Flight Simulation Fidelity for Rotorcraft Design, Certification and Pilot Training

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy

by

Emma Timson

September 2013

ABSTRACT

The benefits of using flight simulators for rotorcraft design, certification and pilot training include reduced costs, increased safety, and control over external parameters such as environmental conditions and operational situations.

The progression of technology and computing power over recent decades has led to the ability to manufacture highly sophisticated flight simulators that can be used to train complex flight operations and accurately predict the behaviour aircraft. However, such sophistication comes at a cost and there is a need to understand the trade-offs between cost and effectiveness to allow the benefits of flight simulation to be transferred to lower cost applications such as initial skills acquisition training.

Assessment of simulator capabilities has traditionally been carried out with focus on the physical similarity of individual components of the simulator systems - motion system, visual system, flight model etc. However, this work is focused on the assessment of the fitness for purpose of the integrated system as a whole. This is referred throughout the thesis as perceptual fidelity, which has been defined as

'The simulator's ability to induce the behaviours known to be essential for operation of the actual aircraft in the performance of a specific task'.

The novel contribution of the work in this thesis is the development of new quantitative metrics and a subjective evaluation technique that could be utilised across the simulation industry for quantification of perceptual fidelity of the overall simulation. It is intended that the methods detailed in this work could be used to support simulator development and augment current assessment techniques where appropriate.

The quantitative measures of perceptual fidelity are based on comparison of ADS-33E PRF style performance metrics and the Attack metric, a control activity measure developed by Padfield et al. In this work, the utility of these metrics was assessed through correlation analyses with pilot subjective opinion. A lack of correlation in multi-axis tasks was seen and, as a result, novel metrics of pilot control strategy and adaptation have been developed in this work that show significant improvement in correlation with pilot subjective opinion.

The subjective assessment methodology developed in this work is based around a new subjective rating scale – the Simulation Fidelity Rating (SFR) scale. The author contributed to the development of the SFR scale along with others at the University of Liverpool and the National

Research Council (NRC) of Canada. This scale has been designed specifically to be industry applicable a d to determine the overall perceptual fidelity of the integrated simulation in a specific role.

A campaign of piloted simulation and flight test trials has been conducted as the core experimental phase of this work. All the pilots completed a series of controlled experiments where a number of flying tasks were completed with a number of varied simulation models. The pilots rated each simulation against a baseline simulation using the SFR scale and their performance and control activity were recorded. This test campaign generated the pilot feedback for the development of the SFR scale and data for development of the quantitative metrics. The tests were also utilised to demonstrate how this assessment methodology can be used in controlled simulation experiments to provide previously lacking supporting evidence to simulator qualification criteria of individual components.

From the analysis of the results, it was found that a more aggressive pilot excites the dynamics of the aircraft to a greater extent, thereby exposing more fidelity issues – leading to poorer SFRs. For similar reasons, an aircraft with degraded HQs was found to cause increased pilot sensitivity to transport delay. Perceptual fidelity was also found to be task dependent, In particular, pilots were found to be more susceptible to changes in off-axis response in the Acceleration-Deceleration manoeuvre than in the Precision Hover manoeuvre. These findings prove that here is a true need for simulation qualification criteria that are based on the intended use of the equipment.

Significant spread was seen in the pilot ratings of perceptual fidelity in a number of cases. This was attributed partly to differing interpretation of the terminology within the Simulation Fidelity Rating (SFR) scale and also to pilot selection of task strategy. Therefore guidance material has been developed by the author from lessons learnt throughout the test campaign. This guidance material is intended to ensure best utility of the SFR scale in the future, to mitigate against the effects of differing interpretation of SFR terminology and variation in pilot task strategy through pilot briefing and correct experimental design.

The SFR scale has been developed in the context of assessing a simulator for the purpose of rotary wing skills acquisition training. However, the methodologies described throughout the thesis are intended to be transferable to more sophisticated training devices for rotary-wing and fixed-wing pilot and crew training as well as for the quantification of the fidelity of certification and design simulators.

ACKNOWLEDGEMENTS

"Yes there were times, I'm sure you knew, when I bit off more than I could chew..."

My Way, Frank Sinatra

This quote seems to sum up the most stressful moments of completing my research and during these times is when my colleagues, friends and family made their biggest contributions to my thesis. I'd like to take a moment to thank all those people who helped me "chew" my way through this PhD.

The greatest thanks must go to my Supervisors, Dr Mark White and Professor Gareth Padfield. Firstly for affording me the opportunity to work with them on what has been and will continue to be a fascinating research field. Secondly, for their invaluable academic guidance, optimism and support. They truly have been the most fantastic mentors for me throughout my studies and, on occasion, pretty good counsellors too. Thanks both. I would also like to thank Dr Philip Perfect for your guidance and collaboration on the Lifting Standards work and support in computer related problems.

Eight very talented and patient test pilots were the cornerstone to my research; particular thanks go to Andy Berryman, Rob Erdos and Stephan Carignan, for their willingness to undergo repeated torture and for their contributions to development of the SFR scale and test methodologies. To Lee Evans, Noddy Holder, Chris Knowles, Russ Cripps and Al Riches, thank you all for taking time out of your busy schedules to help out. I hope you all had fun and return to Liverpool again in the future. A special mention also goes to Bill Gubbels for his continued support with the collaborative work at the NRC.

A personal thanks to Mike Jones for listening to me moan over coffee. It wouldn't have been the same without you mate. To everyone else, past and present, in the school of engineering and particularly in the FS&T lab, Sophie Robinson, Garnet Ridgway, Mike Jump, Linghai Lu, Neil Cameron, Maria White, Denise Bain and Steve Bode, to name but a few. It's been great having such a fantastic group of people to be around every day. My new co-workers have got some big shoes to fill.

I'd also like to thank the friends that I have made through the Royal Aeronautical Society, American Helicopter Society, NRC and GARTEUR AG-21. Each and every one of you has provided me with motivation and words of support and wisdom that have proved invaluable on my journey. Keep in touch.

I have to thank all my family, especially Moomin, Gran & Crawf and Dad for your unwavering belief in me and reminding me what I've worked so hard for. Love you all.

There are not enough thanks in the world for my partner, Kai. He has acted as my emotional crutch (some might say punch bag) over the course of my studies. His understanding, endless patience and uncanny ability to know when I need wine have kept me sane. Thanks babe.

Last, but by no means least, I would like to thank the University and City of Liverpool. This place has been become my home over the last eight years and I have made some fantastic friends here. I hope you all make it out to France soon.

À bientôt Liverpool

The research described in this thesis has been made possible by the Engineering and Physical Sciences Research Council standard research grant EP/G002932/1 and the US Army International Technology Centre (UK) (reference W911NF-11-1-0002).

ACRONYMS

Attps	Attack points per second
Attrt	Attack rate
ACAH	Attitude Command Attitude Hold
AD	Acceleration-Deceleration
ADS	Aeronautical Design Standard
AE	Allowable Error
AG	Action Group
ANOVA	Analysis of Variance
ART	Advanced Rotorcraft Technologies
ASRA	Advanced Systems Research Aircraft
ATC	Air Traffic Control
ATD	Additional Transport Delay
CG	Centre of Gravity
COF	Cut-Off Frequency
COTS	Custom Off-The-Shelf
CT&M	Correct Trend and Magnitude
CSGE	Control System Graphical Editor
CTEF	Cumulative Transfer Effectiveness Ratio
DoF	Degrees of Freedom
Dsp	Control Displacement
DVE	Degraded Visual Environment
EASA	European Aviation Safety Association
EPSRC	Engineering and Physical Sciences Research Council
FAA	Federal Aviation Authority
FBW	Fly-By-Wire
FCMC	Flight Control Mechanical Characteristics
FCS	Flight Control System
FFS	Full Flight Simulator
FLME	FLIGHTLAB Model Editor
FNPT	Flight Navigation Procedural Trainer
FoV	Field of View
FSTD	Flight Simulation Training Device
FTD	Flight Training Device
GARTEUR	Group for Aeronautical Research and Technology in Europe
GM	Gain Margin
GPS	Global Positioning System
GUI	Graphical User Interface
HA	HELI-ACT
HELI-ACT	Helicopter Active Control Technology
HELINV	Helicopter Inverse Simulation
HGS	Helicopter Generic Simulation
HQ	Handling Quality
HQR	Handling Qualities Rating
HQSF	Handling Qualities Sensitivity Function
ICAO	International Civil Aviation Organisation
IG	Image Generator
IHST	International Helicopter Safety Team
IOS	Instructor/Operator Station
ITEF	Incremental Training Effectiveness Ratio
IWG	International Working Group

JAA	Joint Aviation Authority
JAR	Joint Aviation Requirements
JHSAT	Joint Helicopter Safety Analysis Team
LCD	Liquid Crystal Display
LCoS	Liquid Crystal on Silicon
LIVE	Liverpool Virtual Environment
LS	Lifting Standards
MFD	Multi-Functional Display
MTE	Mission Task Element
MTT	Motion Task Team
MUAD	Maximum Unnoticed Added Dynamics
NAA	National Aviation Authority
NASA	National Aeronautics and Space Administration (USA)
NoE	Nap-of-the-Earth
NRC	National Research Council (Canada)
NTSA	National Training and Simulation Association (USA)
OEM	Original Equipment Manufacturer
OMCT	Objective Motion Cue Testing
OTW	Out-The-Window
PC	Personal Computer
PH	Precision Hover
PIO	Pilot Induced Oscillation
PM	Phase Margin
PoM	Proof of Match
PSD	Power Spectral Density
PVS	Pilot-Vehicle System
QTG	Qualification Test Guide
R&D	Research and Development
RCAH	Rate Command Attitude Hold
RMS	Root Mean Square
SATC	Simulated Air Traffic Control
SCAS	Stability and Control Augmentation System
SIMDUCE	Simulator Day UCE
SFR	Simulation Fidelity Rating
SHOL	Ship-Helicopter Operating Limits
STD	Synthetic Training Device
SYCOS	Synthesis through Constrained Simulation
TER	Transfer Effectiveness Ratio
TPV	Task-Pilot-Vehicle
UCE	Useable Cue Environment
UoL	University of Liverpool
VCR	Visual Cue Rating
VMS	Vertical Motion Simulator
V&V	Verification and Validation

NOMENCLATURE

Uppercase

ATD	Additional transport delay [ms]
$A\eta$	Control attack [1/s]
C	Forcing function input [in]
CP	Control Power [deg/s]
C_{PR}	Pitch due to roll cross coupling ratio [nd]
C_{RP}	Roll due to pitch cross coupling ratio [nd]
F	Hypothesis strength [nd]
I_β	2^{nd} moment of flapping inertia [slug-ft^2]
K	Order of Yc(s) transfer function
K_1	Structural pilot model feedback gain [nd]
K_2	Structural pilot model feedback gain [nd]
K_{cc}	Cross-coupling gain [nd]
K_e	Pilot model gain [nd]
K_v	Motion feedback gain in structural pilot model [nd]
K_{turb}	Turbulence gain [nd]
K_β	Flapping spring stiffness [ft-lbf/rad]
M_q	Pitch damping derivative [nd]
M_β	1^{st} mass moment of inertia w.r.t flap hinge offset [slug-ft]
N	Number of rotor blades [nd]
N_A	Number of attack points [nd]
N_f	Number of frequency points in Y(f) [nd]
P_n	Cumulative frequency domain control activity power
P_N	Total frequency domain control activity power
Q	Quickness [1/s]
R	Total rotor radius [ft]
R_p	Pearson correlation coefficient [nd]
T	Total time [s]
T_{tr}	Time to trigger response [s]
T_i	Precision pilot model lead time constant [nd]
T_l	Precision pilot model lag time constant [nd]
T_1	Structural pilot model lead time constant [nd]
T_2	Structural pilot model lag time constant [nd]
U	Longitudinal velocity [ft/s]
V	Lateral velocity [ft/s]
W	Vertical velocity [ft/s]
XA	Lateral Cyclic input [in]
XB	Longitudinal Cyclic input [in]
XC	Collective input [in]
XP	Pedal input [in]
X_u	Sway damping derivative [nd]
Y(f)	Frequency domain control activity signal [in]
$Y_p(s)$	Pilot model transfer function [nd]
$Y_c(s)$	Aircraft model transfer function [nd]

Lowercase

a_0	Lift curve slope [nd]
c	Rotor chord [ft]
e	Error [rad] or [rad/s] and also exponential [nd]
eR	Hinge offset [ft]
f_n	Control deflections per second [1/s]
g	Gravitational acceleration [ft/s^2]
$\dot{h}$	Vertical rate [ft/s]
k	Tau coupling coefficient [nd]
k_1	Acceleration tau coupling coefficient [nd]
k_2	Deceleration tau coupling coefficient [nd]
n_z	Normal acceleration [g]
m	System output [rad] or [rad/s]
m(r)	Rotor mass distribution [slug/ft]
p	Roll rate [deg/s] *and also Probability of the null hypothesis in Chapter 7 [nd]*
q	Pitch rate [deg/s]
r	Yaw rate [deg/s] *and also incremental rotor radius in Chapter 3 [ft]*
s	Laplace transform variable [nd]
t	Incremental time [s]
vxi	Inertial longitudinal velocity [ft/s]
vxb	Body longitudinal velocity [ft/s]
vyi	Inertial lateral velocity [ft/s]
vzi	Inertial vertical velocity [ft/s]
x	Range [ft]
x'	Velocity [ft]
y(t)	Time domain control activity signal [in]

Greek Symbols

ΔC_{RP}	Change in Roll due to pitch cross coupling ratio [nd]
$\Delta\eta$	Change in control deflection [in] or [%]
$\overline{\Delta\eta}$	Mean change in control deflection [in] or [%]
Φ	Phase [deg]
Ω	Rotational rotor speed [rad/s]
γ	Lock number [nd]
ζ	Damping ratio [nd]
$\dot{\eta}_{pk}$	Peak rate of control deflection [%/s]
$\overline{\dot{\eta}_{pk}}$	Mean peak rate of control deflection [%/s]
θ	Pitch attitude [deg]
λ_β	Flapping frequency ratio [nd]
ρ	Ambient air density [slug/ft^3]
σ_n	Cumulative RMS of Y(f) [in]
σ_N	Total RMS of Y(f) [in]
τ	Time to go [s]
τ_e	Effective time delay [s]
τ_p	Phase delay [s]
ϕ	Roll attitude [deg]
ψ	Yaw attitude [deg]
ω_{BW}	Bandwidth frequency [rad/s]
ω_c	Crossover frequency [rad/s]
ω_{co}	Cut-off frequency [rad/s]
ω_n	Natural frequency [rad/s]

CONTENTS

1 BACKGROUND

As the focus of this thesis is flight simulation fidelity, the purpose of this first chapter is to provide the reader with background information regarding flight simulation technology and the benefits associated with the use of flight simulators. Definitions of flight simulation fidelity are given and the current methodologies for assessing simulator fidelity of rotorcraft flight simulators are introduced.

1.1 Flight Simulation Technology

Flight simulation, in the broadest sense, may be defined as a synthetic replication of flight. As early as 1910, the need for such systems was recognised to familiarise pioneering pilots with the control characteristics of aircraft. The first recorded flight simulator was the Antoinette Learning Barrel, shown in Figure 1-1. In this flight training device a pilot was required to use the controls to keep a horizontal reference bar aligned with the horizon as the barrels were moved by human operators to represent pitch, roll and yaw. In hindsight, this representation of the attitudes was incorrect as pilots do not perceive the rotations in this way. However, most early pioneers were not aware of this incongruity.

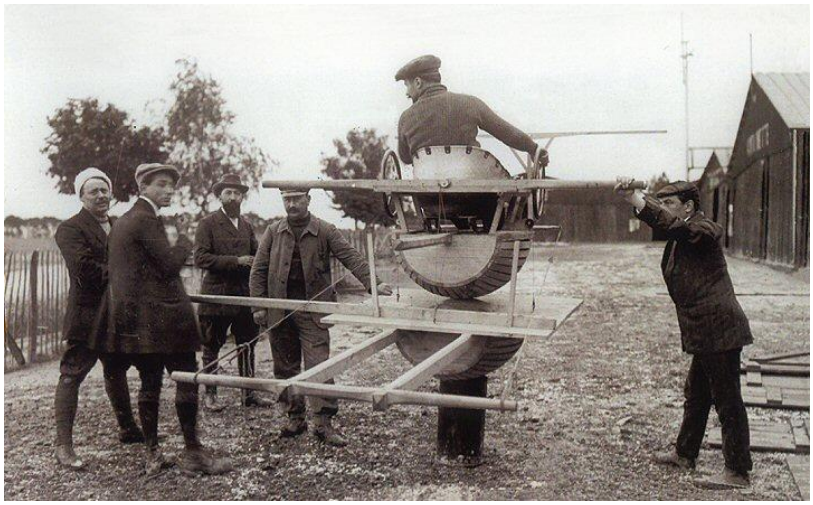


Figure 1-1 - The Antoinette Learning Barrel (Flightglobal.com Historical Photo Archive)

Today, the term 'flight training device' covers a whole host of devices, from effective low cost procedural trainers (Figure 1-2) to full flight simulators with six Degrees of Freedom (DoF) motion platforms (Figure 1-3). Although the complexity of flight simulators vary, all flight simulators will have some combination of the standard simulator components; flight model, cockpit instruments, control forces, visual system, motion system and audio.

Figure 1-2 - CAE Integrated Procedures Trainer (www.cae.com/photos)

Figure 1-3 - CAE NH90 Full Flight Simulator (www.cae.com/photos)

A basic interpretation of the way in which these simulator components interact during real-time simulation is illustrated in Figure 1-4. At the core of the simulation is the mathematical model. The initial aircraft states are set by 'trimming' the aircraft in the desired flight condition. Changes in aircraft states are then computed through the aircraft equations of motion at every time step of the real-time simulation in response to external forces and pilot control inputs. The pilot perceives the changes in aircraft states through a number of cueing systems, shown in Figure 1-4 and described below. The pilot's perception of errors between desired aircraft states and current aircraft states then leads to the pilot making control inputs. These control inputs then feedback into the mathematical model and the aircraft equations of motion are updated and the process begins again.

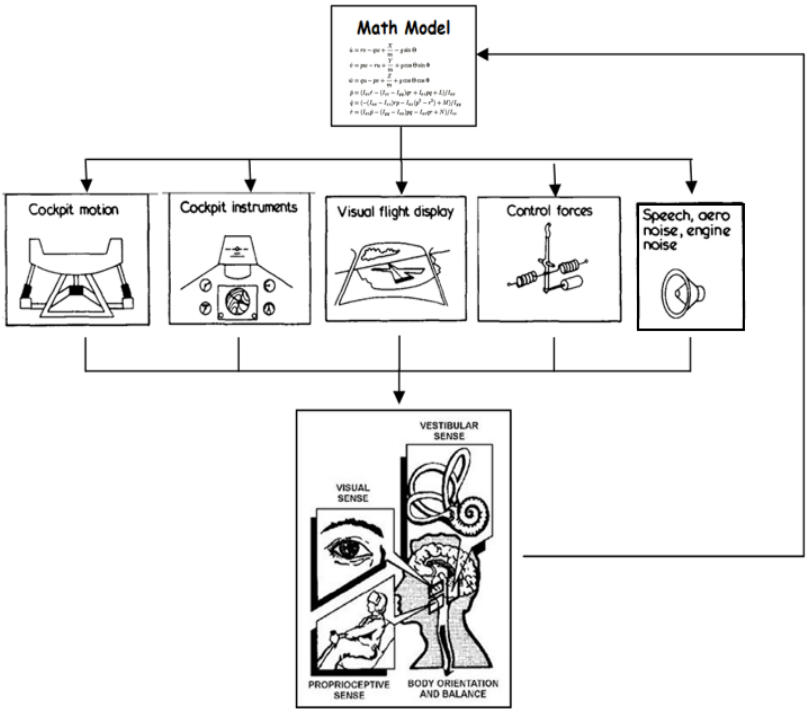


Figure 1-4 - Flight Simulator Components and Flow of Processing (Various Sources)

The mathematical model is a set of equations that compute the aircraft states in real-time. The complexity of the mathematical model can vary from a single Degree of Freedom (DoF) linear representation of a specific aircraft state, to a comprehensive nonlinear model with detailed aerodynamic data. Recent years have seen significant progress in the modelling of aircraft dynamics. This has allowed accurate modelling of rotorcraft in specific operational scenarios, such as helicopters operating to the back of a ship, and modelling of unconventional configurations, such as tilt-rotors. These advances can be generally attributed to the increase in computing power allowing more and more accurate aerodynamic, structural and system data generation as well as rapid computation for real-time simulation.

The visual system simulates the Out-The-Window (OTW) view and provides information to the pilot regarding translational and rotational positions and rates. Helicopter simulation presents a unique challenge for visual cue system technology as helicopter operations are highly varied and often require low altitude and high speed flight. Very low altitude, or Nap-of-the-Earth (NOE), flight requires a high degree of visual cueing to ensure the pilot can be precise in their manoeuvring. High texture, resolution, field of view and scene content are required in to deliver a visual cueing environment that the pilot perceives to be equivalent to flight [1]. For this reason, the requirement for high quality visual cueing is widely accepted. Previous research has led to the suggestion that;

"a non-optimum distribution of field-of-view elements, coupled with a severe lack of near-field detail, compromises the pilot's sensing of translational rates relative to nearby terrain or the landing surface" **[1].**

Visual cue quality has advanced at a significant rate throughout the last few decades as a result of increased computer power, development of real-time visualisation technologies, in the computer gaming industry, and improved projection system technology. Aerial photography has been used to develop detailed, real-world, image databases covering large areas

The cockpit instruments provide further visual cueing to the pilot and are of particular use to the pilot in poor visual conditions. The realism of the instrument panels can have a large impact of pilot immersion and so, in many cases, the Original Equipment Manufacturer (OEM) parts are used for simulator instruments. Where this is not required and/or possible, computer based representation of the control panels are often used.

Cockpit motion delivers vestibular motion cueing to augment the visual cueing of motion by providing information on accelerations and forces to the pilot. In general, cockpit motion is generated using a six DoF hexapod motion platform with acceleration onset cueing followed by washouts below motion threshold levels, although motion platforms with less DoF are also used. Other vestibular cueing features include 'G-seats' and vibration platforms which can be retrofitted to any flight simulator. There is still much debate within the simulation industry as to the benefits of motion cueing and the resources that should be invested into them. Burki-Cohen et al. [2] have conducted numerous studies into the effect of motion on transfer of training for airline pilots. They found there to be no operationally relevant differences between training high workload take-offs and landings in a Full Flight Simulator (FFS) with motion and without motion, irrespective of the vehicle being simulated [2] [3]. However, rotary wing operations and missions are more dynamic and are performed close to the ground and therefore it is widely accepted that pilots rely much more heavily on vestibular cueing than in airline manoeuvring. Studies have confirmed that the presence of motion affects pilot control activity and performance in manoeuvres [4] [5].

The control forces provide the pilot with proprioceptive (force-feel) cues. The realism of control forces is important for training and research simulators because the control forces affect pilot workload [6]. A control loading system connects the pilot controls to the aerodynamic surfaces of the aircraft through either hydraulic or electric actuators.

The complexity of flight simulator audio systems is highly varied. Generic engine and rotor sounds can be used to increase immersion or aircraft specific sounds that respond to changes in aircraft dynamics in real-time can be modelled. Similarly, Simulated Air Traffic Control (SATC) can range from a pre-loaded audio file to an actual human interacting with the simulated environment. The need for audio simulation is highly dependent on the purpose of the simulation and therefore audio systems are highly specialised.

1.2 The Need for Rotorcraft Flight Simulation

Flight simulators are used throughout the lifecycle of aircraft. They are used in Research and Development (R&D) to aid the development of new aircraft designs and upgrades, as well as to develop and evaluate pilot aids, before incurring the safety risks and costs associated with flight testing. Flight simulators are also used to support flight test campaigns for qualification of final aircraft designs. However, the most frequent use of flight simulators is as Flight Simulation Training Devices (FSTDs) to train pilots, flight crews and maintenance engineers.

The use of flight simulators (for pilot training in particular) has a significant economic advantage over the use of aircraft. The United States National Training and Simulation Association (NTSA) have estimated the operating cost of full flight simulators to be 5-20% of the operating cost of aircraft [7]. The NTSA also found the average ratio of cost of actual flight hours to simulation flight hours to be 17/1 over a number of military airframe types (see Table 1-1).

Table 1-1 - Relative Cost of Simulated versus Actual Flight Hour [7]

Airframe	Cost/Actual Flight Hour	Cost/Simulated Flight Hour	Ratio
F-16	$5000	$500	10/1
FA-18A	$3955	$217	18/1
P-3C	$2903	$119	24/1
S-3A	$4360	$143	30/1
SH-60B	$1724	$118	15/1
CH-47	$3000	$435	7/1
		Average Ratio:	17/1

The use of flight simulation not only reduces the cost of training but also plays a large role in aviation safety. A flight simulator provides a safe environment in which to train and also provides pilots with an opportunity to practice procedures that are deemed too dangerous to practice in flight, such as engine failures or unusual attitudes. The Joint Helicopter Safety Analysis Team (JHSAT) report from 2006 [8] details the analysis of 152 U.S. registered civil helicopter accidents that occurred in the Calendar Year 2006 (CY2006) as well as recommendations for preventing future similar accidents. JHSAT reported that 68% of CY2006 helicopter accidents were attributed to pilot judgement (Figure 1-5). It was also found that 18% of all accidents occurred during training (Figure 1-6), highlighting the need for more training and for that training to be conducted in a safer environment.

In September 2005, the International Helicopter Safety Team (IHST) announced a goal to reduce helicopter accidents by 80% by 2016. The 2006 JHSAT report recommended an increased use of flight training devices and simulators with emphasis on training of emergency procedures [8] as the most essential method to help reach this goal (see Figure 1-7).

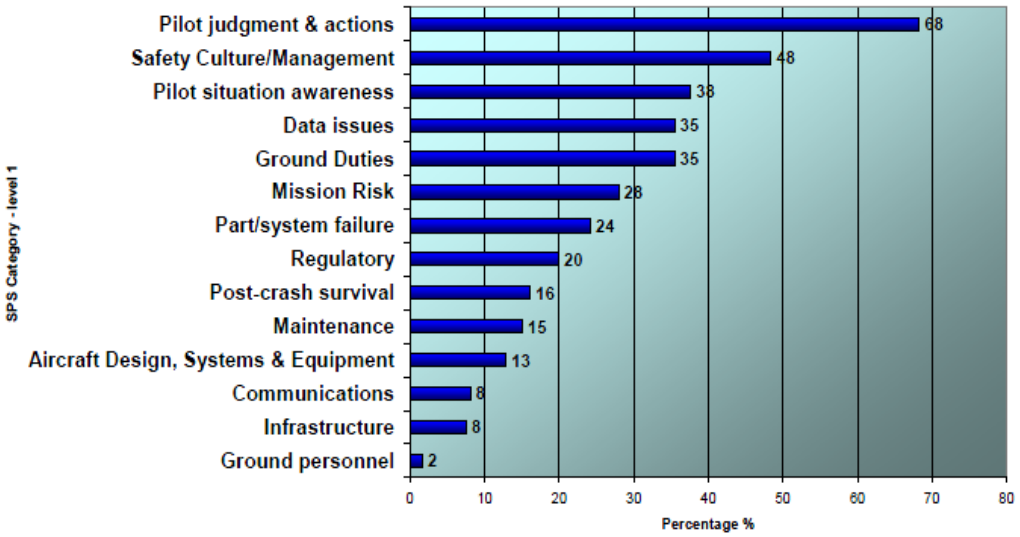


Figure 1-5 - Standard Problem Statement Results [8]

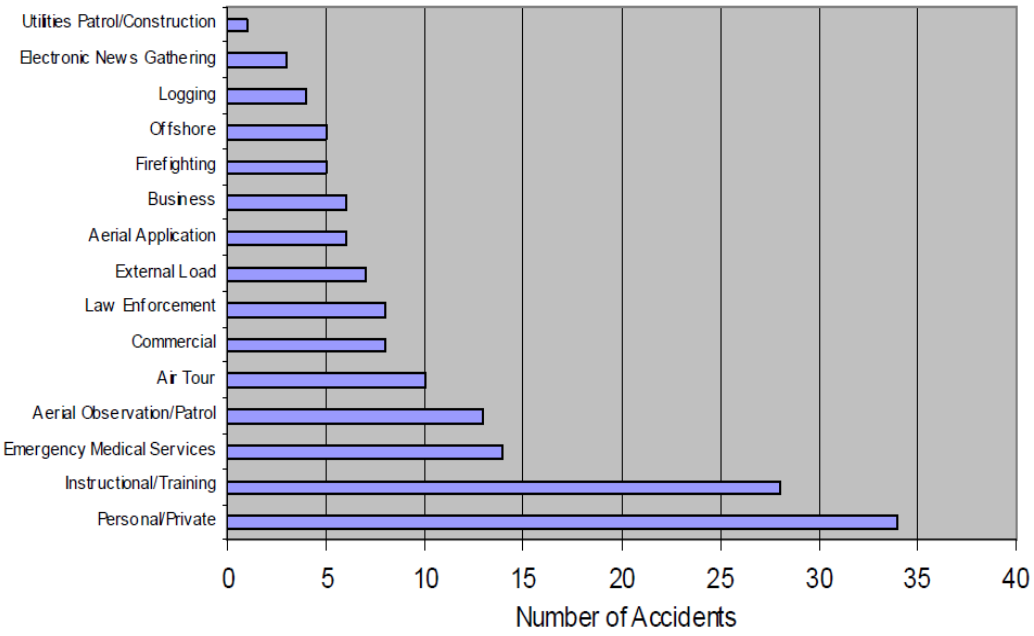


Figure 1-6 - Number of Accidents by Primary Operation [8]

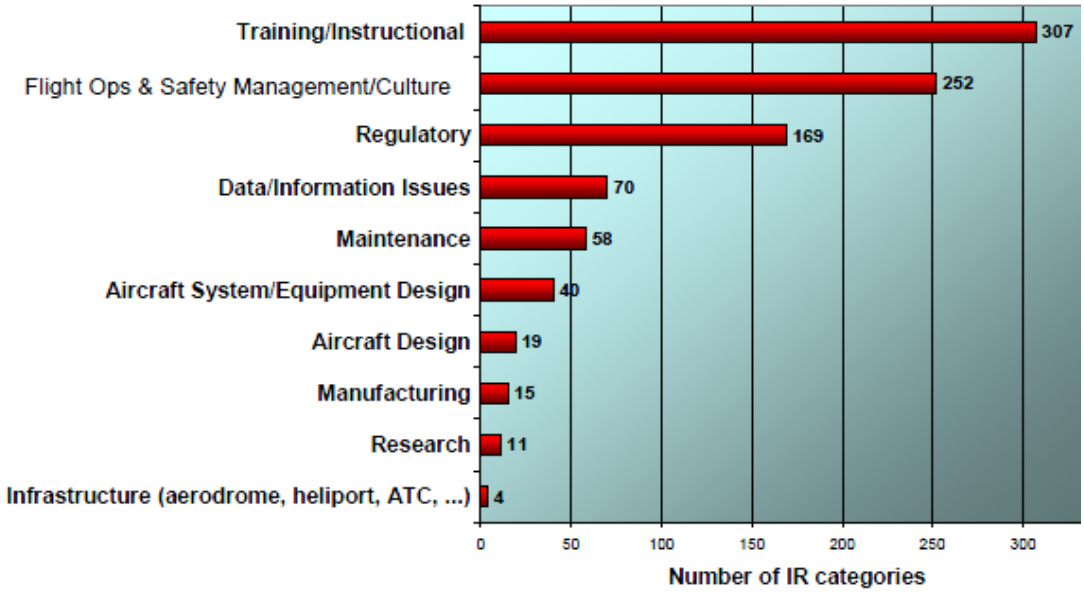


Figure 1-7 - Distribution of Intervention Recommendation Categories [8]

In addition to the cost and safety benefits, the use of flight simulators for training allows the instructor full control of the environmental conditions and operational scenarios. This allows the pilot to repeat critical decision making scenarios that may rarely be experienced in flight. For such reasons, synthetic training has repeatedly been deemed not only to be more cost effective and safer than live training, but can also provide a more effective training environment.

The extent to which flight simulators are utilised has seen significant growth in recent years and despite the economic recession, flight simulator sales are still on the rise [9]. This growth can be attributed, in part, to the advancements in virtual environment technology and computing power.

However, because complex systems are more expensive, there is an inherent trade-off between simulator sophistication and cost. To fully utilise the potential of flight simulation, care must be taken to ensure that choices regarding the components, and the way in which they are integrated, are carefully considered. Too little sophistication may compromise the fitness for purpose of the device and too much sophistication may lead to redundancy in the device which compromises the efficacy of the simulator as a cost saving tool.

1.3 Flight Simulation Fidelity

The term ‘Fidelity’ is often used to reflect the sophistication of simulator hardware and software. One definition of fidelity found in the Oxford English Dictionary is;

"The degree of exactness to which something is copied or reproduced".

Therefore flight simulation fidelity may be inferred as;

The degree to which a flight simulator matches the characteristics of the real aircraft [10].

This is assuming that the 'something' that is being replicated is the aircraft, and a more sophisticated (higher fidelity) simulator would replicate the aircraft more truthfully.

However, there is another interpretation of fidelity that assumes the 'something' to be replicated is the pilot behaviour and performance. This leads to an alternative definition of fidelity as;

The simulator's ability to induce the (trainee) pilot to output those behaviours known to be essential to control and operation of the actual aircraft in performance of a specific task [11].

Both of these definitions are valid and widely used in industry. The practical difference between these two definitions is that, using the first definition, a desktop trainer will always be considered as a ‘low fidelity’ device and a full-flight simulator a ‘high fidelity’ device. However, if the second definition is adopted, then the desktop trainer may be considered as high fidelity for achieving a particular purpose (e.g. instrument training) and a full flight simulator may be deemed a low fidelity device for some purposes (e.g. upset recovery or mission rehearsal).

As well as a lack of a common understanding of the term “fidelity” itself amongst the industry, there are many different descriptors used to distinguish between the two facets of simulation fidelity described above. Examples reflecting the type of fidelity outlined in the first definition above include equipment fidelity, objective fidelity, physical fidelity and predictive fidelity. For the second definition, terms such as behavioural fidelity, functional fidelity, environmental fidelity and perceptual fidelity are seen throughout the literature. Hay summarised this problem well in 1980 [21] but still, there is no formally recognised terminology.

1.4 Flight Simulation Assessment

In 2001, the first formal helicopter flight simulator qualification requirements; Joint Aviation Requirements (JAR)-STD 1H [12], were published by Joint Aviation Authority (JAA). An update to these standards, JAR-FSTD H [13], was published in 2008. The JAA documents were then superseded by the European Aviation Safety Agency (EASA) document CS-FSTD (H) [14] in 2012. CS-FSTD (H) defines 3 types of flight simulation training device; Full Flight Simulator (FFS), Flight Training Device (FTD) and Flight Navigation Procedural Trainer (FNPT). A number of Levels are then used to reflect the fidelity (sophistication) of the device. All simulator types and respective Levels can be seen in Table 1-2.

Table 1-2 - CS-FSTD (H) Flight Simulator Types and Qualification Levels

FFS Level	FTD Level	FNPT Level
Level D	3	MCC (multi-crew co-operation)
Level C	2	III
Level B	1	II
Level A	-	I

The number of training credits that can be obtained in an FSTD is ultimately determined by the training provider. However the FSTD device level has a large impact on the number of training credits that can be gained through synthetic training. In a Level D FFS, the majority (if not all) of type conversion training can be conducted in the simulator. As you move down and/or to the right of the table, the sophistication of the device reduces. With little evidence to support the training benefits associated with the lower level FSTDs, the number of credits that are awarded to such devices is limited. This causes problems for small training centres as the acquisition of a Level D simulator is often not an economically viable option when considering the cost of running small aircraft or having only a small fleet.

A need to determine a relationship between training needs and required sophistication of the FSTD was identified by industry and, in response to this, a Royal Aeronautical Society (RAeS) led International Working Group (IWG) was initiated to develop a new set of criteria that were focused on training needs. From this work, two new manuals for training simulator qualifications have been developed; International Civil Aviation Organization (ICAO) 9625 ed. 3 Vol. I [15] for fixed-wing and ICAO 9625 ed. 3 Vol. II [16] for rotary-wing simulators.

A list of training tasks was defined and the level of representation (none, generic, representative or specific) for each of the 12 simulator fidelity characteristics (Figure 1-8) required to train each task was defined. The ICAO 9625 documents contain several matrices

to allow an operator to determine for what training tasks and training type a current FSTD can be used and also to allow for specification of new FSTDs based on a known training objective. The ICAO 9625 ed. 3 vol. II is, at the time of writing, not officially utilised by any National Aviation Authority (NAA) but is being gradually adopted.

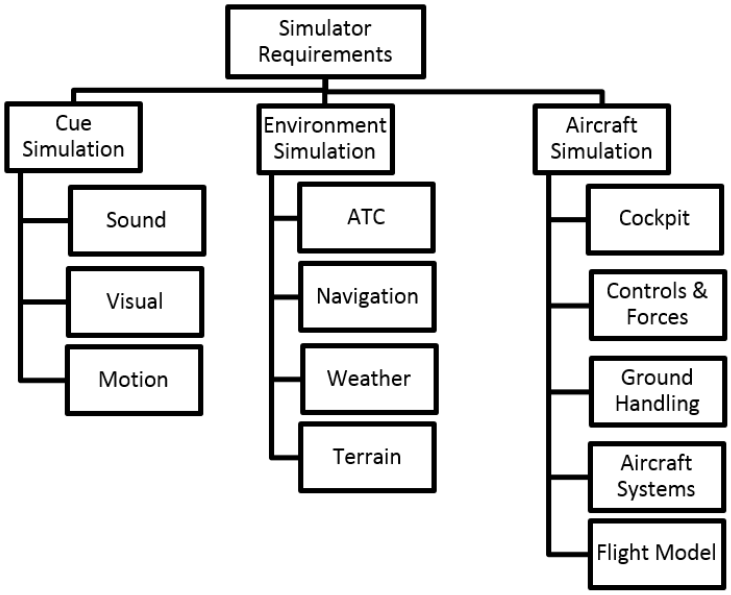


Figure 1-8 - ICAO 9625 Simulation Fidelity Characteristics

Although the development of the ICAO 9625 documents has provided guidance for fidelity assessment based on training needs, a number of longstanding issues remain in the assessment methodology. In the EASA and United States Federal Aviation Authority (FAA) frameworks, the simulator is assessed by considering 'the degree to which a flight simulator matches the characteristics of the real aircraft' (Fidelity definition 1 [10]). Individual simulator components (Figure 1-9) are assessed against flight test data, through a number of quantitative validation tests, to obtain Proof of Match (PoM) evidence. Subjective pilot opinion is then used to verify the realism of the simulation before qualification. In 2002, The Group for Aeronautical Research and Technology in EURope (GARTEUR) Helicopter Action Group, HC-AG12: (Validation Criteria for Helicopter Real-Time Simulation Models), began a critical examination of the JAR-STD 1H criteria for the aircraft flight dynamics models (performance and Handling Qualities (HQ)), which are unchanged in the current EASA CS-FSTD(H) criteria, and revealed a number of shortcomings [17].

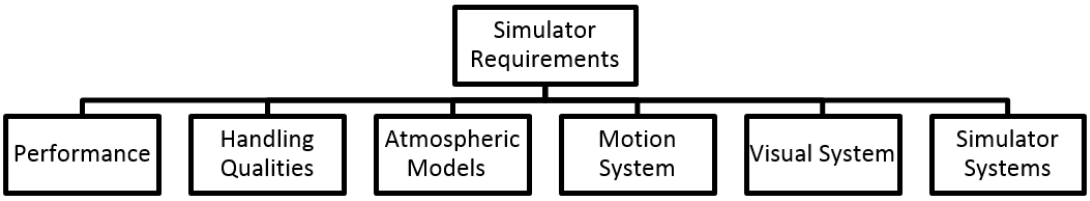


Figure 1-9 - CS-FSTD (H) Simulator Component Breakdown

It was found that physical tuning of the flight model can achieve approximately 80% match with flight test data across the flight envelope and the remaining 20% is achieved through non-physical adjustment of model parameters [17]. The effect of the non-physical tuning may have undesirable effects in areas of the flight envelope that are not checked in the Qualification Test Guide (QTG) and particularly at the edge of the flight envelope. The AG-12 group recommended that a more robust and systematic method for achieving the required match to flight test data was required.

HC AG-12 also established that many of the mathematical model criteria in JAR-STD 1H were derived from the preceding fixed-wing document (JAR-STD 1A [18]) with no experimental justification for their applicability to rotorcraft applications. Therefore meeting the requirements is not a clear indication of whether a mathematical model is sufficiently representative of the real world vehicle. Supporting data and analysis techniques are required to verify that adhering to the current criteria guarantees that a simulator is of sufficient quality for the required purpose.

One of the case studies in the AG-12 work was a sensitivity study in which a landing manoeuvre was simulated with a variety of trajectories, all within the JAR-STD 1H landing validation test limits [12] as shown in Figure 1-10. It was revealed that conducting the landing at the upper and lower boundaries of acceptable deviations from the nominal case resulted in unsafe landing velocities. It was also shown that if a different piloting technique was used then the upper limit case results in safe landing velocities where the nominal case results in unsafe landing velocities. This highlights that the fidelity is dependent on the task. This dependency is not considered in JAR-STD 1H, JAR-FSTD H or EASA CS-FSTD (H). The dependency of fidelity requirements on the training task was noted by the IWG during the development of the ICAO 9625 documents. However, it was beyond the scope of the IWG work to gather supporting data or to address the validity of the standards.

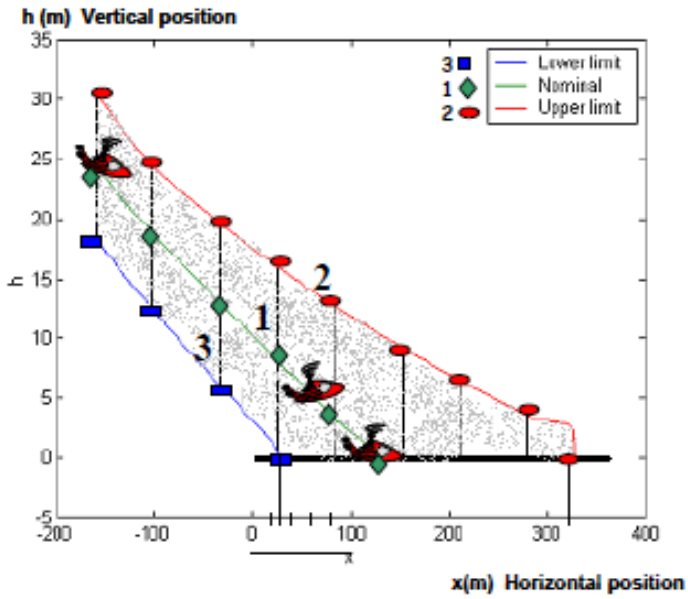


Figure 1-10 - Upper and Lower JAR-FSTD H Allowable Errors in Landing Trajectories [17]

EASA CS-FSTD (H) requires PoM of time histories in terms of response magnitude and overshoot frequency and damping. The work by GARTEUR HC AG-12 found that changes in pitch attitude phase delay were not revealed from time domain PoM tests. However, this change in phase delay has a significant effect on the aircraft handling characteristics and therefore pilot strategy adopted. The off-axis response criteria were also criticised in the GARTEUR HC AG-12 report. The criteria state that, following a four second step input, the off-axis attitude response should be of 'Correct Trend and Magnitude' (CT&M). Subpart C part b.2.ii.C of EASA CS-FSTD (H) states:

"*Where the tolerances have been replaced by 'Correct Trend and Magnitude' (CT&M), the FSTD should be tested and assessed as representative of the helicopter to the satisfaction of the Authority. To facilitate future evaluations, sufficient parameters should be recorded to establish a reference. For the initial qualification of FNPTs no tolerances are to be applied and the use of CT&M is to be assumed throughout*" [14] .

This requirement is open to interpretation and considered to be a weakness of the standards.

The GARTEUR HC AG-12 report suggested that;

"A *fresh examination of cross coupling for simulation fidelity assessment is warranted*"[17]

It was also recommended that the current standards should be supplemented by Aeronautical Design Standard (ADS) Performance Specification, ADS-33E-PRF [19], Handling Qualities cross-coupling and frequency domain criteria.

Unlike training simulators, there are no formal requirements published for design and development or research simulators for any specific or general application. This is in part due to the lack of flight test data available and also due to the wide range of applications. Instead, it is the responsibility of the researcher to justify and provide the evidence that the accuracy of their simulation ensures that the results attained using that simulation will transfer to the operation or design of the aircraft. As a result, the literature is rich with methodologies for simulator fidelity assessment developed by the research community. Such methodologies are the subject of the technical review in Chapter 2.

1.5 Overall Simulation Fidelity

'Lifting Standards: A Novel Approach to the Development of Fidelity Criteria for Rotorcraft Flight Simulators' began as an Engineering and Physical Science Research Council (EPSRC *research grant EP/G002932/1*) funded project at The University of Liverpool (UoL) in response to the findings of the GARTEUR-AG 12 activities. The scope of the Lifting Standards project has since been extended to include work by Hodge, as well as the research presented in this thesis (see section 1.6). The earlier work undertaken in Lifting Standards aimed to bridge the gap between pilot subjective opinion and formal metrics and to develop an objective methodology for assessing the overall fidelity of the pilot-in-the-loop system. The work by Perfect et al [20] as part of the Lifting Standards project defined two complementary subgroups of fidelity based on the definitions given on page 8 from references [10] and [11]:

a. ***Predicted Fidelity*** - the degree to which a flight simulator matches the characteristics of the real aircraft.

b. ***Perceptual fidelity*** - the simulator's ability to induce the trainee pilot to output those behaviours known to be essential to control and operation of the actual aircraft in performance of a specific (pilot-in-the-loop) task.

Around these two definitions, a generalised methodology for fidelity assessment based on predicted and perceptual fidelity was outlined, as shown in Figure 1-11. It can be seen the simulator requirements are first defined based on its intended purpose. The predicted fidelity is then assessed by comparing quantitative data from the aircraft and simulator, including dynamic responses to open-loop control inputs. Improvements made where necessary until the predicted fidelity is deemed acceptable. The perceptual fidelity can then be assessed using subjective opinion and measurements of pilot behaviour in flight and simulation. This process iterates until the simulator is deemed to be fit for purpose.

As noted in section 1.3, there are many other sub categorisations of fidelity found in the literature. It is therefore necessary to define which terminology is used (and what the interpretation of this terminology is) to ensure a common understanding. Because the work in this thesis is founded on the early Lifting Standards work by Perfect et al [20], the terminology of "predicted fidelity" and "perceptual fidelity" are used throughout this thesis, with the definitions as stated above

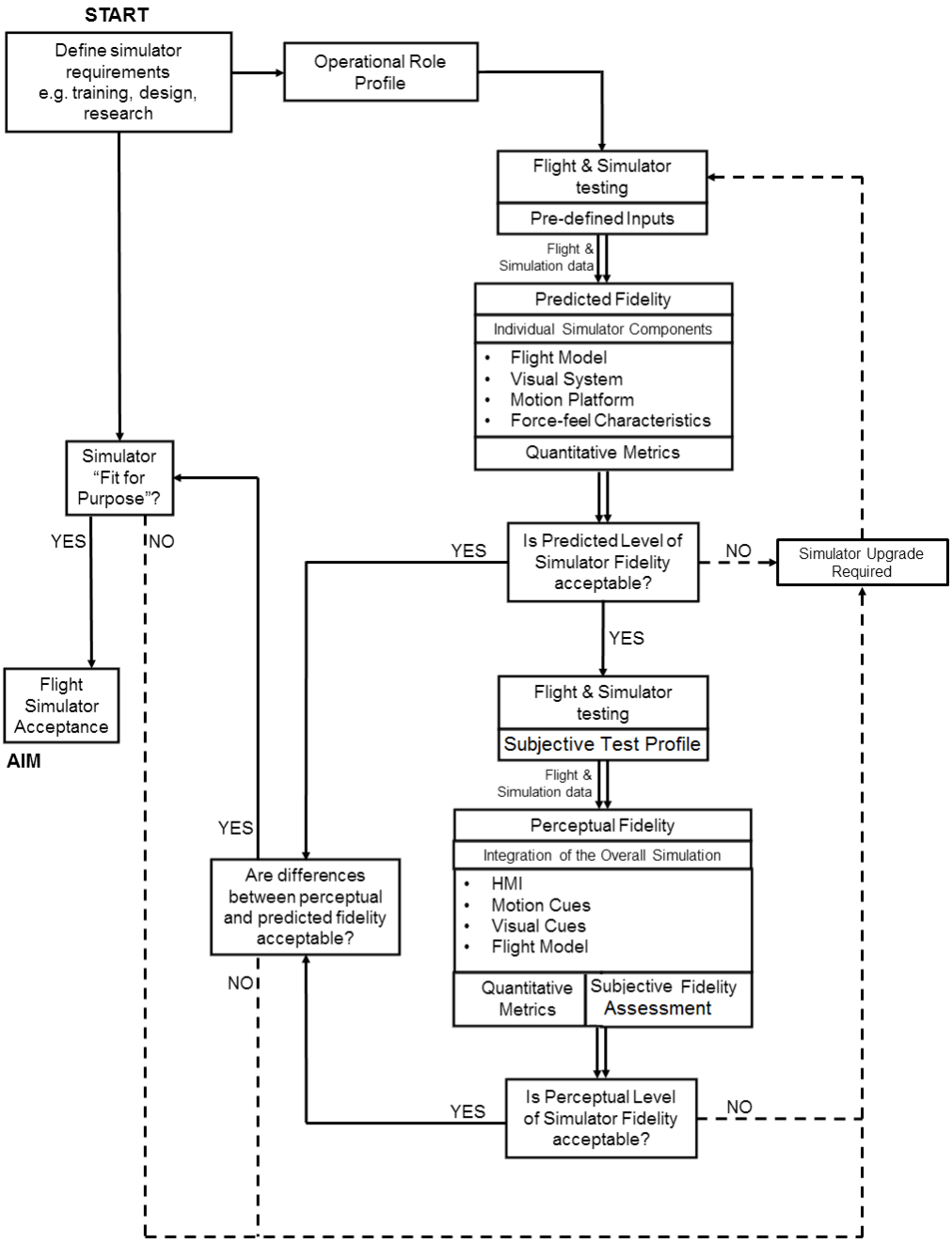


Figure 1-11 - General Approach to Flight Simulation Fidelity Assessment [20]

1.6 Research Scope

The proposal for the research described in this thesis is to continue the work of 'Lifting Standards, through examination of the appropriateness of the current EASA regulations and development of novel metrics and methodologies for assessment of overall simulation fidelity.

The novelty of this thesis is that the quantitative metrics and assessment methodologies will be purposefully designed for industrial application. Rather than clinically measuring the physical differences between the aircraft and the simulator, the metrics and methodologies developed in this work will allow the user to gain an understanding of the fitness for purpose of the integrated pilot-vehicle system through comparisons of pilot behaviour and performance in flight and simulation.

This approach can be compared to the differences in the measurement of mental demand load and task demand load. Reference [21] defines task demand load as "The mental workload imposed by the system to be supervised". Mental demand load is defined as "The mental workload experienced by the operator". Like predicted fidelity, task demand load regards the physical parameters of the system: The same plant will always deliver the same task demand load and the same simulator will always deliver the same predicted fidelity. However, measurements of perceived fidelity (and mental demand load) will vary depending on the human operator.

The benefit of measuring mental demand load and perceived fidelity is that the real-world performance of the integrated pilot-vehicle system is measured. The disadvantage of such methods is the inherent variability in the data obtained. As a result, the scope of this research includes detailed data analysis and cross referencing of subjective and objective data to understand the sources of variability in perceived fidelity in order to develop metrics and methods that are sufficiently robust for industrial application.

Before being able to define the specific aims and objectives of this study (given in section 2.5), a review of the current methods available for assessing flight simulation perceptual and predicted fidelity is required, to ascertain what further research is needed in order to develop a robust, complete and comprehensive methodology for flight simulation fidelity assessment Chapter 2 describes a number of the methodologies uncovered in this technical review and the conclusions drawn from the review set the aims and objectives of this thesis.

In Chapter 3, the simulation hardware and software used for piloted experiments are discussed in terms of their capabilities and Proof of Match (PoM) with aircraft data. This is followed by description of a new subjective rating scale for overall fidelity assessment in Chapter 4, with a case study and guidance for its correct utilisation. Piloted simulation trials that showcase the utility of the rating scale for defining tolerances for quantitative metrics are detailed in Chapter 5. The results from these trials lead onto two empirical studies into measurement of pilot workload and adaptation in Chapters 6 and 7. Finally, Conclusions and recommendations from the research are presented in Chapter 8.

2 TECHNICAL REVIEW

This chapter summarises a number of techniques for assessing both predicted and perceptual fidelity. The advantages and limitations of the techniques are discussed and case studies presented. The conclusions of this technical review set the aims and objectives of the work in this thesis.

2.1 Methods for Assessing Simulator Predicted Fidelity

The predicted fidelity of each simulator component can be individually assessed quantitatively. The challenge regarding predicted fidelity assessment is what the most appropriate metrics are and what values of such metrics are acceptable and what are not. In the following subsections, metrics for assessment of the mathematical model, visual system and motion system are discussed.

2.1.1 Comparative Handling Qualities

The Handling Qualities (HQs) of an aircraft are defined as;

"*Those qualities or characteristics of an aircraft that govern the ease and precision with which a pilot is able to perform the tasks required in support of an aircraft role*" [21].

ADS-33E-PRF defines the handling qualities requirements for military rotorcraft in the US. The handling qualities of an aircraft can be 'predicted' using dynamic response criteria drawn from the response to clinical tests such as pulse, step, doublet and frequency sweep control inputs. HQ metrics have been developed to assess the full range of aircraft response, from low to high frequency and from small to large amplitude (Figure 2-1). Each frequency-amplitude region has an associated handling qualities dynamic response criteria. These criteria are discussed below.

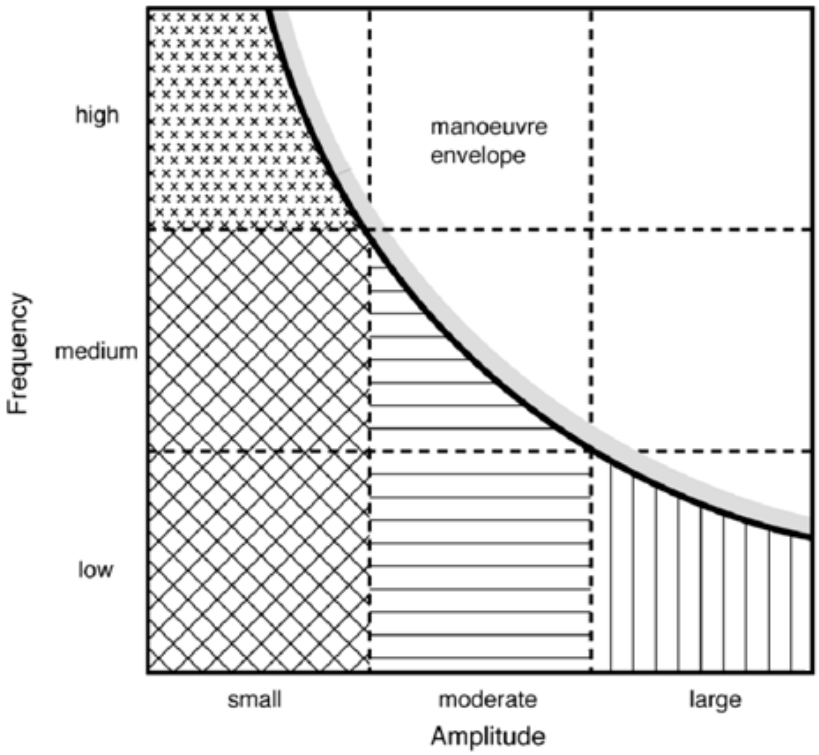


Figure 2-1 - Dynamo Construct for Dynamic Response Criteria [24]

i. Small amplitude, high frequency - bandwidth and phase delay. Bandwidth, ω_{BW}, is a stability measure that defines the range of control input frequencies over which a pilot can apply closed-loop control without threatening the stability of the aircraft. ADS-33E-PRF provides two definitions of bandwidth, depending upon the response type of the aircraft. For a rate response type, it is the lesser of the gain bandwidth (the frequency corresponding to a gain margin of 6dB) and the phase bandwidth (the frequency corresponding to a phase margin of 45° relative to the 180° attitude response phase), shown in Figure 2-2. For an attitude response type, it is equal to the phase bandwidth. The phase delay, τ_p, is a measure of the slope of the phase response beyond 180° phase, and is defined as

$$\boldsymbol{\tau_p = \frac{\Delta\Phi_{2\omega 180}}{57.3 \times 2\omega_{180}}} \qquad \textbf{Equation 2-1}$$

Where $\Delta\Phi_{2\omega 180}$ is the phase change between ω_{180} and $2\omega_{180}$.

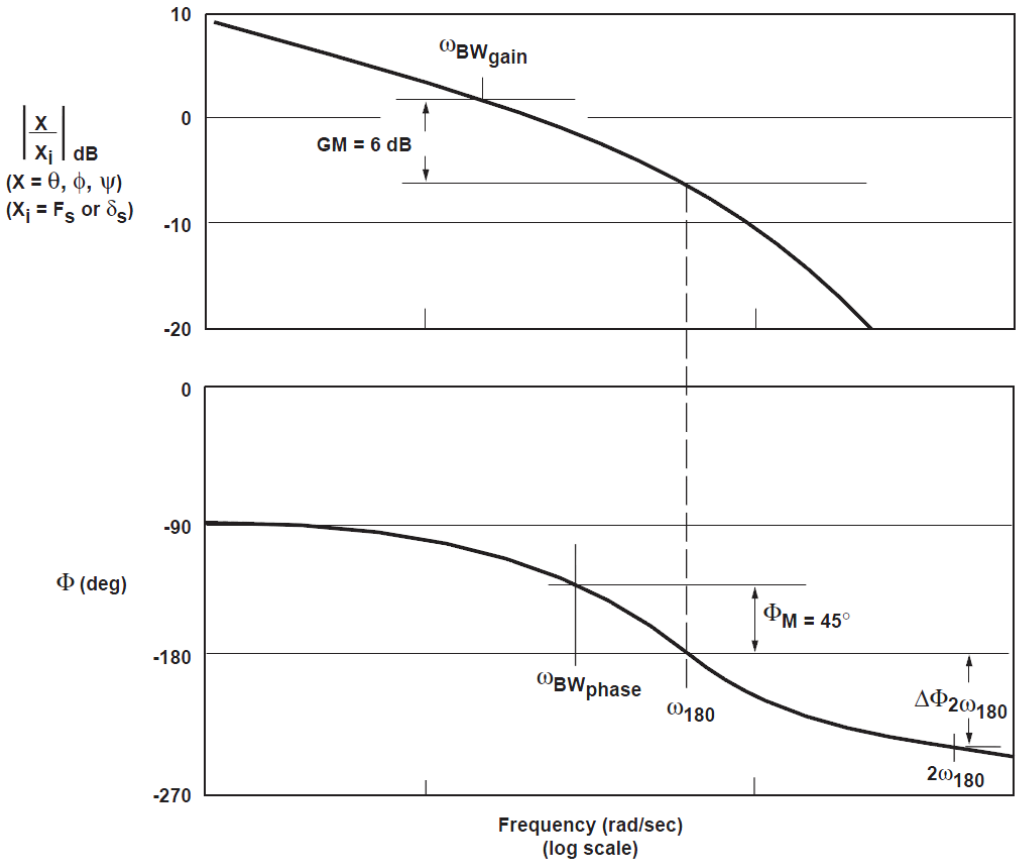


Figure 2-2 - Bandwidth Definition ADS-33E-PRF [19]

ii. Small amplitude, low to medium frequency - open-loop stability. Stability is quantified in terms of the natural frequency, ω_n, and damping, ζ, of the aircraft's natural modes of motion and are determined from eigenvalue analysis.

iii. Moderate amplitudes - quickness. Attitude quickness, Q, provides a measure of the ability to rapidly attain moderate amplitude attitude changes. It is defined as the ratio of peak attitude rate to the attitude change after a sharp input (step for attitude command and pulse for rate command). For example, pitch quickness is defined as

$$\boldsymbol{Q_\theta = \frac{q_{pk}}{\Delta\theta_{pk}}} \qquad \textbf{Equation 2-2}$$

iv. Large amplitudes - control power. Control Power, CP, is defined as the maximum attitude or rate response achievable by applying full control from a trim condition.

There is also a further set of HQ metrics that must be satisfied; the inter-axis coupling metrics. These define the level of acceptable off-axis response to control inputs.

i. Roll/Pitch Coupling is defined as the ratio of the peak off-axis attitude response to the on-axis attitude response 4 seconds after a sharp step input in lateral/longitudinal cyclic. For example, pitch due to roll coupling is defined as

$$\frac{\boldsymbol{\theta_{pk}}}{\boldsymbol{\phi(4)}} \qquad \textbf{Equation 2-3}$$

ii. Pitch due to collective coupling is defined as the ratio of the peak change in pitch attitude response to the peak change in normal acceleration within 3 seconds of a step input in collective;

$$\frac{\boldsymbol{\Delta\theta_{pk}}}{\boldsymbol{\Delta n_{z_{pk}}}} \qquad \textbf{Equation 2-4}$$

iii. Yaw due to collective cross coupling is defined from

$$\boldsymbol{r\ from\ XC\ @\ 1s} = \left|\frac{r_1}{\dot{h}(3)}\right| \qquad \textbf{Equation 2-5}$$

and

$$\boldsymbol{r\ from\ XC\ @\ 3s} = \left|\frac{r_3}{\dot{h}(3)}\right| \qquad \textbf{Equation 2-6}$$

where r3=|r(3)-r1| and r1 is either the first peak in yaw rate response within 3 seconds after application of a collective step input, or, the yaw rate 1 second after the application of a collective step input.

The knowledge of the potential for HQ metrics as fidelity metrics predates both the GARTEUR AG-12 work and the JAR documents. In 1992, at a NASA/FAA Helicopter Simulator Workshop, Key proposed bandwidth as a measure of simulator fidelity, particularly pertaining to the assessment of delays in the system [25]. He suggested that the stick-to-visual bandwidth should not degrade into Level 3, unless the aircraft itself had a Level 3 bandwidth. Application of this criterion would therefore mean that simulations of aircraft with lower (degraded) bandwidth would require a simulation with smaller computational delays. However, the question was left open as to how much handling qualities fidelity is required for transfer of training.

In 1993, Strope et al. used comparison of flight test and mathematical model on-axis bandwidth for verification and validation of a UH-60A with slung-load flight dynamics model [26]. It was stated that frequency domain validation is important due to pilot cueing accuracy

being frequency dependent. It was also noted that time domain errors are still highlighted in frequency domain Verification and Validation (V&V) as phase shifts.

In 1996, Padfield et al. [27] and later McCallum and Charlton [28] proposed the use of the handling qualities standard, ADS-33E-PRF, for deriving fidelity metrics for research simulators; the rationale here being that if the simulator is to be used to optimise handling qualities, then it is logical that the simulator and flight vehicle should possess the same handling qualities for high fidelity. Advani and Wilkinson [29] and Roscoe and Thompson [30] also presented an approach using comparative measures of performance and control activity, correlated with handling qualities ratings given for the same tasks flown in simulation and flight.

The Lifting Standards project at UoL [20] was initiated in 2008 in response to the work undertaken by GARTEUR AG-12 [17] and others [28], [29], [30]. Comparison of flight and simulation ADS-33E-PRF HQ frequency and time domain metrics have been used to assess the predicted simulator fidelity of the flight dynamics model. However, as yet, no boundaries on acceptable and unacceptable differences in handling qualities metrics have been determined. As with HQs these are likely to depend on task.

In 2006, analysis of results from a study at CAE [31] showed that the handling qualities Pitch/Roll coupling parameter showed potential as a fidelity metric for off-axis response. A tentative boundary of cross coupling ratio error was proposed between ± 0.40 and ± 0.50. A recommendation was given for further tests to validate this boundary.

2.1.2 Maximum Unnoticed Added Dynamics

In the early 1980s, an effort was undertaken by engineers at the McDonnell Douglas Corporation to develop 'Maximum Unnoticeable Added Dynamics' (MUAD) envelopes [32]. The aim of these envelopes was to define regions of acceptable levels of mismatch in equivalent-systems in terms of phase and magnitude. Then, any change in the vehicle dynamics that falls within the limits would be too small to be detected by a pilot, implying high perceptual fidelity. The MUAD envelopes were originally drawn from a set of in-flight simulator test data for pitch control of a fixed-wing airplane. Mitchell et al conducted an experiment in a fixed-base helicopter simulator to assess the applicability of the MUAD envelopes for roll axis dynamics of a hovering helicopter [33]. 200 combinations of additional dynamics were tested and pilot opinion on whether the effects were noticeable and if they affected the hovering task were noted. It was found that the MUAD envelopes were too stringent and therefore not appropriate for this application. In response to this, a new set of

envelopes were developed from the generated data and termed 'Allowable Error' (AE) Envelopes to distinguish from the MUAD envelopes. It was found that a higher bandwidth baseline lead to less stringent AE envelopes [33].

2.1.3 Motion Cueing System Quality

The first quantitative criteria for rotorcraft motion cueing fidelity were developed by Sinacori in 1977 [34]. He proposed that the gain and phase distortion between the aircraft model and commanded motion system accelerations at different frequencies could be used to quantify the fidelity of the motion system. The criterion boundaries were generated using pilot subjective opinion. An example plot for a frequency of 1 rad/s (common pilot operating frequency) is shown in Figure 2-3. Three regions; high, medium and low fidelity (clarification given underneath plots) are defined from both the phase distortion (y-axis) and gain (x-axis) of the specific force (left plot) and rotational velocity (right plot).

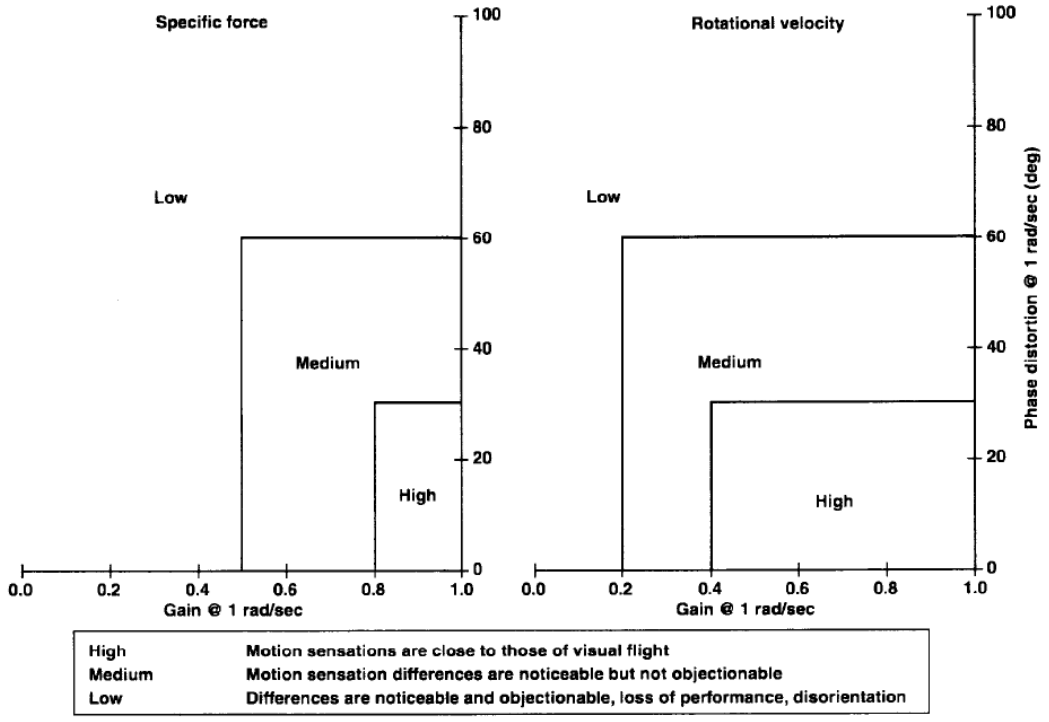


Figure 2-3 – Sinacori Motion Fidelity Criteria at 1 rad/s [34]

However, with this criterion, low phase distortion and high gain are required simultaneously. Research by Stroosma et al. [35] has shown that this is very difficult to achieve even with large motion travel and therefore the Sinacori criterion defined even highly sophisticated motion systems as having low predicted fidelity.

In the update from JAR-STD 1H to JAR-FSTD H a number of quantitative criteria were added for the assessment of motion platforms (and have been carried through to CS-FSTD (H)). These criteria requires that from frequency tests in all six axes between 0.1 and 1 Hz (0.63 to 6.3 rad/s) the phase distortion must be between 0 and -20 degrees and a modulus of ± 2 dB. For the same tests between 1.1 and 3 Hz (6.9 and 18.8 Hz) the phase delay must be between 0 and -40 degrees and a modulus of ± 4 dB. The JAR-FSTD H criteria at 1 rad/s (0.63 Hz) are overlaid on the Sinacori chart in Figure 2-4. It can be seen that the EASA CS-FSTD (H)) requirements are even more stringent than the Sinacori Criteria.

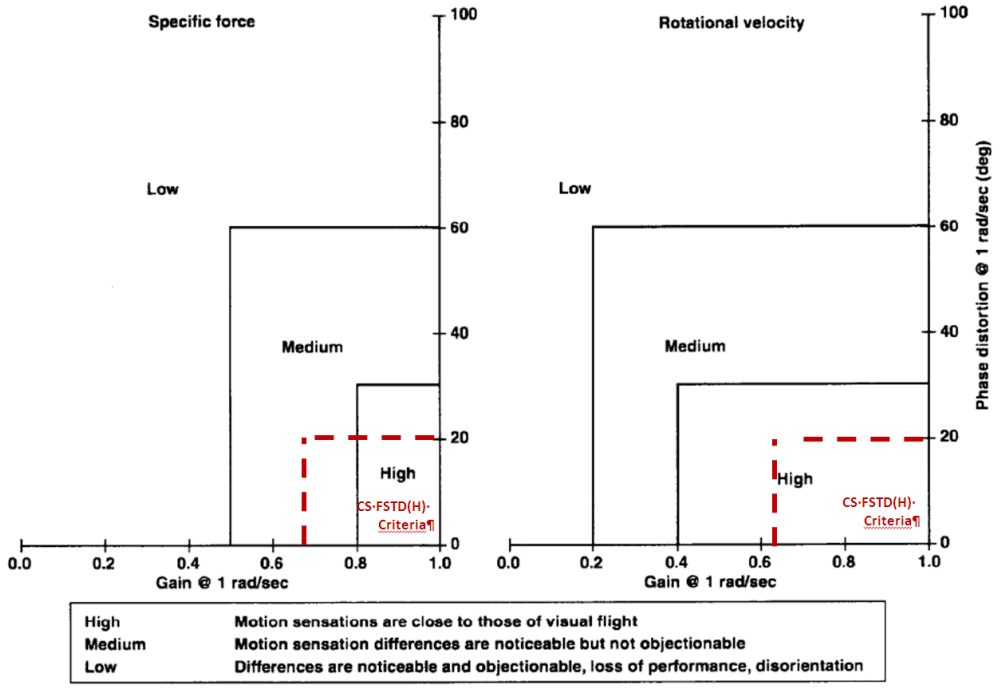


Figure 2-4 - Comparison of EASA CS-FSTD (H) and Sinacori Motion Criteria at 1 rad/s

An alternative method for quantifying simulator motion cue fidelity, The Objective Motion Cue Test (OMCT) was developed by Advani and Hosman in 2006 [36]. The purpose of this test is to objectively measure the frequency response of the complete motion cueing system. The OMCT quantifies the motion cueing using the frequency domain input to, and output from, the motion algorithm. The test is carried out using a sinusoidal input and the phase distortion and modulus of the motion is measured. A Royal Aeronautical Society International Working Group (IWG) responsible for the development of the new ICAO 9625 Vol. II, ed. 3 [16] has appointed a Motion Task Team (MTT) to evaluate and validate the OMCT. As part of this work, a large number of partners have gathered data for fixed wing simulators and validated the results against experienced pilot opinion. This data have now been collated and analysed in the attempt to define OMCT boundaries. However, as yet, no

such data gathering attempt has been undertaken for rotary wing simulators and there is no evidence that the fixed wing results are not transferable to rotary wing.

Heffley et al. conducted a detailed study using human operator control theory and pilot comments to determine motion requirements in failure detection tasks and tracking tasks [37]. It was noted that motion is required to offer a clue of a failure. This is of particular importance in scenarios where the pilot may not have sufficient spare capacity to obtain visual or audio cues of a failure. It was also found from this study that for tracking tasks, rotational motion is more influential than translational motion, although it warns against exclusion of translational motion. Furthermore, it was noted that pilot comments of fidelity changed from one task to another, further highlighting the fidelity requirement dependency on task discussed earlier.

Hodge et al conducted a piloted simulation study into the required vestibular cues for roll-sway and yaw axes in low speed helicopter manoeuvring [38]. Pilot ratings and measures of pilot-vehicle performance showed that rotational cues were the most important in yaw tasks with slight improvement with the addition of translational motion. This study also concluded that for fully coordinated roll-sway manoeuvres, the selection of motion washout filter gains and reduction of phase distortion introduced by the motion drive algorithm was essential for acceptable motion cues.

2.1.4 Visual Cueing System Quality

EASA CS-FSTD (H) includes minimum requirements for visual system parameters such as Field of View (FoV), resolution, and scene content. The requirement for scene content is as follows:

"Sufficient scene content to recognise aerodromes, operating sites, terrain, and major landmarks around the FATO area and to successfully accomplish low airspeed/low altitude manoeuvres to include lift-off, hover, translational lift, landing and touchdown" [14]

The term *sufficient* allows room for interpretation. The reason for this flexibility in the standards is that requirements for scene content are task specific - a NoE precision manoeuvre or lateral sidestep will require higher content and higher texture resolution than a higher altitude, higher speed cruise. However, the visual scene content has a large effect on the pilot strategy in terms of what visual cues are utilised to sense rates and

positions/attitudes. If the visual scene content causes the pilot to use different cues than they would in flight, or cues that simply are not available in flight, then the transfer of training may be compromised.

Few empirical studies have been conducted with regards to visual system requirements although pilot comments on the inadequacy of simulator visual cueing, particularly in low level flight are common.

2.1.5 Transport Delay Measurement

In the current simulation standards [14], transport delay is defined as;

"The total Synthetic Training Device (STD) system processing time required for an input signal from a primary pilot flight control until motion system, visual system or instrument response. It does not include the characteristic delay of the helicopter to be simulated." **[14].**

The total transport delay from control input to visual and motion response must be no more than 100ms for the highest Level of simulator qualification. Previous research at NASA Ames [4] investigated the effect of varying simulator transport delay on Handing Qualities Ratings (HQRs) [19]. The results showed that additional transport delays of only 80ms resulted in degradation of the average HQRs from Level 1 to Level 2 in the Vertical Motion Simulator (VMS) for several tasks. This would suggest that a simulator with an additional 80ms transport delay would lead to compromised training utility. It should be noted that the baseline transport delay in the VMS is only 10ms due to lead compensation filters that are used to eliminate the delays in the motion and visual systems [4].

The total transport delay is dependent on both the visual system delay and the motion system delay. If the motion and visual system transport delays are not correctly synchronised, it is likely that the pilot will experience conflicting visual and vestibular cues. This leads to disorientation which can cause the pilot to feel sick and compromises learning benefits. On this topic, EASA CS-FSTD (H) states

"*Visual scene changes from steady state disturbance shall occur within the system dynamic response limit but not before the resultant motion onset*" [14].

An experiment outlined in reference [25] demonstrated that pilots are unable to ascertain the source of the delay perceived. The delays associated with the motion system were found to be more complex than the visual system due to washout filters. It was suggested that the

visual and motion delays should be matched rather than trying to reduce delays in each system independently. However, Chung, Schroeder and Johnson obtained results [39], which indicated that visual cues should be synchronous with motion cues, or at worst, the visual cues should lead the motion cues.

The contradictions in results regarding transport delay and the EASA CS-FSTD(H) requirements suggests that more research is required to determine the best way to minimise transport delays and synchronise motion and visual cues.

2.2 Methods for Assessing Simulator Perceptual Fidelity

Simulator perceptual fidelity is concerned with the fitness for purpose of the overall simulation, as perceived by the pilot, rather than the degree of similarity of the individual components of the simulator to their equivalents in flight. Therefore perceptual fidelity assessment is focused on the closed-loop behaviour of the vehicle and the pilot. Klyde et al refer to such an approach to fidelity assessment as the flight-centred approach [6]. Such an approach utilises quantitative measurements of the pilot behaviour and task performance of the pilot-vehicle system in conjunction with subjective pilot opinion. Overviews, merits and limitations of various methods for determining each of these four aspects of perceptual fidelity are discussed in the following subsections.

2.2.1 Analysis of Task Performance

If the pilot is not able to achieve comparable performance in the simulator as in the aircraft then the utility of the simulator as a training aid should be called into question. Furthermore, if the performance of the aircraft is different to the performance of the simulator, the handling qualities are, by extension, different and therefore the utility of the simulator as a research tool is compromised. Therefore it is reasoned that the difference in task performance achieved between flight and simulation has potential to be used as a measure of perceptual fidelity.

It may be more intuitive to think of the simulator defects resulting in the pilot achieving poorer performance in the simulator compared to the aircraft. This is often the case, particularly where the cueing environment is degraded in the simulator compared to real-world flight. However, there are also many cases when the simulator deficiencies may result in the pilot achieving better performance than would be permitted in the aircraft. This may occur where the simulator has access to artificial cues that do not exist in the real-world,

where the flight dynamics modelling neglects complex nonlinearities or where turbulence/wake effects are not properly modelled.

ADS-33E-PRF defines a number of Mission Task Elements (MTEs) for the assessment of aircraft handling qualities [19]. The tasks are defined in terms of task objectives, a task description and a set of quantitative performance tolerances for 'desired' and 'adequate' performance. For example, in the Precision Hover MTE, the pilot is required to maintain height within ± 3ft for desired performance and ± 6ft for adequate performance. Similar requirements exist for plan position and heading. This can be useful for fidelity assessment as it can be determined whether the pilot achieved the same 'Level' of performance in flight and simulation.

Numerous studies have been conducted where MTEs have been flown in flight and simulation and the performance compared for assessment of simulation fidelity [6] [40] [41]. However, no limits on acceptable/unacceptable differences in performance have been defined for general use as results from studies to date have been very specific.

The benefit of using MTE's and task performance measures is that the well-defined nature of the MTEs allows for relative simplicity of the data analysis. The disadvantage of such a technique is that these MTEs have been defined to test the limits of the Handling Qualities of the aircraft, not to assess simulation fidelity, and therefore are often more aggressive than training/operational tasks conducted in a training simulator.

ICAO 9625 Vol. II [16] includes a list of 115 rotary wing training tasks. However, there are no performance metrics associated with these tasks at present. It is suggested that 'MTE like' definitions for performance could be applied to these tasks to allow for operationally relevant fidelity assessment tasks to be defined.

2.2.2 Analysis of Pilot Behaviour

The definition of perceptual fidelity is based around the transference of pilot behaviours from the simulator to flight. In the context of a design/development simulator, the effect of modifications on pilot workload is of specific interest. For training simulators, learning of incorrect behaviour or operational strategy can be dangerous if the pilot first experiences the same scenario in flight when in sole command of the aircraft. Therefore, from a simulator assessment point of view, efforts must be made to ensure that the pilot is controlling the simulator in the same way they control the aircraft.

