Cranfield University

Matthew Ebbatson

The Loss of Manual Flying Skills in Pilots of Highly Automated Airliners

School of Engineering

PhD

Cranfield University

School of Engineering

PhD Thesis

2009

Matthew Ebbatson

The Loss of Manual Flying Skills in Pilots of Highly Automated Airliners

Supervisors: Dr. J. Huddlestone and Dr. D. Harris

Academic Years 2005 to 2009

“What is chiefly needed is skill rather than machinery”

Wilbur Wright
13 May 1900

Acknowledgements

I would like to thank all of those people at Cranfield and the contributing organisations who made this work possible. Also, thank you to my friends and family for their kind support and encouragement throughout. Most importantly, thank you to Nathalie, for always having confidence in me.

Abstract

Anecdotal and subjective evidence suggests that the manual flying ability of pilots operating highly automated aircraft is declining owing to a lack of opportunity to exercise such skills in the modern air transport environment. However, there is a paucity of objective evidence to support this safety concern. Consequently, the work presented in this thesis aims to provide empirically derived data to evaluate the extent and causes of the speculated manual skills decline and guide possible intervention strategies.

Initially a cognitive task analysis is undertaken to determine the cognitive demands of performing manual flight in a large jet transport aircraft. Expert pilots report employing highly refined mental models structures which enable them to predict the aircrafts performance whilst causing minimal burden to their mental capacity. The study concludes that when measuring manual flying performance careful consideration must be given to designing a task which challenges both the cognitive and physical aspects of manual flying skill.

Secondly, relatively novel pilot performance measures based upon the frequency analysis of control input data are evaluated. An empirical study finds that these techniques are both reliable and sensitive to manual flying performance. Furthermore, when studying large transport aircraft, such measures of the pilots control strategy are found to contribute valuable information about performance which is missing when just traditional ‘outer-loop’ performance measures are applied. The study concludes that these measures of control strategy are valuable in evaluating manual flying performance.

Finally, the manual flying skills of a sample of pilots of highly automated aircraft are evaluated on a challenging manual flying task. A significant proportion exhibit poor manual flying performance as judged by a type rating examiner. Further analysis reveals that the performance of the pilots is significantly influenced by the amount of recent manual handling experience they have accumulated, rather than their longer-term manual flying experience. Significantly, airspeed tracking ability is influenced which is cited elsewhere as a causal factor in many manual flying skill related accidents. The results support the previous anecdotal and subjective concerns relating to the loss of manual flying skills.

Publications Resulting from this Work

Journal Articles:

Ebbatson, M. (2006). *Practice makes imperfect: common factors in recent manual approach incidents.* Human Factors and Aerospace Safety 6(3), 275-278.

Ebbatson, M., Huddlestone, J., Harris, D. and Sears, R. (2007). *The Application of Frequency Analysis Based Performance Measures as an Adjunct to Flight Path Derived Measures of Pilot Performance*. Human Factors and Aerospace Safety 6(4), 383-394.

Ebbatson, M., Harris, D., Huddlestone, J. and Sears, R. (2008). *Combining Control Input with Flight Path Data to Evaluate Pilot Performance in Transport Aircraft.* Aviation, Space and Environmental Medicine 79(10), 1061-1064.

Ebbatson, M., Harris, D., Huddlestone, J. and Sears, R. (2008). *The Relationship Between Manual Handling Performance and Recent Flying Experience in Air Transport Pilots*. Ergonomics. *In review.*

Conference Proceedings:

Ebbatson, M., Huddlestone, J., Sears, R. and Harris, D. (2006).
Loss of manual flying skills: Evaluating objective performance measures. *27th Conference of the European Association for Aviation Psychology*; September 24th – 28th; Potsdam, Germany.

Industry Presentations:

FORCE Safety Research for Flight Operations, Gatwick, March 2007
GAPAN Technical & Safety Committee, April 2008
RAeS Flight Crew Training Conference, September 2008

Table of Contents

Table of Figures

Table of Tables

Glossary

ACTA	Applied Cognitive Task Analysis
AAIB	Air Accident Investigation Branch
ALAR	Approach and Landing Accident Reduction (tool kit)
ATC	Air Traffic Control
ATPL	Air Transport Pilot Licence
ATQP	Alternative Training and Qualification Programme
ATSB	Australian Transport Safety Bureau
CAA	Civil Aviation Authority (United Kingdom)
CATII	Category Three Approach
CPL	Commercial Pilots Licence
CSV	Comma Separated Variable (file format)
CTA	Cognitive Task Analysis
DFT	Discreet Fourier Transform
DME	Distance Measuring Equipment
EADI	Electronic Attitude Directional Indicator
EFIS	Electronic Flight Instrument System
EHSI	Electronic Horizontal Situation Indicator
FAR	Federal Aviation Regulations
FDM	Flight Data Monitoring
FDR	Flight Data Recorder
FMGS	Flight Management and Guidance System
FMS	Flight Management System
FODCOM	Flight Operations Department Communication
FOQA	Flight Operations Quality Assurance
FSF	Flight Safety Foundation
FTD	Flight Training Device
HFACS	Human Factors Analysis and Classification System
HTA	Hierarchical Task Analysis
Hz	Hertz
ILS	Instrument Landing System
IMC	Instrument Meteorological Conditions
IOS	Instructors Operating Station
IPC	Instrument Proficiency Check
IR	Instrument Rating
JAA	Joint Aviation Authorities
JOT	Jet Orientation Training
LOFT	Line Orientated Flight Training
LPC	Licence Proficiency Check
LR	Logistic Regression
ME	Mean Error
MSC	Mean Power Spectral Density
NG	Next Generation
NTSB	National Transportation Safety Board
OPC	Operational Proficiency Check
PF	Pilot Flying

PM	Pilot Monitoring
PSD	Power Spectral Density
QAR	Quick Access Recorder
QNH	Sea Level Atmospheric Pressure
QRH	Quick Reference Handbook
RMSE	Root Mean Square Error
SDE	Standard Deviation of Error
SME	Subject Matter Expert
SOP	Standard Operating Procedure
SPSS	Statistical Package for Social Sciences
SMS	Safety Management System
TOD	Top of Descent
TOT	Time Outside of Tolerance
TRE	Type Rating Examiner
UK	United Kingdom
VFR	Visual Flight Rules

Chapter 1

Introduction

Given that weather conditions over the UK were relatively benign, the pilots of an Airbus A321 airliner returning to Nottingham East Midlands airport elected to perform a manual approach. This involved disengaging the auto-pilot and auto-thrust systems and ‘hand flying’ the aircraft via its side-stick controller, rudder pedals and thrust levers. The flight director system remained engaged, providing assistive flight guidance information to the crew. However, because they failed to select the approach mode on the Flight Management and Guidance System (FMGS) the flight director did not prompt the crew to descend towards the runway. Consequently, the aircraft was allowed to fly through the Instrument Landing System (ILS) glideslope and became slightly high on the approach. The crew, realising their error, elected to deselect the flight director and continue following the ILS using raw data alone. The slight perturbation in profile was relatively minor and should have been safely recoverable in the distance remaining. However, the handling pilot was unable to perform a stable approach in this manual condition. The aircraft oscillated significantly on the ILS and the airspeed was allowed to bleed off excessively. The aircraft arrived over the runway threshold with low energy, requiring the pilot to command an unusually nose high attitude in order to arrest its rate of descent. The touch down was heavy and the high body angle caused the aircraft to strike its tail on the runway surface, resulting in substantial structural damage. The subsequent accident investigation found the aircraft to be fully serviceable prior to the event. It highlighted several issues with the flight but primarily noted that the pilot’s manual flying skill was inadequate and a significant contributory factor (AAIB, 2002). A review of the recent UK accident and incident reports (Ebbatson, 2006) revealed many similar events, all in which the manual flying ability of the handling pilot was considered to be lacking. This thesis will investigate how differences in exposure to manual flight operations influence the strength of a pilot’s manual flying skill.

1.1 Background

The evolution of the airliner flight deck from primitive wheel-house to highly automated control room has been driven by the concurrent desires for increased efficiency and safety in flight operations. In the past, the approach to the integration of flight deck automation has been predominantly technology driven. Automatic functionality has often been incorporated into aircraft designs as and when it has become available, simply with the view that more is better (Billings, 1997; Hawkins, 1998). The support offered by automation on the flight deck has been touted as a means of freeing up crew capacity and suppressing human error at its source, by substitution of the human. However the validity of this approach has been strongly challenged on human factors principles (Wiener, 1988; Billings, 1997; Parasuraman, Sheridan and Wickens, 2000; Goteman and Dekker, 2003). It is apparent that automation sometimes fails to work as a co-operative crew member and, rather than aiding pilots in their work, it becomes a hindrance which must be ‘worked around’.

During initial training, all pilots are taught the complex psychomotor and cognitive skills required to control their aircraft by physical manipulation of the primary flying controls i.e. basic manual flying skills (see JAR-FCL). Conversely, during routine operation of a modern jet transport aircraft it is more common for the flight path and energy to be controlled by a combination of automated systems (i.e. the auto-pilot and auto-throttle). In this mode of operation the psycho-motor aspect of control is minimal and the cognitive aspect is modified (Damos, John and Lyall, 2005; Latorella, Pliske, Hutton and Chrenka, 2001), with the emphasis of the pilots work shifted towards higher order cognition (i.e. complex decision making & problem solving).

The opportunity for airline pilots to practise basic manual flight is usually minimal, although this is somewhat modulated by the type of carrier, its fleet, operational philosophy and route network. For example, two of the most prominent European low cost carriers have quite distinct operational philosophies, one encouraging routine manual flight over use of the automatics wherever possible and the other vice versa. The only mandatory requirement for manual flying proficiency to be evaluated (outside of initial training and type rating) is during the biannual Operator Proficiency Check (OPC) and annual License Proficiency Check (LPC). Even on these occasions only a small set of manual handling tasks with limited scope are stipulated (see CAA Standards Document 24).

However, the ability of the pilot to revert to basic manual control is essential, for example, in cases where the aircraft's automatic capability is diminished or when reconfiguring the automatics is an ineffective use of the crew's capacity (Amalberti, 1998). It is conceivable that, due to the infrequent opportunity to exercise manual flying skill in modern flight operations, crew may experience "out-of-the-loop unfamiliarity" (Wickens, 2000) and their basic flying ability may diminish over time. The threat of this skill fade is a concern shared by pilots, operators, regulators, manufacturers and researchers alike (Baron, 1988; Childs and Spears, 1986; Parasuraman, Molloy and Singh, 1993; Tenney, Rogers and Pew, 1998).

There is much subjective affirmation of this concern in the manual flying skills literature (see following sections) but a paucity of objective data with which to support it, nor much suggestion as to a viable future strategy for monitoring manual flying ability on a day-to-day basis. In order for regulators and operators to address the safety concern higher quality, empirically derived data must be available which better defines the nature of the potential problem. This research programme aims to provide these data and develop methods to assist other research in the field.

1.2 Rationale for Research

1.2.1 Automated Flight

The late 1970s were a turning point in the evolution of flight deck design. Rising oil prices coupled with advances in microprocessor technology saw the emergence of ‘two crew’ operations which dispensed with the flight engineers position, their role being absorbed by sophisticated automated systems. This transition significantly shaped the modern flight deck environment. The literature divides flight deck automation into three principle types (Billings, 1997), these being information automation, control automation and management automation.

Information automation relates to the presentation of flight related information and is encapsulated by the Electronic Flight Instrument System (EFIS) found in most modern aircraft. When operating aircraft equipped with these systems the crew view highly processed information via computer generated ‘glass’ displays rather than having to gather and mentally process large amounts of raw data from a disparate array of electro-mechanical gauges.

Control automation governs the aircraft’s flight path and energy and when engaged has (limited) authority over the principle flying controls. Modern autopilots and auto throttles have advanced significantly since the early generation jets. They allow the pilot to delegate short term tactical flight goals and are capable of performing sophisticated vertical transitions and automatic landings with incredible reliability and precision.

Management automation relates to the long term strategic planning of the aircraft’s operation and guidance. The Flight Management System (FMS) is perhaps the most significant piece of technology to be incorporated into the modern flight deck. These systems integrate navigational and environmental data and use complex algorithms to continually compute highly optimised flight paths which satisfy a series of strategic goals stipulated by the crew (i.e. fly from waypoint A to waypoint B, cross waypoint C above a certain altitude, climb to an altitude using the most fuel efficient profile etc.). Furthermore the pilots can choose for the FMS to deliver its guidance information directly to the autopilot and auto throttle systems where it can be executed, thereby integrating the automated control and management of the aircraft.

The result of automation has been vast increases in the precision of navigation and the efficiency of aircraft operations. However, it has also brought about a fundamental shift in the way pilots operate their aircraft, redistributing workload and creating fresh opportunity for human error to occur (Sarter, Woods and Billings, 1997; Harris, Hancock, Arthur and Caird, 1995). Pilots have predominantly become aircraft managers rather than direct controllers, spending the majority of their time planning the flight, programming the automation and monitoring its operation rather than actively handling the flying controls (Wood, 2004).

1.2.2 Manual Flying Skill

At a very basic level, manual flying skills may be defined as those which are displaced by the presence of automation on the flight deck. As previously noted automation is responsible for both the processing of flight information and the physical manipulation of the aircraft and therefore it seems logical that in the absence of these automated systems the pilot must employ both cognitive and psychomotor skills to compensate. Certainly the psychomotor aspect is observable in that the pilot must physically actuate the aircrafts primary flying controls to govern its orientation and trajectory in the absence of the autopilot and autothrottle. However, the previous examination of automation functionality shows us that in the absence of information and management automation systems, such as the FMS and electronic flight displays, the pilot must also employ significant cognitive skills. These skills are required in order to assess the aircrafts current condition, predict its future state and plan flight paths which satisfy navigational requirements.

Naturally the demand placed on these manual flying skills is not uniform across the flight profile but is dependent upon the transient nature of the aircraft. For instance during cruising, straight and level, flight the aircrafts state tends to be relatively steady and when properly trimmed few physical inputs are required and little cognitive effort is required to monitor the situation. Contrastingly, during the departure and climb and the approach to landing phases the aircraft's state tends to be highly transitory both in the horizontal and vertical planes. It takes far more physical input to guide the aircraft during this phase and much more cognitive involvement to monitor and predict its path and energy. A more detailed definition of the differences between manual and automated flying skills will be developed throughout the course of the thesis.

1.2.3 Evidence for the Loss of Manual Flying Skills from Pilot Attitude Surveys

As highly automated airliners began to enter operation a great deal of new human factors research also began, much of it focusing on the skills required to operate the novel technology and potential errors which may occur (Curry, 1985; Wiener, Chute and Moses, 1999; Tenney, Rogers and Pew, 1998). Researchers also showed interest in the attitudes of the pilots who were the first to transition to these new types. The earliest signs that pilots were concerned about the potential loss of their manual flying skills owing to the operation of highly automated airliners emerged through this work. Although these surveys of opinion extended for more than a decade few empirical measurements of flying proficiency were collected during this time to substantiate or refute the concerns they presented.

Curry (1985) distributed a cockpit automation attitude measurement scale to recently appointed Boeing 767 pilots who had transitioned from older generation aircraft. Over 80% of the sample 'strongly agreed' that the new generation of flight deck technology could lead to a degradation of manual flying skill, however only 63% of respondents believed their own skills had suffered. This disparity was either an artefact of the pilots' self rating biases, or may indicate that crew of the time were still working around the automatics and operating their aircraft like they had the previous generation of 'manual' airliners. The latter suggestion was supported by a second statistic from the research which reported that 87% of the sampled pilots were seeking to hand fly the aircraft as much as possible on every sector. Similar results were obtained in a study of crew transitioning from the early generation DC-9-10/30 to its highly automated variant, the DC-9-80 (MD-80) (Curry, 1985) and also between electro-mechanical and highly automated variants of the Boeing 737 (Wiener, Chute and Moses, 1999).

Flight crews of this generation were evidently sufficiently confident in their manual flying ability to disengage the automatics from time to time and exercise their skills. It is likely that this is because those crews had a considerable foundation of manual operating experience to fall back on. However the modern pilot demographic is considerably different to that which existed at the time of the aforementioned research. Highly automated airliners are now prolific amongst the airline fleets of developed nations and have been for some time. Even many highly experienced pilots have likely spent the majority of their career operating highly automated types. With smaller turboprops and regional jets (the starting point for many flying careers) now also incorporating highly automated flight decks many of the younger generation of pilots may not have been exposed to ‘manual’ types outside of their *ab-initio* training.

