] Air Safety Institute. "Mastering Takeoffs and Landings." Tech. Rep. Operations and Profiency No. 6, AOPA Air Safety Foundation, Frederick, MD, USA, 2008.

[200] Kim, J., Palmisano, S.A., Ash, A. and Allison, R.S. "Pilot gaze and glideslope control." *ACM Transactions on Applied Perception*, vol. 7(3), pp. 1–18, Jun 2010.

[201] Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A. and Wickens, C.D. "Comparison of Expert and Novice Scan Behaviors During VFR Flight." *11th International Symposium on Aviation Psychology*. Columbus, OH, 2001.

[202] Tyndiuk, F., Thomas, G., Lespinet-Najib, V. and Schlick, C. "Cognitive Comparison of 3D Interaction in Front of Large vs. Small Displays." *Proceedings of the ACM symposium on Virtual reality software and technology - VRST '05*, pp. 117–123. ACM Press, Monterey, CA, 2005.

[203] Bishop, I. and Rohrmann, B. "Subjective responses to simulated and real environments: a comparison." *Landscape and Urban Planning*, vol. 65(4), pp. 261–277, 2003.

[204] Shupp, L., Ball, R., Yost, B., Booker, J. and North, C. "Evaluation of viewport size and curvature of large, high-resolution displays." *Proceedings of Graphics Interface 2006*, GI '06, pp. 123–130. Canadian Information Processing Society, 2006.

[205] Warren, W.H. and Kurtz, K.J. “The role of central and peripheral vision in perceiving the direction of self-motion.” *Perception & Psychophysics*, vol. 51(5), pp. 443–454, 1992.

[206] Lin, J.J.W., Duh, H.B.L., Abi-Rached, H., Parker, D.E. and Iii, T.A.F. “Effects of Field of View on Presence, Enjoyment, Memory, and Simulator Sickness in a Virtual Environment.” *Proceedings of the IEEE Virtual Reality Conference 2002*, VR ’02. IEEE Computer Society, 2002.

[207] Ball, R., North, C. and Bowman, D.A. “Move to improve: promoting physical navigation to increase user performance with large displays.” *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’07, pp. 191–200. ACM, New York, NY, 2007.

[208] Beall, A.C. and Loomis, J.M. “Optic Flow and Visual Analysis of the Base-to-Final Turn.” *The International Journal of Aviation Psychology*, vol. 7(3), pp. 201–223, Jul 1997.

[209] Beall, A.C. *Visual control of the base-to-final turn in fixed-wing aircraft.* Ph.D. thesis, University of California, 1998.

[210] Bowman, D., Datey, A., Ryu, Y., Farooq, U. and Vasnaik, O. “Empirical Comparison of Human Behavior and Performance with Different Display Devices for Virtual Environments.” *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, pp. 2134–2138. SAGE Publications, 2002.

[211] Sabri, A., Ball, R., Fabian, A., Bhatia, S. and North, C. “High-resolution gaming: Interfaces, notifications, and the user experience.” *Interacting with Computers*, vol. 19(2), pp. 151–166, 2007.

[212] Sadin, S., Povinelli, F. and Rosen, R. “The NASA technology push towards future space mission systems.” *39th International Astronautical Congress*. IAF, 1988.

[213] Shanks, A. “Technology management in warship acquisition.” *Technology Management, Acquisition and Planning at INEC 2008*. INEC, 2008.

[214] Sauser, B., Ramirez-Marquez, J., Verma, D. and Gove, R. “From TRL to SRL: The Concept of Systems Readiness Levels.” *Conference on Systems Engineering Research*. CSER, 2006.

[215] Bedford, G. and Kalawsky, R. "Operator Training and Performance Measurement in Remotely Piloted Aerial Systems." *RAeS Autumn Flight Simulation Conference 2012*. Royal Aeronautical Society, 2012.

[216] Bedford, G. and Kalawsky, R. "Operator Training & Performance Measurement in Remotely Piloted Aerial Systems." http://tinyurl.com/avrrc-uav, 2013. (Accessed: 30-Oct-2013).

[217] Laminar Research. "How X-Plane Works." http://www.x-plane.com/desktop/how-x-plane-works/, 2013. (Accessed: 30-Oct-2013).

[218] X-Rite, Inc. "X-Rite i1Display 2 Professional Monitor Calibration i1 solutions." http://www.xrite.com/i1display-2, 2011. (Accessed: 20-Aug-2012).

[219] basICColor GmbH. "basICColor display 5 High End Monitor Calibration and Profiling Software." http://www.basiccolor.de/basiccolor-display-5-en/, 2013. (Accessed: 30-Oct-2013).

Appendix A

List of Publications

Parts of this thesis have been published and orally presented at international, peer-reviewed conferences, as summarized in the following list. The AIAA and IEEE papers are also being considered for journal submission:

Peer-reviewed conference papers/proceedings

- Le Ngoc, L. and Kalawsky, R.S. "Visual Circuit Flying with Augmented Head-tracking on Limited Field of View Flight Training Devices." *AIAA Modeling and Simulation Technologies Conference 2013.* AIAA, Boston, MA, USA, Aug 2013.
- Le Ngoc, L. "Expanding Virtual Mission Training via Low-cost Head-tracking Augmentation." *RAeS Spring Flight Simulation Conference 2013.* Royal Aeronautical Society, London, UK, Jun 2013.
- Le Ngoc, L. and Kalawsky, R.S. "Evaluating Usability of Amplified Head Rotations on Base-to-Final Turn for Flight Simulation Training Devices." *Proceedings of IEEE Virtual Reality 2013.* IEEE, Orlando, FL, USA, Mar 2013.
- Le Ngoc, L. and Kalawsky, R.S. "A Systems Framework to Off-the-Shelf Technology Integration for Flight Simulation Research." *Workshop on Off-The-Shelf Virtual Reality 2013.* IEEE, Orlando, FL, USA, Mar 2013.

Appendix B

Flight Simulation Training Device Features

To assist in the definition of the devices and to provide focus for the training analysis it was decided to breakdown any FSTD into some key components that would lead towards the construction of the FSTD Specification. Consequently twelve FSTD features were defined from a training perspective that, used together, create an FSTD as follows [36] on the next pages:

1. **Cockpit layout & structure:** Defines the physical structure and layout of the cockpit environment, instrument layout and presentation, controls and pilot, instructor and observer seating.

2. **Flight model (aero & engine):** Defines the mathematical models and associated data to be used to describe the aerodynamic and propulsion characteristics required to be modeled in the FSTD.

3. **Ground handling:** Defines the mathematical models and associated data to be used to describe the ground handling characteristics and runway conditions required to be modeled in the FSTD.

4. **Aircraft systems:** Defines the types of aircraft systems simulation required to be modeled in the FSTD. The ATA chapter definitions describe these in more detail (e.g. hydraulic power, fuel, electrical power, etc.). Systems simulation will allow normal, abnormal and emergency procedures to be accomplished.

5. **Flight controls and forces:** Defines the mathematical models and associated data to be used to describe the flight controls and flight control force and dynamic characteristics required to be modeled in the FSTD.

6. **Sound cue:** Defines the type of sound cues required to be modeled. Such sound cues are those related to sounds generated externally to the cockpit environment such as sound of aerodynamics, propulsion, runway rumble, weather effects, etc. and those internal to the cockpit.

7. **Visual cue:** Defines the type of out-of-cockpit window image display (e.g. collimated or non-collimated) and field of view (horizontal and vertical) that is required to be seen by the pilots using the FSTD from their reference eyepoint. Technical requirements such as contrast ratio and light point details are also described. HUD and EFVS options are also addressed.

8. **Motion cue:** Defines the type of motion cueing required that may be generated by the aircraft dynamics and from other such effects as airframe buffet, control surface buffet, weather, ground operations, etc.

9. **Environment ATC:** Defines the level of complexity of the simulated Air Traffic Control environment and how it interacts with the flight crew under training in the FSTD. The focus of this feature is on the terminal area manoeuvring, not on the in-flight cruise phase of flight.

10. Environment Navigation: Defines the level of complexity of the simulated navigation aids, systems and networks with which the flight crew members are required to operate, such as GPS, VOR, DME, ILS, NDB, etc.

11. Environment Weather: Defines the level of complexity of the simulated ambient and weather conditions, from temperature and pressure to full thunderstorm modeling, etc.