Padfield et al formulated a time domain metric called 'control attack' [42] that simultaneously accounts for the rapidity and magnitude of a pilot's control inputs. The attack metric, A_η, is able to characterise each discrete control input in a particular axis during a manoeuvre and is defined as the ratio of the peak rate of control deflection, $\dot{\eta}_{pk}$, over the magnitude of the change in control deflection, $\Delta\eta$, see Equation 2-7. Rapid, small control inputs lead to highest values of attack while slower, larger movements lead to the lowest values. An example of a high attack and low attack control input are shown in Figure 2-5.

$$A_\eta = \frac{\dot{\eta}_{pk}}{\Delta\eta} \qquad \textbf{Equation 2-7}$$

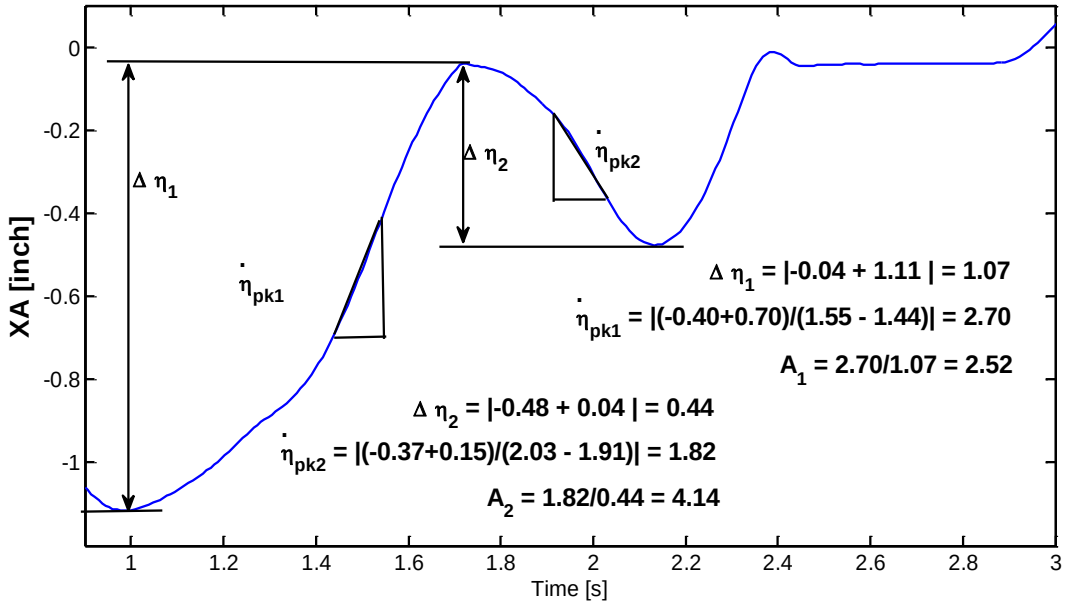


Figure 2-5 - Calculation of Lateral Cyclic (XA) Control Attack

The 'Lifting Standards' work at UoL has utilised the control attack metric and developed a set of three metrics to fully reflect the pilot compensation in a given control axis over the duration of a closed-loop flying task [20]:

i. Attack per second (attps), f_η - The total number of discrete control deflections divided by the manoeuvre time; this is a normalised measure of the activity, a kind of busyness metric.

$$\mathbf{f_\eta} = \frac{N_A}{T} \qquad \textbf{Equation 2-8}$$

ii. Mean Attack Rate, $\overline{\dot{\eta}_{pk}}$ - The mean of the peak rates of the control deflections. This is determined by calculating the attack rate for each attack point and taking the average mean.

iii. Mean Control Deflection, $\bar{\eta}$ - The average amplitude of individual control deflections.

$$\overline{\boldsymbol{\eta}} = \frac{\boldsymbol{\Delta\eta_1+\Delta\eta_2+\Delta\eta_3+\Delta\eta_4+\ldots\Delta\eta_{N_A}}}{\boldsymbol{N_A}} \qquad \textbf{Equation 2-9}$$

A two dimensional chart can be used to graphically present all of the attack metrics, an example of which is shown in Figure 2-6. The individual data points show the range of control activity and also the number of attack points. The mean data point represents a centre of gravity of the pilot control activity in terms of attack and control deflection magnitude. The lines of constant attack rate give an indication of the mean attack rate. The further towards the top left hand corner, the higher the compensation.

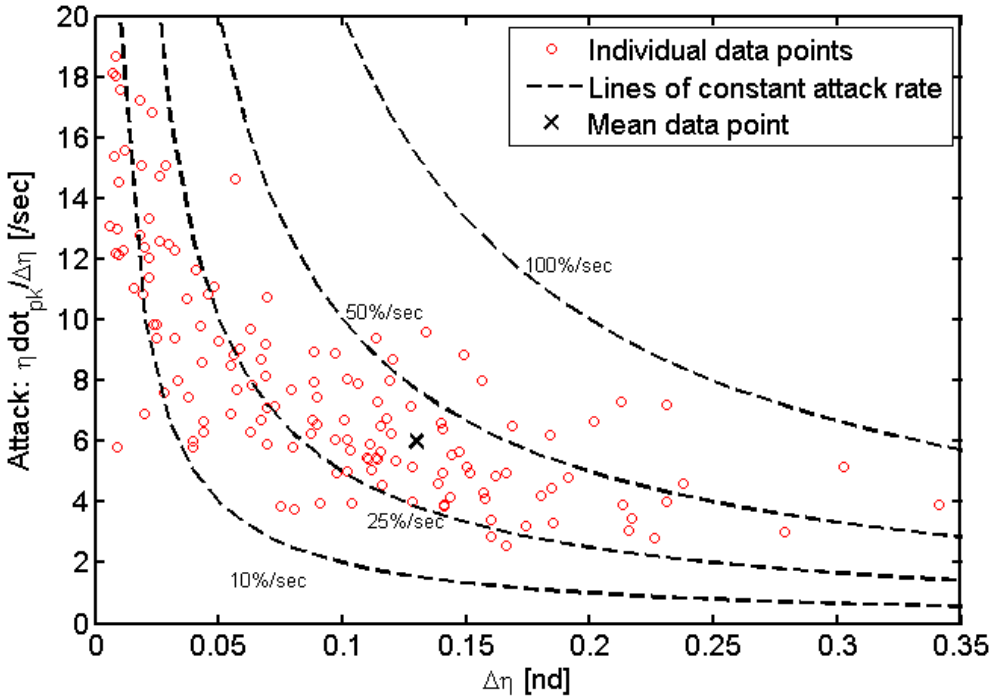


Figure 2-6 - Example of an Attack Chart (for an Acceleration-Deceleration MTE)

The Lifting Standards work has compared the control attack metrics in flight and simulation during a number of MTEs. An example comparative attack chart is shown in Figure 2-7 and shows that the control activity in the lateral axis was higher in the aircraft than in the simulator in the Acceleration-Deceleration MTE. It was concluded that control attack metrics have been shown to be sensitive to differences between flight and simulator and are in agreement with pilot comments [20]. However, no boundaries of acceptable/unacceptable

differences between flight and simulation were proposed for perceptual fidelity assessment in this study due to the very limited database.

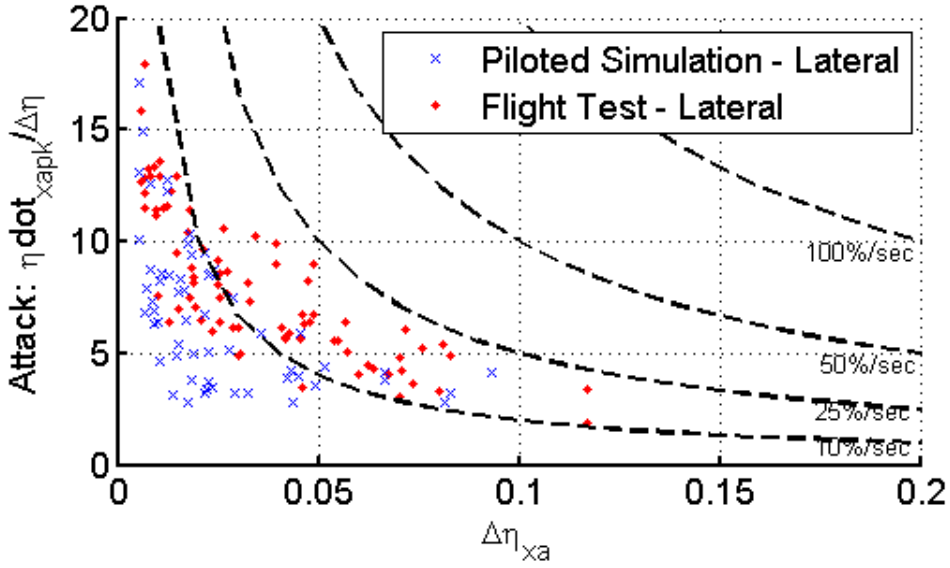


Figure 2-7 - Comparative Control Attack Charts between Flight and Simulation for an Acceleration-Deceleration Manoeuvre [41]

Tischler and Remple developed a frequency domain measure of pilot control activity; cut-off frequency [43]. The cut-off frequency is defined as the frequency at which 50% of the cumulative power of the pilot control signal Power Spectral Density (PSD) is reached, or 70.7% of the Root Mean Square (RMS).

To obtain the PSD of the control signal, the discrete Fourier transform of the time domain signal y(t) must be taken to obtain the frequency domain signal Y(f).

σ_N, is then the square RMS of the squares of the frequency domain magnitudes Y(f), i.e.

$$\boldsymbol{\sigma_N} = \sqrt{\frac{Y(f)_1{}^2+Y(f)^2+Y(f)_3{}^2\cdots+Y(f)_N{}^2}{N_f}} \qquad \textbf{Equation 2-10}$$

and the total cumulative power, P_N, is

$$\boldsymbol{P_N = \sigma^2} = \frac{Y(f)_1{}^2+Y(f)_2{}^2+Y(f)_3{}^2\cdots+Y(f)_N{}^2}{N_f} \qquad \textbf{Equation 2-11}$$

The cut-off frequency, ω_{co} can they be determined:

$$\boldsymbol{\omega = \omega_{co} \quad when \quad \frac{P_n}{P_N} = 0.5 \quad or \quad \frac{\sigma_n}{\sigma_N} = \sqrt{0.5} = 0.707} \qquad \textbf{Equation 2-12}$$

In the same way as the attack metrics, cut-off frequency can be calculated from flight and simulation control time histories and compared as a measure of perceptual fidelity. The cut-off frequency metric has been used in a number of flight vs. simulation comparisons and has shown good correlation with subjective pilot opinion of the fidelity of the simulation [40], [44], [45]. Again, no boundaries of acceptable/unacceptable differences between flight and simulation have been proposed for perceptual fidelity assessment.

The way in which the pilot manipulates the aircraft controls is not the totality of the pilot strategic behaviour. Pilot scan activity, mental workload, mission demands such as communication and system monitoring are all aspects of pilot behaviour. Therefore, psychologists and human factors engineers take a much more empirical view of comparative pilot behaviour for simulator fidelity assessment: if the pilot is truly behaving in the same way in the aircraft in the simulator, then the pilot's psycho-physiological responses will be the same in the aircraft and the simulator. Any changes in perspiration, heart rate, pupil dilation etc. would suggest a difference in stress/workload levels and may also be an indication of the pilot's belief that the simulator provides the illusion of real-world flight is compromised.

Schnell et al utilised the Cognitive Avionics Tool Set (CATS) to measure the effects of simulator fidelity on neuro-cognitive and physiological patterns exhibited by pilots [46]. They hypothesised that CATS could be used to determine flight simulation fidelity criteria. One experiment compared measurements of between an instrumented L-29 Jet Trainer Aircraft and a fixed base flight simulator. The pilot was fitted with a number of sensors including, electroencephalogram electrocardiogram, respiration and eye gaze tracking. The CATS toolbox allowed a high level of integration of the sensors into a single sensor vest to minimise the effect of the sensors on the pilot. In excess of 380 variables were obtained from the pilot sensor data. The results suggested a significant difference in neuro-cognitive behaviour between a fixed-base simulator and flight, particularly heart beats per minute and electroencephalogram power data [46].

There are concern that the practicalities of using such metrics for fidelity assessment are limited as the metrics will be highly dependent on the test individual and calibration of measurement equipment is complex. Additionally, typical equipment for physiological measurement include sensor vests, face masks, skin sensors and helmet mounted equipment. Using such equipment may introduce stress or physical excursion on the pilot, thereby biasing the results.

Thus far, measures of perceptual fidelity based on post analysis of experimental data have been considered. However, in the 1960's human operator theory [47] emerged as a method of analytically assessing the behaviour of a human operator interacting with a dynamic system. Such analytical methods provide the potential for a purely analytical assessment of perceptual fidelity that could be used prior to piloted simulation trials, thereby exposing major fidelity problems early in the development process.

2.2.3 Assessment of the Pilot-Vehicle System

For assessment of the pilot-vehicle system, a mathematical model of the pilot must be generated. The human pilot is a multimodal, adaptive, learning controller capable of exhibiting an enormous variety of behaviour [48]. However, it is possible to model the human pilot with a reasonable degree of success using quasi-linear models; a linear approximation of the pilot plus a remnant function, which accounts for control outputs that are not predicted by the linear model. For an in depth discussion and review of quasi-linear pilot models, reference [48] is recommended. The input to the system, or forcing function can be a random appearing compensatory tracking signal on a single state (the most investigated use of pilot models) or, in the case of a discrete manoeuvring task, time-optimal command(s) for the controlled state(s). The output of the system is then fed back and compared with the command signal. The pilot model then responds to an error by applying a control action. Figure 2-8 shows the basic structure for this type of model, where $Y_p(s)$ is the model of the human pilot and $Y_c(s)$ is the model of the controlled element, i.e. the aircraft.

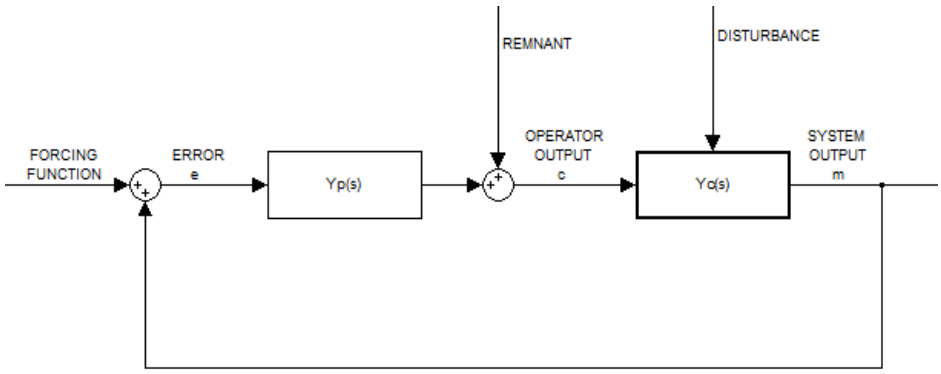


Figure 2-8 - Quasi-Linear Human Operator Model Structure [48]

A classic frequency domain measure with such systems is its crossover frequency - the frequency at which the magnitude of the ratio of the output, m, to the forcing function input, C, (the transfer function) of the system is unity (i.e. 0dB). Below the crossover frequency, the output follows the input, whereas above the crossover frequency the system becomes

essentially open loop [47]. Crossover frequency is determined graphically from the system Bode plot. An example of cross-over frequency determination is shown in Figure 2-9.

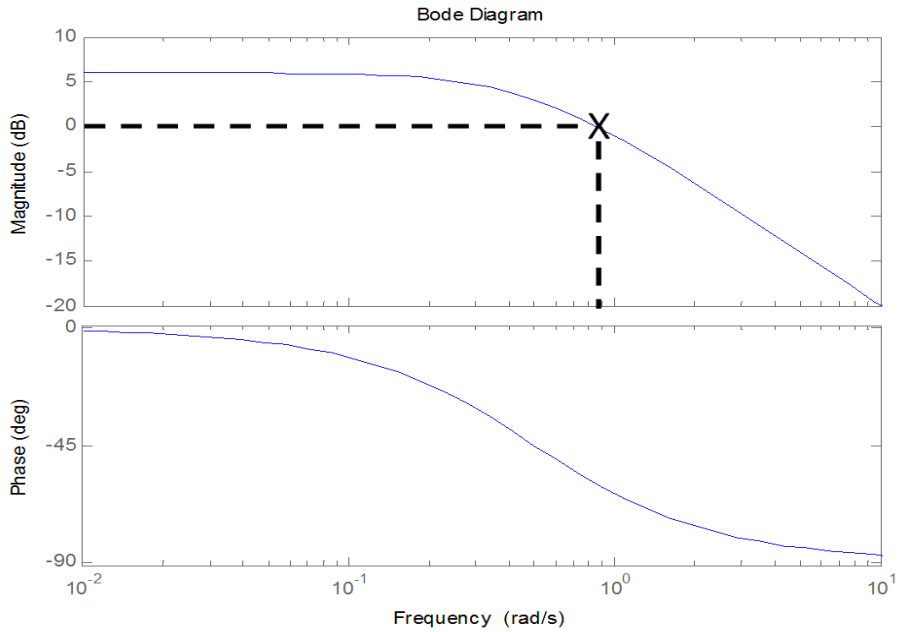


Figure 2-9 - Illustrating Crossover Frequency Determination

The Crossover Model, formulated by McRuer and Krendel [47], is a general criteria for Pilot-Vehicle System (PVS) modelling that states that the pilot adjusts their piloting technique to ensure that the combined PVS characteristics satisfy conditions for closed-loop stability and reasonably low error, i.e. that the resulting transfer function has K_e/s like characteristics around the crossover frequency, ω_c. Graphically, this response is seen as a -20dB/decade slope on the system Bode plot in the region of the crossover frequency. Mathematically, the crossover model is described as;

$$\boldsymbol{Y_p(S) \cdot Y_c(S) = \frac{\omega_c e^{-\tau_e s}}{s}} \qquad \textbf{Equation 2-13}$$

where τ_e is the effective time delay

An in depth analysis of many compensatory tracking tasks by McRuer et al [49] led to the development of a simple pilot transfer function, $Y_p(s)$ that consists of a pilot gain, K_e, a lead and lag time constant, T_l and T_i, and a pure transport delay. The transfer function can be written as;

$$\boldsymbol{Y_p(s) = K_e \cdot \frac{Tl.s+1}{Ti.s+1} \cdot e^{-\tau_e s}} \qquad \textbf{Equation 2-14}$$

The pure transport delay, τ_e, is taken to be 0.14 for K_e/s like characteristics [47]. The pilot gain, lead time constant and lag time constant are selected based on a number of *adjustment*

rules put forward by McRuer et al, including the necessity to compensate for the controlled element $Y_c(s)$ to ensure K/s like characteristics (-20dB/decade slope in the cross over frequency), minimise tracking errors and obtain an appropriate crossover frequency, ω_c [47].

In 1985, Hess developed a more refined pilot model, the structural pilot model [50]. The structural pilot model isolates the contributions of the different pilot sensory systems to the error perception. The motion cues provide information on accelerations which is then fed back into the central nervous system where visual tracking information is received, which then feeds into the neuromuscular system (see Figure 2-10). This means that this pilot model is more suitable than the simple pilot model described above for modelling the human pilot behaviour in discrete manoeuvring tasks.

In 1991, Hess proposed a methodology for assessment of flight simulation (perceptual) fidelity using the structural pilot model. The hypothesis was that task specific perceptual fidelity of the primary control loop could be assessed by determining the similarity of the predicted pilot transfer functions and primary loop closed-loop transfer functions. An example of this process is given in ref [51]. It was noted that similarity of the primary closed-loop transfer function is most easily quantified in terms of bandwidth. As a general guidance, it was suggested that the bandwidths of the primary closed-loop transfer functions should not differ by more than one-fourth of a decade, or a factor of approximately 1.5 [51].

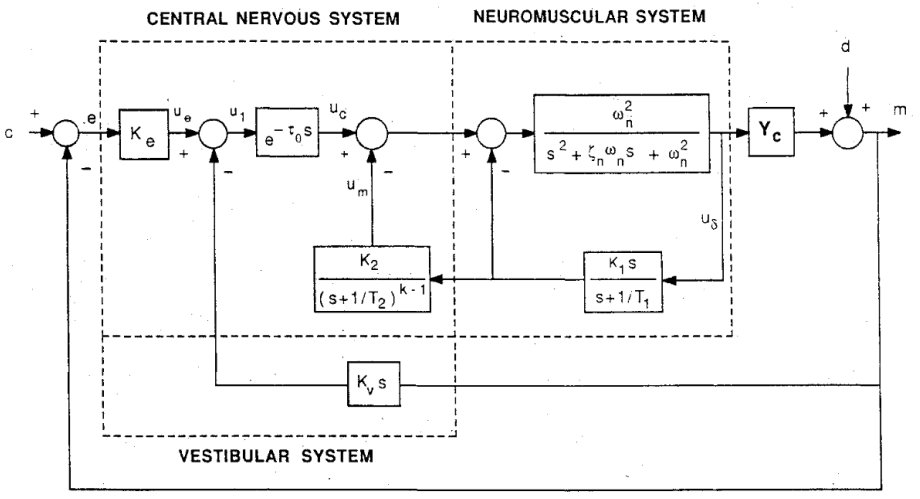


Figure 2-10 - Hess Structural Model of the Human Pilot Including Motion Cues – 1991 [51]

Many variations on the structural pilot model have been used in simulation fidelity research [52], [53], [54] with extensions, multi-loop capability and further parameters added where

necessary for the particular scenario to be modelled. As an example, Figure 2-11 shows the revised pilot model for modelling of Pilot Induced Oscillation (PIO) scenarios.

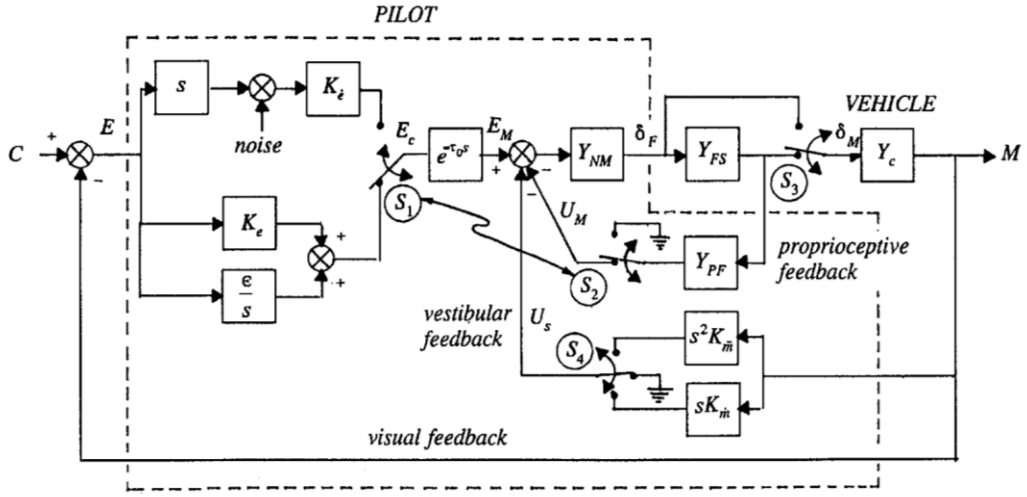


Fig. 1 Revised structural model of the human pilot.

Figure 2-11 - Hess Revised Structural Pilot Model – 1997 [52]

In 1992, Hess proposed a frequency domain metric based on the structural pilot model, named the Handling Qualities Sensitivity Function (HQSF). This metric is defined as;

$$\boldsymbol{HQSF} = \left(\frac{1}{K_e}\right)\left|\left(\frac{U_M}{C}\right)(j\omega)\right| \qquad \textbf{Equation 2-15}$$

where

$$\boldsymbol{U_M} = \frac{K_2}{\left(s+\frac{1}{T_2}\right)^{K-1}} \cdot \frac{K_1 s}{\left(s+\frac{1}{T_1}\right)} \qquad \textbf{Equation 2-16}$$

He proposed that the difference between the areas under the nominal flight vehicle HQSF curve and the simulated vehicle HQSF curve (see Figure 2-12) could provide a measure of fidelity. This metric has been used in several studies to assess the effect of visual cue quality on perceptual fidelity and has shown good correlation with pilot perception [53], [55], [56].

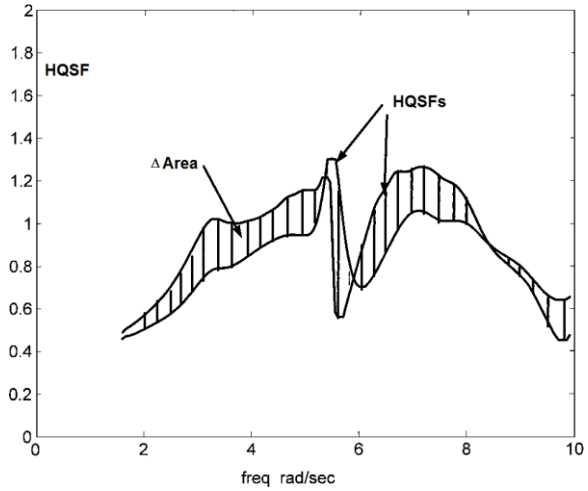


Figure 2-12 - Example Difference in Areas in Nominal and Simulated Vehicle HQSF Functions [55]

Results comparing the fidelity metric for lateral and vertical repositioning tasks showed a significant degradation in the fidelity metric (increased numerical value) for small motion compared to large motion as shown in Figure 2-13 [55]. This sensitivity confirms the possibility that the fidelity metric based on the difference between the HQSF in flight and simulation could be used as a predictive fidelity assessment for motion cue fidelity assessment. Research would be required to assess the acceptable/unacceptable boundary for the fidelity metric from cross reference with subjective opinion.

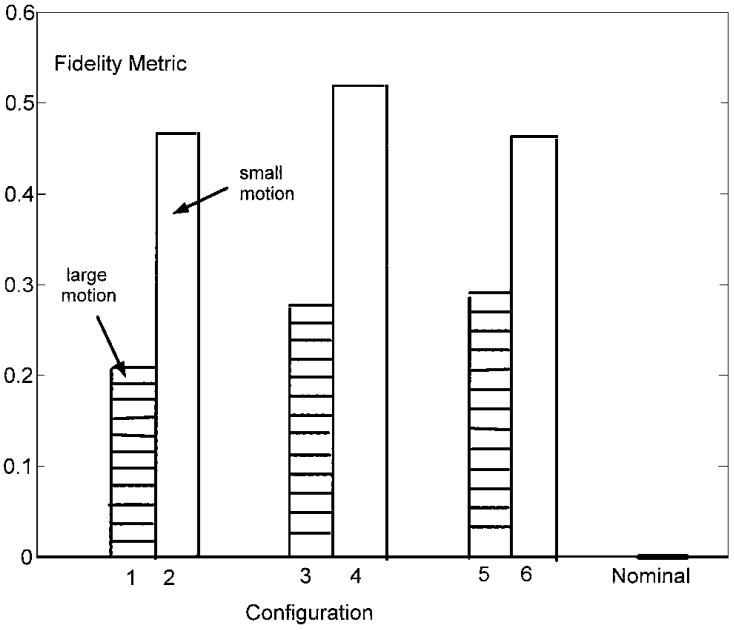


Figure 2-13 - Comparison of Fidelity Metrics for Small and Large Motion Gains [55]

Thompson and Bradley developed an algorithm for Helicopter Inverse Simulation (HELINV) [57], which utilised the analytical methodology described above to assess the control activity required for a Helicopter Generic Simulation (HGS) model to fly a pre-defined trajectory. The time history of the closed-loop control activity could then be extracted from the inverse

simulation. It was found that predictions of pilot workload from HELINV were consistent with piloted simulation trials [57]. Although the HELINV methodology has not as yet been used for simulation fidelity assessment, there is potential utility in comparing flight test control activity to the control activity generated by HELINV to obtain a predictive assessment of closed-loop task specific perceptual fidelity without the need for piloted simulation.

A similar toolbox, the Task-Pilot-Vehicle (TPV) model, was developed by Heffley in 2010. The pilot model was based on the structural pilot model but with a series of nested inner loops beginning with inner rate and attitude loops and then outer velocity and position loops [58]. The toolbox was developed such that any form of linear or nonlinear vehicle model can be used. Reference [58] provides validation data that shows that the TPV model is capable of realistic simulation of complex, multi-segment, tasks. Therefore preliminary investigations into task-specific perceptual fidelity could be conducted by comparing state and control activity from the TPV model with the flight test data before piloted simulation trials.

Numerous other methods for utilisation and development of pilot modelling techniques for fidelity assessment are present in the literature and require further attention than can be given here. Examples include the Synthesis through Constrained Simulation (SYCOS), developed by Bradley and Brindley [59], the Adaptive Pilot Model, developed by Padfield and White [60] and the Cybernetic Approach developed by Stroosma and Pool [54]. All of these methodologies have shown potential for flight simulation fidelity assessment. As of yet, no criteria have been developed regarding changes in pilot model parameters and the research results are only validated for specific cases.

2.2.4 Subjective Assessment

Quantitative analysis can only expose shortcomings in the areas of a system that are measured. Subjective opinion draws on the knowledge that an experienced user has of a system. This subjective opinion is important when assessing the performance/effectiveness of the pilot-vehicle system as the complex nature of the interaction between the pilot and the vehicle can be difficult to fully quantify. Furthermore, subjective ratings can help determine thresholds of acceptable and unacceptable differences in system parameters between flight and simulation.

In the current rotorcraft flight simulation qualification standards, EASA CS-FSTD(H), a standard check-ride flight is defined to assess the capabilities of the FSTD in a "typical training period" [20]. This includes take off, landings, auto-rotation and nap-of-the-earth flight. After this check-ride, the pilot is required to give a subjective assessment of the fidelity

of the simulation with respect to real-world flight experience. The criteria for Level D fidelity in EASA CS-FSTD (H) is:

"When evaluating functions and subjective tests, the fidelity of simulation required for the highest level of qualification should be very close to the helicopter. However, for the lower levels of qualification the degree of fidelity may be reduced in accordance with the criteria contained (in the document)" [14].

This requirement is ill-defined, and open to interpretation by the operator/qualifying body.

Research simulator subjective assessment is even less defined. As a result, subjective rating scales and questionnaires are developed 'in house' for specific studies. One example of such a scale was developed during a NASA programme to validate the General Purpose Airborne Simulator [61] shown in Table 2-1. This is a non-structured scale whereby the pilot chooses which descriptor he feels to best reflect his opinion and awards the correlating rating.

Table 2-1 - Simulation Pilot Rating Scale Developed by NASA for use in Validation Program [61]

Category	Rating	Adjective	Description
Satisfactory	1	Excellent	Virtually no discrepancies. Simulator reproduces actual vehicle characteristics to the best of my memory. Simulator results directly applicable to actual vehicle with high degree of confidence.
	2	Good	Very minor discrepancies. The simulator comes close to duplicating actual vehicle characteristics. Simulator results in most areas would be applicable to actual vehicle with confidence.
	3	Fair	Simulator is representative of actual vehicle. Some minor discrepancies are noticeable, but not distracting enough to mask primary characteristics. Simulator trends could be applied to actual vehicle.
Un-satisfactory	4	Poor	Simulator needs work. It has minor discrepancies which are annoying. Simulator would need some improvement before applying results directly to actual vehicle, but is useful for general handling qualities investigations for this class of vehicle.
	5	Bad	Simulator not representative. Discrepancies exist which prevent actual vehicle characteristics from being recognised. Results obtained here should be considered as unreliable.
	6	Very Bad	Possible simulator malfunction. Wrong sign, inoperative controls, other gross discrepancies prevent comparison from being attempted. No data

Another example of a custom designed fidelity rating scale is the Motion fidelity rating scale developed by Hodge et al at the University of Liverpool [38] (see Figure 2-14). This scale was designed with a similar structure to the HQR scale to determine pilot acceptability of motion cues. Pilots can award ratings in three coarse levels; (i) motion cues are acceptable, sensations are close to real flight or have insignificant deficiencies; (ii) motion cues are acceptable with some noticeable but not objectionable deficiencies; and (iii) motion cues are not acceptable [38]. Further descriptors then give more definition to the ratings. Although ratings obtained using this scale are a useful reflection of a pilot's perception of the realism of motion cues, no information of the fitness for purpose of the motion is gathered.

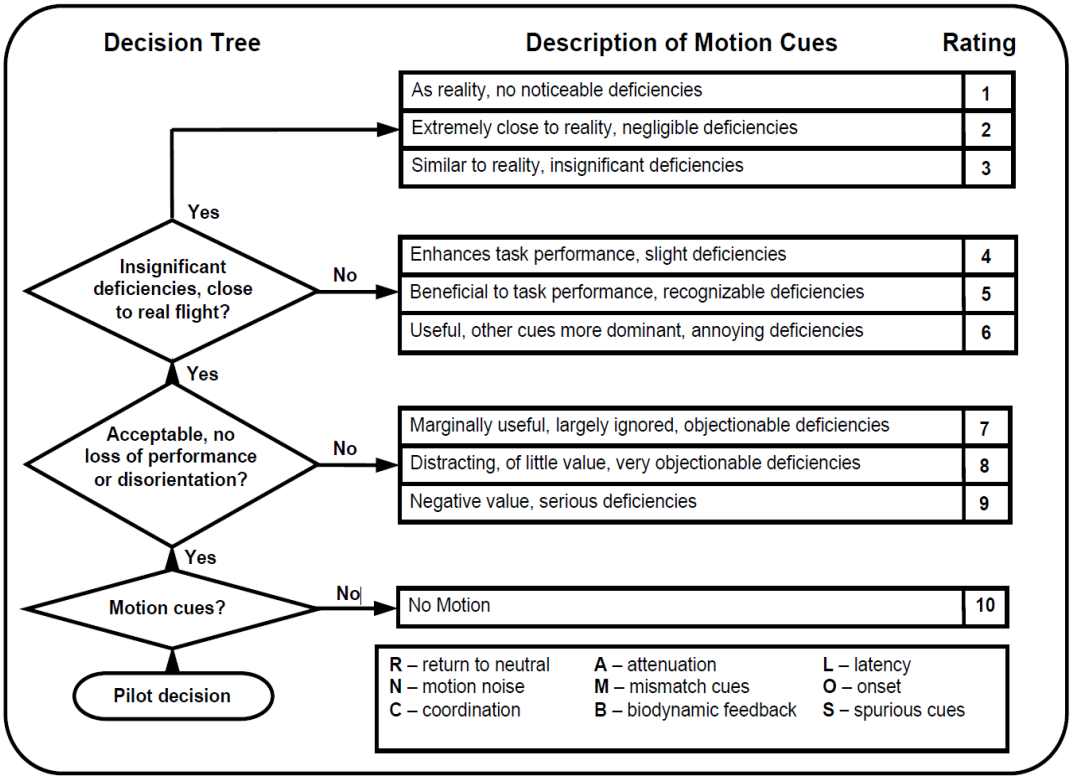


Figure 2-14 - The Motion Fidelity Rating Scale [38]

ADS-33E-PRF [19] includes a subjective Handling Qualities Rating (HQR) scale, developed by Cooper and Harper (Figure 2-15). Comparison of HQRs awarded in flight and simulation has been used in many studies to serve as a formalised assessment of perceived simulation fidelity, for examples see references [40], [51], [62].

Work conducted by the US Army in 1992 investigated the effects of time delays and motion cueing on handling qualities for seven rotorcraft operations: slalom, hover, quick-stop, bob-up, sidestep [4]. It was concluded that the presence of motion improved the handling

qualities ratings by 1/2 to 2 points. Without motion, level 1 handling qualities were not attainable. The variation in handling qualities suggests a poor fidelity match between the no motion and motion cases. This in turn suggests that a fixed-base simulator would exhibit poor fidelity compared to the aircraft for rotorcraft operations. The 1992 US Army report [4] also investigated the effect of varying motion washout and accelerations. It was found that, for precision manoeuvring, reduced washout and decreased motion gains gave improved HQRs. For aggressive manoeuvring it was found that higher onset accelerations with higher washouts was more desirable for aggressive manoeuvres. This shows that motion requirements are task specific and that tuned motion sets for specific tasks may be required.

In an experiment conducted at the NASA VMS facility [1], HQRs awarded by 4 pilots for varied vehicle vertical-response characteristics were significantly modified by changes in motion-cue fidelity although the pilots were able to attribute the changes in handling qualities to the degradation in motion cueing. Again, an impact of motion cueing on handling qualities suggests perceptual fidelity is dependent on motion cueing.

However, it should be noted here that, when using comparisons of HQRs as indicators of fidelity, at no point does the pilot make any comment on the fidelity of the simulation. This methodology assumes that a difference in HQs between flight and simulator is indicative of poor fidelity and similarly that similar HQs between flight and simulation is indicative of good fidelity.

Work at UoL [63] has shown that while dissimilar HQRs are a clear indication of poor fidelity, similar HQRs are necessary but not sufficient to imply high fidelity. This is not surprising when it is considered that a certain level of perceived pilot compensation may be needed for a number of different reasons. Therefore there are instances where the pilot may perceive they have compensated to the same extent in the simulator and the aircraft but in different ways. The HQR scale is not able to convey such scenarios. Comparisons of other workload/handling qualities ratings such as the Bedford Workload scale [64] have also been used in much the same way and are prone to the same lack of sensitivity to fidelity issues.

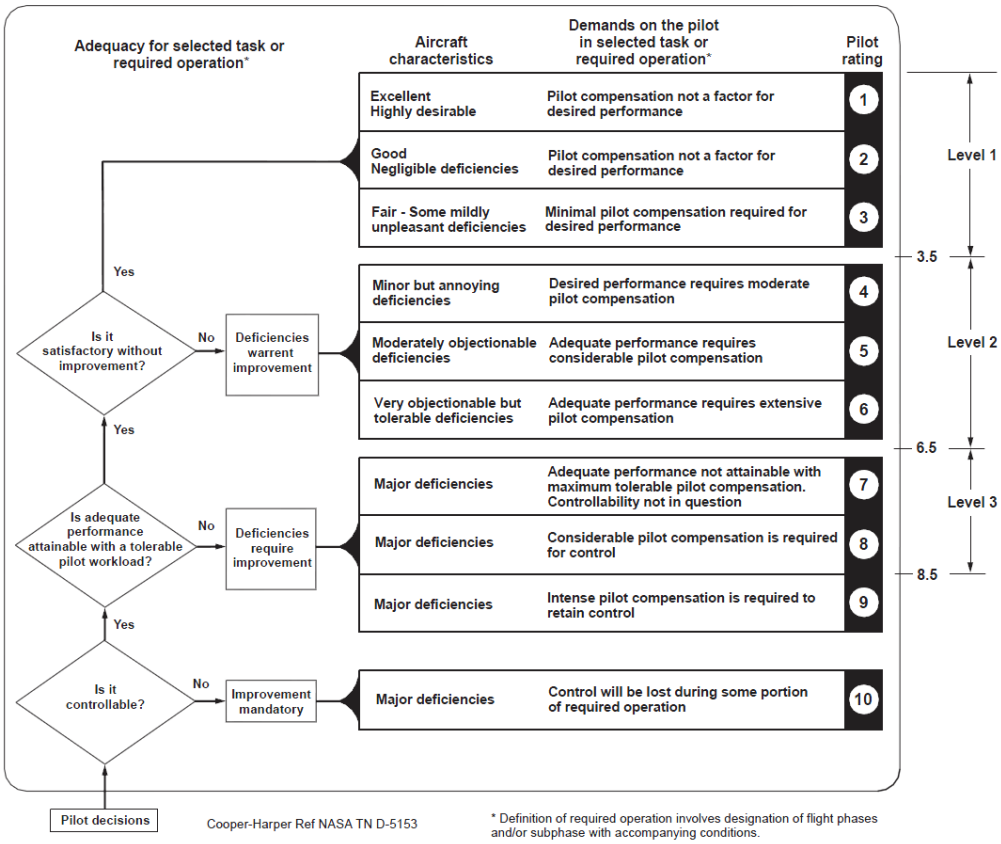


Figure 2-15 - Cooper-Harper Handling Qualities Rating (HQR) Scale [19]

ADS-33E-PRF also includes a subjective rating scale to determine the Useable Cue Environment for a given task. A UCE=1 reflects a Good Visual Environment (GVE) and a UCE=3 reflects a Degraded Visual Environment (DVE). Such ratings are used to determine the appropriate vehicle response type for the task. The subjective ratings used to determine the UCE are Visual Cue Ratings (VCRs). The VCR scale is shown in Figure 2-16. For each MTE to be assessed the pilot awards attitude VCRs for pitch roll and yaw and translational VCRs for vertical and horizontal translational rate. VCRs are taken from 3 pilots. The worst attitude rating and the worst translational rate rating are taken and plotted on the UCE chart (Figure 2-17) to obtain a UCE value from each pilot. The largest UCE value is taken as the UCE for the task. Key proposed that comparing the real-world UCEs with simulated cues, or SIMulator Day UCE (SIMDUCE), could provide a quantitative measure of cueing fidelity [25]. However, the same concerns arise as with comparing HQRs in that the same VCR can be given for a number of different reasons.

The review of available subjective assessment techniques suggests that there is a need for the development of a subjective rating scale for the formalised subjective assessment of flight simulation fidelity. This would remove ambiguity and interpretation from the current

methodology and hence provide more general understanding of the subjective assessment results.

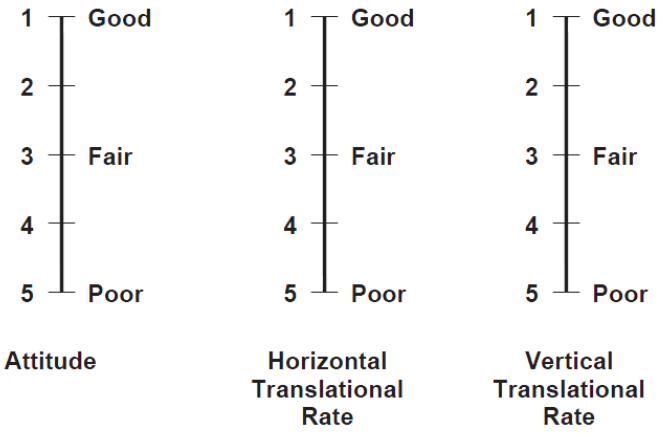


Figure 2-16 - The Visual Cue Rating Scale [19]

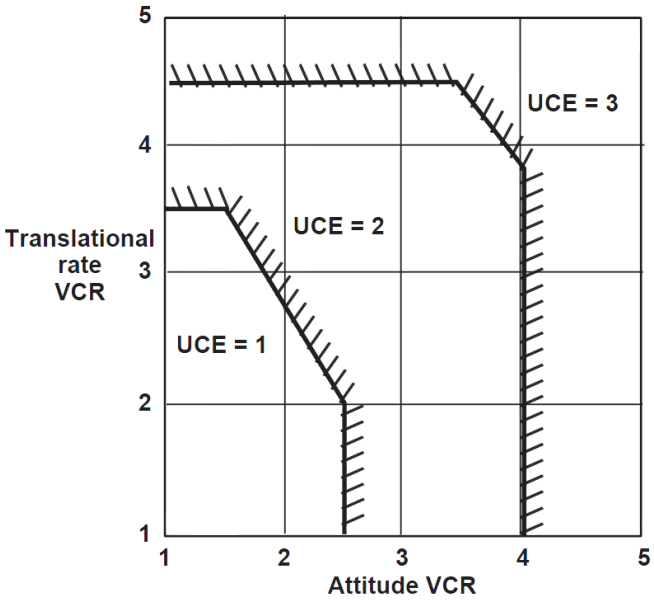


Figure 2-17 - Useable Cue Environments for Visual Cue Ratings [19]

2.3 Methods for Assessing Simulator Training Effectiveness

The utility, or fitness-for-purpose, of a system is often measured in terms of its capability of producing the desired result, i.e. its effectiveness. A commonly used measure of the utility or effectiveness of a simulator is to ascertain the number of flight hours that can be replaced by a simulator thereby quantifying the financial saving realised through training in the FTD.

Roscoe defined Training Effectiveness Ratio (TER) to be a quantitative measure of the time (and therefore cost) savings realised in learning to operate an aircraft by training in an FTD [65]. To calculate the training effectiveness, two participant groups are required; an experimental group and a control group. The experimental group receives training in the FTD. The control group receives no prior training. The compared parameter is either the time or number of trials required to reach a defined level of performance. Alternatively, the error count while performing a task can be used to quantify transfer.

Using time as an example, Cumulative Transfer Effectiveness Function (CTEF) is the ratio of the savings in live training time required to the number of hours trained in the FTD.

$$\boldsymbol{CTEF = \frac{Y_o - Y_x}{Tx}} \qquad \textbf{Equation 2-17}$$

Y_0=Aircraft training time required for control group,

Y_x = Aircraft training time required for experimental group after FTD training,

Tx = FTD training time.

CTER<1 suggests either more hours in the simulator are required than in the aircraft, or that the pilot requires further training in the aircraft after simulator training.

CTER=1 suggests full transfer of training, i.e. no further aircraft time is required for the control group after training and the same number of hours are required in the simulator as in the aircraft.

CTER>1 suggests that the FTD delivers more efficient training than the aircraft itself. This is possible for specific tasks as the pilot may not be required to perform time consuming procedures and can therefore learn faster.

$$\boldsymbol{ITEF = \frac{Y_{x-\Delta x} - Y_x}{\Delta Tx}} \qquad \textbf{Equation 2-18}$$

Y_0-Y_x= time, trials, or error saving

ΔTx = incremental unit in time, trials, or errors.

Y_{x-} = time, trials, or error required by a transfer group to reach performance criterion, after X-ΔX training units.

This methodology has been utilised for a number of rotary wing transfer of training studies [66]. A meta-analysis of 19 fixed and rotary wing studies is described in reference [66]. This study highlighted the drawbacks to the use of transfer of training studies for the quantification of training effectiveness. These include financial and time requirements and the inability of the CTEF and ITEF to correctly model the effect of prolonged training in an FTD that delivers negative transfer. The functions suggest that prolonged exposure in such a device would reduce the negativity of the training (just as the function models diminishing returns of transfer as simulation time increases) due to the symmetry of the function. The CTER also does not correctly model the effects of over training

2.4 Conclusions of Technical Review

This review served to describe the variety of approaches that have been developed for flight simulation fidelity assessment and the current position in terms of readiness of application for general use. The following findings are highlighted from this review as the most significant:

1. The quantitative mathematical model Proof of Match (PoM) criteria in the current training simulator certification specifications, EASA CS-FSTD (H), lack supporting evidence for their validity for qualifying rotary wing simulators. Findings from previous research suggests that, in particular, the transport delay and off-axis response requirements warrant further investigation.

2. Handling qualities metrics, such as inter-axis coupling and bandwidth, as well as flight dynamics allowable error envelopes, have been proposed by many for use in predicted fidelity assessment. While such metrics have shown potential, no boundaries of acceptable/unacceptable differences have been defined.

3. EASA CS-FSTD (H) does not consider that predicted fidelity requirements are dependent on the task to be trained. The development of the ICAO 9625 documents highlighted this need but it was beyond the scope of the IWG to define tolerances of any metrics.

4. Methods for quantifying pilot behaviour are extensive, ranging from direct analysis of pilot control activity to measurement of physiological parameters such as heart rate and pupil dilation. Comparing this behaviour between flight and simulation has shown potential for perceptual fidelity assessment. Although several studies have used

comparative pilot behaviour measures to highlight differences, no acceptable/unacceptable boundaries have been identified.

5. Analytical models of the human pilot have been utilised in a number of research studies examining the difference between the pilot-aircraft system and the pilot-simulator system. Pilot-vehicle model characteristics such as crossover frequency and pilot gain have been used as quantitative measures of this similarity as well as comparisons of metrics such as the HQSF. Although numerous individual studies suggest such metrics have good potential there is little evidence for their generalised applicability.

6. The current requirements for subjective testing are ambiguous. A number of institutions use comparative ratings of workload (HQRs) and visual cueing quality (UCEs). However, these are not deemed to be appropriate for simulation fidelity assessment as a match is achievable when fidelity deficiencies are present. A number of rating scales to quantify the pilot's perception of fidelity have been developed but none have gained formal acceptance for the evaluation of simulator fidelity.

7. Transfer of training studies have been extensively used to quantify the utility of a simulator for training. However, such studies have been criticised for not correctly capturing effects of negative training or overtraining. Furthermore, such methods are resource demanding.

2.5 Research Aim and Objectives

The research described throughout this thesis is intended to support the on-going work of Lifting Standards project, which aims to establish a new approach to predicted and perceptual fidelity assessment. In light of the main findings of the technical review, the primary aim of the research is determined as the development of industry applicable metrics and methods for the assessment of the perceptual fidelity of a fully integrated simulation. The objectives addressed to meet this aim are listed below with a brief summary of how these objectives were realised.

1. **Gather simulation and flight test data to support perceptual fidelity metric development.**

A number of flight and simulation test campaigns will be planned and conducted, at the University of Liverpool and NRC, to ensure all the data required to achieve the other objectives is available for analysis. Testing methodologies are to be designed to mitigate against the effects of pilot learning, anticipation and subjective bias.

2. **Quantify the predicted fidelity of the Bell 412 FLIGHTLAB model to act as a baseline for the perceptual fidelity work.**

ADS-33E-PRF and EASA CS-FSTD (H) metrics will be computed using open-loop aircraft and simulator model responses. These metrics will be used to compare the aircraft and simulation models dynamic behaviour to benchmark the predicted fidelity of the Bell 412 model.

3. **Develop methods for obtaining subjective feedback on perceptual fidelity.**

In response to the lack of a formal procedure for capturing pilot opinion of fidelity (Conclusion 6 of technical review), a subjective rating scale and questionnaire will be developed. These tools will be exercised and developed through piloted simulation and flight trials. Industry workshops are to be held to ensure best utilisation of the subjective scale and questionnaire by industry.

4. **Determine effect of cross-coupling and transport delay errors on perceptual fidelity.**

To address the need identified in Conclusion 1 of the technical review, the newly developed subjective assessment methodology will be used to determine pilot sensitivities to changes in inter-axis coupling and transport delay in a number of closed-loop manoeuvres.

Devise a methodology for defining allowable errors in predictive fidelity criteria such as those in EASA CS-FSTD (H).

Conclusion 1 of the technical review noted a requirement to obtain supporting data for quantitative tolerances. Therefore, the potential for utilising subjective ratings to defining the allowable errors will be examined using the results from the piloted simulation and flight trials.

5. Progress the development of quantitative metrics of perceptual fidelity.

Conclusion 4 of the technical review highlighted a need for development of measures of pilot behaviour. Control attack metrics determined from pilot control activity will be used as a starting point for quantitative perceptual fidelity metrics. If necessary, these metrics will then be further developed to improve correlation with pilot subjective opinion and to better reflect workload and task strategy adaptation in multi-axis flying tasks.

6. Determine the dependencies of perceptual fidelity tolerances.

The utility of pilot modelling techniques was identified in the technical review (Conclusion 5). These methods will be used alongside the control attack metrics to determine in what ways simulation fidelity might be dependent on parameters such as task, pilot strategy and vehicle handling qualities. Gathering evidence of task dependency is of particular importance to encourage the support of the use of training-need based simulator qualification as proposed by ICAO 9625 (Conclusion 3).

3 EXPERIMENTAL FACILITIES

The work in this thesis has been supported by a number of simulation trials using the HELIFLIGHT-R simulator at the University of Liverpool as well as a flight test campaign using the Canadian National Research Council (NRC) Bell 412 Advanced Systems Research Aircraft (ASRA). This short chapter introduces the Bell 412 ASRA and the HELIFLIGHT-R flight simulator used for real-time piloted simulation. The PoM between the UoL mathematical model and Flight test data is included along with discussion of model tuning and rotor modelling effects on predicted fidelity.

3.1 Bell 412 ASRA

The NRC Flight Research Laboratory's Bell 412 ASRA [67] (Figure 3-1) was used as the reference aircraft for the work in this thesis. The Bell 412 is a twin engine, medium weight utility helicopter with a dual seat, side-by-side crew station. In order to allow the ASRA to act as an in-flight simulator, the NRC have made a number of modifications to a standard Bell 412 aircraft. The ASRA has been equipped with instrumentation which allows measurements of attitudes, accelerations, forces, control activity, Global Positioning System (GPS) location, rotor and engine parameters as well as aerodynamic data. All of these data are recorded on board the aircraft for post-flight analysis. The aircraft has also had a full authority digital Fly-By-Wire (FBW) control system installed. This permits the ASRA to be operated with various control laws and stability characteristics that can be altered during flight. The FBW system contains safety trip points (Table 3-1) that cause the experimental fly-by-wire system to disengage and control to be reverted to the safety pilot, who flies the standard mechanical control system with partial authority stability augmentation There are three different Stability and Control Augmentation System (SCAS) types that are currently implemented in the ASRA; an Attitude command Attitude Hold (ACAH), Rate Command Attitude Hold (RCAH) and Rate Damped. The aircraft can also be flown without any SCAS in the nominal 'bare airframe' configuration.

In 2005, a flight test campaign was conducted as part of the HELI-ACT (Helicopter Active Control Technology) project at UoL [68] to collect flight test data for the Bell 412 ASRA in trim conditions and in response to open loop control inputs. As part of the Lifting Standards project there were two more flight test campaigns conducted in 2009 and 2011 respectively. The first of these trials [69] obtained Bell 412 flight test data from a number of MTEs in the bare airframe and ACAH configurations as well as open loop control response tests for the

ACAH configuration. Open-loop and MTE tests for the RCAH and rate damped systems were conducted in the second of the Lifting Standards flight test campaigns.

Figure 3-1 - NRC's Bell 412 Advanced Systems Research Aircraft (ASRA)

Table 3-1 - ASRA Safety Limits [67]

Parameter	Limitation
Torque	Below 105kts: 92% Mast Torque Above 105kts: 85% Mast Torque
Roll Attitude/Rate (above 25 ft)	Above 45 kts: ±65°, ±60°/s 30 – 45 kts: ±45°, ±40°/s Below 30 kts: ±35°, ±35°/s
Roll Attitude/Rate (below 25 ft)	Above 45 kts: ±45°, ±60°/s 30 – 45 kts: ±35°, ±40°/s Below 30 kts: ±25°, ±35°/s
Pitch Attitude/Rate (above 25 ft)	All speeds: ±32°, ±25°/s
Pitch Attitude/Rate (below 25 ft)	Above 30 kts: ±25°, ±25°/s Below 30 kts: ±15°, ±25°/s
Yaw Rate	Above 45 kts: ±25°/s 30 – 45 kts: ±30°/s Below 30 kts: ±40°/s ±10°/s when height is < 10ft

3.2 HELIFLIGHT-R

HELIFLIGHT-R [69] (Figure 3-2) is one of the University of Liverpool's in house research simulators. It was manufactured by Advanced Rotorcraft Technology (ART) and utilises the ART FLIGHTLAB software [70]. The simulator is configured with a dual seat, side-by-side crew station with an on-board Instructor/Operator Station (IOS). However, there is capability to remove this cockpit and replace it with any other crew station configuration. HELIFLIGHT-R also has local and wide area networking capabilities which allow it to be connected to other simulators at the University of Liverpool and also allows for networking with external clients. The system utilises a general purpose Linux based computer for the vehicle model, which is connected to an external, windows based PC. This PC runs software developed at UoL called the Liverpool Virtual Environment (LIVE). LIVE is the operator's interface with the simulation environment and all the components of the simulation; motion, visual, controls, audio and flight model etc. are connected via a local area network. The architecture of the local network is shown in Figure 3-3. Specifications of each of the elements of the HELIFLIGHT-R simulator are detailed below with focus on the impact of the system characteristics on the utility of the simulation for the research described in this thesis.

Figure 3-2 - HELIFLIGHT-R, UoL Research Simulator

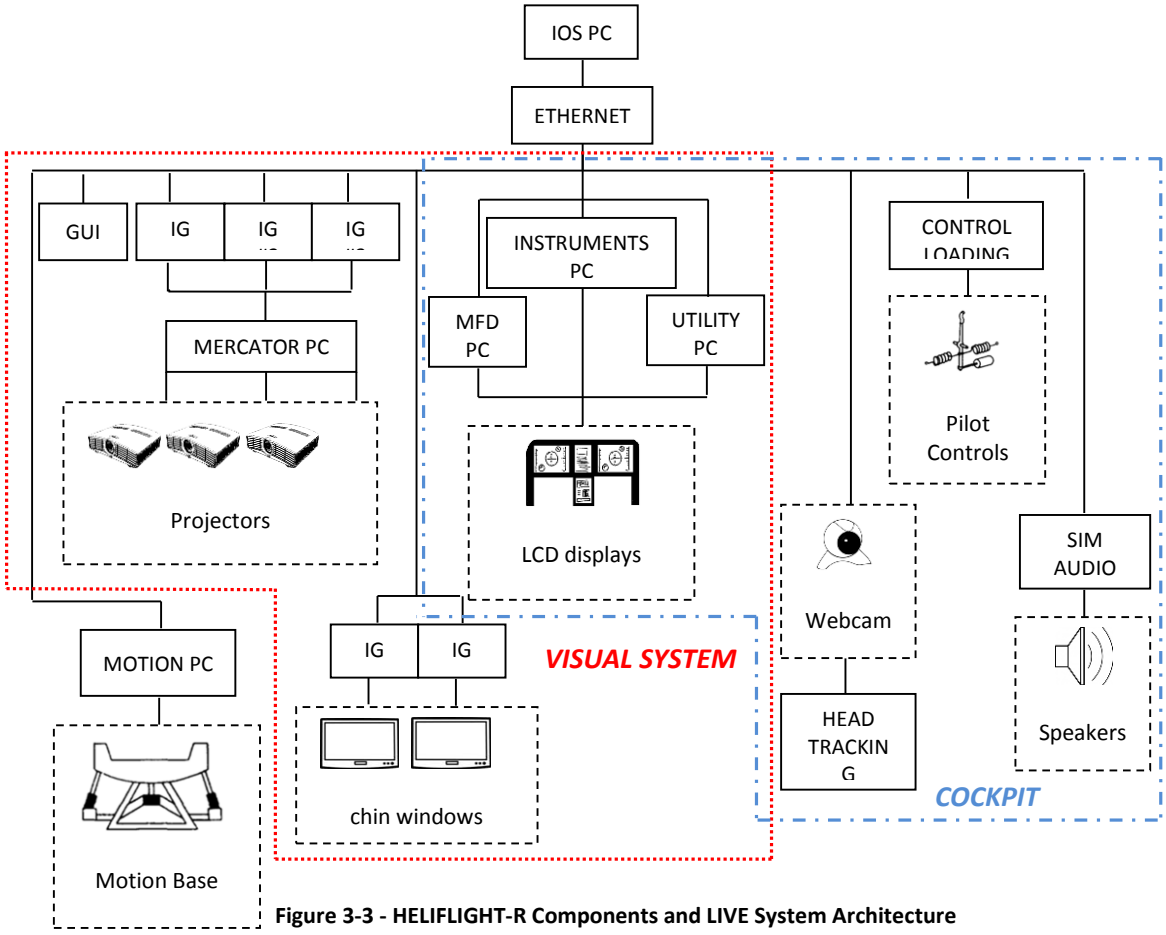


Figure 3-3 - HELIFLIGHT-R Components and LIVE System Architecture

3.2.1 Motion Base

HELIFLIGHT-R utilises a Moog MB/E/6dof/18000kg six DoF electric motion hexapod (Figure 3-2). The motion base capabilities are plotted against EASA CS-FSTD (H) criteria for Level A, B, C and D simulator qualification [14] in Figure 3-4. Each actuator has a 600mm stroke, giving peak accelerations of >300°/s2 in each of the three rotational axes (Figure 3-4a), 0.71g in surge and sway, and 1.02g in heave (Figure 3-4b), although it should be noted that there was no expectation for the capability to meet the Level D criteria due to the short stroke actuators. The motion washout filter settings can be reconfigured to improve pilot perceived motion fidelity for different vehicle types and flight regimes. Previous research [5], [37] has shown that pilot workload and task performance in low level manoeuvring rotorcraft flight is significantly affected by the presence of vestibular cueing. Therefore, the use of the motion platform was considered necessary for this research.

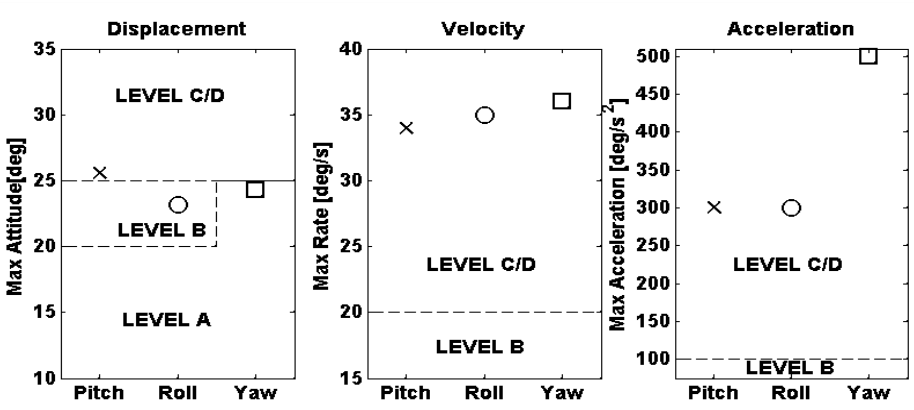


Figure 3-4a - HELIFLIGHT-R Rotational Motion Capabilities in Comparison to EASA CS-FSTD (H) Criteria

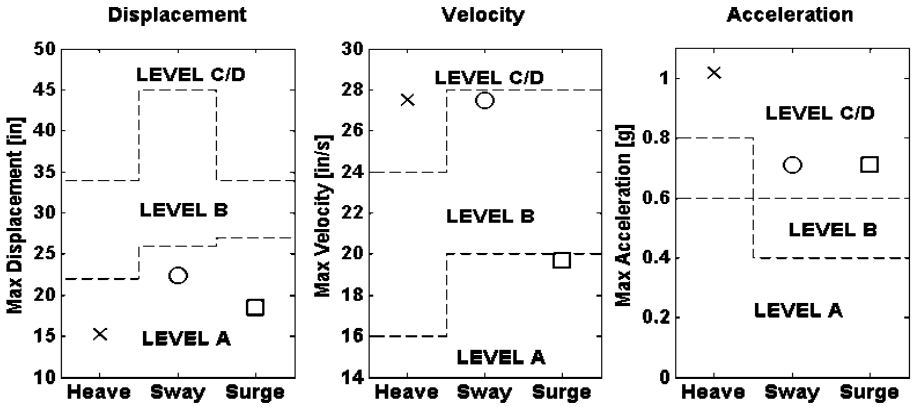


Figure 3-4b - HELIFLIGHT-R Translational Motion Capabilities in Comparison to EASA CS-FSTD (H) Criteria

3.2.2 Visual System

i. Image Generator PCs and Projectors

The Out-The-Window (OTW) visual environment is generated on three Windows PC Image Generators (IGs) using Vega Prime, a Commercial Off The Shelf (COTS) visualisation tool from Presagis. However, it is possible to integrate the FLIGHTLAB software with other run time software such as Landscape [71], X-Plane and Microsoft Flight Simulator run-time environments. The resulting OTW scene then passes through a Mercator IV pixel management system to ensure the projected display is correctly blended and distributed. As part of the early Lifting Standards work, the NRC's handling qualities test course in Ottawa was replicated in Landscape, and later into Vega, using GPS mapping of the test course cones and photographs of the surrounding area. This has improved the place illusion in this research as all the test pilots were familiar with the NRC test course.

The OTW scene is displayed to the pilot via three Canon XEED SX60 Liquid Crystal on Silicon (LCoS) HD projectors onto a 12ft carbon fibre dome (Figure 3-2) to create an uninterrupted 210x65 degree field of view with a 1400 x 1050 video resolution. The comparison between the Bell 412 ASRA and HELIFLIGHT-R operational FoV (limited by instrument panel and aircraft framework) can be seen in Figure 3-5. A curved projection system is the only way to obtain a FoV comparable to that available in the Bell 412 aircraft. The main limitation of the HELIFLIGHT-R simulator is that there is no airframe to obstruct this FoV. Pilots commented that they utilised the airframe as a position and translational rate cue. In the simulator, some pilots were seen to adapt their visual strategy by utilising the edge of the FoV to obtain similar cues.

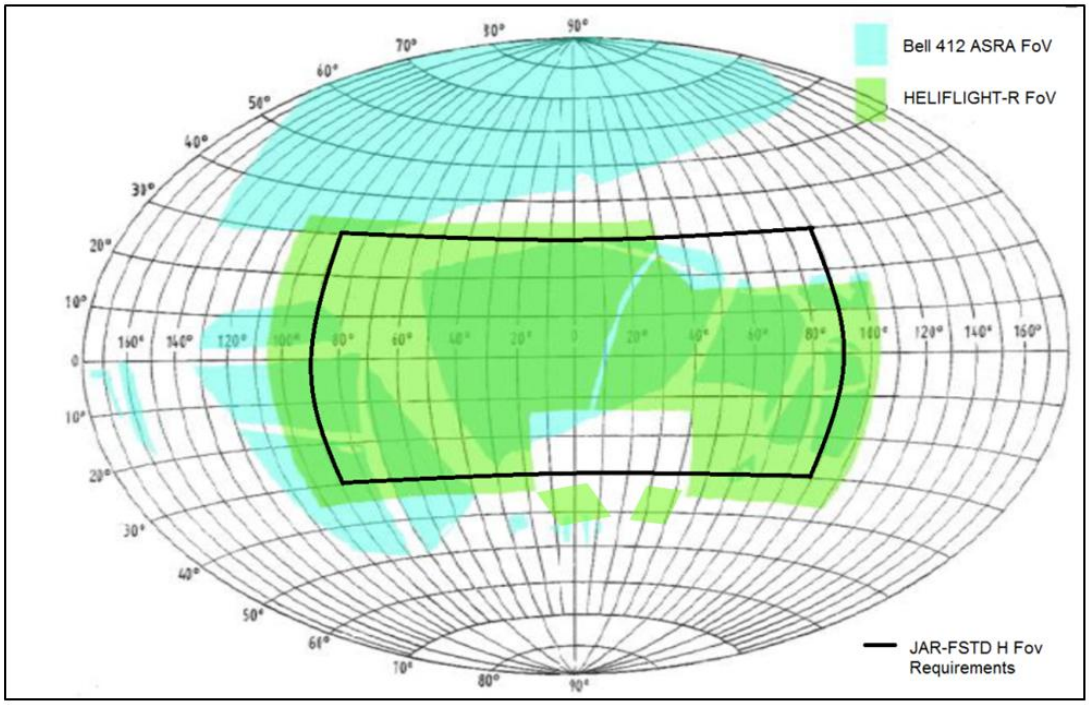


Figure 3-5 - HELIFLIGHT-R and Bell 412 ASRA FoVs from Left Hand Seat with JAR-FSTD H FoV Requirements

ii. Environmental PC

A Graphical User Interface (GUI), shown in Figure 3-6, has been generated at UoL that allows real-time control of the VEGA Prime environment configuration. The image generator PCs pick up data packets from the GUI machine over the Ethernet and use these data packets to set the correct environmental conditions. This allows visual effects such as rain, cloud, time of day, ground effect washout etc. to be added to the OTW visuals.

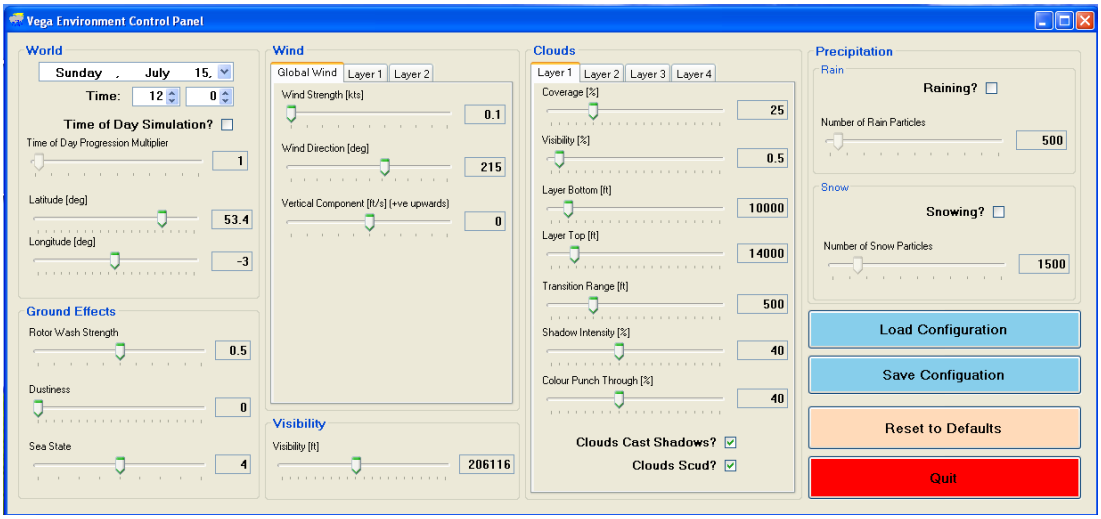


Figure 3-6 - GUI for VEGA Environment Control

iii. Chin windows

Chin windows were installed into HELIFLIGHT-R in 2012 to improve the FoV in response to pilot comments of poor translational velocity cues in precision tasks at low altitude. The chin windows are powered by two separate Windows PC Image Generators that feed directly to Liquid Crystal Display (LCD) flat screen monitors in the foot wells of the simulator cockpit. Upon installation of the chin windows, it was found that they had an impact on pilot VCRs due to improved translational rate cues for precision, low level manoeuvring.

iv. LCD Displays: Instruments, MFDs and Utility Screens

Inside the crew station are four LCD flat panel touch screen displays (Figure 3-7). The left and right screens display the primary flight control instruments. The centre two screens are MFD touch screen and can be used for pilot questionnaires, secondary instruments, task descriptions etc. The three PCs that control the LCDs can be used to allow up to three different screens to be displayed simultaneously.

HELIFLIGHT-R came with a generic helicopter control panel, provided by ART (Figure 3-8). A replica of the Bell 412 instrument panel has been generated (Figure 3-9) using VAPS XT, another COTS product from Presagis. Glass cockpit representation of instruments was deemed acceptable for this research as the work was focused on flying skills not instrument tasks. Pilots commented that they were able to receive the same information from the same place.

Figure 3-7 - HELIFLIGHT-R (Complete Visual System)

Figure 3-8 - Generic Instrument Panel

Figure 3-9 - Bell 412 Instrument Panel Generated in VAPS XT

3.2.3 Inceptors and Control Loading

There are two sets of controls in the HELIFLIGHT-R cockpit and each set of controls consists of three control inceptors; a cyclic stick, a collective and pedals. A Moog Flight Control System (FCS) ECol 8000 four axis control loading system provides a force feedback system through these controls and also allows the pilot control inputs to be back driven into the FLIGHTLAB model to allow for aircraft trim, control augmentation and autopilot modelling. There is an overhead panel with a number of switches that can be reconfigured to interconnect with FLIGHTLAB model variables for in-flight alteration and there is also capability to integrate other inceptors such as throttles, yokes and side-sticks. However, such inceptors currently have no force-feedback applied to them.

The system was delivered with a generic control model already implemented. The Flight Control Mechanical Characteristics (FCMC) settings were tuned to represent the Bell 412 for this work according to pilot subjective opinion due to the lack of FCMC data for flight.

3.2.4 Data Recording and Monitoring

From the control room monitoring station (Figure 3-10). The real-time simulation can be monitored via a number of screens. Three screens (top tier) show the OTW view, another shows an external view of the helicopter (lower right). Two further monitors show a view from the webcam on the pilot (2nd from the right, lower tier) and allow the active instrument panel can be monitored on the (lower left). The final two monitors are for control of the real time simulation environment and monitoring of the vehicle states and pilot control activity in real time.

The vehicle states and pilot control activity can be recorded in discrete, user defined files and audio and central OTW view can be recorded to DVD. The webcam image can also be recorded for eye and head tracking analysis.

Figure 3-10 - Control Room Monitoring Station

3.2.5 Flight Model - FLIGHTLAB

The flight model has been implemented using FLIGHTLAB [70], a COTS modelling, analysis and real-time simulation tool from ART. FLIGHTLAB is currently one of the leading commercial software tools for rotorcraft modelling and analysis. The software has an open architecture which allows the user to customise model components, GUIs and analysis tools to meet user specific requirements. The standard GUIs allow for simple generation of models, control systems and analysis. The main interface functions are described below:

- FLIGHTLAB Model Editor (FLME) - A GUI that allows the user to develop a flight dynamics model through a library of model components (see Figure 3-11). The model components are then interconnected using an architecture which also allows for user created custom

components. Values can then be assigned to parameters that specify each model component to represent a specific vehicle. This was essential for the current research as the flight model dynamics could be easily modified to assess their impact on simulation fidelity.

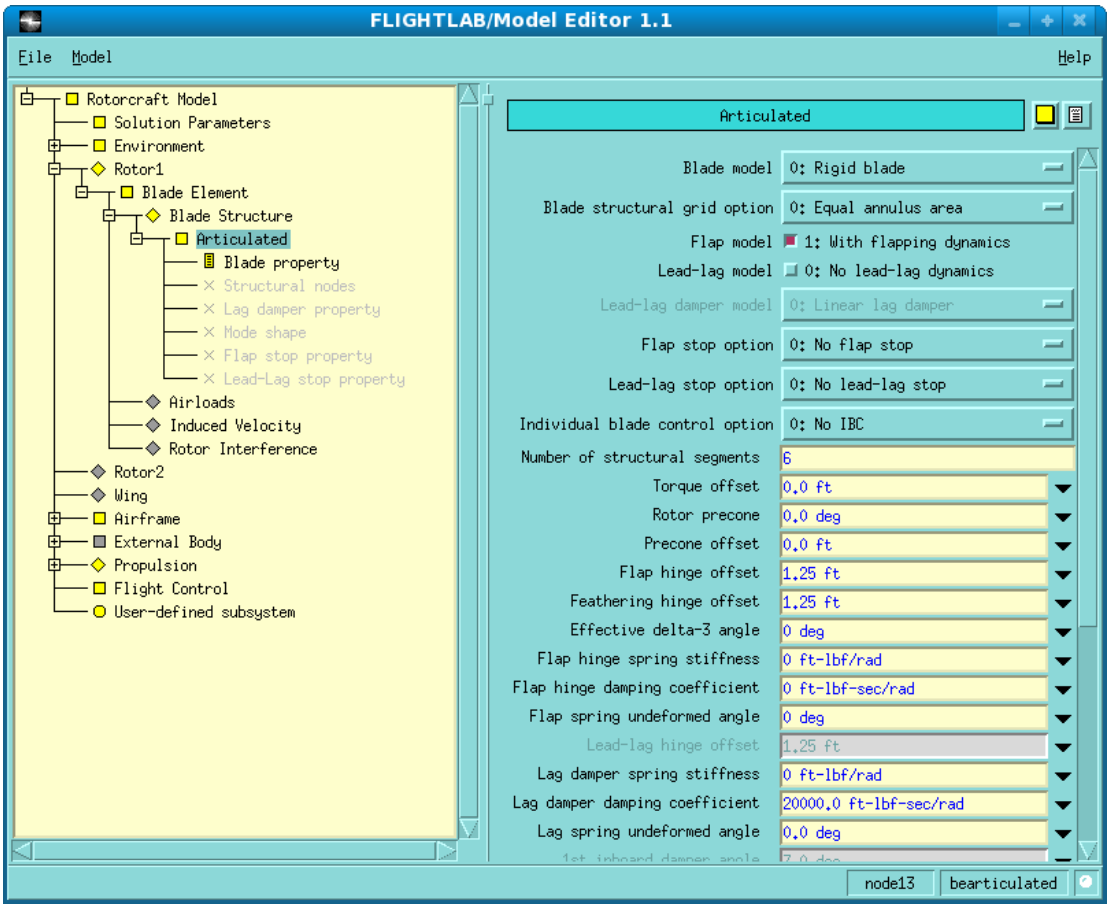


Figure 3-11 - FLIGHTLAB Model Editor (FLME) Graphical User Interface (GUI)

- Control System Graphical Editor (CSGE) - A GUI For the design of the aircraft's SCAS, as shown in Figure 3-12. CSGE includes a library of control system element blocks which can be interconnected in the CSGE GUI in a schematic view (Figure 3-12). Parameters for each control element can be defined and attached to model components in FLME to fully replicate the FBW SCAS of the real aircraft. This allowed simple application of the NRC's FBW SCAS systems into the models used in this research.

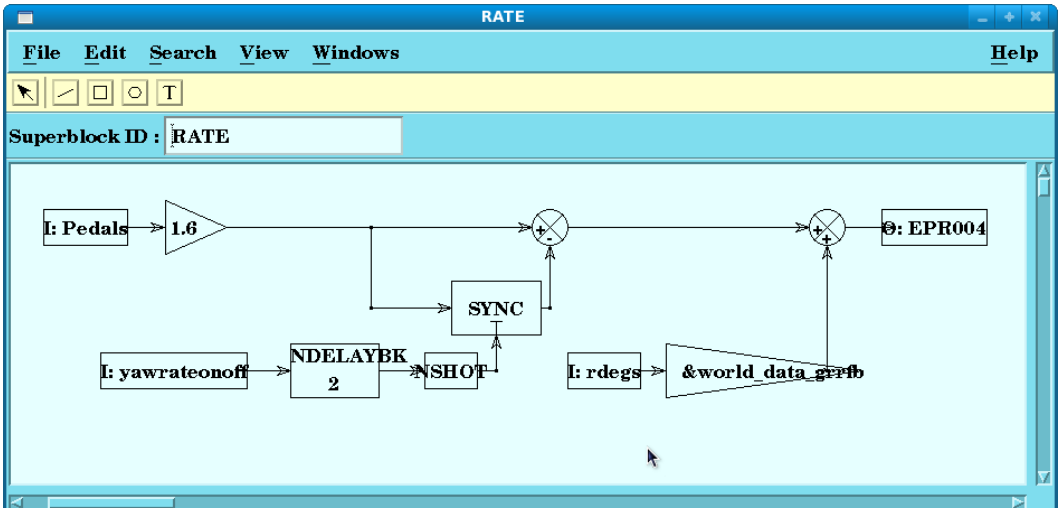


Figure 3-12 - Control Systems Graphical Editor (CSGE) GUI

- Xanalysis - A GUI that allows the user to trim the model to a desired condition, perform a linearisation of the model for stability analysis and carry out investigations of the aircraft linear and nonlinear dynamic response to a predefined control inputs (Figure 3-13 and Figure 3-14). This functionality was heavily utilised in the research to generate time histories of nonlinear dynamic response to control inputs for comparison with flight as well as for handling qualities assessments.