The current generation of pilots may therefore conceivably lack the same foundation of manual handling experience which gave earlier generations the confidence to routinely revert to manual control and maintain proficiency (Curry, 1985). With a lack of experience feeding a lack of confidence to build experience the problem could likely worsen. The phenomenon is exacerbated by other developments in the air transport environment. As highly automated aircraft have demonstrated increasingly reliable and precise navigation, operational procedures have evolved to exploit this capability, allowing higher traffic flows through airspace whilst improving environmental and economic performance. Many airport departure and arrival navigational procedures are now highly complex and principally designed to be flown via the automatics. They can be difficult to fly manually in large high performance aircraft.

Furthermore airlines have introduced Flight Data Monitoring (FDM) programmes to feedback operational performance data into their Safety Management Systems (SMS). Whilst these programmes have been hugely beneficial to flight safety, if they are improperly implemented they could potentially further deter crews from disengaging the automatics since poor manual performance is likely to be detected and questioned.

Owen and Funk (2007) summarised the concern more recently when they undertook an online meta-review of perceived flight deck automation problems, citing evidence from research literature to either support or refute a collated set of issues. Part of this review detailed manual flying skills issues caused by the operation of highly automated aircraft. A total of 31 pieces of evidence were found to support the statement “Pilots may lose psychomotor and cognitive skills required for flying manually or for flying non-automated aircraft, due to extensive use of automation”. Although this is a reasonably large body of evidence, the nature of the data sources limits its objectivity. The data were almost exclusively derived from the compilation of subjective pilot opinion (bar a single incident survey and two citations from accident reports), the majority of which is sourced from just two principle research studies (Curry, 1985; Wiener, 1989). Whilst this evidence provides compelling support for the existence of a safety concern, it does not offer an objective grounds to measure the extent of the problem. Furthermore, it is evident that the assumed definition of ‘manual flying skill’ varies considerably between sources. In some cases just the physical components of skill are considered, relating primarily to the impact of control automation (see Billings, 1997). In other cases the cognitive aspects of manual flying skill are considered, relating to the impact of information and management automation (Billings, 1997).

1.2.4 Evidence for the Loss of Manual Flying Skills from Accident Data

The analysis of past accidents and incidents is an important means of assessing where operational risks exist and guiding intervention strategies.

Previous studies have reported that between 70% and 80% of aviation accidents result from some form of human error (O'Hare, Wiggens, Batt and Morrison, 1994). However, establishing accident causality is notoriously difficult since there are typically multiple convening factors which make up an accident sequence. Consequently, it may be complex to isolate the contributory factors, such as manual handling deficiency, and to distinguish cause from effect. However, frameworks and taxonomies have been created which aim to make the analysis of these events more systematic.

The Human Factors Analysis and Classification System (HFACS) is one such taxonomy which extracts the human factors which contribute to an accident or incident sequence. HFACS has been widely applied and well validated (Wiegmann and Shappell, 2003). It is based upon Reasons 'Swiss cheese' accident model (1990) and structured around four hierarchical levels 1) Organisational influences 2) Unsafe supervision 3) Preconditions for unsafe acts, and finally 4) Unsafe acts of the operator (see figure 1). Each level is sub-divided into more specific elements.

When examining the contributory actions of the flight crew to accidents, as is the case when looking for evidence of manual flying deficiencies, the focus falls on the 'unsafe acts' level of HFACS. Unsafe acts are broken down into three types of error 1) Decision errors 2) Skill-based errors, and 3) Perceptual errors. Manual flying deficiencies are encompassed principally by the 'skill-based error' category which is defined by Weigmann & Shappell (2003) as "stick-and-rudder and other basic flight skills that occur without significant conscious thought". The authors of the methodology cite 'breakdown in visual scanning', 'poor technique' and 'over-controlled the aircraft' as typical aviation skill based errors.

There are distinct performance phases which define the process of skill acquisition. Rasmussen's skill-rule-knowledge framework (Reason, 1990) is a commonly adopted model which defines a tripartite of performance levels. At the lowest level "knowledge-based" performance is applied in novel situations where the performer must used detailed on-line analytical processes to understand a situation and formulate an effective course of action. Errors at this level typically occur because the performer has an incomplete or erroneous knowledge base or insufficient processing resources. Pilots who have had little opportunity to build manual flying knowledge may exhibit this form of erroneous behaviour, becoming overwhelmed by the task demands.

With increasing expertise the performer may then exhibit “rule-based” behaviour, where familiar problems can be diagnosed and solved using stored rules of the form if (state) then (action). The level of conscious activity is somewhat reduced. Errors at this level typically occur because a situation is misdiagnosed and the wrong rule is applied. Finally, at the highest level, expert performers apply skill based behaviour whereby upon diagnosis complex sequences of pre-programmed instructions are executed, largely without any dedicated conscious monitoring. These action structures allow for cognitive efficiency since they place little demand on the information processing channels. The manual flying inputs of a highly skilled pilot will therefore be made largely with little conscious effort and in response to very sophisticated situational assessments. However, the manual control inputs of less skilled pilots will be made very consciously and demand far greater information processing bandwidth.

Meta-cognition is an awareness of ones own cognitive performance, particularly relating to the acquisition of skill. An individual with heightened meta-cognitive ability understands the process of skill acquisition as well as their own position in that process, thus allowing them to enhance their learning performance. Very skilled pilots who are reflective with good meta-cognitive ability may thus be able to recognise inefficiency in their cognitive performance during manual flight and adapt their information gathering and assessment processes to suit.

Using HFACS, a review of the US National Transport Safety Bureau’s (NTSB) commercial aviation accident records for the years 1990 through 1996 (Weigmann and Shappell, 2001) revealed that 63.6% of accidents occurring to FAR part 121 operations (scheduled passenger or cargo airlines operating large transport category aircraft) involved at least one skill-based error. Errors of this category were by far the most prominent in the data set and remained at a fairly consistent level throughout the seven year sample period.

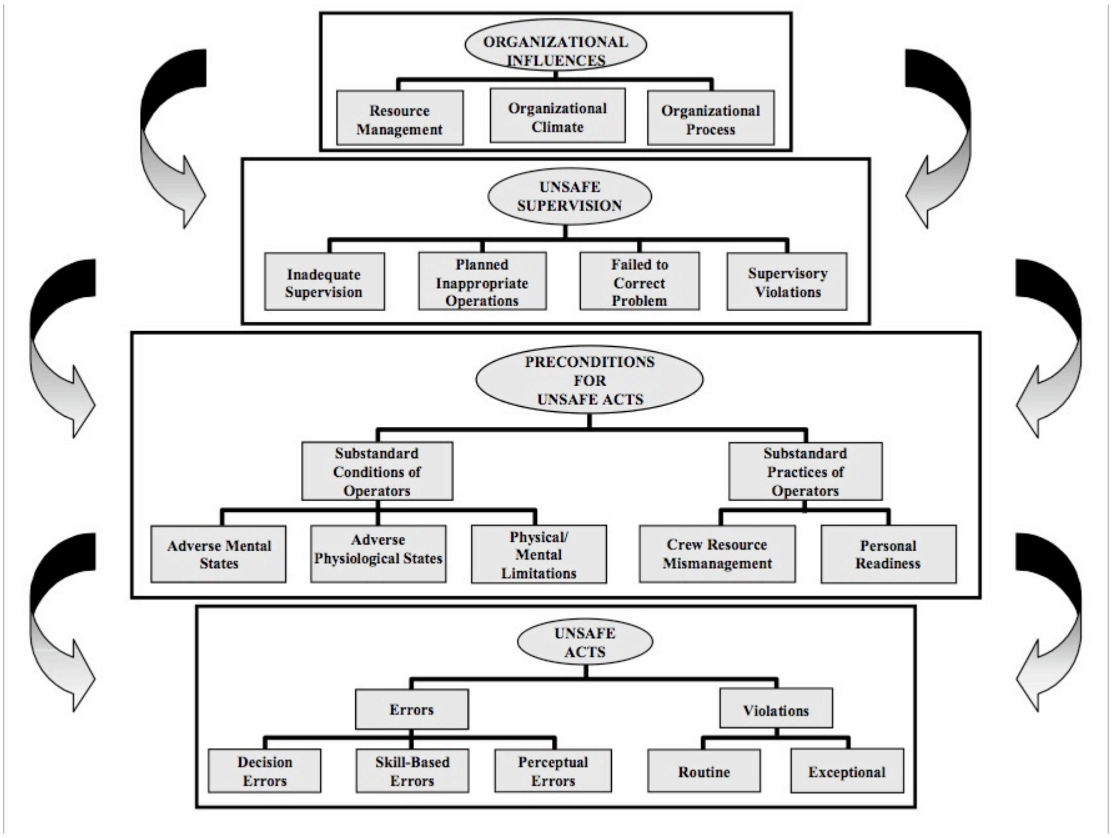


Figure 1 - The HFACS model (reproduced from Weigmann and Shappell, 2003)

A more recent report published by the Australian Air Transport Safety Bureau (2004) analysed accidents occurring to Australian registered aircraft operating over their national territory during the period 1993 to 2002 using HFACS. The use of a common taxonomy allowed the analysis to be compared alongside equivalent data from the US. The analysis showed that 84% of Australian accidents and 77% of US accidents involved at least one skill based error. However it should be noted that the data set was not restricted to FAR part 121 type operations and in fact approximately 80% of the accidents related to general aviation operations. No specific break down is given for the percentage of skill-based accidents occurring to air carrier operations and so it is difficult to make direct comparisons with the earlier NTSB analysis.

The aforementioned HFACS analyses do not explicitly define the number of accidents in which poor manual flying skills were directly attributed as a casual factor. However they do suggest that a significant proportion of accidents occurring to large air transport aircraft involve a skill-based error, giving scope to the proposition that manual flying skill deterioration represents a significant threat to flight safety.

The UK Civil Aviation Authority (CAA) Accident Analysis Group published a review of global fatal accidents occurring over the period 1997 through 2006 to large public transport aircraft (CAP 776). The group use a bespoke taxonomy rather than the HFACS system, allocating primary causal factors, causal factors, causal groups and circumstantial factors. In a similar result to previous studies flight crew related causal factors were listed for 78% of the fatal accidents. More specifically ‘flight handling’ was listed as a primary causal factor in 14% and a causal factor in 29% of all fatal accidents. The group reports that flight handling ‘tended to be associated with inadequate speed, pitch attitude and/or directional control, often following an engine failure, resulting in the aircraft stalling’. When sorted by consequence, 17% of the events involved a loss of control in flight, following non-technical failure (the report cites the example ‘flight crews inadequate speed control’) and 63% of these events involved a flight handling causal factor. Therefore flight crew’s handling of the aircraft was often cited as a contributory factor to fatal accidents in cases where no aircraft malfunction existed.

The CAA report appears to give more direct evidence about the significant role of flight crews’ manual flying skill in large transport aircraft accidents. However the definition of ‘flight handling’ assumed by the report is not explicitly detailed and it is not clear if it also includes flight control actions performed through the auto flight systems. Some caution must therefore be given in the interpretation of these results.

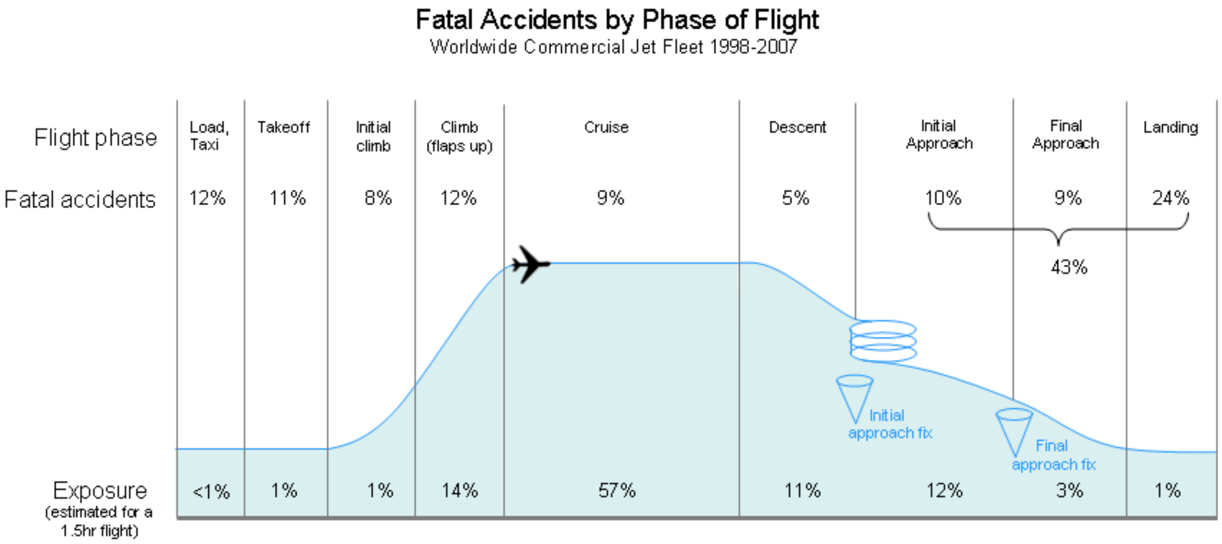


Figure 2 –Distribution of fatal accidents by flight phase for the period 1998 to 2007 (adapted from Boeing 2008)

Boeing (2008) recently conducted an analysis of accident data pertaining to the worldwide commercial jet fleet, including the distribution of fatal accidents by flight phase for the period 1998 to 2007 (see figure 2). Notable is the dominance of the approach and landing phase, which is where 43% of fatal accidents occur despite this only representing approximately 16% of the typical flight period. Contrastingly, the takeoff and climb phase, which also accounts for 16% of the flight period, was associated with 31% of fatal accidents and the cruise phase, which represents a sizable 57% of the total flight period, was associated with just 9% of fatal events. The approach and landing phase thus seems to be a particularly high risk and perhaps demanding flight phase, and when combined with the previously mentioned CAA data, likely to involve loss of control issues. This view is shared by the Flight Safety Foundation (2000) who has long labelled approach and landing accidents as “the biggest killers in aviation” and has consequentially designed the Approach and Landing Accident Reduction (ALAR) toolkit. It seems prudent that in order to tackle the most critical manual flying skills issues we should first look to this phase of flight.

Ebbatson (2006) noted a trend amongst highly automated aircraft types in a review of the recent UK incident and accident data (2000 to 2006) published by the Air Accident Investigation Branch (AAIB). This includes the accident case study reproduced in the introduction to this thesis. Many reports centred on pilots who had deliberately disengaged their aircraft's automatics at an early stage of the approach and had subsequently demonstrated poor manual handling ability. In the bulletins issued by the AAIB there are accounts of at least two very similar incidents occurring relatively recently. On both occasions the autopilot, auto thrust and flight director systems were disengaged at an early stage of the approach, as in the East Midlands incident. Also on both occasions there were significant deviations on the localiser and glideslope, followed by poor airspeed management in the latter stages of the approach resulting in high flare angles and tail strike damage.

Significantly these, and many similar events, often involved highly experienced crew and occurred shortly before the handling pilot was scheduled to undertake a licence or operational proficiency check in the simulator. This gives support to earlier evidence that flight crews confidence in their manual flying ability is diminishing and that some feel they need to practise manual flight, even in sub-optimal conditions (i.e. following an un-briefed reversion to raw data), in order to perform successfully during their proficiency check. Ironically, in these cases the roles of the aircraft and simulator appear to have been reversed, with crews practising in the aircraft in order to perform well in the simulator. The problem seems to be exacerbated by a general decline in the number of simulated training hours made available to crews as airlines are pressured to reduce costs, and face the need to focus the remaining hours on more dominant automation related issues.

The UK CAA recently issued a Flight Operations Department Communication to Aircrew (FODCOM 24/2004) which highlights their concern over this practise, encouraging crews to participate in manual flying but urging that it is conducted in appropriate circumstances and is properly planned and briefed for.

These individual accident reports support the broader anecdotal accounts and subjective evidence (e.g. Curry, 1985; Wiener et al., 1999; Wiener, 1988) of manual flying skill deterioration. It must also be considered that all of the statistical data and evidence evaluated in this section were the subject of aircraft accident reports. Accidents are the most severe consequences of errors, where significant damage to the aircraft, its occupants or other property occurs, and fortunately they are relatively low in frequency. They are the result of the perfect alignment of many latent and active failures. The Heinrich ratio (1959) suggests that for every fatal accident there are 29 less severe accidents and as many as 300 near misses, many of which may go unreported. In summary, the relatively severe manual handling events highlighted in these studies may only represent the tip of a much greater issue.