12. Environment Airports & terrain: Defines the complexity and level of detail of the simulated airport and terrain modeling required. This includes such items as generic versus customized airports, visual scene requirements, terrain elevation, EGPWS databases, etc.

13. Miscellaneous Defines criteria for the following FSTD miscellaneous feature technical requirements:

- instructor station;
- self-diagnostic testing;
- computer capacity;
- automatic testing;
- updates to hardware and software;
- daily pre-flight; and
- system integration (transport delay).

Appendix C

Participant Information Sheet

Loughborough University

The Effects of Amplified Head Rotations on Monoscopic Flight Simulation Training Devices

Participant Information Sheet

Professor Roy S. Kalawsky,	r.s.kalawsky@lboro.ac.uk	01509 635678
Luan Le Ngoc,	l.le-ngoc@lboro.ac.uk	01509 635674

Loughborough University
Leicestershire, UK
LE11 3TU

What is the purpose of the study?

The objective of these experiments is to comprehend the value of augmenting low fidelity simulators with amplified head rotations for expanding their flight training purposes.

Who is doing this research and why?

This study is part of a PhD research project supported by Loughborough University. It aims to explore the effects of augmenting low visual fidelity flight simulators via amplified head rotations compared to higher fidelity setups.

Are there any exclusion criteria?

Previous flight experience of participants is desired and participants will be grouped accordingly (novice, recreational, student pilot, commercial).

Once I take part, can I change my mind?

Yes, after you have read this information sheet and asked any questions you may have, you will be asked to complete the Informed Consent Form. However, at any time prior, during or after the experiments should you wish to withdraw from the study, please indicate so to the principal investigator. Withdrawal from participation is not subject to any cause and you will not be asked nor have the obligation to explain your cause for withdrawal.

How long will it take?

The experiment consists of two parts. The first part takes around 60 minutes to complete and the second part 100 minutes. This includes time for filling in the questions and questionnaires scheduled in the experiments.

What will I be asked to do?

Your primary task is to take control of an aircraft during the final portion of a visual traffic circuit pattern and perform a full stop landing on the runway. The secondary, concurrent task is to lookout for other traffic and press a button on the flight stick when you have spotted a popup object as briefed. After completing each simulator configuration you will need to fill in a workload and sickness assessment. At the end of the experiment, there is a general questionnaire.

Are there any risks in participating?

None.

Will my taking part in this study be kept confidential?

The information you provide will be held and used in accordance with the Data Protection Act 1998. Any information collected about you during the research will be kept strictly confidential and securely stored at Loughborough University. You will be identified by an ID number and information pertaining your name and address removed so that you are not linked to it. All data recordings will be destroyed six years after the completion of this research.

What do I get for participating?

Voluntary entry into a prize draw upon completing the experiments and free simulator flight time after the experiment.

I have some more questions who should I contact?

Luan Le Ngoc, l.le-ngoc@lboro.ac.uk 01509 635674 / 0785 130 6390

What if I am not happy with how the research was conducted?

The University has a policy relating to Research Misconduct and Whistle Blowing which is available online at http://www.lboro.ac.uk/admin/committees/ethical/Whistleblowing(2).htm

Appendix D

Consent Form

Loughborough University

The Effects of Amplified Head Rotations on Monoscopic Flight Simulation Training Devices

INFORMED CONSENT FORM
(to be completed after Participant Information Sheet has been read)

The purpose and details of this study have been explained to me. I understand that this study is designed to further scientific knowledge and that all procedures have been approved by the Loughborough University Ethical Advisory Committee.

I have read and understood the information sheet and this consent form.

I have had an opportunity to ask questions about my participation.

I understand that I am under no obligation to take part in the study.

I understand that I have the right to withdraw from this study at any stage for any reason, and that I will not be required to explain my reasons for withdrawing.

I understand that all the information I provide will be treated in strict confidence and will be kept anonymous and confidential to the researchers unless (under the statutory obligations of the agencies which the researchers are working with), it is judged that confidentiality will have to be breached for the safety of the participant or others.

I agree to participate in this study.

Your name ______________________________

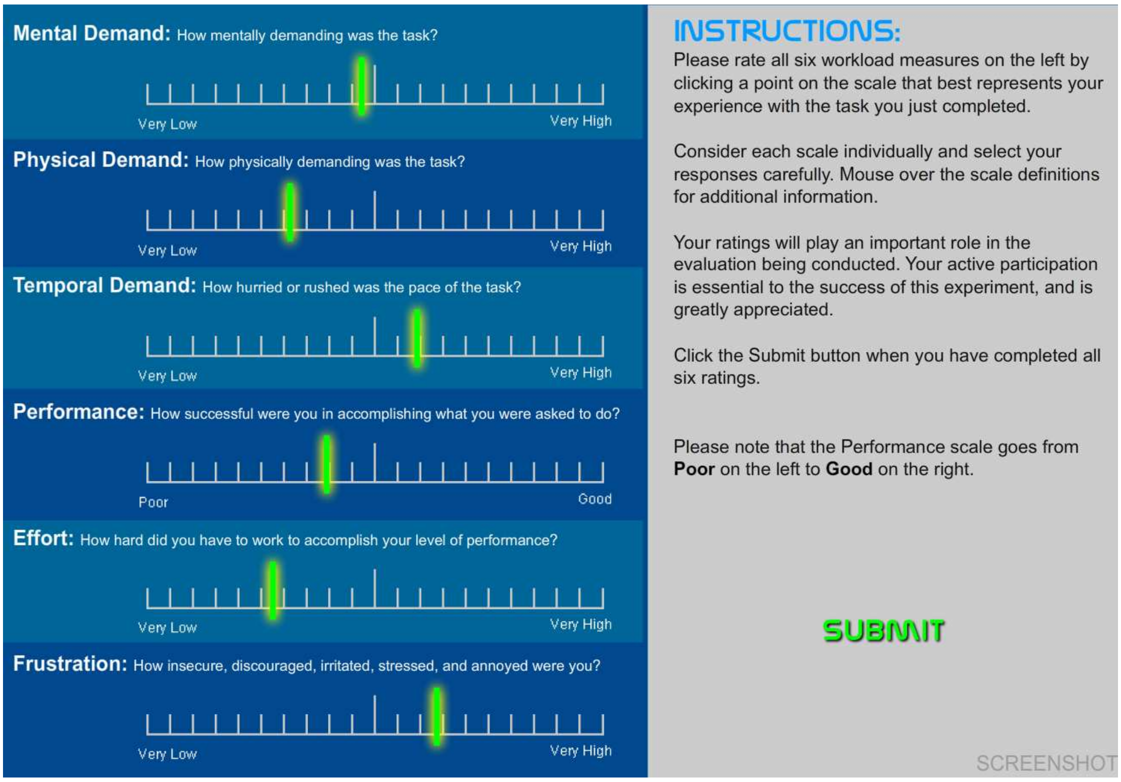
Your signature ______________________________