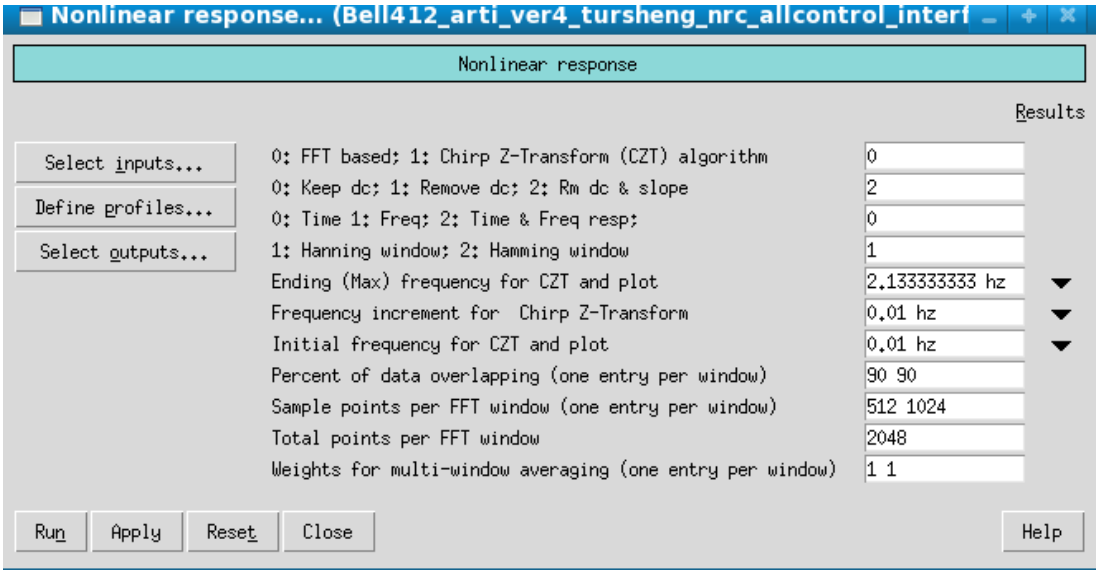


Figure 3-13 - Xanalysis Nonlinear Response Setup GUI

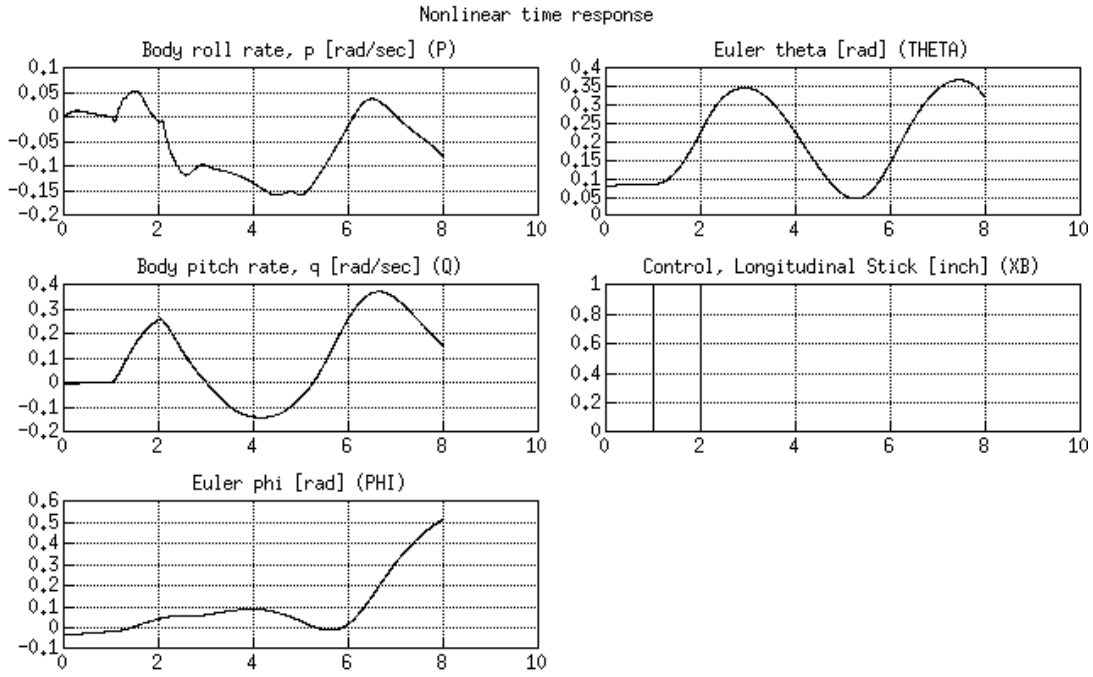


Figure 3-14 - Example of an Xanalysis Nonlinear Response GUI

3.3 Bell 412 ASRA FLIGHTLAB Model

A 90 state FLIGHTLAB flight dynamics model of the Bell 412 ASRA was developed by Manimala et al at UoL as part of the HELI-ACT project [68]. The hingeless main rotor was modelled using articulated rigid blades with an equivalent offset hinge and spring. A Bailey rotor was used to model the tail rotor. A Peters-He finite state dynamic wake model was used to model the inflow to the rotor and generate the rotor interference model that was used to model aerodynamic interferences between the rotor wake and the empennage and fuselage. The Peters-He model uses 4 inflow harmonics and 4^{th} power radial variation – together these provide an interference model with 15 states. The fuselage was modelled using look up tables of forces and moments acting on the fuselage for combinations of angle of attack and sideslip. The data for these tables were obtained from various sources [72], [73]. The inverted Clark-Y aerofoil data for the horizontal stabiliser were available and so implemented into the FLIGHTLAB model. The aerofoil segments used on the Bell 412 vertical fin were not known and so the UH-60 aerofoil (NACA 0021) was used as a substitute. Due to the proprietary nature of the Bell 412 blade aerofoil sections, the main rotor model included aerofoil data from a VR7 aerofoil for the inner blade segments and an 0A309 aerofoil for the outer blade segments. Finally, a simple engine model was used due to lack of engine data.

The moments of inertia and control gearings of the aircraft were tuned through trial and error to improve the match between dynamic response of the FLIGHTLAB model and data obtained from the Bell 412 ASRA during flight testing [68]. It was found that the model matched on-axis responses of flight well across a range of speeds. However the off-axis responses were less well matched. It was suggested that such deficiencies may be improved by increasing the complexity of the wake distortion model and that more accurate engine and blade models may improve the model further.

During the Lifting Standards research [69], the Attitude Command Attitude Hold (ACAH) and Rate Command Attitude Hold (RCAH) SCAS systems were modelled using CSGE and data supplied by the NRC on the control laws. The control systems were then wrapped around the FB412_HA bare airframe model developed by Manimala et al for the functioning SCAS models.

The RCAH model has acted as a reference model for much of the research reported in this thesis. Therefore, in the early stages of this research, data for validation of the RCAH model were collected. Only hover control response data have been collected to date. The comparisons of flight and mathematical model open-loop responses are shown in Figure 3-15 to Figure 3-18 against the JAR-FSTD H Low Airspeed HQ requirements (section 2.2b). The oscillations in the pitch and roll rates were attributed to noise in the ASRA's instrumentation rather than vehicle dynamics.

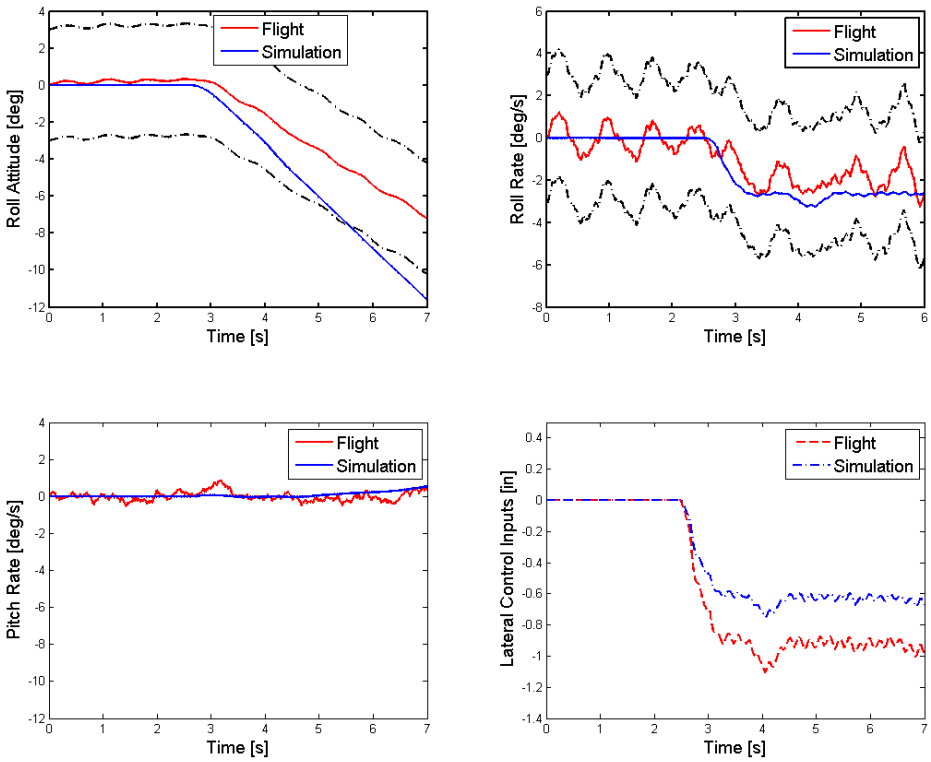


Figure 3-15 - RCAH Lateral Step Response - Hover

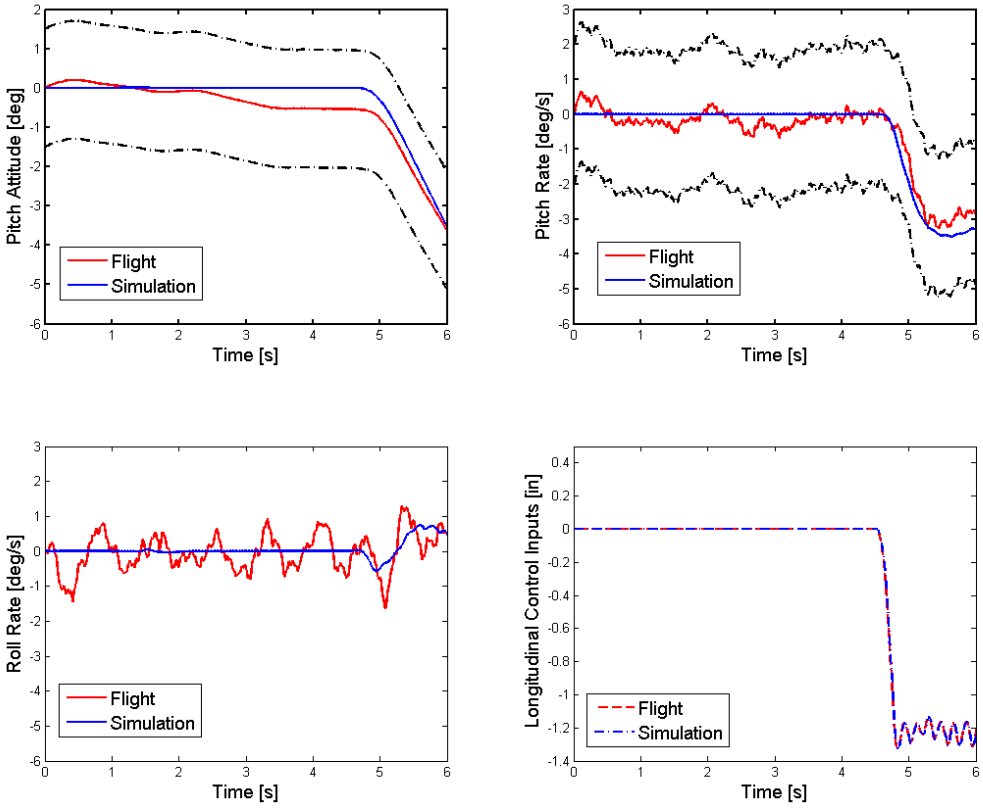


Figure 3-16 - RCAH Longitudinal Step Response - Hover

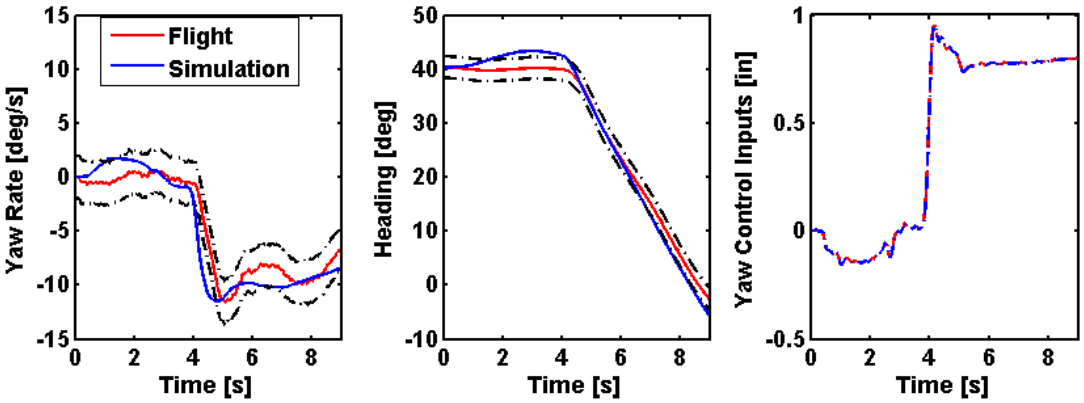


Figure 3-17 - RCAH Pedal Step Response - Hover

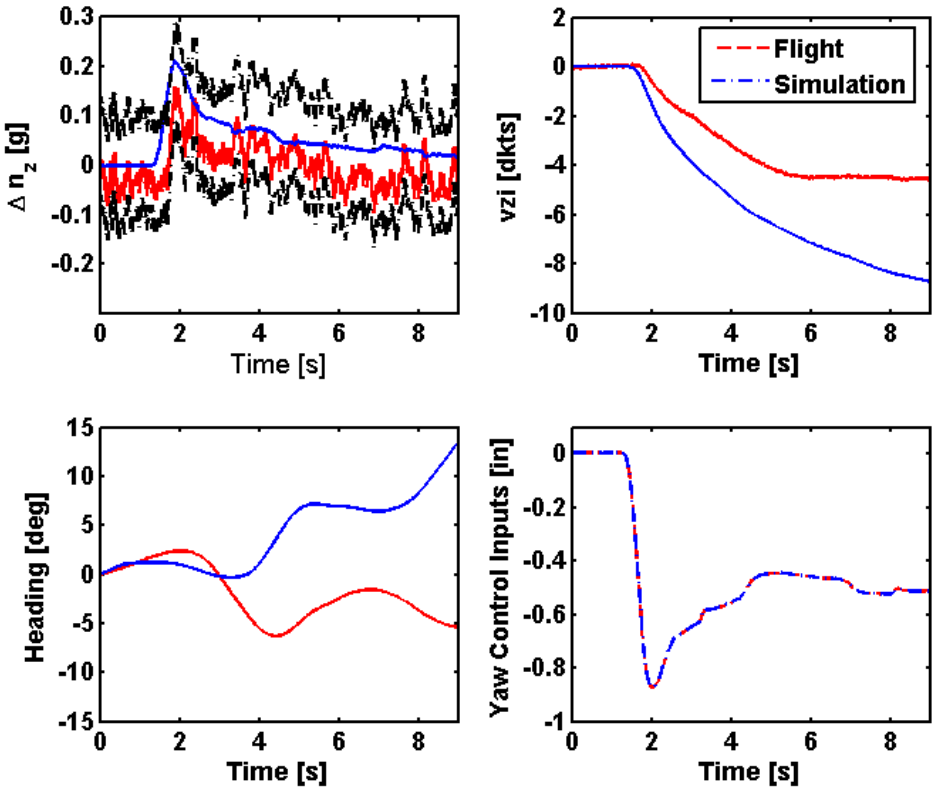


Figure 3-18 - RCAH Collective Step Response - Hover

The responses show Level D on-axis response in pitch (Figure 3-16) and collective (Figure 3-18) axes with borderline matches in on-axis response in the roll (Figure 3-15) and yaw (Figure 3-17) axes. The pitch and roll off-axes responses are of 'correct trend and magnitude'. The yaw and heave off-axis responses are less well matched. This is in agreement with the findings of Manimala et al with regards to the bare airframe mode.

In 2009, as part of the Lifting Standards project, the Bell 412 bare airframe Model was updated to include a turboshaft engine and an interference model based on the inflow model derived from general Peters-He finite state inflow model in an attempt to improve

the match of the off-axis response with flight. Up to date control ranges and trim control positions were obtained from the NRC along with values for aircraft centre of gravity and airframe inertias which had previously been used as tuning parameters.

The differences between the tuned HELI-ACT (HA) FB412 FLIGHTLAB model and the physically correct Lifting Standards (LS) FB412 FLIGHTLAB model are outlined in Table 3-2. The most obvious differences are in the vehicle inertia values which were obtained purely through trial and error to improve the PoM in the HELI-ACT model [68].

Table 3-2 - Differences between HELI-ACT (HA) and Lifting Standards (LS) Bell 412 Models

	FB412_HA	**FB412_LS**
Roll Inertia (slug-ft²)	8000	4073.5
Pitch Inertia (slug-ft²)	25000	16691
Yaw Inertia (slug-ft²)	22000	14679
Vehicle cg (ft)	[11.6 0.04 4.58]	[11.45 -0.07 4.19]
Total mass (lbm)	10000	10200
X-Z product of inertia (slug-ft²)	-1500	-2169.7
Rotor Interference	No	Yes
Engine	Simple	Turboshaft
Initial control positions	[0 0 5.3 7]	[0 0 5 5]
Control ranges	[-6.14 6.33] [-6.11 5.99] [0.21 10.63] [-2.86 3.93]	[-6.1 6.0] [-6.14 6.33] [0.0 10.7]; [-3.92 2.86]

The first task of this research was to quantify the predicted fidelity of both the tuned and updated Bell 412 models as per JAR-FSTD H low speed handling qualities requirements. The flight test data recorded during the first Lifting Standards flight test campaign was used as validation data. The responses of both models and flight vehicle to like-for-like control inputs in lateral cyclic (XA), longitudinal cyclic (XB), collective (XC) and pedal (XP), are shown in Figure 3-19 to Figure 3-26. The dashed black lines represent the JAR-FSTD H tolerances.

From inspection of Figure 3-19 through Figure 3-26, it can be seen that the physically accurate model (FB412 LS) does not match the flight test data as well as the tuned model (FB412 HA). The quickness of the responses to control inputs and pitch and roll control power are much higher with the LS model. Table 3-2 reveals that the main cause for this increased quickness and control power is due to the significantly reduced inertias in the LS model.

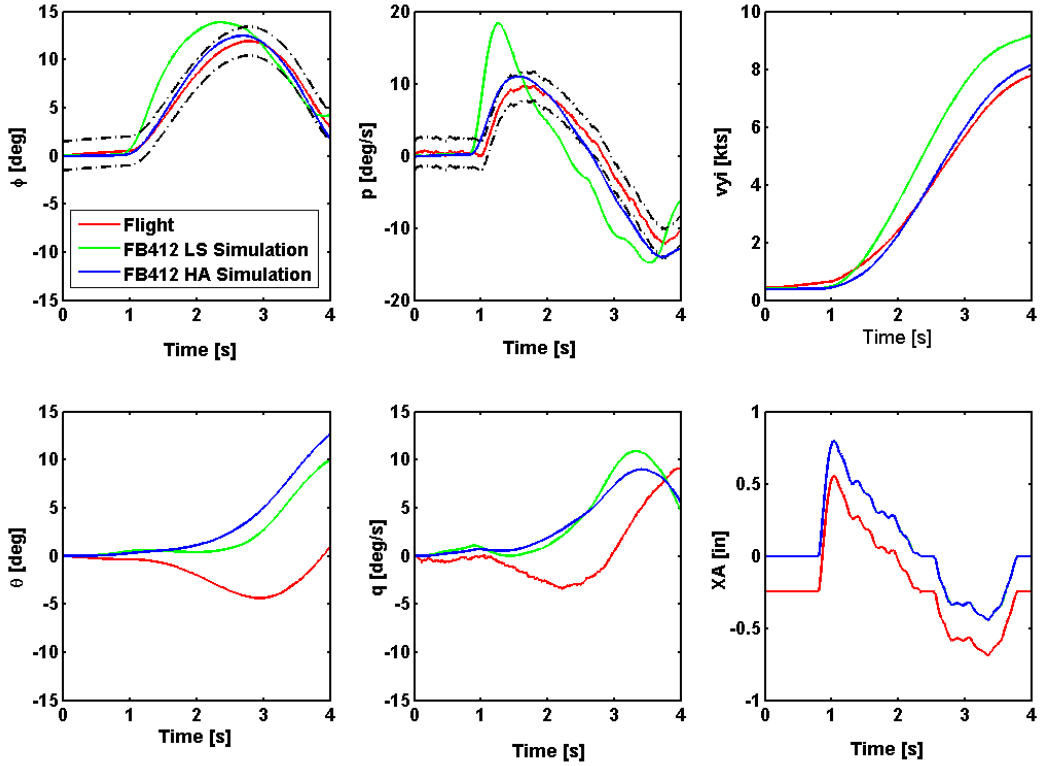


Figure 3-19 – Response to Lateral (XA) Control for HELI-ACT and Lifting Standards Models Against Flight (Hover)

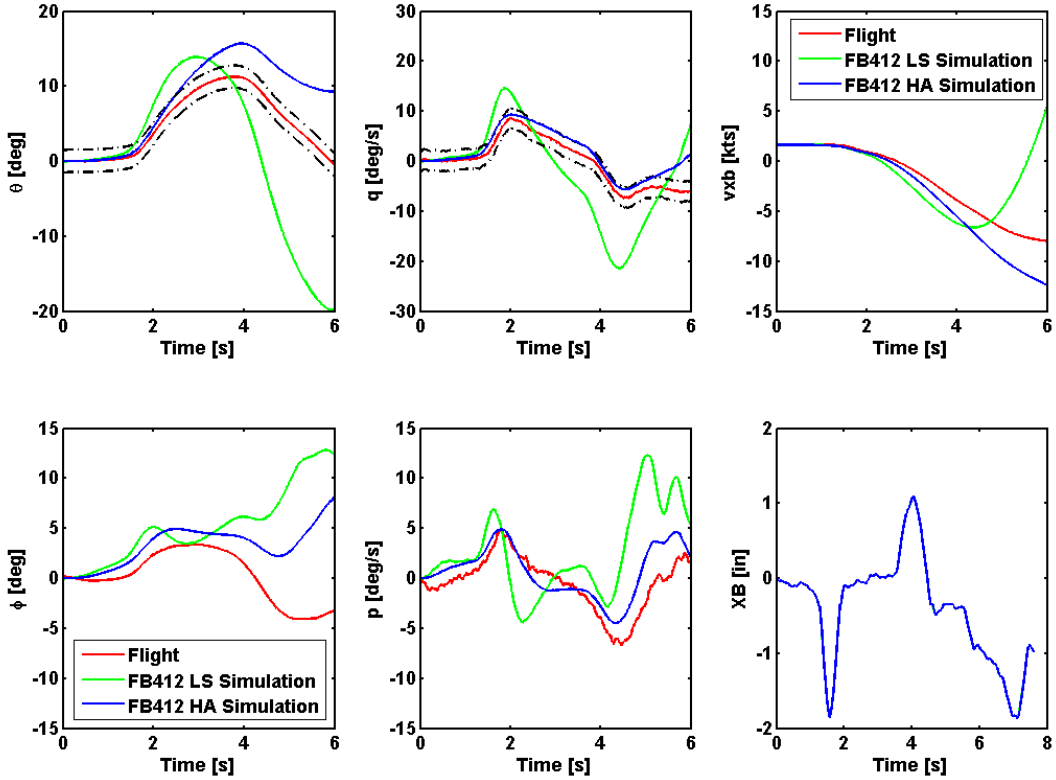


Figure 3-20 – Response to Longitudinal (XB) Control for HELI-ACT and Lifting Standards Models Against Flight (Hover)

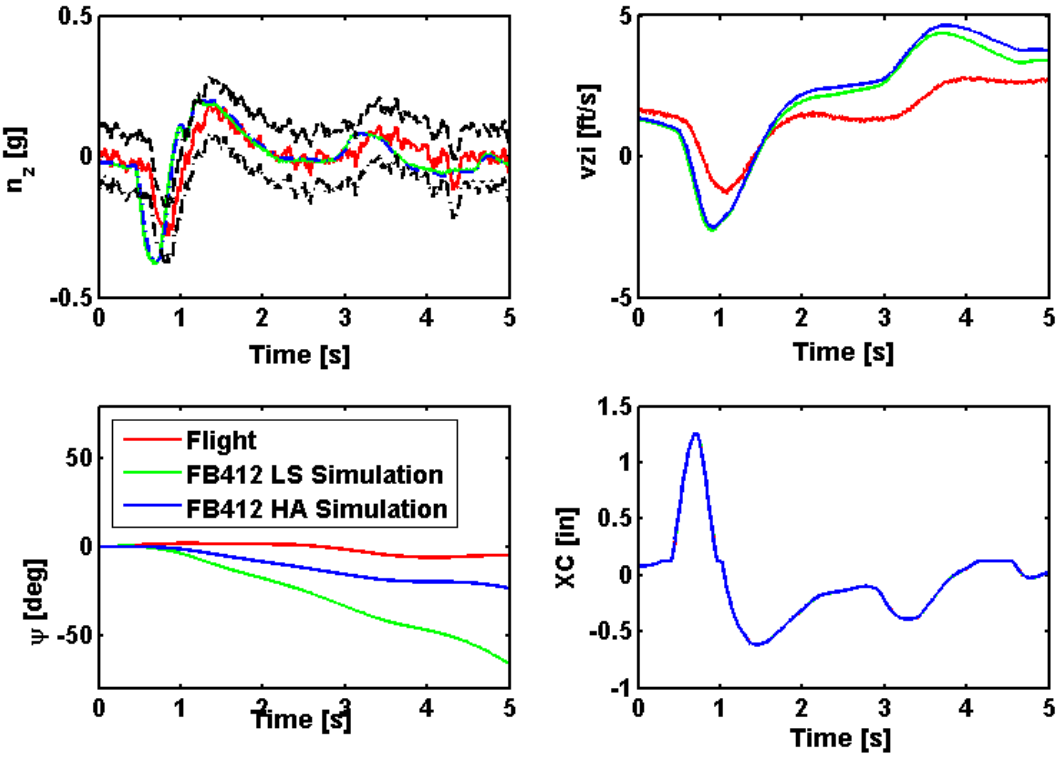


Figure 3-21 – Response to Collective (XC) Control for HELI-ACT and Lifting Standards Models Against Flight (Hover)

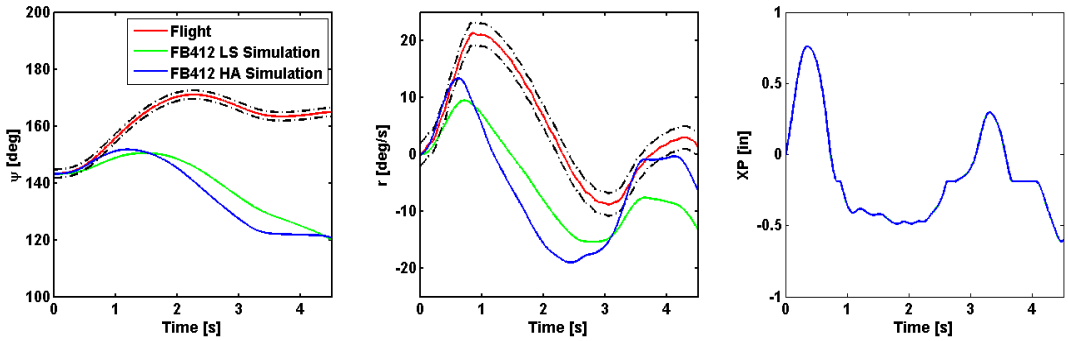


Figure 3-22 - Response to Pedal (XP) Control for HELI-ACT and Lifting Standards Models Against Flight (Hover)

As the confidence in the inertia and Centre of Gravity (CG) values of the LS model was high, it was concluded that these values had been tuned to such a large extent to compensate for other model deficiencies in the HELI-ACT model. If these deficiencies could be identified then it may be possible to obtain a physically correct model that results in a good PoM. This would increase confidence in the model in areas not tested by the PoM requirements.

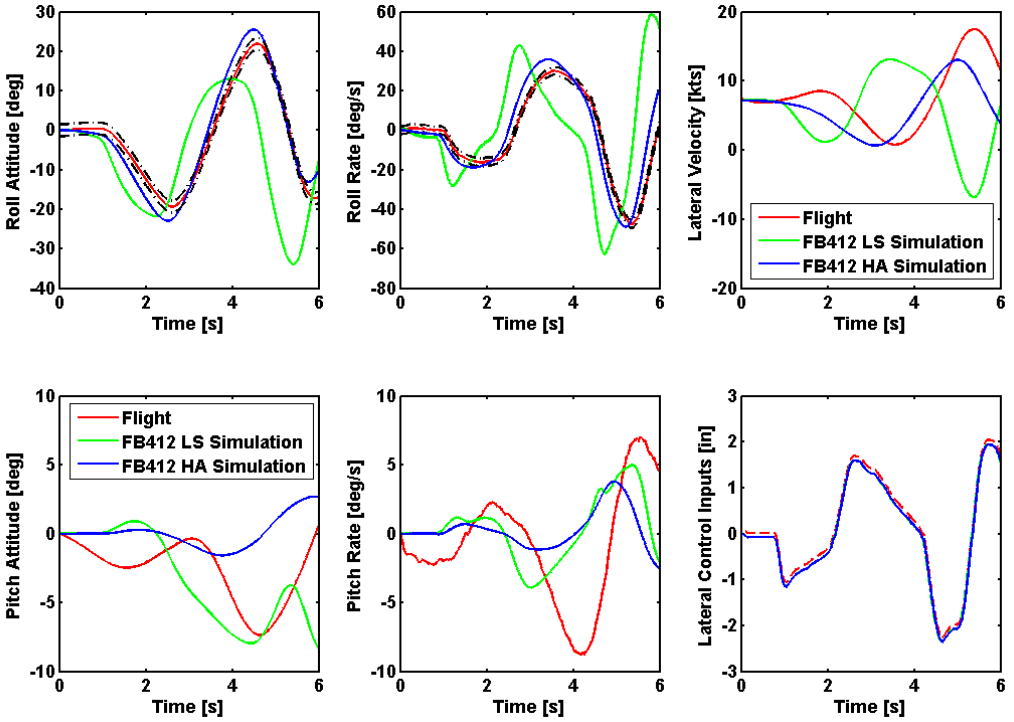


Figure 3-23 – Response to Lateral (XA) Control for HELI-ACT and Lifting Standards Models Against Flight (80 Kts)

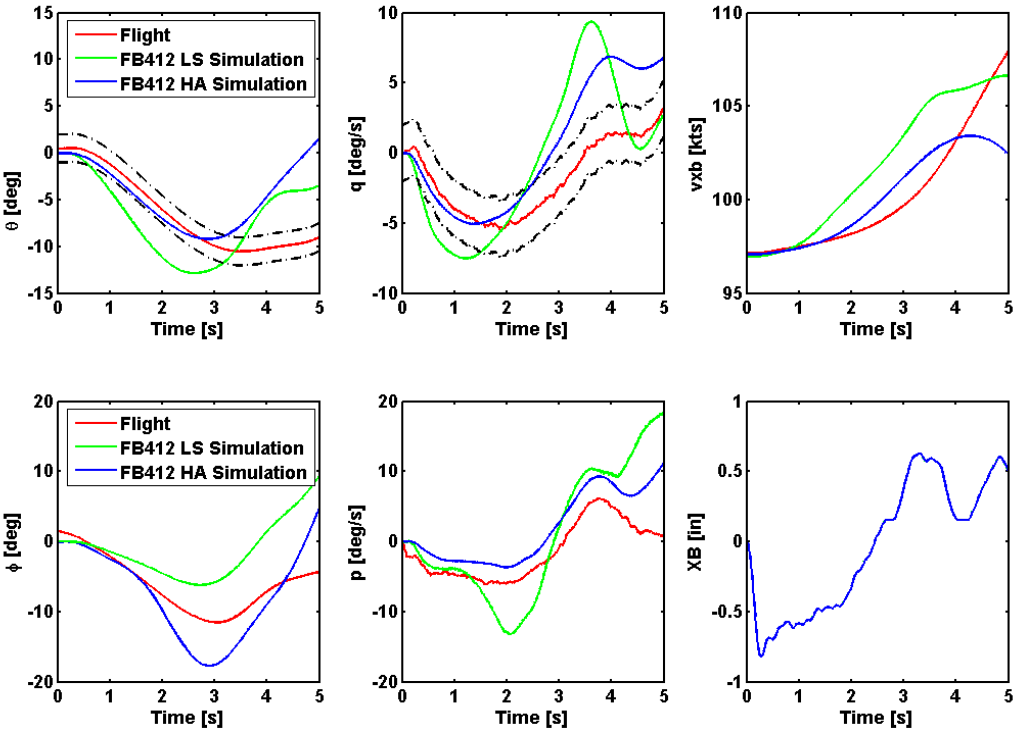


Figure 3-24 - Response to Longitudinal (XB) Control for HELI-ACT and Lifting Standards Models Against Flight (Forward Flight)

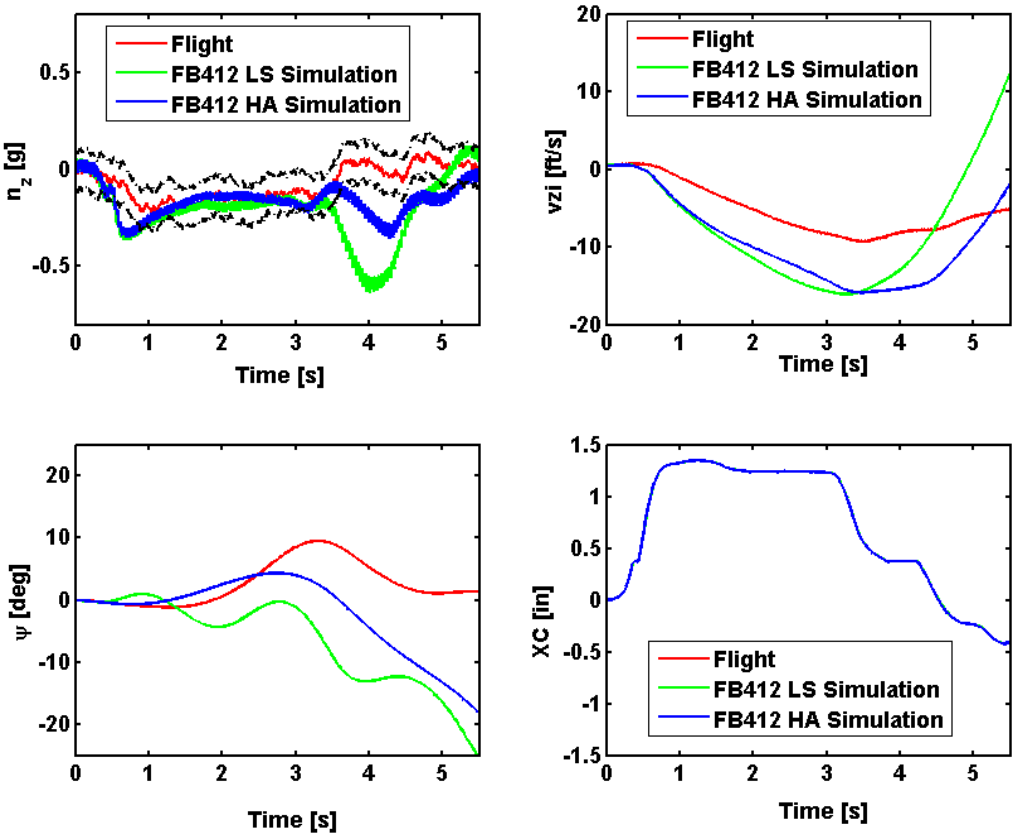


Figure 3-25 – Response to Collective (XC) Control for HELI-ACT and Lifting Standards Models Against Flight (Forward Flight)

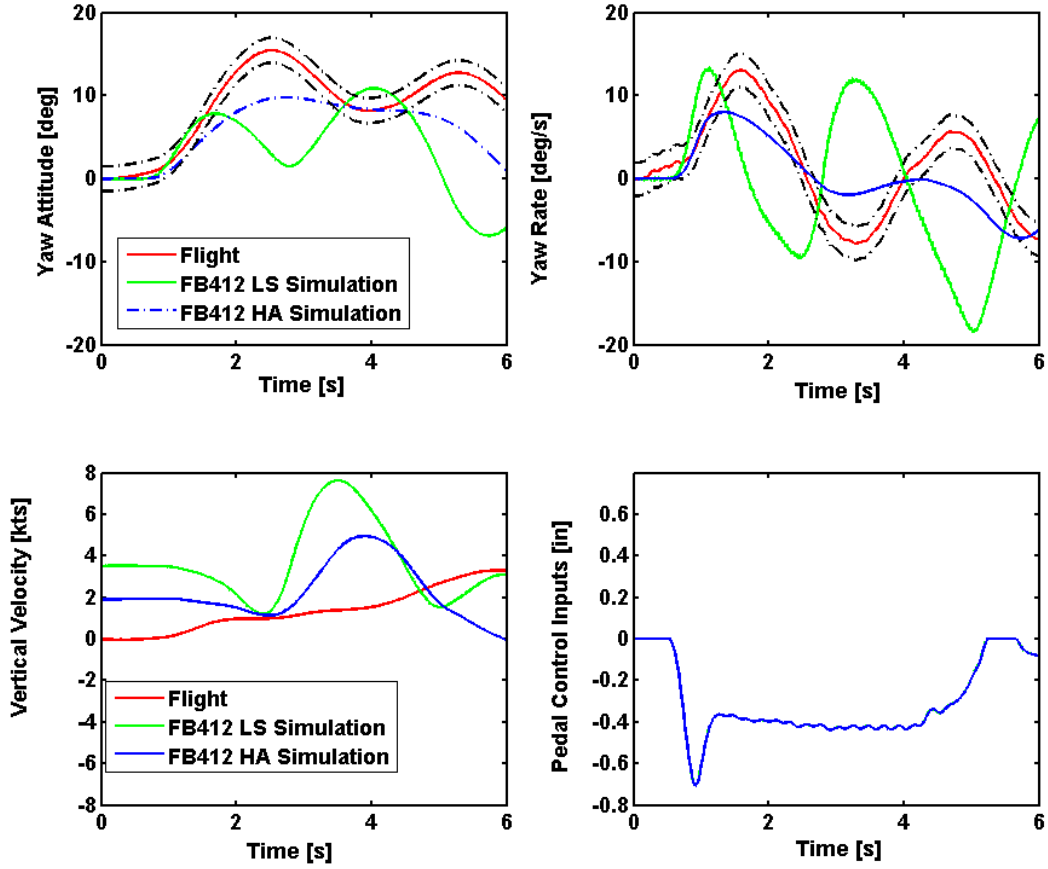


Figure 3-26 – Response to Pedal (XP) Control for HELI-ACT and Lifting Standards Models Against Flight (Forward Flight)

Rotor modelling parameters can have a significant impact on model response quickness. Such parameters include rotor stiffness, λ_β, and Lock number, γ. Therefore, the rotor model component of the Bell 412 FLIGHTLAB model was investigated. The input parameters to the FLIGHTLAB model are shown below with the input values for the FB412_HA model:

Rotor Radius	R	23ft
Rotor Speed	Ω	33.9292rad/s
Spring Stiffness	K_β	- 15381ft-lbf/rad

The Bell 412 ASRA's semi-rigid rotor blades are modelled using an articulated rotor in FLIGHTLAB and so a non-physical parameter, equivalent hinge offset (eR), is also required in the model.

Equivalent Hinge offset	eR	1.058ft

The values for R and Ω were verified through cross reference with other data sources [72], [73], [74]. However, conflicting values for eR were found: reference [68] stated that eR = 1.058ft (as used in FB412_HA). However, reference [74] stated eR = 0.575ft. No alternative references for the value of K_β were found in the literature.

Therefore, it was decided that a third model, FB412 LS_002, would be created with new values of eR and K_β. For this new model, the value of eR was be taken to be 0.575ft, from reference [74] and the value of K_β was determined from the relationship shown below, in Equation 3-1:

$$\mathbf{K_\beta} = \left(\boldsymbol{\lambda_\beta}^{\mathbf{2}} - \mathbf{1} - \frac{\mathbf{eR} * \mathbf{M_\beta}}{\mathbf{I_\beta}}\right) * \mathbf{I_\beta} * \boldsymbol{\Omega}^{\mathbf{2}} \qquad \textbf{Equation 3-1}$$

λ_β is the non-dimensional flapping frequency ratio. A fixed value of $\lambda_\beta = 1.03$ was chosen for the third model due to agreement across several data sources [72], [73], [74].

Values for second moment of flapping inertia, I_β, and first mass moment w.r.t. flap hinge offset, M_β, were obtained from the mass distribution, m(r), and equivalent hinge offset, eR, of the blade, as per Equation 3-2 and Equation 3-3:

$$\boldsymbol{I_\beta} = \int_{\mathbf{eR}}^{\boldsymbol{R}} \boldsymbol{m}(\boldsymbol{r} - \boldsymbol{eR})\boldsymbol{r}^{\mathbf{2}}\boldsymbol{dr} \qquad \textbf{Equation 3-2}$$

$$M_{\beta} = \int_{eR}^{R} m(r - eR)\mathrm{r}dr \qquad \text{Equation 3-3}$$

The mass distribution that was included in the FB412_HA model is shown in Figure 3-27.

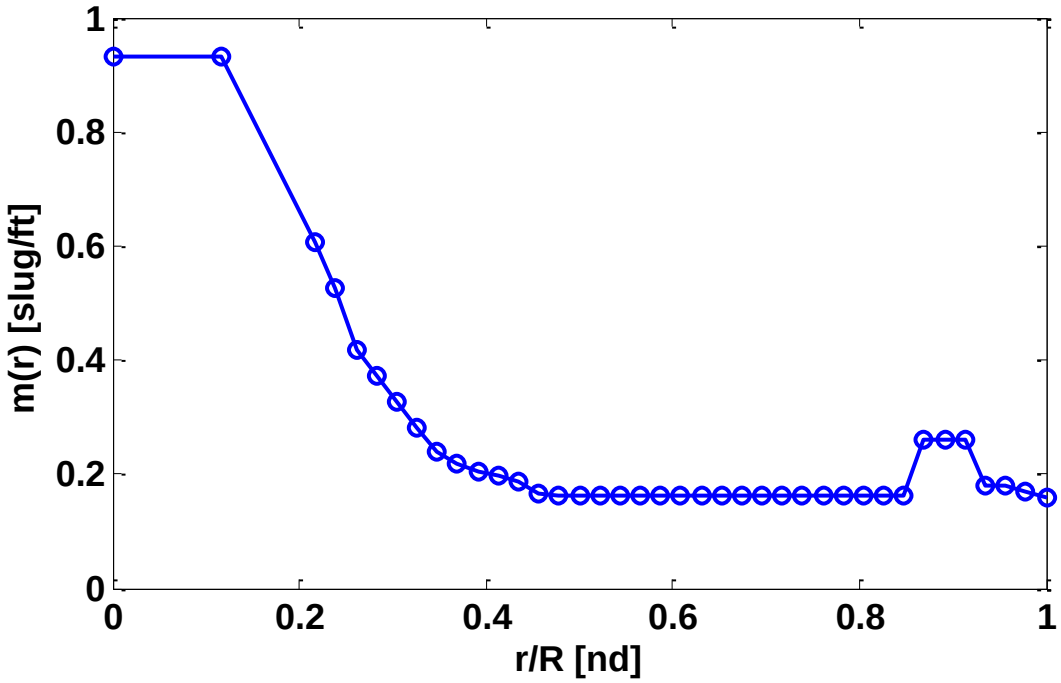


Figure 3-27 - Blade Mass Distribution of FB412 HA and FB412 LS

The step-by-step process for updating the value of K_β is described below:

1) ***A change in eR*** from 1.058ft to 0.575ft effects the values of I_β and M_β, as per Equations 3-4 and 3-5. New values of I_β and M_β were calculated using Equation 3-4 and Equation 3-3 with the mass distribution shown in Figure 3-27 and the new value of eR (0.575ft) to give:

I_β=728.88 slug-ft^2 and M_β=56.6084 slug-ft

2) As a result of the changes in I_β and M_β, ***the value of K_β must then also be changed*** to make Equation 3-3 balance for fixed values for Ω and λ_β.

   Using the changed value of eR, updated values of I_β and M_β and a new K_β value was calculated:

$$K_\beta = \left(1.03^2 - 1 - \frac{0.575 * 56.6084}{728.88}\right) * 728.88 * 33.93^2 = 13630 \quad ft - lbf/rad$$

3) The change in K_β also caused ***a change in γ*** (Equation 3-4) and this change in γ lead to a ***change in S_β*** (Equation 3-5). Neither of these variables are inputs to the FLIGTHLAB rotor model but are included here as they are defining characteristics of a helicopter rotor.

$$\gamma = \frac{\rho c a_0 R^4}{I_\beta} \qquad \textbf{Equation 3-4}$$

$$S_\beta = \frac{{\lambda_\beta}^2 - 1}{\gamma} \qquad \textbf{Equation 3-5}$$

The differences between FB412 LS and FB412 LS-002 can be seen in Table 3-3. The updates to K_β and eR led to a slight decrease in Lock number, γ, and therefore it would be expected that the quickness of the response would be reduced. This is what was needed to improve the POM with flight test data.

As an example, the comparative plots between the two models and flight test data, in response to a longitudinal cyclic input in hover and at 100kts, can be seen in Figure 3-28 and Figure 3-29 respectively.

Table 3-3 - Difference in Rotor Model Parameters in HA FB412 and LS FB412 Models

Parameter		FB412 LS	FB412 LS -002
Second moment of flapping inertia [slug-ft^2]	I_β	682.79	728.88
First Mass Moment w.r.t. Flap hinge offset [slug-ft]	M_β	54.1961	56.61
equivalent hinge offset [ft]	eR	1.058	0.575
Flapping Spring Stiffness [ft-lbf/rad]	K_β	-15381	13630
Lock number [nd]	γ	6.94	6.50

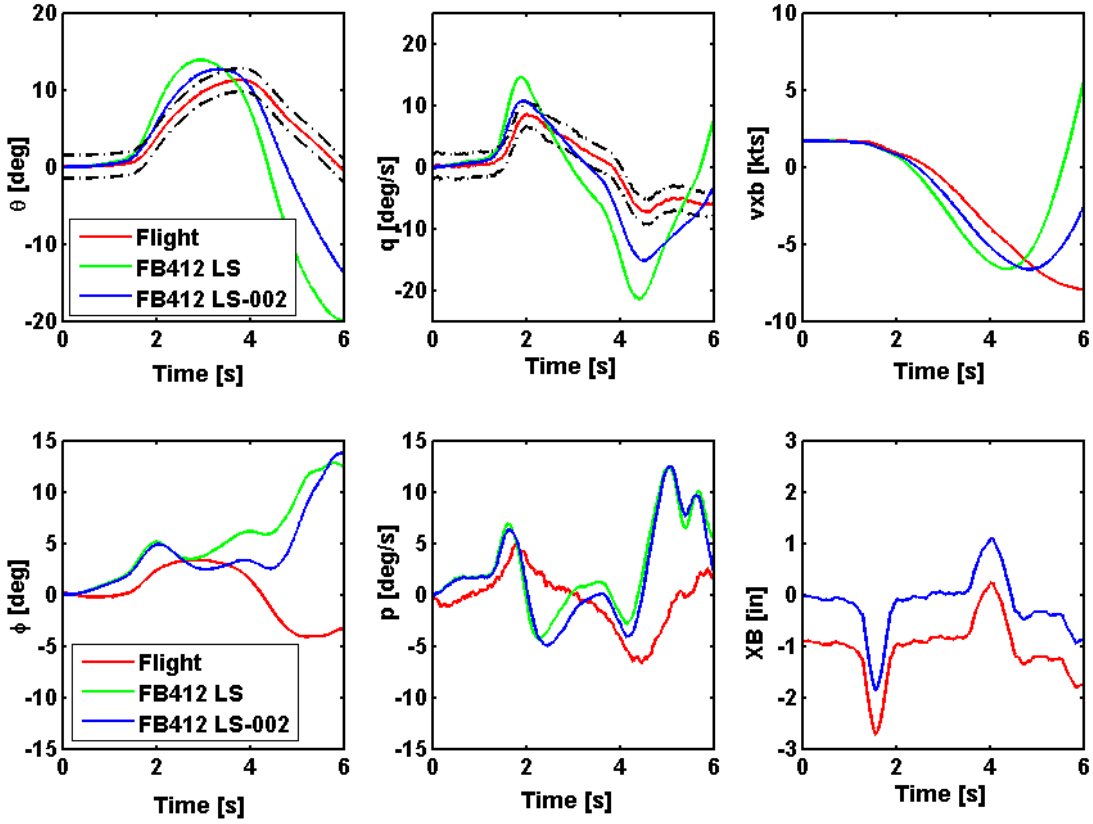


Figure 3-28 - Reduction in Lock Number, γ, to Reduce Response Quickness - Hover, Pitch Response

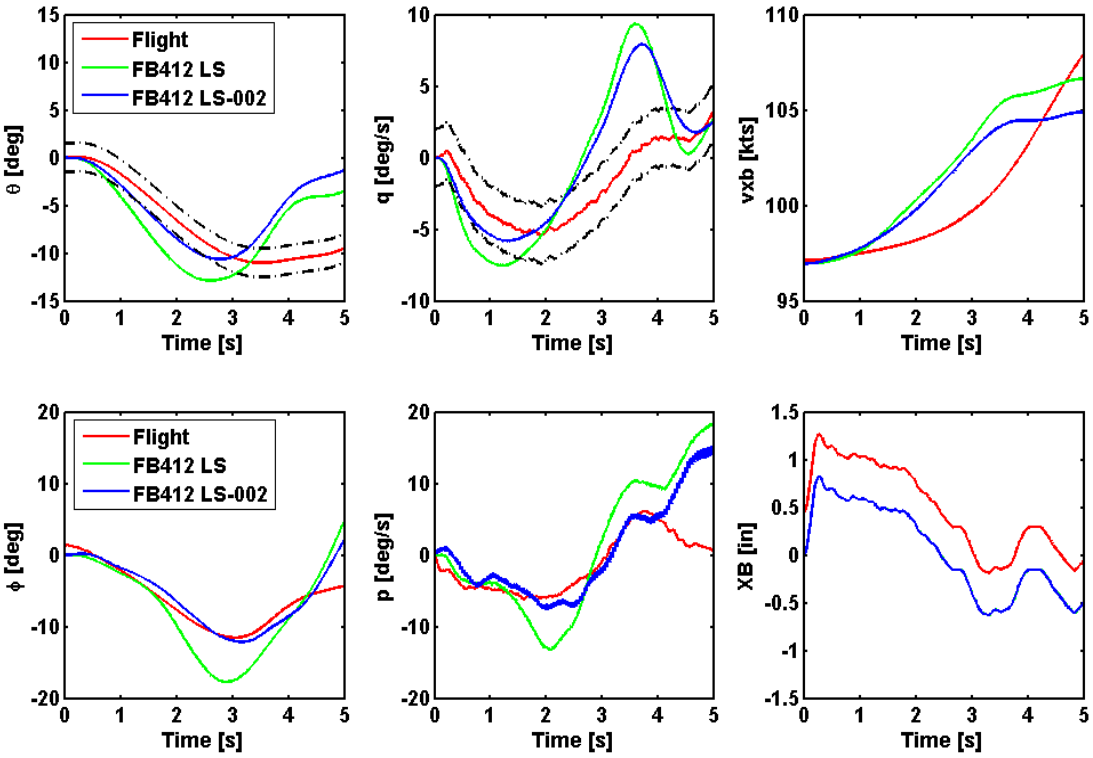


Figure 3-29 - Reduction in Lock Number, γ, to Reduce Response Quickness - 100kts, Pitch Response

Figure 3-28 and Figure 3-29 show that the reduction in Lock number, γ, reduces the peak attitude and rate response. However the initial rate of response is still much greater than that of the flight test data.

It was suggested that perhaps the rotor mass distribution itself could be the cause of the overtly high quickness. Therefore, the mass distribution of the model was investigated.

The FB412_HA rotor model used the NRC value for blade mass as 117.1lbs (3.63 slug) and the yoke mass as 144 lbm (4.472 slug) this gives a total mass over the rotor radius as 261 lbm (8.1086 slug). The mass distribution from the FB412_HA model shown in Figure 3-27, leads to computation of;

Total mass= 6.45 slug

M_β = 54 slug - ft

I_β = 683 slug - ft^2

Therefore the blade is much lighter in the model (6.45 slug) than reported by the NRC (8.11 slug). This leads to values for I_β and M_β that are lower, which in turn leads to a high value of γ and therefore quicker response than the aircraft.

Assuming the blade as a uniform rod, an approximation of blade mass distribution is:

$$I_\beta = \frac{1}{3} * total\ mass * R^2 = 1429.8\ slug\text{-}ft^2 \qquad \textbf{Equation 3-6}$$

The actual I_β value would be expected to be lower than this as more of the mass is concentrated towards the hub, suggesting the mass distribution model in the FB412-HA and FB412-LS models is incorrect. An internal report by Manimala [75] derived a different blade mass distribution to the one that was implemented in the model. This mass distribution model utilised the blade mass distribution profile and blade mass from reference [75], a yoke mass value supplied by the NRC and the assumption of linear mass variation in the yoke. Linear interpolation was used between the yoke and blade profiles (3-6ft from the hub). This is shown in Figure 3-30 and it can be seen that there is a higher mass concentration at towards the hub in this distribution profile.

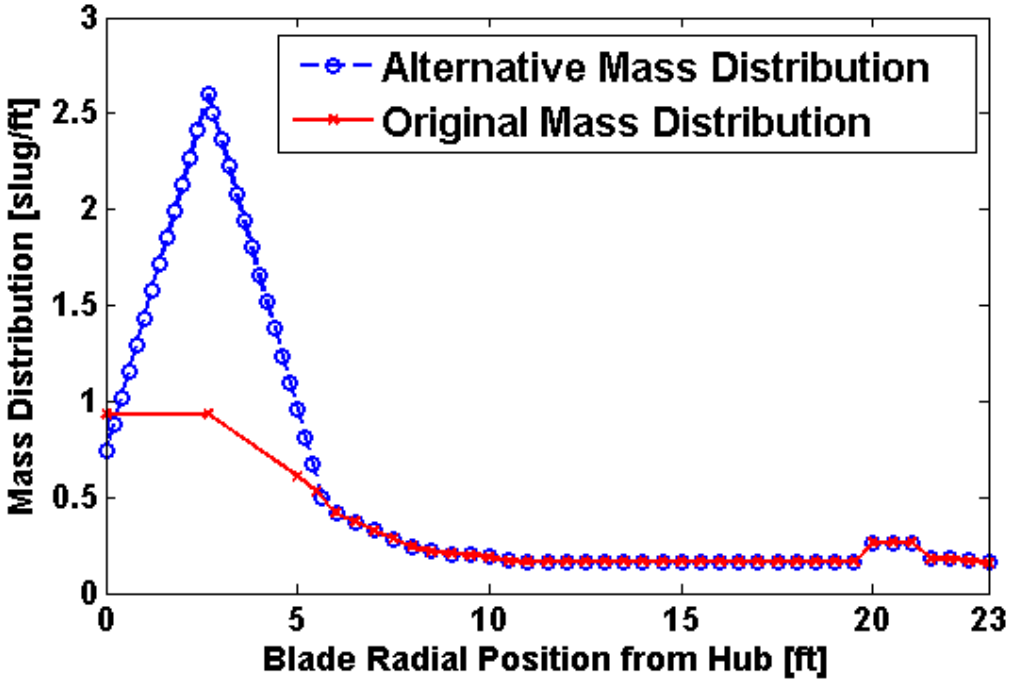


Figure 3-30 - Two Different Blade Mass Distribution Models for the Bell 412 ASRA

The alternative mass distribution accompanied with the alternative hinge offset, leads to computation of;

Total blade mass = 9.4483 slug

M_β = 67.6172 slug - ft

I_β= 764.3056 slug-ft^2

The differences between the original FB412_LS model and the updated FB412_LS-003 model can be seen in Table 3-4.

Table 3-4 - Differences between FB412_LS Model and FB412_LS_003 Model

Parameter		**FB412-LS**	**FB412 LS-003**
Second moment of flapping inertia [slug-ft^2]	I_β	674.2489	764.3056
First Mass Moment w.r.t. Flap hinge offset [slug-ft]	M_β	51.4395	67.6172
equivalent hinge offset [ft]	eR	1.058	0.575
Lock number [nd]	γ	7.029	6.2088
Flapping Spring Stiffness [ft-lbf/rad]	K_β	-15381	8820.6

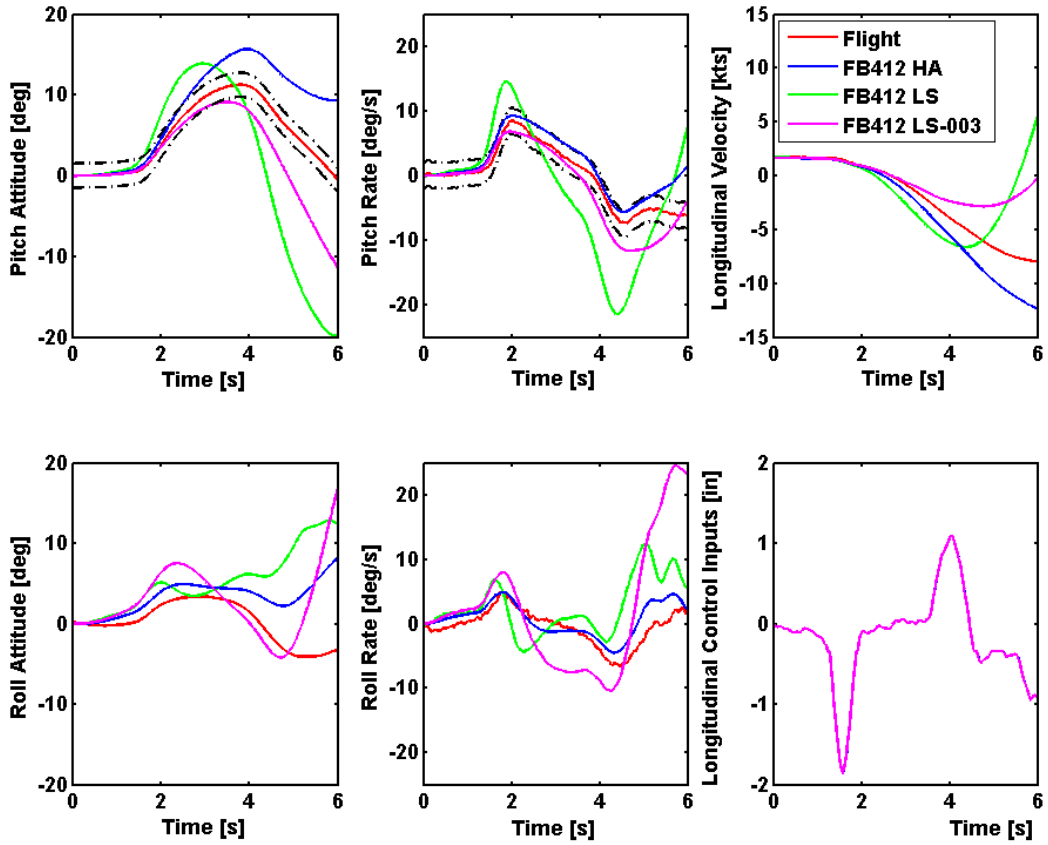


Figure 3-31 - Effect of Blade Mass Distribution on Control Response - Hover, Pitch Response

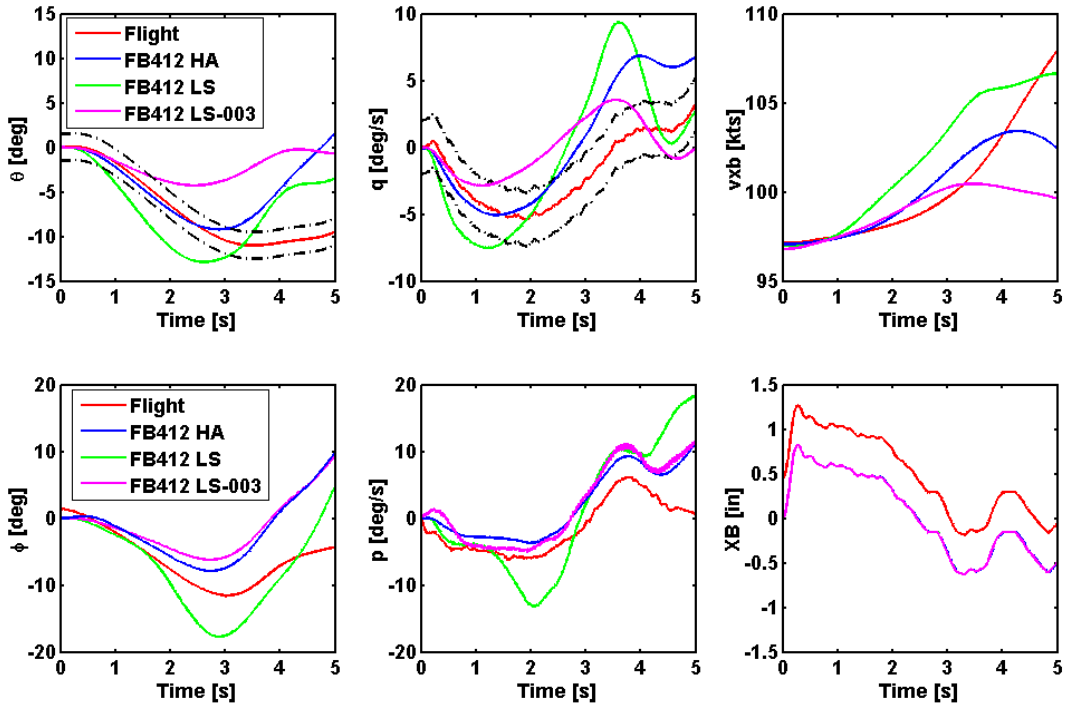


Figure 3-32- Effect of Blade Mass Distribution on Bare Airframe Control Response - Forward Flight, Pitch Response

Examples of effect of the model changes discussed above on PoM are shown in Figure 3-31 and Figure 3-32. The modified Lifting Standards Model (FB412 LS-003) showed a marked improvement compared to FB412-LS and LS-002 (Figure 3-28) in the pitch axis response in hover. In Forward flight, the modification of the blade mass distribution dramatically reduced pitch control power and so the match of L3-003 (Figure 3-32) was degraded from that of LS-002 (Figure 3-29). However, the tuned HELI-ACT model (FB 412-HA) undoubtedly shows the best match with flight test data in both hover and forward flight (Figure 3-31 and Figure 3-32). This study highlighted that often models that are known to not physically represent the aircraft often result in the best PoM. This is a concern as it is unknown whether the model will behave like the aircraft in complex, manoeuvring flight. This is where subjective assessment of simulation becomes key. The next chapter describes the development of a new subjective rating scale for fidelity assessment. This scale has been used to assess the various model configurations to determine whether the tuned model is perceived to be better than the physically accurate model in manoeuvring flight. The results of this subjective assessment are presented as a case study at the end of Chapter 4.

Multiple unknowns may have to be investigated to unearth the root of the poor PoM of the physically correct model. However, this activity was not stated as an aim of the research and therefore it was decided that for this research the HELI-ACT model would be used with updates to a turboshaft engine model and Finite state rotor interference modelling.

3.4 Conclusions

1) The HELIFLIGHT-R facility is an appropriate platform for flight simulator fidelity research for a number of reasons. Firstly, the flexibility of the system and modelling software allows controlled modifications to be made to the simulation to examine their effects on perceptual fidelity. Secondly, the availability of motion cues and a large FoV, along with accurate replication of the Bell 412 ASRA instrument panel and NRC handling qualities test course, ensures a high level of immersion during piloted simulation trials.

2) The research contained within this thesis is based on the Bell 412 Advanced Systems Research Aircraft. Two models were created as part of the HELI-ACT (HA) and Lifting Standards (LS) work using flight test and aerodynamic data. One model (HA model) used tuned values of CG and vehicle inertias to optimise the match between flight and simulation time domain responses. The second model (LS model) included accurate values of inertia and CG.

3) Investigations into the physical accuracy of the rotor blade element model lead to enhancement of the on-axis PoM of the physically accurate, LS model. However, the best overall match was obtained using the tuned HA model and therefore this model was chosen for use in the remainder of the research.

4) Concerns have been raised as to whether a pilot would perceive the 'non-physicality' of the tuned model. Therefore, subjective assessment was proposed to ensure the validity of the tuned model in manoeuvring flight.

4 DEVELOPMENT OF A RATING SCALE FOR SUBJECTIVE ASSESSMENT OF SIMULATION FIDELITY

This chapter describes the development of a new, industrially applicable, rating scale for the subjective fidelity assessment of fully integrated flight simulations. The methodology for application of the resulting Simulation Fidelity Rating (SFR) scale is outlined and the scope of its applicability discussed. A small case study is presented where the final version of the rating scale was used to determine the effect of model tuning on perceptual fidelity. Guidance notes for future utilisation are also included in this chapter.

4.1 Background and Introduction

Subjective pilot opinion is an important component to the validation or qualification of a flight simulator. However, it was highlighted in Chapter 2 that there is no formal methodology for capturing this subjective data. A number fidelity rating scales have been developed [38], [61] and implemented in research projects with varying degrees of success, as discussed in Chapter 2. However, existing scales are either focused on a specific component of the simulation rather than the integrated simulation or they lack true insight into the deficiencies of the simulation.

A need was identified by Padfield et al during Lifting Standards project at the University of Liverpool to address the subjective evaluation gaps in the existing simulator qualification processes, by developing a subjective rating scale that could be industrially utilised for flight simulation fidelity assessment. The utilisation of such a scale would add structure, standardisation and transparency to the subjective assessment of flight simulators. A collaborative project between UoL and the NRC was approved and part-funded by the US Army International Technology Centre (UK) to develop such a scale.

It was intended that the subjective ratings would complement the quantitative comparison of aircraft data currently used (where available) and would act as a minimum requirement where little or no flight test data are available (low cost training devices, research & development simulators etc.).

As part of this collaborative team, the author was involved in the decisions regarding the structure and terminology of the SFR scale. It was also the responsibility of the author to develop a robust methodology for excising the SFR scale to obtain pilot feedback as well as to write guidance material to ensure correct utilisation of the methodology.

4.2 Preliminary Rating Scale Design Considerations

Padfield proposed drawing on the experience of the Handling Qualities (HQ) community in capturing pilot opinion and associated ratings, for the development of the subjective fidelity rating scale. The justification for this is that the HQR scale has been successfully adopted internationally in the handling qualities community and the test pilot and flight test engineer communities are familiar with the HQR scale and formally trained to use it. Therefore, it was decided early on that, as with the HQR scale, the rating scale should be algorithmic and require the evaluation pilot to answer a series of questions to determine what 'Level' their rating lies within.

The aim of the subjective rating scale was to formally evaluate perceptual fidelity of the integrated simulation environment, rather than to obtain fidelity ratings of individual components. Based on this aim, an initial framework and fundamental requirements were drawn up following discussions between the NRC and UoL. The fundamental requirements that were agreed upon were:

1. The scale should be efficient and straightforward to use.

2. The output of the scale should be a measure of the fitness for a specific, well defined purpose of the fully integrated simulation, rather than individual components.

3. The parameters against which the pilot rates the simulation fidelity should be ones that are understood at the cognitive level.

4. Qualifiers of simulation perceived fidelity should be defined such that a common, unambiguous understanding can develop between evaluation pilots, engineers and regulatory bodies.

5. The scale should be sufficiently comprehensive that SFRs, along with supporting pilot comments, reflect and capture all fidelity deficiencies.

Perceptual Fidelity was defined in Chapter 1 as the simulator's ability to induce the trainee pilot to output those behaviours known to be essential to control and operation of the actual aircraft in performance of a specific task. It was therefore agreed that pilot behaviour and performance should be the rating parameters in the scale as they are directly comprehendible by the pilot and have direct impact on the fitness for purpose of the simulation.

Padfield (UoL) and Erdos (NRC) drafted three variants of a Simulation Fidelity Rating (SFR) scale, SFR Scale A, B and C, were drafted against the criteria set above. These are shown in Appendix A (Figures A-1 to A-3). All three scales had four levels of perceptual simulation fidelity (Level 1 being the best and 4 being the worst). In each case, the Level of utility of the simulator for a specific purpose is determined through the answers to questions in a decision tree format. SFR scale issue A (see Appendix A - Figure A-1) required negative answers to move down the levels of utility whereas SFR scale issues B and C (Appendix A - Figures A-2 and A-3) required positive answers to move up. All three scales then used various descriptors of task strategy and performance to distinguish specific ratings.

During discussions between the teams at the NRC and UoL, was agreed that a rating in Level 1 reflects a simulation that is 'fit for purpose'. By extension, a rating in Level 2 means the simulation has limited utility for the specific purpose. Finally, a Level 3 rating means that the simulation has poor perceptual fidelity for the given task and therefore the simulator is not fit for the specific purpose. It should be clearly noted that the specific definition of 'fit for purpose' is inherently dependent on the purpose itself. The purpose of the simulation must be clearly identified and briefed to the test pilot before any meaningful assessment of fitness for purpose can be obtained.

Classes of training simulator purpose were considered to be skills acquisition, skills development and skills assessment. For every purpose, there will be a unique definition of ‘fit for purpose’ and therefore a different description associated with each Level. The definitions of each Level of fidelity for the various training purposes are given in Table 4-1. In the exploratory trial, the simulation purpose was taken to be skills acquisition.

The SFR scale was developed within the remit of assessing rotary wing training simulators. However, it is intended that the SFR methodology can be utilised to assess any training simulator be it a fixed-wing simulator, driving simulator, medical procedures trainer or virtual reality training environment. Furthermore, the application of the SFR scale is not limited to the assessment of training simulators. The SFR scale also has utility for the assessment of the fitness for purpose of design, development, and research simulators. In these cases, the definitions of fitness for purpose are specifically tailored.

For example, proposed definitions of 'fit for purpose' for a design and development simulator being used for cockpit development, and a research simulator used for determination of Ship-Helicopter Operating Limits (SHOL) envelope are given in Table 4-2.

Table 4-1 - Definitions of Fidelity Levels for a Variety of Training Simulator Applications.

Simulator Application	Level 1	Level 2 Fidelity	Level 3 Fidelity
Skills Acquisition Initial Training Attain Skills	100% positive transfer of training - Simulation training is sufficient to allow operational performance to be attained with minimal pilot adaptation. There is complete Transfer of Training (ToT) from the simulator to the aircraft in this task.	Some positive transfer of training - Additional training in the aircraft would be required in order to achieve operational performance in the aircraft. There is partial positive ToT from the simulator to the aircraft in this task and simulator training must be augmented with in-flight training for operational proficiency.	Negative transfer of training would occur, meaning that the simulator is not suitable for training to fly the aircraft in this task.
Skills Development Recurrent Training Retain Skills	Simulation training is sufficient to restore skills for operational performance	Simulation training provides limited improved performance capability. Additional training in flight required.	No positive transfer of training. Simulator is unsuitable for training application
Skills Assessment Check Ride Demonstrate Skills	Simulation is sufficient to comprehensively demonstrate skills associated with qualified performance	Performance in simulator demonstrates limited elements of the required skills for operational performance	Performance in simulator does not serve to demonstrate the required skills.

Table 4-2 - Example Definitions of Fit for Purpose for Application to Design & Development and Research Simulators.

	SHOL research simulator	Cockpit design and development simulator
Definition of 'Fit for Purpose'	Effects of aerodynamics on pilot workload are fully replicated and sufficient ship motion cues are supplied for the specific sea state and wind combinations. High confidence in results obtained using the simulator.	Physical and mental workload associated with specific flight regime fully replicated. Cockpit dimensions and layout representative of aircraft. Full transfer of the effects of cockpit upgrades to aircraft.

4.3 Subjective Assessment Methodology

Again drawing on HQ engineering practices, it was decided that fidelity should be evaluated in clearly defined tasks rather than in an arbitrary flight profile defined by the pilot. Performance criteria, akin to those set out in ADS-33E-PRF MTEs [19] make it easier for the pilot to reflect on their performance, and clear task definition ensures comparison of like-for-like experiences.

The basic fidelity assessment methodology using the SFR scale is defined as follows:

Test pilots fly the task in the aircraft and notes the level of performance achieved and task strategy used. A HQR may also be taken to benchmark the performance and level of pilot compensation. The evaluation pilot(s) then repeats tasks in the simulated vehicle. The pilot would again note his performance, task strategy and award a HQR. The SFR scale and questionnaire would then be utilised to draw comparisons between the two experiences and lead to an SFR rating for the specific task. In the case of qualification of a training simulator, a full list of training tasks with associated SFRs could form a qualified 'training envelope' of the simulator. For a research simulator, the resulting SFRs would form part of the experimental setup validation documentation. A small case study using this method is detailed in section 4.8

A novel application of the SFR scale has also been identified by the author. The SFR scale can be used to aid the development of rationalised quantitative training simulator certification fidelity tolerances. In such a case the definitions of 'fitness for purpose' would be that of the intended FSTD but the nominal vehicle would no longer be the aircraft but would be a baseline simulation (see Table 4-3). Parameters of interest can be incrementally varied in controlled simulation experiments and a pilot would then use the SFR scale and questionnaire to compare each variant to the baseline. The ratings would then be plotted against the varied parameter to draw boundaries of acceptable/unacceptable errors for a specified training task and objective. Ideally, this process would be repeated with several test pilots. This methodology is used in the experimental work described in Chapter 5.

Table 4-3 - The Nominal and Simulated Vehicles for Each Simulator Application Area

	Nominal vehicle	Simulated Vehicle
Training	Aircraft	FSTD
Design and Development	Aircraft	Baseline simulation
Research	Aircraft	Baseline simulation
Development of Metrics	Baseline simulation	Modified simulation

4.4 Limitations of methodology

There are some requirements with regard to pilot selection. For optimal results, the test pilot must be proficient in flying the vehicle and task and also must have operational recency so that a meaningful comparison can be made. It may not always be possible to find a qualified test pilot that also meets these requirements. In such a case extra care should be taken in the pilot brief and debrief and it should be noted in the publication of the results that an operational pilot was used rather than a test pilot.

As with HQRs, Simulation Fidelity Ratings (SFRs) should be taken from at least three pilots. Using three pilots aims to mitigate against the effects of anomalous ratings due to vested interest [77] and pilot strategy However, this may not always be possible. It is imperative that all ratings are included in published results and that rating spread and sample sizes of test pilots are clearly presented.

The scenarios in which the SFR scale can be used are also limited as there is a requirement for real world experience. Therefore the SFR scale could not be used for the assessment of simulators of concept aircraft that have not been flown.

A subjective fidelity rating determines goodness and also the severity of deficiencies, but does not highlight what the simulation deficiencies may be. Therefore, an accompanying questionnaire was developed to capture pilot commentary in a standardised format for the test engineer to determine what aspects of the simulation need to be improved. The questionnaire is shown in Appendix A, Table A-1. The questionnaire allowed the pilot to consider each component of the simulation and each degree of freedom which aids the pilot in introspecting on the experience before awarding a rating.

In February 2011, a week long exploratory trial was conducted at the University of Liverpool in collaboration with the NRC to assess the utility of each variant of the SFR scale for application to training simulator evaluation and then the further develop the chosen variant and assessment methodology.

4.5 Exploratory Trial

In order to develop the SFR scale, the three variants had to be exercised in a range of realistic scenarios. The evaluation pilot for the exploratory trial was Robert Erdos, who is the chief test pilot at the NRC and has a great deal of experience of flying in the Bell 412 ASRA. This ensured that he would be suitable for rating the fitness for purpose of the HELIFLIGHT-R Bell

412 simulation for training the flying skills required for a range of MTEs. Prior to the exploratory trial, the test pilot flew four MTEs in the Bell 412 ASRA in Ottawa to award HQRs and ensure recent experience in the relevant MTEs.

The trial lasted one week and the pilot flew the same four MTEs that were flown in Ottawa with four different FBW control configurations - Bare Airframe, Rate Damped, ACAH and RCAH. On the first day of trials, the pilot utilised each variant of the SFR scale (Appendix A, Figures A-1 to A-3) and the original draft of the Simulator Fidelity Questionnaire (Appendix A, Table A-1) to assess the fitness for purpose of the various HELIFLIGHT-R simulations for training flying skills for the individual MTEs in the corresponding Bell 412 ASRA configurations. The pilot's feedback from these initial assessments on the utility of each of the variants of the SFR scale were as follows:

SFR issue A - UoL Scale 1 (Appendix A - Figure A.1)

- There is a need to clearly differentiate between specific and general purpose.
- The term ‘deficiencies’ was confusing to the pilot as he was unable to introspect on the specific deficiencies in the simulator. There was also confusion as to whether the term related to deficiencies in the simulator HQs because the same term is used to mean HQ deficiencies in the HQR scale.
- The inclusion of comparative workload and handling qualities in Level 2 and 3 of the scale is misleading. In the precision hover the same overall level of workload and HQ level was perceived but for very different reasons - deficiencies in the simulation can cancel themselves out in this way.

SFR issue B - UoL Scale 2 (Appendix A - Figure A.2)

- Semantically, ‘task strategy’ would be better than ‘control technique’ as it is not only the pilot's hands and feet that are active.
- The question "Is the simulation satisfactory without improvement?" lead to some confusion. The same question is asked of the aircraft in the HQR scale pertaining to its HQs. There is possibility that the pilot may misinterpret the question and respond regarding the simulator HQs rather than its perceptual fidelity.
- Same comments as UoL Scale 1 on workload and HQ Levels.

SFR issue C - Erdos Scale (Appendix A - Figure A.3)

- Noted the need to define the purpose before making a decision on whether the 'fidelity is satisfactory for purpose'.
- Definitions of comparable and representative performance are needed.
- Found that initial questions lead to Level 2 but the descriptors that are most applicable are in Level 1 as comparable performance was perceived.
- Level 2 Fidelity can occur with comparative performance if the task strategy is different - a two dimensional problem.

After this initial feedback, a discussion session was held to develop a new SFR scale that incorporated the best parts of each SFR scale variant. The semantics and structure were then further modified where appropriate to deal with ambiguity and incompatibility (Table 4-7). This lead to SFR scale issue 3, see Figure 4-3. It should be noted that this is not the final version of the SFR scale. The SFR scale was further modified in light of feedback from other pilots after more testing (See Section 4.7). The final version of the SFR scale is shown in Figure 4-4. The evolution of the SFR scale was iterative and therefore different versions were used by different pilots. Table 4-7 summarises which version was used by each test pilot. All results presented have been updated to correlate with the ratings of the final version of the SFR scale (Figure 4-4).

It was noted that in the HQR scale, the pilot's ratings lie on a diagonal line. For example, the pilot cannot rate good performance with high compensation (see Figure 4-1). A similar figure of task strategy adaptation and comparative performance can be constructed (see Figure 4-2).

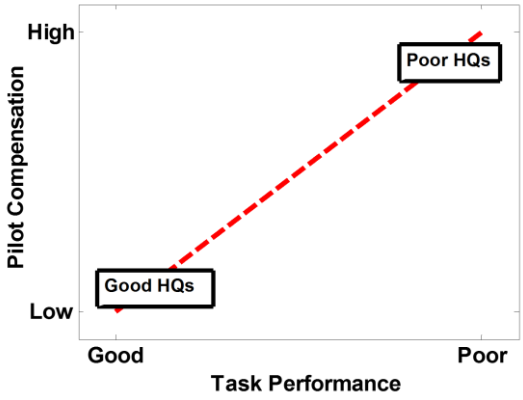


Figure 4-1 - Traditional HQRs Force Ratings on a Diagonal

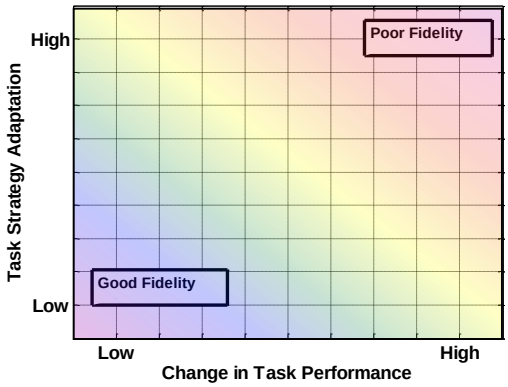


Figure 4-2 - All Combinations Considered in SFR Scale

However, for a fidelity assessment, the entire matrix of possibilities is required. To illustrate this with an example; a simulator with poor far field visuals but good ground texture may allow the pilot to achieve a nap-of-the-earth task to the same performance standard in the aircraft but using a very different scan strategy. In this case, the pilot would be in the top left hand corner of Figure 4-2.