1.2.5 Evidence of Manual Flying Skill Loss from Experimental Work

Surprisingly, given the amount of the aforementioned subjective data and anecdotal evidence (Wood, 2004), very little objective experimental work has been conducted to evaluate the loss of manual flying skills concern. Arthur, Bennet, Stanush and McNelly (1998) compiled a meta review of generic skill decay research finding that, whilst in general all skills will fade without sufficient frequency or quality of practice, complex, open-loop, predominantly cognitive based skills are likely to decay more rapidly than simple, closed-loop, predominantly psychomotor based skills. However, only two studies have been conducted in the aviation domain and they did not form part of this meta review. It has already been noted that many pilots report a loss of confidence in their manual flying ability and it is possible that this in itself may affect meta-cognitive processes, such as the focusing of attention or management of capacity. There is insufficient evidence in the literature to confidently predict how the effect of deteriorated confidence or other emotional factors brought about through inexperience may impact the performance of manual flying skill. However, a study of young pilots (Terelak,

1993) found that increased levels of anxiety generally had a negative impact on the learning of psychomotor flying skills.

In contrast to the findings of Arthur et al. (1998), Viellette (1995) compared the manual handling performance of pilots operating electromechanical (with little automation) and EFIS equipped (with sophisticated automation) variants of the same basic jet transport aircraft. The study evaluated performance on a course tracking and instrument approach task in a full flight simulator. It was found that the group operating the EFIS equipped variant demonstrated significantly lower tracking performance in a number of dimensions compared to the group operating the traditional electromechanical variant.

Whilst the results of this trial make a valuable contribution there are a number of limitations in its method acknowledged by the researchers. Primarily, the RMSE performance metric employed is relatively insensitive (see Hubbard, 1987) and much progress has been made in the development of performance metrics and multivariate statistical techniques since this study was performed. Also no detailed data of the individual pilot's operating experience was collected and the unlikely assumption is made that all pilots in the EFIS group have the same level of recent manual handling experiences and that other differences in the pilot's career background did not significantly influence their performance. However, the researchers noted that they found considerable variation in performance in the EFIS group suggesting that individual differences in career background and automation exposure may have influenced performance. It was suggested that future studies explore the contribution of these factors. Furthermore, given the shift in pilot demographic since the research was performed, the utility of the results in the modern environment are questionable.

More recently Young, Fanjoy and Suckow (2006) evaluated the manual handling performance of a cross section of pilots undertaking airline interview simulator checks. Detailed information about the participating pilot's career background and automation exposure was collected allowing for it to be correlated with performance, along with a survey of the pilot's instrument procedures knowledge and self-reported scan proficiency. There was some evidence from the study that those pilots who placed greater faith in the automatics had weaker manual flying skills. However performance was evaluated using a subjective observer rating system for which no validation of reliability was provided. Consequently few statistically significant results were presented.

1.2.6 Adapting Training

Recent changes in regulations allow airlines to be more flexible with their training, modifying their programmes to be relevant to their particular operation, rather than to meet generalised criteria. This scheme is know as the Alternative Training & Qualification Programme (ATQP). However, in order to justify such modifications the airline must put forward a robust safety case based on strong objective evidence. The training and assessment of manual flying skills may be an area that could benefit from modification under ATQP in future training programmes for highly automated airliners. Unfortunately, as noted in a CAA review of Flight Crew Reliance on Automation (Wood, 2004) and affirmed in this review, there is currently a lack of objective data to demonstrate the postulated decline in manual flying ability, nor a definitive method with which to provide these data.

1.3 Scope of Research

The research focussed exclusively on commercially operated jet transport aircraft equipped with highly automated flight decks. For the purpose of this research the term ‘highly automated’ is considered to indicate a flight deck equipped with auto-pilot, auto-throttle, flight-director, flight management system and electronic flight information system (including, as a minimum, an electronic attitude directional indicator (EADI) and electronic horizontal situation indicator (EHSI) or equivalent). The research focused on the primary flight control and management tasks and did not consider secondary manual tasks, such as aircraft’s system control. The research is orientated towards air transport operations in developed nations.

1.4 Research Objectives

In light of the literature discussed in this chapter the following research objectives were defined.

1. Clarify the definition of manual flying skills and evaluate the cognitive mechanisms of manual flight.

2. Determine the most appropriate means of objectively measuring manual flying skill proficiency.

3. Assess the manual flying performance of a broad sample of pilots operating highly automated aircraft on a valid and relevant manual handling task.

4. Evaluate the effects of differences in the career background and recent manual handling exposure of the pilots on their manual flying performance.

1.5 Structure of the Thesis

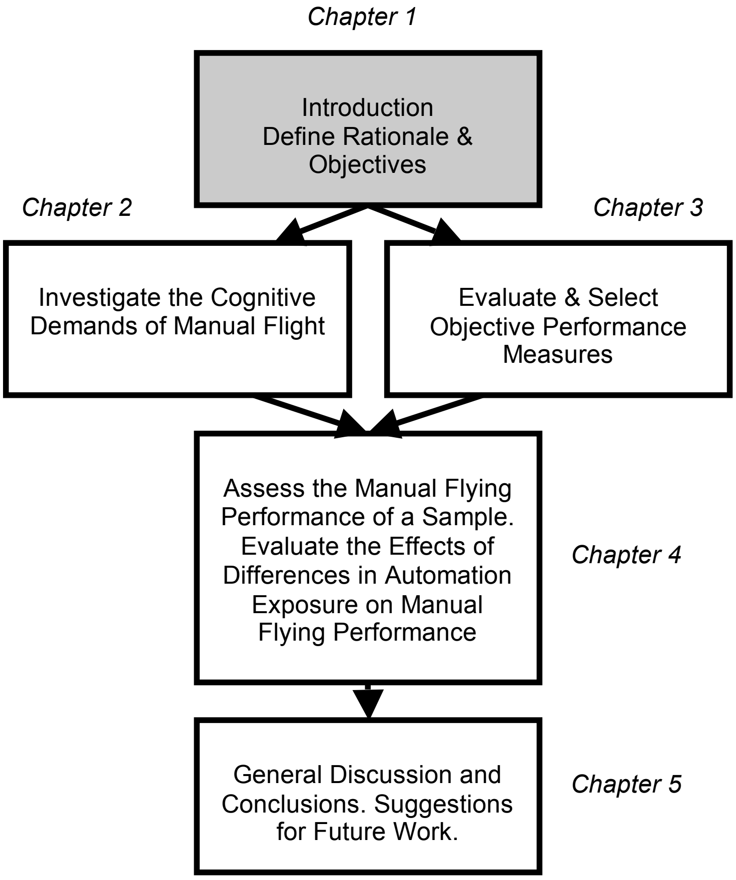
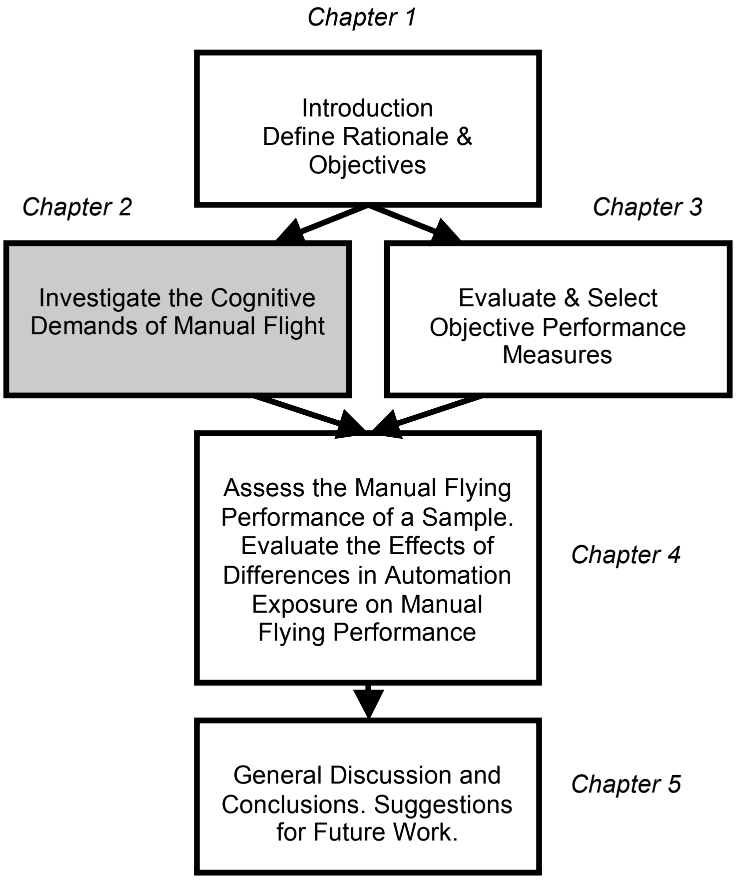


Figure 3 – Structure of the Thesis

The diagram in Figure 2 is presented as an overview of the Thesis structure and to indicate where each research objective is fulfilled. The shaded element indicates the reader's current position in the Thesis. The diagram is reproduced at the beginning of every major section.

Chapter 2

Study I: Cognitive Manual Flying Skills



2.1 Introduction

Manual flight is generally defined as a condition whereby the pilots operate the aircraft without the support of its primary automatic systems and exert control by manipulating the primary flying controls i.e. inceptor, rudder pedals and thrust levers (Parasuraman, Sheridan and Wickens, 2000). This leads to a distinctive and quantifiable difference in the physical skill requirements of manual versus automatic flight, with the former clearly requiring far more sophisticated psycho-motor ability. This overt characterisation of manual flight can lead to the assumption that the only skills threatened by 'out of the loop unfamiliarity' (Wickens, 2000) are motor skills.

However, automation is also designed to support the pilot's cognitive functions (Sarter, Wickens, Mumaw, Kimball, Marsh, Nikolic and Xu 2003). Sophisticated instrument displays which present highly processed information via systems such as the flight director and flight management system all work to relieve the pilot's information processing requirements and support decision making. Ideally the pilots should remain cognitively engaged in order to cross check the automation using raw data sources. However the exceptional reliability of modern automatics, the complexity of navigational procedures and other operational pressures may reduce the level of cognitive engagement (Parasuraman et. al., 1993; Wood, 2004). There is therefore potential that the cognitive skills required for manual flight may be redundant during automatic operation and that these skills may also be vulnerable to decay. The purpose of this chapter is to better understand the cognitive processes that underpin and shape manual handling proficiency.

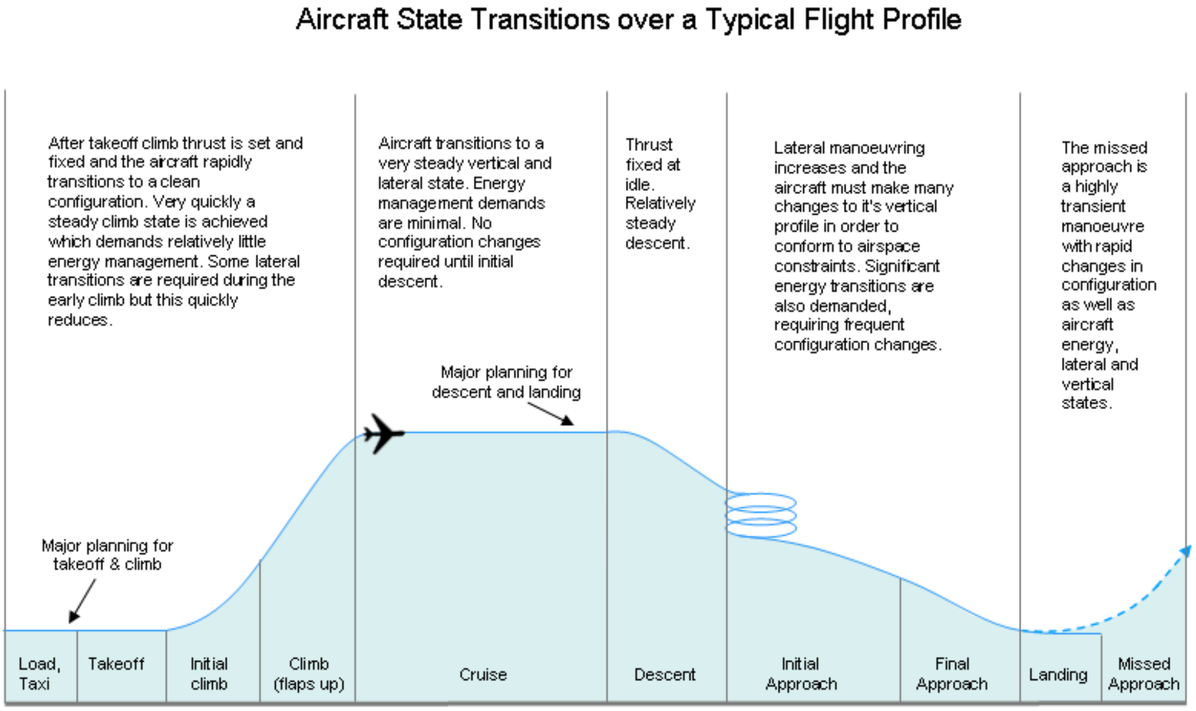


Figure 4 – Aircraft state transitions over a typical flight profile

Figure 4 presents an overview of the aircraft state transitions which occur during a typical flight. Whilst the aircraft is in a steady state the complexity of the cognitive problem is reduced and the pilot can employ a relatively simple mental model to determine how the aircraft will respond and what information should be monitored. Transitory states however are far more complex since the pilot's mental model must incorporate two aircraft states and the risk of it failing is correspondingly increased (see later sections for a more detailed discussion of mental models and aircraft control).

The diagram shows that during the initial climb the aircraft rapidly enters a relatively steady climbing state. There is typically no need to modify the aircrafts thrust or configuration during this phase. There may be an occasional requirement to stop the climb and adopt level flight due to airspace requirements but generally all climb segments will be flown with the same fixed thrust, with the pilot seeking to maintain steady climb airspeed. Without automatic assistance the cognitive burden of the pilot is focused on the lateral navigation of the aircraft. The pilot must scan and integrate the primary flight and navigational data to build a horizontal mental model and determine the aircrafts current horizontal position, it's future horizontal position, how that relates to the horizontal goals and restrictions, and what course adjustments may be required. The cruise phase is also highly steady with configuration and thrust essentially fixed, and with minimal requirement for changes in the aircrafts altitude or course.

By contrast the descent, approach, landing and missed approach (if executed) phases are highly transient. Lateral manoeuvring increases significantly and the aircraft must make significant energy changes, adopting varying configurations and descent profiles, in order to conform to a variety of airspace and traffic constraints. The pilot's mental model must therefore incorporate a multitude of aircraft states and be highly dynamic, increasing the chances of cognitive failures occurring. This latter phase of flight therefore appears, from a high level, to be the more cognitively challenging and prone to error, supporting the data which show it to generate a greater number of accidents (see section 1.2.4).

2.1.1 Models of Human-Aircraft Control

Manual aircraft control requires the pilot to employ both open-loop and closed-loop control behaviour (Baron, 1988). Open-loop behaviour is independent of feedback (i.e. a golfer driving a ball) and involves the execution of pre-programmed motor schema to effect large changes in the aircrafts orientation, path or location. Closed-loop control is used to track and maintain a target state by monitoring feedback channels. In this control mode the pilot monitors and adjusts their performance in order to reduce any discrepancy between the desired aircraft state and the observed aircraft state. This is often termed ‘pursuit tracking’. Feedback is delivered primarily via the flight instrumentation and outside field of view although vestibular, somatic, proprioceptive and auditory cues are also utilised. The continuous closed-loop control requirement of manual flight is therefore highly demanding of the pilot’s physical and cognitive capacity.

Human factors research has developed process control models to describe in more detail how pilots achieve manual aircraft control. The series model (McRuer, 1982) presented in figure 3 is widely cited and demonstrates the hierarchical nature of the control process. For example, whilst the pilot ultimately wishes to satisfy high level flight goals, such as flying level at 6,000ft, the control system only allows for direct manipulation of the aircraft’s basic six degrees of freedom i.e. body attitude and translational rates. Consequently the pilot must manipulate lower order parameters (e.g. attitude, airspeed etc.) in order to satisfy higher order goals (e.g. altitude, path etc.). The pilot must close several control loops concurrently. In the series model these control loops are shown nested within each other. The ‘inner’ attitude control loop is closed in order to satisfy the ‘outer’ flight path control loop. Owing to the control system design employed in large transport aircraft there are generally significant time lags between the pilot making a control input and the occurrence of an effect in an outer loop parameter i.e. a control wheel input causing a lateral displacement of the aircrafts position (this will be discussed further in chapter 3). Consequently such aircraft require the pilot to anticipate the aircrafts likely and desired flight path in order to achieve effective control. The associated mental projection is highly demanding of cognitive resources (Moray, 1999).