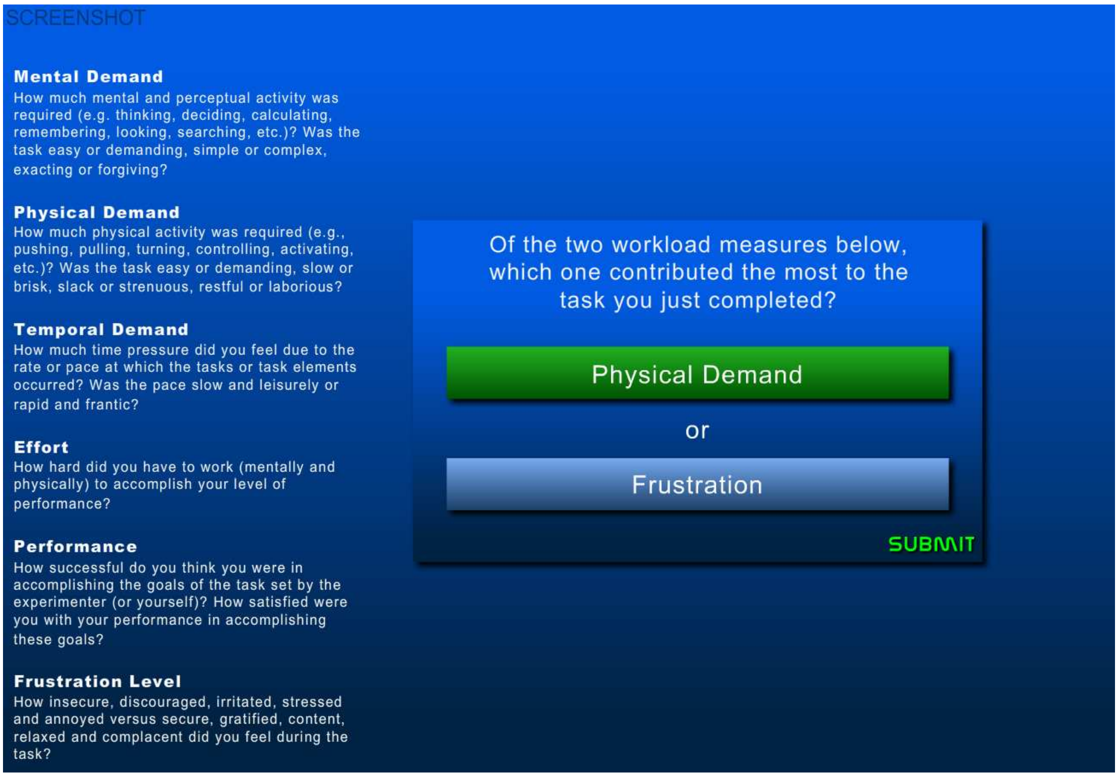
Signature of investigator ______________________________

Date ______________________________

Appendix E

Digital NASA-TLX Rating Sheet




SCREENSHOT
Mental Demand
How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand
How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand
How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Effort
How hard did you have to work (mentally and physically) to accomplish your level of performance?
Performance
How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Frustration Level
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?
Of the two workload measures below, which one contributed the most to the task you just completed?
Physical Demand
or
Frustration
SUBMIT

Appendix F

SSQ Form

AVRRC - SIMULATOR SICKNESS QUESTIONNAIRE - 2012

Please fill this form after completing each simulator display configuration.
* Required

Participant ID *
enter your first name (confirm with experimenter in case of duplication)

DISPLAY CONFIGURATION *
select display configuration you have just flown

- ○ Single monitor
- ○ Triple monitor
- ○ Triple projector

SIMULATOR SICKNESS QUESTIONNAIRE *
please match symptom with applicable severity

	None	Light	Moderate	Severe
General discomfort	○	○	○	○
Fatigue	○	○	○	○
Headache	○	○	○	○
Eye strain	○	○	○	○
Difficulty focusing	○	○	○	○
Salivation increasing	○	○	○	○
Sweating	○	○	○	○
Nausea	○	○	○	○
Difficulty concentrating	○	○	○	○
Fullness of the head	○	○	○	○
Blurred vision	○	○	○	○
Dizziness with eyes open	○	○	○	○
Dizziness with eyes closed	○	○	○	○
Vertigo	○	○	○	○
Stomach awareness	○	○	○	○
Burping	○	○	○	○

Appendix G

Post-Experiment Questionnaire

AVRRC - GENERAL QUESTIONNAIRE - 2012

post-experiment questionnaire
* Required

A. Pilot Characteristics

Name *
First name + Surname

Date of Birth *
DD-MM-YYYY

Gender *
Male

Vision *
- ○ Uncorrected
- ○ Corrected with glasses
- ○ Corrected with lenses

Hand *
- ○ Left
- ○ Ambidexter
- ○ Right

Current crew position *
select N.A. if non-airline crew
- ☑ N.A.
- ☐ Captain
- ☐ First Officer
- ☐ Second Officer
- ☐ Other:

Extra qualifications/experience *
tick any applicable
- ☐ Instructor
- ☐ Military

- ☐ Test pilot
- ☐ Flight engineer
- ☐ Other:

Flight experience *
total nr of flight hours

Flight log
list relevant a/c types & flight hours best of your knowledge

Do you have any prior experience with VR head-tracking? *

- ○ Yes
- ○ No

Do you have any experience with 3D first person gaming? *

- ○ None
- ○ Some
- ○ Alot

B. Debrief

Handling Qualities Rating Scale *
Please rate the aircraft dynamics (ease of flight control) via Cooper-Harper rating sheet

1	2	3	4	5	6	7	8	9	10
○	○	○	○	○	○	○	○	○	○

How would you rate the realism level of the outside visuals? *
very realistic

Comments on outside visuals

Please indicate your #1 simulator configuration preference *
most liked
Single monitor

Please indicate your #2 simulator configuration preference *
second most liked
Single monitor

Please indicate your #3 simulator configuration preference *
least preferrred
Single monitor

Do you have any remarks on the experimental setup?

Do you have any remarks regarding the experiment that this questionnaire has not addressed?

Submit
Never submit passwords through Google Forms.

Powered by Google Docs

Appendix H

Private Pilot Practical Test Standards

The Flight Standards Service of the Federal Aviation Administration (FAA) developed practical test standards as the standard requirements when conducting the practical test portion for the private pilots' exam [113]. These are also used as standards during flight training. The following adapted excerpts are the task and test standards which was used to form the basis of the checkride during the research experiments.

Straight-and-Level Flight

Objective: To determine that the applicant:

1. Exhibits satisfactory knowledge of the elements related to attitude instrument flying during straight-and-level flight.
2. Maintains straight-and-level flight solely by reference to instruments using proper instrument cross-check and interpretation, and coordinated control application.
3. Maintains altitude, ±100 feet; heading, ±20°; and airspeed, ±10 knots.

Turns to Headings

Objective: To determine that the applicant:

1. Exhibits satisfactory knowledge of the elements related to attitude instrument flying during turns to headings.
2. Transitions to the level-turn attitude using proper instrument cross-check and interpretation, and coordinated control application.
3. Demonstrates turns to headings solely by reference to instruments; maintains altitude, ±200 feet; maintains a standard rate turn and rolls out on the assigned heading, ±10°; maintains airspeed, ±10 knots.

Performance Manoeuvre: Steep Turns

Objective: To determine that the applicant:

1. Exhibits satisfactory knowledge of the elements related to steep turns.
2. Establishes the manufacturers recommended airspeed or if one is not stated, a safe airspeed not to exceed V_A.
3. Rolls into a coordinated 360° turn; maintains a 45° bank.
4. Performs the task in the opposite direction, as specified by the examiner.
5. Divides attention between airplane control and orientation.
6. Maintains the entry altitude, ±100 feet, airspeed, ±10 knots, bank, ±5°; and rolls out on the entry heading, ±10°.

Ground Reference Manoeuvres: Rectangular Course

Objective: To determine that the applicant:

1. Exhibits satisfactory knowledge of the elements related to a rectangular course.
2. Selects a suitable reference area.
3. Plans the manoeuvre so as to enter a left or right pattern, 600 to 1,000 feet AGL at an appropriate distance from the selected reference area, 45° to the downwind leg.
4. Applies adequate wind-drift correction during straight-and-turning flight to maintain a constant ground track around the rectangular reference area.
5. Divides attention between airplane control and the ground track while maintaining coordinated flight.
6. Maintains altitude, ±100 feet, maintains airspeed, ±10 knots.

Appendix I

Technology Readiness Levels

Originally developed by NASA [212], Technology Readiness Levels (TRL) as shown in Figure I.1 are a technology management tool to assess the maturity of particular technology during development and provides a scale to compare technology. It has then since been adopted in many domains [213]. In defence acquisition and research, the use of TRL is similar to the NASA concept but with only slight differences in the higher levels for accepted system maturity into responsible operational use [214]. Table I.1 shows both the Department of Defense (DoD) and the Ministry of Defence (MoD) viewpoint of TRL in their procurement process with descriptions.