A matrix was developed at the end of the exploratory trial to determine where the ratings and Levels appeared in a matrix of possible combinations of comparative task performance and task strategy adaptation (see Table 4-4). Numbers in the top left of each cell denote SFRs. It was seen that the combinations of negligible compensation and less than comparable performance are not accounted for in the scale. These are the combinations that arise when a poor cueing environment limit the pilot adaptability. This issue was left unresolved at the end of the exploratory trial and was to be contemplated and amended during follow up trials. The finalised matrix that resulted from the follow up trials in section 4.7 is shown in Figure 4-7.

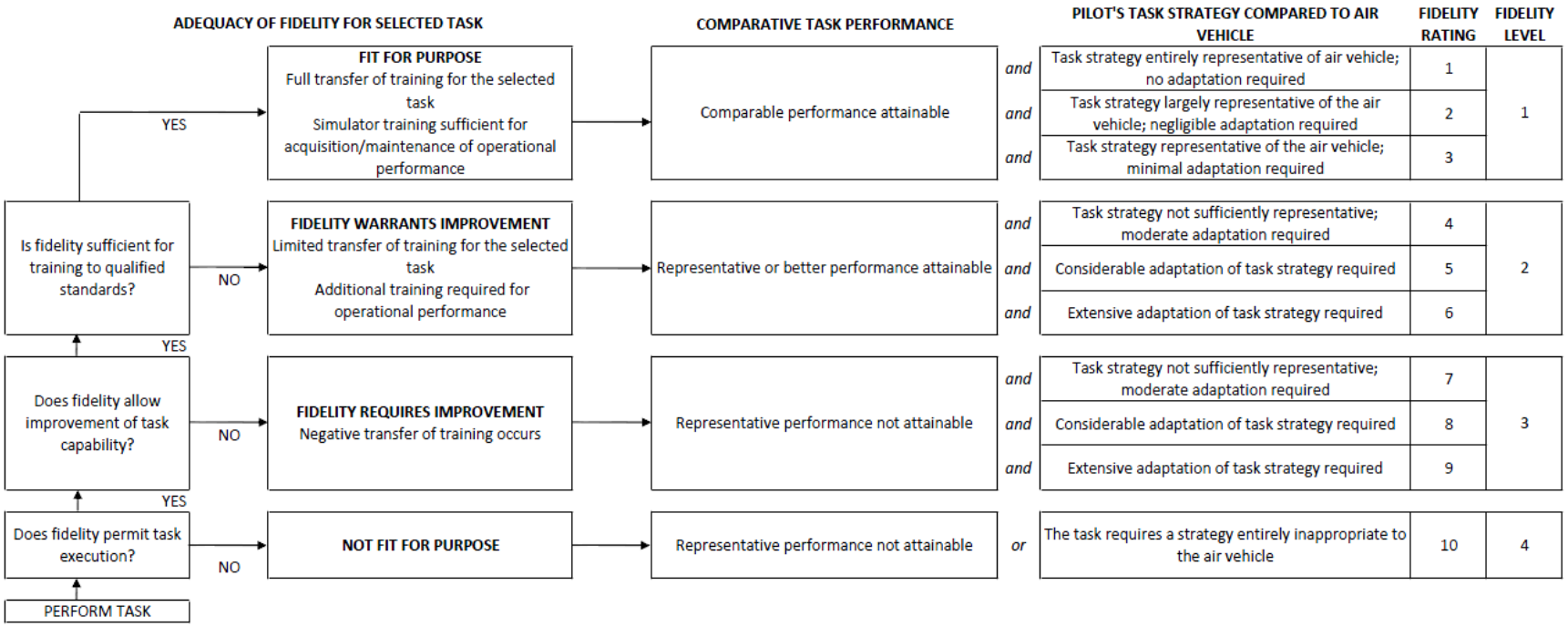


Figure 4-3 - Simulation Fidelity Rating (SFR) Scale - Issue 3 (*Not Final Version*)

Table 4-4 - Initial SFR Matrix

		ADAPTATION			
		Negligible/Minimal	Moderate	Considerable	Extensive
PERFORMANCE	Comparative	1,2,3 LEVEL I	4	5	6
	Representative	? LEVEL II	4	5	6
	Not Representative	? LEVEL III	7	8	9

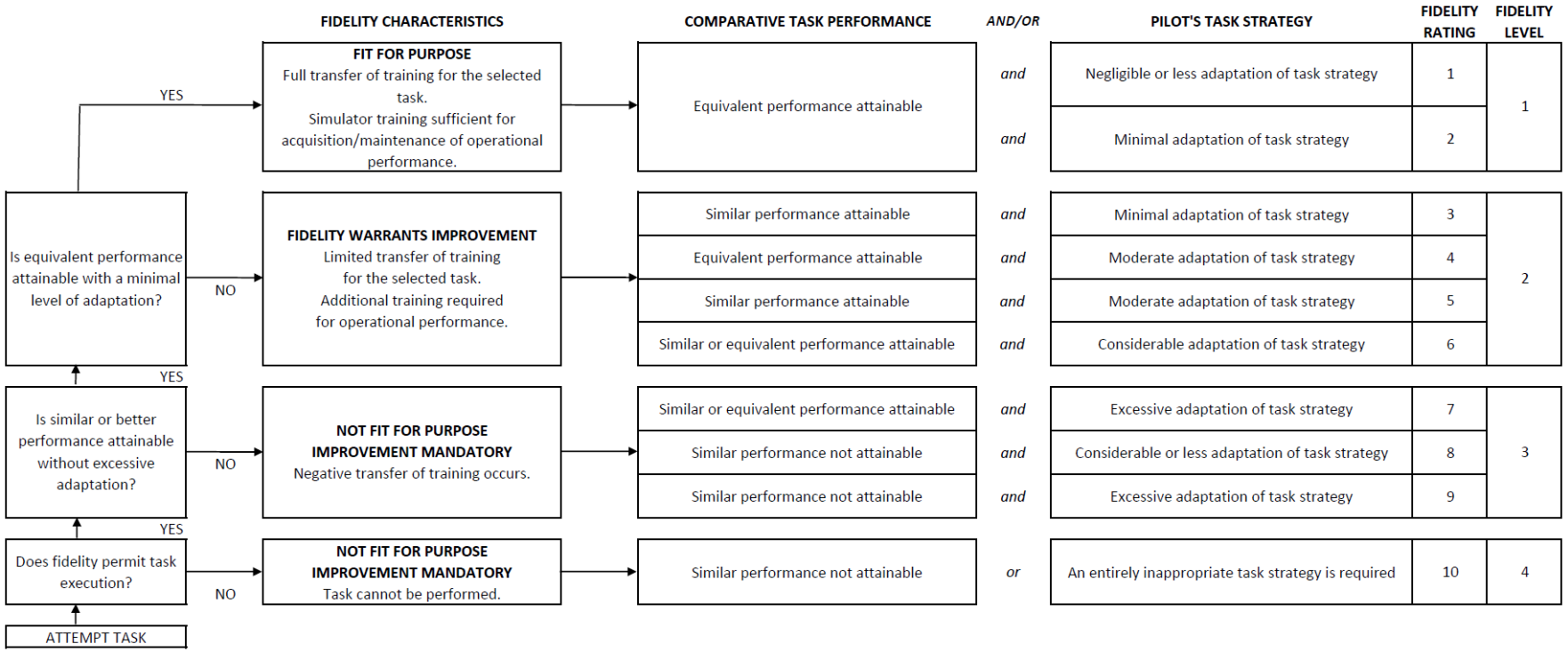


Figure 4-4 - The Simulation Fidelity Rating Scale FINAL VERSION

4.6 Guidance Material

In the exploratory trial, the accompanying guidance material began to take shape. This included guidance regarding the interpretation of terminology in the SFR scale as well as considerations (beyond definition of the purpose) to ensuring validity of the ratings obtained. This initial guidance material is outlined below.

1. *Interpretation of Terminology*

Pilots are not fully able to ascertain the source of fidelity deficiencies. However, they are able to introspect on his performance and the way in which he behaved. The term 'task strategy' was chosen over 'control technique' or 'workload' as it encompasses a broader range of activities. It was decided that the way in which the task strategy differed between flight and simulation would be referred to as task strategy adaptation. The descriptors of task strategy adaptation were agreed upon as 'none', 'negligible', 'minimal', 'moderate', 'considerable' and 'extensive', following the compensation descriptors in the HQR scale.

It was noted that adaptation of task strategy can be subtle, but may represent a significant factor in the assessment of fidelity, for example, the artefacts in the visual scene the pilot uses for cues or the shaping of control inputs. Task strategy adaptation refers to all aspects of the pilot's behaviours related to the task; not only the pilots control inputs but also scan strategy, cognitive workload and utilisation of cues. Examples of adaptation can include such behavioural variations as:

- Feeling a necessity to make reference to an instrument.
- Dissociating or neglecting ineffectual cues (such as poor motion cueing).
- Variations in proprioceptive cues associated with differences in flight control, mechanical characteristics or seating position.
- Changes in head movement caused by differences in cockpit field of view.

The evaluation pilot should consider variations in parameters such as, but not limited to, those depicted in Figure 4-5.

Figure 4-5 - Components of Task Strategy

When using tasks with the ADS-33E-PRF MTE format for fidelity assessment, the well-defined 'desired' and 'adequate' performance requirements associated with such manoeuvres can be used to quantify task performance in flight and simulation. Thus, changes in task performance can be clearly monitored and inferred. The comparative performance descriptors were agreed upon as 'comparable performance attainable', 'representative performance attainable' and 'representative performance not attainable'. Guidance as to the definitions is given in Table 4-5. While this is a subjective decision, the pilot was encouraged to request performance feedback from the test engineers if unsure themselves.

Experience has found that it is necessary to clearly brief the evaluation pilot with regards to these descriptors. In particular, it must be made clear that an improvement in performance is as much a negative result in terms of fidelity as a degradation in performance. For example, if the pilot achieves adequate performance in the aircraft but achieves desired in the simulator, then 'comparable' performance between flight and simulation has not been attained. This is an important briefing point as the pilot may lean to awarding more leniently in cases where the handling qualities are favourable due to vested interest - this is due to an inherent tendency to believe that if the simulator flies well then it must be a good simulator. However, if the simulator flies better than the aircraft, then there is potential for under training (in a training simulator) or underestimation of pilot workload (in design and research simulators), which could be dangerous.

Table 4-5 - Guidance for Pilots on Terminology of Comparative Performance Descriptors

Comparable Performance Attainable	The same level of task performance (desired, adequate etc.) is achieved for all defined parameters in simulator and flight with only small variations across the defined parameters.
Representative Performance Attainable	There are no large single variations in task performance, or, there are no combinations of multiple moderate variations across the defined parameters.
Representative Performance Not Attainable	Any large single variation in task performance, or multiple moderate variations, will put the comparison of performance into this category.

2. Test Conditions

When assessing the simulator against the aircraft the test pilot noted that a valid SFR must entail an assessment of the simulator under the same conditions as the truth data was flown in the aircraft. The test conditions include parameters such as lighting, winds, turbulence, and vehicle loading. Therefore, where actual flight precedes simulator flight, these conditions should be noted and replicated in the simulation. Where simulation experience precedes flight experience, the best effort should be made to forecast conditions or use 'typical' conditions for the environment to be flown in.

3. Pilot Training and Experience

It was agreed that qualified test pilots would make the most suitable fidelity evaluation pilots as they have been trained to reflect on not only the performance and behaviour of the aircraft but also introspect on their own behaviour. It is widely acknowledged that a degree of specialised training and experience is required for a pilot to meaningfully apply the HQR scale. Similarly, pilots applying SFRs require a degree of training to award meaningful SFRs.

4. Duration of Simulator Exposure

In the exploratory trial the test pilot was required to give ratings with reference to the real aircraft. The pilot commented that calling upon recollections of flying the aircraft as a frame of reference became more difficult as he became more accustomed to the simulation through repeated exposure. This lead to the conclusion that there is a limited time frame in which the pilot can give meaningful SFRs.

After this exploratory trial, guidelines for application of the SFR scale began to be developed and pilot briefing documents were written. The SFR methodology was then ready to be used in earnest with more test pilots to ensure robustness to use by the wider test pilot community.

5. Task Definition

The task may be defined as the training manoeuvre/procedure, accompanied by a set of performance requirements. In a HQR evaluation, a Mission Task Element (MTE) specification consists of the target manoeuvre profile alongside a set of 'desired', and 'adequate' performance tolerances for each element of the manoeuvre profile (e.g. height, airspeed, heading etc.), where the achievement of a certain category of performance assists the pilot with determining the Level of handling qualities of the aircraft. The same style of task definition is adopted for an SFR evaluation, where the comparison of the achieved desired/adequate/beyond adequate performance between flight and simulator assists the evaluating pilot with the judgement of comparative performance.

The requirements for the assessment tasks are that it be aggressive, operationally relevant, nap-of-the-earth, multi-axis repeatable and that the performance can be easily determined. The Quick-Stop Turn is a new manoeuvre that was developed as part of the research specifically for simulation fidelity assessment. It was designed to be a combination of the Level turn, downwind approach and descent training tasks. This task was developed and trained in flight by two qualified test pilot and three engineers. The radius of the turn, required end point and performance tolerances were altered to increase aggression. The aggression was set to a point where the pilots could achieve desired performance but required multiple training runs to do so.

The Quick-Stop Turn course is shown in Figure 4-6.The manoeuvre begins with the aircraft in stabilised flight at 50ft and 30kt airspeed by point the start point. After which the pilot initiates a banked level turn, maintaining height and airspeed throughout the turn. The pilot must not exceed the longitudinal position marked by the cones. The pilot must maintain height and airspeed while turning into the wind until passing through the first gate. The pilot must roll out of the turn level with the Accel-Decel centre line cones and come to a hover by the same longitudinal position the turn was started at, denoted by the cones at 10ft. The Deceleration tolerances take effect as the pilot passes through the second gate. The performance criteria are summarised in Table 4-6.

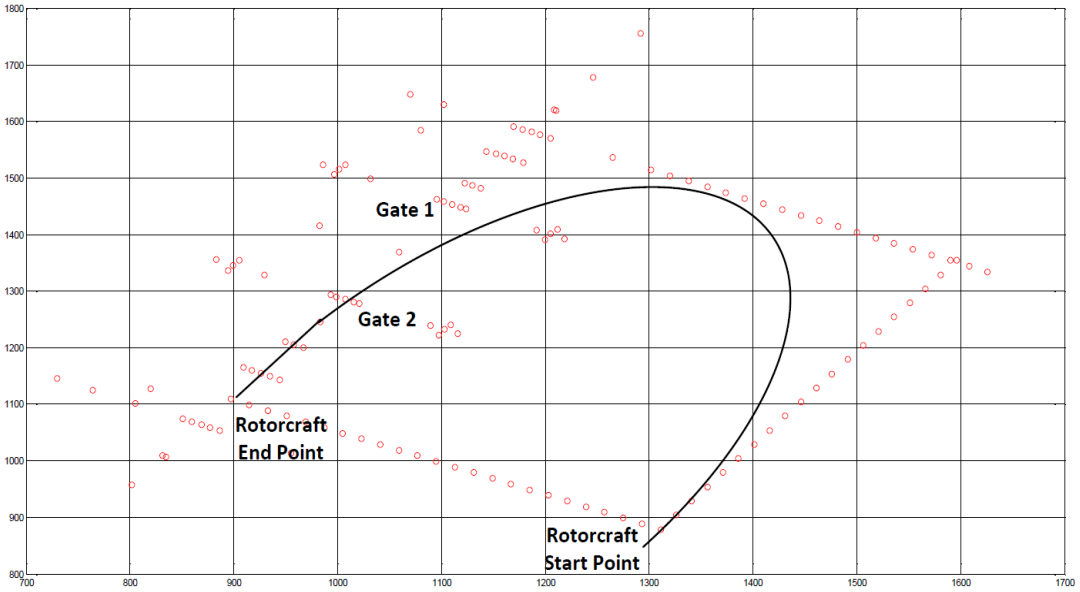


Figure 4-6 - Quick-Stop Turn Course

Table 4-6 - Quick-Stop Turn Performance Criteria

Performance Criteria	Desired	Adequate
Maintain heading within ±X ° during deceleration	10	15
Maintain lateral track within ±X ft during deceleration	10	20
Maintain airspeed within ±X kts throughout 150 degrees of turn	5	10
Maintain height throughout 150 degrees of turn	5	10
Longitudinal tolerance on the turn is plus zero, minus X feet	20	50
Longitudinal tolerance on the final hover point is plus zero, minus X feet	20	50

4.7 Further Testing

SFR Scale issue 3 (Figure 4-3) evolved into the final version of the SFR scale (Figure 4-4) as a result of the piloted simulation trials with seven new test pilots to support the definition of quantitative fidelity tolerances for pitch/roll coupling error and transport delay. These tests are fully described in the next chapter. What is of interest here is that the SFR scale evolved with response to feedback from these test pilots. Table 4-7 summarises this evolution and the major changes are detailed below.

Table 4-7 - Evolution of the SFR Scale

SFR iss.	Comments	Changes made
A,B, C	**All three scales trialled by pilot A on 31/01/11** Scale B more effective than Scale A as working from bad to good avoids over lenient ratings. Erdos scale layout is more intuitive for the pilot.	Erdos Scale layout adopted, questions in first row tweaked for 'general' and 'specific' purpose. 'and' and 'or's modified in level 2 and 3
1	**This scale was used with pilot A on 01/02/11 & 02/0211** Agreement that Level 1 fidelity implies fitness for purpose. For this to occur the performance must be comparable and the strategy adaptation minimal. Questionnaire also updated	Section on 'fidelity deficiencies' entered into the scale and performance and task strategy columns merged
2	**This scale was briefly used with Pilot A on 03/02/11** Extra column is to mimic the 'system characteristics' column in the HQR scale to aid engineer in detecting deficiencies.	Alteration in initial questions to focus on training benefit, use of 'sufficient' and 'improvement of task capability' . 'ands' and 'ors' modified.
3	**This was the scale used with Pilot A and Pilot B** Definitions of comparable and representative performance outlined. Agreed that for the simulation to be level 2 all training must be positive or neutral, no negative transfer of training. Development of a matrix highlighted missing combinations of adaptation and comparative performance.	Initial questions changed to mirror the HQ questions. restructuring of performance in each fidelity level to incorporate the content of the fidelity matrix.
4	**This was the scale used with Pilot B** In this scale there is a new hierarchy developed - that unrepresentative performance is more of a deficiency than extensive adaptation. However, this remains an ongoing debate within the team. There are also concerns within level 3 of this scale in that it is very detailed for a simulation which is unfit for purpose	Initial questions changed to require pilot to reflect on comparative performance and task strategy adaptation. Level 3 simplified from 5 options to 3
5	**This was the scale used with Pilot B** Unrepresentative performance remains a larger deficiency than extensive adaptation in this scale. Debate still ongoing	The word 'extensive' has been replaced with 'intolerable' in the pilot adaptation column as 'extensive' did not seem to correspond with the idea of 'maximum tolerable'
6	**This scale was used by Pilot C and Pilot D** Scale 6 aims to reduce confusion within level 2 by rephrasing 'representative or comparable performance' to 'representative performance'. However, it was felt that this deviated from the want to cover all the cells in the matrix and may cause confusion	'Similar' and 'equivalent' now replace 'representative' and 'comparable' for performance descriptors due to vagueness and 'excessive' replaces 'intolerable'. Control strategy column now limited to adaptation with no mention of 'representative' strategy
7	**This scale was used by Pilot C and Pilot D** Semantic changes made due to visit from the RWTES. Pilots did not feel that 'tolerable adaptation' was meaningful. Representative became a confusing word to the pilots as it was used for differing levels of 'goodness' in the performance and strategy columns. The debate over which is more critical, lack of representative performance or excessive adaptation is still ongoing	SFR 1 and SFR 2 now become SFR 1 (negligible or less adaptation) SFR 3 is now level 2. SFR 4a becomes 4 and SFR 4b becomes 3
8	**This scale was used by Pilot A, B, E, F and G has been published for external use.** In level 2 it was deemed that an increase in adaptation was more of a concern than loss of performance.	_____

i. Semantics of the SFR scale evolved to ensure that a common, unambiguous understanding developed (issue 1 through to issue 6)

ii. It was noted in the exploratory trial that SFR scale issue 3 (Figure 4-3) did not account for all combinations of task strategy adaptation and comparative performance. To assess this discrepancy, a simulation fidelity matrix (see Figure 4-7) was assembled to include all possible combinations of task strategy adaptation and comparative performance. It was decided which combinations should lie in what Level of Fidelity and an individual rating assigned to each combination (issue 3 to issue 4).

iii. The original requirements for the scale stated that the output from the scale should be a measure of the simulator's utility. Asking the pilot to reflect directly on the fitness for purpose of the simulation at the first stage of the decision-making algorithm goes against this requirement. Therefore, it was decided that the questions leading into the various fidelity levels should be focused on the pilot cognitively rating their comparative task performance between flight and simulation and required task strategy adaptation between the flight and simulation (issue 4 to issue 5)

iv. It was suggested that there is little distinction between none and negligible adaptation and therefore having two SFRs to cover these options is superfluous. SFR 2 and SFR 1 were therefore combined (issue 6 to issue 7).

v. It was proposed that in Level 2 fidelity, a deviation in task strategy represents a larger deficiency than a deviation in task performance, though the opposite is suggested to be true in level 3 due to safety concerns about loss of control. (issue 3 to 4 and issue 7 to issue 8)

The changes in semantics and structure of the SFR scale lead to the finalised SFR matrix, shown in Figure 4-7, and the final version of the SFR scale (issue 8) shown in Figure 4-4. The finalised questionnaire is shown in Appendix A in Table A-2. It should be noted that all SFRs presented in this work that were collected using versions of the SFR scale previous to issue 8 have been converted to the equivalent SFR as per the scale shown in Figure 4-4.

		ADAPTATION OF TASK STRATEGY			
COMPARATIVE TASK PERFORMANCE		Negligible/ minimal	Moderate	Considerable	Excessive
	Equivalent	1,2	4	6	7
	Similar	3	5	6	7
	Dissimilar	8	8	8	9

Figure 4-7 - The Simulation Fidelity Matrix (Final)

The next chapter details the piloted simulation trials and presents the SFRs alongside quantitative analysis of pilot control activity and performance. Following this will be a discussion around the interpretation of these results and lessons learned regarding best practices of the SFR scale.

4.8 Case Study - Effects of Model Tuning on Perceptual Fidelity

The SFR scale has been utilised with a test pilot to determine whether the effects of rotor modelling parameters, discussed in chapter 3, are perceived by the pilot and captured in SFRs.

Each model was tested in three different flying tasks. The Quick-Stop Turn (as detailed in section 4.6), the ADS-33E-PRF. The Precision Hover MTE [19] and the Bob-Up MTE, which is a variation on the ADS-33E-PRF Vertical Remark MTE [19] to allow for dimensions of the NRC test course and resulting in a 25ft height change (for suitable task aggression). The Precision Hover MTE requires the pilot to transition at 45° from one hover point to another and then maintain the new hover position for 30 seconds. The Bob-Up MTE requires the pilot to translate vertically from one hover to another and maintain the new hover position for 2 seconds before returning to the original position.

The test pilot flew three Sorties. In each Sortie the pilot flew one of the three manoeuvres with all three models. The test matrix was randomised to mitigate against the effects of learning and assimilation in the results.

Sortie	Manoeuvre	Model 1	Model 2	Model 3
1	Precision Hover	LS	LS Mod	all control
2	Bob-Up	LS Mod	all control	LS
3	Quick-Stop Turn	All control	LS Mod	LS

4.8.1 Comparative HQRs

The pilot in general found all three simulations to be more difficult than flight. He commented on tendency of all the models to PIO more than the aircraft, particularly in the lateral axis.

Overall, the tuned FB412_HA model was perceived to have the most similar handling qualities as flight and the FB412_LS model the most disparate. The modifications made to the rotor model of the LS model, to generate the LS_MOD model, led the pilot to perceive that the handling qualities were closer to flight in the hover. However, these modifications had no effect on perceived handling qualities in forward flight manoeuvring.

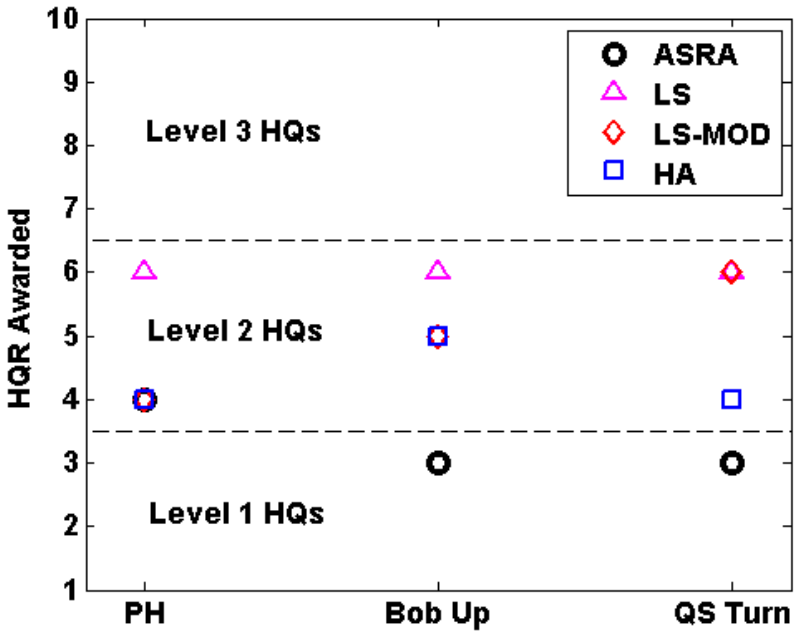


Figure 4-8 - SFRs Awarded for 3 Bell 412 ASRA Model Variants as Initial Skills Trainer in 3 Flying Tasks

4.8.2 SFRs

The SFRs shown in Figure 4-9 indicate that the pilot perceived the HA model to have much more training utility than the LS model, particularly in hover flight. The SFRs show that the pilot perceived the simulation with the LS model to have negative transfer of training for hover tasks. In both hover tasks, the modification to the rotor model lead to a significant increase in perceived fidelity. However, the forward speed SFR was not improved, with the pilot noting considerable adaptation rather than moderate adaptation. This agrees with the predicted fidelity assessment presented in Chapter 3. The variation in SFRs across the tasks confirms that perceptual fidelity is dependent on the task.

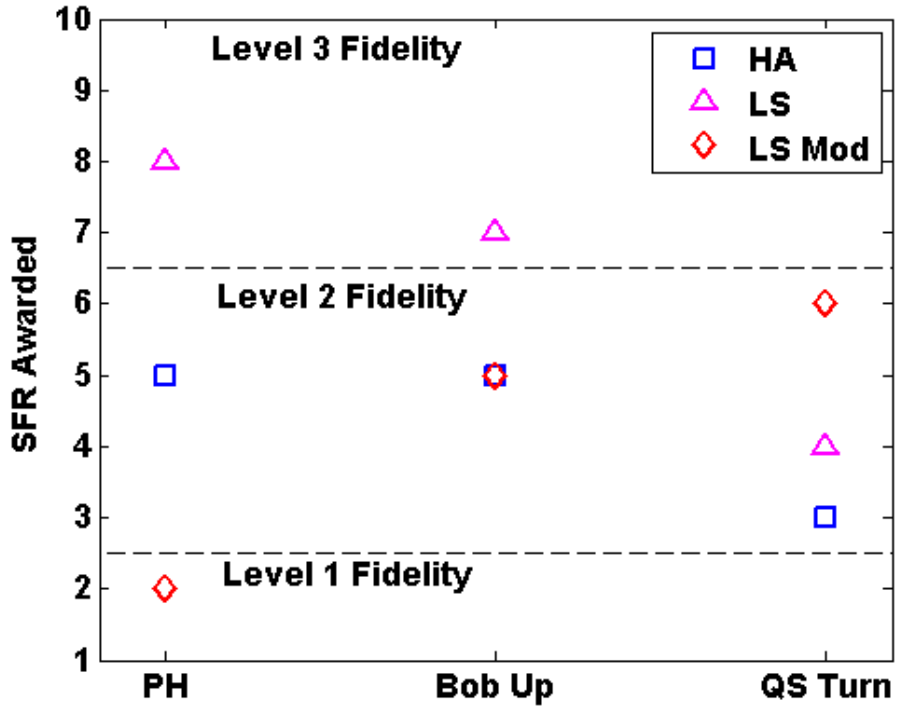


Figure 4-9 - SFRs Awarded for Three Variants of the Bell 412 ASRA Bare Airframe Model

4.9 Conclusions

1) A novel subjective rating scale for the assessment of the perceptual fidelity of integrated simulations has been developed collaboratively with others at the University of Liverpool and the National Research Council of Canada. The Simulation Fidelity Rating (SFR) scale provides a subjective assessment tool to augment quantitative fidelity assessment or to act as an alternative where no quantitative data is available. This chapter has detailed the author's contributions to the development of this scale as well as the methodology, limitations and guidance material for its most effective use including guidance of pilot briefing and task definition.

2) A questionnaire has also been developed by the collaborative team to complement the use of the SFR scale. This provides a formal template for the gathering of detailed feedback that is important for identification of deficiencies that can be improved.

3) An original application of the SFR scale that was identified by the author has been introduced in this Chapter and is used in the experimental work of Chapter 5. This application uses the SFR scale in controlled simulation experiments to determine the relationship between changes to a chosen simulation parameter and perceptual fidelity of the overall simulation.

4) A case study conducted by the author has been presented where the SFR scale has been used to determine the relative perceptual fidelity of three model variants of the Bell 412 ASRA bare airframe vehicle (as introduced in Chapter 3). It was concluded that the pilot generally preferred the tuned (HA) model over the physically accurate baseline (LS) model. The modifications made to the physically correct model improved perceptual fidelity for low speed tasks but degraded perceptual fidelity for high speed tasks. This agrees with the predicted fidelity assessment presented in Chapter 3. In the precision hover task, the LS-mod model was preferred to the tuned model. This case study highlights the utility of the SFR scale and also confirmed that perceptual fidelity is task dependent.

5) Guidance material has been developed by the author to ensure the best utilisation of the SFR scale by others. It has been highlighted that the definition of 'fit for purpose' is dependent on the simulator application (training, design, and certification). Correct definition and briefing of this definition to the assessment pilot is crucial to the successful application of the SFR scale. Thus, a number of examples of definitions of fitness for purpose have been provided. Guidance for pilots regarding the interpretation of the terminology in the SFR scale has been included to aid pilot briefing sessions.

5 PILOTED SIMULATION TRIALS

This Chapter describes the methodology used in a series of piloted simulation trials aimed at assessing the effect of flight model variations on perceptual fidelity. The subjective and objective results from eight test pilots are presented and discussed. Challenges associated with analysis of the pilot-in-the-loop data highlighted in this chapter provide the motivation for the studies presented in subsequent chapters in this thesis.

5.1 Trial Aims and Objectives

In Chapter 2 it was highlighted that the current EASA CS-FSTD (H) quantitative acceptance criteria [14] lack engineering justification for application to rotary wing simulation. Two test cases in particular were highlighted for further investigation – error in off axis response and transport delay.

The effect of inter-axis coupling on perceptual fidelity was chosen for specific investigation as the only guidance provided by the EASA criteria is that the off-axis dynamic response proof of match be of ***"correct trend and magnitude".*** This test case could be used to gather evidence for the development and justification of acceptable/unacceptable limits of errors in off-axis response in response to open-loop control inputs.

The effect of transport delay on perceptual fidelity was also chosen for investigation due to results of a previous study [4] conflicting with the current criteria of 100ms or less for Level D acceptance. This test case could be used to gather evidence to advise the need to update this criterion in future issues of the criteria.

Another aim of this trial was to utilise the newly developed SFR scale to obtain pilot feedback and validate appropriate sensitivity and robustness through comparative analysis with quantitative data and through pilot feedback sessions. This study was suitable for execution of the SFR scale as the aim was to determine the effects of errors in the flight dynamics model on the perception of the fidelity of the integrated system, rather than the effects on the perception of the fidelity of the flight dynamics model itself.

To gather the required data, piloted simulation trials were conducted where, in a random order, increases in either the inter-axis coupling or transport delay were implemented in the flight model. Pilots awarded SFRs against the nominal, baseline simulation as well as HQRs [19]. Time histories of pilot control activity and aircraft states were also recorded for

quantitative comparisons between the baseline and modified simulation. Both the subjective results and quantitative measures of pilot control activity and task performance were analysed to determine at what levels of increase in off-axis response and transport delay the utility of the simulation for training becomes compromised.

5.2 Flight Model Modifications

To ensure that the effects of the flight model on perceptual fidelity were isolated from any other limitations of the HELIFLIGHT-R Simulator [69], a baseline simulation (FB412-RCAH) was utilised as the reference case for these trials. A number of modified flight model simulations were then used as the test cases.

The EASA CS-FSTD(H) criteria stipulate that for a Level D simulation, the transport delay must be 100ms or less after control movement for Level C and Level D acceptance and 150 milliseconds or less for Level A and Level B acceptance. The document defines the transport delay to be;

"The total Synthetic Training Device (STD) system processing time required for an input signal from a primary pilot flight control until motion system, visual system or instrument response. It does not include the characteristic delay of the helicopter to be simulated." [14]

The nominal HELIFLIGHT-R transport delay at the time of testing was approximately 100ms (now reduced to 85) [78]. Therefore, in the context of these trials the characteristic delay of the helicopter to be simulated is 100ms. The transport delay to be investigated is therefore the Additional Transport Delay (ATD). The ATD was implemented using delay blocks in each axis of the control system via CSGE, adding delays of a defined number of time steps between the pilot control inputs being made and the signals being entered into the FBW system, the lateral axis case is shown as an example in Figure 5-1. Each second of real-time simulation consists of 128 time steps, hence 13 time steps will be approximately 100ms transport delay.

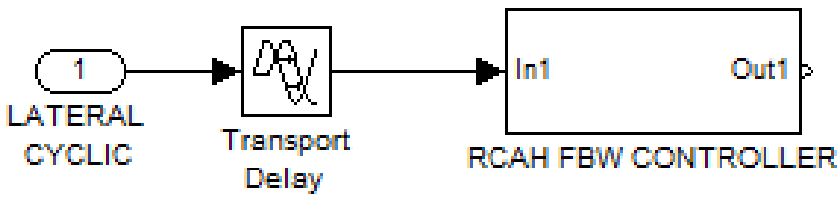


Figure 5-1 - CSGE Implementation of ATD

The bandwidths and phase delays of the various transport delay flight models were calculated using frequency sweeps in FLIGHTLAB, taking into account the 100ms characteristic transport delay of the HELIFLIGHT-R simulator. The pitch and roll results are shown against the ADS-33E-PRF [19] 'all other MTEs' criteria (Figure 5-2 and Figure 5-3). The baseline and ATD=100ms cases lie well within Level 1. The ATD=200ms case is borderline Level 1/Level 2 and the ATD=300ms case lies within Level 2 for both roll and pitch bandwidth. In line with the suggestion that differing handling qualities suggest poor fidelity, as discussed in Chapter 2,the change in HQ level for the ATD≥200ms cases would suggest a fidelity deficiency. It was decided that the levels of variations in models should lead to HQs that span across all three Levels. However, simulation trial workup suggested that models with ATD>300ms caused disorientation and therefore no predicted Level 3 models were used in this trial. It may be the case that cumulative deficiencies in roll and pitch lead to the poorer assigned handling qualities.

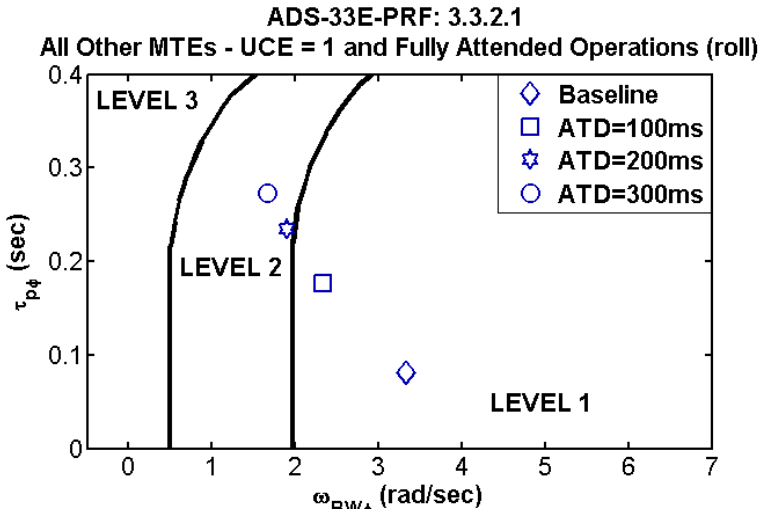


Figure 5-2 - Effect of ATD on Roll Bandwidth

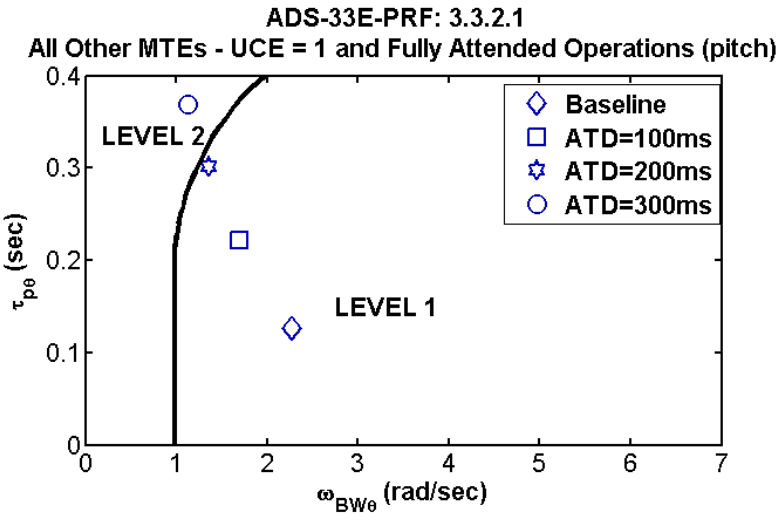


Figure 5-3 - Effect of ATD on Pitch Bandwidth

EASA CS-FSTD(H) requires that, following a step input (e.g. in longitudinal cyclic), the on-axis response (e.g. the pitch rate and pitch attitude) should be within the tolerances shown as the broken lines in Figure 5-4, and the off-axis response (e.g. roll response) should be of 'correct trend and magnitude' (CT&M) [14]. The ambiguity of this requirement has been highlighted as a shortcoming of rotary wing simulation criteria by industry [31] and was therefore chosen as a second test case for investigation in this research.

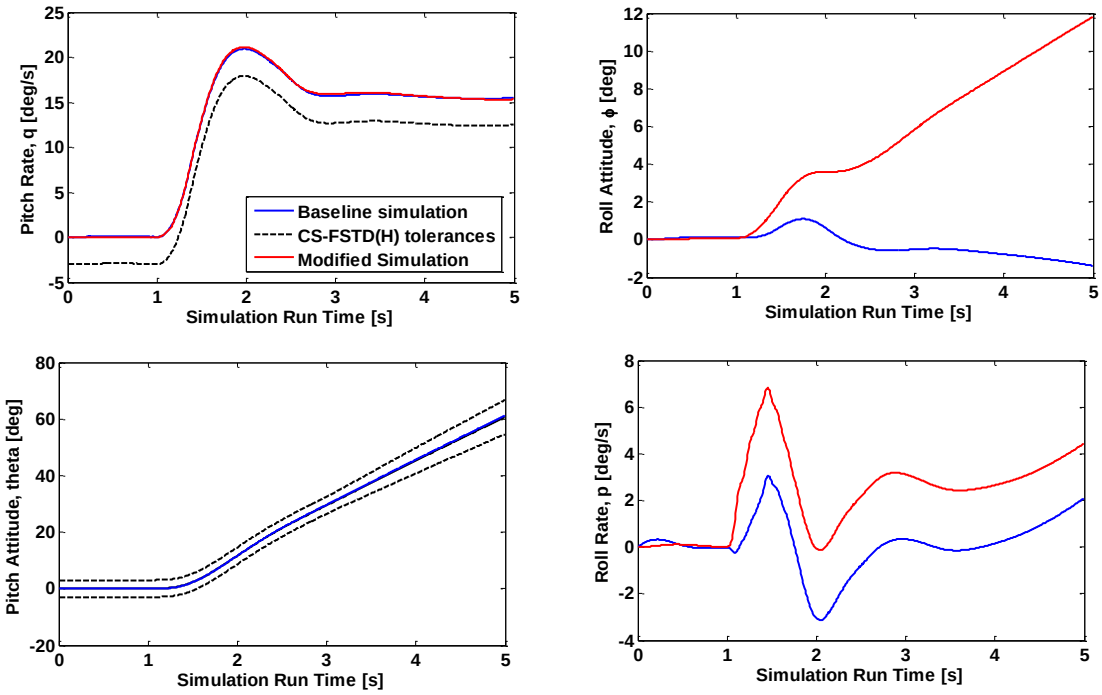


Figure 5-4 - On-Axis and Off-Axis Open Loop Responses to a 2 Inch Step Input in Longitudinal Cyclic (XB)

Off-axis response is largely dictated by inter-axis coupling between two axes, e.g. pitch and roll. ADS-33E-PRF defines the pitch due to roll (or roll due to pitch) inter-axis couplings for aggressive agility as the ratios of peak off-axis response, θ_{pk} (or ϕ_{pk}) to on-axis response to a step control input 4 seconds after the input, ϕ (4) (or θ(4));

$$C_{PR} = \frac{\theta_{pk}}{\phi(4)} \qquad \text{Equation 5-1}$$

$$C_{RP} = \frac{\Delta\phi_{pk}}{\theta(4)} \qquad \text{Equation 5-2}$$

The baseline RCAH model has negligible pitch/roll cross coupling and therefore negligible off-axis response (see Figure 5-4). The additional pitch/roll cross coupling was implemented in the FLIGHTLAB Control System Graphical Editor (CSGE) by directing a proportion of the longitudinal control input, via a proportional gain K_{cc}, into the roll axis as well as the pitch axis and vice versa (as shown in Figure 5-5). This method of implementation ensures that only the off-axis response is affected by the variation in coupling ratio. However, there is a limitation that this modelling technique and does not include the rate effects and higher frequency effects due to rotor dynamics that occur in practice. The unpredictability caused by these effects can be a significant cause of perceived HQ deficiencies.

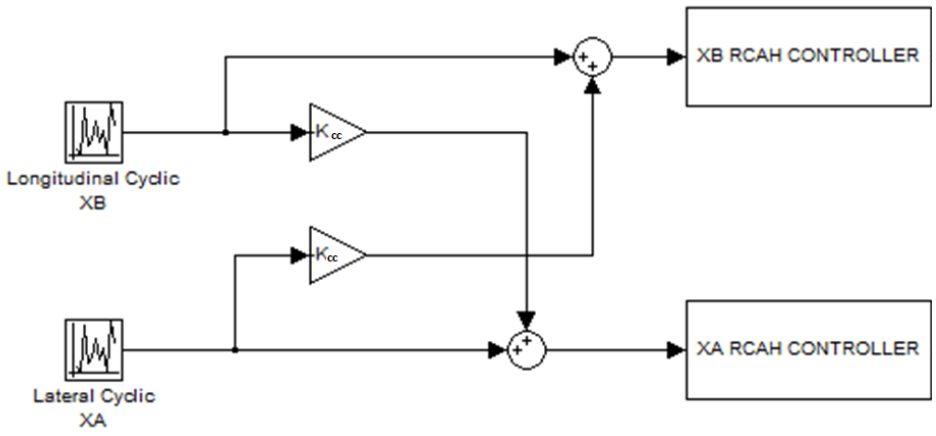


Figure 5-5 - CSGE Implementation of Pitch/Roll Cross Coupling

ADS-33E-PRF requires that for Level 1 HQs C_{RP} and C_{PR} are no greater than 0.25 for Level 1 HQs and no greater than 0.60 for Level 2 HQs. Figure 5-6shows altering the cross coupling gain, K_{cc}, affects the handling qualities (C_{PR} and C_{RP}) of the aircraft in a uniform manner. To obtain a range of models across HQ Levels it was decided that cross coupling gains, K_{cc}, of 0.1, 0.2 and 0.3 should be used. However, during the trials, it was found that the pilots were

less sensitive to changes in C_{PR} and C_{RP} than the predicted HQs, due to the cross/coupling model limitations. Therefore the values of K_{cc} were varied from 0.2 to 1.0 (C_{RP}=0.4 to 2.0) to ensure a range of SFRs was obtained from each pilot as a major aim of this work was to exercise the SFR scale.

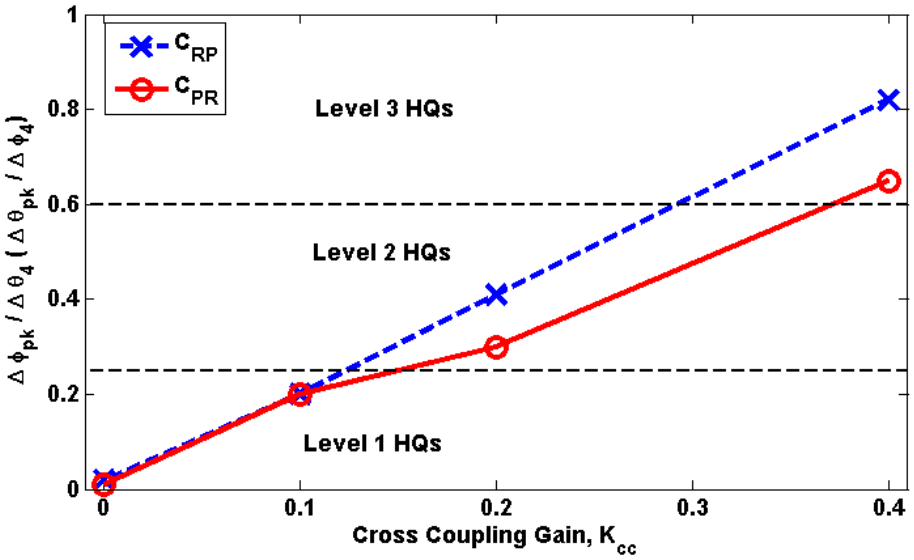


Figure 5-6 - The Effect of Pitch/Roll Cross Coupling on Handling Qualities

5.3 Tasks

Two tasks were used in the trial to assess the hypothesis that fidelity requirements are dependent on training task, as noted by the IWG during the development of ICAO 9625 [16] and by GARTEUR HC AG-12 [17]. As the dynamics of the flight model were the parameters under investigation, flying tasks were appropriate. In particular, ADS-33E-PRF Mission Task Elements (MTEs) were chosen as they have clear performance requirements and are well defined. The Precision Hover (PH) MTE and Accel-Decel (AD) MTE both contain a number of different phases - guidance, capture and stabilisation. However, the two manoeuvres were chosen to complement each other due to their difference. The majority of the PH MTE is in the stabilisation phase and requires a high degree of multi-axis control. Contrastingly, the majority of the AD MTE is spent in the guidance phase and is predominantly a single axis (longitudinal cyclic) task.

The Precision Hover (PH) MTE was chosen due to its multi-axis and closed-loop, compensatory tracking elements. The Acceleration-Deceleration (AD) MTE was chosen to be the second task as a contrasting, predominantly single-axis, and guidance manoeuvre.

The PH MTE is a limited agility ADS-33E-PRF MTE that involves transition to hover and hover maintenance phases. The test course used is shown in Figure 5-7. It should be noted that this is modified from the standard ADS-33E-PRF setup in terms of longitudinal distances between hover point, reference point and hover board point, due to space limitations. The performance criteria are unchanged and detailed in Table 5-1. The manoeuvre is initiated from the hover. From here, the aircraft translates at 45° relative to the heading of the rotorcraft at a ground speed of between 6 and 10 knots, at an altitude of less than 20 feet. The ground track should be such that the rotorcraft will arrive over the target hover point. The hover should be captured in one smooth manoeuvre following the initiation of deceleration – ***"it is not acceptable to accomplish most of the deceleration well before the hover point and then to 'creep up' to the final position'' [19].***

The relationship between the heights of the pole and the hover board is such that, when over the target hover point and aligned with both the marker on the pole and the hover board, the rotorcraft will be at the reference height of 10 feet.

In AD-33E-PRF, it is required that the manoeuvre be flown in moderate winds from the critical direction in the GVE. However, due to the SCAS system, this was deemed superfluous as the RCAH system would trim out any steady state winds that could be modelled.

Table 5-1 - Precision Hover Performance Criteria (GVE)

Criteria	**Desired Perf.**	**Adequate Perf.**
Attain stabilised hover within X seconds of initiation of deceleration	5	8
Maintain a stabilised hover for at least X seconds	30	30
Maintain the longitudinal and lateral position within ±X feet on the ground	3	6
Maintain altitude within ±X feet	2	4
Maintain heading within ±X °	5	10
There shall be no objectionable oscillations during the transition to, or during the hover	Applies	Does not Apply

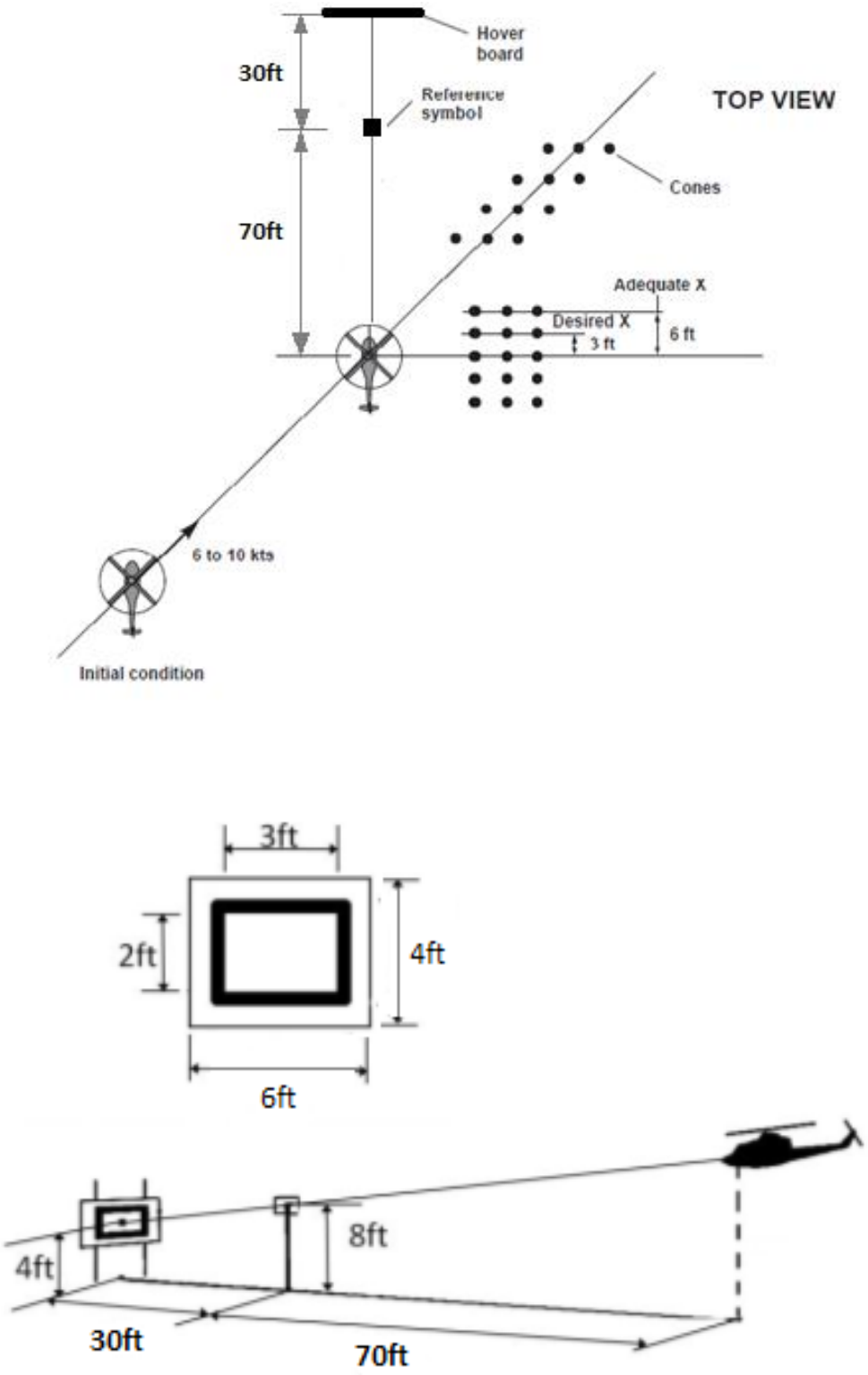


Figure 5-7 - Precision Hover Course [19]

The Accel-Decel (AD) manoeuvre profile is shown in Figure 5-8. The desired track is indicated by a series of cones along the centre-line of the course, with markers to the left and right indicating the boundaries of desired and adequate lateral tracking performance. The available test course distance at the NRC was 800ft. The performance criteria are detailed in Table 5-2 and remain unchanged from the ADS-33E-PRF criteria apart from the requirement for 30° nose up attitude at the end of the manoeuvre. This was relaxed due to the FBW safety trip limits and FoV issues. The manoeuvre is started from a stabilised hover. To initiate the MTE, the pilot should rapidly increase power to approximately 95% of the maximum continuous power, maintaining altitude constant using pitch attitude, and hold collective constant during the acceleration to an airspeed of 40 knots (Relaxed from the ADS-33E-PRF requirement of 50 knots due to space constraints on the NRC flight test course). Upon reaching the target airspeed, the pilot should initiate a deceleration by aggressively reducing the power and holding altitude constant. The peak nose-up attitude should occur just before reaching the final stabilised hover.

Table 5-2 - Performance Criteria for the Accel-Decel MTE

Criteria	Desired	Adequate
*Within X seconds from initiation of the manoeuvre, achieve at least the greater of 95% maximum continuous power or 95% maximum transient limit that can be sustained for the required acceleration, whichever is greater.	1.5	3
Maintain altitude below X feet	50	70
Maintain lateral track within ±X feet	10	20
Maintain heading within ±X °	10	20
Significant increases in power are not allowed until just before the final stabilised hover	Applies	Applies
*Achieve a nose-up pitch attitude during the deceleration of at least X ° above the hover attitude. The maximum nose-up attitude should occur shortly before the hover	30	10
Longitudinal tolerance on the final hover point is +0, -X feet	21	42
Rotor RPM shall remain within the limits of X without undue pilot compensation	OFE	SFE

* May be relaxed due to ASRA FBW safety trip limits

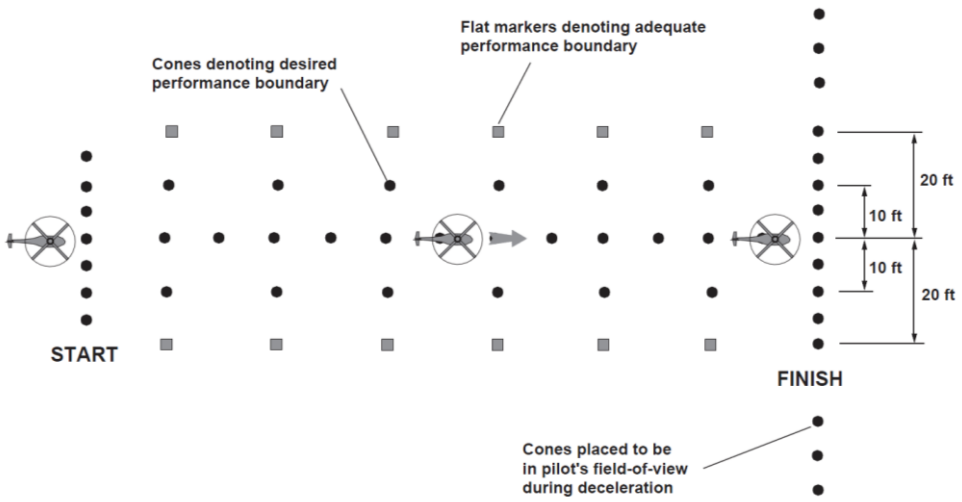


Figure 5-8 - Acceleration-Deceleration Course [19]

5.4 Experimental Methodology

The HQR scale, SFR scale and SFR questionnaire were used to obtain pilot subjective opinion on Handling Qualities of the baseline simulation and modified simulations and also to rate the fitness for purpose of the baseline simulation for training the pilot for flying in the modified simulation (to assess perceptual fidelity). Eight test pilots were used during this test campaign.

Each pilot completed at least four runs of the manoeuvre in the baseline configuration to ensure that they were able to achieve consistent task performance. At this point a HQR for the baseline configuration was awarded by the pilot. The pilot was then informed that a 'modification had been made to the simulation' but was not told the nature of the change, whether it be to the flight model, motion/visual system, control feel etc., and the manoeuvre was re-flown. The SFR questionnaire was then completed and an SFR was obtained after the first run in the modified simulation (recall from Chapter 4 that it is considered that the pilot's greatest sensitivity to variations in simulator changes will be upon first exposure to a specific model/vehicle). The pilot was then asked to complete three more runs in the modified configuration to allow for adaptation of task strategy, prior to awarding a HQR for the modified simulation (and check on the SFR). As well as subjective ratings, data from the simulation trials were analysed to compare performance and pilot strategy between the simulations.

As discussed in the previous chapter, clear pilot briefing of the purpose of the simulation and the scenario in which the ratings are to be given are vital to ensure valid SFRs. For these piloted trials, the pilots were given the following brief:

Purpose:

The purpose of the training in the FTSD is to teach the pilot the flying skills necessary to complete a defined manoeuvre to operational proficiency (either the Precision Hover or Acceleration-Deceleration manoeuvre).

Scenario:

You are to be trained in a rotary wing FSTD (represented by the baseline HELIFLIGHT-R simulation) to fly a defined manoeuvre to the defined standards. Once optimum performance has been achieved, within the limitations of the vehicle dynamics, the training will be considered to be complete. You will then be required to perform the same manoeuvre in the real aircraft (represented by a modified HELIFLIGHT-R simulation). You are to comment on the extent to which the training (represented by the baseline HELIFLIGHT-R simulation) prepared you for execution of the flying task in the real aircraft (represented by a modified HELIFLIGHT-R simulation), highlighting required adaptation of task strategy and changes in performance. You are then asked to use the Simulation Fidelity Rating (SFR) scale to quantify the utility of the FSTD (represented by the baseline HELIFLIGHT-R simulation) as a trainer for the aircraft (represented be a modified HELIFLIGHT-R simulation) for the defined task.

For clarity of terminology, the term 'run' has been used to mean a single completed manoeuvre. The term 'test point' has been used to mean a specific combination of pilot, model variant and manoeuvre. As such, each test point nominally consisted of 8 runs.

Due to time and resource limitations, each pilot was not able to complete each test point. In some cases, pilots required 5 or 6 runs in a model variant until content that consistent performance and strategy had been attained. In other cases, SFRs were taken but there was not sufficient time to conduct repeat runs in the modified models. The right most two columns in Table 5-3 show the number of pilots that flew each test point.

Table 5-3 - Number of Pilots that Completed Each Test Point

Model Type	Model	Number of Pilots who flew test point – Precision Hover	Number of Pilots who flew test point – Accel-Decel
N/A	Baseline	8	8
ΔC_{RP}[nd]	0.2	1	0
	0.4	4	5
	0.8	0	3
	1.0	4	6
	1.6	3	0
	2.0	4	2
ATD [ms]	47	3	0
	100	7	6
	200	5	5
	300	4	3

Three pilots (Pilots A, B and G) flew the RCAH B412 in Flight and Simulation to obtain SFRs, within the same scenario. The awarded SFRs are shown in Table 5-4 and the accompanying pilot comments are shown in

Table 5-5. The agreement between the pilots was good although Pilot B rated consistently lower than Pilots A and G. The time between aircraft and simulator exposure for pilot A and Pilot G was less than one week whereas for Pilot B there was a two month separation.

Additionally, Pilot B had considerable prior experience with the HELIFLIGHT-R simulator, but only a small amount with the Bell 412. Pilot A and Pilot G, on the other hand, was very experienced with the ASRA, but had not flown the HELIFLIGHT-R simulator prior to the SFR evaluations. Both the increased aircraft experience and greater recency of Pilots A and G made them more sensitive to the fidelity issues of the simulator than Pilot B.

Table 5-4 - SFRs Awarded by Pilots for HELIFLIGHT-R as a Skills Acquisition Trainer for Bell 412 ASRA

	SFR awarded for PH	SFR awarded for AD
Pilot A	4	5
Pilot B	2	3
Pilot G	4	5

Table 5-5 - Accompanying Pilot Comments from SFR Questionnaires

	PH	AD
Pilot A	Worse performance in terms of aggression and transition speed, issues with rate perception and target acquisition, busier in lateral cyclic. Cones seem closer together than in Ottawa. Height cues are soft and cockpit has a wider FoV	Struggling to pick up visual cues on the capture - very under-cued, less busy on the cyclic than in aircraft. Decrease in aggression of pitch capture required. Motion cues were also a bit off. Desired in aircraft but adequate in sim. Issues dominated by simulator cueing environment.
Pilot B	UNKNOWN	UNKNOWN
Pilot G	Initially thought PH was an SFR 3 but through discussion realised that 2 controls were easier and two were harder and this does not cancel. Yaw and collective were less active than the aircraft but there was a low grade pitch PIO that wasn't evident in the aircraft.	Accel Decel end capture was different due to pitch up and recovery cueing. Lack of sink in the simulator during acceleration-phase. Collective strategy was easier than in aircraft - less control shaping required.

Upon reflection of the pilot comments and the awarded SFRs, along with the quantitative PoMs presented in Chapter 3, the simulation was deemed fit for purpose for the development of flight model fidelity criteria

In the remaining sections of this chapter, the results of the quantitative and subjective analyses of the simulation trials are presented and discussed.

5.5 Results

The means and spreads of HQRs and SFRs awarded by all test pilots in each test condition are presented as functions of ΔC_{RP} and ATD in this section. The dashed lines represent the boundaries between the various levels of handling qualities and simulation fidelity, defined by the respective rating scales. Because the SFR scale, like the HQR scale, is not a linearly ordinal scale (i.e. a rating of 4 is not implicitly twice as bad as a rating of 2). Therefore the median average has been used rather than the mean as the author believes this provides a more faithful representation of the data. The individual pilot ratings can be found in Appendix B.

Task performance has been quantitatively assessed by analysing aircraft state time histories during the MTEs. Pilot behaviour is much more complicated to fully measure and many methods for doing so were outlined in Chapter 2. In this work, the time histories of control

activity in lateral cyclic (XA), longitudinal cyclic (XB), collective (XC) and pedals (XP), were used for pilot behavioural assessment.

The effect of flight model variations on task performance and control activity was assessed across all runs. The data files for each run were manually inspected for anomalies such as aborted runs and runs where the start or end of the run had not been recorded. The data files were then cropped to ensure that only the manoeuvre itself was captured. This was an important step as any additional run time where pilots were inactive on the controls or repositioning could bias the analysis.

The precision hover MTE is comprised of two distinct phases; the transition to the hover board and the 30 second hover maintenance phase. The performance requirements in these two phases are different and furthermore, the task strategy required in each phase is also disparate. Therefore, the cropped PH data files were split into two files to allow the control activity and performance for the PH transition and maintenance phases to be analysed separately. The point at which the transition ended and the hover maintenance began in each test run was determined subjectively from inspection of the lateral and longitudinal velocity time histories and other aircraft states where necessary. The example in Figure 5-9 serves to highlight that the transition phase includes the deceleration to the hover board.

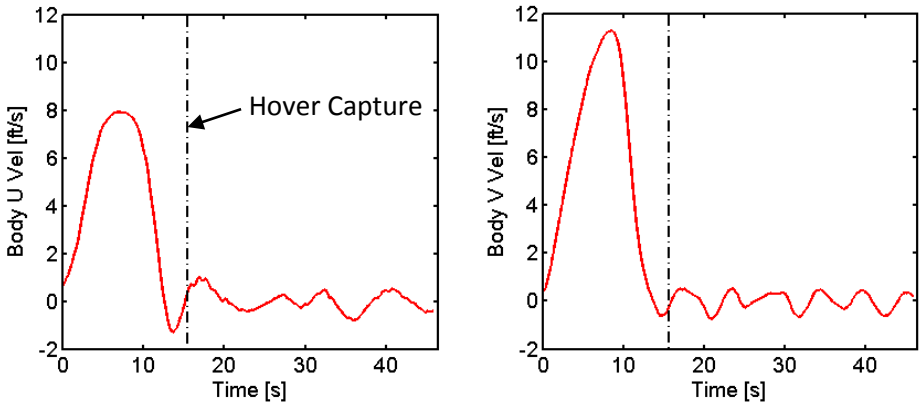


Figure 5-9 - Example of Procedure for Dividing the Precision Hover into Two Phases; Body-Referenced Aircraft U and V Velocities Relative to the Ground

For the task performance assessment, measurements that represented the ADS-33E-PRF performance criteria were used. Where the pilot aimed to keep a parameter constant (e.g. lateral position during hover), the Root-Mean-Square (RMS) of the parameter from the mean was determined to provide an indication of the variation in the parameter. All the metrics used for assessment of task performance are shown in Table 5-6. The deceleration

time was taken as the difference in time between the end of the transition data file and the time at which maximum ground speed occurred.

Table 5-6 - Performance Metrics for Quantitative Analysis of Task Performance

AD	Max pitch attitude	RMS lateral position	RMS heading	Overshoot of stop point
PH (transition)	Max ground speed	Deceleration Time	-	-
PH (maintenance)	RMS lateral position	RMS longitudinal position	RMS Height	RMS Heading

The control activity from each run has been presented for each axis: lateral cyclic (XA), longitudinal cyclic (XB), collective (XC), and pedals (XP), in terms of a number of closed-loop control activity metrics that were introduced in Chapter 2:

1. Number of control points per second (attps), f_{η}
2. Mean control attack rate (attrt), $\bar{\dot{\eta}}_{pk}$
3. Mean control deflection (dsp), $\bar{\eta}$
4. Cut-off frequency (COF) , ω_{co}

The MATLAB functions that were developed for calculating these values can be found in Appendix D.1 and Appendix D.2.

To ensure that the data were captured in their entirety, the control activity for each test condition has been presented in terms of box plots with overlaid mean values for each metric. An example is shown in Figure 5-10. All metrics for all test cases are shown in Appendix C. The main points of the analysis are drawn out below using examples from the full data set.

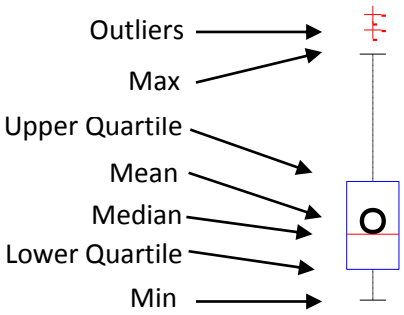


Figure 5-10- Example Data Presentation - XP Attack Per Second, F_n, In Precision Hover Transition

5.5.1 Predicted vs. Assigned Handling Qualities

The baseline model was predicted to be Level 1 for both the Precision Hover and Accel-Decel MTEs (Figure 5-2, Figure 5-3, and Figure 5-6). However, Figure 5-12 shows the spread of baseline HQRs for the Precision Hover manoeuvre to be between HQR=2 and HQR=5. 39 HQRs were obtained for the baseline model. 87% of these ratings included desired performance and 72% of these ratings were Level 1; there were five HQR=5 ratings. Pilot H awarded a HQR=5 during his first test point on the first day of exposure to HELIFLIGHT-R simulator, suggesting proficiency may not have been reached and therefore the HQR is not valid. Similarly, HQR=5 ratings given by Pilot D and Pilot F were from the first sorties of their respective trials. Therefore, in general, the assigned HQRs were aligned with the predicted HQRs for the PH MTE.

Figure 5-11 shows the spread of baseline HQRs for the Acceleration-Deceleration manoeuvre to be between HQR=3 and HQR=6. 26 HQRs were obtained, 85% of which lay within Level 2. Pilots noted that the fly-by-wire pitch rate trip limits of the aircraft were often breached during the deceleration phase of the AD MTE, and they found it difficult to achieve consistent performance due to lack of cueing, particularly in the vertical and lateral axes, at hover capture. This may be the cause of the discrepancies between the predicted (Level 1) and assigned (Level 2) HQRs.

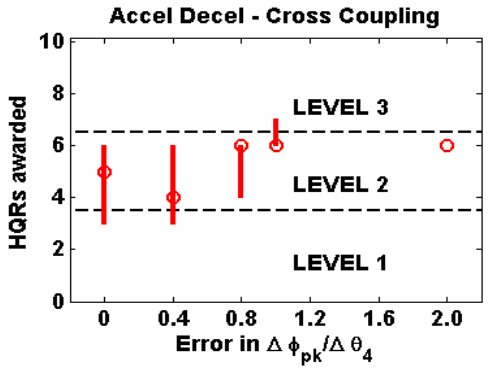


Figure 5-11 - HQRs Awarded in the Accel-Decel as a Function of Cross Coupling Error Parameter

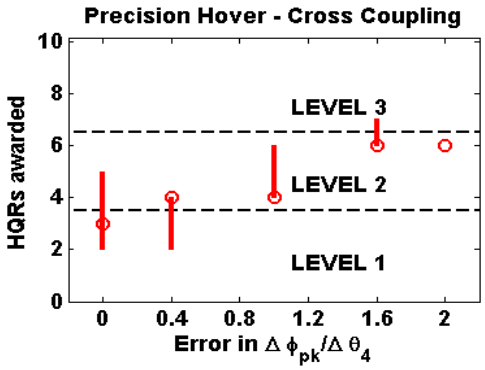


Figure 5-12 - HQRs Awarded in the Precision Hover as a Function of Cross Coupling Error Parameter

It was predicted that models with $\Delta C_{RP} > 0.3$ would exhibit Level 3 HQs (Figure 5-6). However, the required variations in ΔC_{RP} to obtain any Level 3 HQRs was much higher (Figure 5-11 and Figure 5-12). In addition, it was predicted that at ATD=200ms, the HQs would be borderline Level 1/Level 2 and ATD=300ms would result in Level 2 HQs (Figure 5-2 and Figure 5-3). However, the pilots were more sensitive than this with the HQRs for the 200ms lying well within Level 2 and the HQRs for ATD=300ms pushing into Level 3 (Figure 5-13 and Figure 5-14).

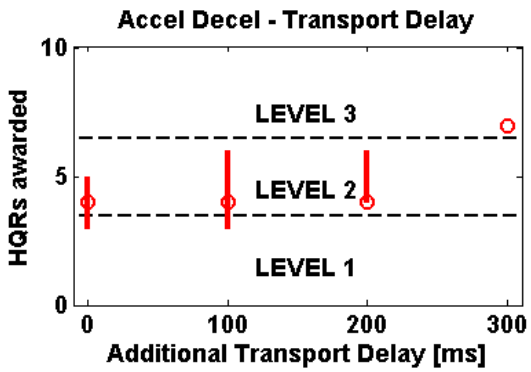


Figure 5-13 - HQRs Awarded in the Accel-Decel as a Function of ATD

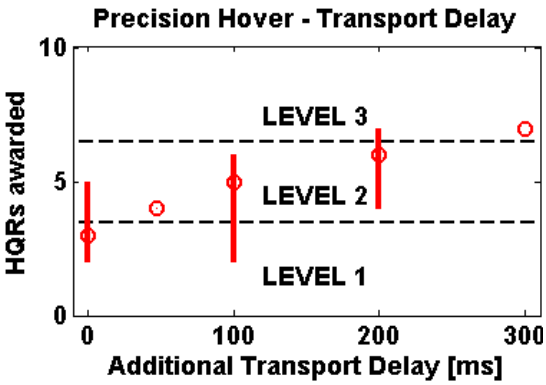


Figure 5-14 - HQRs Awarded in the Precision Hover as a Function of ATD

Inspection of Figure 5-15 reveals that the change in SFR between incremental test cases is larger than that of the equivalent changes in HQRs. A specific example can be drawn (refer to Appendix B, Table B-6 and Table B-8); Pilot A awarded a HQR=4.5 (a special case where desired performance is achievable but requires more than moderate compensation) for the baseline Accel-Decel task and awarded a HQR=5 for the ATD=100ms AD task. However, the pilot awarded an SFR=5 for this test point which implies that the simulation was not fully fit for purpose. This confirms that similar HQRs are not sufficient to suggest high fidelity, as noted by Padfield et al [63].

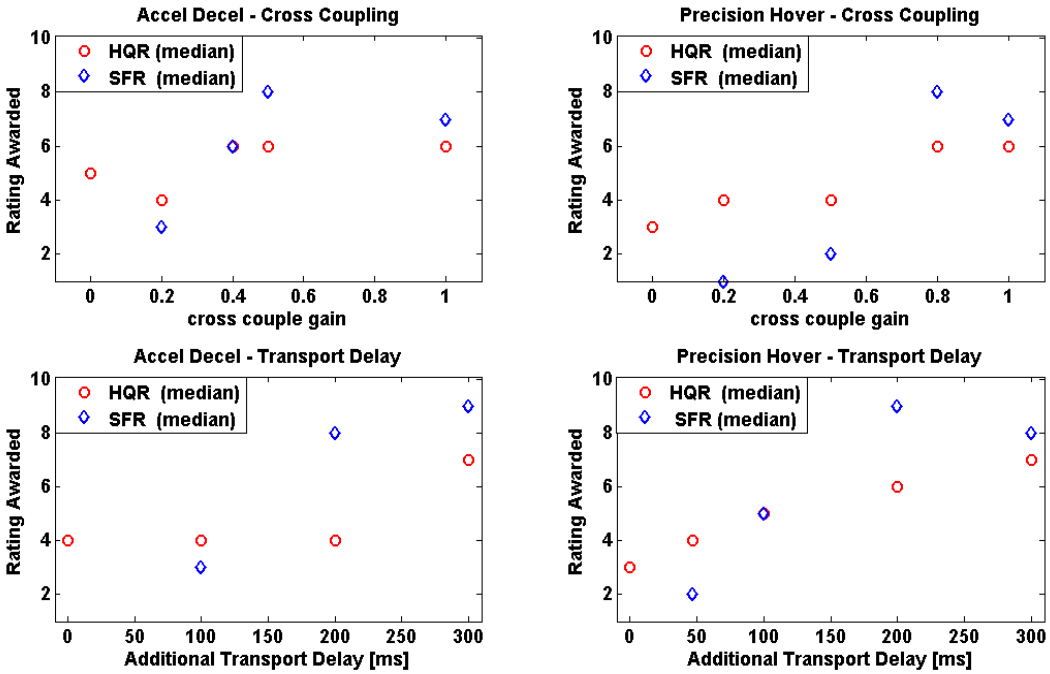


Figure 5-15 -Comparing Sensitivity of HQRs and SFRs

5.5.2 The effect of Additional Transport Delay (ATD) on perceptual Fidelity

At the heart of this research was the motivation to provide evidence for defining new metrics and tolerances for simulation fidelity. Therefore, the extent of modelling variation at which the SFRs awarded degraded from Level 1 to Level 2 fidelity (SFR=2 to SFR=3) and Level 2 to Level 3 Fidelity (SFR=6 to SFR=7) was of interest. The SFRs awarded by all pilots for variations in ATD are presented in terms of medians and spread for the AD and PH MTEs in Figure 5-16 and Figure 5-17 respectively. From inspection of Figure 5-16 and Figure 5-17, pilots appear to be similarly perceptive to transport delay in the Precision Hover and Accel-Decel MTEs and in general found an ATD of 100ms to compromise simulator fitness for purpose (low speed flying skills acquisition for the specific task). As none of the test points used yielded all Level 1 ratings, smaller ATD values would need to be investigated to determine an acceptable value for ATD to ensure fitness for purpose. It is also clear that there is a large amount of spread in the ratings awarded for specific test points. The causes and implications of such spread is discussed in section 5.6.1

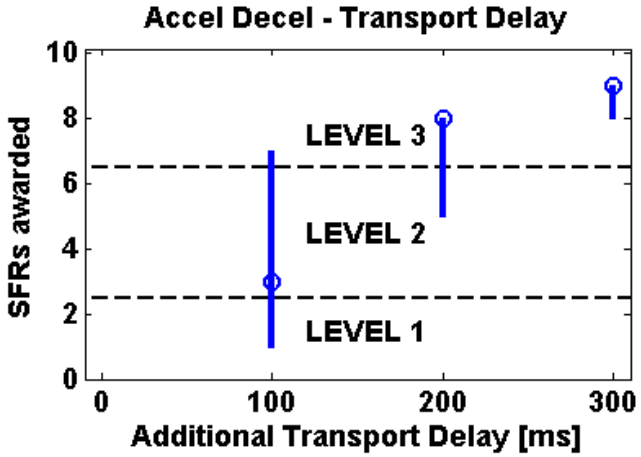


Figure 5-16 - SFRs Awarded in the Accel-Decel as a Function of ATD

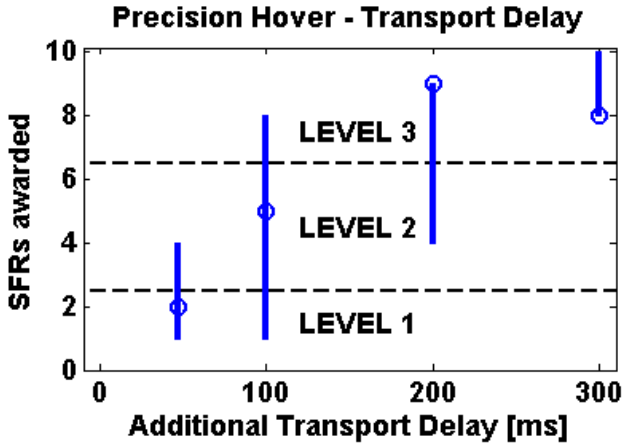


Figure 5-17 - SFRs Awarded in the Precision Hover as a Function of ATD

In Figure 5-18, percentage of ratings in Level 1, Level 2 and Level 3 fidelity are presented for each test point. Figure 5-18 shows that the majority of SFRs for ATD=100ms in both the AD and PH MTEs were in either Level 2 or Level 3 fidelity region (simulation warrants improvement/ simulation not fit for purpose). However, a simulation with this much additional transport delay compared to the helicopter to be simulated (in this case the baseline simulation) would be compliant with EASA CS-FSTD(H) Level D transport delay criteria. This suggests that that meeting the quantitative criteria does not necessarily guarantee a simulation that is fit for purpose, which supports the findings of the GARTEUR AG-12 work.