Unfortunately models such as the series model of control lack detail of the higher level cognitive processes which underpin them. For example, in the model there are two confluence points where feedback of the observed aircraft state is compared to the desired aircraft state in order to determine their relative error and to select a future course of action. However, the cognitive skills and strategies which are necessary to generate the requisite data and perform this ‘black box’ function are not detailed. Also, the desired flight path is specified as an input to the model. In automatic operation this information would usually be provided by the flight management system. However in manual flight there is a substantial degree of cognitive processing which must be undertaken to derive the required goals. Again the processes by which this is achieved are not detailed in these models yet form a crucial component of the ‘skill’ of manual flying.

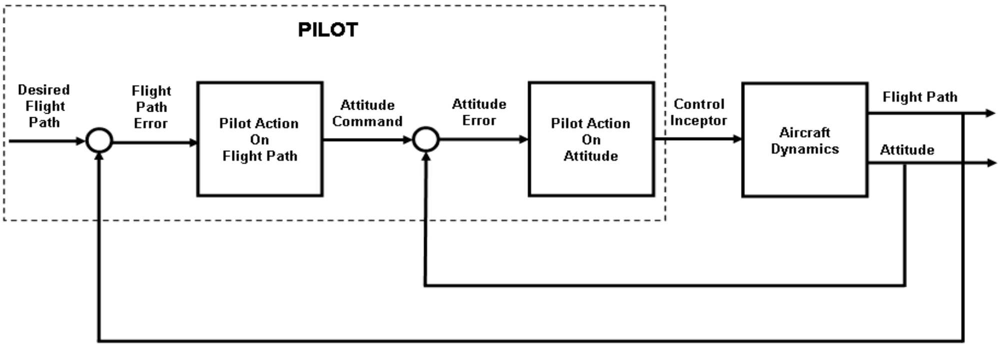


Figure 5 – The series model of pilot control showing the nested inner and outer control loops (adapted from McRuer, 1982)

2.1.2 Mental Models

It is theorised that for a human to have effective control over any process they must possess a mental model of that system (Moray, 1990; Sarter et al., 2003). A mental model is a user’s memory of the structure of a system which may be used to simulate how it will respond to a control input or environmental change (Delzell, Johnson and Liao 1998; Hegarty, 2004; Klein and Crandall, 1999). The quality of a pilot's mental model is a key determinant of their flying performance since it guides their attention and allows them to prioritise tasks. The development of strength in the mental models structure is related to experience and any weaknesses in it may be directly related to failures in task performance. Strong mental models are critical under conditions of stress since performers will tend to revert to familiar routines and those most easily recovered from memory, which are generally one and the same.

Humans generally find it simpler to control simple linear systems but higher order (2nd and 3rd) rates of change with significant time lags are characteristic of translational phases of flight. To accomplish control the pilot must be able to think ahead of the aircraft. Accordingly, the mental model acts as a mechanism of mental projection and enables anticipatory control of a system. This is congruent with Endsley's (2006) definition which states that the highest levels of situational awareness are achieved when the controller can anticipate the future state of the system. Mental models therefore play an important role in problem solving, judgement, decision making and planning for the pilot. They are simplifications of the real system which rely on abstraction and mental rules of thumb (heuristics). Carley and Palmquist (1992) propose that expertise is signified by more efficient mental model structures which are simplified in some areas, where unnecessary system complexities are removed in favour of generalised rules, and expanded in others, where more detailed system knowledge is beneficial. Flach and Jaques (2003) published work which showed how a pilot's mental model of the inner control loop problem was refined with expertise. It was shown that on a precision approach pursuit tracking task experienced pilots effectively 'lock out' several of the aircrafts degrees of freedom in order to simplify the control problem. For example, controlling the descent profile by 'fixing and forgetting' the aircrafts thrust setting and manipulating just the pitch attitude.

There is large body of research which has examined the cognitive demands of using automation on the flight deck (Harris, Hancock, Arthur and Caird 1995; Holder and Hutchins, 2001; Sarter, Woods and Billings, 1997; Sarter et al., 2003). However very little objective research has been undertaken to establish the basic cognitive mechanisms and mental models which pilots use to operate large transport aircraft manually, and how these mechanisms and models may be subject to decay (Childs and Spears, 1986). Accordingly a study was undertaken using objective cognitive task analysis techniques to audit the cognitive processes employed by pilots in a challenging manual flight scenario. The results of this study were intended to inform further stages of the research programme and to ensure that the scenarios used to elicit and measure manual handling ability were valid from a cognitive perspective.

2.2 Methodology

2.2.1 Cognitive Task Analysis

It is often desirable to fragment a task into a structure of smaller subtasks so that a better understanding of the underlying processes can be formed. The decomposition can be performed to guide training and workspace design, to identify potential errors in task execution or simply understand how experts perform the task (Kirwan, 1992). Such techniques are given the general label of task analysis. Traditionally task analysis has been conducted in the behavioural domain, whereby analysts explore the observable actions which are performed by the operator and describe the rules that determine when those actions are performed. An example of this technique is Hierarchical Task Analysis (HTA). Using this technique actions are grouped in a very tight structure, each forming the goal for a subordinate network of actions. Traditional task analysis techniques build detailed pictures of the physical aspects of task performance but offer very little insight of the cognitive functions underpinning that performance.

As many aspects of work, such as piloting modern aircraft, have become less physical and more information processing orientated there has been an increased emphasis in understanding the cognitive elements of task performance (Flach and Jaques, 2003). Correspondingly task analysis techniques have developed to encompass the cognitive domain. The field of cognitive task analysis is still relatively immature and although a variety of methodologies are described in the literature, generally they offer more of a guiding philosophy than a prescriptive set of rules and procedures.

However, the Applied Cognitive Task Analysis (ACTA) methodology (see Militello, Hutton, Pliske, Knight, Kline and Randel, 1997) offers a prescriptive and streamlined approach to the data collection and analysis procedure. Although originally conceived to guide industry practitioners with a limited research background it has found popularity in the applied research field in diverse domains such as fire fighting, naval radar operation and aviation (see Latorella, Pliske, Hutton and Chrenka, 2001). The ACTA technique was adopted for this study because it is was supported by detailed instructional material, yields quality information of cognitive demands with minimal resource requirements and has been well demonstrated in literature. The methodology consists of three complementary data collection phases 1) Task diagram phase 2) Knowledge audit phase and 3) Simulation interview. These are structured around one on one interview sessions with subject matter experts (SME).

The following sections describe how the ACTA methodology was applied to elicit the critical cognitive components of the manual flying task during the approach to landing phase. The approach phase was chosen as it involves some of the more complex flight path and energy management tasks and is the most likely phase for manual flight to be undertaken in an operational environment. Additionally the findings of this research are intended to compliment those of Flach et al (op cit.) who also studied the approach to landing phase.

2.2.2 Participants

The ACTA process requires highly skilled subject matter experts as its data source. For this study expert flight crew were assumed to be, at a minimum, command qualified. Furthermore, senior captains who had been appointed to training duties or had a background in test flying were actively sought. These experts were more likely to have developed the self analytical skills necessary to report their own behaviour and cognition to the analyst. Multiple SMEs were interviewed to provide reliability and validity, overcome individual biases, provide redundancy protection against poor quality interviews and ensure the full breadth of the task was examined.

Interviews were conducted in quiet, private rooms either at the participant's place of work or at Cranfield University. Sessions were conducted with only the researcher and the SME present and, with permission, they were audio recorded. British Psychological Society ethical conduct standards were adhered to throughout the course of the research and all data was de-identified.

A total of nine participants contributed over 13 hours and 80,000 words of transcribed interview data. All were senior captains, with four holding the chief pilot position and two having a training role. Average flying experience was approximately 10,500 hours. The aircraft types currently operated by the participants were the MD-11, B747-200, B747-400, B767, B757, B737 and A320. The participants also had considerable experience on other commercial types including the B777, DC-10, B747-300, A310, RJ100, Gulfstream 100, Learjet 25/35, ATR42/72 and older generation types including the B707, HS Trident, HS 748, BAC 1-11, Bristol Britannia and Bristol Freighter.

2.2.3 Phase One – Introductions and Task Diagram

The opening phase of the interview involved the collection of biographical data using a pro-forma and a semi-structured discussion around the potential decay of manual flying skills and influencing factors (see interview protocol in appendix A). The interviewer presented working definitions of 'automatic flight', 'manual flight', 'approach phase' and 'cognitive skill' so that these would be consistent through the research sessions.

In line with the ACTA methodology a 'task diagram' was elicited from the SME. The task diagram is a simple representation of how the SME conceptually structures the task of interest. The researcher asked the SME to think about how they manually control the aircraft during the approach to landing phase and to list between four and eight steps that fully described that process. The researcher recorded the steps on a flip chart visible to both parties. The SME was then asked to indicate which of the tasks steps involved challenging cognitive elements and these steps were highlighted accordingly.

Ultimately a simple conceptual map of the task was produced representing the structure adopted by the SME and indicating where all the cognitively demanding elements were concentrated. This diagram served as a guiding framework for the remainder of the interview session.

2.2.4 Phase Two – Knowledge Audit

The knowledge audit phase of the interview identified the specific areas of cognitive expertise which enable experts to deliver superior performance on the manual approach to landing task. For each highlighted step in the task diagram the interviewer administered a series of scripted probe questions to elicit examples of cognition in areas such as situational assessment, diagnosis, prioritisation, self assessment etc. The full interview schedule including cognitive probe questions can be found at Appendix A.

The ACTA process labels the elicited examples as cognitive demands. The cognitive demands were listed down the first column of a table printed on a flip chart. For each cognitive demand the analyst enquired as to why this cognitive demand was challenging, what cues and strategies may be employed to satisfy it and what errors a novice may make, recording this information in corresponding columns of the table structure. The resulting tabulated data sets are known as cognitive demand tables and are reproduced fully in Appendix B.

2.2.5 Phase Three – Simulation Interview

During the final research phase the SME performed a simulation of the manual approach and landing task and then in a subsequent interview was asked to describe how and why they did what they did. This technique is used to provide an understanding of the expert’s problem solving processes in context and can uncover details not revealed by the knowledge audit. Although high fidelity simulations clearly offer rich environments for analysis, past studies have shown that less resource intensive, low fidelity simulations are capable of yielding information of equal quality and validity in the cognitive task analysis setting (Latorella et al., 2001). The ACTA suggested format of a paper simulation exercise was adopted for this study since it required minimal resources to develop and deliver and was easily portable to the various interview locations. A paper simulation is a written walkthrough of a task scenario where several perturbing events are introduced. The participant is required to read through the simulation and think about how they may react to or resolve the presenting problems.

The paper simulation described a reasonably challenging but nonetheless likely approach scenario to be conducted under manual control, divided into sequential segments (Appendix C). In each segment a problem was introduced, including ATC re-routes, energy management challenges and environmental condition changes. The scenario was developed with reference to the Flight Safety Foundations (FSF) Approach and Landing Accident Reduction (ALAR) toolkit (2000) which lists key factors which lead to approach and landing incidents. The SME was instructed to read through the scenario, pausing after each segment to consider what their thoughts, decisions, judgements or actions may be, if any, at that stage. Printed approach charts and flight briefing material specific to the scenario were provided.

The analyst asked the expert to verbally ‘walk through’ the scenario and to draw attention to any pertinent events, decision points or judgements that occurred. The analyst then asked the expert to detail the actions, salient cues and potential errors associated with each critical point in turn. This information was added to the cognitive demand tables elicited during the knowledge audit.

2.2.6 Treatment of Data

The task diagram statements and cognitive demands tables from the interview sessions were transposed into Microsoft Excel worksheets. The interview audio recordings were transcribed to complement the written notes taken during the sessions, producing over 85,500 words of data. Repeated passes of the audio recordings were made to check for transcription accuracy, and identify intonation subtleties, inference, emphasis and possible misleading statements. For each participant in turn the transcript and written notes were scanned to identify the individual cognitive demands brought to attention during the session and to uncover any which were passed over during the interviews. Any newly identified cognitive demands were added to the ACTA prescribed cognitive demands tables.

Where there was uncertainty of meaning in the transcript or written notes the audio recording was consulted. If the meaning was still questionable a second independent researcher experienced in cognitive task analysis and familiar with the aviation domain was consulted. If the meaning could not be agreed the statement was excluded from the analysis.

2.3 Results and Discussion

The results of this study demonstrate that manual flying proficiency is heavily dependent upon many cognitive processes which are potentially redundant during automated flight. In particular, cognitive resource intensive mental arithmetic operations are central to the execution of the aircraft/environment model which is used to anticipate control requirements and to manage the aircrafts energy efficiently. Expert pilots report employing heuristics to ease the cognitive burden of the manual flying task and identify typical shortcomings in the mental model structure of less able pilots.

2.3.1 Task Diagram

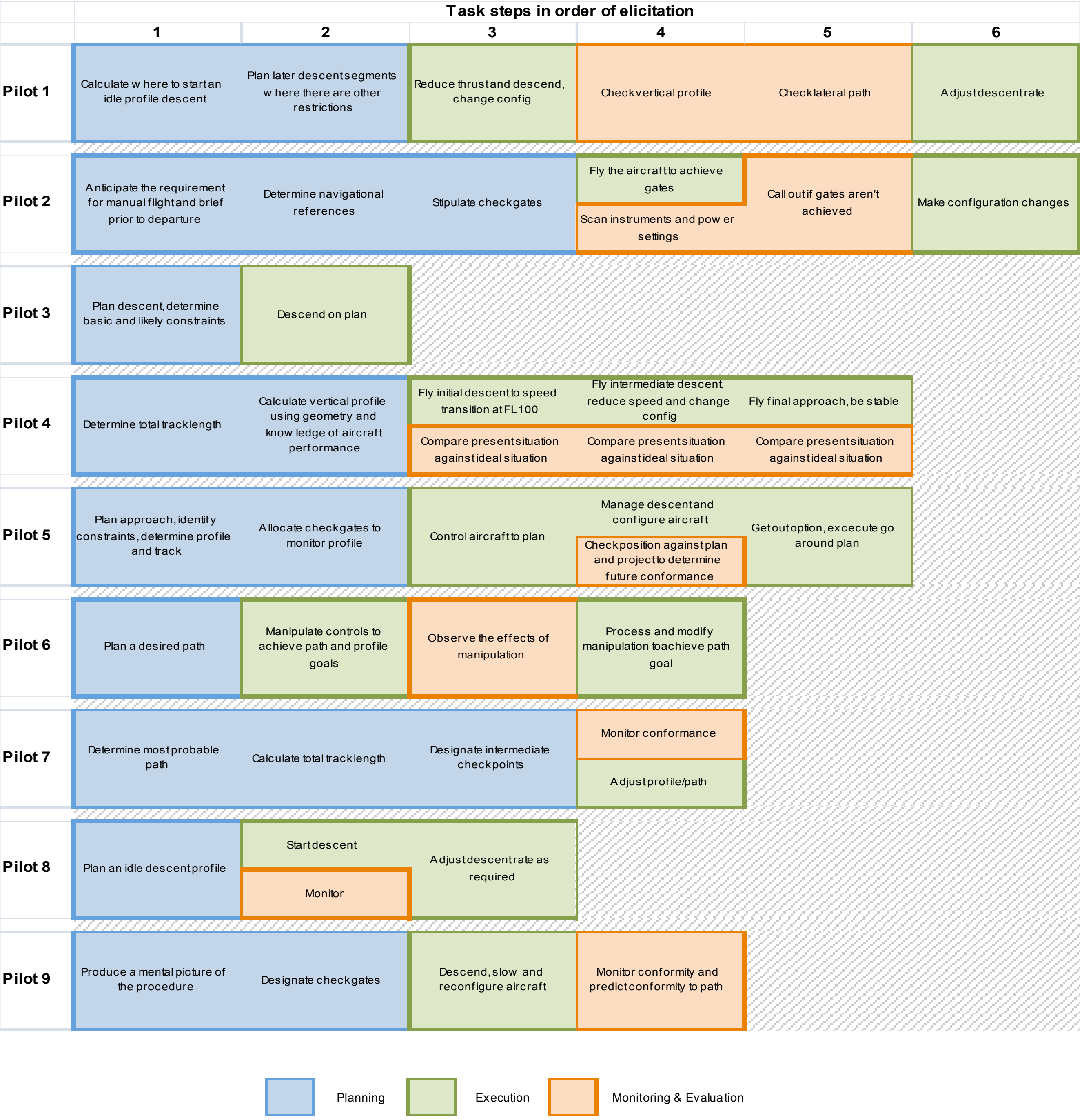
The task diagram phase identified how the individual SMEs conceptually structured the manual approach to landing task. To compare and contrast the representations of each SME the task diagram steps were assembled into a common figure (see figure 4).