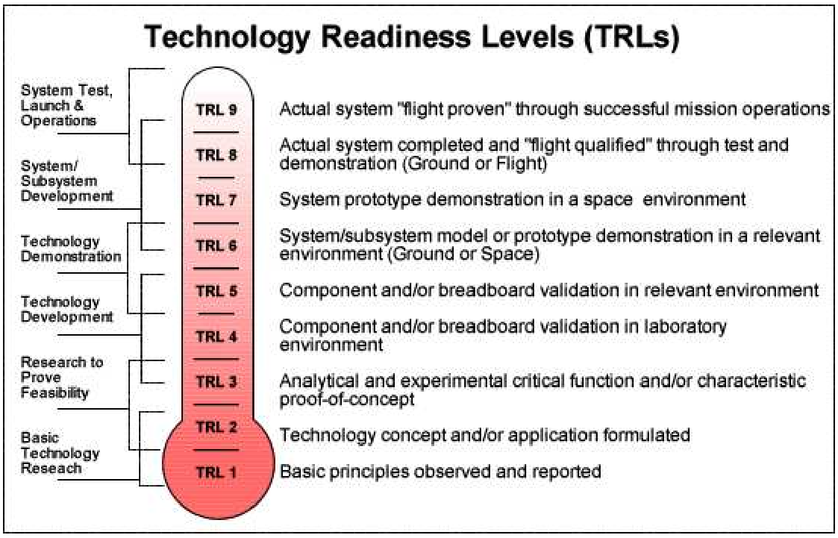


FIGURE I.1: NASA Technology Readiness Levels

Table I.1: Technology Readiness Levels from MoD/DoD [35, 160]

Technology Readiness Level	Description
01. Basic principles observed and reported.	Lowest level of technology readiness. Scientific research begins to be translated into technologys basic properties.
02. Technology concept and/or application formulated.	Invention begins. Once basic principles are observed, practical applications can be invented. The application is speculative and there is no proof or detailed analysis to support the assumption. Examples are still limited to paper studies.
03. Analytical and experimental critical function and/or characteristic proof of concept.	Active R&D is initiated. This includes analytical studies and laboratory studies to physically validate analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative.
04. Component and/or breadboard validation in laboratory environment.	Basic technological components are integrated to establish that the pieces will work together. This is relatively low fidelity compared to the eventual system. Examples include integration of ad hoc hardware in a laboratory.
05. Component and/or breadboard validation in relevant environment.	Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so that the technology can be tested in simulated environment. Examples include high fidelity laboratory integration of components.
06. System/subsystem model or prototype demonstration in a relevant environment.	Representative model or prototype system, which is well beyond the breadboard tested for level 5, is tested in a relevant environment. Represents a major step up in a technologys demonstrated readiness. Examples include testing a prototype in a high fidelity laboratory environment or in a simulated operational environment.
07. System prototype demonstration in an operational environment.	Prototype near or at planned operational system. Represents a major step up from level 6, requiring the demonstration of an actual system prototype in an operational environment. Examples include testing the prototype in a test bed aircraft.
08. Actual system completed and qualified through test and demonstration.	Technology has been proven to work in its final form and under expected conditions. In almost all cases, this level represents the end of true system development. Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications.
09. Actual system proven through successful operations.	Actual application of the technology in its final form and under mission mission conditions, such as those encountered in operational test and evaluation. Examples include using the system under operational mission conditions.

Appendix J

Other Simulator Supported Research Projects

The design and construction of the research flight simulator described in Chapter 5 was not only in support of the work done in this thesis, but also formed a collaborative effort with two other research projects backed by industry cases. Although there were overlaps in functional requirements, there were still uncertainties in exacting requirements posed by these other projects. Despite these challenges, the final simulator design and build was successful in supporting all three projects and their experiments. The following short summaries of these two other projects provide insight into the variety of research the simulator was required to support influencing its flexibility in operational use.

Operator Training & Performance Measurement in Remotely Piloted Aerial Systems

With the development and use of military and civilian unmanned systems, there is a rapidly growing need for training programs and performance measurement in support. Research has been carried out into the viability of a competency based training system for use with unmanned aerial systems and semi-automated decision process based, performance measurement [215, 216].

Aims & Objectives:

- Create a time and competency based structure related to operator task processes.
- Develop a performance measurement system relying on operator decision processes rather than actions.
- Create a semi-automated data collection and analysis system.
- Form the basis for future operator training, licensing and performance measurement.

Expected Impact:
This research could influence the global development of both commercial and military remotely piloted aerial systems operator training and licensing systems.

Sponsor:
EPSRC/BAE Systems

Simulating overload of aircrew attention for optimal training benefit

With simulation becoming a key factor in training of not only aircrew but expanding into other domains, there is a need to get the most gain out these resources when they are used. This project therefore examines the relationship between workload and performance in regards to the optimal training benefit in a simulated environment [161].

Aims & Objectives:

- Study if trainees in a flight simulator can have their workload levels repeatedly and consistently loaded up to a high level without overloading them.
- Determine how a saturated workload loading process compares to a non-loading process and whether this can lead to greater gains for training.
- Create a procedure to load up a subject to high workload levels without over saturation.

Expected Impact:
Guide future simulator training by enabling increased productivity from the same resources, exploiting small changes in preparation for a simulated event to yield greater performance gains.

Sponsor:
EPSRC/BAE Systems

Appendix K

Market Requirements Analysis

It is fundamental to understand the marketing potential for any research carried out using the constructed flight simulator. This appendix provides the results of the brief market requirements analysis estimate discussed in Section 5.2.1 built on Bedford's initial analysis [216]. The market is split into three main categories: Academia, Industry and Government. These three groups represent the potential areas for marketable products produced by research conducted upon the simulator. Four subcategories of proposed interest to these groups were inferred: fidelity, adaptability, interoperability and cost.

Figure K.1 is a visual representation of the market analysis breakdown. Each area's importance for each subcategory has been graded with symbols ranging from – to ++ with – representing a low level of importance and ++ representing a high level of importance. The results (at the time) imply that the ideal simulator system would mimic the requirements of the prominent industry sponsor (supporting the two projects of Appendix J with the minimum requirements obviously matching the Loughborough University category.