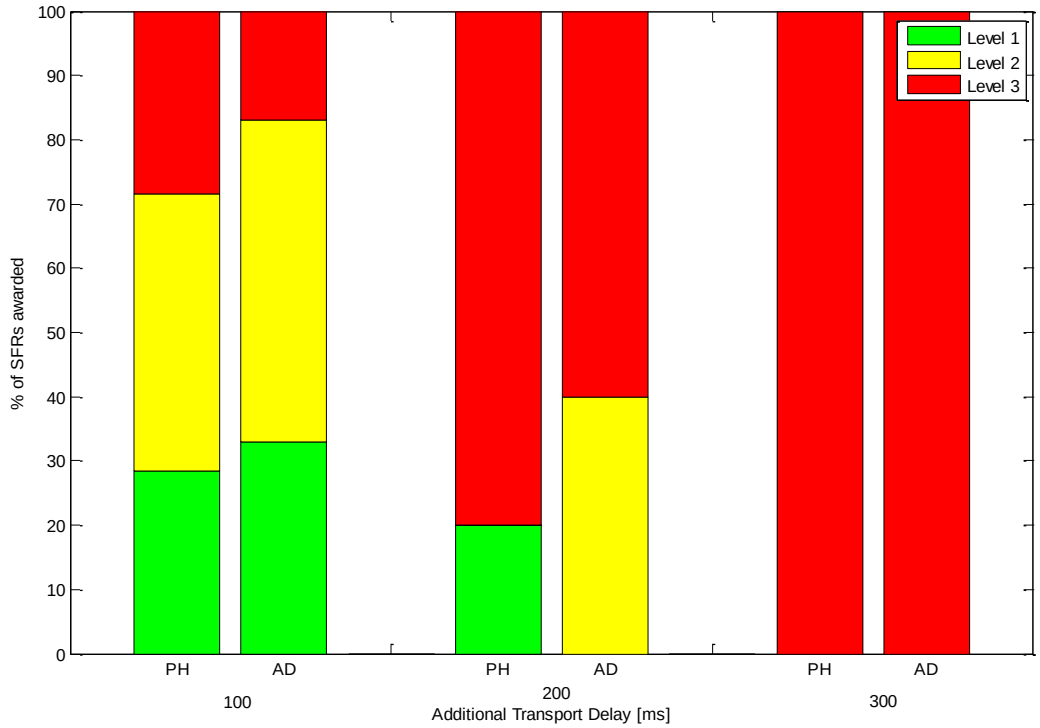


Figure 5-18 - % of SFRs Awarded in Various Fidelity Levels for Different Transport Delays

To determine what aspect of the pilots' performance and control strategy was influencing the SFRs that were awarded, a quantitative analysis of task performance and pilot control activity was conducted.

It was expected that the overshoot of the stop point in the AD MTE would significantly increase as ATD increased as the aircraft would not begin to decelerate immediately upon input of aft longitudinal cyclic and therefore, if the pilots responded when they thought they should, the aircraft would have overshot the stop point. Figure 5-19 shows the overshoots of the stop point on the first attempts for all pilots and it can be seen that the pilots overshot the end point on the first run more as the transport delay was increased.

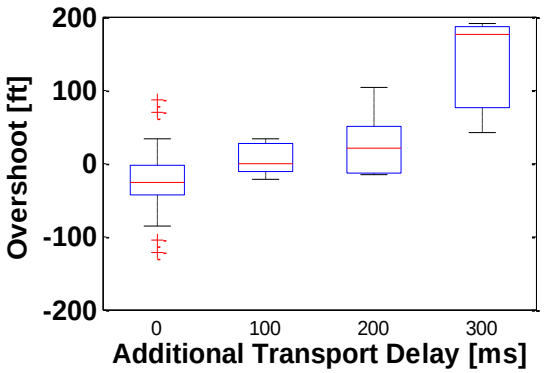


Figure 5-19 - Accel-Decel Overshoot Performance as a Function of ATD (1st Runs Only)

However, if all runs are considered, there is no significant correlation between mean overshoot and ATD (see Figure 5-20) and the mean overshoots are significantly reduced at high ATD, demonstrating that the pilots successfully adapted their adaptation to maintain performance after repeat exposure.

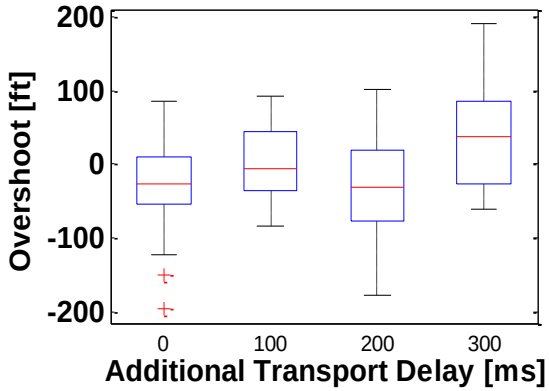


Figure 5-20 - Accel-Decel Overshoot Performance as a Function of ATD (All Runs)

Figure 5-20 and Figure 5-21 show that the pilots' task strategy adaptation took the form of using less control inputs (reduction in f_n) of larger magnitudes (increase in $\bar{\eta}$) in the primary control axis; longitudinal cyclic (XB) at high levels of ATD. The same trends were also seen in the other control axes (see Appendix C, Figure C-2). This implies that the backed out of the loop in the presence of the transport delays.

A reduction in the spread of control activity as ATD increases is also seen in Figure 5-21 and Figure 5-22. It is suggested that, with a good HQ vehicle, pilots can adopt a variety of successful control strategies. However, with poor HQs the range of successful control strategies is more limited, forcing the pilot strategies to converge. This convergence is seen in all the primary control axis metrics (see Appendix C, Figure C-2).

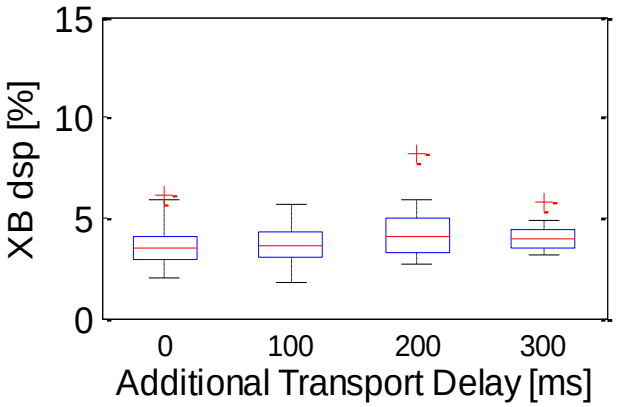


Figure 5-21 - Increase In Longitudinal Cyclic (XB) Mean Input Magnitude in the Accel-Decel as a Function of ATD

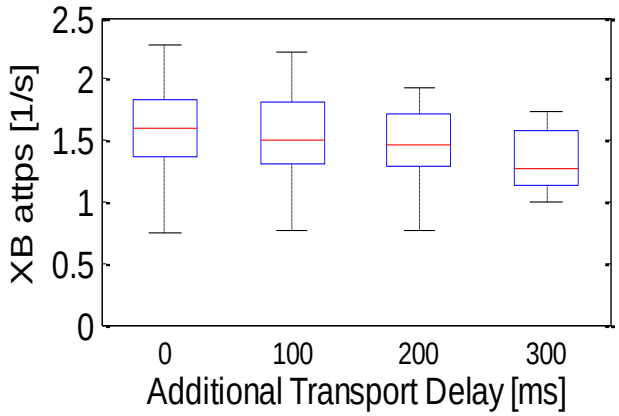


Figure 5-22 - Reduction in Number of Longitudinal Cyclic Attack Points per Sec as a Function of ATD (AD MTE)

The performance in the PH MTE was not significantly affected by an increase in ATD up to 200ms. However, at ATD=300ms the pilots were decelerating much less aggressively (higher deceleration time - Figure 5-23). Figure 5-24 shows the lateral cyclic (XA) and longitudinal cyclic (XB) control activity from all runs as a function of ATD in the transition phase of the PH MTE. The effect of ATD on the pilot control activity in the transition phase is that, up to 200ms, the pilots remained engaged with the task by increasing workload to maintain performance. This is reflected in the metrics by an increase in $\bar{\dot{\eta}}_{pk}$ and $\bar{\eta}$ metrics (Figure 5-24). At ATD=300ms it appears as if the pilots were no longer able to successfully close the loop and so backed out of the loop and used a much more gentle approach to avoid excitation of PIOs. This is seen as a reduction in control metrics (Figure 5-24) and increase in the deceleration time (Figure 5-23).

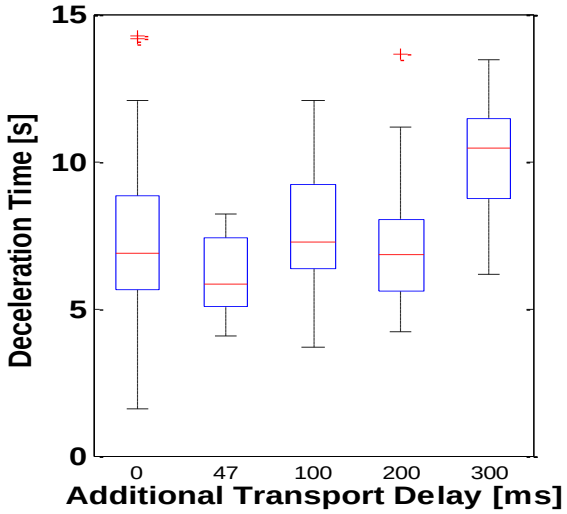


Figure 5-23 - Deceleration Time as a Function of ATD

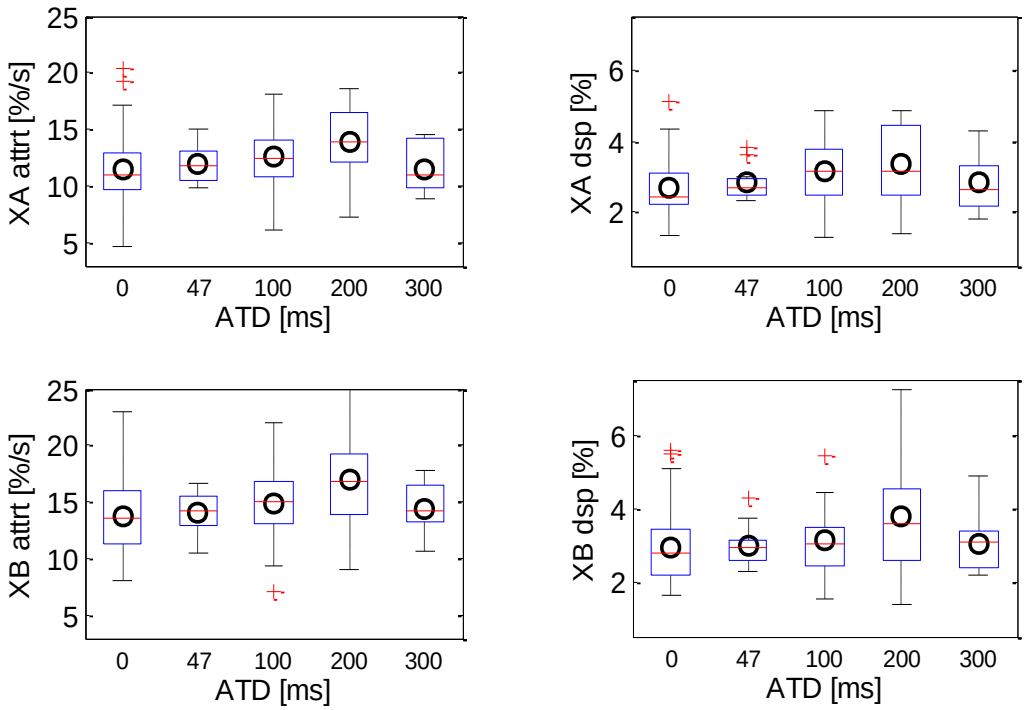


Figure 5-24 - Cyclic Mean Control Rates and Magnitudes in the Transition Phase as a Function of ATD

In the hover maintenance phase, a degradation in station keeping was observed, which correlated with a reduction in cut-off frequency in lateral cyclic (XA), longitudinal cyclic (XB), and collective (XC), as shown in Figure 5-25. This illustrates that the pilot was backing out of the loop to avoid PIOs at the expense of accuracy of the hover.

A further observation from Figure 5-25 is that, as with the AD MTE, the increase in ATD leads to a convergence of pilot strategies in terms of cut-off frequency as the pilots are forced to use a more open-loop strategy to avoid PIOs.

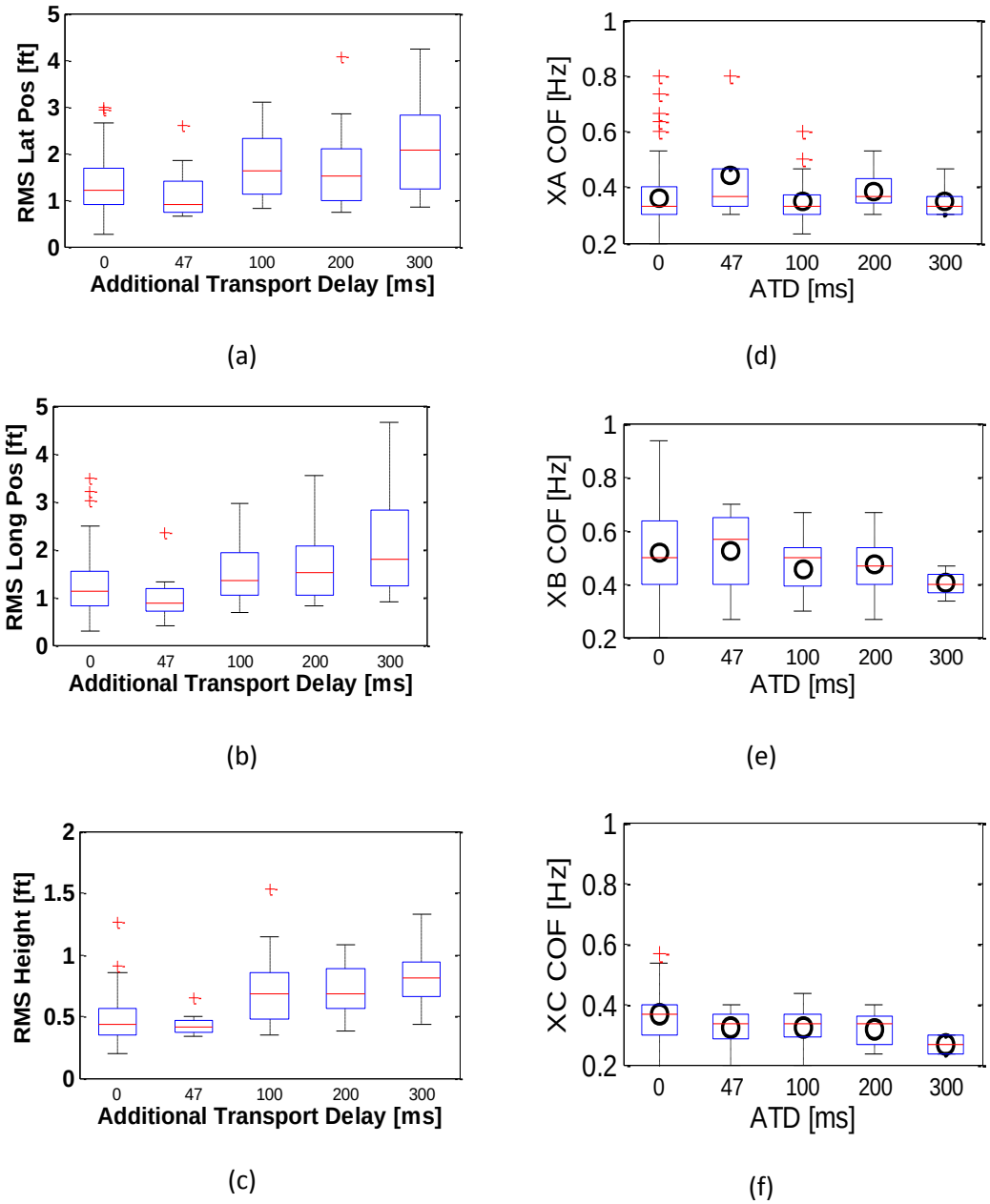


Figure 5-25 - Reduction in Cut-Off Frequency (D-F) and Corresponding Degradation of Station Keeping Performance (A - C) as a Function of ATD.

The similarity of the results from quantitative analysis of the effect of ATD on control strategy in the AD and PH MTEs suggests that ATD has a similar effect on perceptual fidelity in both tasks. This is in agreement with the similarity of the pilot ratings in the two tasks and implies that the effect of transport delay on perceptual fidelity is not heavily task dependent.

5.5.3 The Effect of errors in Inter-axis coupling on perceptual Fidelity

From the results presented in Figure 5-26, it is suggested that the tolerance for the fidelity metric based on ADS-33E-PRF pitch/roll inter-axis coupling metric for Level 1 fidelity for the Precision Hover task may lie just above ΔC_{RP} =0.4 (and most probably below ΔC_{RP} =0.8). This value would certainly be smaller for the Accel-Decel although not enough data was collected to stipulate where the boundary may be (Figure 5-27). It should also be noted that it cannot be certain that these results would be the same had the same extent of cross coupling been achieved through a different method of implementation.

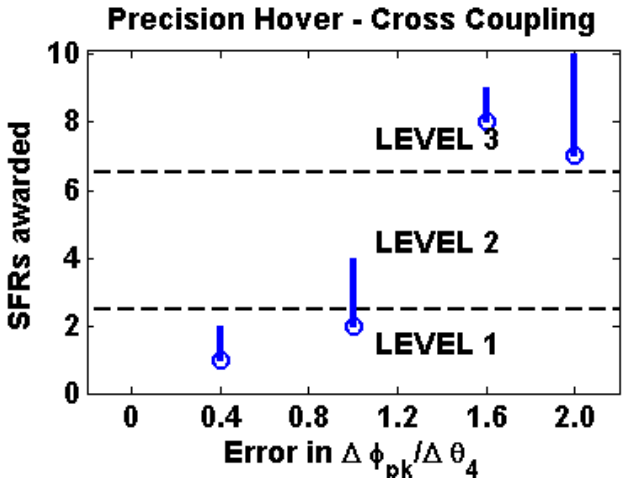


Figure 5-26 - SFRs Awarded in the Precision Hover as a Function of ΔC_{RP}

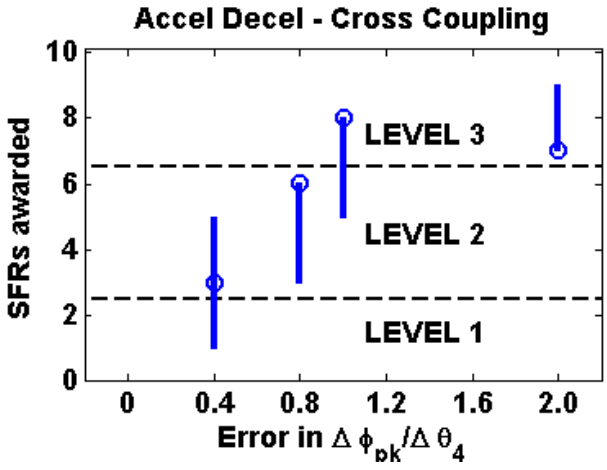


Figure 5-27 - SFRs Awarded in the Accel-Decel as a Function of ΔC_{RP}

Figure 5-28 shows that for a cross coupling error parameter, ΔC_{RP} =0.4, 100% of the SFRs awarded lay in Level 1 for the PH MTE whereas 60% of the SFRs were Level 2 in the Accel-Decel. Similarly, for a coupling error parameter value of 1.0, 50% of the SFRs awarded in the PH MTE were Level 1 and 50% were Level 2 whereas 83% of the ratings for the AD case were

Level 3. This suggests that pilots are more sensitive to cross coupling errors in the Accel-Decel than in the Precision Hover. It is suggested that this may be because the Accel-Decel is a more aggressive manoeuvre that excites the cross coupling to a larger extent. This result illustrates that perceptual fidelity is task dependant, which is not currently considered in the EASA CS-FTSD (H) standards.

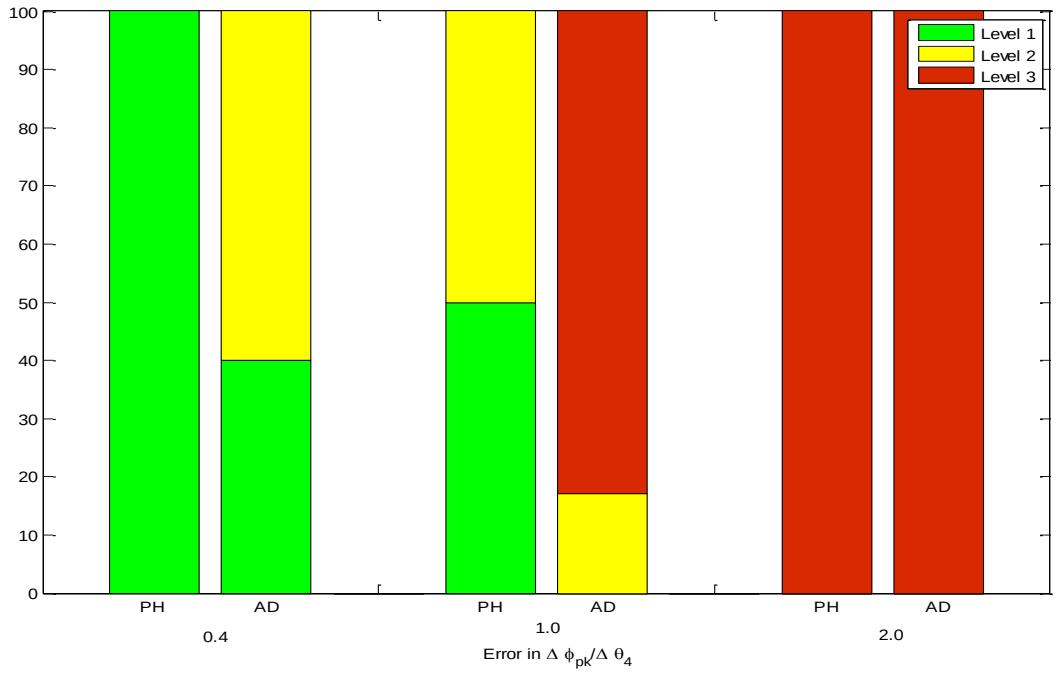


Figure 5-28 - % of SFRs Awarded in Various Fidelity Levels for Different Cross Coupling Errors

The most significant effect of ΔC_{RP} on the AD performance was an increase in overshoot of the end position of the manoeuvre, as shown in Figure 5-29. It would be intuitive to expect that increase in C_{RP} would lead to poorer lateral track maintenance. However, this was not that case (see Figure 5-30). The longitudinal position cues available to the pilot at the Accel-Decel are limited at the stop point for two reasons. Firstly, the pilot is expected to stop directly over a set of cones, when they are not visible via chin windows or out the side of the cockpit. Secondly, the nose up pitch attitude reduces the ability of the pilot to use cues such as building or trees to obtain this information. For this reason, was the stop position performance suffered as a result of degradation of handling qualities as the pilot attention was drawn into management of lateral track errors. This is confirmed by inspection of the pilot control activity.

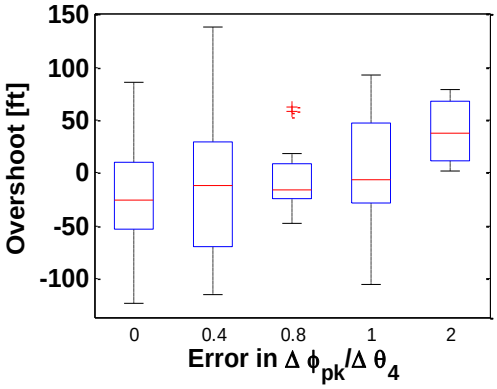


Figure 5-29 - End Position Overshoot as a Function of ΔC_{RP}

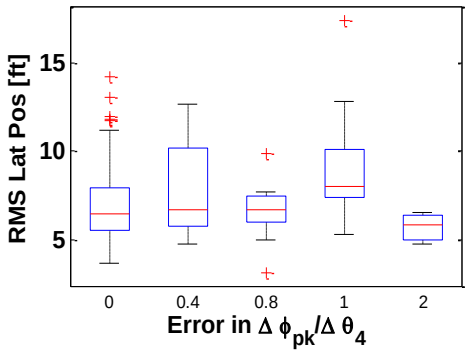


Figure 5-30 - Lateral Position Error as a Function of ΔC_{RP}

Figure 5-31 shows the relationship between increase in cross-coupling error and control activity metrics in lateral cyclic (XA), longitudinal cyclic (XB), collective (XC) and pedals (XB). Inspection of Figure 5-31 reveals that an increase in ΔC_{RP} leads to an increase in number of attack points per second, f_n, the rate at which control inputs are applied, $\bar{\dot{\eta}}_{pk}$, and control input magnitude, $\bar{\eta}$, in the lateral cyclic (XA) control activity with no significant increases in longitudinal cyclic activity. In terms of control strategy, the Accel-Decel MTE requires a longitudinal cyclic forward pulse input to select a nose down attitude for the acceleration phase and a longitudinal cyclic backward pulse input for the deceleration phase. In the case of zero turbulence, the control activity in other axes is entirely compensatory to correct for inter-axis coupling and aircraft instabilities. Therefore, as there are no lateral cyclic task demands in the AD MTE, the increase in XA control activity is purely due to the requirement to suppress lateral positional errors introduced by inputs in longitudinal cyclic. It was this increase in lateral control activity meant that the lateral track maintenance was not degraded.

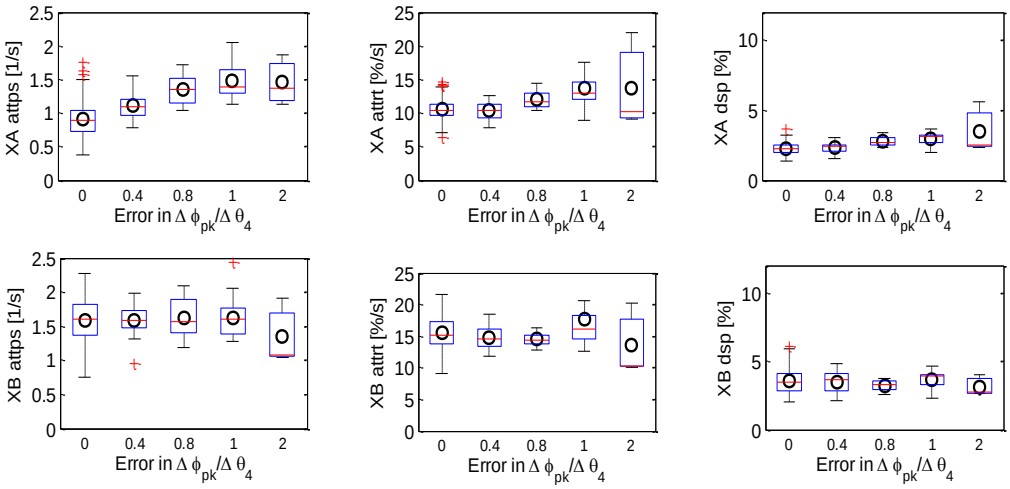


Figure 5-31 – Effect of increasing Δ C_{RP} on Accel-Decel Control Strategy

Analysis of the precision hover performance found that an increase in C_{RP} had no significant impact on the performance in either phase of the manoeuvre apart from an increase in deceleration time at the highest cross-coupling value as shown in Figure 5-32. However, the analysis of the control activity did not show any clear cause for this increase in deceleration time. The lack of correlation between increase in ΔC_{RP} and control activity and performance metrics suggests that this flight model variation is not readily exposed in the PH MTE.

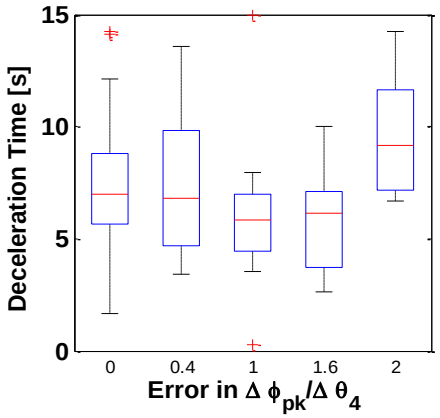


Figure 5-32

Whereas, for the ATD cases, the results from the ATD and PH MTEs were similar, the results from the quantitative analysis of the effect of inter-axis coupling suggest that variation of ΔC_{RP} had a larger effect on perceptual fidelity in the AD MTE than in the PH MTE. This is in agreement with the poorer SFRs in the AD than the PH and confirms that, in some cases at least, fidelity requirements are dependent on the training task.

5.6 Discussion

5.6.1 Spread in Ratings

There was a significant spread in the SFRs awarded, most notably for the ATD=100ms PH and AD test points, where ratings range across all three fidelity Levels (Figure 5-17). This was seen as concerning as such a large spread reduces the confidence in the results. Furthermore, ATD=100ms was a critical test point as it is the current EASA Level D criteria for simulator.

The bane of any subjective rating scale is that the users are prone to personal bias and variation in their interpretation of the terminology in the scale. In the SFR scale, there is particular ambiguity with regards to the definitions of the descriptors of task strategy adaptation as they are purely qualitative and therefore without experience and training, pilots may infer different meanings of the descriptors. Guidance material including a set of reference cases to aid evaluation pilots calibrate what is meant by ‘minimal’, ‘moderate’, ‘considerable’ and ‘excessive’ task strategy adaptation may help to reduce this variance.

Another source of variance in pilot ratings is due to variation in pilot strategies. Different pilots fly tasks in subtly, or in some cases significantly, different ways, depending on proficiency and flying experience. This variation in pilot control activity could have had a direct effect on the SFRs awarded because more active, or 'higher gain', pilots are likely to excite more of the aircraft's dynamics, thereby exposing more fidelity issues, ultimately leading to poorer SFRs. For example, if a pilot uses a slower transitions to the hover more slowly and/or uses a two stage deceleration in the PH MTE then the task aggression is reduced: therefore reducing the extent to which the aircraft dynamics, and thereby the transport delay, are excited. Similarly, during the hover maintenance, the pilot is required to maintain lateral and vertical position within ±3ft. In these trials, all of the models had a RCAH SCAS. The attitude hold augmentation allows the pilot to maintain accurate plan position with very little workload in the absence of environmental disturbances. This was noted by pilot A;

"The hover task was undemanding and so could be executed with a low level of control activity. By retaining that strategy the effect of time delay did not become evident, but the strategy was not adapted in response to the delay" (R. Erdos).

However, if the pilot engages with the controls, trying to further increase positional accuracy (beyond what is required by the task definition) then the effect of the transport delay will become more apparent.

From this idea that a high level of control activity increases the possibility of exposure of fidelity issues, it may be inferred that an aircraft with degraded handling qualities (which requires more compensatory control activity) may also cause fidelity issues to be exposed that wouldn't be exposed in a Level 1 HQ aircraft. To date, the SFR scale has only been used to rate against a nominally Level 1 handling qualities baseline aircraft. Repeat tests using baselines with Level 2 and Level 3 handling qualities would enable investigation of this hypothesis.

In terms of application of the scale in industry for simulator qualification, complications with simulator acceptance based on SFRs would arise in situations where large amounts of spread were seen in the ratings. If the scale were to be employed for such a use, a decision would have to be made in light of case by case analysis of the SFRs. It is clear that it would not be appropriate to side with the most flattering SFR, although it is possible this may be tempting when pushing for simulator qualification. Due to the small sample size it is suggested that any ratings that do not lie within the midst of the majority should be analysed in detail with reference to the SFR questionnaire and supplementary pilot comments/feedback to inform decisions on whether they are valid or anomalous. It is then proposed that the poorest valid SFR should be taken as the defining rating. The justification for this reasoning is that if all pilots are to be trained using the simulation, it only takes one pilot having a disparate experience in the aircraft than in the simulation to lead to the use of an inappropriate strategy, and possibly task failure or even worse.

As noted in chapter 4, the final training envelope of the simulator will then be defined as a list of tasks with a corresponding Level of Fidelity to each task. This list should clearly state the environmental conditions in which the training task has been evaluated with comments on the validity of training. As the operational requirements change, new training tasks must be evaluated and added to the training envelope definition list.

Spread in the awarded SFRs may be minimised through careful task definition, trial design and pilot briefing/training. Improvement of task design could mitigate against variations in pilot strategy by requiring a specific approach and ensuring scenarios do not allow pilots to

take advantage of SCAS to disengage with the task. Care should be taken here to ensure that the task remains operationally relevant and not over prescribed, as noted in section 4.6.

Pilot briefing and training would allow for further instruction on the intended strategy to be taken while conducting the task and would also allow for discussion of the terminology within the scale with examples to aid with calibration of the meanings of the descriptors. Considered trial designs including test matrix randomisation and repeat test points can ensure that ratings were not skewed due to effects such us over exposure to the simulation environment. Real time performance feedback capabilities during trials aids the pilots in determining comparative performance and ensures the pilot has flown the task per the task definition.

5.6.2 Correlation between metrics and controlled variables

In section 5.5, some qualitative trends were found between the quantitative performance and control metrics and the variation in controlled parameters, ATD and $\Delta C_{RP.}$ To identify the significance of these correlations for both MTEs, Pearson correlation statistical analysis tests [79] were used. The correlation between two variables, x and y, in a dataset of length, n, is given by R_p where;

$$\boldsymbol{R_p} = \frac{\boldsymbol{n}\ \sum(\boldsymbol{xy}) - \sum(\boldsymbol{x})\sum(\boldsymbol{y})}{\sqrt{\left(\boldsymbol{n}\ \sum(\boldsymbol{x^2}) - \sum(\boldsymbol{x})^{\boldsymbol{2}}\right) * \left(\boldsymbol{n}\ \sum(\boldsymbol{y^2}) - \sum(\boldsymbol{y})^{\boldsymbol{2}}\right)}} \text{ [79]} \qquad \textbf{Equation 5-3}$$

A perfect positive (negative) correlation (see Figure 5-33) results in R_p=1 (-1). No correlation results in R_p=0. A correlation of R_p= ±0.4 and larger is considered moderate and R_p=± 0.7 and larger is considered to be a strong correlation.

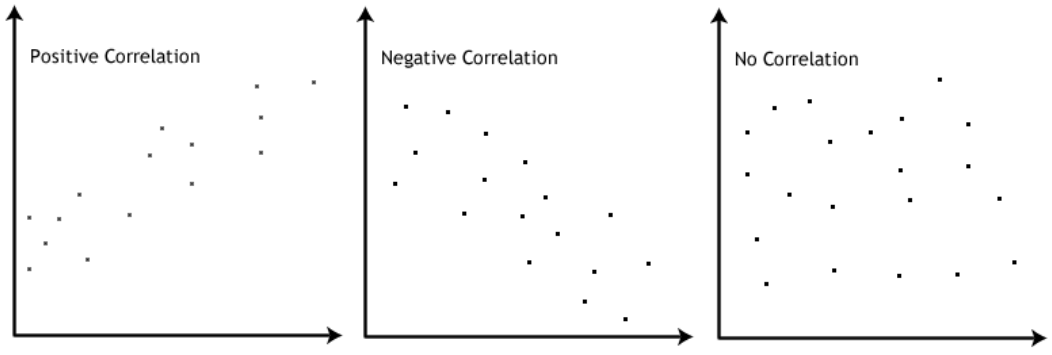


Figure 5-33 - Pearson Correlation Examples

Although some trends have been identified in objective analyses, the only metrics that showed R_p>0.4 were XA f_n, $\bar{\dot{\eta}}_{pk}$and $\bar{\eta}$ in the Accel-Decel cross coupling tests (Figure 5-31). This general lack of correlation is attributed to the large spreads in the quantitative metrics

across various runs of test points. There are a number of possible causes for such spread in experiments using pilots. These causes have been considered and are discussed below.

Spread in pilot control behaviour is often assumed to be due to learning effects and pilot intra-run variability. To investigate whether such effects were a cause of the spread in the quantitative metrics, the control attack metrics over 30 baseline PH runs for Pilot F (most completed runs) have been used as an example.

Figure 5-34 show the mean attack per second (f_n), mean attack rate ($\bar{\dot{\eta}}_{pk}$) and mean control deflection ($\bar{\eta}$) in lateral cyclic (XA) and longitudinal cyclic (XB) for each of Pilot F's baseline PH MTEs in the order that they were flown over a period of 3 days. There is no apparent learning trend in $\bar{\dot{\eta}}_{pk}$ or $\bar{\eta}$ (Figure 5-34). Instead, a continuous intra-run variability can be seen. Such variations suggest that the pilot is not varying this aspect of their control purposefully or to achieve any improvement in task performance. These variations arise due to inherent human limitations in repeatability. Convergence in both lateral and longitudinal control axes can be identified in both the f_n metrics (from Figure 5-35) from the narrowing spread in values as run number increases. The f_n metric might imply that the pilot is in fact learning, however, from closer inspection of Figure 34 the pilot is either using fast, small inputs or slow, long inputs. Either of these strategies results in the same f_n but are still two very different strategies that show no true convergence over time. It is concluded that the pilot is aware of a range of strategies that are successful and therefore does not need to converge beyond identification of this range.

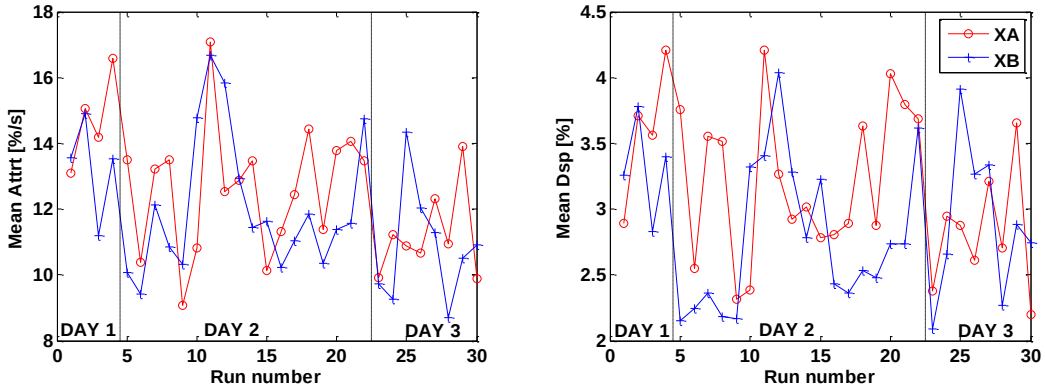


Figure 5-34 - Pilot F - Intra-Run Variation in Rapidity of Cyclic Control Inputs in the Precision Hover Transition

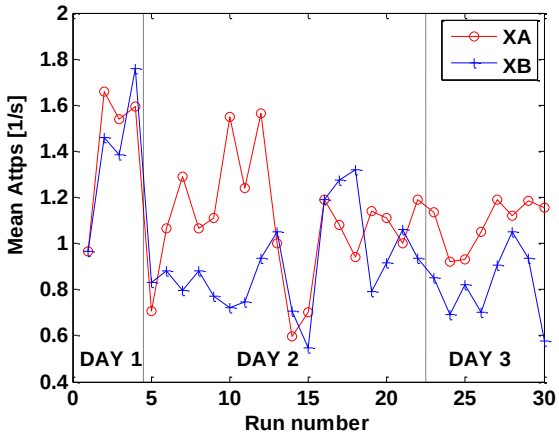


Figure 5-35 - Pilot F - Long Term Convergence of Number of Cyclic Control Inputs in the PH Transition

The results of the metric analyses, shown in full in Appendix C, contain information from eight different evaluation pilots. There is an inherent spread in behaviour amongst individuals because of factors such as flying proficiency, experience and recency as well as physiological factors and training. Figure 5-36 shows the pilot to pilot variation in baseline PH transition control activity metrics. Pilot A's metrics show that he uses a fewer, smaller and slower control inputs than the other pilots. Pilot B is at the other end of the pilot spectrum, using a lot of large, rapid cyclic control inputs. In this sense, Pilot B may be referred to as a 'high-gain' pilot and Pilot A referred to as a 'low gain' pilot. Pilot F, who was discussed previously, is somewhat ‘typical’ of the pilots used.

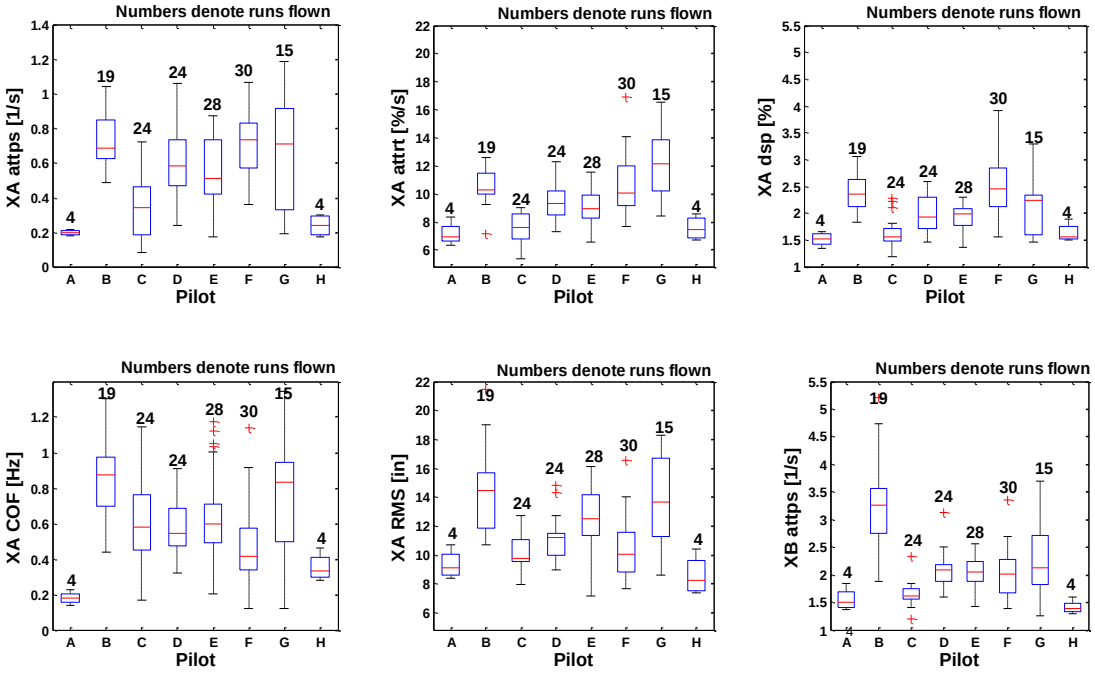


Figure 5-36 - Variation in Pilot Control Attack Metrics for Baseline PH Transition Phase

Although all the test pilots were briefed to fly the manoeuvres as per the ADS-33E-PRF Specifications, the MTE specification still allows for some variation in the way the task is approached. One of the ADS-33E-PRF requirements for a Precision Hover transition is that the approach ground speed should be between 6 and 10 knots. Figure 5-37 shows velocity profiles for Pilots A and B flying a PH transition. Pilot A transitioned at the bottom end of the approach speed where pilot B flew a much more aggressive transition, but both techniques are within the specification of the MTE.

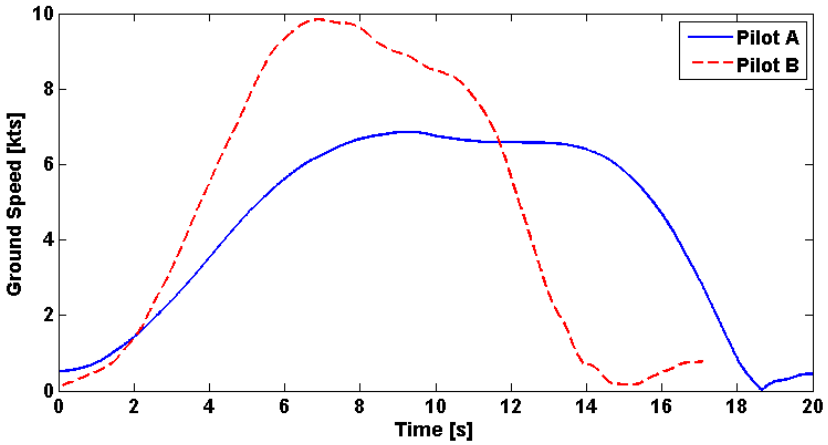


Figure 5-37 - Comparison of Pilot A and Pilot B Precision Hover Transition Velocity Profile

The corresponding control activity can be seen in Figure 5-38. The control traces show that Pilot B was using more and larger control inputs. This is reflected in the attack charts shown in Figure 5-39 and the associated mean attack metrics. The HQRs awarded by pilot A and Pilot B were HQR=2 (pilot compensation not a factor) and HQR=3 (minimal pilot compensation) respectively.

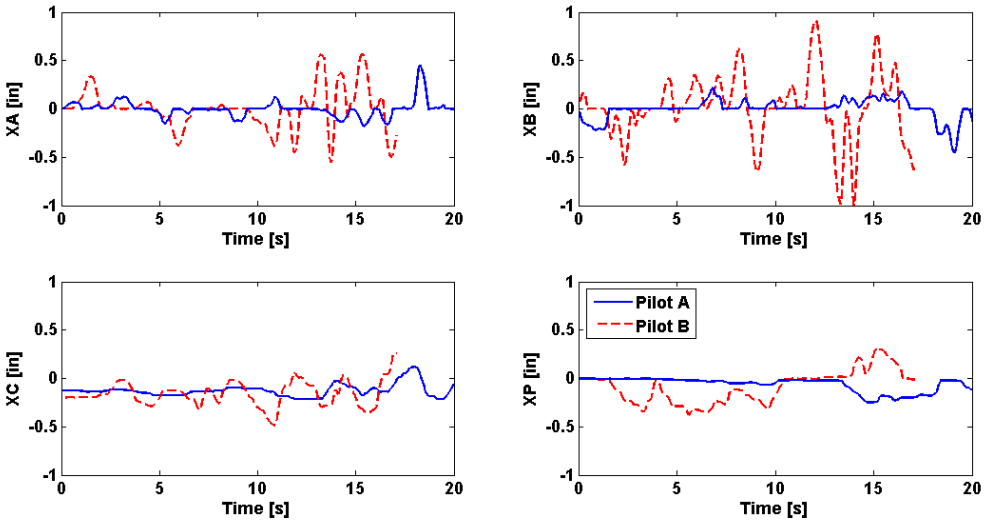


Figure 5-38 - Comparison of Pilot A and Pilot B Control Activity during a Precision Hover Transition

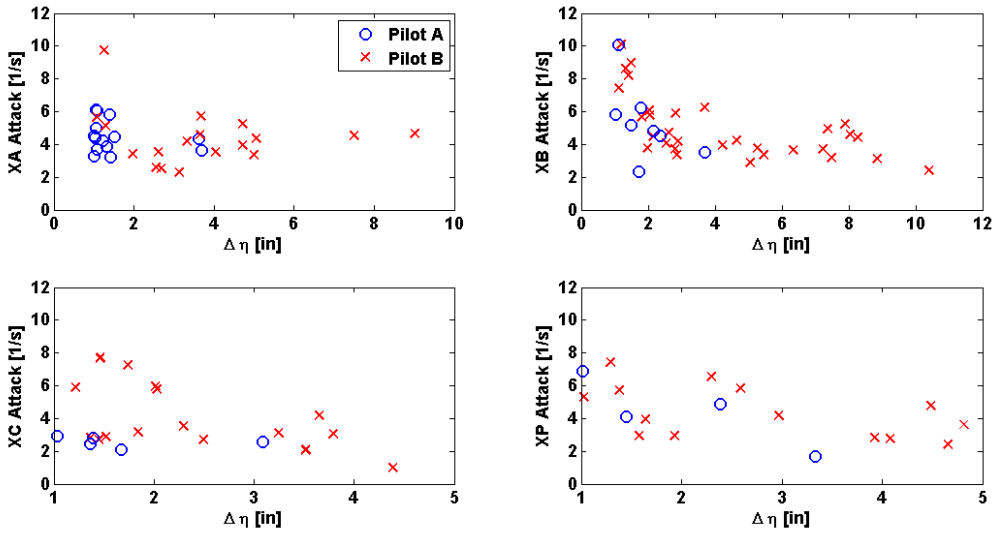


Figure 5-39 - Attack Charts of Pilot A and Pilot B Control Activity in PH MTE Transition Phase

Due to piloted simulation trial time limitations and varying pilot sensitivities to modelling errors, not all test points were completed by all eight pilots (see Table 5-3). Such incomplete test matrices can lead to artificial skews in the data. For example, the only cross-coupling model modification flown by Pilot A was ΔC_{RP}=2.0. Because Pilot A uses a relatively low attack strategy (see Figure 5-36), the quantitative attack metric data at ΔC_{RP} =2.0 has been artificially skewed down (refer back to Figure 5-31) As a result, the lateral cyclic (XA) control activity in the Accel-Decel at ΔC_{RP} =2 does not follow the general trends of the metrics in Figure 5-31.

5.6.3 Correlation between objective and subjective measures of fidelity

If the control activity metrics reflect pilot task strategy, then changes in those control activity metrics may have potential for quantifying task strategy adaptation and therefore perceptual fidelity. Correlations of changes in metrics with SFRs can be used to determine acceptable/unacceptable levels of change in the lateral cyclic control activity metrics. However, a statistical analysis such as the Pearson test was not possible to investigate the relationship between the metrics and SFRs due to the discrete and non-linearly ordinal nature of the SFRs. Therefore, only a qualitative inspection of the data was possible.

The metrics that showed moderate correlation with increasing flight model errors (section 5.6.2) have been investigated as potential fidelity metrics. Figure 5-40 shows the % change in the XA control activity metrics between the last baseline run and first modified model run as a function of the SFR awarded for AD, inter-axis coupling variation test points. The change in XA control activity metrics, particularly the % change from baseline of the average displacements, show positive correlation with SFR awarded. This suggests that the change

in off-axis control activity may be a suitable metric for assessing inter-axis coupling perceptual fidelity. If further tests were conducted to obtain statistically significant data, the point of intersection of the trend line between the % change in off-axis control activity and SFRs with the Level 1/Level 2 boundary could define a quantitative perceptual fidelity validation criteria.

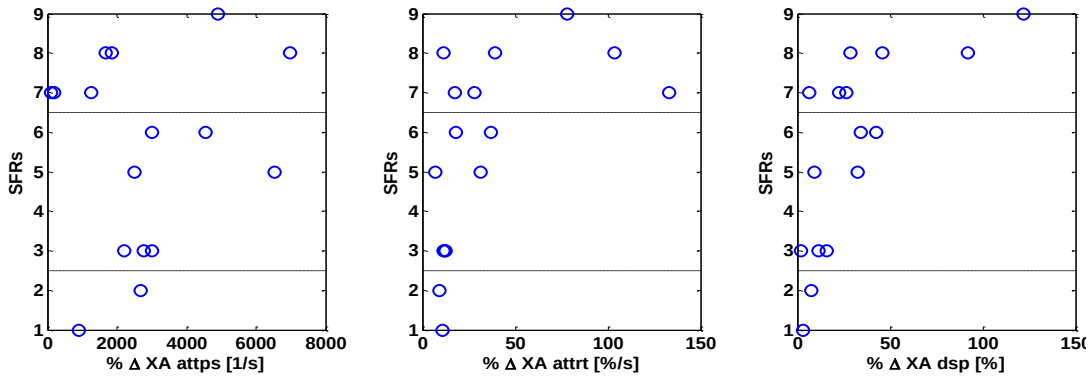


Figure 5-40 - % Change in XA Attack Metrics - Correlation with SFRs (AD Coupling Rating Runs)

Although positive correlations are implied from the data shown in Figure 5-40, there is no exclusivity between each fidelity level and % changes in metrics and therefore it is not possible to define tolerances on acceptable/unacceptable changes in these metrics.

A complication of the quantitative analysis is that pilots do not necessarily adapt in the same way after the introduction of a flight model variation. To explore this further, the pilot questionnaires for the ATD=100ms PH test points were studied. The key points drawn from the questionnaires for the ATD=100ms PH test points are given in Table 5-7.

Table 5-7 - 100ms Additional Transport Delay Pilot Comments - Precision Hover

Pilot	SFR	Comments
A	1	No perceived differences
B	3	Aggression on the capture only, only point where fidelity is an issue.
C	8	The capture was more difficult, tried to back out of the loop for the hover. Generally more aggressive. Considerable adaptation in lateral and longitudinal cyclic and pedals.
D	5	The lateral positioning during the stabilisation phase was degraded. A lot of left stick required
E	2	Slightly busier in lateral and longitudinal cyclic.
F	4	Lateral cyclic inputs were magnified by about 2.
G	7	The lateral cyclic was easily excited - had to back out. Adaptation in multiple axes.

It can be seen that the pilot perception of adaptation is varied and the pilots used different descriptors such a "busier", "magnified" and "backed out the loop".

To determine whether these perceptions correlated with changes in the control activity metrics, the SFRs of this test point for each pilot were plotted as a function of the percentage change in control activity metrics in lateral cyclic (XA) and longitudinal cyclic (XB), as shown in Figure 5-41.

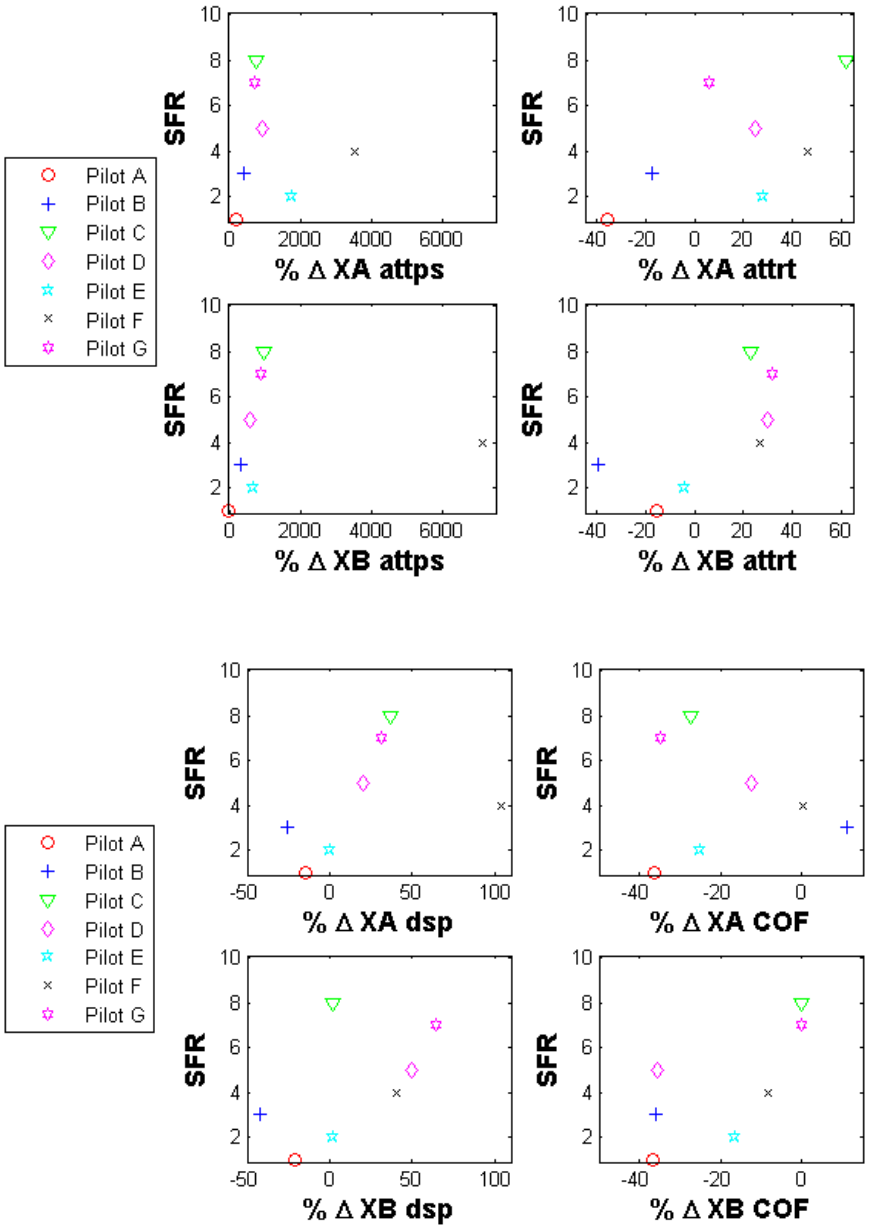


Figure 5-41 - PH Maintenance, ATD=100ms, Link between Change in Control Activity and SFR

A number of observations were made on inspection of Figure 5-41:

1. Pilot C and pilot G both noted having to "back out of the loop". Significant reductions in XA cut-off frequency are seen for both pilots. This decrease in cut-off frequency confirms that the pilots did back out of the loop - using larger, lower frequency control inputs.

2. Pilots D and F both noted degradation in the lateral cyclic particularly. However, the metrics in Figure 5-41 show that both of these pilots adapted considerably in all attack metrics to a similar extent in both longitudinal and lateral cyclic (XA and XB). The increased perception of adaptation in the XA despite the same level of adaptation of XA and XB control activity indicates that these pilots may be more perceptive of changes in the lateral axis than the longitudinal axis. It is proposed that this may be due to stronger lateral cueing in the precision hover MTE than longitudinal cueing.

3. Figure 5-41 shows that Pilots A and B reduced their control attack rate and control deflection and longitudinal cut-off frequency in light of the time delay suggesting their workload decreased. However, Pilots A and B did not perceive any fidelity issues in the hover maintenance phase of the PH MTE at ATD=100ms. This may suggest that a reduction in workload is not as perceivable as an increase in workload.

4. Pilot E noted that he was busier in lateral cyclic. From Figure 5-41, a significant increase in XA f_n and $\bar{\ddot{\eta}}_{pk}$is seen with relatively small changes in other metrics. Again, where metrics reduced, the pilot did not make comment.

These observations highlight that pilot task strategy is a complex, multi-dimensional construct and when the pilot rates adaptation of task strategy he is accumulating information reflected by several metrics in all of the active control axes.

To further highlight this, an example where multiple pilots awarded the same SFR for a given test point has been used. Pilots B, E and F all awarded an SFR=9 for the PH, ATD=200ms test point, which corresponds to ‘excessive’ task strategy adaptation. Figure 5-42 shows the percentage change in lateral and longitudinal cyclic metrics between the last baseline and first ATD=200ms cases for the three pilots that all awarded SFR=9.

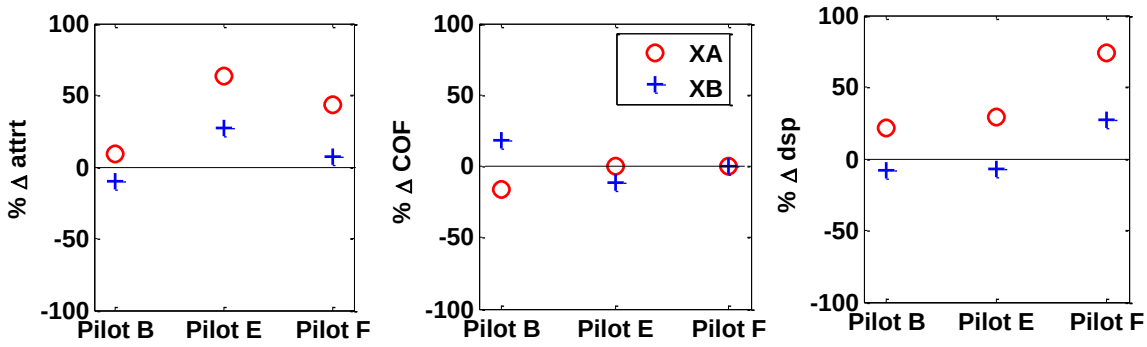


Figure 5-42 - PH Maintenance Lateral (XA) and Longitudinal (XB) Metrics for 3 SFR=9 Cases (ATD=200ms)

A number of observations were made from Figure 5-42:

1. All pilots increased their attack rate and control deflection magnitude in XA.

2. Pilot E and F showed little adaptation of their cut-off frequencies in either axis, but showed increases in both attack rates and control deflection magnitude. Whereas, pilot B's cut-off frequency was significantly adapted but his attack rate and control deflection magnitude adaptation were relatively low.

3. Pilot B adapted in opposite senses in lateral and longitudinal axes but to similar extents in both axes across all metrics, whereas Pilot F adapted in the same sense in both axes but to a larger extent in XA.

4. Pilot E's adaptation was more complex - the pilot adapted his strategy by using more rapid inputs in both axes but used smaller inputs in XB while using larger inputs in XA.

5. Pilot B's attack rate increased and control deflection size was comparable (attack similar). Pilot E's attack rate increased more than deflection size (increase in attack). Pilot F's deflection magnitudes increased more than the attack rates (reduction in attack)

These observations illustrate that it is not possible to fully quantify pilot task strategy adaptation using only one of the control activity metrics. The overall pilot perception of task strategy adaptation can be due to numerous small changes across all control axes, one significant change in a particular control axis or a combination of the two. It is suggested that a metric that reflects pilot control activity magnitude and frequency/rapidity across all axes would be a more appropriate metric on which to base a perceptual fidelity metric. To aid the development of such a metric, a study is needed that looks specifically into pilot control activity in multi-axis tasks, such as the PH MTE, to gain a better understanding into pilot control strategy and how it changes.

5.7 Validation in flight

To support validation of the methodology used in the current study, several test points were repeated in the ASRA airborne simulator by pilot A and Pilot B with RCAH SCAS engaged (therefore nominally the same baseline vehicle). It was expected that if the pilot awarded the same SFR for a given test point in flight and simulation, then the methodology was robust. The comparable results are shown in Table 5-8 and Table 5-9. Generally, the SFR awarded in the ASRA was in the same fidelity level as the SFR awarded in HELIFLIGHT-R (indicated by green and yellow). However, there was poor agreement between SFRs awarded in the ASRA and HELIFLIGHT-R for both pilots for the Accel-Decel, ATD=100ms test point. Pilot A found transport delay errors more noticeable in flight than in the simulator but pilot B found the same error more noticeable in the simulator than in flight. The pilot comments from flight and simulation from both pilots are summarised below to provide further information on these discrepancies.

Table 5-8 - Pilot A - Flight vs. Flight and Sim vs. Sim Results (Updated to Align with SFR Scale 8 ratings)

	Flight SFR	Sim SFR
Precision Hover ΔC_{RP}=2	7	7
Precision Hover ATD=100ms	2	2
Accel Decel ATD=100ms	2	6

Red=bad agreement
Yellow =reasonable agreement
Green = poor agreement

Table 5-9 - Pilot B - Flight vs. Flight and Sim vs. Sim Results (Updated to Align with SFR Scale 8)

	Flight SFR	Sim SFR
Precision Hover ΔC_{RP}=2	8	10
Precision Hover ATD=100ms	3	4
Accel Decel ATD=100ms	8	1
Accel Decel ΔC_{RP}=0.4	1	1

Red=bad agreement
Yellow =reasonable agreement
Green = poor agreement

Pilot A simulator comments (ATD=100ms, AD): Initial acceleration felt slightly slower, deceleration and hover stabilisation were worse. There was a sluggish longitudinal response during the deceleration. Issues were mainly in the last half of the run. (HQR=4.5 to HQR=5)

Pilot A flight comments (ATD=100ms, AD): A little wobbly at the end, no distinct difference. Slight sense that the heading error was bigger at the beginning. Roll wobble at the end. (HQR 5 to HQR=4)

Pilot B sim comments (ATD=100ms, AD): Can't think of anything in the handling that leaps out as being detrimental. Allowed the aircraft to sink at the end but the perception was that I wasn't flying anything different.

Pilot B flight comments (ATD=100ms, AD): Perception that max speed occurred later in the run, pulled back before pitch started, similar performance not attainable as overshot the end point and a lot of stick movement. (HQR 3 to HQR 7).

The conditions in flight were different for each pilot; Pilot A flew with a 13kt, 32° relative to course heading, whereas Pilot B flew with a 13kts, 287° relative to course heading. These winds were not present in the simulation as the simulator tests were carried out first.

Pilot B had an effective cross wind in flight which he did not have in the simulator, which may have exacerbated the poor HQs, thereby making the pilot more susceptible to fidelity issues. This explains why the pilot may have awarded a worse SFR in flight than in simulation. Conversely, Pilot A had a large component of tailwind in the aircraft that was not present in the simulator. This could have made the task aggression lower and therefore reduced the exposure to the transport delay.

5.8 Conclusions

Several piloted simulation trials were conducted in the HELIFLIGHT-R facility at UoL to determine the effect of flight model variations on perceptual fidelity. The results have been presented and discussed within this chapter. The main conclusions of these trials are summarised below:

1) The majority of the evaluation pilots perceived that a simulation with 100ms additional transport delay had limited or no utility for training initial flying skills for low speed tasks - namely the Acceleration-Deceleration and Precision Hover MTEs. This was poorer than expected by the current EASA CS-FSTD (H) criteria, which allows a simulator transport delay of up to 100ms above that in the aircraft for the highest Level of FSTD qualification. In response to this result and results obtained in previous studies, a larger scale study should be undertaken, using data from several simulators and several aircraft configurations, to determine whether these criteria should be modified in the next issue of the standards.

2) The pilots were more sensitive to pitch/roll inter-axis coupling errors in the Accel-Decel than in the Precision Hover. The ratings suggest that an acceptable /unacceptable level of inter-axis coupling error ΔC_{RP} <0.4 for AD and 0.4< ΔC_{RP} <0.8 for PH. However, more data is required to fully define what the criteria should be. Further development of this metric will continue as part of a new GARTEUR action group, GARTEUR HC AG-21, on recommendation from this research.

3) Significant variance in awarded SFRs was seen for specific test cases. This has been largely attributed to different interpretation of task strategy adaptation descriptors and variation in pilot control strategy. It has been proposed that in parallel to the effect of pilot aggression on sensitivity to flight model variations, aircraft with poorer HQs may also cause the pilot to be more sensitive to flight model variations. Therefore further testing should be done to assess this hypothesis. Guidance material for the design of fidelity assessment tasks and trial setup, as well as clear guidance, training and briefing has been detailed to minimise the variance in pilot strategy and therefore pilot perception of fidelity.

4) Pearson statistical correlation tests showed moderate positive correlation between off-axis attack metrics and increasing cross coupling in the Accel-Decel as well as between lateral cyclic (XA) attack metrics and transport delay in the Precision Hover maintenance phase.

5) It was found that percentage changes in the attack metrics between the last baseline runs and first modified runs showed some, but not sufficient, correlation with the SFRs awarded. The overall pilot perception of task strategy adaptation was found to be due to numerous small changes across all control axes, one significant change in a particular control axis or a combination of the two. It has been proposed that quantification of pilot task strategy adaptation requires metric(s) that reflect control activity in terms of both frequency and magnitude across all active control axes.

The next two chapters of this thesis detail work that has been conducted to investigate the findings of these piloted simulation trials. Firstly, an exploratory trial was conducted to further develop multi-axis, multi-dimensional metrics that better reflect pilot perception (in response to conclusion 6). Following this, a methodology for predictive assessment of perceptual fidelity is introduced within the context of investigating the effect of baseline handling qualities on sensitivity to flight model variations (in response to conclusion 4).

6 DEVELOPMENT OF QUANTITATIVE METRICS FOR FIDELITY ASSESSMENT

In Chapter 6, methods for measuring workload are discussed within the constraints of an exploratory piloted simulation trial conducted in the HELIFLIGHT-R facility. The aim of this trial was to better understand pilot workload in a multi-axis flying task to inform the development of quantitative metrics for perceptual fidelity assessment. The sensitivity of the quantified metrics to changes in simulated atmospheric turbulence levels was assessed, as well as the correlation of these metrics with subjective pilot opinion in the form of HQRs. Finally, new multi-axis metrics for the assessment of workload and task strategy adaptation are presented.

6.1 Pilot Workload and Compensation Background

The ability to measure pilot workload, and particularly compensation, is central to understanding the way in which pilot behaviour differs between flight training devices and real-world flight and therefore, the utility of a flight training device. However, the knowledge of how to quantify pilot behaviour in terms of workload remains elusive.

Adoption of the handling qualities methodology throughout this thesis led to the use of two measures of quantitative perceptual fidelity assessment in Chapter 5, namely:

1) The difference in task performance achieved between flight and simulation and;

2) The difference in the pilot's behaviour, specifically task strategy, required to achieve the task.

Measurement of task performance in flying tasks was relatively straight forward as task performance requirements can be defined through the use of ADS-33E-PRF style Mission Task Elements (MTEs) [19]. This performance was quantified for the baseline and modified simulations and then directly compared for fidelity assessment.

Measurement of the pilot's task strategy, and therefore adaptation of this strategy for fidelity assessment, was more difficult. In the handling qualities framework (and thus far in this thesis) pilot task strategy has been quantified in terms of the control attack metrics (see section 2.2.2).

Cooper and Harper define pilot compensation as;

"The measure of additional pilot effort and attention required to maintain a given level of performance in the face of less favourable or deficient vehicle characteristics" [21].

The total workload is then comprised of;

"The workload due to compensation for aircraft deficiencies plus the workload due to the task" [21].

The research detailed in Chapter 5 concluded that the relationship between individual attack metrics and SFRs was not sufficient for criteria to be proposed for fidelity assessment. It was noted that quantification of pilot task strategy adaptation requires metric(s) that reflect control activity in terms of both frequency and magnitude across all active control axes.

Figure 6-1 shows a number of sinusoidal control inputs that vary in frequency, amplitude or both. The mean data points are shown on an attack chart in Figure 6-2. In case 1, the pilot maintains the frequency of control inputs but reduces the peak amplitude. This change in workload is captured by the control deflection metric and attack rate metric but not the attack or attack per second metrics. Conversely, if the pilot reduces the frequency of the inputs but maintains peak altitude, this change in workload is captured by all the metrics except the control deflection magnitude. There are also instances where the pilot will apply the same rate of control input and of the same magnitude but will not be continuously active on the controls. This would only be captured by the attack per second metric. The attack rate metric does not capture all different types of compensation as the same attack rate can be applied to generate a lot of small inputs or a few large inputs. In this case however, the attack per second, control deflection and attack would all capture the difference. This confirms that all three attack metrics (fn, $\overline{\dot{\eta}_{pk}}$ and $\bar{\eta}$) are required to fully reflect the pilot's level of compensation as changes in pilot compensation can occur in a number of different ways and not metric alone can capture each possibility.

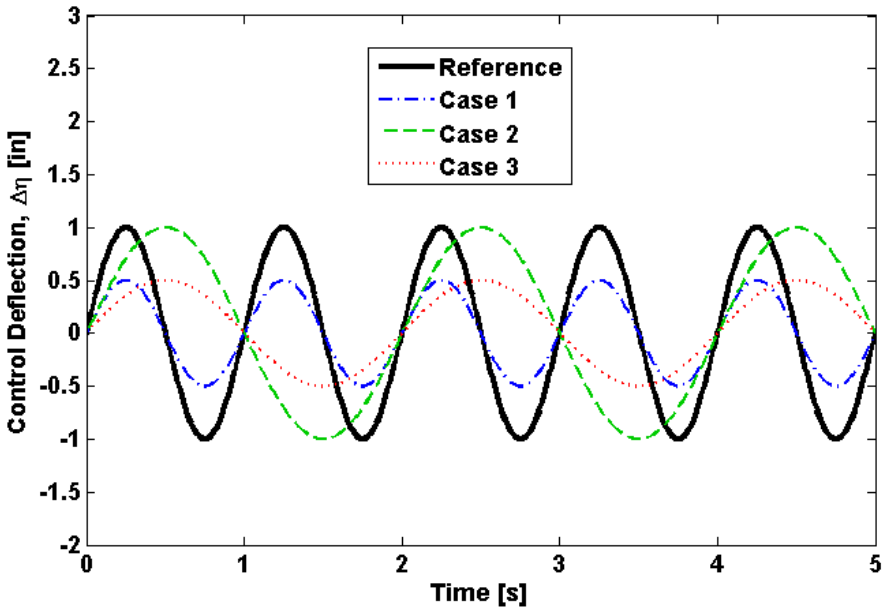


Figure 6-1 - Different Types of Changes in Workload

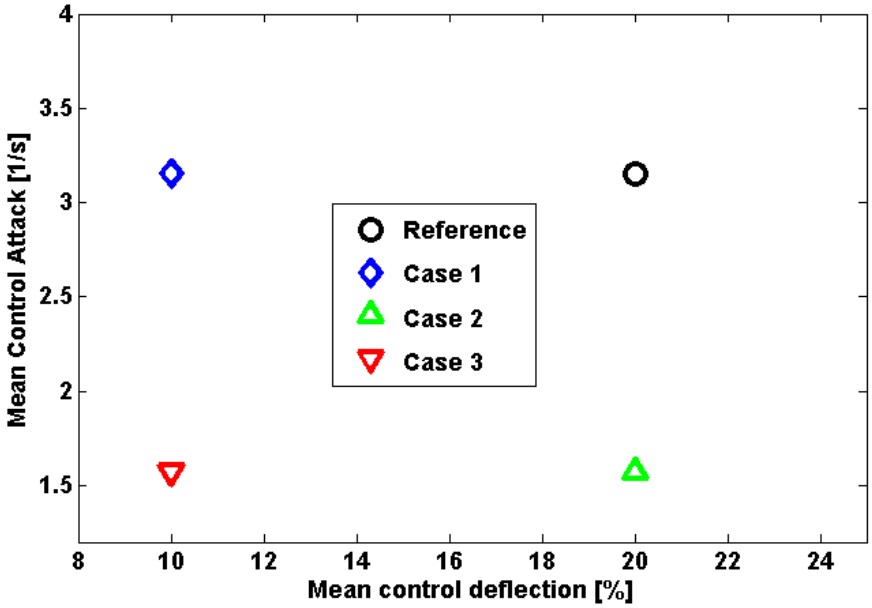


Figure 6-2 - Position of Each Test Case on the Attack Chart

Another reason that the control attack metrics may have proved inadequate to reflect pilot opinion is that the way the pilot manipulates the controls may not be the full extent of their workload. There is mental processing that is continuously required to inform the correct manipulation of the controls in response to pilot perception of vestibular, visual and auditory cues. Roscoe proposed a definition of pilot workload that highlights this:

"Pilot workload is the integrated mental and physical effort required to satisfy the perceived demands of a specific flight task" [83]

Roscoe's definition is widely used in the human factors community for assessment of pilot workload, not only for flying tasks but for combined flight and mission tasks as well. In such cases, workload has been quantified through direct measurement of physiological quantities such as heart rate [84], eye movement [85], and brain activity [86]. However, there are challenges associated with the measurement of physiological parameters including equipment calibration, intrusiveness of equipment and complex post processing.

Another approach used for the measurement of workload is based on the concept of spare capacity [87]. The postulation is that the workload due to the primary task can be measured in terms of the pilot's spare capacity for performing a secondary task in conjunction with the primary task. As the pilot's workload due to the primary task increases, their spare capacity for undertaking a secondary task is reduced. The workload can therefore be assessed by the degree to which the secondary task performance degrades in the dual-task situation relative to when each task is performed independently. Multiple Resource Theory suggests that rather than having a single resource with limited capacity, discrete resources can be allocated as the task demands [88]. Researchers are therefore advised to take care to ensure that a secondary task is designed with discrete stimuli that occupy the same resource pools as the primary task. For example, if the primary task was communicating with air traffic control, the secondary task should be audio dependent with cognitive elements. However, care should also be taken that the secondary task measures do not intrude on the primary task. This can be a significant problem when measuring physiological parameters of the pilot, for example, the use of head mounted eye tracking device may inherently alter the pilot's scan strategy and also alter physiological measures due to the physical and mental stress of wearing the measuring equipment. The secondary task should be designed to ensure that the secondary task performance is sensitive to any variations in primary task workload being assessed.

In response to the spread seen in the simulation trials described in Chapter 5, it was decided to conduct a fundamental experiment to better understand pilot workload in multi-axis tasks. An isolated study was therefore conducted to develop metrics that can reflect pilot opinion of workload and accurately capture differences in this workload. Such metrics could then be used as indicators of perceptual fidelity. The metrics to be developed were the attack metrics that were utilised in Chapter 5.

Due to the multi-axis nature of rotary wing flight, even for relatively simple manoeuvres, it was hypothesised that a control metric that fully reflects pilot workload would need to be

multi-axis too. To support the development of such a control activity metric based on the control attack metrics an exploratory piloted simulation trial was designed with the following requirements:

1. Determine sensitivity of attack metrics to changes in pilot workload. For the attack metrics to be successful descriptors of pilot strategy, they must be sensitive to changes in workload.

2. Explore how pilot workload is divided amongst axes in a multi-axis flying task: Is workload proportional to the number of axes engaged? Does the pilot work harder in one axis than another? Do pilots perceive different levels of workload in different axes?

3. Assess whether changes in attack metrics, pilot perception of change in workload and changes in measurements of spare capacity agree with one another.

The exploratory trial was split into two phases. The first phase was designed to look specifically at the effect of multi-axis control on pilot workload as well as aid development of multi axis metrics. The aim of the second phase was to further assess the correlation of metrics with subjective opinion and to determine whether measurements of spare capacity agreed with those of control activity and pilot perception. The adapted Precision Hover MTE (detailed in section 5.3) was chosen for examination in this study due to its multi-axis nature.

The following section describes the simulation facilities and methodology used in the simulation trials. The results are then presented and discussed with conclusions in the final section of the paper.

6.2 Piloted Simulation Trials

In the first phase of the trials, only one test pilot was available for testing. The pilot flew standard precision hovers (all four controls active) as well as single (lateral cyclic or longitudinal cyclic active) and dual axis (lateral and longitudinal cyclic active) precision hovers in order to investigate the effects of multi-axis control and division of pilot workload amongst the control axes. For the lateral/longitudinal only versions of the task, only lateral/longitudinal displacements were flown while yaw and heave axes were locked. Each control configuration was flown in three varying levels of difficulty through increasing levels of simulated atmospheric turbulence (see Table 6-1). Turbulence was chosen (rather than transport delay or inter-axis coupling as used previously) because turbulence increases

workload in a predictable manner, therefore providing a controlled testing environment. The test matrix was randomised with at least four runs for each test point to mitigate against learning and anomalies. The control activity and aircraft state data were collected during the test runs and the pilot awarded HQRs for each of the 12 test conditions (see Table 6-1). The turbulence was modelled as a number of sum of sine wave signals that were applied to the airframe as translational and rotational forces. The intensity of the turbulence was varied using a simple gain, K_{turb}, applied to the sum of sine waves signal.

Table 6-1 - Phase 1 Exploratory Trial Test Matrix

Test Point	1	2	3	4	5	6	7	8	9	10	11	12
Axes engaged	Lat	Long	Lat & Long	All	Lat	Long	Lat & Long	All	Lat	Long	Lat & Long	All
Turbulence	No Turbulence				Med Turbulence (K_{turb}=0.3)				High Turbulence (K_{turb}=0.6)			

The relative turbulence imposed in each translational axis was determined by a historical set of auxiliary gains in the HELIFLIGT-R software that had been previously tuned to pilot subjective opinion. These gains are shown in Table 6-2. An open-loop simulation was performed of the aircraft in hover to determine an example time-varying turbulence input signal and the corresponding aircraft state responses. These time histories are presented in Figure 6-3. It can be seen that the disturbances in roll are smaller than in pitch which leads to corresponding smaller aircraft roll rates and attitudes. Despite larger yaw rate turbulence forces, the magnitude of the response of the aircraft is similar to that in pitch.

The change in aircraft longitudinal position is larger than changes in lateral position, despite much smaller surge components of the turbulence. The yaw attitude diverges because the SCAS system has no attitude retention in this axis.