A simple coding scheme was superimposed upon the task diagram data. Task steps were categorised as planning, execution, or monitoring & evaluation activities. Task steps which simultaneously described two different categories of activity were subdivided. Conversely, neighbouring steps which described the same category of task were 'band-boxed' together. Again, a second analyst experienced in CTA and the aviation domain was asked to perform the same coding process in parallel and independently. Where the assigned coding category differed between analysts a third, similarly experienced, analyst was asked to code the case. Ultimately if a consensus could not be reached the data was excluded from the analysis.

It is notable that the SME's were consistent in the way they conceptualised the manual approach to landing task and that in general they described the task using a simple closed-loop structure (i.e. plan → execute → monitor → adjust → plan) as per the process control model presented in the introduction to this chapter.

2.3.2 Cognitive Process

A total of 63 unique cognitive demand descriptions were elicited across all participants for the manual approach to landing task. Full reproductions of the cognitive demands tables are given at Appendix B. The data were coded into ten emergent categories and as before this template was presented to two other analysts for comparative coding. Again, where a case could not be coded it was excluded from the analysis. These categories and the frequency of their occurrence in the data are presented in figure 5. The majority of the cognitive demands were associated with the vertical profile and energy management aspects of the approach task rather than the lateral aspect, suggesting that the former is a more cognitively complex activity.



Task steps in order of elicitation

	1	2	3	4	5	6
Pilot 1	Calculate where to start an idle profile descent	Plan later descent segments where there are other restrictions	Reduce thrust and descend, change config	Check vertical profile	Check lateral path	Adjust descent rate
Pilot 2	Anticipate the requirement for manual flight and brief prior to departure	Determine navigational references	Stipulate checkgates	Fly the aircraft to achieve gates Scan instruments and power settings	Call out if gates aren't achieved	Make configuration changes
Pilot 3	Plan descent, determine basic and likely constraints	Descend on plan				
Pilot 4	Determine total track length	Calculate vertical profile using geometry and knowledge of aircraft performance	Fly initial descent to speed transition at FL100 Compare present situation against ideal situation	Fly intermediate descent, reduce speed and change config Compare present situation against ideal situation	Fly final approach, be stable Compare present situation against ideal situation	
Pilot 5	Plan approach, identify constraints, determine profile and track	Allocate checkgates to monitor profile	Control aircraft to plan	Manage descent and configure aircraft Check position against plan and project to determine future conformance	Get out option, excecute go around plan	
Pilot 6	Plan a desired path	Manipulate controls to achieve path and profile goals	Observe the effects of manipulation	Process and modify manipulation to achieve path goal		
Pilot 7	Determine most probable path	Calculate total track length	Designate intermediate checkpoints	Monitor conformance Adjust profile/path		
Pilot 8	Plan an idle descent profile	Start descent Monitor	Adjust descent rate as required			
Pilot 9	Produce a mental picture of the procedure	Designate checkgates	Descend, slow and reconfigure aircraft	Monitor conformity and predict conformity to path		

🟦 Planning 🟩 Execution 🟧 Monitoring & Evaluation

Figure 6 – A composition of each subject matter expert's task diagram. indicating how they conceptually structured the manual approach to landing task at a high level. Task steps are coded into planning, execution or monitoring & evaluation activities.

The detailed cognitive demands, coupled to the structure of the task diagrams, gave a strong description of the cognitive process by which the pilots operate the aircraft, together with strategies and heuristics which are employed by experts to minimise the cognitive burden and achieve superior performance. The following account of the process is derived from that data.

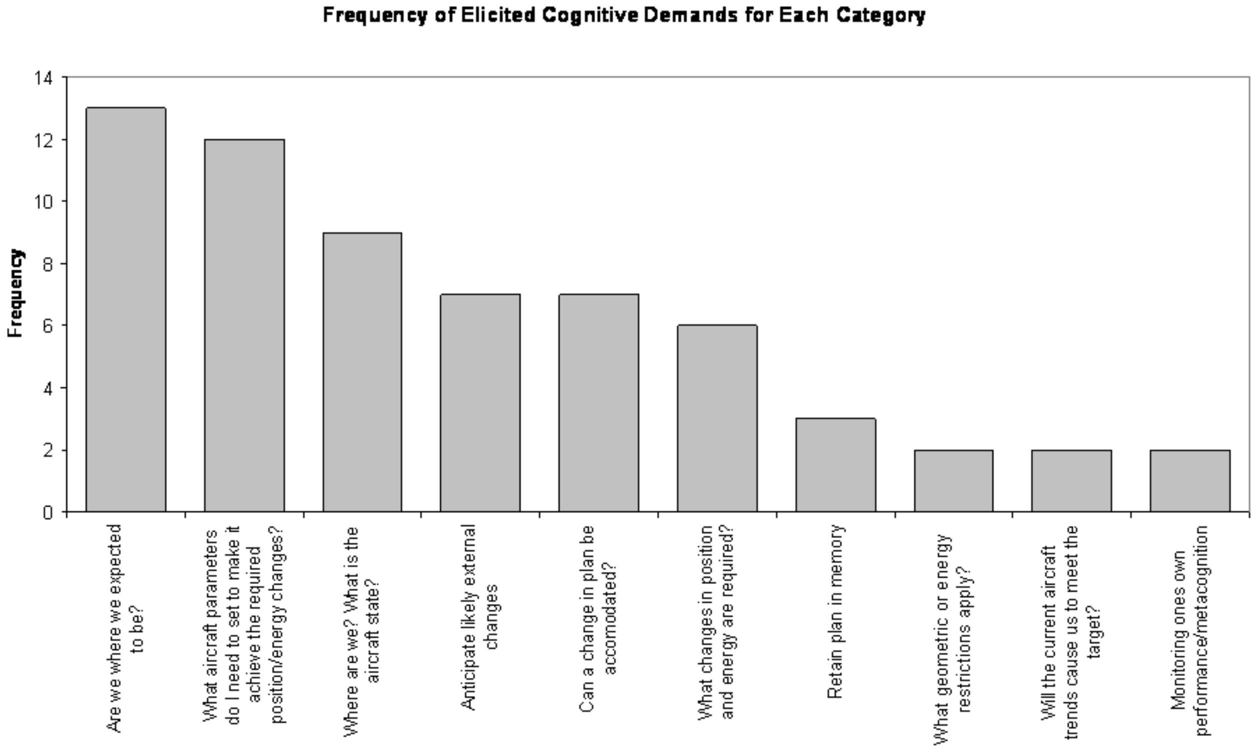


Figure 7 - Cognitive demands categories and their frequency of occurrence

The bulk of the cognitive demands related to the initial approach planning phase, prior to the top of descent. However, the SME reported re-visiting parts of this planning activity during execution of the approach to accommodate changes or rectify errors. Initial activity focussed on identifying the applicable geometric and energy constraints such as crossing restrictions, minimum safety altitudes, airspeed restrictions etc. and selecting suitable navigation aids. This included consulting their operating experience to anticipate any likely ATC restrictions or routings that were idiosyncratic to that airfield or approach.

Primarily the SME would attempt to fit a clean idle descent profile (typically using a 3nm per 1,000ft of altitude model) anchored to the final approach fix back through the geometric and energy restriction (establishing the top of descent (TOD) point). However, where restrictions violated this profile the SME undertook a mental modelling and simulation process to determine the most efficient means of satisfying the constraints. This involved computation of the speed, track length and altitude exchanges required to meet the gate and fitting this information against the SMEs mental model of the aircrafts performance in the current environment. The reported mental models were essentially a series of lookup values and simple algorithms which related track length, ground speed, altitude and time for various aircraft configurations and environmental conditions (wind speed was generally factored into the aircraft performance model via the groundspeed to airspeed relationship).

The SME relied heavily on heuristics to simplify the execution of the model. For example “a typical descent at 210kts from 7,000ft requires 23 track miles”, “lead in distance for a standard rate 90 degree turn is groundspeed over 100”, “add 1 extra track mile per 3,000ft of descent when using engine ant-ice” etc.. The SME would also ‘lock out’ degrees of freedom in their calculation by making generalisations, i.e. average terminal manoeuvring speed will be 180kts which equates to about 3 miles every minute, rather than making complex calculations for several segments at different speeds. This supports the work of Flach and Jaques (2003) discussed in the introduction to this section.

The process of running the aircraft/environment model relies on the development of efficient estimation techniques and again draws on mental arithmetic operations which significantly burden both working and long term memory (Baddeley, 1986; Beilock, Kulp, Holt and Carr, 2004). However the mental rules of thumb and simplifications outlined here by the SME bypass most of the mental arithmetic, reducing processing requirements and freeing up capacity for other flying tasks. The expert pilots indicated that those inexperienced in manual flight would often exhibit poor performance because they either became saturated by the modelling process, generated inaccurate information or would bypass the process entirely, using a ‘one size fits all’ approach and ‘relying on luck for it to work out’. It was suggested by several SME that a lack of manual flying experience and reliance on highly processed information from the flight management system can prevent the formation of heuristics, efficient mental strategies and in general the optimisation of the mental model structure.

> *“(Poorly performing manual pilots) haven’t got enough information yet hardwired into their brain to accurately predict what the aircraft’s going to do next”*
>
> (Excerpt from SME interview)

The SME also indicated that constructing a mental simulation of the approach helped to retain the specific details in memory and reduced the need to make notes or refer back to the approach charts. These observations are in line with the suggestions of mental model theory cited in the introduction (Hutchings, 1995) and more general theory that information is held better in memory following some degree of processing.

When a satisfactory approach had been constructed the SME again reported using mental lookup tables to execute the plan. Long term memory stored key attitude and thrust reference values that would cause the aircraft to achieve the planned translational rates in its current configuration. Again, SME comment suggested that dependence upon the flight director appears to reduce the formation of this ‘lookup data’ and many less manually experienced pilots ‘hunt around’ to find appropriate attitudes or thrust reference values.

> *“It would seem that people just don’t know where to put the nose, (they) haven’t got a sort of safe number in their mind and don’t know what to do with the attitude of the aeroplane”*
>
> (Excerpt from SME interview)

The most significant cognitive demands (judged by the frequency of elicitation) related to the monitoring of the aircraft’s position and energy against the planned approach. Principally this involved establishing the aircrafts current position and energy state. The SME pilots reported the importance of considering the efficiency of the scan pattern and adapting it to the informational requirements of that stage of flight. They considered that novices would often attend to irrelevant information, using a fixed scan pattern which made inefficient use of their already limited capacity (Damos, John and Lyall, 1999). Determining the aircrafts position from raw data also required some mental arithmetic operations but was made less complex if thought was given to the selection of appropriate navigation aids.

> *"You can make a mental calculation and say this is or this isn't going to work, and you can say to the (ATC) 'that's not enough miles, can I have another 6 or 7' "*
>
> (Excerpt from SME interview)

Comparing the aircrafts current position against the planned approach was accomplished through two mechanisms, both of which demanded resources for mental arithmetic and drew upon the aircraft/environment mental model. In one mode the SME reported interpolating the planned flight path geometry to produced discreet spatial check gates which could be compared against the computed aircraft position i.e. "the profile puts us at 3,000ft 3 miles from the fix, so we should be at 4000ft 6 miles from the fix". In a second mechanism the SME would extrapolate the current aircraft trends and establish if they were likely to intersect with the next major gate as planned. Discrepancies in profile were usually transformed into track mile error. Once more the expert pilots reported using a mental lookup table which they used to evaluate whether an error in profile could be recovered or not given the environmental conditions and aircraft configuration. Error tolerances usually reduced in magnitude with closing proximity to the airfield. Intolerable profiles errors were resolved by re-visiting the planning phase.

> *"You're always sort of flying in a tube...the diameter of this tube becomes smaller and smaller towards the runway"*
>
> (Excerpt from SME interview)

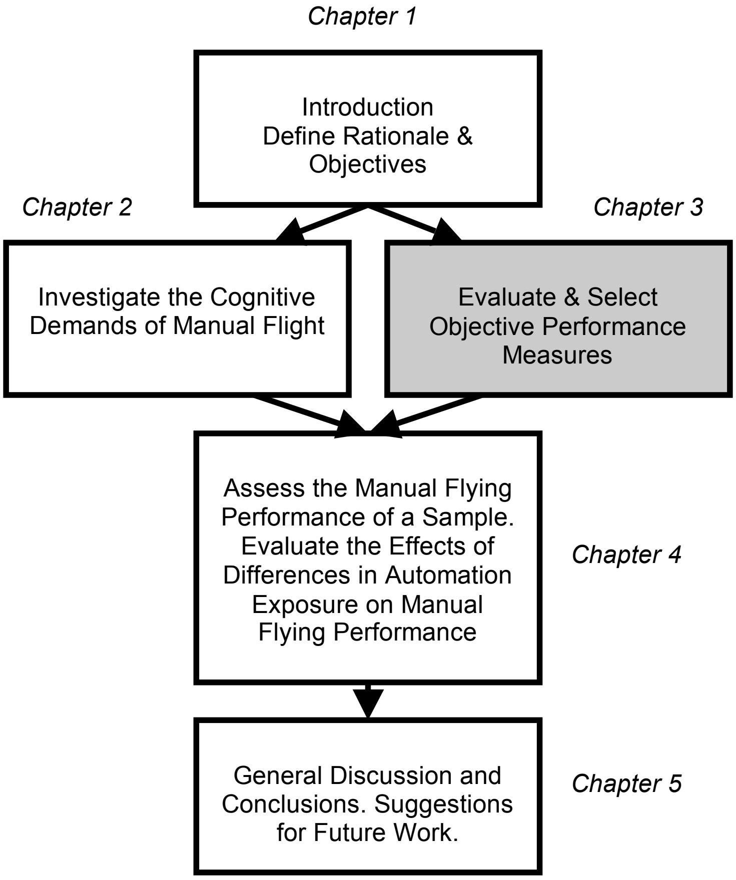
2.4 Chapter Conclusions

The cognitive mechanism outlined in this research should not be surprising to air transport pilots experienced in manual flight. However, it has been elicited using objective techniques and emphasises the central role of a robust aircraft performance mental model in executing almost all of the critical planning and monitoring functions during raw data manual flight. Primarily the management of the vertical profile and aircraft energy is heavily dependent on mental computation and can demand significant working and long term memory capacity if an efficient model, simplified through heuristics, is not available.

The formation of efficient mental models is a function of experience. Without manual flying exposure a pilot's mental model of the performance of their aircraft is unlikely to become optimised and will remain cumbersome and inefficient to use. Correspondingly they will have little spare capacity with which to control the aircraft or perform secondary tasks. It may be that under certain conditions pilots are slower to acquire this knowledge structure, or perhaps learning simply fails to occur at all. Likewise, it is unclear whether once developed this structure can truly be forgotten, or whether it is actually a failure in retrieval which characterises 'skill fade'. Nevertheless, when developing tasks to evaluate the manual handling proficiency of a pilot it is therefore vital to consider the cognitive aspects of this skill. The chosen scenario must challenge the pilot's knowledge and ability to manipulate critical aircraft performance and environmental information.

Chapter 3

Study II: Evaluating and Selecting Manual Flying Performance Measures



3.1 Introduction

A performance measurement technique, whether it is used to select capable job candidates, gauge the efficacy of a training program, or monitor operational safety, must have integrity, assessing individuals consistently against a common interpretable scale. This chapter evaluates the various means of assessing pilot manual flying expertise and selects those which are most suitable for researching the causes of manual flying skill variation.

3.2 Performance Measurement in Flight Operations

In commercial flight operations the assessment of a pilot's manual handling performance is mostly entrusted to the judgement of expert human observers. Principally, line performance assessment occurs during the biannual Operator Proficiency Checks (OPC) and annual Licence Proficiency Checks (LPC) administered by a suitably trained Type Rating Examiner (TRE). This method of assessment has many advantages in the operational environment. Primarily it is practical since the required measurement resources, expert flight crew, already exists within the airline and can be relatively easily administered. Secondly it has a great deal of face validity, drawing upon the human's innate ability to perceive often very subtle differences in performance. Thirdly it is an established practice which is well understood, documented and trusted within the airline environment.