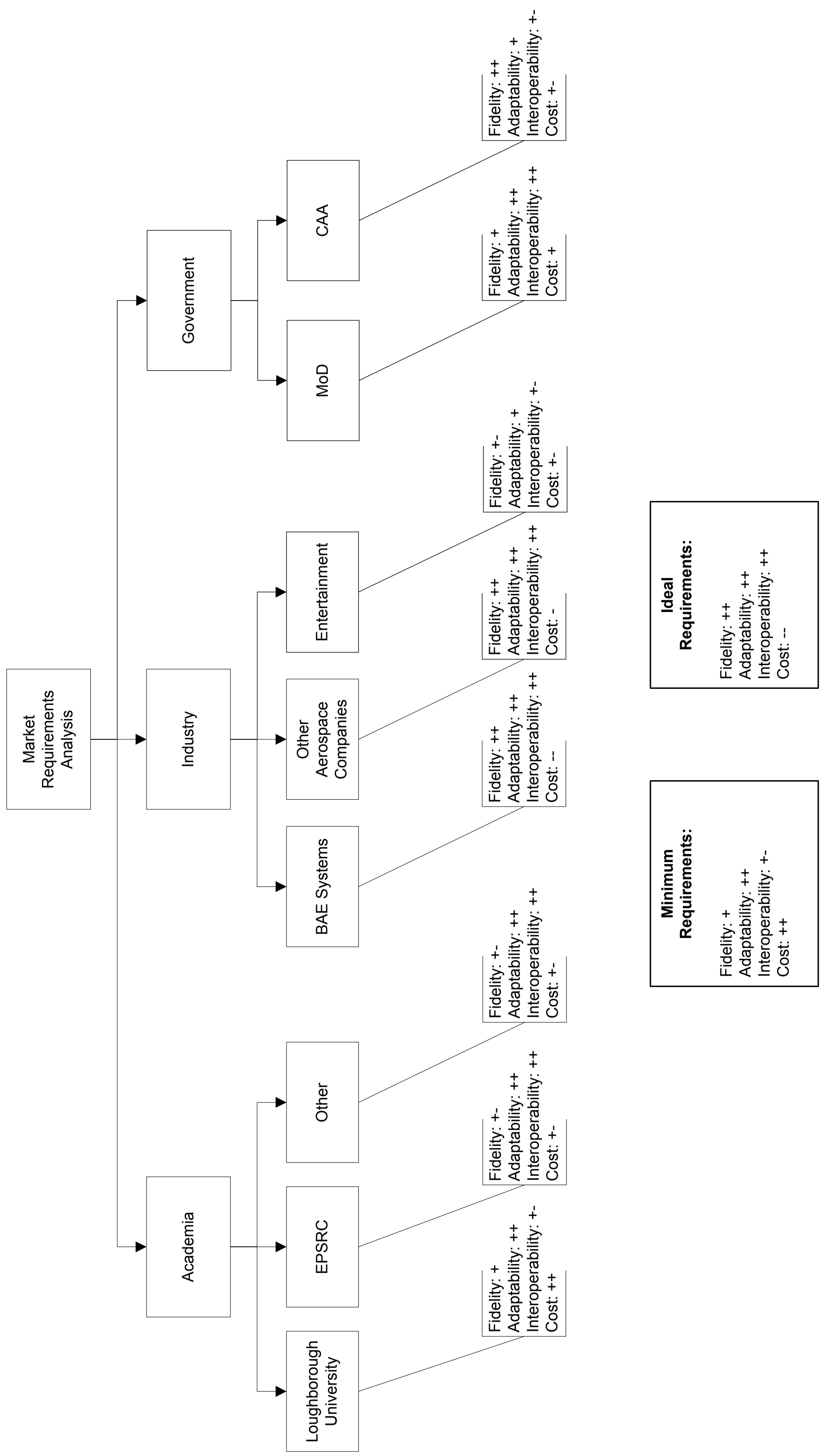


Figure K.1: Market Requirements Analysis

Academia

This category represents the academic market. Marketing focus will not center round the ethos of a marketable product but in developments made within the research field, the potential of the system to aid future research and to increase the standing and capabilities of the relevant departments in both the commercial and academic fields.

Loughborough University

The proposed system will be used with three doctoral projects (this thesis being departmental independently funded together with the two BAE CASE studentships in Appendix J). It is in the universities best interest for these projects to be successful in terms of institutional academic standing as well as standing within the industrial market as a leading developer of industrial solutions.

Fidelity With research focusing on pilot/operator use of simulators it is important that a reasonable level of fidelity is reached in terms of operation and visual displays. It is not necessary however, or in some cases possible, to reach industrial/military levels of fidelity as discussed in Section 3.2.

Adaptability Due to the system being required to perform well over multiple research projects, with differing needs and tasks, it is highly desirable that the system is adaptable to any process required from it. This also applies to potential future research needs that are not currently defined.

Interoperability The system itself is likely to be comprised entirely of COTS (Commercial-Off-The-Shelf) software and hardware to reduce the level of expenditure required; this leads to the standard of interoperability between the proposed system and other industry/government systems to be lower than would have been desired. This stemmed from the Project Objective Statement in Section 5.2.

Cost Due to the system being academically funded the cost of the system is highly important in determining the levels of the three previous factors. To reduce costs the system will be comprised of COTS hardware and software, again an objective stated at the start in Section 5.2.

Engineering and Physical Sciences Research Council (ESPRC)

The EPSRC provide funding set amount of funding for the three projects involved with the proposed system; they are not necessarily looking for a marketable product created in conjunction with the system. They will, however, wish for the projects to be successful.

Fidelity Not essential to the EPSRC as their fundamental requirement is for the research to be completed successfully; it is likely that all that is required is for the proposed system to meet the minimum requirements of the researchers.

Adaptability It is likely that the proposed system will be used in conjunction with future ESPRC funded projects so a high level of adaptability is likely to be required.

Interoperability Due to the nature of the ESPRC being a nationwide funding organization a high level of interoperability is likely to be desired; this will allow further co-operation with other academic institutions as well as with industrial partners.

Cost The ESPRC do not directly contribute to the funding of the simulator and any money used would be part of a set budget already allocated for research in the whole. Cost, therefore, is not applicable in this case.

Industry

The following sub-categories represent the industrial aspect of the proposed system; these include not only CASE student sponsors (in the form of BAE Systems) but also potential investors, employers and interested parties.

BAE Systems

With two research projects (Appendix J) also making use of the constructed simulator, they were a major factor influencing the design and function. It definitely influenced some of the conceptual designs during the design synthesis in terms of hardware and software to match the industry partner's own simulation architecture for easy transfer of research.

Fidelity To meet BAE systems standards it is highly desirable for the systems fidelity to be extremely high, this is to try and match the current levels of fidelity currently expected of BAE systems simulators. The

BAE simulators are used primarily for pilot training and are, therefore, at the highest end of the fidelity scale. This level of fidelity is, on budget, unrealistic for the proposed system but any effort it harmonization makes it desirable for research transfer and securing future projects.

Adaptability The proposed system is to be used with three differing research projects, this requires a high level of adaptability. The system is also highly likely to be used with future industry collaborations so, again, a high level of adaptability to future needs is very desirable.

Interoperability The research currently planned in conjunction with the system is industry funded; this means that BAE funded research is likely to be used by BAE upon completion. Adaptability and applicability of the research is therefore highly likely to be based on the degree to which the research is interoperable with current BAE systems; it comes down to the research platform (i.e. the proposed system) also being highly interoperable with BAE's current systems where possible.

Cost As the funding for the proposed system is not being sourced from BAE systems the cost of the system is not of much relevance to the company. Due to the system also being COTS orientated and being of comparatively low cost compared to a BAE simulator again cost of replicating the simulator at BAE is also not of great relevance.

Other Aerospace Companies There is potential for other (aerospace companies) to be interested in the research being performed on the proposed system. At completion of this thesis, another aerospace industry partner has indeed been using the simulator for a fourth research project.

Fidelity Similarly to BAE systems, other competing companies use extremely high fidelity systems; the proposed system must try and reflect this high level of fidelity, both in terms of software and hardware, where possible.

Adaptability There is potential for the proposed system to be used in conjunction with other aerospace companies in the future and this would require a high level of adaptability of the system to ensure that future research requirements are met.

Interoperability Again, similarly to BAE systems, the generated research is highly desired to be interoperable with a company's current systems to allow for a high level of research applicability.

Cost At this point in time there is no funding being received from any other external aerospace company that can be applied to the simulator. If the company wish to replicate the simulator then the comparative cost of the proposed COTS system is likely to be much lower than current systems in operation with the company.

Entertainment Though small there is marketing potential within the entertainment sector; there is a large community of recreational simulation enthusiasts who would have interest in the proposed system as well as the companies who currently supply this community with hardware and software.

Fidelity Due to the entertainment industry having rather mixed requirements as to quality of interface the fidelity is not of massive importance. Many of the consumers cannot afford extremely high end hardware or software and the industrial sector do not produce software up to the standard of the aerospace industry.

Adaptability The is not a massive requirement for the proposed system to perform well in multiple disciplines but it is likely that the system could be used for varying tasks so there is a need for a reasonable level of adaptability but it is not crucial

Interoperability As with adaptability it is unlikely that there will be a great need for interoperability within the entertainment sector as commercial systems will be built for a specific task rather than there being a need for the system to be operable for multiple environments.

Cost Though cost is likely to be a large factor in the commercial industry the range of needs, based mostly around fidelity, will vary hugely; the system will be built purely on the available budget whether that budget is large or small.

Government

Much of the intended research will be primarily used within the industrial sector but there may be repercussions directly associated with the research as well as with anything developed using the research as a basis within the government and military sectors.

UK MoD (Ministry of Defence) Much of the proposed research could have an effect on the UK MoD; the two BAE projects are both based in the military sector so it is likely that any useful developments that originate from the research will be used within the MoD.

Fidelity The level of fidelity required by the MoD is likely not to be of the same standard as that required by BAE. Whereas, BAE will not only use the product for function but also as a demonstration piece to raise the company profile, the MoD is likely to require a level of fidelity at which the task (training etc) can be completed without the need to raise their profile

Adaptability It is highly likely that the MoD will wish for a system that can perform well in multiple disciplines and for multiple platforms, this would lower expenditure and raise capability in the long run by not having to rely on many spate, specialist systems. It would also be desirable for the system to be easily upgraded and compatible with new systems.

Interoperability A high level of interoperability would lower training and development costs which, in the current economic climate, would be highly desirable to the MoD.

Cost As previously mentioned the cost of the system is likely to be a factor in the MoD's decision whether to implement the proposed system and the associated research developed in conjunction with the system. Being the likely end consumers for the previously mentioned research and system and with competition from other industrial companies making developments and selling products in a similar area to the proposed system a lower cost option would likely prove to be more attractive than a higher cost, more specialized system.

CAA (Civil Aviation Authority) The some of the intended research is also likely to have implications for the Civil Aviation Authority with the UAV training research, as an example, having the limited potential to help provide a basis for a generic form of UAV operator licensing and a set standard for training.