Table 6-2 - Auxiliary Turbulence Force Gains for Each Axis

U	V	W	p	q
1	1	1	0	0
surge2pitch	surge2roll	sway2pitch	sway2roll	sway2yaw
0.3	-0.075	0.075	0.1	0.1

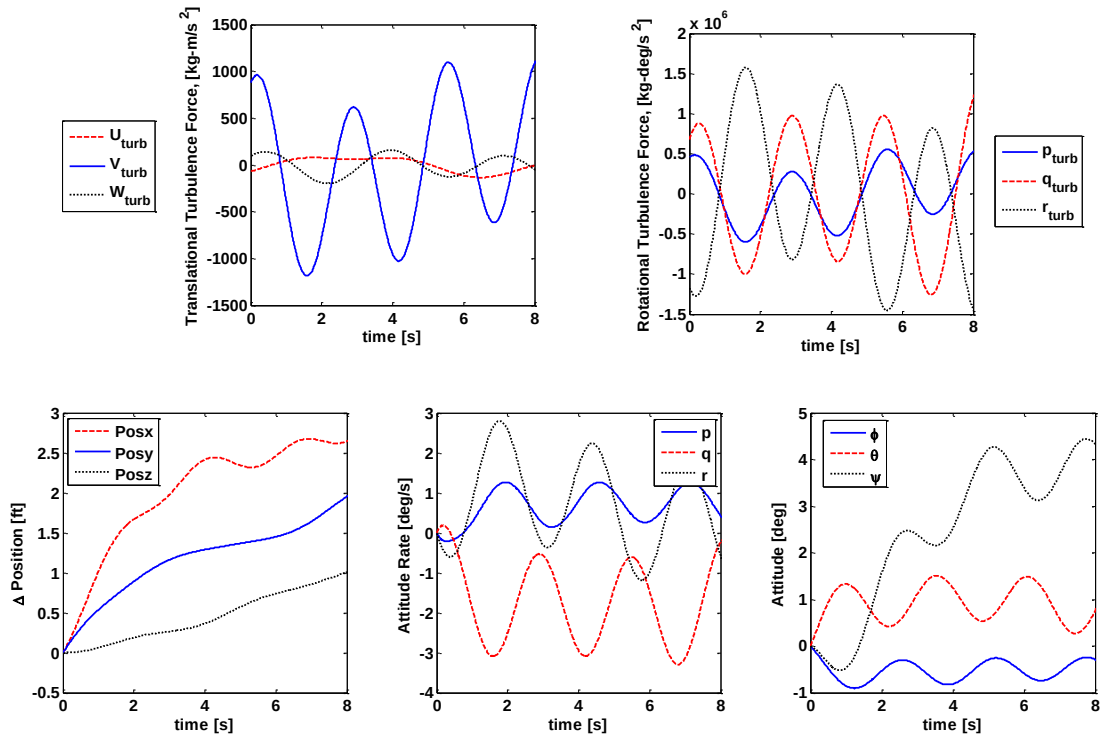


Figure 6-3 - Turbulence Inputs and Corresponding Open-Loop Aircraft Responses.

In the second phase of trials, two test pilots were exposed to three varying levels of difficulty through increasing levels of simulated atmospheric turbulence during a dual-axis Precision Hover (lateral and longitudinal cyclic only). It was intended that the variable turbulence would be tuned to yield Level 1, Level 2 and Level 3 HQRs from each pilot. However, Pilot B was more sensitive than pilot A to the imposed turbulence, awarding a HQR=7 where pilot A awarded a HQR=4 for K_{urb}=0.6. With no turbulence, Pilot B awarded HQR 4 (Level 2) and, with the maximum value of K_{turb} tolerable to the simulator motion system, Pilot A awarded HQR=6 (Level 2). It was also not possible to find conditions that resulted in identical HQRs being awarded by each pilot. The values shown in Table 6-3 were the best matches achieved in the test conditions and represent 'easy', 'medium' and 'difficult' test conditions for phase two of the trial.

Table 6-3 - Turbulence Gains, K_{turb}, Experienced by Each Pilot

	EASY	MEDIUM	DIFFICULT
compensation	minimal/moderate	considerable	extensive/max tolerable
Pilot A	0.2 (HQR 3)	0.5 (HQR 4.5)	1.2 (HQR 6)
Pilot B	0.1 (HQR 4)	0.3 (HQR 5)	0.6 (HQR 7)

In this phase of testing, the pilots were required to perform a secondary task to the best of their ability while maintaining performance in the primary flying task (a dual-axis PH task).

The secondary task was designed to occupy the visio-spatial resources of the pilot, drawing his attention from the primary visual cues as well as requiring cognitive and physical effort. The task required the pilot to continually scan the visual scene for flocks of birds (see Figure 6-4). On recognition of the birds the pilot was required to press the cyclic trigger. This recorded the recognition time of the pilot in the dataset. The pilot was also required to compute and verbally report the number of birds observed in the flock. The timings of the flock entering, direction from which the flock appeared, number of birds within the flock and number of times a flock of birds appeared during a given test point were randomised to avoid any learning effects influencing the results. The test matrix was randomised and included control points where the pilot only had to perform the primary or secondary task alone.

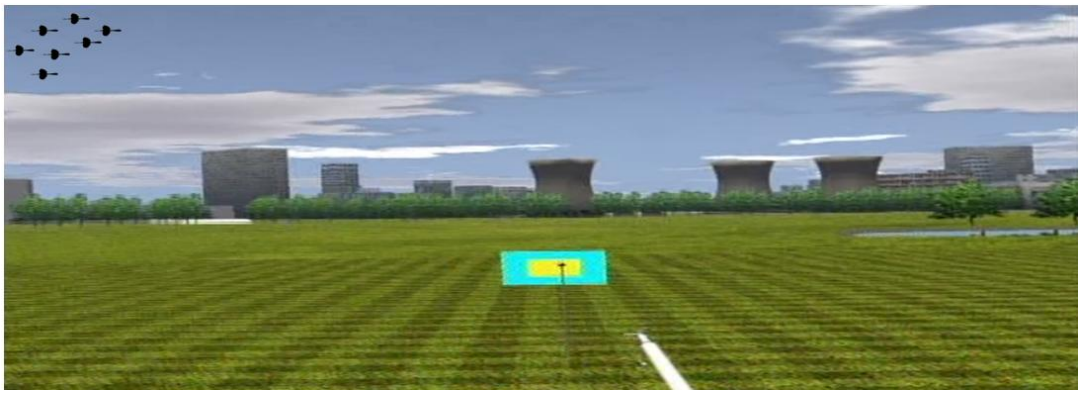


Figure 6-4 - OTW View with Impression of Flock of Birds in Peripherals

6.3 Results and Discussion

In this section the results from the exploratory trial are presented and the implications of the results discussed. For the control activity analysis, the precision hover MTE has been divided into two segments, the transition segment and the hover maintenance segment, to assess whether one segment was dominating the pilot's workload. The point at which the transition ended and the hover maintenance began in each test run was determined subjectively from inspection of the lateral and longitudinal velocity traces and other aircraft states where necessary. The example in Figure 6-5 serves to highlight that the transition phase includes the deceleration to the hover board. Performance has been assessed in terms of hover maintenance of plan position only.

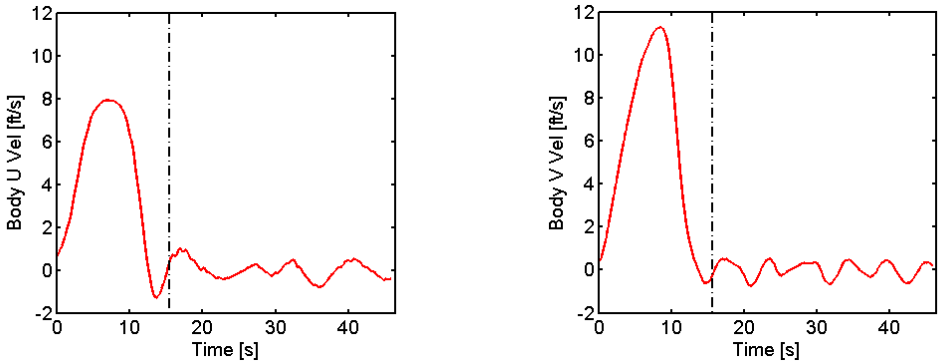


Figure 6-5 - Example of Procedure for Dividing the Precision Hover into Two Phases; Body-Referenced Aircraft U and V Velocities Relative to the Ground

In order to obtain pilot subjective opinion of compensation, HQRs were taken. The validity of mean HQRs is uncertain due to the nonlinearity of the HQR scale. Therefore in this study, modal (most frequent occurring) HQRs are presented and the spread of ratings shown to avoid misinterpretation of the HQRs. Where the pilot felt that desired performance required considerable compensation, a rating of HQR=4.5 was accepted.

The Analysis of Variance (ANOVA) technique [89] has been used to assess the statistical significance of the results. In an ANOVA test, the average of each group of data is compared to the average of the dataset. F represents the strength of the hypothesis that the data in each group is significantly different from any other group (group locality exists) and must be >>1 for confidence of group locality. The probability, p, is the probability of the null hypothesis being true (i.e. no group locality) and the hypothesis is deemed to be significant if $p<0.05$. The minimum number of groups in an ANOVA test is 3 and there must be at least 3 test points in each group.

Phase 1: The effect of multi-axis control on pilot workload

Figure 6-6 shows the modal HQRs awarded by the test pilot in the three levels of atmospheric turbulence with various axes of control engaged. The bars represent the modal HQR awarded with the maximum and minimum indicated by the solid horizontal lines. Figure 6-6 confirms that HQRs increase with increasing turbulence. Also to be expected, the HQRs show that the pilot perceived the single-axis PH tasks to require less compensation than the dual-axes and full control PH tasks across all turbulence levels. The inclusion of the pedal and collective axes into the task does not generally increase the pilot's perception of workload above that perceived for the dual-axes PH task at low and intermediate turbulence levels. At the high turbulence level there is an increase in perceived workload associated with the added control of the collective and pedals. The pilot also found the lateral only task to be marginally more difficult than the longitudinally only task at high turbulence.

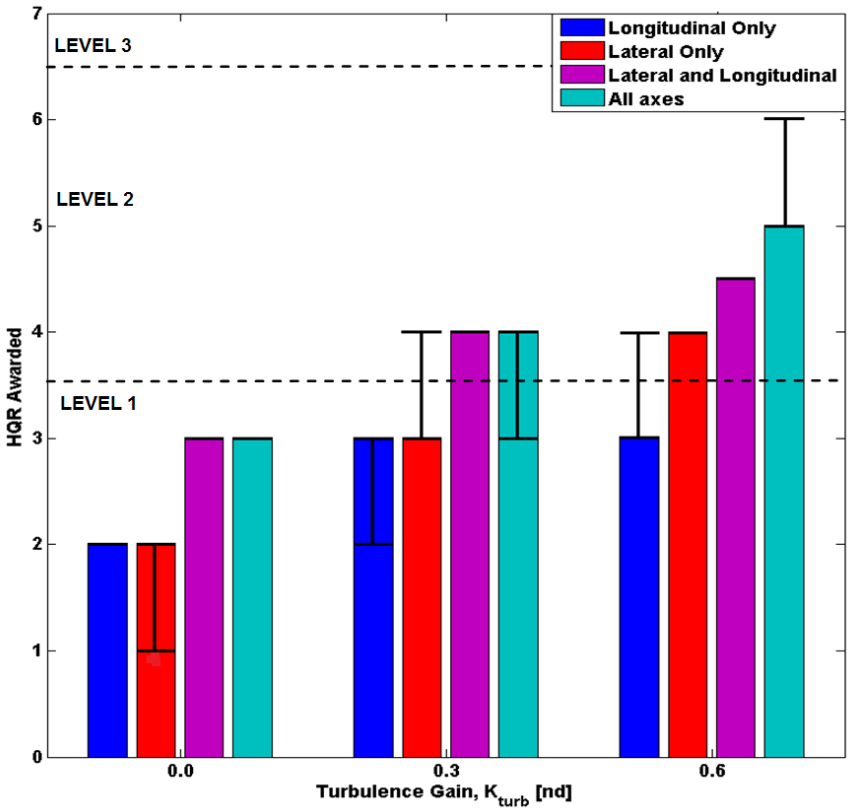


Figure 6-6 - Effect of Turbulence Level and Number of Axes Engaged on Average HQRs Awarded

It would be expected that as turbulence increases, the requirement to suppress errors in prescribed trajectory/position maintenance would increase, thereby increasing pilot workload. This increase in workload with increasing turbulence was clearly perceived by the test pilot (Figure 6-6). Figure 6-7 and Figure 6-8 show the control activity metrics as a

Function of turbulence gain for the all-axes engaged runs. It can be seen that this increase in workload is reflected in the control attack metrics; with higher levels of turbulence the pilot is using more frequent (increase in attack/second - Figure 6-7a and Figure 6-8a), more rapid (increase in attack rate - Figure 6-7b and Figure 6-8b) and larger (increase in control deflection - Figure 6-7c and Figure 6-8c) control inputs in all axes in both the hover and transition phases. Pilot control activity in pedal increases disproportionately. This is because there are a higher relative turbulence forces in yaw (Figure 6-3) and there is no attitude retention in the yaw control axis. Therefore, this component of the turbulence requires more cancelation by the pilot to avoid yaw motion. Collective activity does not increase significantly as the relative turbulence forces in the heave axis are low. In the hover phase, the XB and XA activity increase together. This means that the pilot reverts to a 'stick stirring' strategy; controlling the cyclic as a single controller, in this phase of the manoeuvre. This is not seen to the same extent in the transition phase. The reason for this is discussed later in this section.

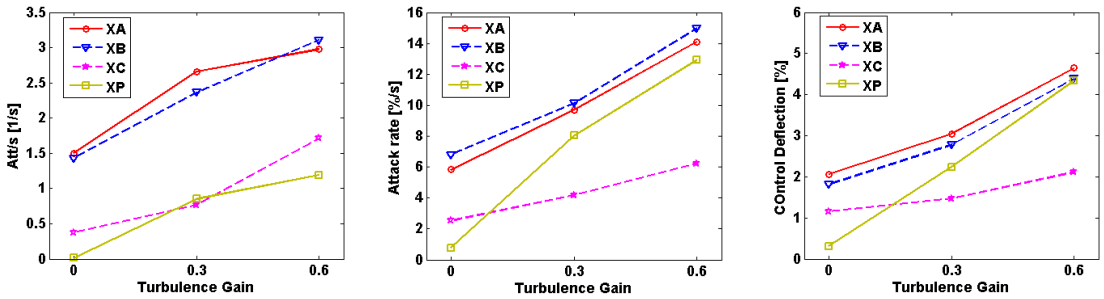


Figure 6-7 - Increase in Control Attack Due to Increasing Turbulence for the Hover Segment of the All-Axes PH Task

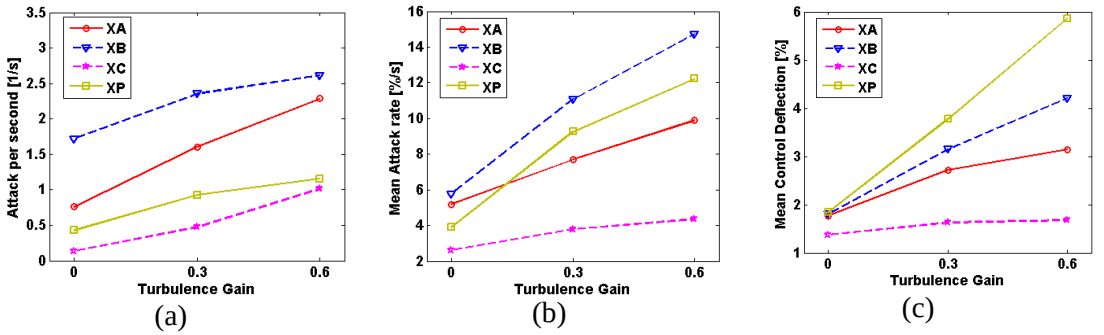


Figure 6-8 Increase in Control Attack Due to Increasing Turbulence for the Transition Segment of the All-Axes PH Task

To assess the significance of the relationship between the increase in metrics and the increase in turbulence, an ANOVA test was used. Three groups were used (representing the different turbulence levels) and there were at least 4 test points (repeats of test condition) in each group. Table 6-4 and Table 6-5 show the ANOVA hypothesis strength, F and probability of null hypothesis, p, for each control attack metric. The results confirm that the relationship between the control metrics and turbulence gain is highly significant for both the transition and hover manoeuvre apart from the collective metrics in the transition segment.

Table 6-4 - ANOVA - Significance of Change in Control Attack Metrics Due to Increasing Turbulence - Hover

	Attack per Second		Mean Attack Rate		Mean Control Deflection	
XA	F=49.40	p=1.4e-5	F=44.35	p=2.2e-5	F=37.98	p=4.1e-5
XB	F=51.47	p=1.2e-5	F=68.95	p=3.49e-5	F=29.43	p=0.0001
XC	F=15.33	p=0.0013	F=32.16	p=8.0e-5	F=42.55	p=2.6e-5
XP	F=68.95	p=3.5e-6	F=22.70	p=0.0003	F=29.43	p=0.0001

Table 6-5 - ANOVA - Significance of Change in Control Attack Metrics Due to Increasing Turbulence - Transition

	Attack per Second		Mean Attack Rate		Mean Control Deflection	
XA	F=39.79	p=3.4e-5	F=23.82	p=0.0003	F=15.24	p=0.0013
XB	F=17.89	p=0.0007	F=35.22	p=5.5e-5	F=16.62	p=0.001
XC	F=25.97	p=0.0002	F=35.95	p=0.0586	F=0.13	p=0.8767
XP	F=1.52	p=0.2706	F=48.17	p=1.6e-5	F=82.93	p=1.6e-6

The statistical analysis of the results from phase one of the exploratory trial (Table 6-4 and Table 6-5) have shown that the control attack metrics have statistically significant sensitivity to changes in workload due to increases in atmospheric turbulence. This confirms that the attack metrics are appropriate for quantitative perceptual handling qualities assessment. In the majority of instances, the hover metrics showed a higher hypothesis strength, F, than the transition metrics.

To further validate the meaningfulness of the attack metrics as a measure of pilot workload, a second ANOVA test has been used to assess the correlation between the control attack metrics and the HQRs awarded. In this test each group represents a level of compensation perceived (from the HQR descriptors) therefore the number of data points in each group could not be controlled and were variable. The number of groups for each test was also variable. Therefore only the lateral axis case has been considered as the longitudinal-only

task yielded 6 HQR=3, 6 HQR=2 and 1 HQR=4 and so insufficient data were available for an ANOVA analysis.

The results from the ANOVA test showed there was no significant correlation between the transition phase metrics and the HQRs ($p>0.05$ in all cases). However, the hover phase control attack metrics showed highly significant correlations across the board. Figure 6-9 shows the ANOVA hypothesis strengths for hover XA attack metrics vs. HQRs for various axes engaged cases. The results show that the hypothesis strengths of all three metrics are reduced for multi-axis tasks. This is because when the pilot's workload is divided between multiple axes, his perception of compensation is now influenced not only by the lateral axis but by all the active axes, therefore weakening the direct relationship between the lateral axis metrics and the perceived workload. In order to measure compensation in a multi-axis task, a metric that encompasses the compensation in all axes had to be developed.

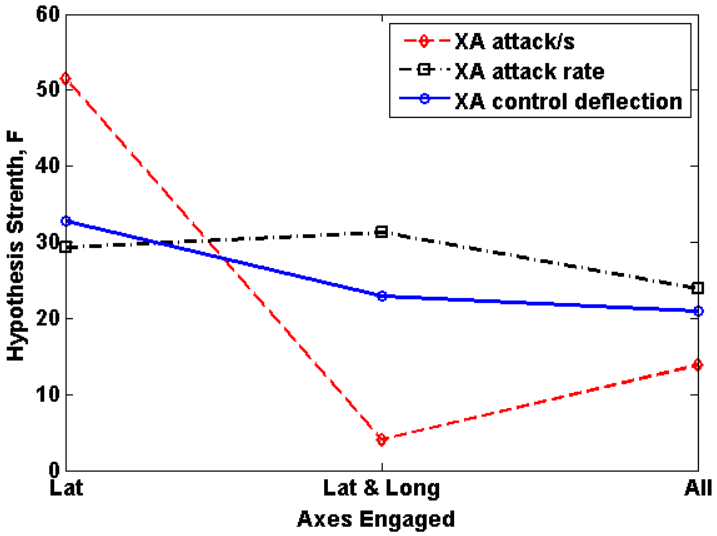


Figure 6-9 - Decreasing Hypothesis Strength between HQRs and XA Mean Deflection in Hover Segment

Figure 6-10 shows the hypothesis strength of a correlation between the HQRs and the sum of the individual axis metrics. It can be seen from Figure 6-10 that the addition of the individual axis metrics has a marked effect on the correlation strength between the HQRs and the metrics.

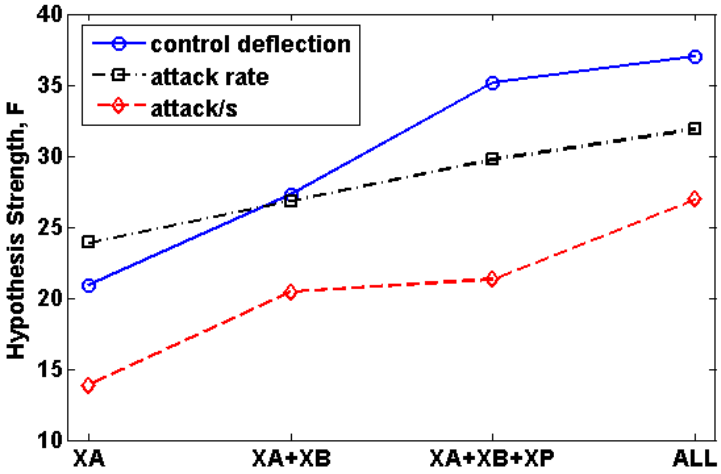


Figure 6-10 - Increase of Hypothesis Strength for All Axes Precision Hover (Hover Segment) or Additive Metrics

However, the magnitude or frequency of control inputs alone cannot tell the story of pilot control activity. A single metric that is two dimensional, representing size and rapidity of inputs was therefore conceived. In Figure 6-11, the additive mean control deflection ($\bar{\eta}_{XA}$+ $\bar{\eta}_{XB}$+ $\bar{\eta}_{XC}$+$\bar{\eta}_{XP}$) is plotted against the additive mean attack rate ($\overline{\dot{\eta}_{pk}}_{XA}$+ $\overline{\dot{\eta}_{pk}}_{XB}$+ $\overline{\dot{\eta}_{pk}}_{XC}$+ $\overline{\dot{\eta}_{pk}}_{XP}$) for the hover segment. It can be seen that this two dimensional representation gives a strong indication of the increase in workload due to turbulence and the introduction of more axes. Figure 6-11 shows an increase in both attack rate and control deflection as HQR increases and there is a clear separation of the Level 1 and Level 2 HQRs awarded. From these preliminary results, it is suggested that this metric may be used to determine levels of acceptable and unacceptable pilot control attack in the precision hover manoeuvre for Level 1 HQs. An initial boundary between Level 1 and Level 2 handling qualities is proposed in Figure 6-11. Further research will be undertaken to ascertain whether this metric also has potential for perceptual (simulation) fidelity evaluation and, if so, how much separation would compromise simulation fidelity.

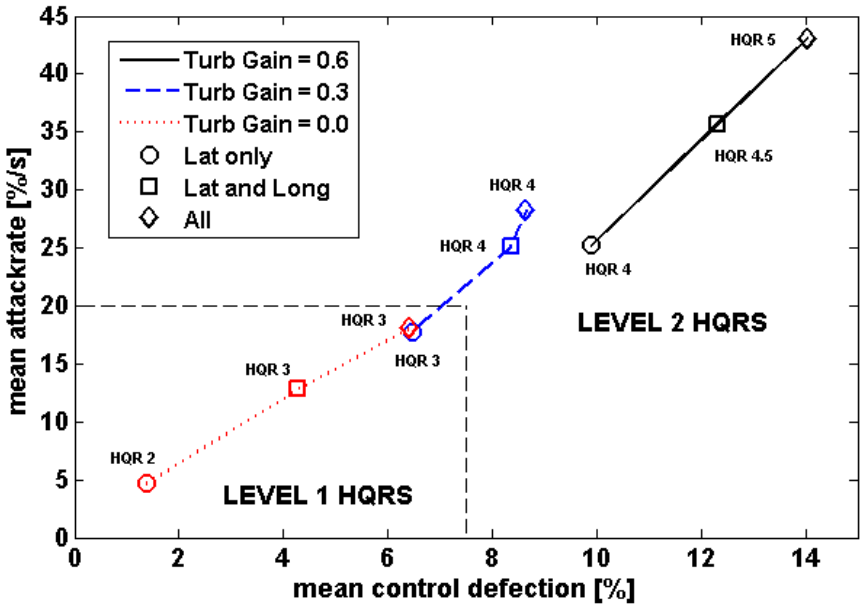


Figure 6-11 - Proposed Additive Metric Plot with Tentative Level 1-2 Boundary (Hover Segment Metrics)

Although the additive metrics show promise, it is unlikely that the contribution of each axis to compensation is equal. Figure 6-7 shows that, in a non-turbulent environment, the lateral and longitudinal cyclic attack metrics are larger than collective and pedal in the hover maintenance segment of the manoeuvre, suggesting the workload in the hover maintenance is dominated by the cyclic. However, in the transition phase, the pedal activity is much more dominant. This is because, rather than flying with the heading of the aircraft at zero using co-ordinated lateral and longitudinal cyclic, the pilot adopted a strategy whereby he yaws in the direction of the board and then uses longitudinal cyclic to fly towards the board. This is illustrated in Figure 6-12 from the co-ordination of XB and XP and explains the lack of stick-stirring observed from the metrics in Figure 6-8. This is a strategy that was adopted by the pilot where strict briefing on the strategy to be adopted was lacking. In the ADS-33E-PRF task, the pilot interpreted the requirements to "maintain heading" only within the hover maintenance segment of the task.

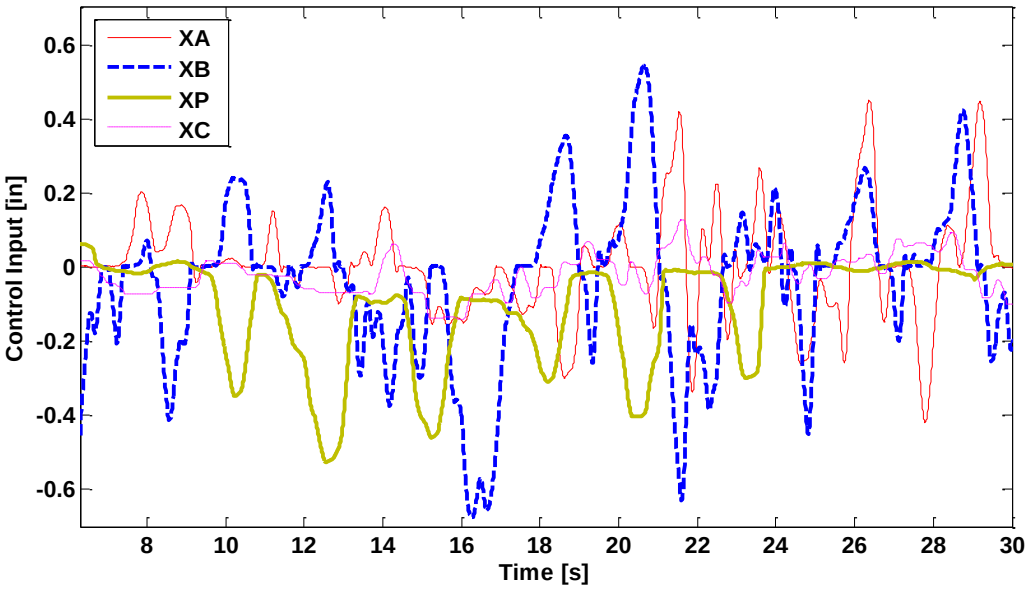


Figure 6-12 - Co-Ordination of XB and XP during Transition Segment

To try and further understand how the control of different axes influence the perception of workload and compensation, the data from the various axes engaged cases has been analysed. Figure 6-13 shows the median of the lateral and longitudinal control metrics across all high turbulence runs for various control configurations in the hover segment of the Precision Hover. The pilot is using more frequent, and more rapid, longitudinal cyclic inputs (XB) in the longitudinal only task than he is using lateral cyclic inputs (XA) in the lateral only task with similar magnitudes. As there are no guidance task demands in the hover segment of the PH, the control activity is entirely compensatory, due to turbulence combined with vehicle and/or task cue deficiencies. The source of these control metric results is shown in example control time histories in Figure 6-14, where it can be seen that there are less XA control inputs than XB (lower attack per sec) and that the XA inputs have lower gradients (lower attack rates).

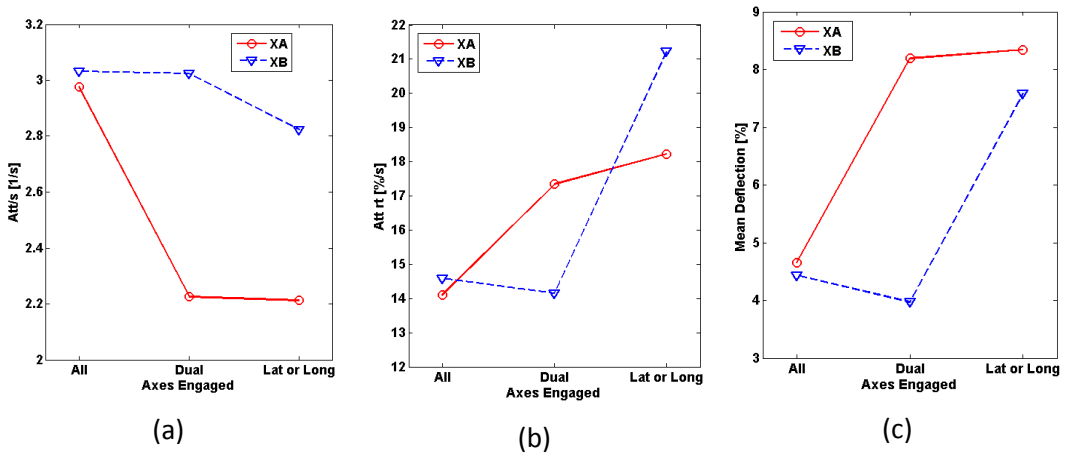


Figure 6-13 - Hover Maintenance Phase Lateral and Longitudinal Cyclic Metrics for High Turbulence

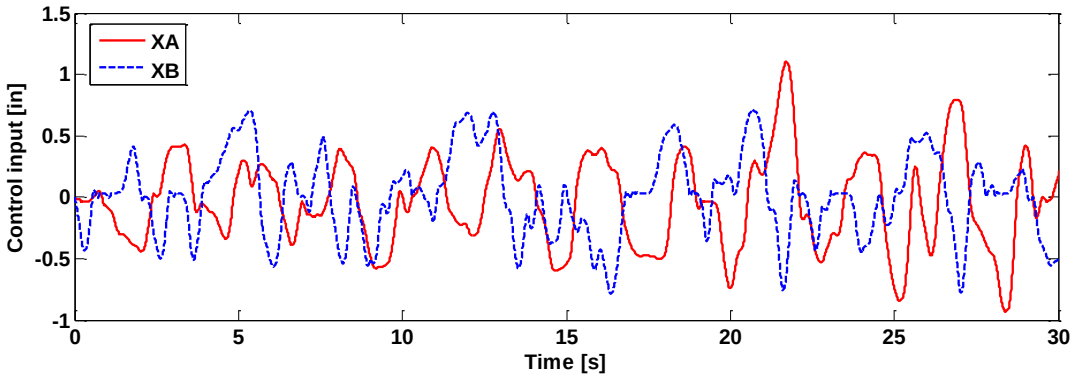


Figure 6-14 - Example High Turbulence, Lat Only and Long Only, Control Time Histories (Hover Maintenance)

There are a number of reasons for the differences in compensation between the two axes. Firstly, the aircraft is more responsive (higher quickness and rates obtained) to lateral cyclic control inputs than longitudinal control inputs as shown in Figure 6-15 . The reason for this is the design of the RCAH SCAS system. Pitch inertia is larger than the roll inertia therefore it is easier to get a faster response and more control power in the roll axis. The maximum roll attitude of the Bell 412 ASRA is approximately 60 degrees and the maximum pitch attitude is approximately 30 degrees. From a HQ perspective, pilots prefer to be able to obtain the same attitude in each axis at the same time with approximately the same control input. This is known as control harmonisation. To achieve this, the roll axis is quickened, so that the same steady state attitude is obtained after the same time from the same control input. Therefore, to achieve the same rate response (to compensate for the same rates of disturbance), less rapid and smaller control inputs required in XA than in XB. The second source of variation in compensation is that, because a higher level of pitch rate turbulence

is being applied to the airframe for a given turbulence condition (see Figure 6-3), there is less compensation required in the lateral axis to correct for turbulence.

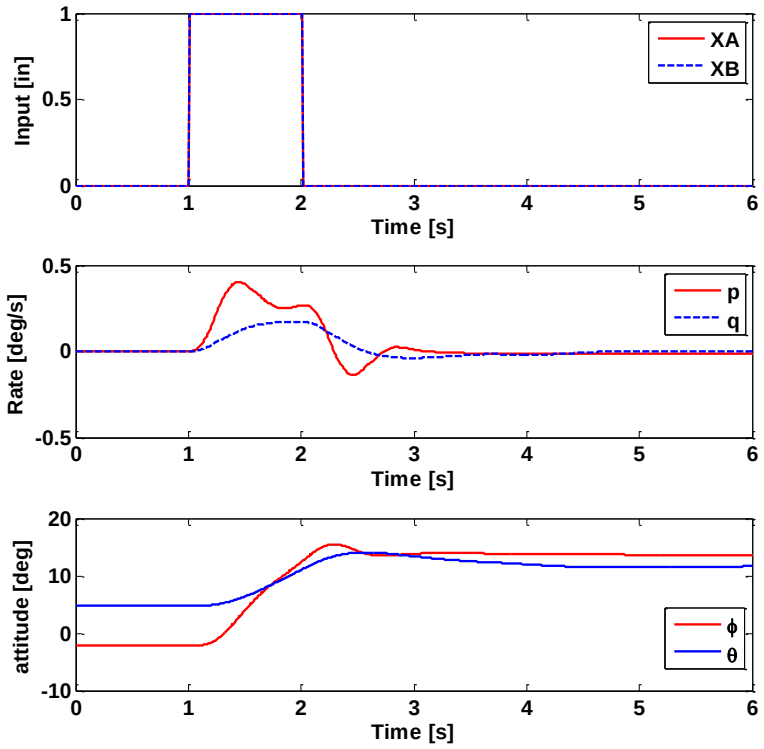


Figure 6-15 - Pitch and Roll Rate Response to a 1s, 1 Inch Pulse in XA and XB

Despite the higher control activity in the longitudinal only task, the pilot did not *perceive* a higher compensation (Figure 6-6). On first exposure to the high turbulence lateral and longitudinal only tasks, the pilot rated HQR=4 in both cases.

To determine whether there was any differentiation between the two ratings of HQR=4, the pilot was asked;

"Which did you find more difficult out of longitudinal only or lateral only if any?" (P. Perfect).

To which the pilot replied

"I think they are comparable. I wouldn't easily rate one harder than the other. However, lateral felt more natural because there was an attitude retention and lateral position control loop at work, whereas longitudinal I evolved into a position to stick transfer function that is effective but unusual" (R. Erdos).

In the Precision Hover task, the pilot was supplied with cues for lateral position and roll attitude from the hover board in the centre of his vision (see Figure 6-4). For pitch attitude,

the pilot had cues from the hover board. For longitudinal position, the pilot could utilise the course cones to the left and right (Figure 5-7), 5ft grids on the grass and a parallax effect between the pole and the hover board (see Figure 6-4).

The pilot perceived that in the lateral only manoeuvre, he used only the hover board for roll attitude and lateral position retention. However, in the longitudinal only task he perceived that he was using the board for pitch attitude control and the 45 degree cones to the left for longitudinal position reference.

Further research would be required to determine whether the differences in the cueing in the lateral and longitudinal axes are the cause of the discrepancy between perceived workload and control activity.

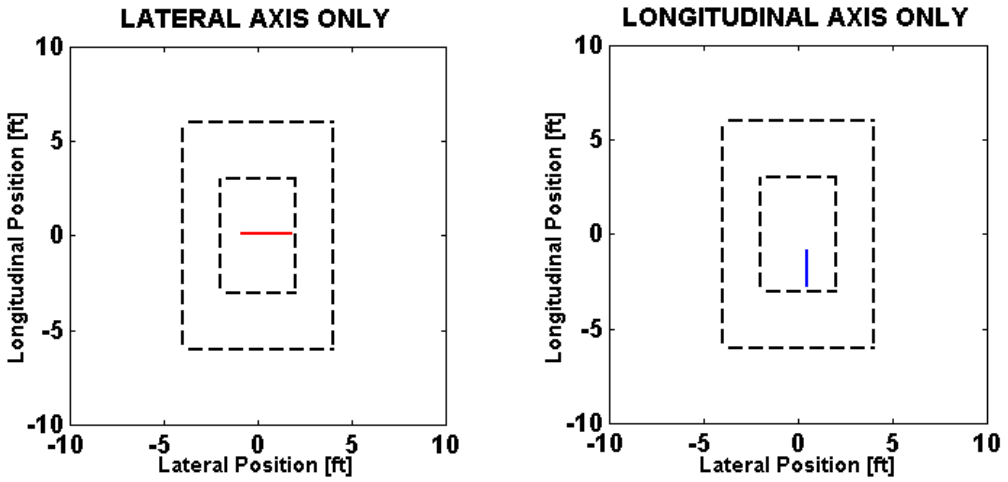
Figure 6-13 also reveals that when the pilot's workload becomes divided between longitudinal and lateral cyclic (dual-axis task) at high turbulence, much smaller (reduction in $\bar{\eta}$) and slower (reduction in attack/sec) inputs are being made in the longitudinal cyclic than for the single-axis task with slightly more active time on the control (increase in attack/sec), whereas the lateral cyclic activity remains similar.

The pilot stated that;

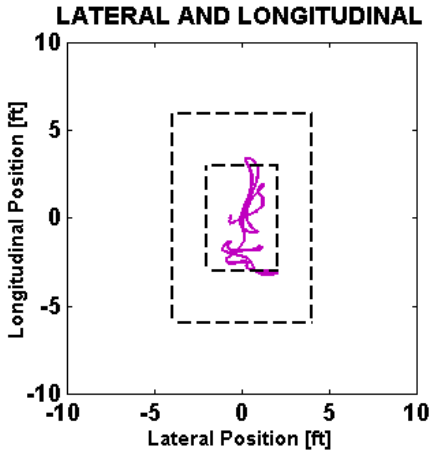
"Lateral errors convolute the cueing in the longitudinal axis" (R. Erdos).

This is in reference to the fact that the pilot was using the 45 degree cones for longitudinal positioning and as the lateral position changed the perspective of the 45 degree cones changed also. This may be the cause for the reduction in longitudinal control activity seen in Figure 6-13: If the pilot was not confident in his longitudinal cues while there were lateral errors, it would be expected that he would concentrate on reducing lateral errors before making any changes to longitudinal position. A more detailed study into this effect would be required to confirm this hypothesis, which is beyond the scope of this work.

The effect of going from single axis to dual axis control on task performance at high turbulence can be seen in Figure 6-16. In the longitudinal only task the pilot is using less of the desired range than in the lateral only task. In the dual-axis task the lateral performance remains similar but the longitudinal performance is degraded. This reflects the reduction in XB activity and lack of change in XA activity seen in Figure 6-13.



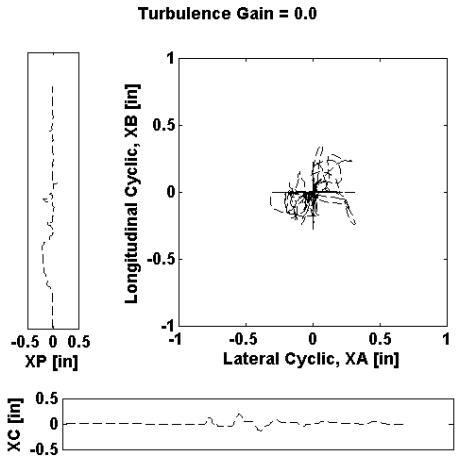
(a) (b)



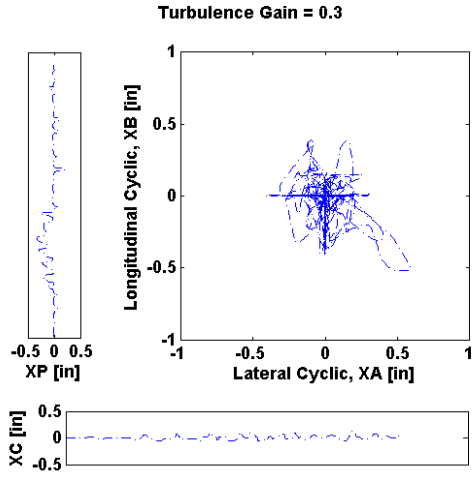
(c)

Figure 6-16- High Turbulence - Effect of Dual Axis Control on Performance

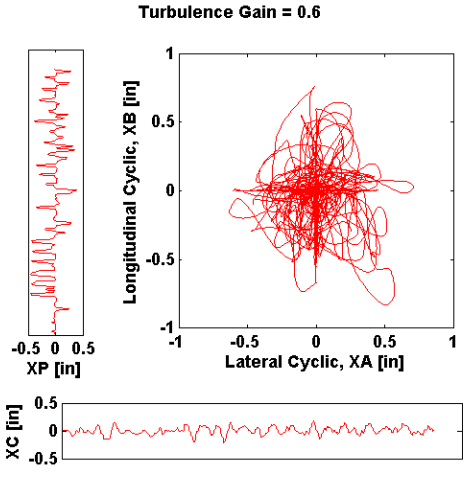
Figure 6-13 also shows that the addition of pedal and collective control causes the pilot to use more, but smaller and slower inputs in lateral cyclic, with little effect on the longitudinal cyclic strategy. The pilot is able to use pedal as well as lateral cyclic to minimise lateral position errors and therefore lateral cyclic activity reduced. This results in a strategy whereby the pilot controls the lateral and longitudinal cyclic in unison by stick-stirring and makes further lateral corrections using pedal, thereby reducing the number of control variables to three rather than four. The stick stirring behaviour is confirmed by the very similar attack metrics in XA and XB in the all-axes active case and the cross-plots in Figure 6-17 that show the lateral and longitudinal cyclic movements throughout the all-axes engaged PH tasks in all turbulence levels.



(a)



(b)



(c)

Figure 6-17 - Cross-Plot of XA and XB Activity for All-Axes Precision Hover Task.

The study into the effects of control axis dominance has shown that despite a requirement for higher compensation in the longitudinal axis, the pilot did not perceive higher workload. When controlling lateral and longitudinal cyclic simultaneously, it was found that the control activity in the longitudinal cyclic decreased but the lateral cyclic control was unaffected. Furthermore, the addition of pedals and collective affected that lateral cyclic activity much more than the longitudinal cyclic activity.

It was also confirmed that the compensation required in the Precision Hover is dominated by cyclic control activity at low turbulence but the extent of this domination reduces with increased turbulence due to the way the turbulence was introduced. Therefore, it is proposed that the efficacy of the additive attack metrics may be optimised through weightings for each axis of control.

The dominance of different axes was also found to be different in the transition phase to the hover phase. This suggests that weightings for the different control axes is dependent on the type of task - guidance or compensatory for example. Section 6.4 discusses how such weightings may be determined.

Phase 2-Multiple Pilots and Measures of Spare Capacity

In this section, results for the dual-axis engaged Precision Hover MTE are presented for the two participating test pilots. The increase in workload for each pilot was analysed in terms of the degradation of secondary task performance and increase in control attack metrics and the correlation between the metrics is discussed. The primary task performance was presented in terms of hover plan position. The secondary task performance was measured in terms of percentage of correct responses to bird sightings and the statistical spread of time to trigger response.

Figure 6-18 and Figure 6-19 show the primary task performance in terms of the lateral and longitudinal positioning during the hover maintenance phase for pilot A and pilot B respectively. The dashed lines represent the desired performance tolerances (Table 5-1). For both pilots, the performance became poorer as the turbulence increased, as expected, but the addition of the secondary task did not significantly affect the primary task performance for either pilot in terms of level of performance. This confirms that the pilot has successfully prioritised the primary task and allowed the secondary task performance to degrade as spare capacity is consumed by the primary flight task as briefed.

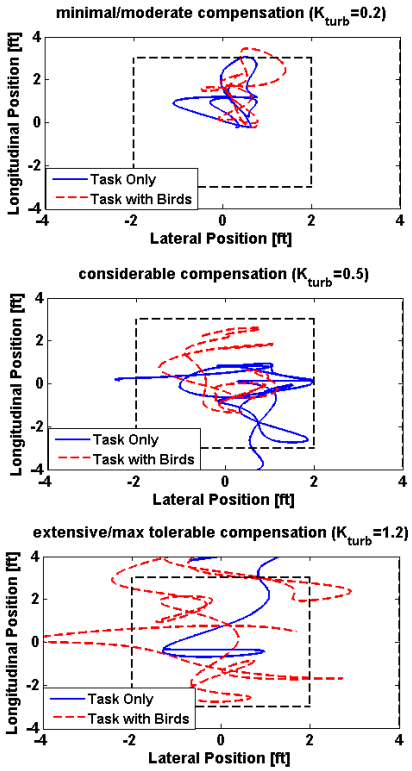


Figure 6-18 - Pilot A Position Keeping Performance With and Without Birds

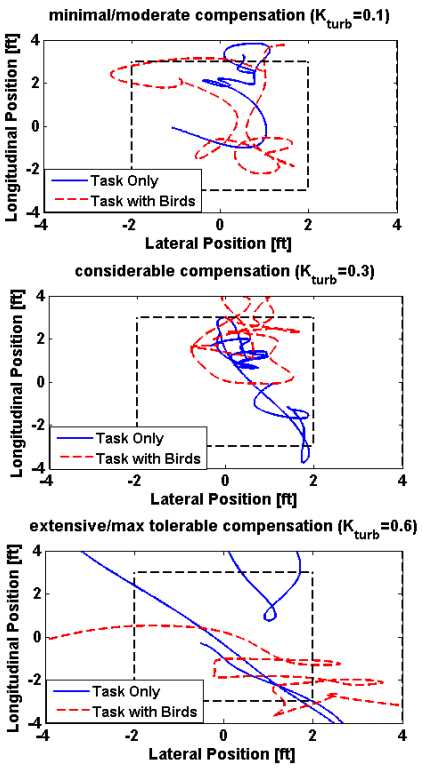
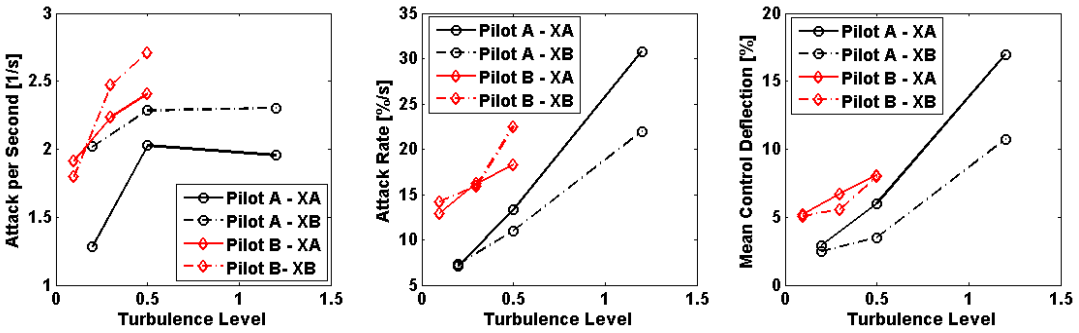


Figure 6-19 - Pilot B Position Keeping Performance With and Without Birds

Lateral and longitudinal cyclic control attack metrics from the runs with birds are shown in Figure 6-20, showing the expected increase in control attack with turbulence as in phase one of the exploratory trial (Figure 6-7 and Figure 6-8). Figure 6-20 highlights the difference in control strategy between the two pilots and shows that pilot B uses a higher control attack strategy for a given turbulence than pilot A.



(a) (b) (c)

Figure 6-20 - XA and XB Control Metrics (With Birds)

As turbulence increases, both pilots use larger and more rapid control inputs. However, the different gradients of the plots in Figure 6-20 suggest that pilots respond to the increase in

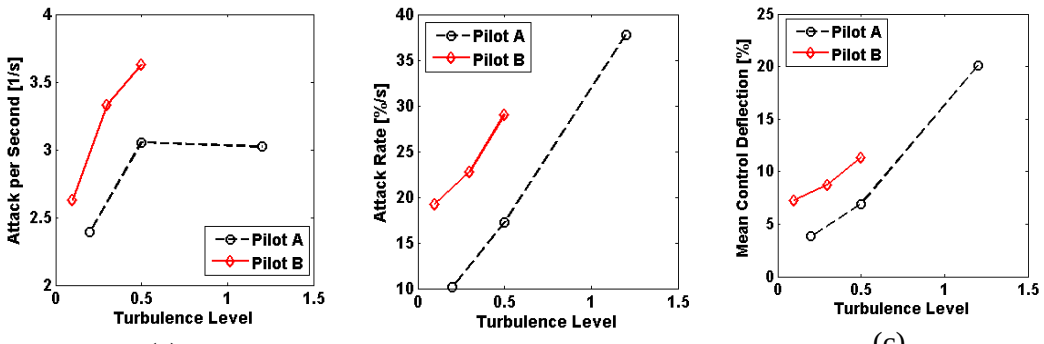
turbulence in slightly different ways. There is a clear plateau in attack/sec for pilot A in Figure 6-20a. An increase in attack/sec in combination with increasing mean control deflection and attack rate reflects that the pilots are active on the controls for a higher proportion of the task. Therefore, the plateau in attack/sec at illustrates that pilot A's activity becomes saturated at high turbulence.

Initially, Pilot A's XA attack/sec, attack rate (Figure 6-20b) and control deflection (Figure 6-20c) increase more rapidly than Pilot B's. However, Pilot B's XB attack/sec, attack rate (Figure 6-20b) and control deflection (Figure 6-20c) increase more rapidly than Pilot A's. This suggests that where pilot A's workload increase is dominated in XA, pilot B's is dominated in XB. This difference in response from different pilots shows that a single-axis metric cannot capture the same change in workload of different pilots.

In light of the 'stick-stirring' behaviour observed in the exploratory trial, the cyclic was analysed as a single control action, by taking the resultant amplitude of the lateral and longitudinal cyclic vector:

$$\boldsymbol{CYC = \sqrt{XA^2 + XB^2}} \qquad \textbf{Equation 6-1}$$

The cyclic attack can then be determined from the time history of the cyclic control in the same way as the attack of the individual lateral and longitudinal sticks. The cyclic attack metrics can be seen in Figure 6-21. Here it can be seen that overall, Pilot A and Pilot B show roughly the same trend in increasing attack with increasing turbulence with pilot B having a consistently higher workload.



(a) (b) (c)

Figure 6-21 – Integrated Cyclic (CYC) Control Metrics (With Birds)

The difference between the two pilots is further highlighted in Figure 6-22, where the additive metric plots for both pilots from the data obtained in phase two of the exploratory trial are shown. The ratings from Pilot B are much closer together than those of pilot A. This

suggests that a small increase in turbulence has a stronger effect on pilot B's perception of workload with a relatively small change in control activity compared to Pilot A, highlighting Pilot B's increased perceptual sensitivity to the effects of turbulence. In terms of the proposed Level 1/Level 2 workload boundaries, the metrics associated with Level 1 HQR test points lie within the proposed boundary and all metrics associated with Level 2 HQRs awarded lie outside this boundary. This increases the confidence in the validity of the proposed boundary for handling qualities assessment as it holds for two pilots with diverse control strategies.

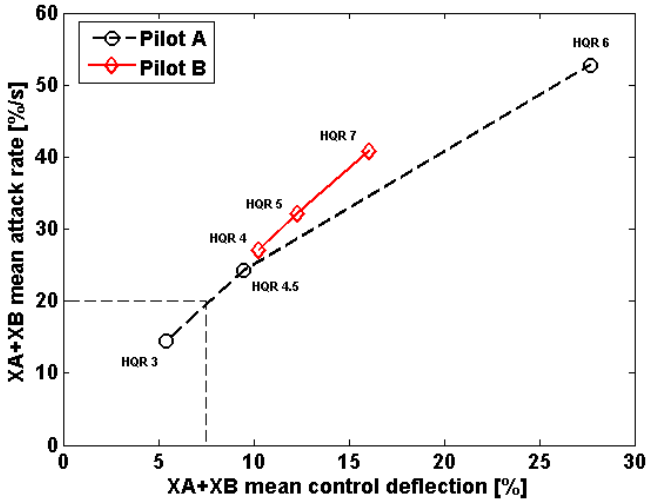
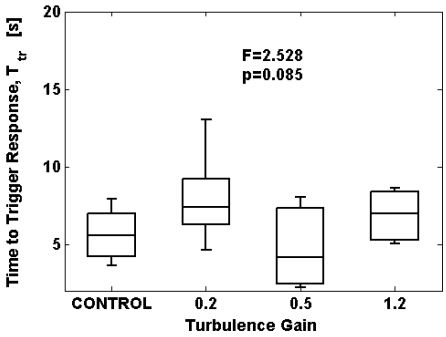


Figure 6-22 - Proposed Additive Metric Plot (Hover Segment Metrics) for Both Pilots

The time between the birds entering the visual scene and the pilot pressing the cyclic trigger to confirm identification of the birds has been named 'time to trigger response', denoted by T_{tr}. Figure 6-23 and 6-24 show time to trigger response for pilot A and pilot B for varying levels of turbulence and the control case (no primary task) respectively. It was found that the effect of task difficulty on time to trigger response for both pilots was not statistically significant (ANOVA test $p>0.05$). It is suggested that due to the requirement of the pilot to scan the entire visual scene, the time taken for the pilot to identify the pilot, and therefore the time to trigger response, was more sensitive to where the pilot was looking when the birds entered the scene than his workload. Figure 6-25 shows that the secondary task performance, in terms of percent of correct recall of the number of birds in the flock, degrades significantly with increasing turbulence gain for both pilots. Pilot A's reduction in secondary task performance is not as marked as pilot B. This suggests that pilot A has a lower workload in flying the task and therefore a higher spare capacity to scan the visual scene and compute the number of birds in the flock. This result is in agreement with the lower control activity used by pilot A compared to Pilot B (Figure 6-21).



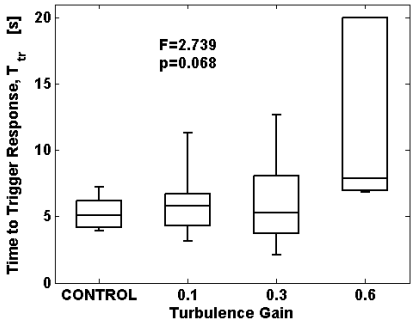


Figure 6-23- Effect of Turbulence on T_{tr} - Pilot A **Figure 6-24 - Effect of Turbulence on T_{tr} - Pilot B**

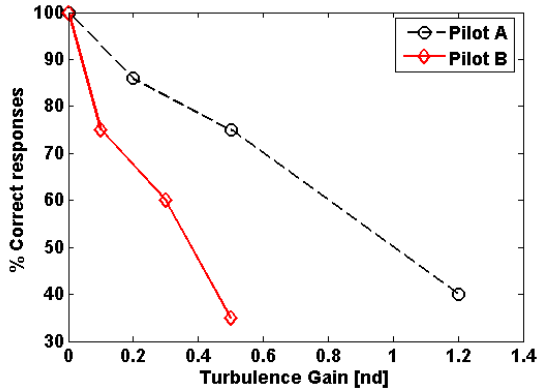


Figure 6-25 - The Effect of Turbulence on Performance of Recalling the No. of Birds in Flock

Although statistical analysis of time to respond to birds in the visual scene showed no correlation with the increase in workload (Figure 6-23 and Figure 6-24), the percentage of correct responses to bird sightings did show correlation to the pilot workload.

One of the objectives of this trial was to determine the potential of measures of spare capacity for quantitative workload assessment. 'Spare Capacity' has no quantifiable definition, however, the secondary task performance, measured as % correct responses, is an indirect indication of spare capacity - the higher the % correct responses, the higher the spare capacity. Essentially, 0% correct answers implies no spare capacity and 100% correct answers implies full spare capacity.

Figure 6-26 shows that the control activity results aligned well with the measures of secondary task performance. The general correlation implies that the control activity metrics fully reflect the pilot workload. However, one anomaly was seen, circled in Figure 6-26. This was the HQR = 7 test case (Pilot B, K_{turb}=0.6). It is suggested that at high levels of workload the relationship between control activity and workload begins to breakdown as the pilot tries to back out of the loop to avoid loss of control, therefore control activity measures are lower than expected by the general trend.

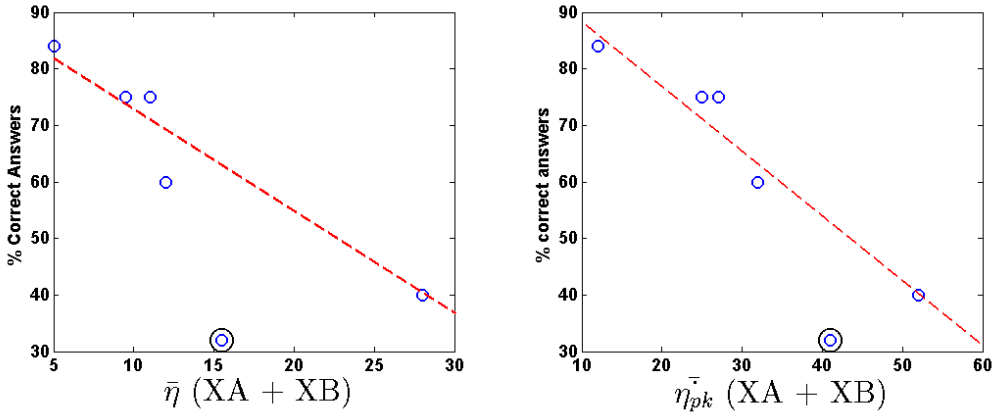


Figure 6-26 - Correlation of Attack Metrics with Secondary Task Performance (Hover)

The potential for the use of % correct (or incorrect) responses in a visio-spatial task for Precision Hover HQ analysis was assessed by examining the relationship between the % of incorrect responses and the awarded HQR, as shown in Figure 6-27. From inspection of where the data crosses the HQ boundaries, it may be inferred that score above approximately 80% indicates Level 1 HQs and a score below approximately 40% indicates Level 3 HQs. Further studies may provide further refinement and validation of these preliminary approximations and it should be noted again that such boundaries may be task and vehicle specific.

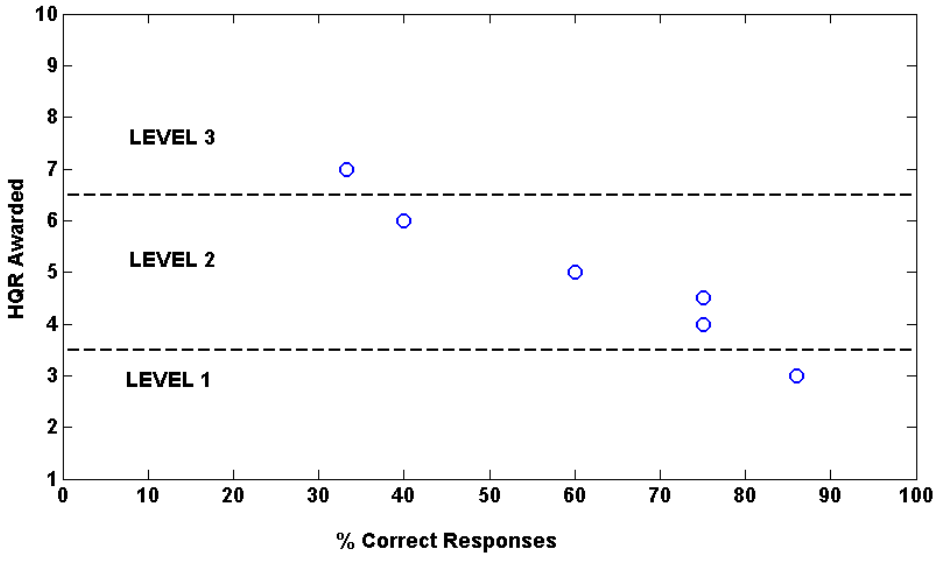


Figure 6-27 - Relationship between HQR and % Correct Responses to Bird Sightings

6.4 Metric weightings and the Adaptation Metric

It has been shown in this chapter that pilot workload can be more accurately measured by considering all axes of control simultaneously. However, it has also been noted that it may be possible to obtain a metric that better reflects pilot opinion of compensation/workload/adaptation by considering that each axis has a different dominance of the pilot perception.

The motivation for the work presented in this chapter was to be to develop improved metrics for measurement of task strategy adaptation. The lessons learned from this study prompted the realisation that information of the dominance of each control axis on the pilot perception of adaptation was contained within the pilot questionnaires (see Table 6-6).

Table 6-6 - Section of Task Strategy Adaptation from SFR Questionnaire

Task Strategy (Flight dynamics)	representative strategy - Negligible strategy adaptation required	Minimal strategy adaptation required	Moderate strategy adaptation required	Considerable strategy adaptation required	Excessive adaptation required	Completely dissimilar strategy required
Lat. Cyclic						
Long. Cyclic						
Collective						
Pedals						
Comments *highlight worst case*						

When completing the questionnaires, the pilots broke down task strategy adaptation into the four control axes. This information has been used to assign a weighting score, W, to each axis depending on the assigned level of adaptation in each axis. The assigned weightings as shown in Table 6-7

Table 6-7 - Weightings Applied to Each Axes Dependent on the Questionnaire Responses

Level of adaptation awarded	Negligible	Minimal	Moderate	Considerable	Excessive	Completely dissimilar
Weighting, W	0	1	2	3	4	5

The Adaptation metrics can then be calculated by first multiplying the percent change in a metric in each axis by the appropriate weighting function and then summing the resultant weighted metrics in each axis. For example, the adaptation in number of attack points per second can be calculated as follows;

$$Adaptation_{f_n} = |\%\,\Delta\, f_{n_{XA}}| * W_{XA} + |\%\,\Delta\, f_{n_{XB}}| * W_{XB}|+\%\,\Delta\, f_{n_{XC}}| * W_{XC}|+\%\,\Delta\, f_{n_{XP}}| * W_{XP}$$

Equation 6-2

The adaptation in all three attack metrics have been calculated for all pilots' PH, transport delay test cases from Chapter 5. Pilot B did not complete SFR questionnaires so his adaptation could not be measured. The adaptation corresponding to the SFR 10 awarded by pilot G was also not included as the pilot was not able to complete the run.

To demonstrate the benefit of the weighting methodology, the adaptation metric has been calculated with weighting factors and without. Figure 6-28, Figure 6-29 and Figure 6-30 show the correlation between the adaptation metrics (with and without weighting factors) and the awarded SFR.

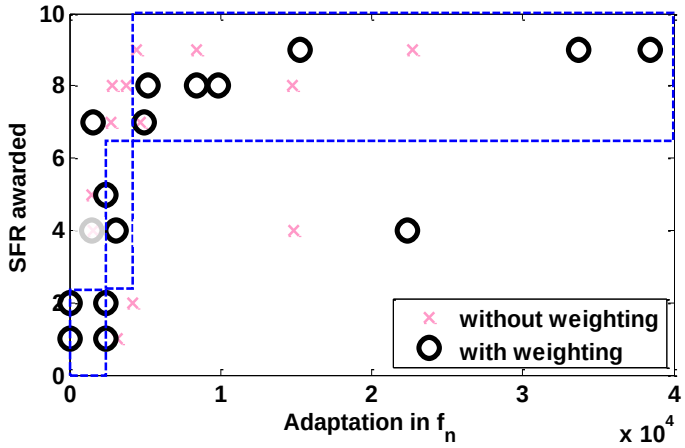


Figure 6-28 - Correlation of SFRs with Adaptation Metric for F_n (Precision Hover ATD Test Points)

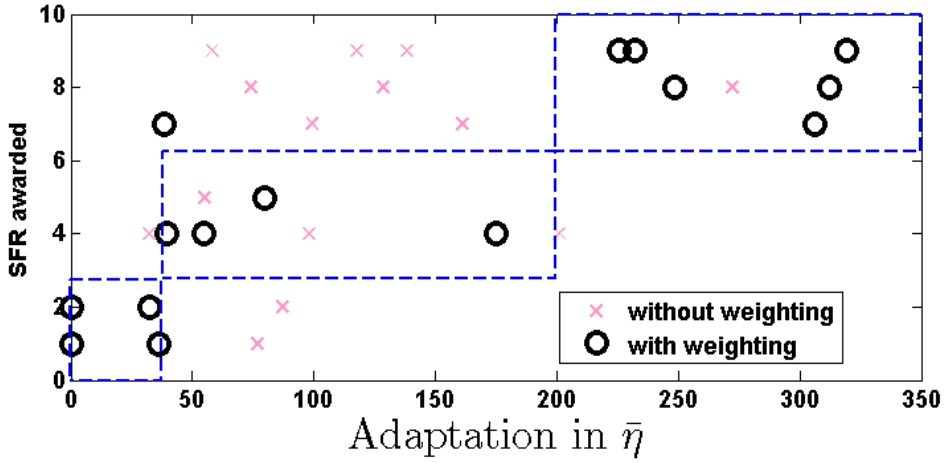


Figure 6-29 - Correlation of SFRs with Adaptation Metric for η (Precision Hover ATD Test Points)

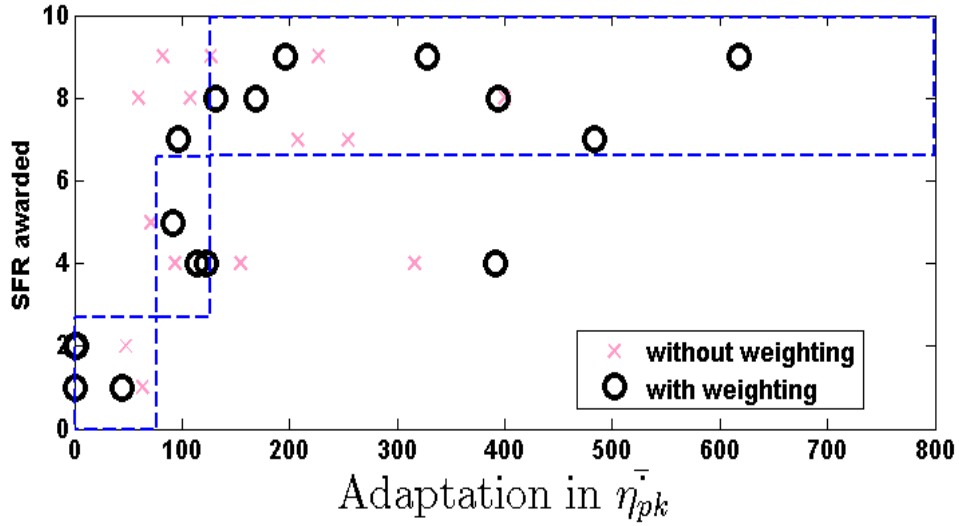


Figure 6-30 - Correlation of SFRs with Adaptation Metric for η dot pk (Precision Hover ATD Test Points)

The utility of this metric is further highlighted by plotting the two dimensional mean attack rate vs. control deflection plot - Figure 6-31. Apart from 2 points (highlighted by black circles), there are clear distinctions in the areas occupied by ratings in different fidelity levels. This allows initial Level 1/Level 2 and Level2/Level 3 fidelity boundaries to be proposed as represented by the lines in Figure 6-31. This exclusivity increases the confidence that the adaptation metrics have potential for quantitative perceptual fidelity assessment. The Level 3 rating that lies within the Level 1 region was awarded by Pilot A for ATD=300ms. It was noted in Chapter 5 that pilot A used an open-loop control strategy and took advantage of the RCAH system which meant that he did have to adapt his strategy very much. However, when giving his ratings he noted that he was thinking about the utility of the simulator for a line pilot who may not be as familiar with the aircraft and therefore would not adopt such a

strategy. The Level 2 rating that lies within the Level 3 region was awarded by pilot F and was the very last test point on the last day of a three day test campaign. Therefore this anomaly can be attributed to fatigue and desensitisation causing the pilot not to perceive the adaptation he was making.

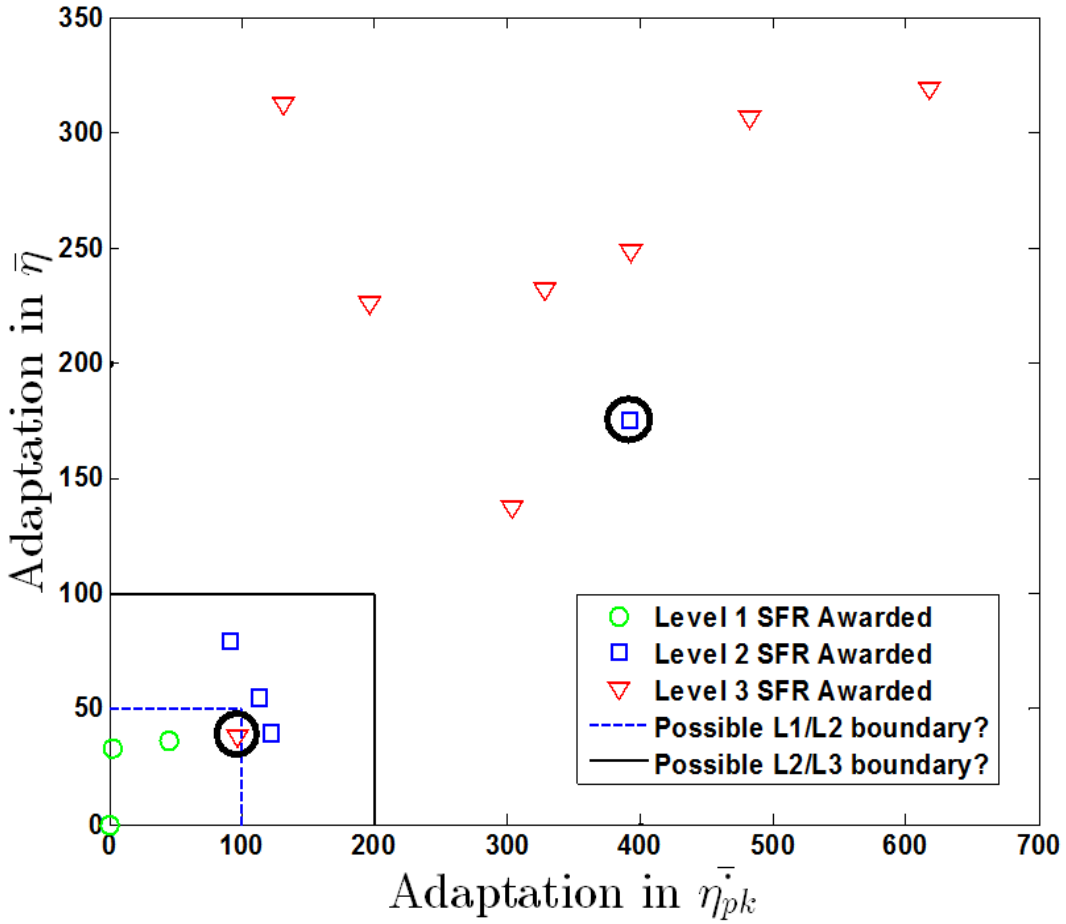


Figure 6-31 - 2-dimensional, Adaptation Metric for PH ATD Test Points Showing Regions of L1, L2 and L3 SFRs

Despite the success of the weighting methodology here, it should be noted that there are a number of disadvantages associated with using weightings in this way and care should be taken in using such methods. Below is a brief discussion of the disadvantages and how the method presented may be improved in the future.

Using weighted methods significantly increases the complexity of the data analysis and reduces visibility of the data after processing. Therefore the necessity and validity of the use of weighted variables should be considered carefully.

Before choosing to use weighted coefficients, the assumption must be made that all weighted variables are independent of one another. If comfortable with this first

assumption, it must then be chosen what value of weight should be used. In this study, the weightings have been determined directly from subjective opinion which means that any unreliability in the subjective results would be amplified in the quantitative results. As the number of independent variables increases (increase in number of control axes in this case), the result relies more and more on this subjective data.

Although they have not been used in this study, there are quantitative methods for weighting that are available in the literature which could minimise these effects, such as dominance weights and relative weights techniques [90].

6.5 Conclusions

Chapter 6 has reported results from an exploratory piloted simulation trial to examine the utility of control activity metrics and secondary task performance as measures of pilot compensatory workload for in multi-axis tasks. The control 'attack' metrics are based on the extraction of discrete control movements, as functions of amplitude and rate of control application. The principal conclusions that can be drawn from the work to date, and further research opportunities are identified below.

1) Investigation of individual control axes revealed that Pilot A was more sensitive to workload in the lateral axis than in the longitudinal axis. In dual-axis control configurations the pilot prioritised the performance in the lateral axis as turbulence increased. In a full authority configuration the pilot adopted a stick-stirring strategy; controlling the lateral and longitudinal cyclic as a single control in the hover but augmented lateral cyclic activity with pedal in the transition.

2) The control attack metrics showed sensitivity to increase in turbulence level and differing piloting strategies and statistically significant correlation with pilot perception of workload in terms of HQRs.

3) It was found that the correlation between the metrics and HQRs for multi-axis flying tasks can be further improved by summing the metrics in each axis engaged to create a multi-axis metric.

4) To fully reflect the pilot workload, both the rate and magnitude of the control inputs should be considered. A two-dimensional metric of mean attack/s against mean control deflection has been developed and a tentative boundary between Level 1 and Level 2 handling qualities has been proposed for this metric in the precision hover manoeuvre.

5) The use of a secondary task has shown potential for measurement of spare capacity and workload. Measurement of the percent of correct responses to bird sighting in the visual scene was degraded in correlation with attack metrics increasing and HQRs degrading. From the data available, boundaries on % correct answers to indicate the Level of Handling Qualities were approximated.

6) By using weighting functions based on pilot questionnaire responses, an adaptation metric has been developed that corresponds well with awarded SFRs in a multi-axis flying task (The PH) for variations in transport delay. Potential of this metric for perceptual fidelity assessment has been demonstrated. The disadvantages associated with using weighting metrics, such as assumptions of independent variables and increased significance of errors in subjective opinion, have also been discussed to avoid miss use of such methods and to promote further work to improve the technique described within this chapter.

7 PILOT MODELLING

In this chapter, Hess's modelling method for closed-loop fidelity assessment has been utilised in an inverse simulation to determine the effect of baseline vehicle handling qualities on the sensitivity of the 'ideal pilot' to flight model errors. The development and definition of the mathematical trajectory of a closed-loop Acceleration-Deceleration manoeuvre is presented and the pilot model and aircraft model are validated against piloted simulation data. The results of an offline study into the effects of Handling Qualities on fidelity requirements are then presented and discussed.

7.1 Aims

The spread seen in the subjective results presented in Chapter 5 was partly attributed to variation in pilot strategy. The explanation for this was that a pilot using a higher gain strategy will excite the aircraft dynamic to a greater extent and therefore is more likely to perceive fidelity issues when there are deficiencies in the flight model.

It was inferred from this that an aircraft with poor handling qualities (namely reduced stability) would lead to increased excitation of the flight dynamics and therefore an aircraft with poor HQs would cause the pilot to be more sensitive to flight modelling errors.

If this hypothesis were found to be true, the implication would be that flight model certification criteria should be dependent on the handling qualities of the aircraft to be simulated.

Thus far in the research, sensitivity to flight model variations of a Level 1 aircraft has been investigated. The aim of this chapter of the research was therefore to carry out a number of simulations with Level 1, Level 2 and Level 3 HQ baseline vehicle to test the hypothesis stated above.

7.2 Methodology

In 1981, Heffley noted that

"*Human operator control theory, combined with adequate quantitative description of training objectives, offers powerful potential for determining simulator fidelity requirements*" [90]

It was decided that rather than embarking on another set of piloted simulation trials, this study would be carried out offline using a pilot model and inverse simulation techniques. This methodology allows for efficient preliminary testing of hypotheses without the need for sophisticated models, hardware and evaluation pilots.

Figure 7-1 shows a simple schematic of the inverse simulation technique applied for a single control loop. The inverse simulation process works by defining a mathematical representation of a manoeuvre, in terms of time varying aircraft states to give a command trajectory. This command trajectory is then fed into a pilot model and then the output from the pilot model (the control activity) is fed into the aircraft model. A feedback loop allows the system to minimise the error between aircraft output trajectory and commanded trajectory. The control activity required to minimise this error can then be extracted from the pilot model as a time history.

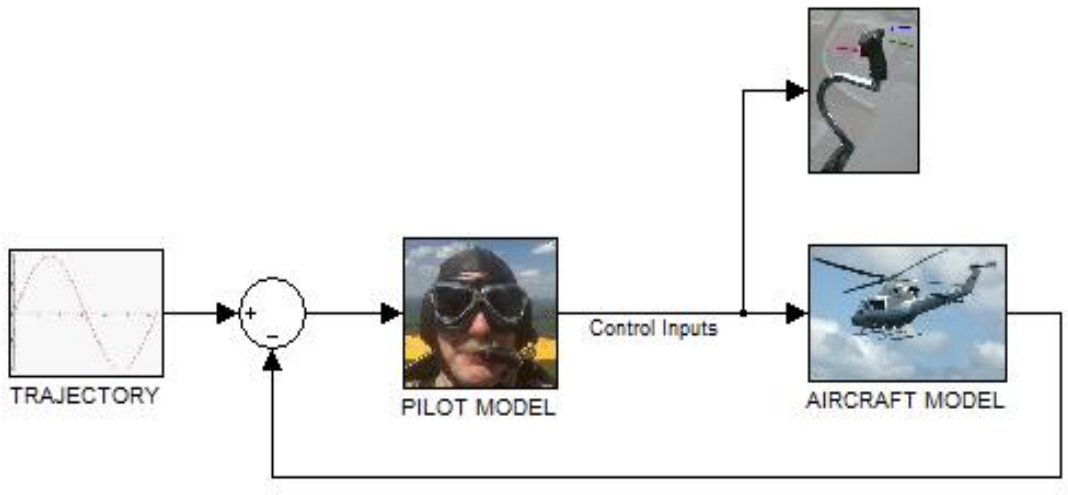


Figure 7-1 - Schematic of Inverse Simulation

In 1991, Hess proposed the use of such a technique for closed-loop fidelity assessment [51]. This methodology was described in Chapter 2 and so will not be covered in detail here. The methodology consists of 4 stages:

1. Select the task and vehicle to be selected

2. Develop mathematical models of the nominal and simulated vehicles.

3. Develop a control strategy – including time-optimal commands for discrete manoeuvres.

4. Select pilot model parameters

For the study into the effects of baseline HQs on sensitivity to flight modelling errors, three baseline linear aircraft models have been used; a predicted Level 1 HQ model, a predicted Level 2 HQ model, and a predicted Level 3 HQ model.

Each baseline model has then been modified using the same extents of Additional Transport Delay (ATD) as used in the piloted simulation trials covered in Chapter 5: ATD=100ms, ATD=200ms and ATD=300ms. This lead to a total of 3 baseline models and 9 modified models.

The task chosen for analysis was the Acceleration-Deceleration (AD) manoeuvre [19], which has been simplified to require longitudinal cyclic only control. This assumes an aircraft with zero cross coupling and a height hold function.

For each aircraft model, the optimum pilot strategy for the AD, defined through a set of optimised pilot model characteristic values, was calculated. The attack metrics were also calculated from the control time history as a further measure of control strategy.

It was hypothesised that the change in control strategy between the Level 1 models will be less than the change in control strategy between the Level 2 and Level 3 models.

The following subsections describe the development and validation of each of the components of the inverse simulation.

7.2.1 The Aircraft Model

With the intention of simplifying the inverse simulation in the first instance of its use, a 3 DoF linearised hover model of the Bell 412 has been used for the Level 1 baseline aircraft model. Only the longitudinal states, q, θ and u have been considered with only longitudinal collective, XB controller.

The nonlinear FLIGHTLAB model of the bare airframe has been linearised in the hover to obtain state space matrices. The resulting A matrix is shown in Table 7-1. The force derivatives are divided by aircraft mass and the moment derivatives by Aircraft moment of inertia, e.g. X/I_{xx}. There are limitations in using a linearised hover model for Accel-Decel analysis, including that many of the derivatives vary considerably with forward velocity, U. Future research should therefore consider the application of this methodology using a full nonlinear model. However, the simplicity of a linear model was desirable and the trade-off in model accuracy was deemed acceptable due to the comparative nature of the study.