However, human judgement is naturally subjective and the discrimination of individual examiners may be coloured somewhat by factors other than the student's manual handling performance i.e. by aspects of the student's personality or appearance, their performance on a different aspect of the test or by the examiner's mood or circumstance. Whilst tools such as behaviourally anchored rating scales and structured grading criterion (e.g. the tolerances for manual handling performance listed in Standards Document 24 (2005) reproduced at Appendix D) can focus human judgement, observational techniques remain insufficiently objective to form the sole means of performance assessment in a sensitive research study of manual flying ability. Truly consistent, sensitive and scientifically viable assessment can only be achieved through the numerical analysis of flight data records.

Mathematical algorithms which evaluate performance are uninfluenced by human emotion and thus generate consistent and predictable results. In aviation the required performance data on which these techniques operate is relatively abundant, with time series records for hundreds of parameters generated by both the aircraft and simulator (primarily via data feeds to the Flight Data and Quick Access Recorders). Because numerical performance evaluation may be applied post-event and without the presence of human observers it has allowed the process to migrate away from the OPC/LPC environment and into routine line operations, facilitating today's Flight Data Monitoring/Flight Operations Quality Assurance (FDM/FOQA) programmes.

3.2.1 Flight Data Monitoring & Discreet Event Measures

The routine analysis of flight data (i.e. FDM/FOQA) has revolutionised the way in which airlines can assess their operational performance and has hugely benefitted flight safety. It is now either a mandatory or recommended practice in most ICAO member states (for the UK see CAP 739) for aircraft exceeding 27 tonnes in weight. However, because the demands placed on FDM/FOQA systems are significant, with very large volumes of data passing through them every hour of the day, the level of analysis undertaken on the data is relatively shallow and the performance measurement techniques applied are quite simplistic.

A simple means of objectively assessing performance is to register when a flight parameter of interest deviates beyond a specified tolerance threshold for the phase of flight i.e. “altitude deviates more than 100ft from the assigned flight level during the cruise”. This form of measurement is often called ‘event analysis’ since it records the occurrence of a pertinent event in the flight and it forms the basis of most Flight Data Monitoring/Flight Operations Quality Assurance (FDM/FOQA) systems (Chidester, 2003; van Es., 2002). Often fairly complex triggering logic can be used to define the event so that specific flight safety issues are targeted, for example the logic “airspeed greater than 170kts with a flap setting of between 25 and 30 degrees” may be used to detect a flap over-speed event and direct an engineering inspection of the subject aircraft. The limitation of this system is that in order to build the event set the analyst must have some preconception of where operational problems may lie and what tolerances define acceptable flight from unacceptable flight. Typically this is driven by company standard operating procedure (SOP) and engineering limitations.

Layers of events with progressively more conservative tolerances may be specified so that an event severity can be deduced, (i.e. minor deviation, moderate deviation, severe deviation) and the analysis focused. Event severity can be non linear and be dependent on secondary parameters, for example a glideslope deviation of one dot low may only be considered a minor event if it were to occur early on the approach and at high altitude, whilst the same deviation may be given a very high severity if it was detected late on the approach and at low altitude.

The discreet event analysis technique is well suited for detecting occurrences of extremely abnormal performance where the barriers which defend against incidents and accidents have been significantly eroded. Its strength lies in its ability to distil practical information rapidly from very large numbers of flights, as is the requirement of FDM/FOQA systems. However, because event analysis is discreet (i.e. an event is either triggered or it is not) it offers a relatively shallow analysis of the each individual flight. For example, Van Es (2002) reports a hard landing investigation conducted via an airlines flight data monitoring system. From a sample of 8,000 flights only 25 were found to have touch down load factors beyond a specified value that constituted a hard landing. Whilst these eventful flights proved to contain very valuable information about hard landings within the operator, over 99 percent of the sampled flights were discarded. Deeper analysis of the discarded data may reveal more subtle abnormalities (‘near misses’) and provide trending information which would point to the causes of hard landings.

Event type measures are also used to assist performance assessment during the OPC/LPC. CAA Standards Document 24 (2005) gives guidance to examiners assessing manual handling performance during these proficiency checks (Appendix D). Within the document are a set of tolerances which govern heading, track, airspeed, altitude, glideslope and localiser tracking deviation for various flight phases and aircraft conditions. It is suggested that sustained deviation beyond these tolerances by the performer signifies inadequate manual handling proficiency. Although the system is intended to be implemented by human observation, and it is offered mainly to support the examiner’s broader judgement, the guidelines are essentially event measures. However, these measures in isolation do not capture many of the aspects of manual handling ability which are considered to be important by the examiners, such as the smoothness, efficiency, co-ordination and anticipation of control (see chapter 2). If these measures were simply translated into an FDM event set they would offer a very limited and crude assessment of pilot handling performance which would likely be insensitive in a scientific study.

This ‘loss of manual flying skills’ research will sample a much smaller number of pilots and flying hours than is typical of an airlines FDM system. Consequently it is not expected that a great number of serious manual flying deviations will be observed. It is more likely that the research will identify subtle differences in performance which show a general ‘creep’ towards the fringes of acceptability. However, it should be considered that serious performance errors may often be rooted in these subtle deviations due to the way that they impact the subsequent decision making process.

FDM event sets and Standards Document 24 tolerances thus have limited application for the fine grained analysis of manual handling performance that is required of this research study. The following sections look to the research domain for more sensitive measures of performance which would be better suited to this study.

3.3 Performance Measurement in Research

A major challenge faced by performance analysts is to develop numerical metrics which successfully replicate the sensitivity and broadness of human perception. The following sections evaluate the various types of numerical performance metrics which are in use in the applied research field. It is important to remember that typically (although with some exceptions) these metrics are designed to assess a single dimension of performance, such as altitude tracking accuracy or localiser tracking smoothness. To produce a broader assessment of manual flying ability it is necessary to somehow combine a battery of metrics which simultaneously assess the multitude of performance dimensions. Developing weightings and strategies for combining individual metrics into a conglomerate is another significant challenge of numerical performance assessment and ultimately is the route to providing an automated performance assessment tool, the focus of much of the applied research (McDowell, 1978; Rantanen, Johnson and Talleur, 2004).

3.3.1 General Properties of a Performance Metric

From a review of performance assessment literature (e.g. Johnson and Rantanen, 2005; Rantanen, Talleur, Taylor, Bradshaw, Emanuel Jr., Lendrum and Hulin, 2001) it is apparent that certain qualities are associated with any capable performance metric. The attributes presented in table 1 are not domain specific but are certainly applicable to pilot performance measurement.

Table 1 - Attributes of a performance metric

Attribute	Descriptive
Reliable	The metric should give a consistent measurement over time and not be heavily influenced by environmental noise.
Valid	The metric should actually be measuring what it purports to measure rather than a loosely associated surrogate.
Interpretable	The metric value should be meaningful to the analyst. There should be a strong theoretical construct linking the measured quantity to performance.
Sensitive	The metric should be able to differentiate between a useful number of performance levels. At the very least it should discriminated adequate from inadequate performance.
Applicable	The metric should be capable of application to the operating environment. For example if the metric is to be used for real time simulator debriefing it must be able to be derived from the data in minimal time, probably autonomously.
General	The metric should be able applicable to different scenarios, fleets etc.

3.3.2 Scalar Measures of Tracking Performance

Research studies generally favour the use of scalar measures of pilot performance over discreet measures (as in FDM/FOQA) since the inherent increase in processing time is generally not a constraint and the greater depth of analysis offered by the higher resolution of measurement is beneficial. Performance is generally assessed by periodically recording the difference between critical flight parameters and their datum during a phase of flight. For example the difference between the aircraft's airspeed on the approach and the reference speed for its weight and configuration.

These 'time series' of errors can then be reduced into performance statistics. For example, the Mean Error metric (ME) computes the arithmetic mean of these error data so as to describe the average deviation from the target, commonly referred to as tracking accuracy. The Standard Deviation of Error metric (SDE) assesses variability around the mean thus describing the variability, or smoothness, of parameter tracking. These measures are typically applied to the principle flight path and aircraft state parameters such as airspeed, altitude, course deviation and glideslope deviation which are clear indicators of performance on well prescribed tasks such as an ILS approach.

Essentially these measures quantify the pilot's success at closing the outer control loop (see figure 3). Some studies have developed more sophisticated measures which look at the aircraft's velocity of divergence and compute the expected time to exceed tolerance (essentially a measure recoverability) or the time spent outside tolerance (a slightly more sophisticated variant of event analysis using continuous data). However, these measures have not been widely adopted, perhaps because they are relatively complex to compute and have less face validity when compare to more conventional metrics.

Another commonly applied metric in research is the Root Mean Square of Error (RMSE or RMS) which attempts to give a global assessment of tracking accuracy where smaller values generally indicate better performance.

However, RMSE has the disadvantage that it produces identical values for quite disparate performances. For example, being consistently high, consistently low, or at the correct mean height but with great variations in height keeping can all result in the same RMSE value (see Hubbard, 1987). Taken in combination ME & SE completely define RMSE yet offer greater diagnostic ability individually, and so they are generally used in preference where brevity of feedback is not a key factor, such as in this research study.

3.3.3 Advantages of Measuring Control Strategy

The measurement of key aircraft state and position parameters is clearly relevant to the measurement of performance. However, it is suggested that this may not be sufficient to precisely characterise differences in manual flying performance in this research setting. The problem arises due to the nature of control systems in large transport aircraft.

The mechanics of manual aircraft control was briefly touched upon in chapter two. Essentially all aircraft control systems are hierarchical in nature. Movements of the control inceptor cause a change in the aircraft flight control surface deflections which cause a change in the aircraft’s attitude which in turn cause a change in its translational rates which ultimately cause a change in its position, the desired effect. The relationship between each stage of the process is mediated by time constants, control powers and other factors which are a product of the control system design and which dictate its rate of response. This differs depending on the aircrafts form and function. For example light fighter jets generally have responsive control systems so that changes in inceptor position produce an almost immediate high rate of change in the aircraft’s attitude and position making them agile and manoeuvrable. However, large jet transport aircraft are designed quite differently. Their control systems are generally much more docile with inceptor changes generating less immediate and lower rate changes in attitude and position. This is done primarily to preserve passenger comfort and avoid over stressing the airframe.

Since the response rate of a large transport aircraft is limited to a relatively low value it is possible for the control input rate of the pilot to surpass it. Consequently the pilot may be demanding changes through the inceptor which the aircraft is unable to achieve. This gives rise to a potential disassociation between the pilot's control strategy and the behaviour of the aircraft.

The previously described performance metrics are all targeted at aspects of the aircraft's behaviour i.e. its speed or position in space. However, these metrics do not describe the demands which were made by the pilot and, owing to the nature of the control system, quite dissimilar input demands may result in very similar aircraft behaviours. For example, it is possible for a pilot to achieve the same flight path in a precision tracking task by making either low amplitude, well timed control inputs which precisely cancel out external disturbances, or by making more frequent, higher amplitude, ineffectively timed control inputs. The latter control strategy may have produced the same result but many of the input demands had no real effect on the aircraft and the total input energy expended by the pilot was far higher. Often this is reported by examiners as 'over control' of the aircraft.

Performance research (Baron, 1988) suggests that the level of energy used to control a system is of equal importance to the result of that control. For example two pilots may follow the localiser datum precisely whilst flying an ILS approach and from this perspective both would be judged to have achieved the same level of flying performance. However, if analysis of their control inputs revealed that one pilot made relatively few, low amplitude inputs to achieve this standard of tracking whilst the other was extremely active on the controls, making frequent large amplitude control reversals, then they may be viewed quite differently. The pilot who used less energy to achieve the same tracking result employed a more physically efficient control strategy and is more skilled in controlling the aircraft. It could also be argued that the use of a more skilled control strategy is consistent with the pilot operating in the autonomous rather than conscious control mode (Hawkins, 1998) and thus reflects on the amount of mental capacity demanded by the manual flying task. Whilst it is clearly desirable for the pilot to achieve key flight path and aircraft energy targets, it is also clearly undesirable that the pilot should expend the majority of their physical and cognitive capacity to accomplish this since secondary flight tasks will suffer.

Previous studies in which pilot performance has been evaluated using the traditional ME, SDE, and RMSE metrics have tended to involve light, agile aircraft (e.g. Davenport and Harris, 1992; Rees and Harris, 1995). Consequently the pilots control strategy is highly associated with the changes in the aircraft state and the traditional means of measuring performance from the outer loop parameters is acceptable. However, it is argued that since this research study will focus on large transport aircraft it is imperative that measures which evaluate the pilot's control strategy are used to augment these more traditional measures. By this means it will be possible to evaluate both the efficiency of the control process (i.e. how much effort was put in), and its success at closing the outer control loop (i.e. the product of that effort).

3.3.4 Approaches to Measuring Control Strategy

Despite the fact that many studies of pilot flying performance report collecting flight control input data, rarely has there been any analysis of pilot control strategy in such a setting (Veillette, 1995). It is not certain why this should be the case but it is suggested that a lack of a clear methodology for measuring and evaluating control input data may be a contributory factor. Primarily, unlike outer loop tracking performance, interpreting control input strategy is less intuitive and requires a greater degree of technical analysis to achieve.

One very simplistic approach which is suggested in literature (Baron, 1988) is to apply the RMSE metric to time series records of control inceptor displacement (see figure 6). Used in this way the RMSE metric can give a reasonable quantification of control input energy. It has in fact been proposed as a measure of pilot physical workload in some FDM systems. However, the metric fails to capture any information about the frequency at which control inputs were made and this property, along with control input amplitude, are typically cited as the critical aspects of the control strategy (see chapter 2).

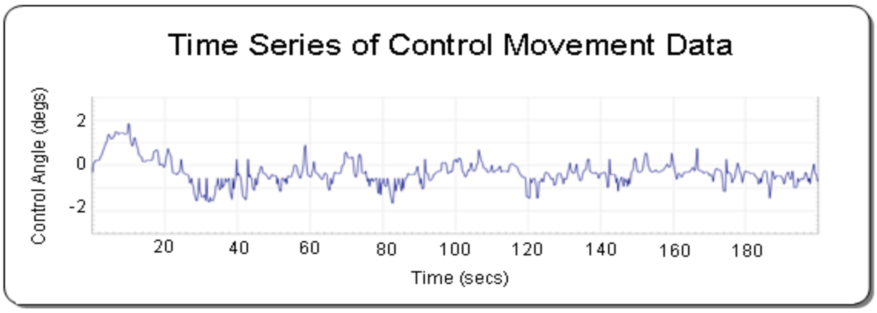


Figure 8 – Example time series of control inceptor displacement

More sophisticated approaches to characterising control strategy are described in the literature. McDowell (1978) initiated the development of control strategy measures to evaluate pilot performance based upon frequency analysis techniques. He used a series of analogue electronic filters to estimate how control input energy was distributed amongst a series of frequency bands for novice, intermediate and experienced pilots flying a Cessna T-37 light military training aircraft. It was found that the more experienced pilots generally concentrated most of their control input energy at the higher frequency end of the spectrum, particularly in the roll axis. It was concluded that there were changes in pilot's control movement power spectra (distribution of control input energy over frequency) as a function of skill level, and that measures of this property could be used effectively to discriminate pilot skill/experience level. Whilst McDowell (1978) utilised quite cumbersome analogue electronic filters to collect his data modern digital signal processing techniques make transforming time series records of control input displacement into the frequency domain a relatively straight forward task. Using these techniques it is possible to replicate McDowell's approach to the measurement of control strategy, but it requires some prior knowledge of power spectral density and discreet Fourier transforms.