Fidelity The CAA will require a high level of fidelity to allow the highest level of realism possible to be used for licensing and training purposes; this would allow for higher training standards.

Adaptability A high level of adaptability would allow for multiple training programs to be run on a single platform. A high level of adaptability will also allow for software improvements as well as training development with new and future systems.

Interoperability Given that the CAA cover multiple platforms and levels of licensing which is globally applicable, in some cases, any potential system needs to be interoperable with differing equipment and differing country standards.

Cost It is likely that the CAA will be indifferent about the proposed system costs as much of the training for licensing is outsourced to flight schools with the final examinations then being overseen by the CAA.

Appendix L

Simulator Features and Statement of Compliance

This appendix describes the minimum feature fidelity level requirements derived from [36] via the fidelity selection done in Section 5.2.2 in accordance with the process described in Section 3.2.2. The statement of compliance by the final design is highlighted in bold after each technical requirement shows how the requirement was met or not and if so the reasons behind it.

1. Cockpit Layout & Structure

1.R An enclosed or perceived to be enclosed cockpit/flight deck, excluding distraction, which will represent that of the aeroplane derived from, and appropriate to class, to support the approved use.
The pilot station shown in Figure 4.6 has a replica fighter-type cockpit shell enclosure. Simulator operations in both the cockpit station and the immersive projector lab shown in Section ?? are always performed in a darkened environment to prevent distractions.

1.1.R FSTD instruments and/or instrument panels using electronically displayed images with physical overlay or masking and operable controls representative of those in the aeroplane are acceptable. The instruments displayed should be free of quantization (stepping). With the requirement for only a spatially representative cockpit/flight deck, the physical dimensions of the enclosure may be acceptable to simulate more than one aeroplane or class of aeroplane in a convertible FSTD. Each

FSTD conversion should be representative of the aeroplane or class of aeroplane being simulated which may require some controls, instruments, panels, masking, etc to be changed for some conversions. If the FSTD is used for VFR training, it should be a representation of the aeroplane or class of aeroplane comparable to the actual aeroplane used for flight training.
Not applicable to the letter, because being a research flight simulator catering to a wide variety of projects and aircraft, it is not vital represent even an actual aircraft class compared to flight training use.

1.2.1.R Flight crew member seats should represent those in the aeroplane being simulated.
Like 1.1.R, the seat in the cockpit facility is actually a replica of a real ejector seat, whereas the projector station uses an office chair with comparable comfort to a padded aircraft seat.

1.2.2.R In addition to the flight crew member seats, there should be an instructor station seat and two suitable seats for an observer and authority inspector.
Figure 4.6 shows the instructor control station adjacent with ample room to wheel in additional office chairs for extra observers.

1.3.R Lighting environment for panels and instruments should be sufficient for the operation being conducted.
Darkened environment, all panels/instruments are virtually simulated on the displays so lighting is not a problem.

2. Flight Model

2.R Aerodynamic and engine modeling, aeroplane like, derived from and appropriate to class to support the approved use. Flight dynamics model that accounts for various combinations of drag and thrust normally encountered in flight corresponding to actual flight conditions, including the effect of change in aeroplane attitude, sideslip, thrust, drag, altitude, temperature.
Flight simulation software suite X-Plane 9.70 uses blade element theory and provides a 6-DOF flight model [217].

3. Ground Handling

3.R Represents ground reaction and handling aeroplane like, derived from and appropriate to class.
Ground handling modeled by flight simulation software [133].

4. Aeroplane Systems

4.R Aeroplane systems should be replicated with sufficient functionality for flight crew operation to support the approved use. System functionality should enable sufficient normal and appropriate abnormal and emergency operating procedures to be accomplished.
Since the aircraft was the flagship showpiece of the simulator software, its modeling and simulation of aircraft systems was done in cooperation with the aircraft manufacturer to represent the systems at the time so this covered the systems listed by 4.1-4.5.R [36].

5. Flight Controls

5.R Aeroplane like, derived from class, appropriate to aeroplane mass to support the approved use.
5.1.R Control forces, control travel and surface position should correspond to that of the aeroplane or class of aeroplane being simulated. Control travel, forces and surfaces should react in the same manner as in the aeroplane or class of aeroplane under the same flight and system conditions. Active Force feedback required if appropriate to the aeroplane installation.
Force feedback was not an option due to limited resources. The flight control stick and throttle used is an actual replica of a fighter flight control system and the simulator has the option to mount a force pressure side stick for aircraft that use this.

5.3.R,R1 Control systems should replicate the class of aeroplane operation for the normal and any non-normal modes including back-up systems and should reflect failures of associated systems. Appropriate cockpit indications and messages should be replicated.
Fault warning panels are provided both in the virtual cockpit and when

the instrument display is used available through the electronic display system [132].

6. Sound Cues

6.G Significant sounds perceptible to the flight crew during flight operations to support the approved use. Comparable engine and airframe sounds.
6.1.G Significant cockpit/flight deck sounds during normal and abnormal operations, aeroplane class-like, including engine and airframe sounds as well as those which result from pilot or instructor-induced actions.
6.2.G The sound of a crash when the simulated aeroplane exceeds limitations.
6.3.G Environmental sounds are not required. However, if present, they should be coordinated with the simulated weather.
6.4.G The volume control should have an indication of sound level setting which meets all qualification requirements.
6.5.G Sound not required to be directional.
All of above required features are available through X-Plane 9.70 [133].

7. Visual Display Cue

7.R Continuous field of view textured representation of all ambient conditions for each pilot, to support the approved use. Horizontal and vertical field of view to support the most demanding manoeuvres requiring a continuous view of the runway.
Taken under advice, since the purpose of this thesis is to study how to overcome field of view limitations!

7.1.R Continuous visual field of view providing each pilot with 200 degrees horizontal and 40 degrees vertical field of view.
Disregarded, since in conflict with thesis goal of investigating reduced field of views.

7.2.R Surface resolution demonstrated by a test pattern of objects shown to occupy a visual angle of not greater than 4 arc minutes in the visual display used on a scene from the pilots eye point.
7.3.R Light-point size not greater than 8 arc minutes.

7.4.R Surface Contrast ratio not less than 5:1

Unable to demonstrate compliance due to unavailability of test resources.

7.5.R Light-point contrast ratio not less than 10:1

Using a colorimeter [218] to measure, the computer displays for the visuals had a calibrated contrast ratio of 2400:1 whereas the projectors had a contrast ratio of 300:1.

7.6.R Light-point brightness - not less than $20cd/m^2$

Using the colorimeter [218] and profiling software [219], the displays were calibrated at a white point brightness of $120cd/m^2$ whereas the projectors achieved $30cd/m^2$. This ensured the overall light emanating from the total screen area was comparable between both stations.

7.7.R Surface brightness should be demonstrated using a raster drawn test pattern. The surface brightness should not be less than $14cd/m^2$.

Unable to demonstrate compliance due to unavailability of test resources.

7.13.R A test is required to demonstrate that the visibility is correct on final approach in CAT II conditions and the positioning of the aeroplane is correct relative to the runway.

Not applicable to experimental flying task of Chapter 4.

10. Motion Cue

None

9. Environment - ATC

None

10. Environment - Navigation

None, however simulation software has built-in capability to support up to Specific fidelity level as follows:
10.S Navigational data with the corresponding approach facilities to support the approved use. Navigation aids should be usable within range or line-of-sight without restriction, as applicable to the geographic area. 10.1.S Navigation Data Base sufficient to support simulated aeroplane systems for real world operations. 10.2.S Complete navigation data base for at least 3 airports with corresponding precision and non-precision approach procedures including regular updates. 10.3.S Instructor controls of internal and external navigational aids 10.4.S Navigational data with all the corresponding standard arrival and departure procedures. 10.5.S Navigation aids should be usable within range or line-of-sight without restriction, as applicable to the geographic area.