Table 7-1- Bell 412 Longitudinal Hover Linearised Reduced 3 DoF A Matrix

A matrix	θ	u	q [rad/s]
θ'	0.0000 [1/s]	0.0000 [rad.s/ft]	0.9696 [nd]
X	-31.1173[ft/rad]	-0.0141 [1/s]	1.5933 [ft/s.rad]
M	-0.0077[1/s^2]	0.0062[rad/s.ft]	-1.3334 [1/s]

Table 7-2 - Bell 412 Longitudinal Hover Linearised Reduced 3 DoF B, C and D Matrices

B matrix	XB (θ_{1s})
θ	0.0000 [nd]
u	-0.6851 [ft/s.rad]
q	0.3097 [1/s]

C matrix	q	θ	u
θ	0	0	0.9703
u	0.9703	0	0
q	0	0.9703	0

D matrix
0
0
0

To determine the validity of the 3DoF linear model, data obtained from a pilot flying the Accel-Decel manoeuvre in the HELIFLIGHT-R facility with the full nonlinear FLIGHTLAB Bell-412 ASRA bare airframe model was used as a reference. The control time history from the piloted simulation data was fed into the 3DoF linear model (Figure 7-2) and the resulting pitch attitude, rate and forward speed responses were compared with the reference data. The results of this test are shown in Figure 7-3.

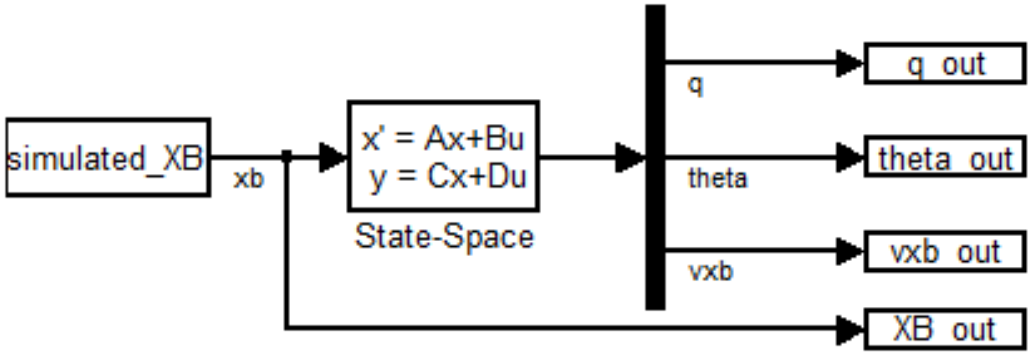


Figure 7-2 - 3 DoF Pitch State Space Model Forward Simulation Fed with Piloted Simulation Control Activity

It can be seen that the linear model responses agree well with the nonlinear responses in terms of trend and magnitude. The errors between the two sets of data are defined by the Root-Mean Square error (RMS) of the time varying signals:

$$RMS_\theta = 7.12\ °,\ RMS_q = 3.89\ °/s,\ RMS_u = 5.36\ kts$$

The main differences are reduction in control sensitivity in the middle section of the manoeuvre. The reason for this is that the control sensitivities are constant in the linear model but vary with forward velocity in the full nonlinear model. Despite this, the match was considered to be acceptable for use in this study.

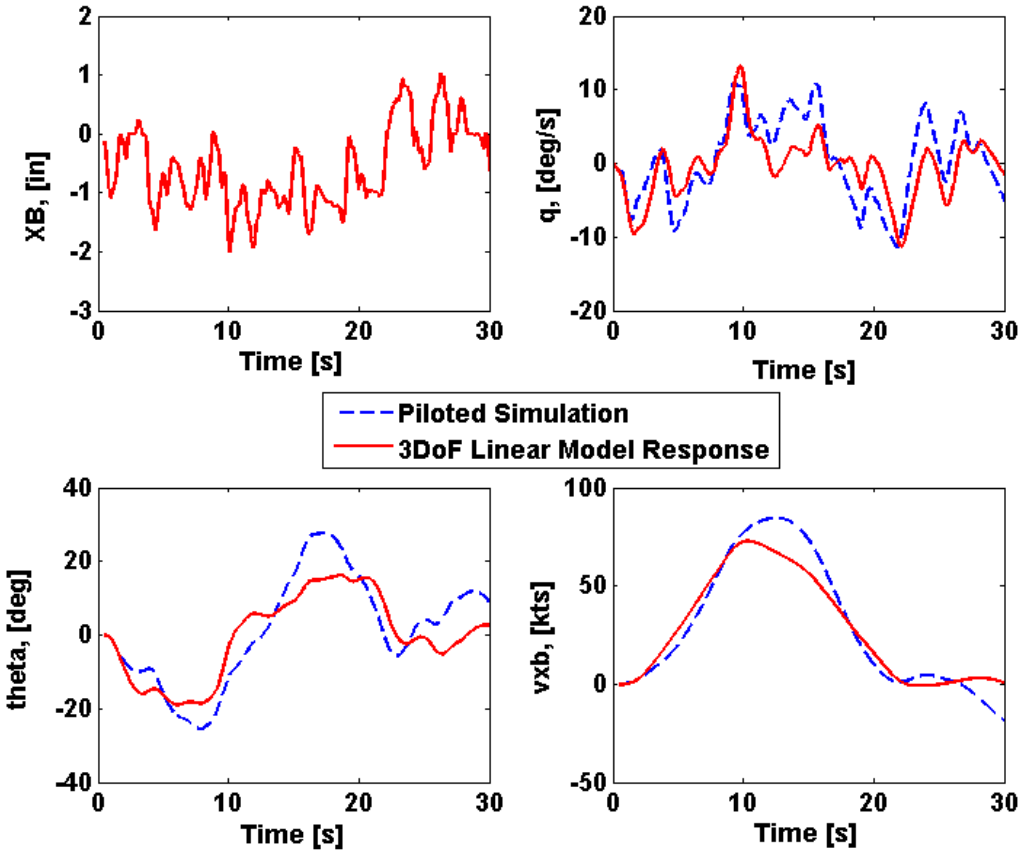


Figure 7-3 - Validation of the Linear Model against the Piloted Nonlinear Simulation

The Level 2 and Level 3 baseline models were obtained by modifying the pitch stability derivative, M_q. It was found that the baseline model had Level 1 pitch stability, by halving the value of M_q, a Level 2 pitch stability was obtained and Level 3 pitch stability occurred when that value is halved once more. This is shown against the ADS-33E-PRF pitch stability criteria for hover in Figure 7-4.The pitch bandwidth of all the models was also calculated using a one degree the one degree of freedom (pitch rate, q) model. The results of this are shown in Figure 7-5.

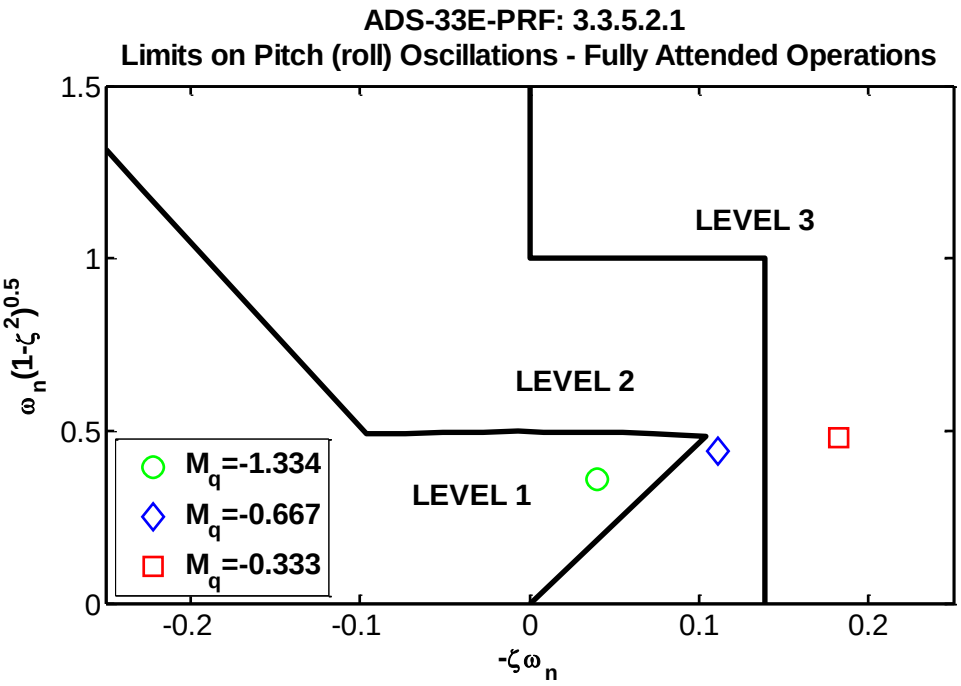


Figure 7-4 - Effect of M_q on Longitudinal Phugoid Stability in the Hover.

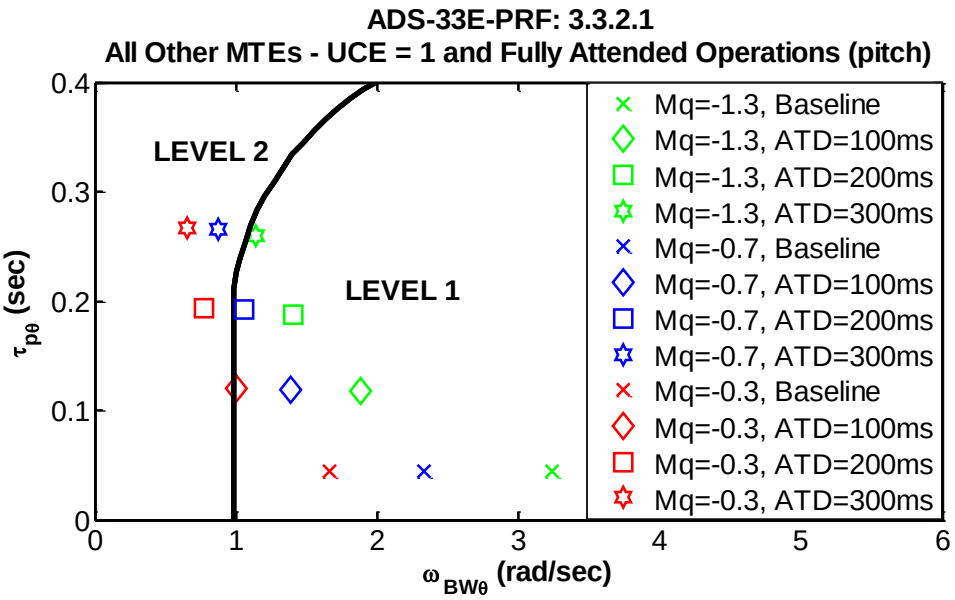


Figure 7-5 Effect of M_q and Transport Delay on Pitch Bandwidth in the Hover.

7.2.2 The Mathematical Trajectory of the Accel-Decel

The simplified, single axis Accel-Decel can be defined by a mathematical description of the longitudinal trajectory, in terms of both the longitudinal velocity and the longitudinal position. Traditionally, polynomials are used to define the trajectory. However, in this study, tau theory [92] has been used to define the mathematical trajectory.

Padfield developed the theory of how pilots utilise information regarding 'time to close on a goal', τ, when closing physical gaps in flight [93]. Tau theory is covered extensively in the literature and so will not be covered in detail here.

Suffice to say, the instantaneous, time varying range, x(t) and velocity x'(t) can be described in terms of normalised parameters, $\hat{x}$ and $\hat{x}'$, and the tau-coupling coefficient, k, as shown in Equation 7-1 and Equation 7-2 respectfully. The value of k determines the aggression of the manoeuvre. For the Accel-Decel, variation of k leads to changes in the point in the manoeuvre the deceleration begins. This is illustrated in Figure 7-6.

$$\hat{x} = -(1 - \hat{t}^2)^{\frac{1}{k}} \qquad \textbf{Equation 7-1}$$

$$\hat{x}' = \frac{2\hat{t}}{k}(1 - \hat{t}^2)^{\left(\frac{1}{k}-1\right)} \qquad \textbf{Equation 7-2}$$

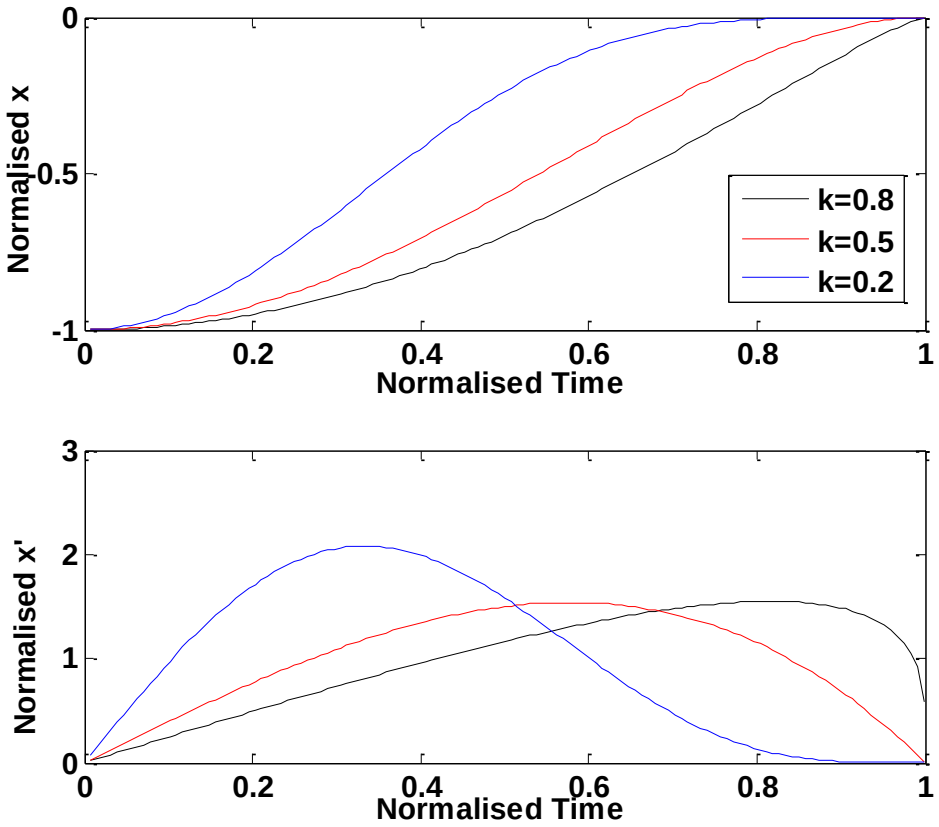


Figure 7-6 - Effect of Tau Coupling Coefficient, K, on Range and Velocity Trajectory

In the first attempt, it was assumed that the pilot is closing the gap of the distance between the start and finish point of the manoeuvre. A MATLAB code was used to optimise the value of k in Equation 7-2 to minimise the RMS error between the piloted simulation vxb data and the calculated tau guide. The optimum tau gain was found to be k=0.35 which gave an RMS in $\hat{x}'$ of 0.11. The match using this tau guidance is shown in Figure 7-7.

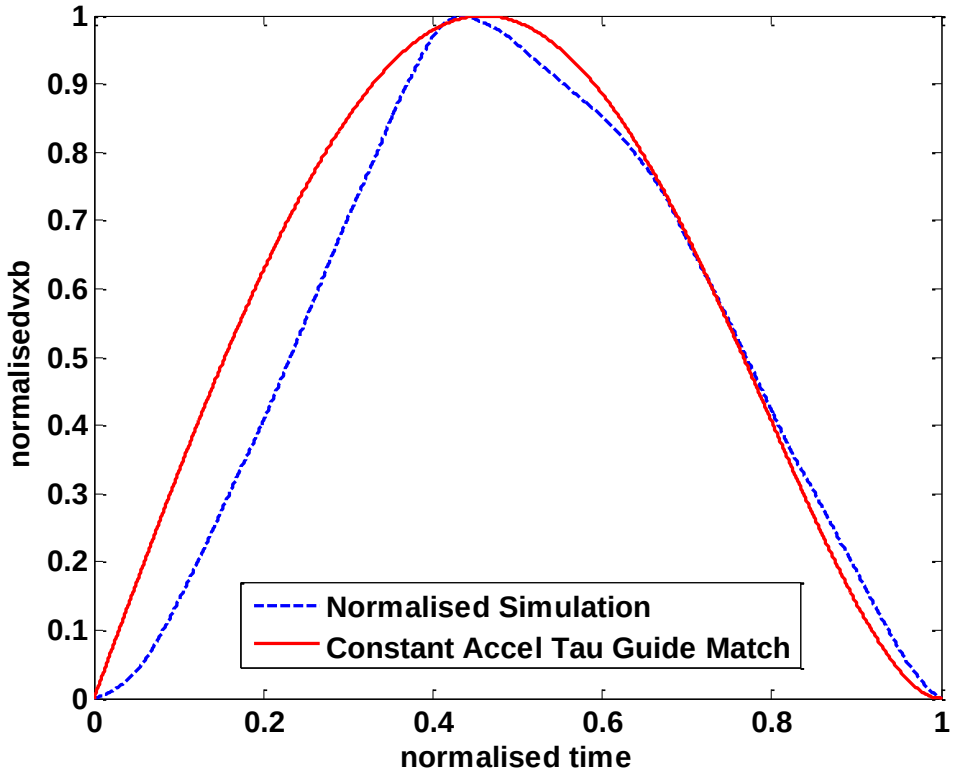


Figure 7-7 - Accel-Decel Tau Guide Optimised Match to Piloted Simulation Data

It was also required that the pitch attitude, rate and longitudinal position could be calculated for the tau guide to ensure the pilot was flying a true Accel-Decel. The pitch attitude required for a given speed (or distance) is dependent on the aircraft dynamics. Padfield et al noted a simple approximation to the relationship between range and pitch attitude:

$$Y_{A\theta} = \frac{X}{\theta} = \frac{-g}{s(s-X_u)} \qquad \textbf{Equation 7-3}$$

Therefore, if the range trajectory, x is known, then the pitch attitude trajectory can be defined as

$$\theta = \frac{-1}{g}(xs^2 - xsX_u) \qquad \textbf{Equation 7-4}$$

In Simulink, Equation 7-4 can be constructed using differential blocks as shown in Figure 7-8.

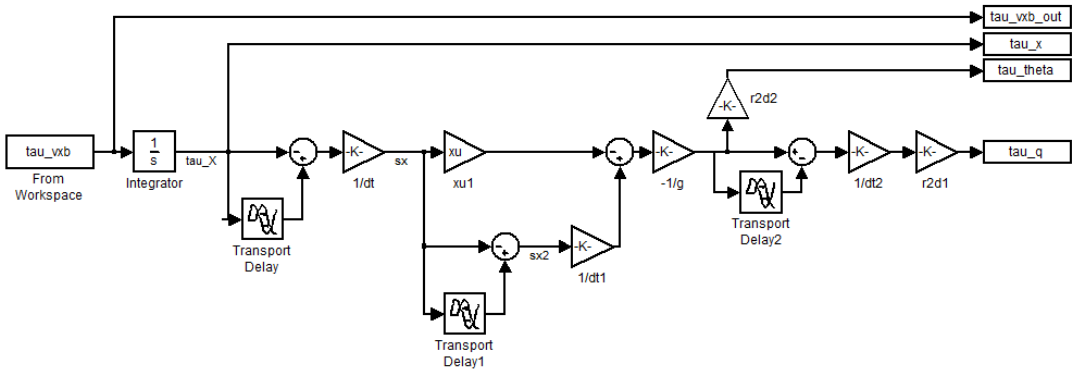


Figure 7-8- Converting Tau Speed Trajectory to Pitch Attitude Trajectory.

The x, u and resulting tau pitch trajectories for tau gain, k=0.35 are shown in Figure 7-9 against piloted simulation data for an AD MTE run in the Bell 412 bare airframe in HELIFLIGHT-R.

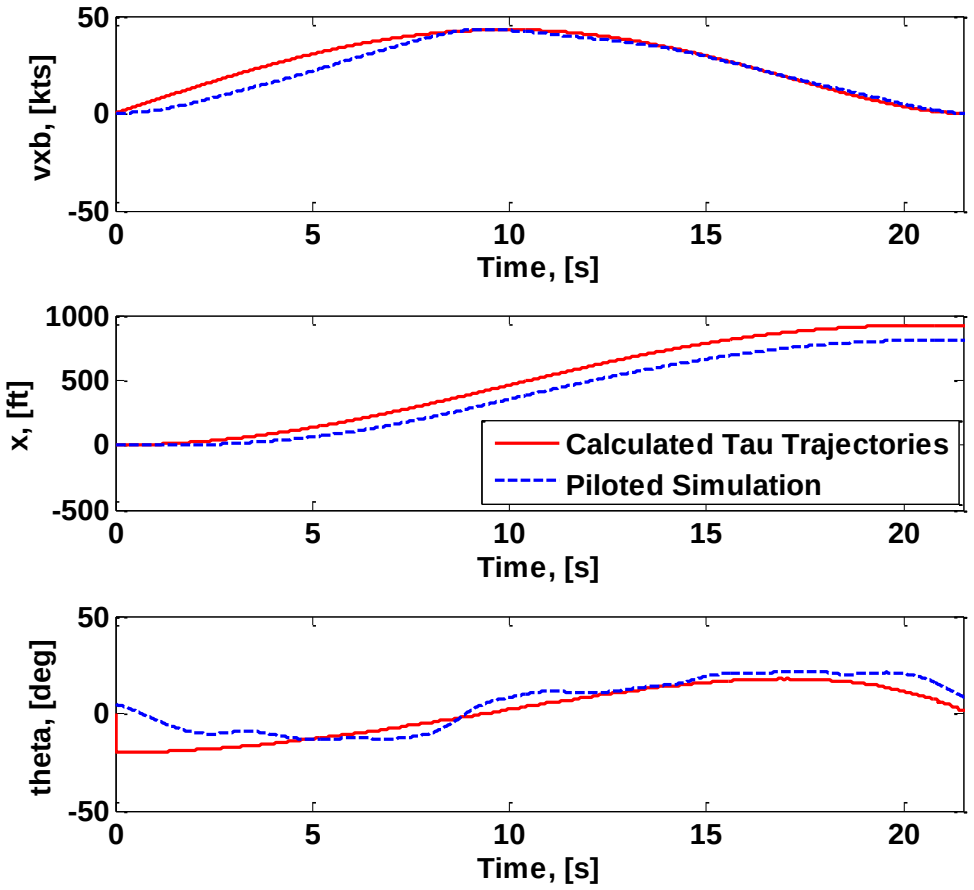


Figure 7-9 - Validation of Tau Trajectories against Piloted Simulation Data

The forward velocity, vxb, trace shows the tau guide overestimates the initial acceleration of the pilot and therefore the distance travelled is over estimated by approximately 20%. The pitch response becomes reasonably accurate towards the latter stages of the manoeuvre, however the pitch attitude at the manoeuvre offset is incorrect. The initial abrupt response was due to the abrupt start to the manoeuvre according to the tau guide.

To further improve the tau guide, the manoeuvre was reconsidered. It was reasoned that in the Accel-Decel, the pilot is not actually closing a ***range*** gap as originally assumed in Equations 7-1 and 7-2. The manoeuvre description briefs the pilot to reach a target velocity, not position and then return to zero velocity. Therefore, the pilot is closing two ***velocity*** gaps, one after the other, each with a different tau coupling coefficient dependent on the pilot strategy - k_1 for the acceleration phase and k_2 for the deceleration phase. Therefore the tau guidance equations for x' becomes;

$$\widehat{x'}_{acceleration} = 1 - (1 - \hat{t}^2)^{\frac{1}{k_1}} \qquad \textbf{Equation 7-5}$$

$$\widehat{x'}_{deceleration} = (1 - \hat{t}^2)^{\frac{1}{k_2}} \qquad \textbf{Equation 7-6}$$

A MATLAB code was then written to optimise the values of k_1 and k_2 such that the error between a reference simulation flight velocity trajectory and the tau approximation was minimised.

This involved the following steps:

1. Normalise the piloted simulation time and forward velocity trajectory
2. Find the point where the manoeuvre begins (accelerates through 0), where the maximum velocity is reached, normalised vxb=1, and where the manoeuvre ends (decelerates through 0).
3. Create two data files of time and forward velocity for the acceleration and deceleration phases of the manoeuvre and normalise the two data files.
4. Vary tau coupling coefficients, k_1 and k_2, until the RMS error between the normalised simulation and tau guidance trajectory velocities are minimised.

5. Reconstruct the manoeuvre using the optimised values of k_1 and k_2.

This lead to values of k_1=0.45 and k_2=0.6. The x, u and resulting tau pitch trajectories for these gain values are shown in Figure 7-10 against piloted simulation data for an Accel-Decel in the Bell 412 bare airframe in HELIFLIGHT-R. The two stage velocity tau guide resulted in a much closer match of the vxb trajectory with an RMS in $\hat{x}'$ of 0.0247 (previously 0.1105). The resulting over estimation in the range flown was reduced from almost 20% to 3%.

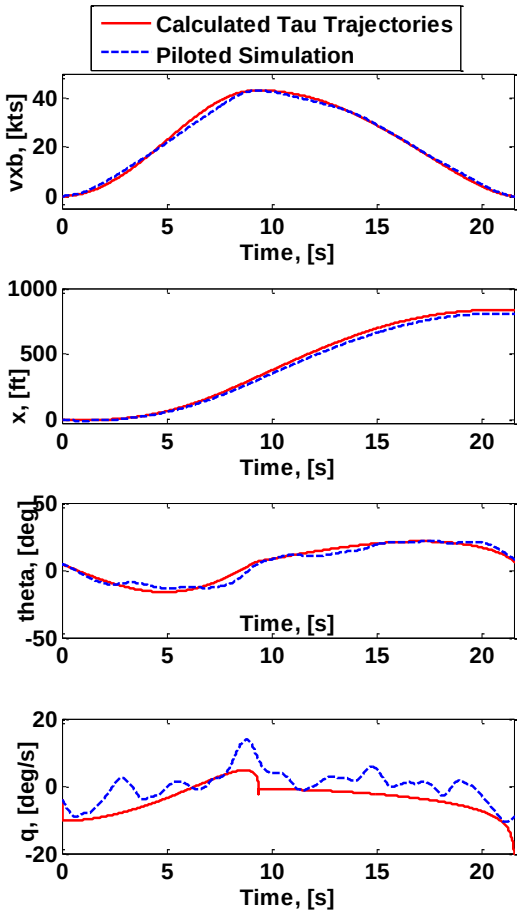


Figure 7-10 - k1=0.45, k2=0.6 Match between Tau Approximation and Piloted Simulation

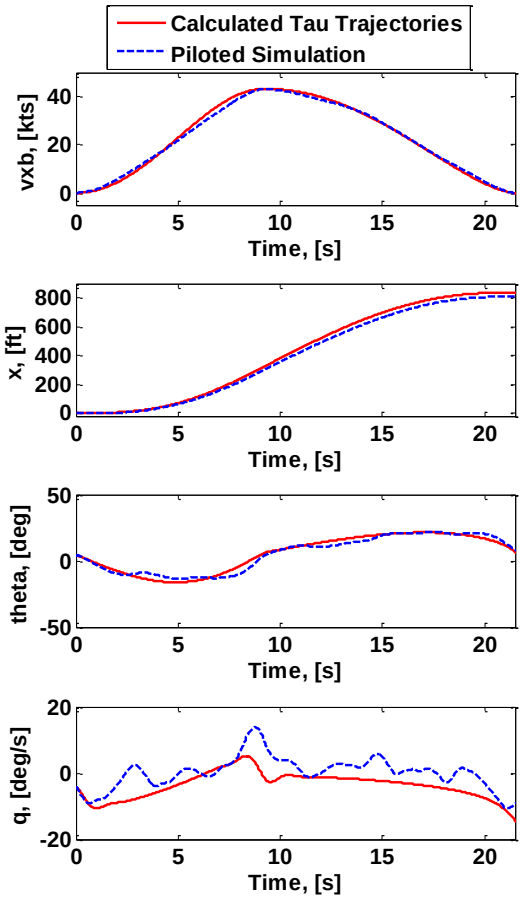


Figure 7-11 - k1=0.45, k2=0.6 Match between Tau Approximation and Piloted Simulation - SPLINED

The theta trajectory, and by extension the q trajectory, have singularity errors due to the differentiations. To mitigate this effect the output data was smoothed using the 'spline' function in MATLAB. The resulting match is shown in Figure 7-11 and shows a much more reasonable pitch rate trajectory. These trajectories are the final set of trajectories used for the study described in this chapter.

7.2.3 The Pilot Model

The last component of the inverse simulation to be developed was the pilot model. A multitude of pilot models exist and have merits for various applications. This is discussed in detail in Chapter 2 and so will not be covered here. For this research, Hess' structural pilot model (shown in Figure 7-12) [51] was used. The selection of the structural pilot model parameters represented in Figure 7-12 is dependent on the order, K, of the primary control loop controlled element, $Y_c(s)$, dynamics, as shown in Table 7-3. Care should be taken here, and throughout the thesis, not to confuse the tau guidance theory nomenclature with that of the pilot modelling aspect. Uppercase 'K' has been to denote pilot model parameters and lowercase 'k' to denote tau parameters.

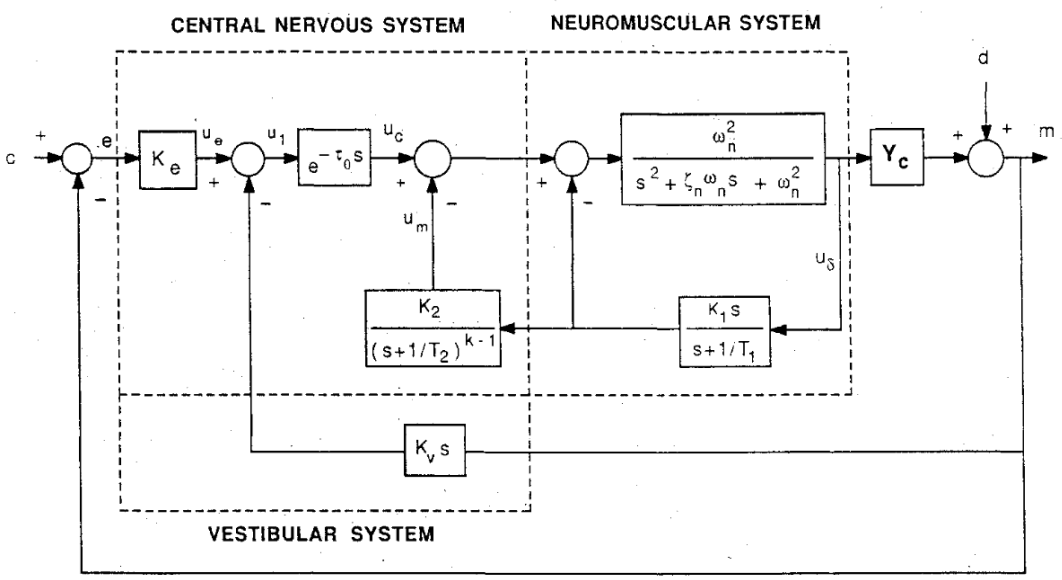


Figure 7-12 - Hess' Structural Pilot Model [51]

Table 7-3 - Structural Pilot Model Parameter Selection [51]

K	K_e	K1	K2	T1 [s]	T_2 [s]	τ0 [s]	ζ_n	ω_n [rad/s]
0	1.0	1.0	2.0	5.0	a	0.15	0.707	10.0
1	1.0	1.0	2.0	5.0	b	0.15	0.707	10.0
2	1.0	1.0	10.0	2.5	a	0.15	0.707	10.0

In the Accel-Decel manoeuvre, the pilot controls speed through selection of a pitch attitude angle. Therefore, in order to determine the correct value of $Y_c(s)$ system order, K, for the structural pilot model. The relationship between pitch attitude, θ and longitudinal cyclic

control, θ_{1s}, had to be determined. The 3 degree of freedom longitudinal equations of motion are given below.

$$\begin{bmatrix} \dot{u} \\ \dot{q} \\ \dot{\theta} \end{bmatrix} - \begin{bmatrix} X_u & X_q - U & -g\cos\Theta_e \\ M_u & M_q & -g\sin\Theta_e \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} u \\ q \\ \theta \end{bmatrix} = \begin{bmatrix} X_{\theta_{1s}} \\ M_{\theta_{1s}} \\ 0 \end{bmatrix} \theta_{1s} \quad \textbf{Equation 7-7}$$

Therefore, in the hover (U=0)

$$\dot{u} - X_u u - X_q q + g\cos\Theta_e \theta = X_{\theta_{1s}} \theta_{1s} \quad \textbf{Equation 7-8}$$

$$\dot{q} - M_u u - M_q q + g\sin\Theta_e \theta = M_{\theta_{1s}} \theta_{1s} \quad \textbf{Equation 7-9}$$

$$\dot{\theta} - q = 0 \quad \textbf{Equation 7-10}$$

Trim pitch attitude in hover in the Bell412 is approximately 5°. Therefore small angle approximation of $g\sin\Theta_e = 0$ and $g\cos\Theta_e = 1$ is assumed.

$$\dot{u} - X_u u - X_q \dot{\theta} + g\theta = X_{\theta_{1s}} \theta_{1s} \quad \textbf{Equation 7-11}$$

$$\ddot{\theta} - M_u u - M_q \dot{\theta} = M_{\theta_{1s}} \theta_{1s} \quad \textbf{Equation 7-12}$$

Taking the Laplace transform leads to;

$$(s - X_u) U(s) = X_{\theta_{1s}} \Theta_{1s}(S) + s X_q \Theta(s) - g\Theta(s) \quad \textbf{Equation 7-13}$$

$$s^2 \Theta(s) - M_u U(s) - s M_q \Theta(s) = M_{\theta_{1s}} \Theta_{1s}(S) \quad \textbf{Equation 7-14}$$

Substituting for U(s) gives

$$s^2\Theta(s) - M_u \left(\frac{X_{\theta_{1s}} \Theta_{1s}(S) + s X_q \Theta(s) - g\Theta(s)}{s - X_u} \right) - s M_q \Theta(s) = M_{\theta_{1s}} \Theta_{1s}(S)$$

Equation 7-15

Multiplying by ($s-X_u$) and multiplying out gives

$$Y_c = \frac{\Theta(S)}{\Theta_{1s}(S)} = \frac{(M_{\theta_{1s}} s - M_{\theta_{1s}} X_u + M_u X_{\theta_{1s}})}{s^3 - X_u s^2 - s M_u X_q + g M_u - (s^2 - X_u s) M_q} \times 0.9703$$

Equation 7-16

The Bode plot of this transfer function is shown in Figure 7-13. Inspection of the Bode plot shows that the system has a Gain Margin (GM) of 57dB and a Phase Margin (PM) of 130 degrees. The pilot model parameter selection methodology outlined by Hess dictates that the control sensitivity of the controlled element (nominally 0.9703) should be varied so that the open loop cross over frequency is such that GM=4dB & PM>45° or GM>4dB and PM=45° [51]. For this to be possible for the transfer function $Y_c(s)$, a control sensitivity of 530 was required. This gave an unrealistic crossover frequency of approximately 80 rad/s.

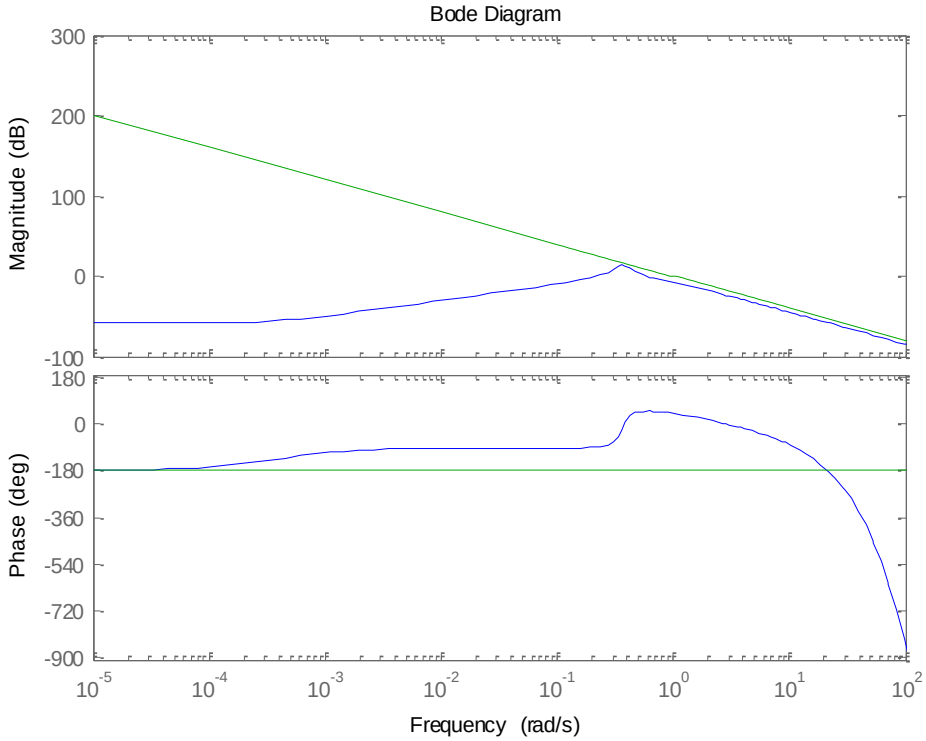


Figure 7-13 - Bode Plot of $Y_c(S)$ 3DoF Transfer Function Showing 2nd Order Approximation – GM = 57, PM=130

The lack of success in the pilot model lead to speculation concerning the controlled variable. A pitch rate commanded inverse simulation was therefore developed to determine if this provided a better representation of reality. The updated transfer function, $Y_c(s)$, is given in Equation 7-17.

$$Y_c(s) = \frac{\mathrm{q}(S)}{\Theta_{1s}(S)} = \frac{\mathrm{s}\theta(S)}{\Theta_{1s}(S)} = \frac{(M_{\Theta_{1s}}s^2 - M_{\Theta_{1s}}X_u\mathrm{s} + M_uX_{\Theta_{1s}}\mathrm{s})}{s^3 - X_u\mathrm{s}^2 - sM_uX_q + gM_u - (s^2 - X_u\mathrm{s})M_q} \times 0.9703$$

Equation 7-17

In this case, the phase and gain margin criteria were met by increasing the control sensitivity of 0.9073 (deg/s)/inch by a factor of 2.2.The Bode plot of this transfer function is shown in Figure 7-14 and illustrates that, because the pilot is controlling q directly, the transfer function of Q(s) from Θ1s can be approximated to 1st order around the crossover frequency, ω_{co}, of approximately 0.5 rad/s.

Therefore K=1 in the structural pilot model, which leads to the pilot model structure shown in Figure 7-15. Note that for 1st order transfer functions, the lead time constant, T_2, is not applicable.

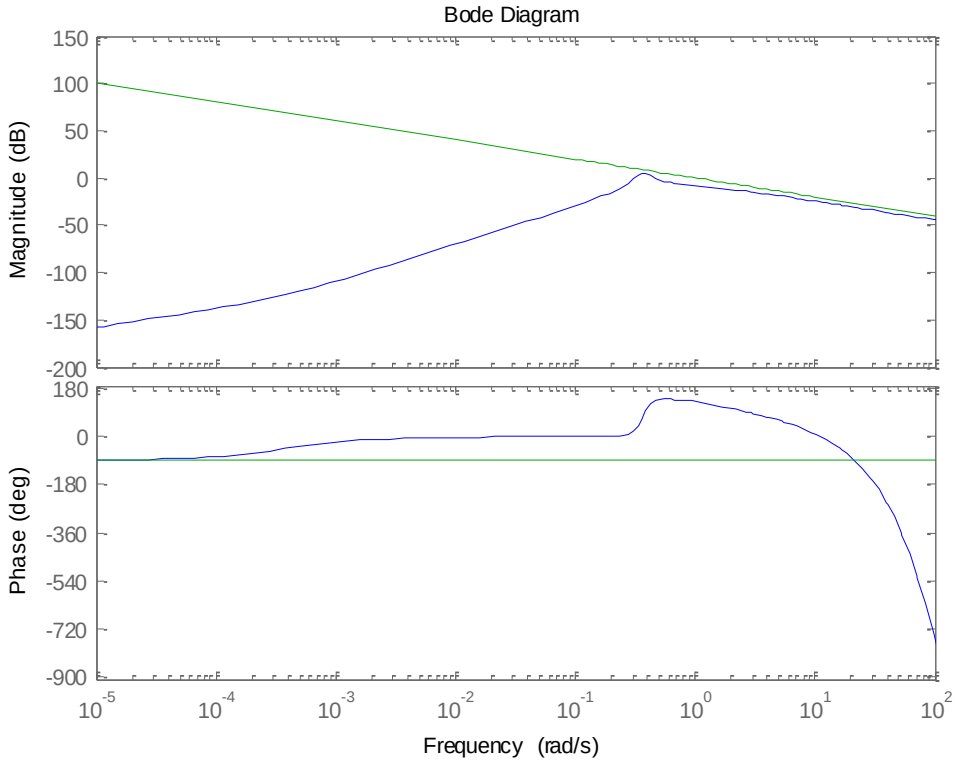


Figure 7-14- Yc 3DoF Transfer Function Bode Plot Showing 1st Order Approximation. PM = 45, GM=35

The next stage of the model parameter selection methodology, as defined by Hess, was to determine a value of motion gain, K_v, (see Figure 7-12) that ensured the lowest damping ratio of any of the oscillatory roots of the transfer function m/u_e was equal to 0.15. However, it was found that these damping values were not sensitive to K_v but were sensitive to the pilot gain, K_e. At this stage it was decided not to include motion gain in the pilot model for the sake of simplicity.

Finally, was the requirement to vary the pilot gain, K_e, so that the crossover frequency of the closed-loop model adhered to the GM and PM criteria (GM=4dB & PM>45° or GM>4dB and

PM=45°) [51]. This resulted in a Gain limited system (PM=173°, GM=3.97dB, COF=0.4 rad/s) with K_e=-70. It should be noted here that the lowest damping ratio was found to be 0.217 (not 0.15) as required by the Hess method.

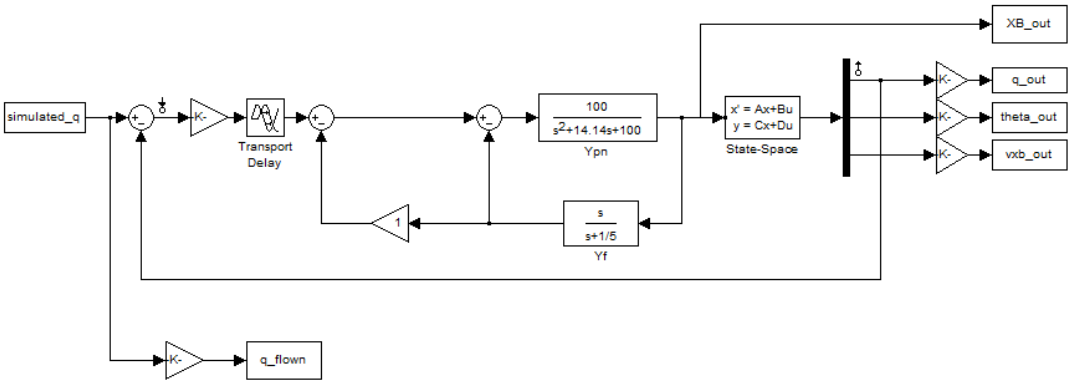


Figure 7-15 - Structural Pilot Model, K=1 for Pitch Rate Feedback.

The validity of the pilot model was assessed by comparing the longitudinal aircraft state trajectories and longitudinal control activity from piloted simulation and the inverse simulation. These comparisons are presented in Figure 7-16.

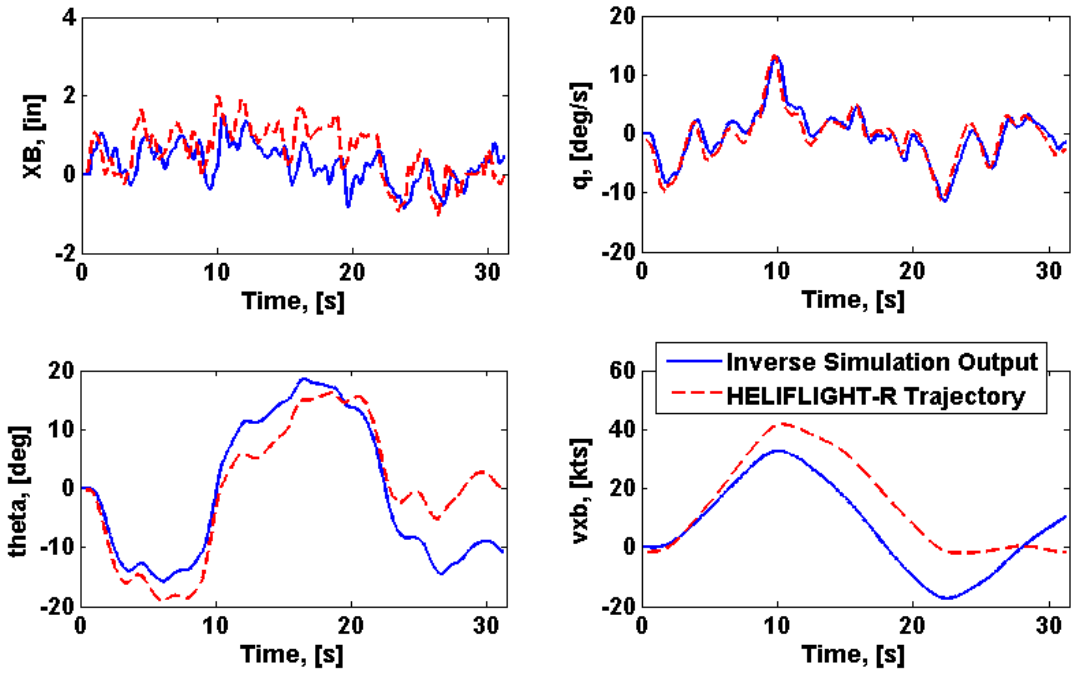


Figure 7-16 - 1st Order Pitch Rate Tracking: K_e=-70, PM=173deg, GM=3.97db, COF=0.4 rad/s, RMSe=1.4880 deg/s

Figure 7-16 shows the pilot model is able to track the pitch rate command using a realistic control strategy, indicating that the relationship between the pilot model and aircraft model was valid. Therefore the inverse simulation was deemed sufficient for the task of

determining the sensitivity of the pilot model to changes in vehicle dynamics and transport delays. There are however limitations with this model. Because a 3 DoF model has been used but only one of these variables has been controlled, the errors in the pitch rate signal magnify when solving for θ. These errors accumulate further when solving for longitudinal velocity, u. Furthermore, in reality, the relationship between the states is not constant over the range of speeds from 0kts to 40kts. The derivatives M_u and X_u vary significantly over this speed range. In the linear model the derivatives are constant and optimised for hover therefore the quality of the match degrades in the middle of the run when the aircraft is at 40 kts.

Figure 7-17 shows the match obtained against a tau pitch rate trajectory with K_e=-70. It should be noted here these are 'ideal' inputs of the pilot to fly a tau based Accel-Decel. Again, the drift of the pitch attitude and forward speed can be seen due to the lack of control of these variables. Further work into the development of a more comprehensive model with multi-loop control would ensure more realistic following of these trajectories. However, the development of such a model was deemed beyond the necessary scope for this preliminary investigation.

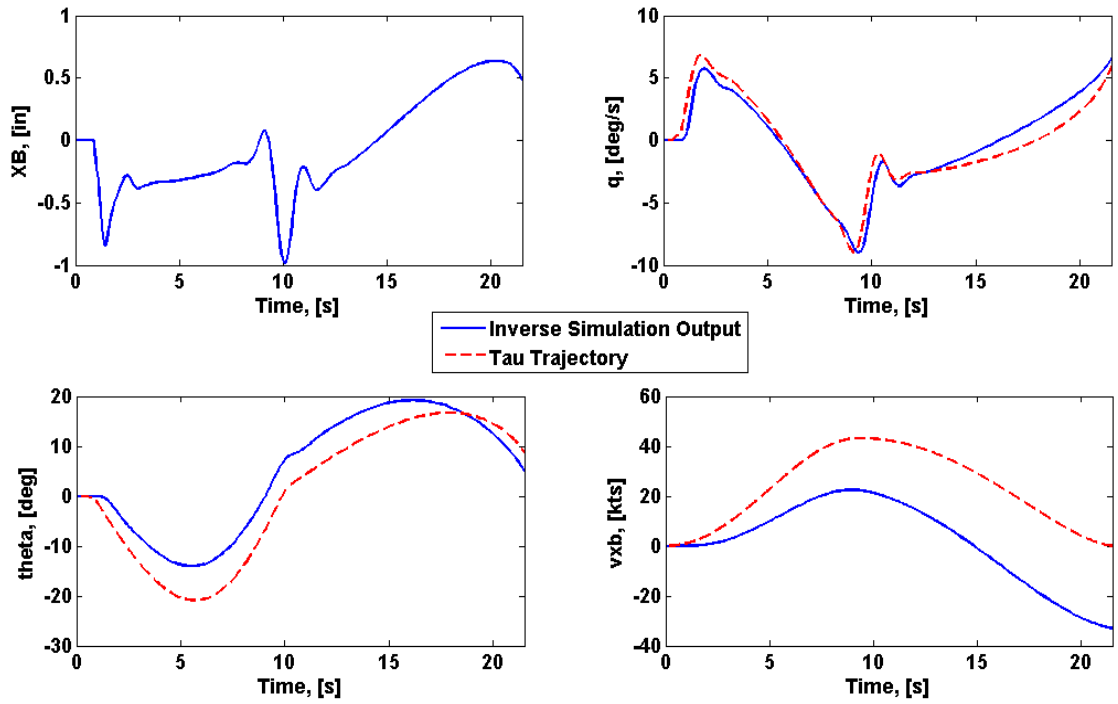


Figure 7-17 - Match between Commanded Tau Pitch Rate Trajectory and Inverse Simulation Outputs for K_e=-70.

7.3 Pilot model optimisation code

A MATLAB code was written to optimise the value of pilot gain, K_e, for each test point between the values of K_e=-1 to -100 in increments of 1. The full MATLAB code can be found in Appendix D.4. The basic functionality of this code is detailed below.

The RMS error between the commanded tau trajectory and output trajectory from the inverse simulation was calculated for each K_e value and the crossover frequency. The minimum RMS error value was then used to determine the optimum value of pilot gain, K_e. To ensure that the phase margin and gain margin criteria were met, thee PM and GM were also calculated for each value of K_e from the transfer function of the pilot-vehicle transfer function, $Y_p(s)Y_c(s)$. If the phase and gain margin criteria were not met, the RMS error for that K_e value was over written with an infinite value to ensure it was not selected as the optimum value.

The next section presents the results from the study into the effect of baseline handling qualities on pilot sensitivity to transport delays.

7.4 Results and Discussion

Figure 7-18 shows the absolute optimised values of pilot gain K_e for all test points, grouped by baseline HQ level (M_q value). The corresponding crossover frequencies and phase margins are shown in Figure 7-20, and Figure 7-22 respectively. These metrics are also presented in terms of % change from the baseline models in Figure 7-19, Figure 7-21, and Figure 7-23. The trends seen in these results as functions of M_q and ATD are discussed below the figures.

7.4.1 Effect of reduction of M_q and increasing transport delay.

Figure 7-20 shows the crossover frequency, ω_{co}, of the closed-loop system, $Y_p(s)Y_c(s)$, as a function of variations in the additional transport delay (ATD) and baseline HQs (variation in M_q). It is clear that the closed-loop crossover frequency increases with reduction in Mq (Increasing Level). This is because the pilot is having to work at higher frequencies to suppress pitch instability. The increased tendency for instability of the pilot-vehicle system is confirmed the reduced phase margin for degraded HQ vehicles (Figure 7-22). In order to maintain stability (gain margin =4 dB) the pilot gain has to reduce in as the handling qualities degraded.

The effect of increase in transport delay had only a slight tendency to reduce the pilot-vehicle crossover frequency (Figure 7-21). The reduction in phase margin (Figure 7-23) means that the pilot-vehicle system was more susceptible to instabilities with higher transport delays. Again, the pilot model maintained stability by reducing the pilot gain (Figure 7-19).

It was also possible to look at how the pilot control activity itself changed with increasing transport delay. The attack generated by the pilot model would be much less than an actual pilot because the pilot model represents the 'ideal' task strategy. However, if the pilot model was able to accurately predict changes in pilot control activity due to model variations, the pilot model could be used for preliminary assessment of perceptual fidelity issues. Therefore, a piloted simulation trail was run where the pilot flew all twelve linear model variants (three baselines and 9 variants) in the Accel-Decel MTE. The attack charts were then obtained from the control time histories and compared with those generated by the pilot model. The results of this are shown in Figure 7-24. The piloted simulation data (red) shows that as the transport delay is increased, the pilot increases the size of his control inputs slightly. This trend also occurs with decreasing M_q. As expected, the pilot model attack and displacements are smaller than that of the pilot himself. The results from the pilot model replicate (to a lesser extent) the trend of increasing η for decreasing M_q. However, the same trend increasing transport delay is not seen. Instead, the pilot model results show an increasing number of attack points with increasing transport delay. This inability to replicate the change in control activity is attributed to the limitation of the inverse simulation to single loop. The pilot model assumed the pilot is only tracking pitch rate. This is a predominantly vestibular feedback loop. However, in reality, the pilot is closing all three loops using a high degree of visual feedback. Further work into the extension of the inverse simulation to include these loops might improve the capability of producing realistic control activity

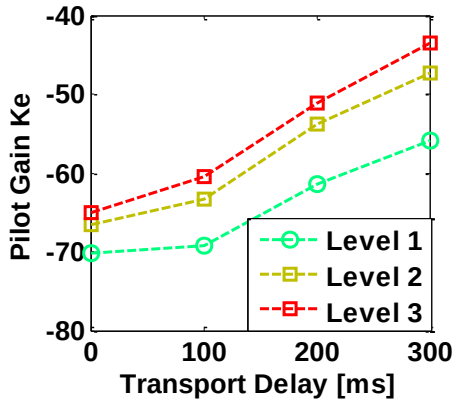


Figure 7-18 - K_e as a Function of ATD and Baseline HQ Level

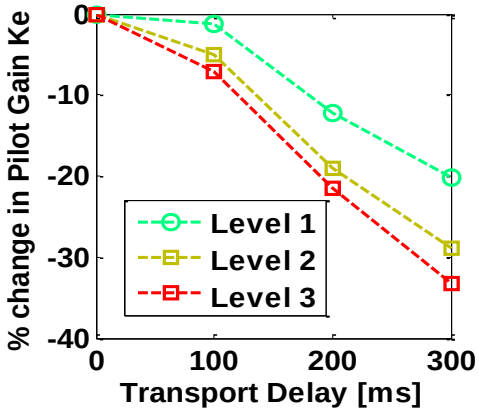


Figure 7-19 - Sensitivity of Ke to Variations in Mq and ATD

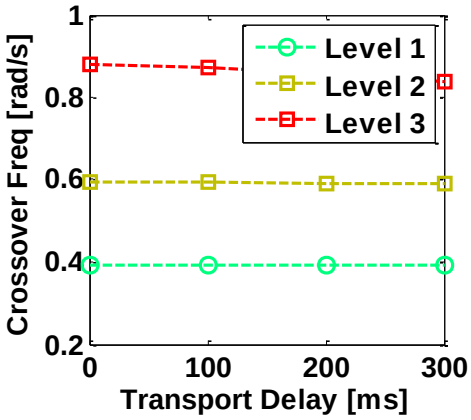


Figure 7-20 - YpYc Ω_{co} as a Function of ATD and Baseline HQ Level

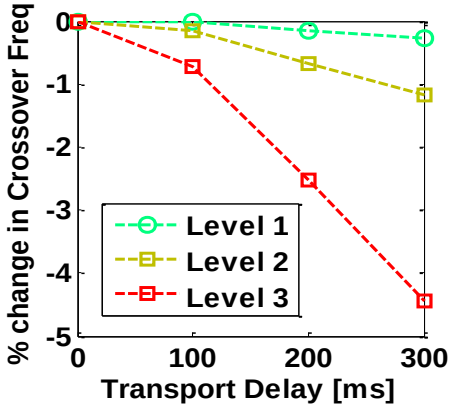


Figure 7-21 - Sensitivity of YpYc Ω_{co} to Variations In Mq and ATD

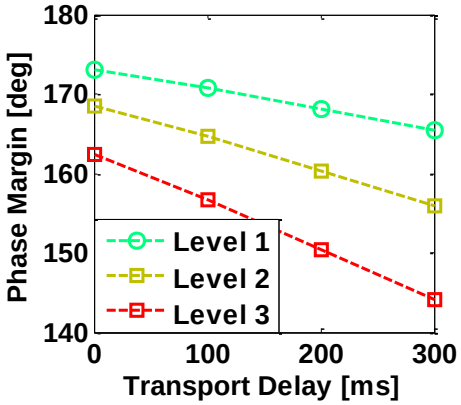


Figure 7-22 - YpYc PM as a Function of ATD and Baseline HQ Level

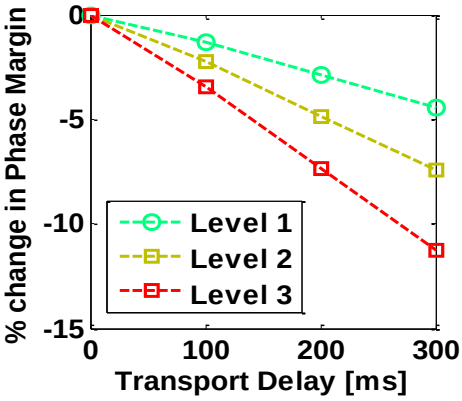


Figure 7-23 - Sensitivity of YpYc PM to Variations in Mq and ATD

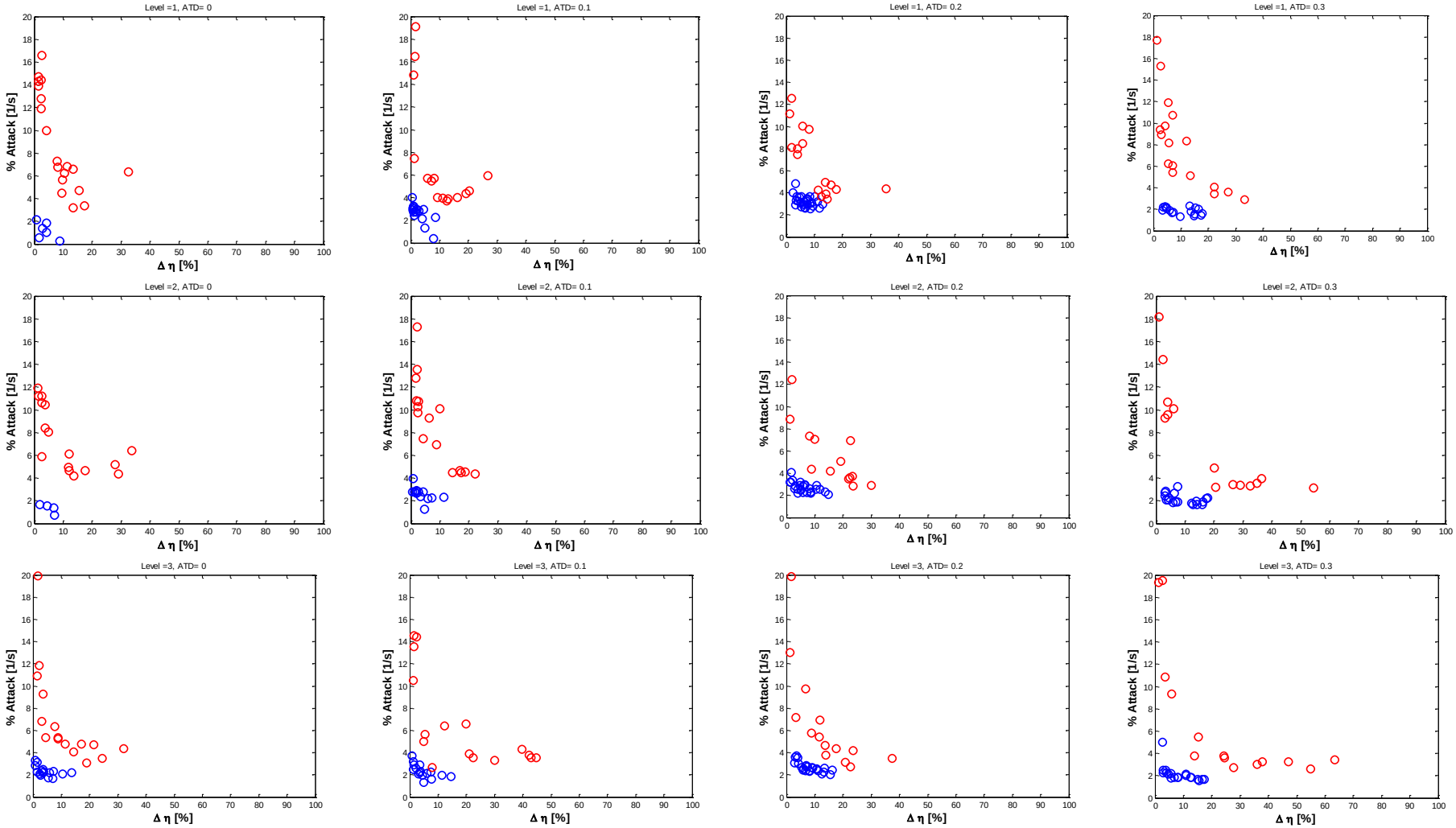


Figure 7-24- Flown (Red) Vs Predicted (Blue) Attack Charts for Increasing Time Delay (Left to Right) and Decreasing M_q (Top to Bottom)

7.4.2 The effect of baseline handling qualities on sensitivity to transport delay.

The difference in sensitivity of the pilot model to ATD for the different models is shown in Figure 7-21, Figure 7-23 and Figure 7-19. These figures show that, in order to maintain stability of the pilot-vehicle system and minimise the error between the pitch rate trajectory, the pilot model gain, K_e, had to decrease more for a degraded handling qualities environment. This reduction in gain lead to a small reduction in pilot-vehicle crossover frequency which indicates that the pilot (model) was using a more open-loop strategy in light of degraded pitch stability. The phase margin of the Pilot-Vehicle System (PVS) also decreased more significantly with increase in ATD for the degraded vehicles than for the Level 1 vehicle.

It has been shown that the PVS with the highest baseline K_e value showed the highest sensitivity. In this study, the K_e value was altered to compensate for modifications to the vehicle. From this result it can also be inferred that a higher gain pilot with a given vehicle would also be more sensitive to changes in transport delay. This is in agreement with the findings of chapter 5 and further highlights the dependency of perceptual fidelity on pilot strategy. Further research into pilot calibration may mitigate against spread in perceptual fidelity results in the future. One proposed method for pilot calibration is to determine their natural attack curve from a frequency sweeps in each axis. The continuation of this research was beyond the scope of the work, however, a small test was conducted in the longitudinal axis using one test pilot and a member of staff at UoL to highlight the method for future research. The results of this are shown in Figure 7-25 and it can be seen that the attack curve of the non-test pilot is significantly lower than that of the test pilot. Therefore, using this method, the non-test pilot would be considered to be a lower aggression pilot.

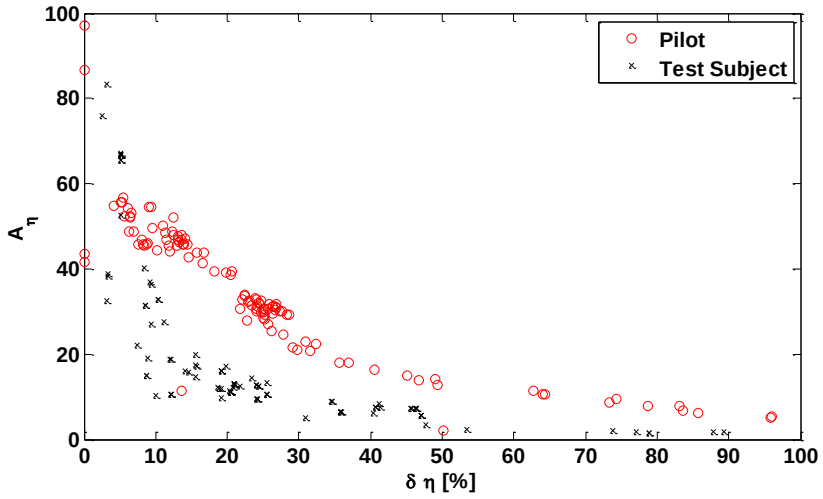


Figure 7-25 -Using Attack Charts from Frequency Sweeps for Pilot Calibration

7.5 Conclusions

This chapter has presented the development of a predictive perceptual fidelity assessment tool. This methodology has then been utilised to determine whether pilot (models) are more sensitive to flight model errors when flying a vehicle with degraded HQs. The main findings from this chapter are summarised below.

1. A mathematical representation of the Acceleration-Deceleration manoeuvre has been developed using tau guidance theory. The approximation is improved by considering that the pilot is closing velocity gaps rather than range gaps and that the acceleration profile is not necessarily symmetrical to the deceleration profile.

2. A predictive method for pilot-in-the-loop sensitivity studies has been presented. Using this method, it has been demonstrated that the parameters of the pilot model were more sensitive to the addition of aircraft transport delay in degraded handling qualities vehicles.

3. The results from this study have highlighted that, as hypothesised, Pilot model parameters were found to be more sensitive to changes in flight model for poor HQ baselines. This highlights that perceptual fidelity is dependent on nominal vehicle handling qualities. Furthermore, the poorer HQ models lead to a PVS with a higher pilot gain. Therefore it can be inferred that a higher pilot gain correlates with increased sensitivity, as discussed in chapter 5.

4. The models used in this study were validated against piloted simulation data and the correlations were deemed sufficient for use in this study. However, development of the inverse simulation with multi-loop control and nonlinear components would improve the utility of the simulation, particularly in the ability of the model to predict changes in pilot control activity in light of flight model changes.

8 OVERALL CONCLUSIONS AND RECOMMENDATIONS OF RESEARCH

This concluding chapter draws together the overarching conclusions of the work presented in this thesis. The research has been focused on the NRCs Bell 412 ASRA aircraft. Several piloted simulation trials were conducted at UoL in the HELIFLIGHT-R simulation fidelity to gather the data required for the research in this thesis. A flight test was also conducted in the ASRA to collect flight test data against which to baseline the fidelity of the FLIGHTLAB model used in this research.

Recommendations for future work motivated by these conclusions are stated to aid the continued development of the methodologies that have been detailed in this work as well as other areas of interest within the field of simulation fidelity.

8.1 Conclusions

1) The author has contributed to the development of a new subjective fidelity rating scale; the SFR scale. It was developed to provide a formalised subjective fidelity assessment methodology. A questionnaire was developed to capture pilot narrative to support the quantitative ratings. Flight and Simulation trials conducted in this research have shown how the subjective ratings can be used to complement quantitative analyses or provide an assessment alternative where little or no data is available.

2) The SFR scale was utilised in the main experimental phase of this research to determine the effect of inter-axis coupling and transport delay errors on perceptual fidelity. The results obtained showed that SFRs were sensitive to such changes (significantly more so than HQRs), thereby showing potential utility for the SFR scale to be used to define quantitative fidelity tolerances. The limited results obtained showed the effect of transport delay on perceptual fidelity to be greater than that predicted by the EASA CS-FSTD (H) criteria. However, more SFRs would need to be collected to determine whether these criteria should be altered.

3) In light of the lack of off-axis response PoM criteria, the sensitivity of pilots to changes in ADS-33E-PRF inter-axis coupling metric was investigated in Chapter 5. Data analysis of the results showed that pilot subjective opinion of perceptual fidelity is significantly sensitive changes in this parameter. Therefore it has been proven that the ADS-33E-PRF inter-axis coupling metric has strong potential to act as a quantitative metric for predictive fidelity where there is currently no off-axis

requirement. However, due to limited data sets, acceptance tolerances for this metric have not been determined.

4) A case study was conducted using the SFR scale to compare the utility of 3 variants of the Bell 412 bare airframe simulation in 3 different tasks. The results showed that, not only were the SFRs of each variant dependent on the task flown, so was the relative ranking of the three variants with one another. Pilot ratings of perceptual fidelity were also found to be task specific in Chapter 5: Pilots were more susceptible to changes in off-axis response in the Accel-Decel manoeuvre than in the Precision Hover manoeuvre. These findings prove that here is a true need for simulation qualification criteria that are based on the intended use of the equipment.

5) Although Initial data analysis from the trials detailed in Chapter 5 revealed some statistical significance between the changes in quantitative attack metrics and the SFRs for the Accel-Decel manoeuvre, there was no significant correlation seen for the Precision Hover MTE. The reason for this was the highly multi-axes nature of the Precision Hover MTE cannot be captured using metrics that only represent pilot activity in a single axis.

6) Chapter 6 detailed the work that was done to develop multi-axis, multi-dimension metrics of pilot workload and adaptation to tackle this issue. These metrics were weighted using pilot opinion of the adaptation in each axis independently. They showed significantly higher correlations with the SFRs awarded than single axis metrics, enough to warrant consideration as perceptual fidelity measures. However, it was also noted that when using this weighting method the assessment process becomes complex and subjective errors can increase in significance.

7) A large amount of spread was seen in the SFRs awarded for some test points in the trials described in Chapter 5. This was attributed partly to differing interpretation of the terminology within the SFR scale with regards to task strategy adaptation. Guidance material was developed (Chapter 4) to aid the interpretation of this terminology and help other users in pilot briefing and task design for mitigation of spread in subjective ratings.

8) It was also found through data analysis that the spread in SFRs was also due to pilot selection of task strategy. In Chapter 5 it was reasoned that a more aggressive pilot excites the dynamics of the aircraft to a greater extent, thereby exposing more

fidelity issues. In Chapter 7, it was shown using a pilot model that perceptual fidelity is dependent on the pilot vehicle system. A more unstable baseline vehicle required higher pilot gain and caused the pilot model to be more sensitive to changes in transport delay. From this result it can be inferred that a pilot using a higher gain in a particular vehicle will also be more sensitive to transport delay changes. These results provide essential insight into the importance of pilot briefing and task design to ensure reliable flight simulation fidelity assessment.

8.2 Recommendations for Further Work into Simulation Fidelity

As an outcome of the work carried out, there are a number areas of interest for further research into flight simulation fidelity. Stated below are a several recommendations for future research in this research area.

1) The data obtained in this research with regards to sensitivity to off-axis response has highlighted the potential of the ADS-33E-PRF inter-axis coupling metric. However, substantial data is required to develop a case for the addition of such a metric to current qualification standards. As a result, GARTEUR AG-21 aim to utilise several facilities available to the group to repeat the tests with more incremental test points to generate the required data to determine accurate and well justified off-axis response acceptance criteria.

2) The majority of the pilot subjective ratings gathered in this work have been from ADS-33E-PRF MTEs. It was noted in Chapter 2 that such MTEs may not have the correct aggression Level for fidelity assessment and guidance for the design of such tasks was given in section 4.6 with an example. It is envisaged that similar tasks could be developed for each training task listed in ICAO 9625. However, this was beyond the scope of this research. The assignment of designing a template and guidance material for fidelity assessment task design has been taken on by GARTEUR AG-21 from recommendation of this research and the continued development of such tasks is encouraged by the wider community.

3) The inverse simulation that was created for research in this thesis should be further developed including extension to multi-loop control and inclusion of nonlinear effects. This would enhance the capabilities of such a model for use in the prediction of realistic pilot control activity. In turn this would allow simulator control attack

metrics to be determined using the inverse simulation for preliminary perceptual fidelity assessment.

4) Common concerns raised by pilots in flight simulators are visual cue quality in simulators, pertaining to content and texture particularly as well as discontinuity between motion and visual cues. A study whereby motion and visual parameters are modified in a controlled manner and the SFR scale utilised with several pilots as it was in Chapter 5 may be used to determine quantitative requirements for visual and motion system software that are currently not available.

5) The trials designed in this study were carefully designed to mitigate against the effects of pilot learning on the results; test matrices were randomised and pilots were allowed at least four runs in the baseline simulation so that a) he was familiar enough with the vehicle to use it as a reference and b) so that effects of modifying the simulation were not masked by the effects of the pilot still learning to fly the baseline. However, pilot learning curves themselves may provide information about perceptual fidelity. After SFRs were awarded in the modified simulations in Chapter 5, the pilots were allowed more runs until they felt their strategy had converged. In some cases, the pilots were able to successfully adapt their strategy very quickly. In other cases this was not so. New studies, or analysis of the data already available, may show that pilot strategy learning curves can be used as an indicator of perceptual fidelity.

6) The University of Liverpool has a number of fixed and motion research simulators. A study is envisaged that could provide useful empirical evidence in support the development of requirements for flight training devices for skills acquisition training. Student pilots could be split into groups, one group trained on the aircraft, one on a low cost flight training device and another on the high fidelity full flight simulator. All pilots would then complete the training task in the aircraft after receiving the same training course in one of the 3 trainers. Comparative measurements of control activity and pilot behaviour could then be used and cross referenced with SFRs awarded by test pilots and feedback from the instructors in each facility to assess simulation fidelity.