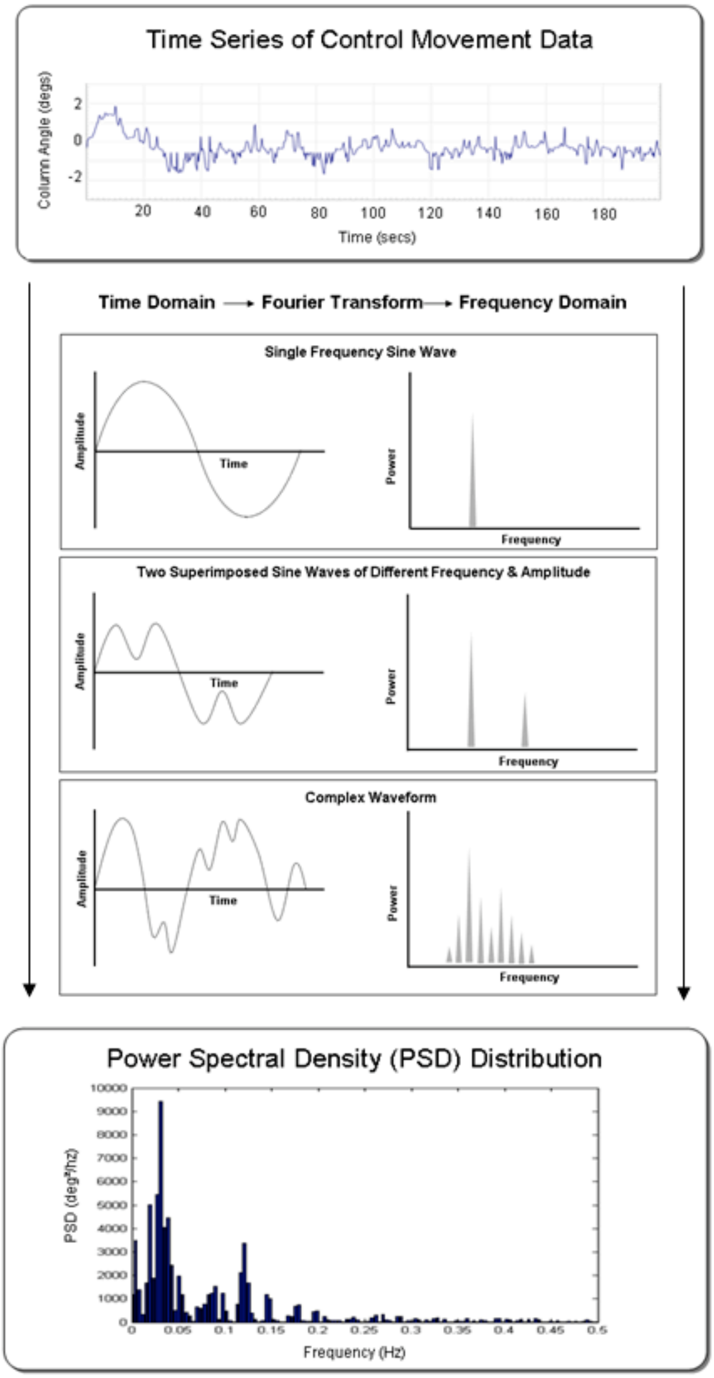


Figure 9 – An illustrative diagram showing the transposition of time series data into the frequency domain using a fast fourier transform algorithm

3.3.5 Power Spectral Analysis

The power spectrum shows how the power of a signal (energy per unit time) is distributed over a frequency range and thus by examining it, it becomes possible to determine how much of the signals power falls into a given frequency bin. This is essentially the approach of McDowell. In digital signal analysis the typical means of computing the power spectrum is by performing a discreet Fourier transform. The discreet Fourier transform (DFT) identifies periodicities in a series of measured data and measures the relative strength of that periodicity (Press, Flannery, Teukolsky and Vetterling, 1989). Essentially such algorithms work on the assumption that any complex waveform can be expanded into a superposition of Sines and Cosines of varying amplitude, frequency and phase (see figure 7). The discreet Fourier transform, F_n, of a series of data f_k with N data points is given by;

$$F_n \equiv \sum_{k=0}^{N-1} f_k e^{-2\pi i n k / N}$$

With appropriate scaling the coefficients of the DFT give the power spectral density (PSD) of the time series data (see figure 8), expressing power per unit frequency (note that phase information is lost in the PSD distribution). Peaks in the distribution show strong periodicities in the signal and indicate where power is concentrated. Although scaling and interpretation of this process can be complex many maths processing packages such as Matlab^{TM} incorporate streamlined DFT functions. By integrating the PSD data between two frequency limits it is possible to determine the amount of control input power within that frequency band (see figure 9). A simple Matlab script file which was developed for this purpose is reproduced in appendix E.

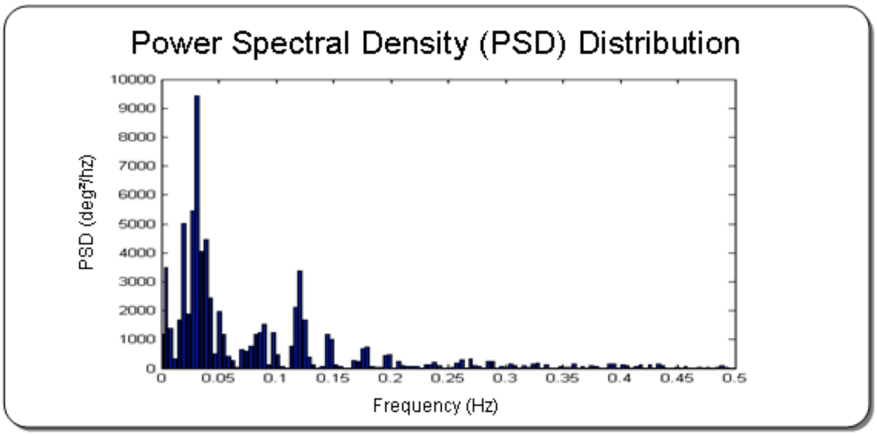


Figure 10 – Example Plot of Power Spectral Density distribution (periodogram) for a control input signal

For analysis purposes frequency bands must be chosen which capture the typical regions of variability in the PSD distribution amongst pilots of differing skill. Whilst this is best established through inspection of data there are also research studies in which similar banding has been proposed and shown to discriminate performance (e.g. McDowell, 1978). It should be considered that these will be specific to the type of aircraft, the task and the environment in which it is conducted. It is possible that the presence of environmental noise may limit the use of these measures in the real world but they are quite suitable for comparing across pilots in a simulated study where environmental noise is highly constrained.

It is worth noting at this juncture that there is a very well established practice of using frequency analysis for the design and evaluation of aircraft control systems. In laboratory environments the actions of the pilot in controlling a know disturbance function can be evaluated using frequency analysis and visualised as bode plots to determine how effective they were (McRuer and Jex, 1967). Furthermore the development of the ‘human transfer function’ aimed to produce a pilot response model which could be used to evaluate the stability of various control system designs. These research strands have been very productive in the design and engineering role, but the techniques employed by them are not suitable for the assessment of pilot performance in a less well controlled and more open ‘real world’ environment. Consequently they have not been detailed further within this research.

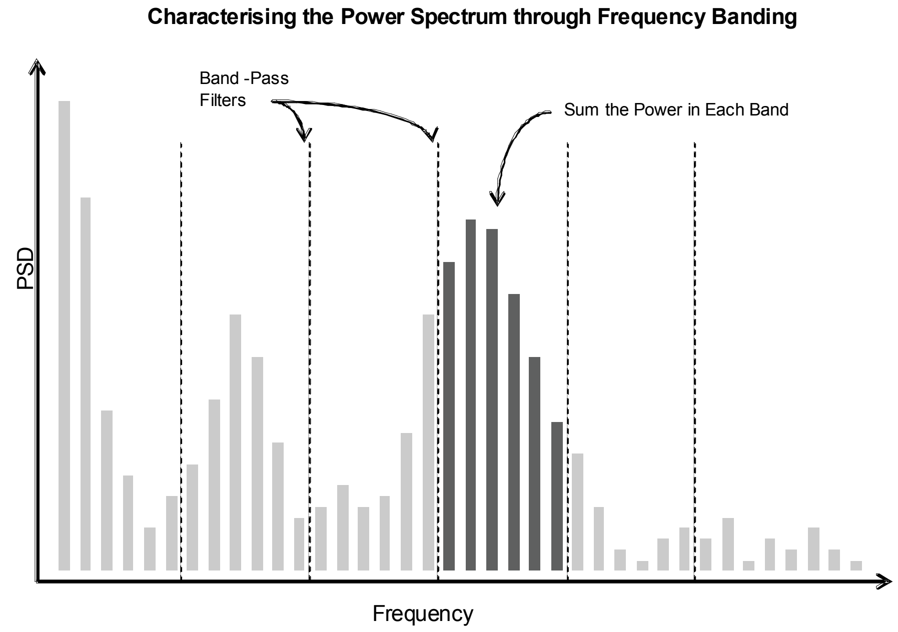


Figure 11 - Illustration of frequency banding of the power spectrum in order to measure differences in control input strategy

3.3.6 An Alternative Characterisation of Control Strategy

Recently Rantanen, Johnson and Talleur (2004) reviewed objective pilot performance measurement techniques in an attempt to develop an automated pilot proficiency scoring system. Their work concentrated heavily on the benefits of frequency analysis of time series data and they proposed several new metrics to quantify differences in PSD distributions which were associated with performance. They did not adopt a band-pass technique as previously outlined. Instead their technique involved first subjecting the PSD data to a high pass filter, removing any spectral components which did not reach a critical magnitude and were therefore considered noise. Various dimensions of the distribution of the remaining components were then analysed, forming the performance metrics (see figure 10). The average amplitude of these significant components and their spread in amplitude were taken as two such measures (labelled the mean magnitude of spectral components (MSC) and the standard deviation of the magnitude of spectral components (DSC) respectively).

The researchers hypothesised that more skilled pilots would distribute their control input power more evenly over the frequency range and generally use smaller inputs, thus their filtered PSD distributions would have more evenly less variable and smaller spectral powers. Also measures of the mean, median and spread in frequencies of the significant spectral components were computed (labelled FMGC, MEDF and FDGC respectively). The researchers theorised that more skilled pilots would have a greater spread in the frequency of their spectral components which would be shifted towards the higher end of the frequency spectrum (see Johnson et al., 2004; Johnson and Rantanen, 2005; Rantanen et al., 2001; Rantanen et al., 2004).

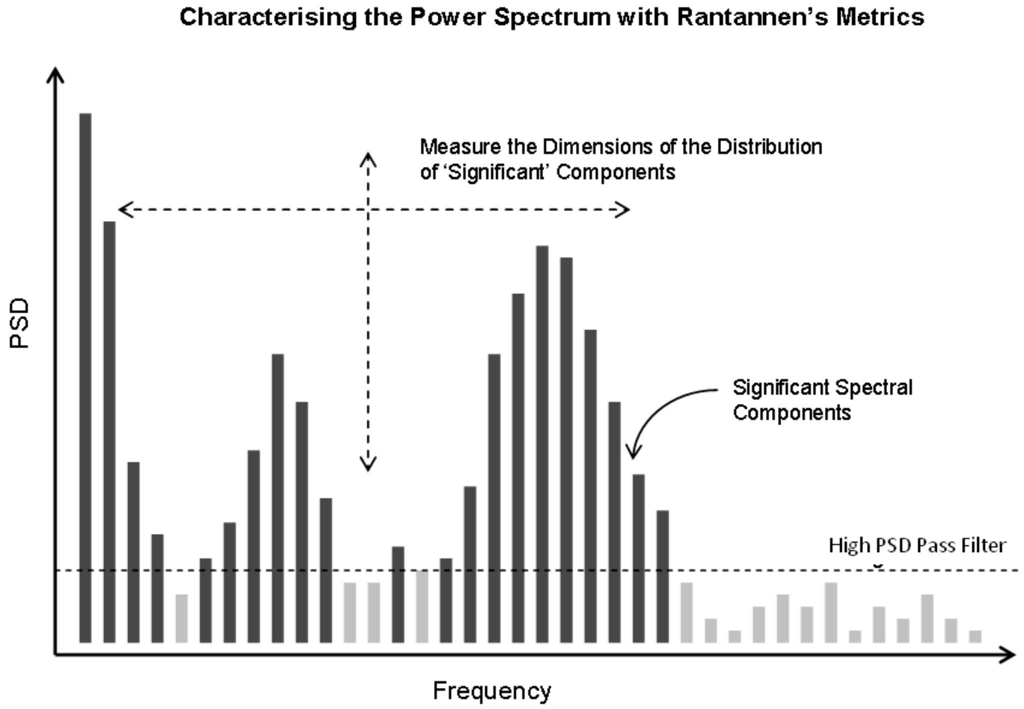


Figure 12 - Illustration of the Rantanen et al. approach to characterising control strategy from the power spectra

The researchers applied their proposed metrics in a series of flight trials, collecting data from a sample of pilots performing an instrument proficiency check (IPC) in a light aircraft. The study demonstrated that the metrics were capable of discriminating between pilots who had passed or failed the IPC. However, due to limitations in the data collection system, control input data was unavailable and the metrics were applied only to outer loop parameters (see McRuer 1982 and Figure 3), such as course deviation and glideslope deviation indications. Also, the individual elements of the test were isolated rather than integrated into a real time scenario and the overall cognitive demand (i.e. the requirement to anticipate future manoeuvres, energy changes etc.) was relatively low. Since it is expected that manual flying failures may result from the effects of limited cognitive capacity this particular lack of fidelity may have impacted the study significantly.

Effectively the researchers measured aircraft performance rather than pilot performance, the former being mediated strongly by the stability characteristics of the machine, and have not measured differences in control strategy *per se.* Furthermore since the study was performed upon light aircraft which have significantly different stability characteristics to large transport aircraft it could not be assured that the metrics would perform well in the setting of this research.

The metrics proposed by Rantanen et al (2004) are built upon a strong theoretical basis and show promise. However the properties which they measure are somewhat unintuitive and complex to derive from the PSD distribution (careful tuning of a low pass filter is required in order to properly separate the signal from noise without sacrificing information), violating some of the principles outlined in table 1. However the underlying theory of extracting control strategy information by characterising the PSD distribution seems to be valid and is in keeping with other research approaches.

3.4 Performance Measurement Literature Summary

There is strong evidence from the literature that the ME and SDE metrics can provide good discrimination of pilot performance when applied to critical outer loop flight parameters (i.e. airspeed, localiser deviation, altitude etc) on a well prescribed task (i.e. where the parameters the pilot is trying to track are explicitly defined). Consequently such measures are certain to be included in the future stages of this research to measure pilot manual flying performance. However the literature also recommends that measures of control strategy are used to augment these more traditional measures in order to gauge the level of input effort exhibited by the pilot. Basic application of the RMSE metric to control input data fails to capture important aspects of the control strategy, namely the frequencies at which those inputs were made.

However frequency analysis techniques may be used to classify different control strategies. There appears to be two primary techniques employed to quantify the power spectrum. The most commonly applied method is to simply ‘band-pass’ the data and determine how much signal power falls into various frequency bins (e.g. McDowell etc). Alternatively Rantanen et al (2004) have recently developed a series of metrics which give a general description of the power spectral density distributions shape. Whilst the former strategy has been reasonably well demonstrated in literature the latter has only found limited application. Neither have been applied for the measurement of manual flying control strategy variation in large transport aircraft which have significantly different control systems with large amounts of lag. Consequently it is difficult, based on literature alone, to select a suitable control strategy measurement technique to take onto future parts of this research programme. It was also considered that there may be some advantage to conducting frequency analysis on the first or second order derivatives of control displacement data (thus analysing control input velocity or acceleration respectively) since signal noise may be reduced. However it was decided to reserve this exploration for future trials.

In order to validate that the described measurement techniques could adequately discriminate between subtle differences in pilots' manual handling ability, and to gauge which measurement strategy was most effective, an empirical study was undertaken. The metrics were applied to data collected from students undertaking a Jet Orientation Course (JOC) in a fixed based simulator device. From instructor assessment data the evaluated students were known to have exhibited a positive manual handling performance increment as a result of the training course. The study aimed to evaluate how successfully the various measurement techniques could discriminate this performance difference. The most capable metrics for the subsequent phases of this research programme were selected on the basis of this analysis.

3.5 Study Aims & Objectives

The following empirical study aimed to demonstrate that frequency analysis based metrics could quantify a meaningful change in control strategy as student pilots undergoing jet transport flight skills training acquired manual handling expertise. Furthermore it tests the hypothesis that as this expertise develops control strategy changes will be characterised by an overall decrease in control input power and a shift in dominance from the low frequency bands to the higher frequency bands, i.e. a more efficient strategy, as proffered by previous research. If observed this would mean that the pilot's control inputs were being made more frequently, signifying that the observe→process→respond cognitive loop is being closed more rapidly and therefore possibly demanding less mental resource. The two methods of quantifying control strategy (frequency banding and Rantanen's metrics) will be contrasted to determine which is more sensitive and capable of discriminating this performance change.