11. Environment - Weather

11.G Basic atmospheric model, pressure, temperature, visibility, cloud base and winds to support the approved use. The environment should be synchronised with appropriate aeroplane and simulation features to provide integrity.
11.1.G Simulation of the standard atmosphere including instructor control over key parameters.
11.2.G The FSTD should employ windshear models that provide training for recognition of windshear phenomena
11.3.G Visibility effects as observed on the visual system should be simulated and respective instructor controls provided.
11.4.G The following features should be simulated with appropriate instructor controls provided: 1) surface wind speed direction and gusts, 2) intermediate and high altitude wind speed and direction, 3)thunderstorms and microbursts, 4) turbulence.
X-Plane [133] provides all of above features, especially since it was designed for desktop usage.

12. Environment - Airports & Terrain

12.G Generic airport model/s with topographical features to support the approved use.
12.1.1.G Visual cues to assess sink rate and depth perception during takeoff and landing should be provided. This should include: 1) Surface on runways, taxiways, and ramps. 2) Terrain features.
12.3.1.G The FSTD should provide for accurate portrayal of the visual environment relating to the FSTD attitude.
12.4.1.G The system should include a generic airport available in daylight, twilight (dusk or dawn) and night illumination states.
12.4.2.1.G Daylight Capability
12.4.2.2.G The system should provide full-colour presentations and sufficient surfaces with appropriate textural cues to successfully accomplish a visual approach, landing and airport movement (taxi).
12.4.2.G Total scene content should be sufficient to identify the airport and represent the surrounding terrain.
12.7 *Visual System for reduced FOV*
12.7.1.G The system should provide a visual scene with sufficient scene content to allow a pilot to successfully accomplish a visual landing. Scenes should include a definable horizon and typical terrain characteristics such as fields, roads and bodies of water and surfaces illuminated by airplane landing lights.
X-Plane [133] provides all of above features, especially since it was designed for desktop usage.

13. Miscellaneous

13.1.R Instructor station should provide an adequate view of the pilots panels and forward windows.
The control station has replicate views of all the pilot station visuals and instruments in addition to monitoring displays (maps, systems) needed to supervise the experiment.

13.2.R Instructor controls for all required system variables, freezes, resets and for insertion of malfunctions to simulate abnormal or emergency conditions. The effects of these malfunctions should be sufficient to correctly exercise the relevant operating manuals.

The control station has full control of simulator running and stopping, failure of systems etc as built-in features by X-Plane [133].

13.8.R A transport delay test may be used to demonstrate that the FSTD system response does not exceed 200ms.

Due to limited time and resources, a true complete system latency test has not been performed. However, during the design of the simulator this was kept in mind and information on individual components was sought to minimize latency.

Appendix M

Functional Analysis Diagrams

Supporting the outputs of the requirements analysis in Section 5.2, high-level functions are visualized using the two diagram tools of a Functional Flow Diagram in Figure M.1 and the corresponding Functional Breakdown Structure of Figure M.2.