APPENDICES

Appendix A. Evolution of the SFR Scale

Table A-1- The Simulation Fidelity Questionnaire (Not Final Version)

"Simulator Fidelity Rating" Questionnaire

Task:_____ Time:_____ Date:_____

Pilot:_____ DVD Time:_____ DVD Number:_____

Aircraft Configuration:_____ Filename:_____

Task Conditions:_____

Simulation Purpose:_____

Comparison for 'truth':_____

Aggressiveness	Simulation Much Lower	Simulation Lower	Identical	Simulation Higher	Simulation Much Higher
Comments					

Task Performance		Far worse performance achieved in simulation	Worse performance achieved in simulation	Achieved Performance Identical	Better performance achieved in simulation	Far better performance achieved in simulation
Yaw	N/A					
Roll	N/A					
Pitch	N/A					
Heave	N/A					
Lateral Pos.	N/A					
Longitudinal Pos.	N/A					
Vertical Pos.	N/A					
Speed	N/A					
Overall	N/A					
Comments						

Simulator Characteristics	No Observable Deficiencies	Minor Deficiencies	Moderate Deficiencies	Major Deficiencies	Not possible to compare
Yaw					
Roll					
Pitch					
Heave					
Comments					

Control Strategy	Strategy Identical	Minor strategy changes required	Moderate strategy changes required	Major strategy changes required	Completely dissimilar strategy required
Lateral Cyclic					
Longitudinal Cyclic					
Collective					
Pedals					
Comments					

Workload	Workload Very Significantly Lower	Workload Significantly Lower	Workload Insignificantly Lower	Workload Identical	Workload Insignificantly Greater	Workload Significantly Greater	Workload Very Significantly Greater
Comments							

HQR	1	2	3	4	5	6	7	8	9	10
Simulation										
Truth										
Comments										

SFR	1	2	3	4	5	6	7	8
Comments								

Influencing Factors	–	-	o	+	++
Flight Dynamics Modelling					
Visual Cueing					
Motion Cueing					
Aural Cueing					
Control Inceptors					
Vibration					
Cockpit Interface					

Additional Comments	

Table A-2 -The Simulation Fidelity Questionnaire (Final Version)

Task:_____Time:_____Date:_____

Pilot:_____ DVD Time: _____DVD Number:_____

Aircraft Configuration:_____Filename:_____

Task Conditions: _____

Simulation Purpose:_____

Comparison for 'truth': _____

Task Performance/ Aggressiveness *(only rate the states featured in task definition)*		Far worse (dissimilar)	Worse (similar)	Equivalent	Better (similar)	Far better (dissimilar)
Roll	N/A					
Pitch	N/A					
Yaw	N/A					
Lateral Pos.	N/A					
Longitudinal Pos.	N/A					
Heave/Vertical Pos.	N/A					
Speed	N/A					
Overall	N/A					
Aggressiveness						
Comments *highlight worst case and dominating phases*						

Task Strategy (Flight dynamics)	Negligible	Minimal	Moderate	Considerable	Excessive	Completely dissimilar
Lat. Cyclic						
Long. Cyclic						
Collective						
Pedals						
Comments *highlight worst case*						

Task Strategy (Cueing Environment)	Negligible	Minimal	Moderate	Considerable	Excessive	Completely dissimilar
Visual Cues						
Motion Cues						
Aural Cues						
Inceptors						
Vibration						
Cockpit						
Comments *highlight worst case*						

HQR	1	2	3	4	5	6	7	8	9	10
Baseline										
Modified										
Comments *highlight main influencing factor(s)*										

SFR	1	2	3	4	5	6	7	8	9	10
Comments *highlight main influencing factor(s)*										

Additional Comments

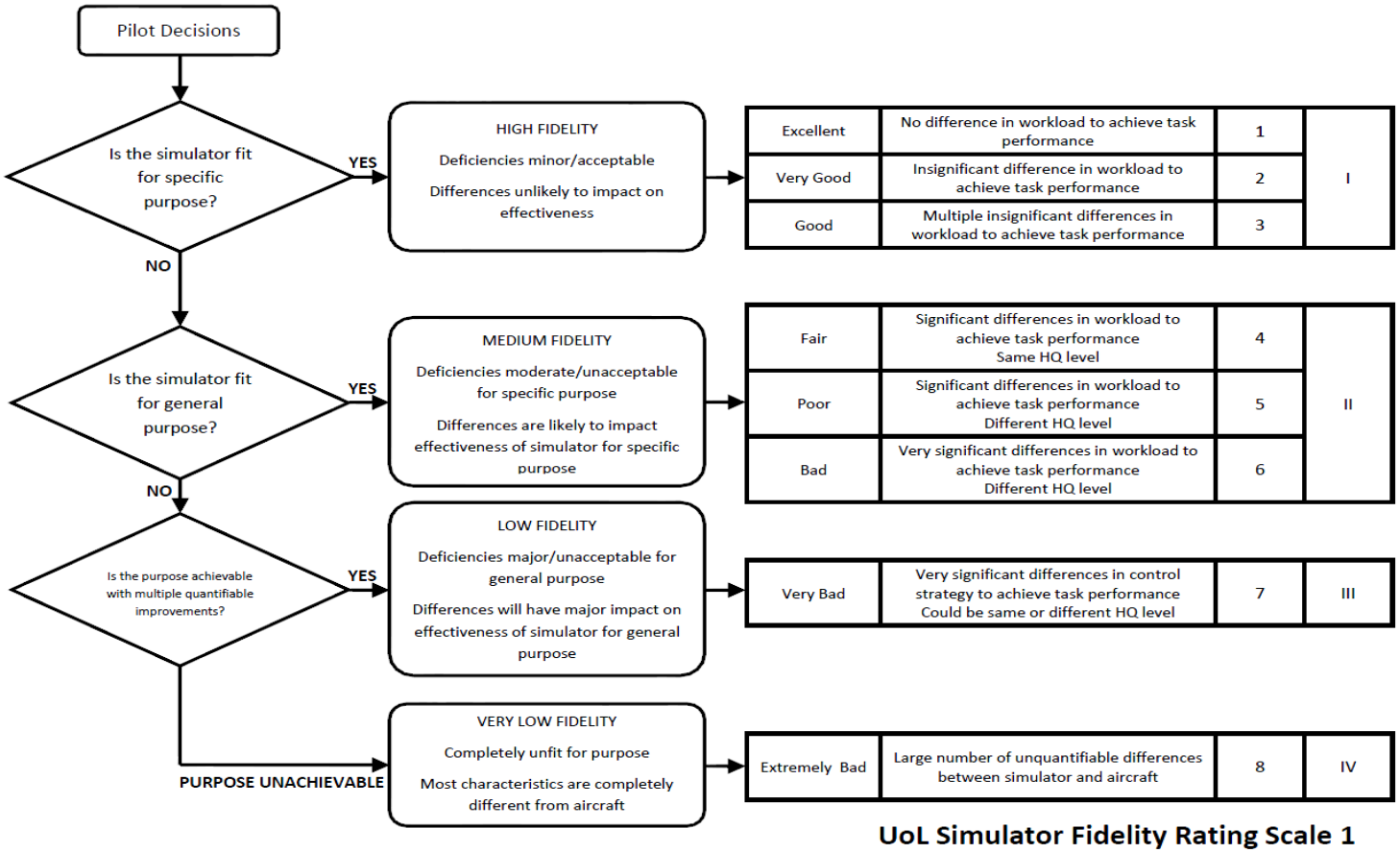


Figure A-1 - SFR Issue A - Uol Scale 1

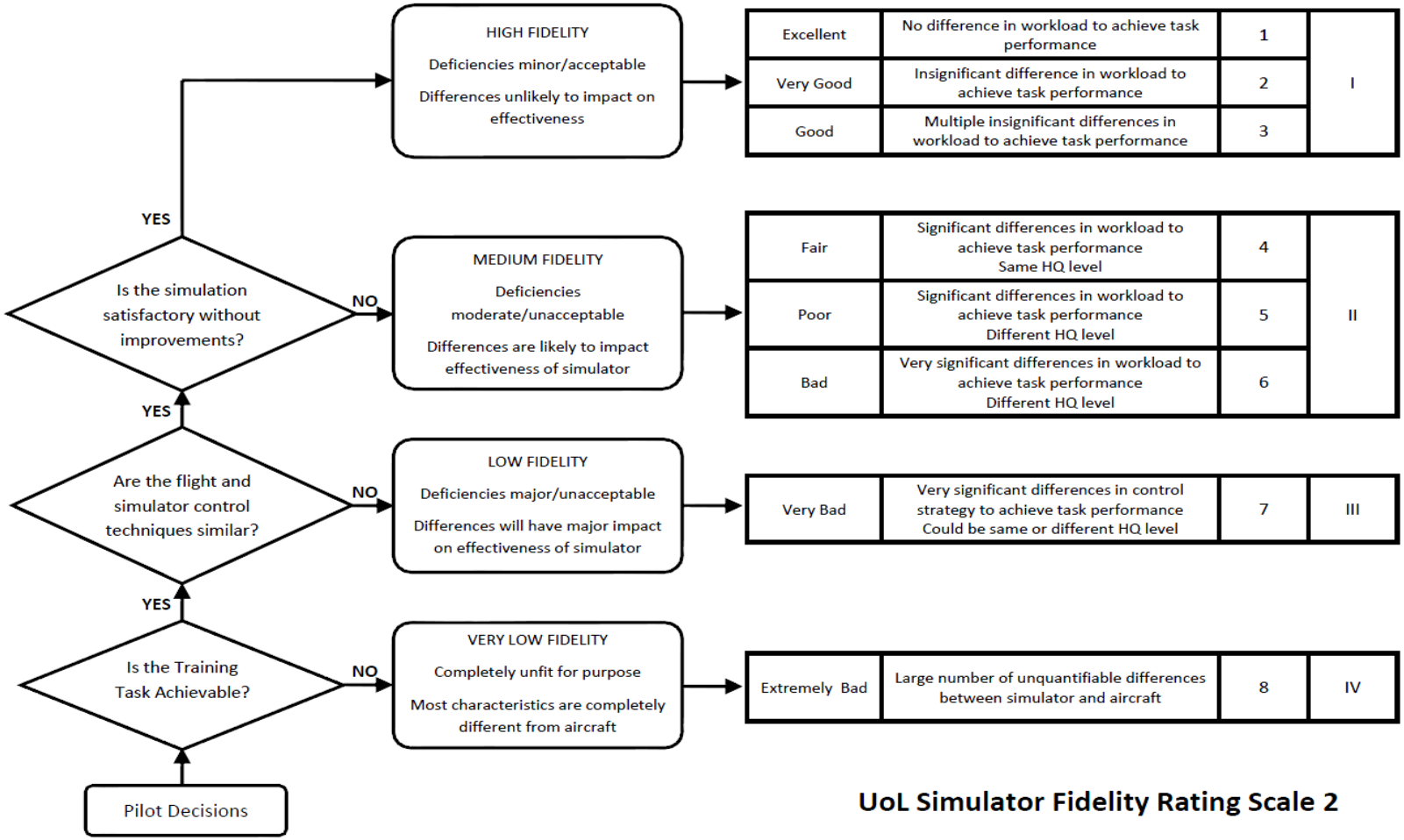


Figure A-2 - SFR Issue B - Uol Scale 2

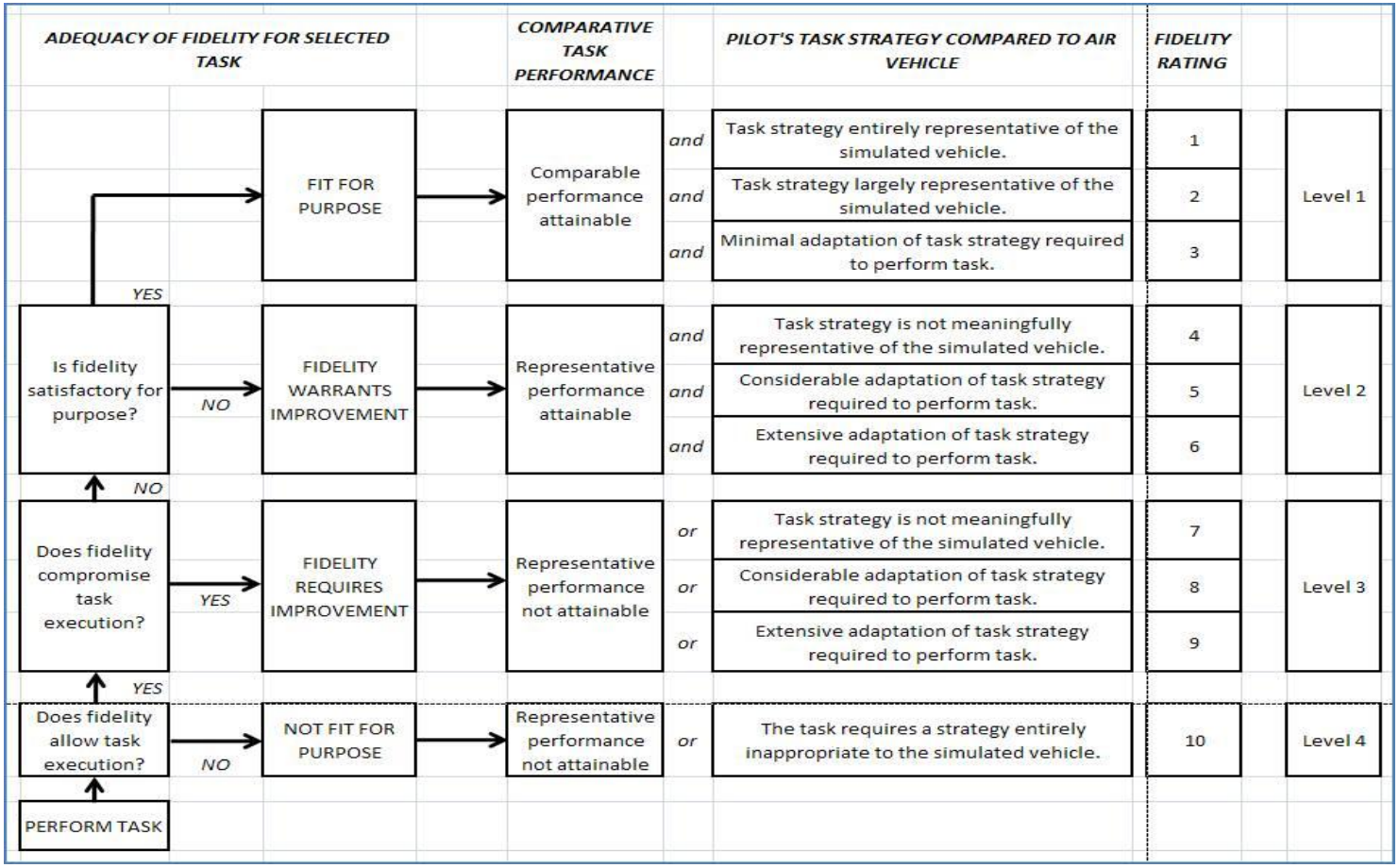


Figure A-3 - SFR Issue C - NRC/UoI Scale

Appendix B. Raw Rating Data

Table B-1 - Precision Hover Cross Coupling - Assigned HQRs

Precision Hover Cross Coupling- Assigned HQRs										
ΔC_{RP}	0.2		0.4		1.0		1.6		2.0	
Model	Base	Mod	Base	Mod	Base	Mod	Base	Mod	Base	Mod
Pilot A	2	2	x	x	x	x	x	x	2	4
Pilot B	x	x	x	x	x	x	x	x	4	6
Pilot C	x	x	2	3	x	x	x	x	3	7
Pilot D	x	x	3	2	5	5	x	x	3	6
Pilot E	x	x	3	4	3	4	3	6	x	x
Pilot F	x	x	3	4	4	6	5	7	x	x
Pilot G	x	x	x	x	2	4	x	x	x	x
Pilot H	x	x	x	x	x	x	5	4.5	x	x

Table B-2- Accel-Decel Cross Coupling - Assigned Hqrs

Accel-Decel Cross Coupling - Assigned HQRs								
ΔC_{RP}	0.4		0.8		1.0		2.0	
Model	Base	Mod	Base	Mod	Base	Mod	Base	Mod
Pilot A	x	x	x	x	x	x	3	6
Pilot B	3	4	4	4	4	7	4	8
Pilot C	5	5	x	x	4	5	x	x
Pilot D	4	3	x	x	3	6	x	x
Pilot E	4	4	5	6	5	6	x	x
Pilot F	5	6	6	6	5	6	x	x
Pilot G	x	x	x	x	5	6	x	x
Pilot H	x	x	x	x	x	x	x	x

Table B-3- Precision Hover Cross Coupling - Assigned SFRs

Precision Hover Cross Coupling - Assigned SFRs					
ΔC_{RP}	0.2	0.4	1.0	1.6	2.0
Pilot A	1	x	x	x	7
Pilot B	x	x	x	x	10
Pilot C	x	1	x	x	7
Pilot D	x	1	3	x	8
Pilot E	x	2	2	8	x
Pilot F	x	2	4	9	x
Pilot G	x	x	2	x	x
Pilot H	x	x	x	8	x

Table B-4 - Accel-Decel Cross Coupling - Assigned SFRs

Accel-Decel Cross Coupling - Assigned SFRs					
ΔC_{RP}	0.2	0.4	0.8	1.0	2.0
Pilot A	x	x	x	x	7
Pilot B	x	1	6	8	9
Pilot C	x	2	x	5	x
Pilot D	x	3	x	8	x
Pilot E	x	3	3	8	x
Pilot F	x	5	6	7	x
Pilot G	x	x	x	7	x
Pilot H	x	x	x	x	x

Table B-5 - Precision Hover Time Delay - Assigned HQRs

Precision Hover Time Delays - Assigned HQRs								
ATD [ms]	47		100		200		300	
	Baseline	Mod	Baseline	Mod	Baseline	Mod	Baseline	Mod
Pilot A	x	x	2	2	2	4	2	5
Pilot B	x	x	3	5	3	5	x	x
Pilot C	x	x	3	4	x	x	3	6
Pilot D	x	x	5	6	x	x	3	7
Pilot E	3	4	2	5	3	6	x	x
Pilot F	5	4	3	4	4	6	x	x
Pilot G	3	4	4	6	3	7	3	9
Pilot H	x	x	x	x	x	x	x	x

Table B-6 - Accel-Decel Time Delay - Assigned HQRs

Accel-Decel Time Delays - Assigned HQRs						
ATD [ms]	100		200		300	
	Baseline	Mod	Baseline	Mod	Baseline	Mod
Pilot A	4.5	5	4.5	6	4.5	7
Pilot B	3	3	x	x	x	x
Pilot C	5	4	3	4	x	x
Pilot D	4	4	4	6	4	7
Pilot E	4	6	x	x	x	x
Pilot F	4	5	6	6	x	x
Pilot G	x	x	x	x	4	8
Pilot H	x	x	7	8	x	x

Table B-7 - Precision Hover Time Delay - Assigned SFRs

Precision Hover Time Delays - Assigned SFRs (normalised to SFR 8)				
ATD [ms]	47	100	200	300
Pilot A	x	1	4	8
Pilot B	x	3	9	x
Pilot C	x	8	x	9
Pilot D	x	5	x	8
Pilot E	1	2	9	x
Pilot F	2	4	9	x
Pilot G	4	7	7	10
Pilot H	x	x	x	x

Table B-8 - Accel-Decel Time Delay - Assigned SFRs

Accel-Decel Time Delays - Assigned SFRs (normalised to SFR 8)			
ATD [ms]	100	200	300
Pilot A	5	7	9
Pilot B	1	x	x
Pilot C	3	8	x
Pilot D	2	5	8
Pilot E	7	x	x
Pilot F	3	6	x
Pilot G	x	x	8
Pilot H	x	8	x

Appendix C. - Quantitative Analysis Results

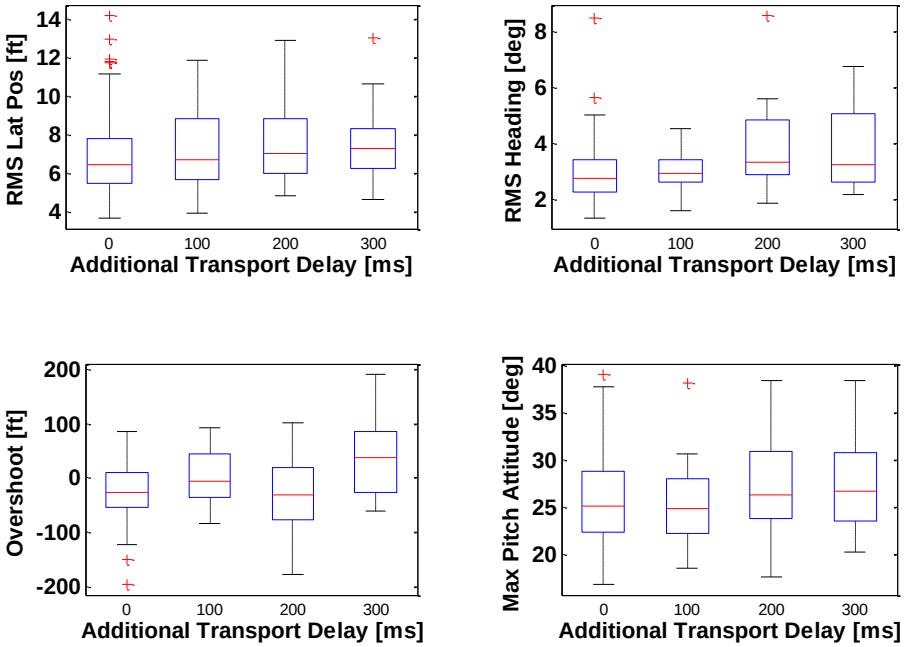


Figure C-1 - Accel Decel Performance as a Function of ATD

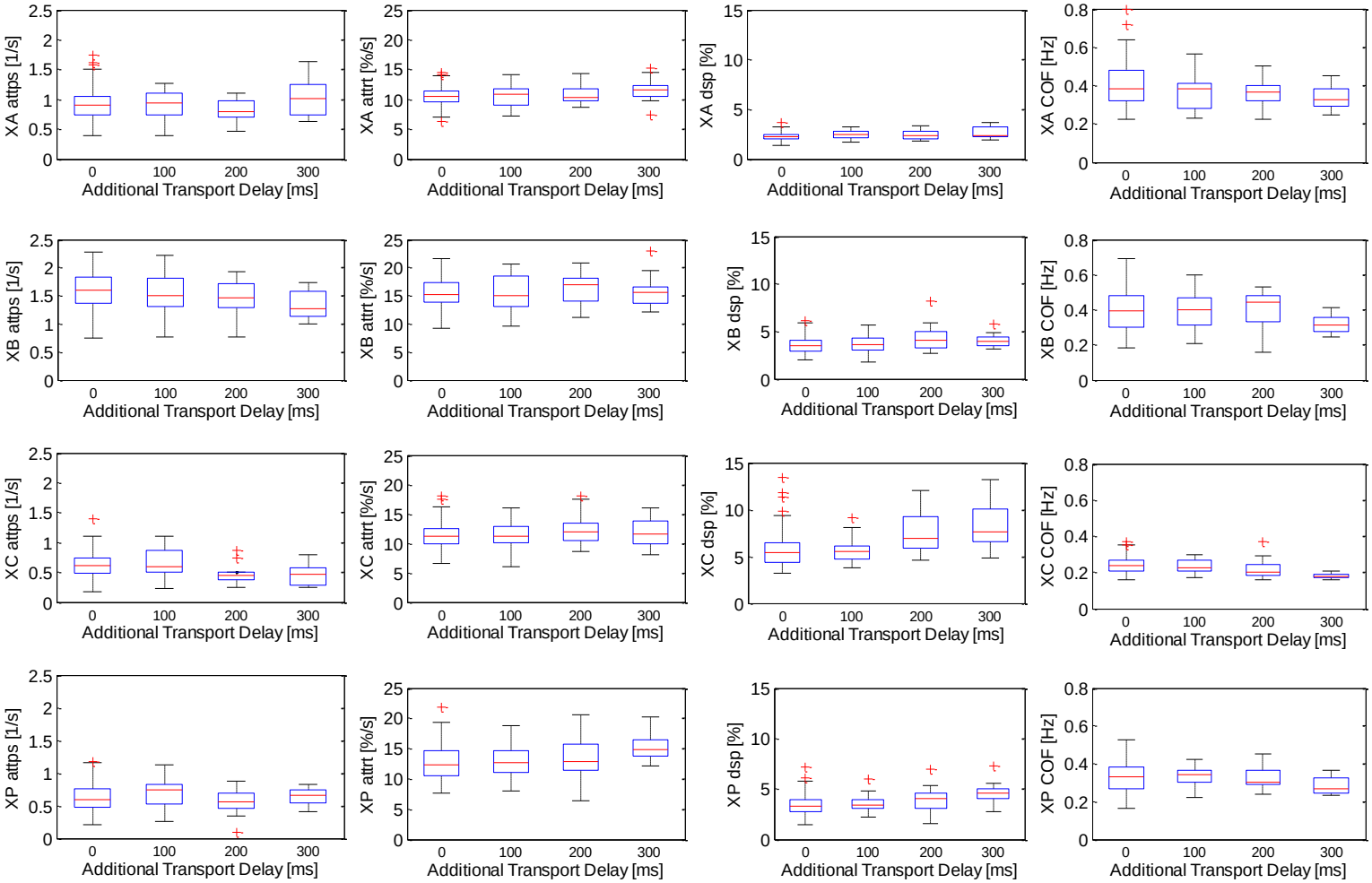


Figure C-2 – Accel-Decel Control Activity Metrics for Various Additional Transport Delay (ATD) Models

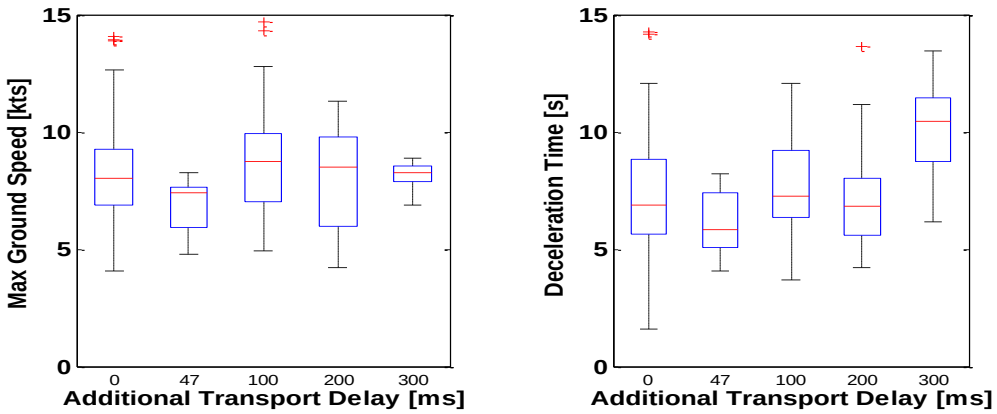


Figure C-3 - Precision Hover Transition Task Performance as a Function of ATD

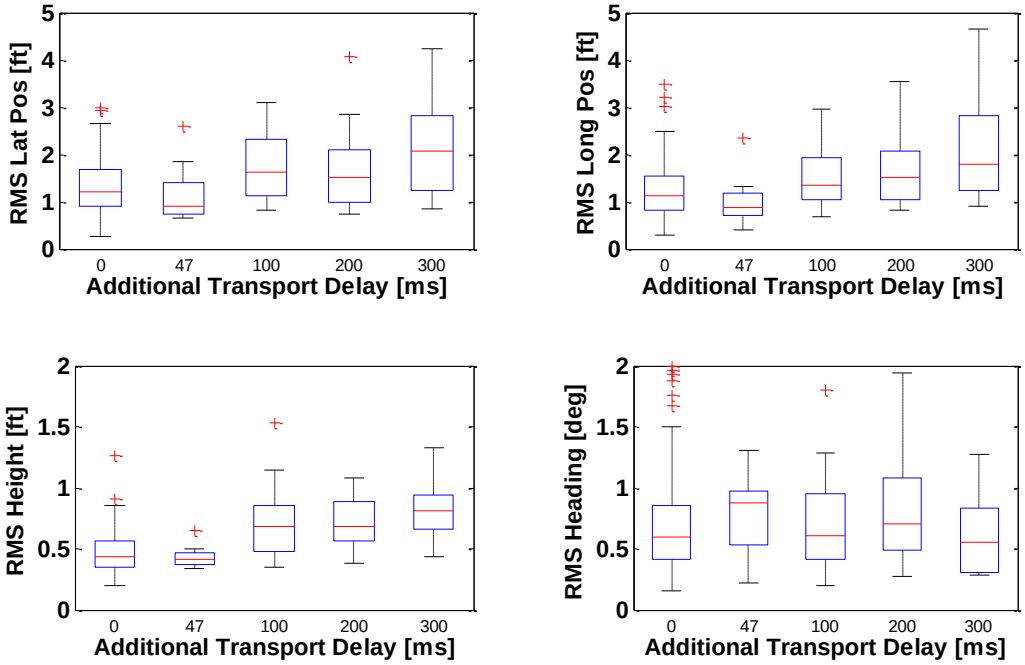


Figure C-4 - Precision Hover Maintenance Performance as a Function of ATD

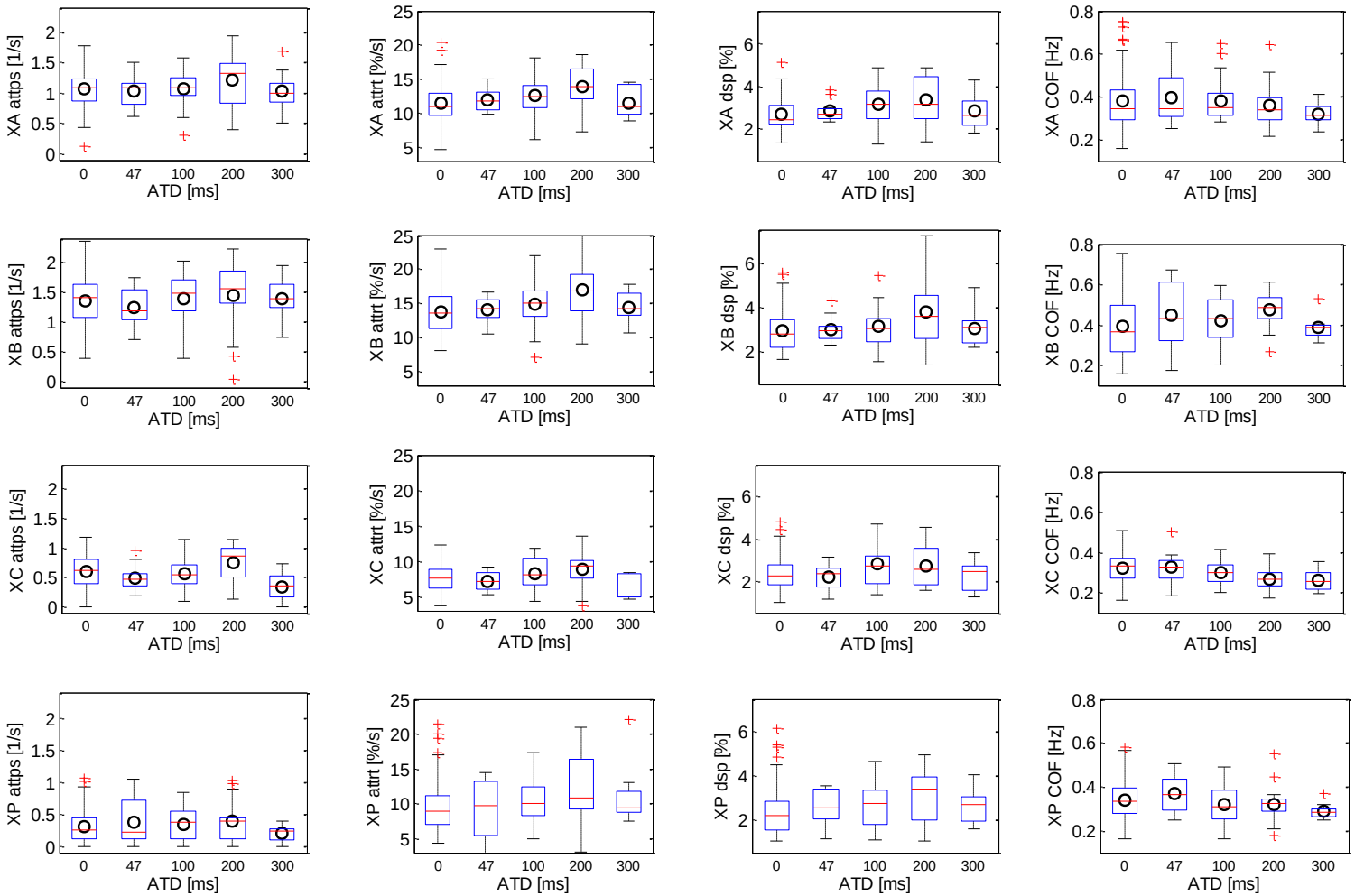


Figure C-5 - Precision Hover (Transition) Control Activity Metrics for Various Additional Transport Delay (ATD) Models

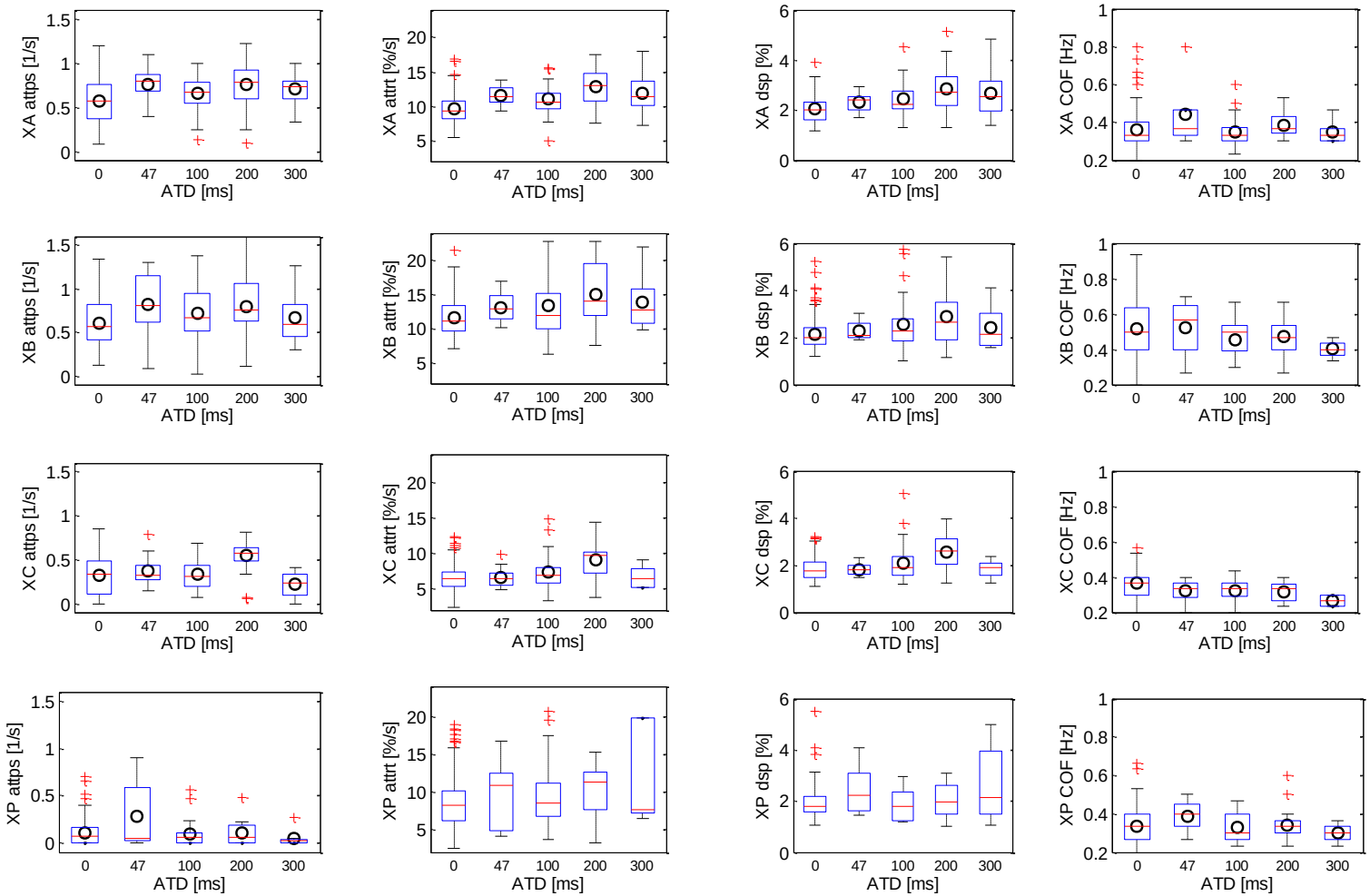


Figure C-6 - Precision Hover (Maintenance) Control Activity Metrics for Various Additional Transport Delay (ATD) Models

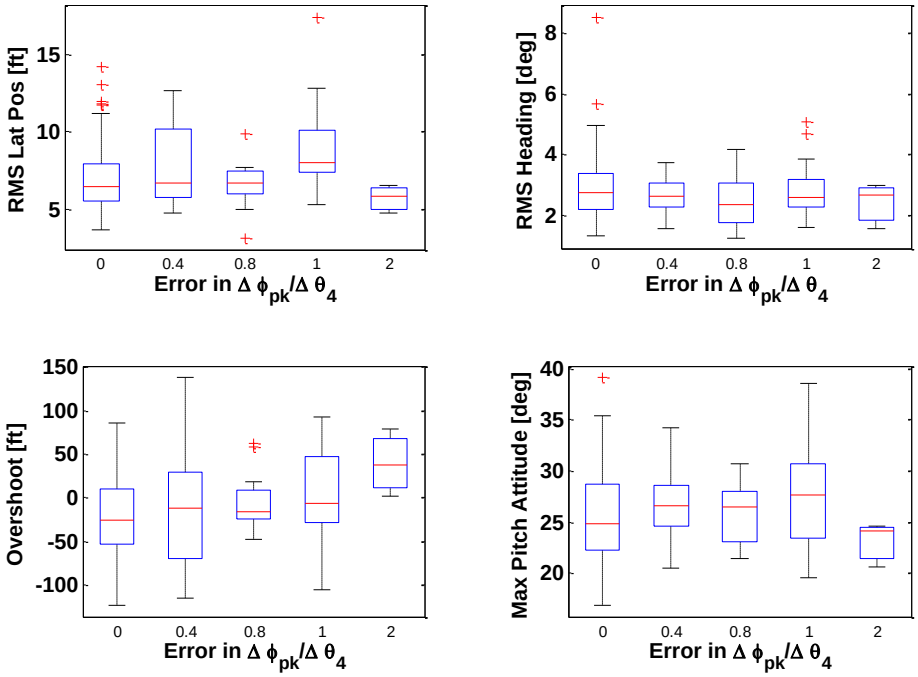


Figure C-7 - Accel-Decel Performance Aa a Function of Cross-Coupling Error

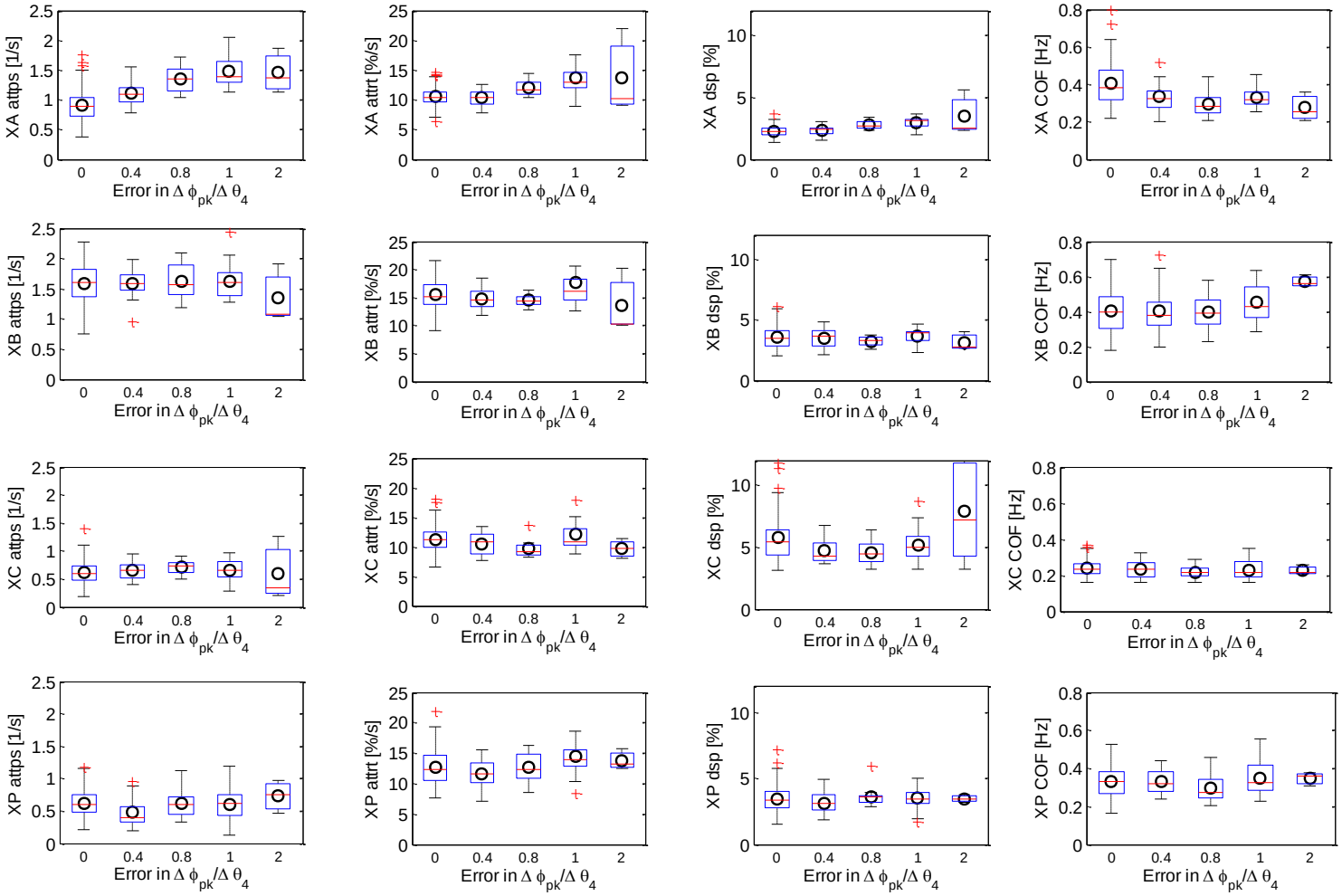


Figure C-8 – Accel-Decel Control Activity Metrics for Various Pitch/Roll Coupling Models

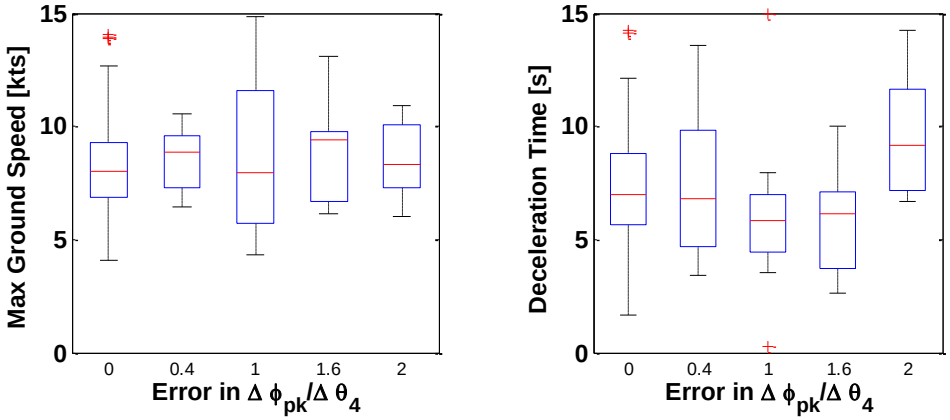


Figure C-9 - Precision Hover Transition Performance as a Function of Cross Coupling Error

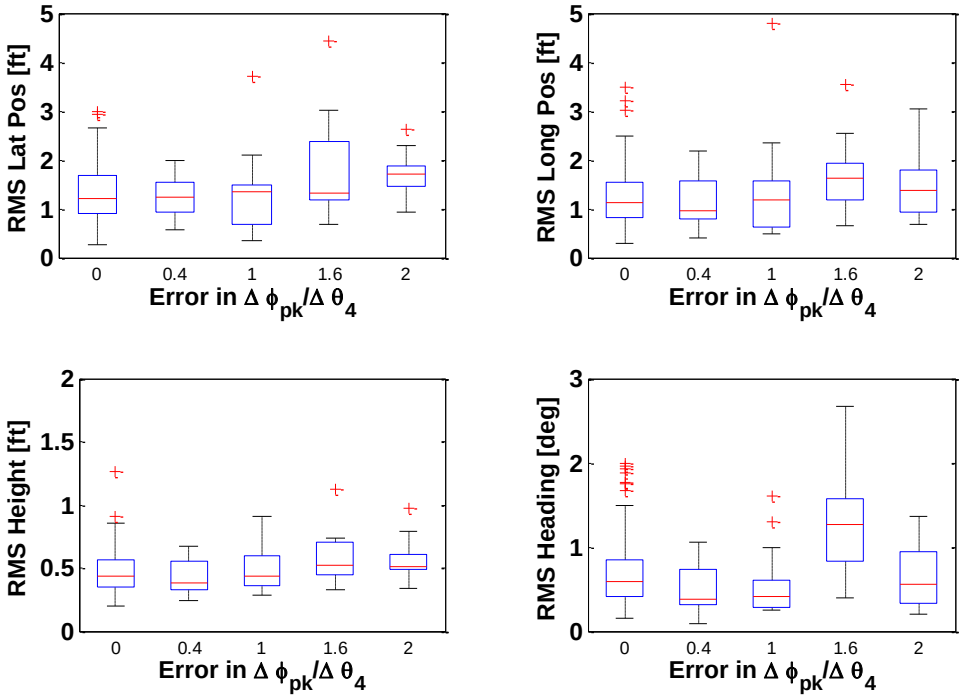


Figure C-10 - Precision Hover Maintenance Performance as a Function of Cross Coupling Error

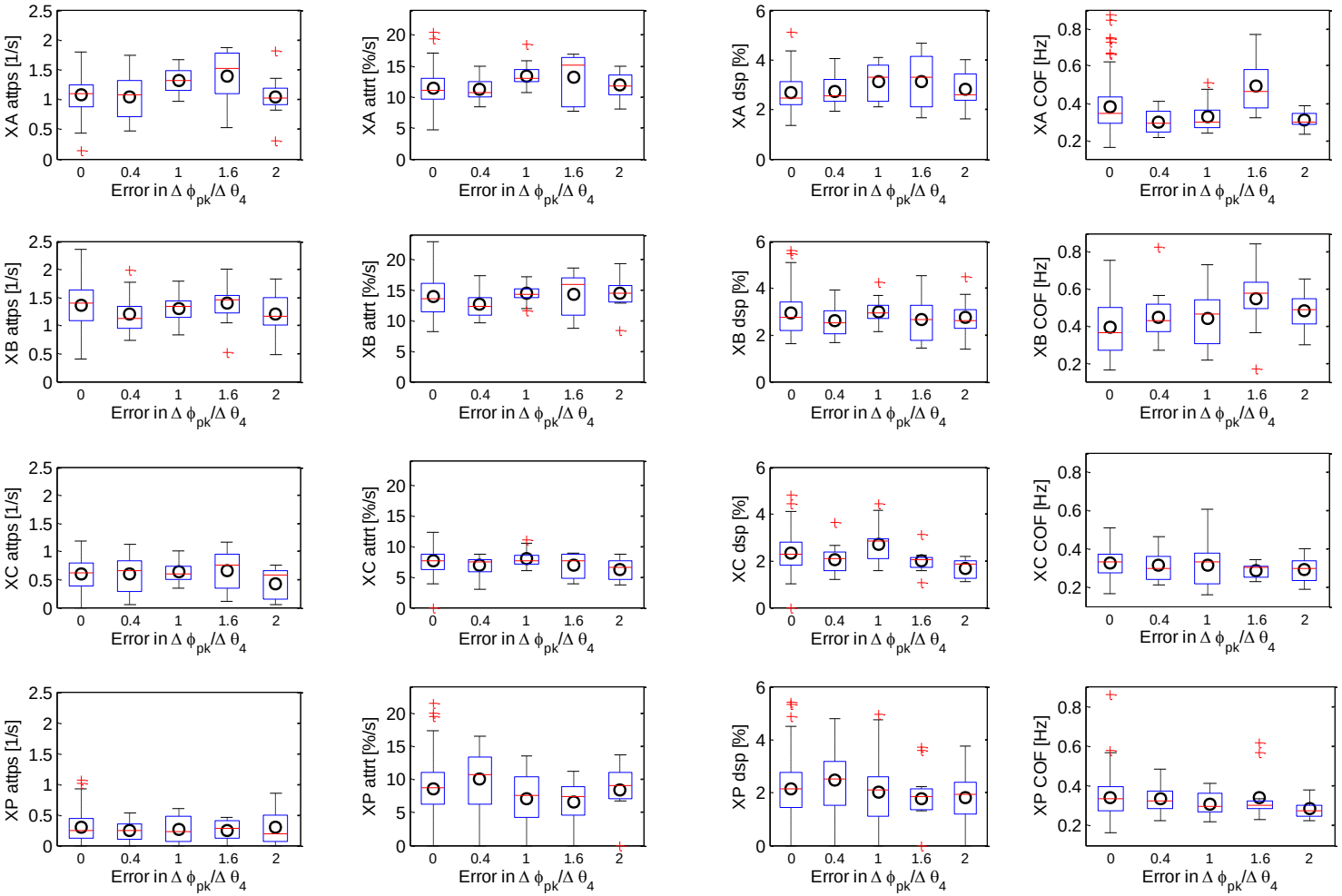


Figure C-11 - Precision Hover (Transition) Control Activity Metrics for Various Pitch/Roll Coupling Models

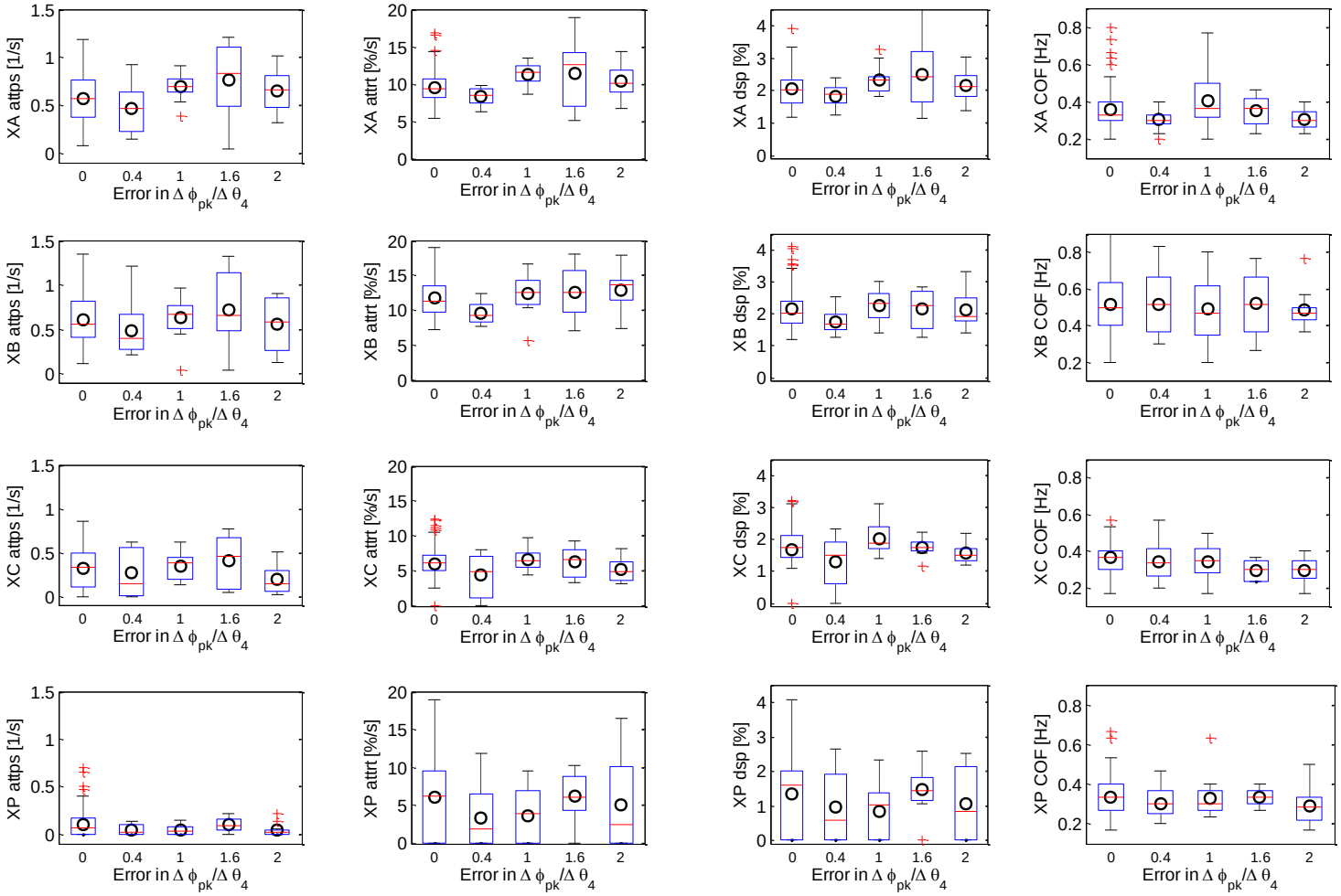


Figure C-12 - Precision Hover (Maintenance) Control Activity Metrics for Various Pitch/Roll Coupling Models

Appendix D. MATLAB Codes

Appendix D.1 - Calculating cut-off frequency

This MATLAB function reads in a time history control trace, $\eta(t)$ and computes the Fast Fourier Transform (FFT) of the signal into a frequency domain signal $\eta(Y)$. The RMS of $\eta(Y)$ is then calculated. Following this, the cut-off frequency, ω_{co}, is calculated by determining at what frequency the cumulative RMS is half of the total RMS.

```
function
[cutoff,f,autospectrum,f2,sigma21bysigma2tot,sigmasquaredtot]
=calc_cutoff_frequency_tischler(time,signal)

time=time-time(1);
dt = (time(end)-time(1))/(length(time)-1);
Fs = 1/dt;
signal = signal - mean(signal);
%find the discrete fourier transform
dft = dt*fft(signal);
f = Fs/2*linspace(0,1,((length(dft)/2))); %Hz
%manipulate
for i=1:1:length(time)
    %find the autospectrum/power spectral density distribution
    autospectrum(i) = (2/time(end)) * (abs(dft(i))^2);
end
%need to apply frequency limits to the calculations
flower = 0.159; %1rad/s
fupper = 1.59;  %10rad/s
cutoffpoints1=find(f<=fupper);
cutoffpoints2=find(f>=flower);
f2=f(cutoffpoints2(1):cutoffpoints1(end));
autospectrum2=autospectrum(cutoffpoints2(1):cutoffpoints1(end));
%calculate the mean-square of the total signal
sigmasquaredtot = sum(autospectrum2)/(2*pi); %convert to rad
sigma21bysigma2tot = 0;
for i=1:1:length(autospectrum2)
    sigma21bysigma2tot(i) = sigma21bysigma2tot(end) +
autospectrum2(i);
end
sigma21bysigma2tot = sigma21bysigma2tot / (2*pi) / sigmasquaredtot;
%iterate through the frequency spectrum to find the point where we
reach a total of 0.5 of the total mean-square signal
for i=1:1:length(autospectrum2)
    sigmasquaredsignal1 = sum(autospectrum2(1:i))/(2*pi);
    if (sigmasquaredsignal1/sigmasquaredtot) < 0.5
        %continuing search...
    else
    %cutoff frequency lies between this point and the previous one
        cutoff = f2(i); %in Hz
        break
    end
end
```

Appendix D.2 - Calculating control attack

This function reads the control time history, η , as well as the control limits. This allows the control signal to be normalised between -0.5 and 0.5. A control rate vector, $\dot{\eta}$, is then created using a for loop. 'if' statement logic is used to determine the points in the vector at which a control reversal occurs (i.e. the sign of the vector $\dot{\eta}$ changes). This creates a vector of control reversal points, etapeaks1. The length of this vector is equal to the number of attack points in the control time history. The discrete changes in control input magnitude, $\Delta\eta$ and the peak rate, $\dot{\eta}_{pk}$, between each point in the etapeaks1 vector are then determined and averaged across their length to determine $\bar{\eta}\ and\ \bar{\eta}_{pk}$.

```
function [attack,deleta,np,meanattack,numsecs]=calc_control_attack
(time,eta,limits)

dt=mean(diff(time));
%total number of seconds of data
numsecs=ceil(max(time-time(1)));
%differentiate control position and normalise
range=limits(2)-limits(1);
for i=2:1:length(time);
    etadot(i)=((eta(i)-eta(i-1))/range*100)/dt;
end
etapeaks1=0;
%find indices of peaks in control position
for i=2:1:length(etadot)
    point1=etadot(i);
    point2=etadot(i-1);
    if point1<=0 && point2>0 || point1>0 && point2<=0
        etapeaks1(length(etapeaks1)+1)=i;
    end
end
etapeaks1=etapeaks1(2:length(etapeaks1));
%change in control position between each peak
for i=2:1:length(etapeaks1)
    deleta1(i)=(eta(etapeaks1(i))-eta(etapeaks1(i-1)))/range*100;
end
deleta1=abs(deleta1);
%peak rate of change of control position between each peak
for i=2:1:length(etapeaks1)
    etadotpeak(i)=max(abs(etadot(etapeaks1(i-1):etapeaks1(i))));
end
%calculate attack
    for i=2:1:(length(deleta1))
        attack1(i)=etadotpeak(i)/deleta1(i);
    end
%remove values where deleta=0
    indices=find(deleta1>0.01);%only consider inputs larger than 1%
    deleta=deleta1(indices);
    attack=attack1(indices);
    etapeaks=etapeaks1(indices);
    np=length(attack);
    meanattack(1)=mean(deleta);
    meanattack(2)=mean(attack.*deleta)/meanattack(1);
end
```

Appendix D.3 - Tau guide match optimisation

```
function[k1,k2,total_err_fcn1,total_err_fcn2,t_max_vxb]=tau_match_fu
nction(vxb,vxb_norm,time,tp,thr,thr2,a)

%Finding beginning and end of Accel-Decel
j1=1;
for i=1:1000;
    %finding where speed crosses zero (accel)
    if vxb_norm(i)<thr &&vxb_norm(i+1)>thr
begin(j1)=i;
j1=j1+1;
    end
end

j2=1;
for i=1500:length(vxb_norm)-1;
    %finding where speed crosses zero (decel)
    if vxb_norm(i)>thr2 &&vxb_norm(i+1)<thr2
finish(j2)=i;
j2=j2+1;
    end
end
if exist('begin')==1;
begin=begin(1);
else
    begin=1;
end
clear vxb_norm

finish=finish(1);

%Cropping file to only include the accel-decel
time_crop=time(begin:finish);
vxb_crop=vxb(begin:finish);

%Normalising Data
for i=1:length(time_crop)
 time_norm(i)=(time_crop(i)-time_crop(1))./max(time_crop-
time_crop(1));
 vxb_norm(i)=(vxb_crop(i)-vxb_crop(1))./max(vxb_crop);
end

%Creating separate data sets for accel and decel phases
split=find(vxb_norm==max(vxb_norm));

time_norm1=time_norm(1:split)/time_norm(split);
vxb_norm1=vxb_norm(1:split);

time_norm2=(time_norm(split:end)-time_norm(split))/(time_norm(end)-
time_norm(split));
vxb_norm2=vxb_norm(split:end);

 k1=0.1;
 k2=0.1;
 for ii=1:18
  for i=1:length(time_norm1);
```

```
%Calculating acceleration tau guide for varying k1 values
 x_vel_accel(i)=(-(1-time_norm1(i).^2).^(1/k1))+1;
 %Determining error between tau guide and flown trajectory
 err_fcn1(i)=abs(vxb_norm1(i)-x_vel_accel(i)./max(x_vel_accel));
  end
%Calculating RMS error
 total_err_fcn1(ii)=sqrt(sum(err_fcn1.^2)/length(err_fcn1));
 k1=k1+0.05;

  for i=1:length(time_norm2);
 %Calculating acceleration tau guide for varying k1 values
 x_vel_decel(i)=((1-time_norm2(i).^2).^(1/k2));
 %Determining error between tau guide and flown trajectory
 err_fcn2(i)=abs(vxb_norm2(i)-x_vel_decel(i)./max(x_vel_decel));
  end
  %Calculating RMS error
  total_err_fcn2(ii)=sqrt(sum(err_fcn2.^2)/length(err_fcn2));
 k2=k2+0.05;
 end
 %Finding k1 and k2 values resulting in lowest RMS error
 k1_number=(min(total_err_fcn1));
 k1=(find(total_err_fcn1==k1_number))*0.05;
 k2_number=(min(total_err_fcn2));
 k2=(find(total_err_fcn2==k2_number))*0.05;

 %Defining optimum tau guide
 for i=1:length(time_norm1);
 x_vel_accel(i)=(-(1-time_norm1(i).^2).^(1/k1))+1;
 end
 for i=1:length(time_norm2)
  x_vel_decel(i)=((1-time_norm2(i).^2).^(1/k2));
 end
```

Appendix D.4 - Pilot model optimisation

This code runs the inverse simulation model shown in Figure 7-15 for varying values of pilot gain, ke for all combinations of transport delay and M_q used in the study in Chapter 7. The transfer function of Y_cY_p is calculated for the value of K and the crossover frequency, PM and GM found from the magnitude and phase of the bode plot. The RMS error between commanded pitch rate and pitch rate output is then computed for cases where the transfer function meets the PM and GM criteria. Where the criteria are not met, a large value for RMS error is stored in the matrix. After all requested values of Ke have been tested, the K value corresponding to the smallest RMS error is found and outputted as the optimum k value along with the associated pilot model parameters.

```
% load in the aircraft space state matrices
load A_matrix
load B_matrix
load C_matrix
load D_matrix
%load tau trajectory
load tau_trajectory
% Set pilot model variables
    PTD=0.15;
    k1=1;
    k2=2;
    T1=5;
    T2=0;
% Set up test point matrix
atd=[0 0.1 0.2 0.3];
mq=[-1.3334 -1.3334/2 -1.3334/4];

% Run optimisation for each combination of transport delay and Mq
value
for iatd=1
    for imq=1:3
    ATD=atd(iatd);
    A_matrix(3,3)=mq(imq);

    kenum=1; % position in vector
for ke=-1:-1:-100;% pilot gain in pilot model

%Run inverse simulation
model='structural_pilot_model_tau_q_command';
% Create the linearization I/O of pilot-vehicle system (YpYc)
ios(2) =
linio('structural_pilot_model_tau_q_command/Demux',1,'out');
ios(1) = linio('structural_pilot_model_tau_q_command/Sum2',1,'in');
% Linearize the model
Y= linearize(model,ios);
% Define the transfer function of YpYc
[num,den]=ss2tf(Y.a,Y.b,Y.c,Y.d);
Ytf=tf([num],[den],'InputDelay',PTD+ATD);
%Determine the frequency response magnitude and phase of YpYc
[mag phase w]=bode(Ytf);
```

```
dby=db(mag(1,:));
pp=(phase(1,:));

% determine a) frequency where mag=0db
%           b) phase where mag=0db
%           c) freq where phase-180deg
%           d) mag where phase=180deg

for i=2:length(dby)
    if dby(i)<0 && dby(i-1)>0
        yy1=dby(i-1);
        yy2=dby(i);
        w1=w(i-1);
        w2=w(i) ;
    end

    if pp(i)<-180 && pp(i-1)>-180
        yp1=pp(i-1);
        yp2=pp(i);
        w3=w(i-1);
        w4=w(i);
    end
end

    if exist('yy1')==1
        grad=(yy2-yy1)/(w2-w1);
        f0db=w2-yy2/grad; %frequency where mag=0d
        f1=find(w==w1);
        f2=find(w==w2);
        pp1=pp(f1);
        pp2=pp(f2);
        grad2=(pp2-pp1)/(w2-w1);
        phase_0db=pp2-grad2*(w2-f0db);%phase where mag=0db
    else
        phase_0db=0;
    end

    if exist('yp1')==1
        grad3=(yp2-yp1)/(w4-w3);
        freq_180deg=w4-(yp2+180)/grad3; %freq where phase=180deg
        f3=find(w==w3);
        f4=find(w==w4);

        yy3=dby(f3);
        yy4=dby(f4);
        grad4=(yy4-yy3)/(w4-w3);
        mag_180deg=grad4*(freq_180deg-w4)+yy4; %mag where
phase=180deg
    else
        freq_180deg=0;
        mag_180deg=0;
    end
 %Calculate crossover frequency
for i=2:length(dby)
    if dby(i)<0 && dby(i-1)>0
        yy1=dby(i-1);
        yy2=dby(i);
        w1=w(i-1);
        w2=w(i);
    end
```

```
    if exist('yy1')==1
        grad=(yy2-yy1)/(w2-w1);
        wc(kenum)=w2-yy2/grad;
    else
        wc(kenum)=NaN; %cases where magnitude doesn't reach unity
    end
end
for i=2:length(pp)
    if pp(i)<-90 && pp(i-1)>-90
        bw(kenum)=w(i-1);
    else
        bw(kenum)=NaN; %case where phase never crosses -180 deg
    end
end
%Calculate phase and gain margins
phase_margin(kenum)=180-abs(phase_0db);
if length(mag_180deg)>0
gain_margin(kenum)=abs(mag_180deg);
else
    gain_margin(kenum)=NaN;
end
 clear mag phase w dby pp ppp yy1 yy2 yy3 yy4 pp1 pp2 yp1 yp2 f1 f2
 clear w1 w2 w3 w4 grad grad2 grad 3 grad4 L sum_error error
clear c_info c_info2 c_info3 phase_0db freq_180deg mag_180deg f0db
% Run inverse simulation and calculate RMS error if phase and gain
margin criteria are met
if phase_margin(Tlnum,kenum)>43 && phase_margin(kenum)<47 &&
gain_margin(kenum)>4
sim('structural_pilot_model_tau_q_command');
error=q_flown.data-q_out.data;
error_squared=error.^2;
sum_error_squared=sum(error_squared);
L=length(error_squared);
RMS(kenum)=sqrt(sum_error_squared/L);
elseif phase_margin(kenum)>45 && gain_margin(kenum)>3.8 &&
gain_margin(kenum)<4.2
sim('structural_pilot_model_tau_q_command');
error=q_flown.data-q_out.data;
error_squared=error.^2;
sum_error_squared=sum(error_squared);
L=length(error_squared);
RMS(kenum)=sqrt(sum_error_squared/L);
%assigning large value of RMS error where criteria are not met
else
RMS(kenum)=10^10;
end
kenum=kenum+1;
end % end of ke loop
% find position of minimum in RMS vector
[kpos]=find(RMS==min(RMS));
% determine equivalent pilot gain, RMS, GM, PM and wco
    ke=kpos*-1;
    ke(iatd,imq)=ke;
    err(iatd,imq)=RMS(Tlpos,kpos);
    PHASE_MARGIN(iatd,imq)=phase_margin(kpos);
    GAIN_MARGIN(iatd,imq)=gain_margin(kpos);
    CROSSOVER_FREQ(iatd,imq)=wc(kpos);
    end
end
```

REFERENCES

[1] Bray, R.S., *"Visual and Motion Cueing in Helicopter Simulation"*, NASA Technical Memorandum, TM-86818, September 1985.

[2] Burki-Cohen, J. S., "*The Effect of Simulator Motion Cues on Initial Training of Airline Pilots"*. American Institute of Aeronautics Modelling and Simulation Technologies Conference, San Francisco, CA, USA, 15-18th August, 2005.

[3] Sparko, A. L., Burki-Cohen, J. S., *"Transfer of Training from a Full Flight Simulator vs. A High Level Flight Training Device with a Dynamic Seat"*, American Institute of Aeronautics Modelling and Simulation Technologies (MST) and Global Navigation and Control (GNC) Conferences, Toronto, Canada, 2nd - 5th August, 2010.

[4] Mitchell, D. G., Hoh, R. H., Atencio, A. Jr., Key, D. L., *"Ground Based Simulation Evaluation of the Effects of Time Delays and Motion on Rotorcraft Handling Qualities"*. US Army AD-A256 921, 1992.

[5] Hodge, S., Perfect, P., Padfield, G., White, M., *"Optimising the Vestibular Cues Available from a Short Stroke Hexapod Motion Platform"*. Proceedings of the 67th American Helicopter Society Forum, Virginia Beach, VA, May 3-5th 2011.

[6] Klyde, D. H., McCauley, M., Crane, P. M., *"A Flight Centred Approach to Assess Dynamic Flight Simulator Force Cueing Fidelity"*. Proceedings of the 49th AIAA Conference, 4-7 January 2011, Orlando, FL.

[7] http://www.trainingsystems.org/publications/simulation/roi_effici.cfm

[8] U.S. Joint Helicopter Safety Analysis Team, "*Calendar Year 2006 Report to the International Helicopter Safety Team*". July 2010.

[9] Strachan, I., *"World Civil Full Flight Simulator Census*". Article in Civil Aviation Training Magazine, Issue 4/2012, pp. 52-80. Halldale, 2012.

[10] Rehmann, A.J., 1995, *"A Handbook for Flight Simulation Fidelity Requirements for Human Factors Research"*. DOT/FAA/CT-TN95/46.

[11] Heffley, R. K., et al, *"Determination of motion and Visual System Requirements for Flight Training Simulators"*. Systems Technology Inc., Technical Report No. 546, August 1981

[12] Anon., *"JAR STD-1H, Helicopter Flight Simulators"*. Joint Aviation Authority, 2001.

[13] Anon., *"JAR-FSTD H, Helicopter Flight Simulation Training Devices"*. Joint Aviation Authority, 2008.

[14] Anon., *"EASA CS-FSTD(H) : Certification Specifications for Helicopter Flight Simulation Training Devices"*. European Aviation Safety Agency, June 2012.

[15] Anon., *"Manual of Criteria for the Qualification of Flight Simulation Devices, Volume I – Aeroplanes"*. International Civil Aviation Organization , ICAO 9625 Third Edition -2009.

[16] Anon. *"Manual of Criteria for the Qualification of Flight Simulation Devices, Volume II – Helicopters"*. International Civil Aviation Organization , ICAO 9625 Third Edition - Draft. 2011.

[17] Padfield, G.D., Casolaro, D., Hamers, M., Pavel, M., Roth, G. and Taghizad, A., *"Validation Criteria for Helicopter Real-Time Simulation Models: Sketches from the Work of GARTEUR HC-AG12"*. European Rotorcraft Forum, 2004.

[18] Anon., *"JAR-STD-1A - Aeroplane Flight Simulation Training Devices"*. Joint Aviation Authority, 2003.

[19] Anon**.** *"ADS-33E-PRF, Handling Qualities Requirements for Military Rotorcraft"*. U. S. Army, 2000.

[20] Perfect, P., White, M., Padfield, P., *"Rotorcraft Simulation Fidelity: New Methods for Quantification and Assessment"*. Aeronautical Journal, Royal Aeronautical Society, February 2013.

[21] Stassen, H. G., Johannsen, G., Moray, N., "Internal Representation, Internal Model, Human Performance Model and Mental Workload", Automatica, Vo.l 26., No. 4, pp811-821, 1990

[22] Hays, Robert. T., *"Simulation Fidelity: A Concept Paper".* US Army Research Institute for the Behavioral and Social Sciences Technical Report 490, 1980.

[23] Cooper, G. E., Harper, R. P., *"The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities"*. NASA TN D-5153, April 1969.

[24] Padfield, G.D., *"Helicopter Flight Dynamics. The Theory and Application of Flying Qualities and Simulation Modelling"*, 2nd ed. 2007, Oxford: Blackwell Publishing.

[25] Key, D., "*Bandwidth and Simduce as a Simulator Fidelity Criteria"*. Proceedings of the NASA/FAA Helicopter Simulator Workshop, pp 147 - 160 93N30690, 1992

[26] Strope, K., Gassaway, B., Harding, J., *"Frequency Domain Verification & Validation of a UH-60A FLIGHTLAB Model with a CONEX Sling Load"*. Proceedings of the 61st American Helicopter Society Forum, Grapevine, TX, June 1-3, 2005.

[27] Padfield, G.D., Charlton, M.T., McCallum, A.T., *"The Fidelity of hi-fi Lynx on the DERA Advanced Flight Simulator using ADS-33 Handling Qualities Metrics".* DRA/AS/FDS/TR96103/1, pp 1-152, Dec 1996.

[28] McCallum A.T., Charlton M.T., *"Structured Approach to Helicopter Simulator Acceptance"*. The Challenge of Realistic Rotorcraft Simulation, RAeS, London, November 2001.

[29] Advani, S. K., & Wilkinson, C. H., *"Dynamic Interface Modelling and Simulation - A Unique Challenge".* Royal Aeronautical Society Conference on Helicopter Flight Simulation. London, UK 2001.

[30] Roscoe, M.F., Thompson J.H., *"JSHIP's Dynamic Interface Modelling and Simulation Systems: A Simulation of the UH-60A Helicopter/LHA Shipboard Environment Task"*. Proceedings of the 59th Annual Forum of the American Helicopter Society, Phoenix, AZ, May 2003

[31] Spira, D., Agnerian, A., Boulianne, M-A., *"Validation of High Fidelity Helicopter Simulation Models: an A/MH-6M Case Study"*. Proceedings of the 62nd American Helicopter Society Forum, Phoenix, AZ, May 9-11, 2006.

[32] Wood, J. R., et al, *"Definition of Acceptable Levels of Mismatch for Equivalent Systems of Augmented CTOL Aircraft"*. Report Number MDC A6792, McDonnell Aircraft Company, September 17th. 1984.

[33] Mitchell, D. G., He, C., Strope, K., *"Determination of Maximum Unnoticed Added Dynamics"*. Proceedings of the AIAA Atmospheric Flight Mechanics Conference, Keystone, CO, 21-14 August 2006.

[34] Sinacori, J.B., *"The Determination of Some Requirements for a Helicopter Flight Research Simulation Facility"*. NASA Technical Report, 1097-1, Systems Technology Incorporated, September 1977.

[35] Gouverneur, B., Mulder, J., van Paassen, M., Stroosma, O., *"Optimisation of the SIMONA Research Simulator's Motion Filter Settings for Handling Qualities Experiments"*. Proceedings of the AIAA Modelling and Simulation Technologies Conference and Exhibit, Austin, TX, 11-14 August 2003.

[36] Advani, S., Hosman, R., *"Revising Civil Simulator Standards - An Opportunity for a Technological Pull"*. Proceedings of the AIAA Modelling and Simulation Technologies Conference and Exhibit, Keystone, CO, 21-24th August 2006.

[37] Heffley, R. K., Clement, W. F., Ringland, R. F., Jewell, W. F., Jex, H.R., McRuer, D.T., Carter, V.E., *"Determination of Motion and Visual System Requirements for Flight Training Simulators"*. US Army Research Institute for Behavioural and Social Sciences, AD A117555, 1981.

[38] Hodge, S. J., Perfect, P., Padfield, G. D., White, M. W., *"Optimising the Vestibular Cues Available from a Short Stroke Hexapod Motion Platform"*. Proceedings of the 67th American Helicopter Society Forum, Virginia Beach, VA, May 3-5th 2011.

[39] Chung, W.W. Y., Schroeder, J. A., and Johnson, W.W., *"Effects of Vehicle Bandwidth and Visual Spatial-Frequency on Simulation Cueing Synchronization Requirements"*. AIAA Paper 97-3655, Aug. 1997.

[40] Atencio, A. Jr., *"Fidelity Assessment of a UH-60A Simulation on the NASA Ames Vertical Motion Simulator"*. NASA TM 104016, 1993

[41] Perfect, P. et al, "*Integrating Predicted and Perceived Fidelity for Flight Simulators*". Proceedings of the 36th European Rotorcraft Forum, Paris, France, September 2010.

[42] Padfield, G.D., Charlton, M.T., Jones, J.P., Bradley, R., *"Where does the workload go when pilots attack manoeuvres?*". 20th European Rotorcraft Forum, Amsterdam, The Netherlands, Sept 1994

[43] Tischler, M. B. and Remple, R. K., *"Aircraft and Rotorcraft System Identification: Engineering Methods with Flight Test Examples"*. AIAA Education Series, Reston, VA, 2006.

[44] Blanken, C. L., and Pausder, H. J., *"Investigation of the Effects of Bandwidth and Time Delay on Helicopter Roll Axis Handling Qualities"*. Journal of the American Helicopter Society, Vol. 39, No. 3, 1994, pp. 24-33.

[45] Lusardi, J., Blanken, C., and Tischler, M. B., *"Piloted Evaluation of a UH-60 Mixer Equivalent Turbulence Simulation Model"*. Proceedings of the 49th American Helicopter Society Forum, May 2003.

[46] Schnell, T., et al., *"Physiological Based Simulator Fidelity Design Guidance"*. Proceedings of the MODSIM World 2011 Conference and Expo; pp. 238-247, 2012.

[47] McRuer, D., Krendel, E., *"Mathematical Models of Human Pilot Behavior"*. AGARD-AG-188, Systems Technology Inc., January 1974.

[48] McRuer, D. T., Jex, H. R., *"A Review of Quasi-Linear Pilot Models".* IEEE Journal of Transactions on Human Factors in Electronics, Vol. HFE-8, No. 3, September 1967.

[49] Mcruer, D. T., Graham, G., Krendel, E., & Reisener, W., *"Human Pilot Dynamics in Compensatory Systems - Theory, Models and Experiments with Controlled Element and Forcing Function Variations"*. AFFDL-TR-65-15, Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, July 1965.

[50] Hess, R. A., *"A Model-Based Theory for Analyzing Human Control Behavior"*. Journal of Advances in Man-Machine Systems Research, Vol. 2, edited by W Rouse, JAI Press, London, 1985, pp. 129-176.

[51] Hess, R. A., Malsbury, T., *"Closed-Loop Assessment of Flight Simulator Fidelity"*. Journal of Guidance, Vol. 14, No. 1, Jan-Feb 1991.

[52] Hess, R. A., "*Unified Theory for Aircraft Handling Qualities and Adverse Aircraft-Pilot Coupling"*. Journal of Guidance, Control and Dynamics, Vol. 20, No. 6, pp. 1141-1148, November-December 1997.

[53] Zeyada, Y., Hess, R.A., *"Computer-Aided Assessment of Flight Simulator Fidelity"*. Journal of Aircraft, Vol. 40, No. 1, January-February 2003.

[54] Pool, Daan M., "*Objective Evaluation of Flight Simulator Motion Cueing Fidelity through a Cybernetic Approach"*. Dissertation, Delft University of technology, ISBN 978-94-6186-032-3, September 2012

[55] Hess, R. A., Siwakosit, W., *"Assessment of Flight Simulator Fidelity in Multiaxis Tasks Including Visual Cue Quality"*. Journal of Aircraft, Vol. 38, No. 4, July-August 2001.

[56] Hess, Malsbury, *"Flight Simulator Fidelity Assessment in a Rotorcraft Lateral Translational Maneuver"*. Journal of Guidance, Control and Dynamics, Vol. 16, No. 1, January-February 1993.

[57] Thomson, D., Bradley, R., *"The use of Inverse Simulation for Assessment of Helicopter Handling Qualities"*. The Aeronautical Journal, September 1997.

[58] Heffley, R. K., *"Application of Task-Pilot-Vehicle (TPV) Models in Flight Simulation"*. Proceedings of the 66th American Helicopter Society Forum, Phoenix, AZ, May 11-13, 2010.

[59] Bradley, R., Brindley, G., *"Progress in the Development of a Robust Pilot Model for the Evaluation of Rotorcraft Performance, Control Strategy and Pilot Workload"*. Proceedings of the 8th European Rotorcraft Forum, Bristol, UK, Sept 2002.

[60] Padfield, G.D., White, M.D., *"Measuring Simulation Fidelity through an Adaptive Pilot Model"*. Aerospace Science and Technology, Vol. 9, (5), pp. 400 – 408, July 2005.

[61] Szalai, K. J., *"Validation of a General Purpose Airbourne Simulator for Simulation of Large Transport Aircraft Handling Qualities"*. NASA TN D-6431, October 1971.

[62] Ferguson, S. W., Clement, W. F., Cleveland, W. B., Key, D. L., *"Assessment of Simulation Fidelity Using Measurements of Piloting Technique in Flight"*. Proceedings of the 40th American Helicopter Society Forum, Arlington, Virginia, US, and 16-18 May 1984.

[63] White, M. D., Perfect, P., Padfield, G., *"Progress in the Development of Unified Fidelity metrics for Rotorcraft Flight Simulators"*. Proceedings of the 66th American Helicopter Society Forum, Phoenix, Arizona, US, 11-13 May 2010.

[64] Roscoe, A. H., Ellis, G. A., *"A Subjective Rating Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use"*. NASA TR 90019, Royal Aerospace Establishment, March 1990.

[65] Roscoe, S., *"Incremental Transfer Effectiveness"*. Air Force Office of Scientific Research TR-72-1664, February 1971

[66] Holman, G. L., *"Training Effectiveness of the CH-47 Flight Simulator"*. US Army, Research institute for the Behavioural and Social Sciences, Report 1209, 1979.

[67] Gubbels, A.W., Ellis, D.K., *"NRC Bell 412 ASRA FBW Systems Description in ATA100 Format"*. Institute for Aerospace Research, National Research Council Canada, Report LTR-FR-163, 2000.

[68] Manimala, B., Walker, D., Padfield, G., *"Rotorcraft Simulation Modelling and Validation for Control Design"*. The Aeronautical Journal of the RAeS, Vol. 111, No. 1116, Feb 2007, pp 77-88.

[69] White, M., Perfect, P., Padfield, G., Gubbels, A., Berryman, A., "*Acceptance Testing and Commissioning of a Flight Simulator for Rotorcraft Simulation Fidelity Research*". Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering Volume 227 Issue 4, pp663 – 686, April 2013

[70] DuVal, R.W., "*A Real-Time Multi-Body Dynamics Architecture for Rotorcraft Simulation*". The Challenge of Realistic Rotorcraft Simulation, RAeS Conference, London, November 2001.

[71] Bickerstaff, I. H. *"Portrait of Landscape: a visualisation solution for military aircraft development"*. Presented at 1998 IMAGE Conference, Scottsdale, Arizona, 2-7 August1998.

[72] C., R., Brackbill, *"Helicopter Rotor Aeroelastic Analysis Using a Refined Elastomeric Damper Model"*. PhD Thesis in Aerospace Engineering submitted to the Pennsylvania State University, December 2000.

[73] W. L., Cresap and A. W., Myers, *"Design and Development of the Model 412 Helicopter"*. Journal of the American Helicopter Society, Vol. 36, May 1980.

[74] Prouty, R.W., *"Helicopter Performance, Stability and Control"*, pp. 689, Krieger Publishing Company, 2005.

[75] Manimala, B., *"Analysis of the Blade Mass Distribution for the Bell-412"*. Flight Science and Technology internal report, University of Liverpool, October 2004.

[76] M., Nakadate, H., Taguchi, J., Takaki, *"Design and Test Evaluation of FBR Bearingless Main Rotor"*. Journal of the American Helicopter Society, Vol. 46, 2, April 2001.

[77] Hoh, R., *"Lessons Learned Concerning the Interpretation of Subjective Handling Qualities Pilot Rating Data"*. Proceedings of the AIAA Atmospheric Flight Mechanics Conference, Portland, OR, August 20-22, 1990.

[78] Hodge, S., Perfect, P., *"Transport Delays Measured in the HELIFLIGHT-R Simulator".* Flight Science and Technology internal report, University of Liverpool 2012.

[79] Sedqick, P., *"Pearson's Correlation Coefficient"*, British Medical Journal Online, 7 July 2012, ISSN 17561833.

[80] Heffley, R., et al, *"Determination of Motion and Visual requirement for Flight Training"*. U.S Army Research for Institute for the Behavioural and Social Sciences Technical Report 546, August 1981.

[81] Cooper, G. E., and Harper, R. P., Jr., *"The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities"*. NASA TN D-5153, National Aeronautics and Space Administration, Washington, DC, April 1969.

[82] Timson, E., Perfect, P., White, M., Padfield, G., Erdos, R., Gubbels, A., *"Pilot Sensitivity to Flight Model Dynamics in Rotorcraft Simulation"*. Proceedings of the 37th European Rotorcraft Forum, Gallarate, September 2011.

[83] Roscoe A., H., and Ellis, G., A., *"A Subjective Scale for Assessing Pilot Workload in Flight: A Decade of Practical Use"*. Royal Aerospace Establishment Technical Report 90019, March, 1990.

[84] Roscoe, A., *"Heart Rate as an In-Flight Measure of Pilot Workload"*. Technical Memorandum FS(B) 464, March 1982

[85] Ellis, K., "*Eye Tracking Metrics for Workload Estimation in Flight Deck Operations"*. Dissertation from University of Iowa, 2009.

[86] Lindholm, E., et al, *"Physiological Assessment of Aircraft Pilot Workload in Simulated Landing and Simulated Hostile Threat Environments".* U. S. Air Force Systems Command, 1984.

[87] Rolfe, J. M., *"The Secondary Task as a Performance Measure: A survey of the Literature".* Royal Air Force Institute of Aviation Medicine Report Number 474, June 1969.

[88] Wickens, C., *"Multiple Resources and Performance Prediction"*. Theoretical Issues in Ergonomics Science, Vol. 3, Iss. 2, pp.159-177, 2002.

[89] Snedecor, George W.; Cochran, William G. (1967), *"Statistical Methods"*, 6th edition. pp. 321.

[90] Nathans, Oswald & Nimon, “Interpreting Multiple Regression Results”, Journal of Practical Assessment, Research & Evaluation, Vol 17, No 9, April 2012. ISSN 1531-7714

[91] Heffley, R. K., et al, *“A Compilation and Analysis of Helicopter Handling Qualities Data, Vol. One: Data Compilation”*. NASA CR-3144, March 1979.

[92] Lee, D.N., *“A Theory of Visual Control of Braking Based on Information about Time-to-Collision”*. Perception, 1976, 5, pp 437-459.

[93] Padfield, G. D., *“The Tau of Flight Control”*. The Aeronautical Journal of the RAeS, Vol. 115, No. 1171, September 2011.