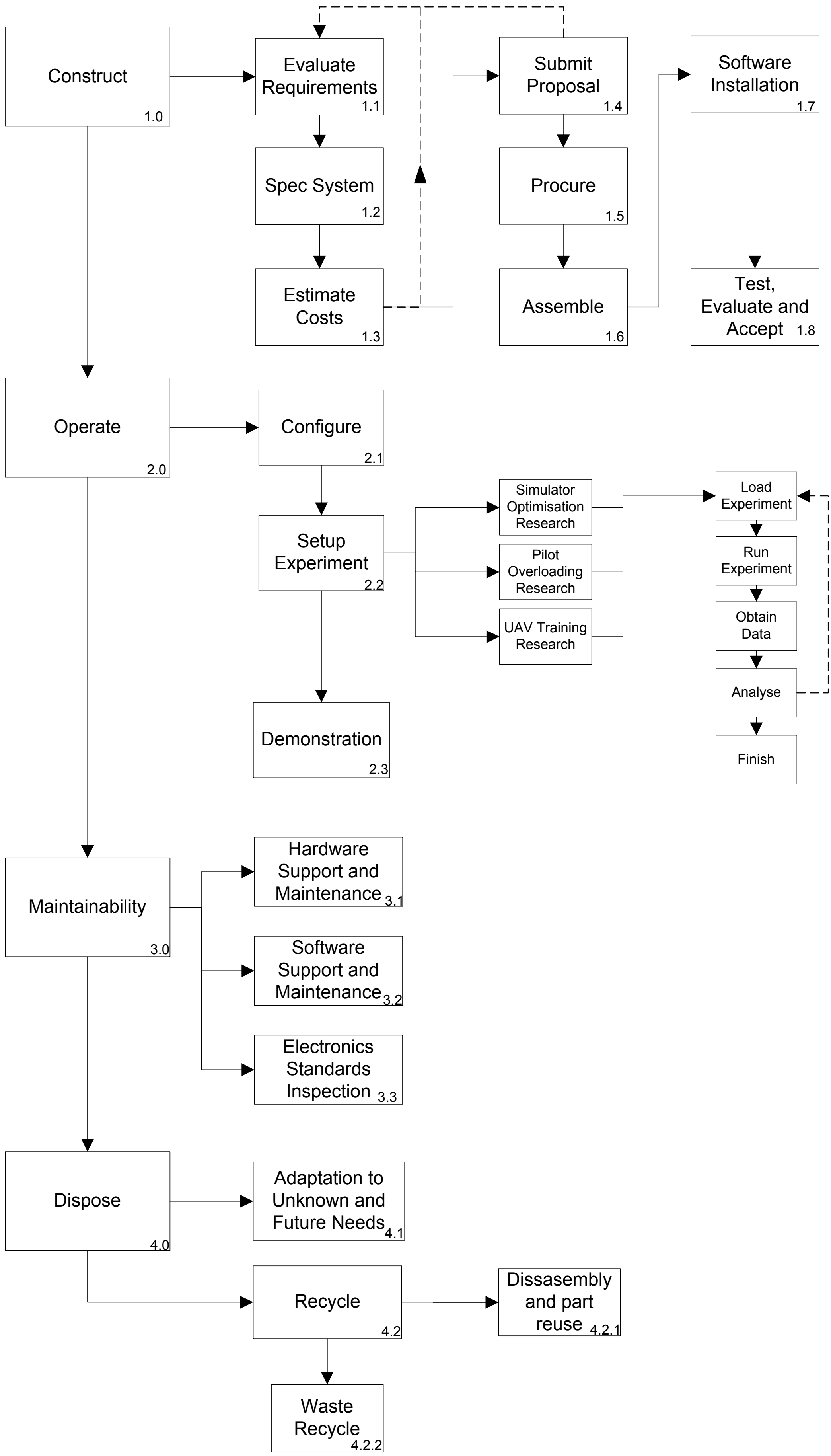


FIGURE M.1: Functional Flow Diagram

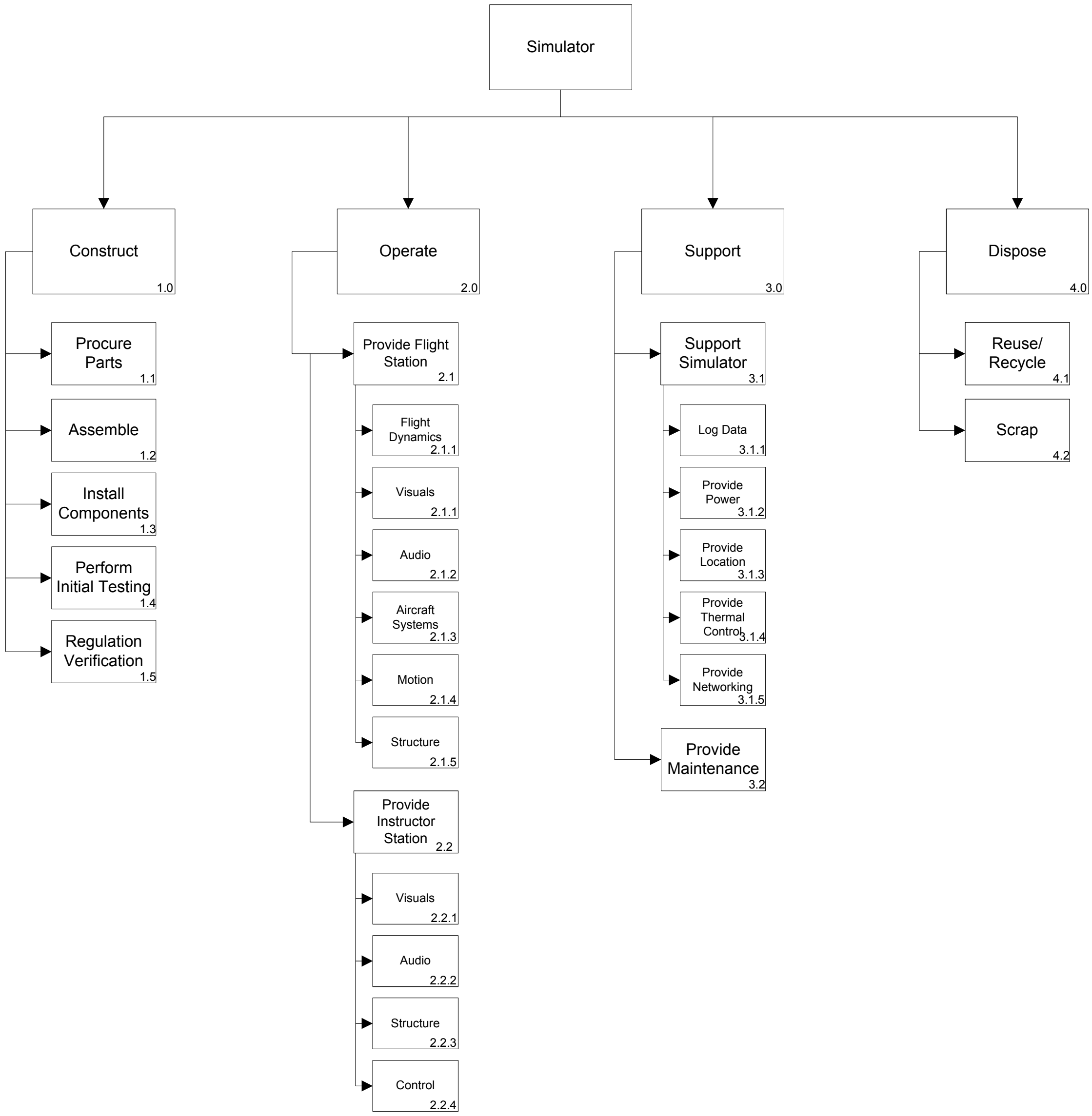


Figure M.2: Functional Breakdown Structure

Appendix N

Simulator Component Inventory

Table N.2 lists the components that made up each pilot station's PC system compiled with prices found on Amazon.com on 08-03-2013. As some of the components were discontinued by the latest price inventory, an equivalent alternative is provided in between brackets. Each pilot station is powered by one of these single PC systems, and although relatively powerful, these are all consumer grade systems, allowing easy access to components and future upgrades will remain low-cost. With a single machine this also reduces network complexity and latency issues to consider when splitting various simulation subsystems to be processed and synchronized by multiple computers. With identical twin stations hardware-wise, a solid state drive for fast disk read access times was featured and it also serves as a data cartridge system enabling hot swapping stations for rapid reconfiguration and development work.

Table N.1: Pilot station mounting options

Option	**Cost**
DIY/Generic Large Desk	$200
Commercial simulator frame	$600
Reused cockpit shell	$750

Table N.2: Pilot station PC components

Type	**Item**	**Cost**
CPU	Intel i7 2700K 3.5GHz @ 5GHz	$ 325
Cooler	Coolit Eco II FatBoy (or Corsair H100)	$ 110
Memory	Kingston 8GB DDR3 1600MHz	$ 51
Motherboard	Asus Maximus IV Extreme-Z	$ 360
GPU	nVidia GTX580 3GB (or GTX680 2GB) x2	$ 924
HDD	Seagate Barracuda 500GB SATA3	$ 60
SSD	Crucial M4 128GB	$ 116
PSU	Corsair 1050W	$ 200
Optical	DVD-RW	$ 15
Case	Coolermaster CM690 II	$ 92
Display	BenQ EW2730V (or BenQ 2750HM) x3	$ 720
Touchscreen	Iiyama T2250MTS (or ViewSonic TD2220)	$ 300
HOTAS	Thrustmaster HOTAS Warthog	$ 437
Rudders	Saitek Pro Flight Combat Rudder	$ 172
Headtracker	Naturalpoint TrackIR 5 Pro	$ 145
Webcam	Logitech Quickcam Sphere AF	$ 40
Total		**$4067**

Table N.3: Alternative Control Station components

Item	**Cost**
HP Z800 Workstation (Xeon E5530 2.4GHz, 3GB)	$ 800
AMD HD6850 GPU x2	$ 240
Matrox TripleHead2Go Digital Edition	$ 330
17" generic LCD monitors x7	$ 450
Total	**$1820**

Table N.4: Triple projector components

Item	**Cost**
Projection Design evo SX+ (or equivalent) x3	$2000
Pixelwix 18 ft diameter curved screen	$5000
Total	**$7000**

Table N.5: Consumer 3-channel projector warp/blend software

Product	**Cost**
Immersive Display Pro	$289
Warpalizer	$450
Nthusim Plus	$489

Appendix O

X-Plane Datarefs

The automatic, custom datalogging plugin made for X-Plane uses the following data reference variables from the simulation as shown in Table O.1. The dependent measures are derived from these raw simulation parameters. The NASA-TLX workload, Simulator Sickness Questionnaire score and secondary task measures are obtained separately outside the simulation engine.

TABLE O.1: X-Plane recorded datarefs

Dataref /sim/	**Name**	**Type**	**Units**	**Description**
flightmodel/	local_x	double	meters	Location in OGL coord.
position/	local_y	double	meters	Location in OGL coord.
	local_z	double	meters	Location in OGL coord.
	lat_ref	float	degrees	Latitude of OGL origin
	lon_ref	float	degrees	Longitude of OGL origin
	latitude	double	degrees	Latitude of aircraft
	longitude	double	degrees	Longitude of aircraft
	elevation	double	meters	Altitude above MSL of a/c
	theta	float	degrees	Aircraft pitch angle
	phi	float	degrees	Aircraft roll angle
	psi	float	degrees	True heading
	magpsi	float	degrees	Magnetic heading
	local_vx	float	m/s	Velocity in OGL coord.
	local_vy	float	m/s	Velocity in OGL coord.
	local_vz	float	m/s	Velocity in OGL coord.
	local_ax	float	m/s^2	Acceleration in OGL
	local_ay	float	m/s^2	Acceleration in OGL
	local_az	float	m/s^2	Acceleration in OGL
	alpha	float	degrees	Angle of attack
	beta	float	degrees	Sideslip angle
	vpath	float	degrees	Vert. flight path angle
	hpath	float	degrees	Horiz. flight path angle
	groundspeed	float	m/s	Aircraft ground speed
	indicated_airspeed	float	knots	Indicated Air Speed
	indicated_airspeed2	float	knots	Indicated Air Speed
	true_airspeed	float	m/s	True airspeed
	magnetic_variation	float	degrees	Local magnetic variation
	M	float	Nm	a/c angular momentum
	N	float	Nm	a/c angular momentum
	L	float	Nm	a/c angular momentum
	P	float	deg/s	Roll rate
	Q	float	deg/s	Pitch rate
	R	float	deg/s	Yaw rate
	P_dot	float	deg/s	Roll acceleration
	Q_dot	float	deg/s	Pitch acceleration
	R_dot	float	deg/s	Yaw acceleration
	Prad	float	rad/s	Roll rate
	Qrad	float	rad/s	Pitch rate
	Rrad	float	rad/s	Yaw rate
	q	float	quaternion	Rotation OGL to a/c coord.
	vh_ind	float	m/s	vertical speed
	y_agl	float	meters	Altitude Ground Level
joystick/yoke_	pitch_ratio	float	ratio	Pitch axis deflection
	roll_ratio	float	ratio	Roll axis deflection
	heading_ratio	float	ratio	Rudder axis deflection
engine/ENGN_	thro	float	ratio	Throttle setting
	N1	float	%	N1 speed of max
	N2	float	%	N2 speed of max
graphics/	head_psi	float	degrees	Pilot's head heading
view/	head_the	float	degrees	Pilot's head pitch
pilots_	head_x	float	meters	Pilot's head rel. c.g.
	head_y	float	meters	Pilot's head rel. c.g.
	head_z	float	meters	Pilot's head rel. c.g.
flightmodel2/	vertical_deflection_mtr	float	meters	Vert. gear deflection
gear/tire_	vertical_force_n_mtr	float	N	Vert. gear force