3.6 Method

3.6.1 Participants

The manual handling performance of 15 cadet pilots (all male, aged between 18 and 25) was evaluated whilst they undertook a 40-hour Jet Orientation course on a Boeing 737NG. All students had similar levels of flying experience (approximately 180 hours in light singles and twins) at the commencement of the course. None of the students had any flying experience of large jet transport types prior to undertaking the conversion course. All of the sampled students subsequently passed the Jet Orientation Training (JOT) course without the requirement for remedial training and were judged by the examiners to have significantly enhanced their manual flying ability. The research process was approved by the Cranfield School of Engineering Ethics Board which adheres to the guidance for ethical conduct promulgated by the British Psychological Association.

3.6.2 Equipment

The study was undertaken on a fixed-base JAA approved (Level 2) Flight Training Device (FTD) simulating a Boeing 737NG series aircraft. The FTD incorporated a 180 degree directly projected outside visual display, six screen Electronic Flight Instrument System (EFIS) and flying controls with electrically generated control loading.

A data logging computer running bespoke software was integrated into the system to collect flight data from the exercises. The logging system recorded 92 flight parameters, including ILS tracking data, and position data from all the primary flight inceptors, at a sampling frequency of 4Hz. Flight data were stored as comma delimited text files with time and date encoded filenames to allow for their identification whilst preserving participant anonymity.

3.6.3 Task

The Jet Orientation Training syllabus required students to fly manual precision instrument approaches at a number of intervals throughout the course. The orientation course immediately followed the student's initial CPL(IR) training. Each student's performance was sampled twice, once within the initial period of training and once during the final stages of the training programme, so that a longitudinal comparison could be made.

The conditions of the approach were standardised for each student on each run so that weather, aircraft and traffic conditions would be consistent. The students were asked to fly a manual approach (i.e. without autopilot, flight director or autothrottle assistance) in instrument meteorological conditions (IMC) to a minimum decision height of 200ft using the Instrument Landing System (ILS) for guidance. In several previous studies this task proved to give the best discrimination of performance between pilots (see Rantanen et al, 2004).

3.6.4 Performance Measures

Flight data acquired from the trials was collated and imported into the Matlab data analysis suite. A bespoke M-file was produced to compute the ME (mean of Error) and SDE (Standard Deviation of Error) metrics for the outer loop parameters (localizer deviation, glideslope deviation and airspeed deviation) giving measures of accuracy and smoothness respectively. The M-file routine also performed a Discreet Fast Fourier Transform operation on the principle control movement data records (control wheel angle, control column angle, rudder pedal angle) from which the PSD was computed. Unfortunately records of thrust lever angle and commanded thrust were rendered unusable due to an error with the data acquisition computer.

It is important to consider the effects of aliasing, windowing and end-effects when analysing the time series data. The maximum frequency of periodicity which may be detected in signal is half the sampling frequency of that signal as defined by the Nyquist-Shannon sampling theorem (Shannon, 1998). Effectively this occurs because any variability of higher frequency takes place between the periods of measurement and thus will be unobservable. The effect is known as aliasing. FDRs and QARs typically sample data at 4hz making the maximum analysable input frequency 2hz, or two control reversals per second. In the studies of Rantanen *et al.* (2004) control input activity was rarely seen to exceed 2hz. Furthermore, since these studies were performed in light, agile, aircraft it is expected that there will be even less activity beyond 2hz in an aircraft with a much larger and heavier control inceptor, as will be used in this study. Consequently the effect of aliasing was not considered to present a significant threat to the study.

Whilst aliasing threatens to mask the high frequency portion of the data, windowing effects threaten to discard the low frequency data. When a period of data is sampled a window is formed. If the periodicity of the signal exceeds the length of the sampling window valuable information will be lost. Whilst extending the window might change the proportional magnitude of any signal effect, small windows may mean that the signal is lost in noise. In this study the window size is mediated by the relatively short duration of the task. However, it was possible to permit a window length of 90 seconds which would allow relatively long period oscillations to be detected, but this may also mean some shorter period oscillations were masked.

Finally, Discrete Fourier Transform algorithms generally assume that the signal upon which they operate is continuous. This means that a data sample should contain a perfect number of signal oscillations and its end point should match perfectly to its start point. In reality this is rarely the case and imperfections know as “signal end effects” exist which create nose in the spectral plot. Prior to performing a Fourier transform it is possible to apply various filters (a Hamming window is a well known example) which can reduce the impact of signal end effects. However, selecting an appropriate filter requires detailed knowledge of the signal and if done improperly can actually introduce further noise. Since there was little prior data for the selection of a filter in this study a simple boxcar filter was applied (essentially minimal suppression of signal end effects).

The spectral plots for each case were inspected visually in the first instance. It was noted that in all cases there was no significant spectral structure beyond 0.25Hz. Two other research analysts familiar with the context of the research were asked to look through the spectral plots and independently suggest the frequency which generally represented the upper bound of significant spectral activity. In both cases the analysts concluded that 0.25Hz appeared to be the ‘cut-off’ frequency. Therefore the frequency range 0Hz to 0.25Hz was divided into five equal frequency bands; very low frequency (0Hz to 0.05Hz), low frequency (0.05Hz to 0.10Hz), medium frequency (0.10Hz to 0.15Hz), high frequency (0.15Hz to 0.20Hz) and very high frequency (0.20Hz to 0.25Hz). Control input power within these bands was computed by integrating the PSD curve between the frequency limits using a Matlab routine. These values formed the frequency band metrics which were named VLFB, LFB, MFB, HFB and VHFB respectively. Additionally, the Rantanen measures of control strategy were computed from the PSD data (see section 3.3.6), again using a Matlab routine. The high pass filter in this case was set with reference to previous research on the subject by Johnson, Rantanen and Talleur (2004). All metric values were then imported into the SPSS package so that statistical comparisons of early and late training performance could be made using paired t-tests.

3.7 Results

3.7.1 Flight path tracking accuracy and smoothness (outer loop parameters)

Using a paired t-test (see table 2) no significant differences were observed in the mean tracking error (ME) of the ILS localiser early and late in the training course. Similarly there were no significant differences in the standard deviation (SDE) of Localiser tracking early or later in the training course.

No significant differences were observed in students' performance early and late in the training course for mean tracking error on the ILS glideslope. Furthermore there were no significant differences in the smoothness of glideslope tracking early or late on the training course (see table 2).

Similarly, there were no significant differences between the mean airspeed error early and late on the training course. However, performance later on the course did demonstrate significantly lower standard deviations of airspeed error. This was indicative of greater stability in the control of the target approach speed (see table 2).

Table 2 - Arithmetic Mean of Error (ME) and Standard Deviation of Error (SDE) for ILS outer control loop parameters broken down by early or late course assessment. Highlighted row indicates a statistically significant result.

ILS Tracking Performance	Early in Course		Late in Course				
	M	σ	M	σ	t	df	Sig.
Tracking Accuracy - ME							
Localiser (dots)	0.064	1.359	-0.778	2.029	1.331	14	0.206
Glideslope (dots)	-0.362	0.282	-0.266	0.357	-0.854	14	0.408
Airspeed (kts)	16.293	7.633	8.92	3.641	-2.142	14	0.050
Tracking Smoothness - SDE							
Localiser (dots)	0.041	0.037	0.013	0.031	2.007	14	0.064
Glideslope (dots)	0.006	0.006	0.006	0.007	-0.159	14	0.872
Airspeed (kts)	0.415	0.276	0.416	0.358	-0.005	14	0.996

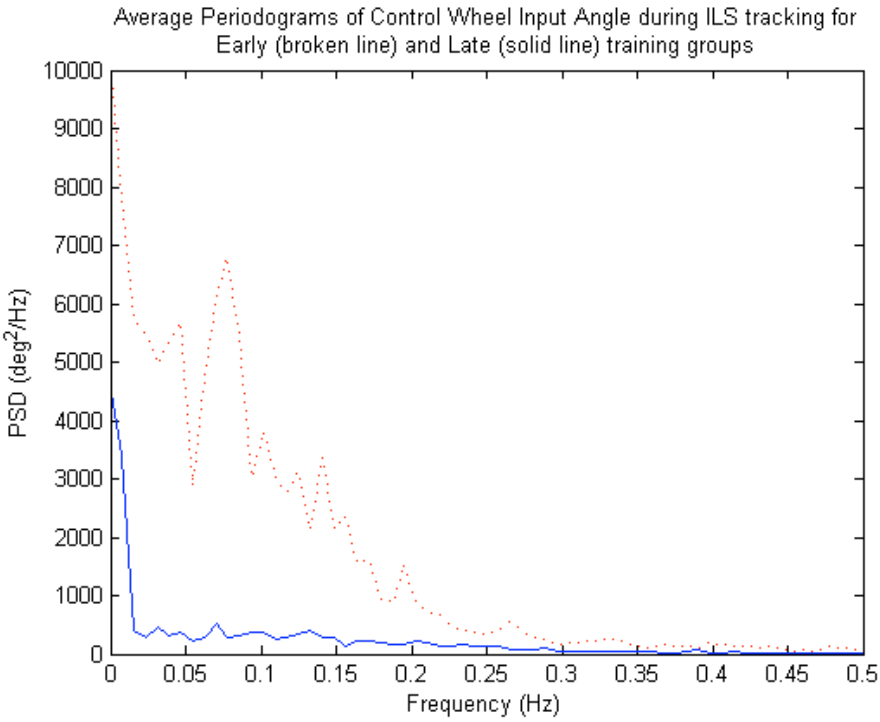


Figure 13 - Averaged periodograms of Control Wheel Inputs for Students Early and Late in Training

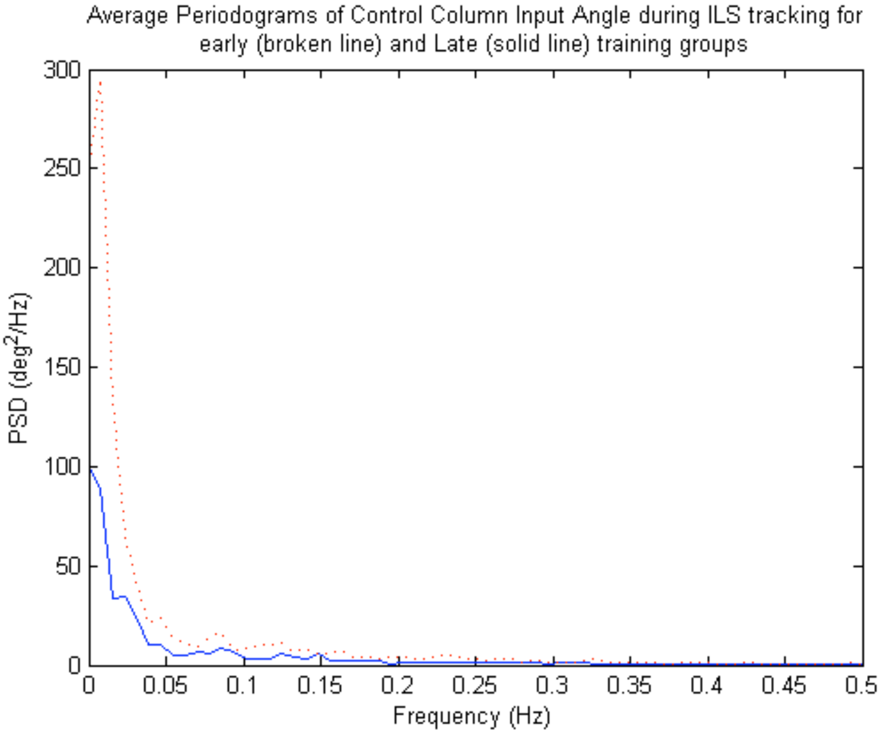


Figure 14 - Averaged periodograms of Control Column inputs for Students Early and Late in Training

3.7.2 Summary Frequency Measures of Control Strategy

Visually there were distinct differences in the roll and pitch control strategies adopted by the students early and late in the training course, as demonstrated by the averaged Periodograms presented in figures 11 & 12. Control input power appeared to reduce across all frequencies as training progressed. In all cases control power appeared to be concentrated towards the low frequency end of the spectrum

Applying the summary frequency metrics devised by Rantanen et al. it was observed that the magnitude of significant spectral components for roll input reduced later in training (see table 4). This result indicates that roll control inputs generally reduced in amplitude with the development of manual flying expertise. The analysis technique also indicated that the spread of frequencies of these components was significantly broader later in training.

The summary frequency metrics indicated that the general amplitude of pitch control inputs was somewhat reduced later in training (see table 4). In general the Rantanen measures offered a similar description to the change in control strategy as the frequency band-pass technique

3.7.3 Frequency Band-Pass Measures of Control Strategy

Using the frequency band-pass analysis method roll control input power (lateral control wheel movements) was found to be significantly lower across all of the frequency bands for students late in their training. This indicates that as expertise developed the effort expended in controlling the aircraft in roll generally reduced (see table 3).

Table 3 – Band-pass frequency analysis metrics for Primary Flight Control Inputs during the ILS tracking task broken down by early or late course assessment. Highlighted rows indicate statistically significant ($p<0.05$) results.

Control Input Strategy - Frequency Band Metrics							
	Early in Course		Late in Course				
	M	σ	M	σ	t	df	Sig.
Control Wheel Power (degs^2)							
Very Low Frequency Band	34183	36177	9134	26064	2.282	14	0.039
Low Frequency Band	31588	39603	2039	2278	2.855	14	0.013
Mid Frequency Band	21151	17070	1902	2461	4.744	14	0.000
High Frequency Band	14135	11889	1776	1701	4.334	14	0.001
Very High Frequency Band	5200	3796	873	1180	4.374	14	0.001
Control Column Power (degs^2)							
Very Low Frequency Band	782	724	280	369	2.233	14	0.042
Low Frequency Band	92	48	42	54	2.681	14	0.018
Mid Frequency Band	64	41	30	47	2.148	14	0.050
High Frequency Band	41	36	21	23	1.735	14	0.105
Very High Frequency Band	22	15	9	8	2.936	14	0.011
Rudder Power (degs^2)							
Very Low Frequency Band	3024	4515	2129	2592	0.593	14	0.282
Low Frequency Band	119	191	67	92	0.939	14	0.182
Mid Frequency Band	68	139	19	21	1.392	14	0.093
High Frequency Band	76	178	15	28	1.290	14	0.109
Very High Frequency Band	25	54	6	11	0.715	14	0.109

The frequency band-pass analysis method also identified that pitch input power (fore-aft movements of the control column) was significantly reduced in the very low, low, mid and very high frequency bands for students late in their training. Again, this indicates that as expertise developed the pitch control strategy changed, with less power expended over most of the input frequencies (see table 3).

The frequency band analysis method identified no significant differences between the yaw input power (rudder movements) at any frequency band for students at either stage of their training. This indicates that there were no identifiable changes to the yaw control strategy associated with manual flying expertise on this task.

Table 4 – Summery frequency metrics for Primary Flight Control Inputs during the ILS tracking task broken down by early or late course assessment. Highlighted rows indicate statistically significant ($p<0.05$) results.

Control Input Strategy – Summary Frequency Metrics							
	Early in Course		Late in Course				
	M	σ	M	σ	t	df	Sig.
Control Wheel Input							
MSC (degs²/hz)	448	361	71	126	4.121	14	0.001
DSC (degs²/hz)	1775	1582	416	1103	2.944	14	0.011
FMGC (hz)	0.122	0.029	0.175	0.100	-1.884	14	0.080
FDGC (hz)	0.078	0.016	0.122	0.016	-2.394	14	0.031
MEDF (hz)	0.087	0.033	0.120	0.071	-1.359	14	0.196
Control Column Input							
MSC (degs²/hz)	4	3	2	2	2.595	14	0.021
DSC (degs²/hz)	28	27	11	14	2.071	14	0.057
FMGC (hz)	0.067	0.036	0.073	0.045	-0.326	14	0.749
FDGC (hz)	0.056	0.030	0.058	0.032	-0.124	14	0.903
MEDF (hz)	0.022	0.017	0.041	0.040	-1.556	14	0.142
Rudder Input							
MSC (degs²/hz)	13	19	9	10	0.756	14	0.462
DSC (degs²/hz)	129	196	91	115	0.574	14	0.575
FMGC (hz)	0.024	0.058	0.014	0.016	0.652	14	0.525
FDGC (hz)	0.019	0.037	0.013	0.015	0.526	14	0.607
MEDF (hz)	0.011	0.038	0.004	0.009	0.708	14	0.491

3.7.4 Descriptive Power of the Metrics

For each performance metric the partial Eta Squared statistic was computed. This value represents the amount of between conditions variance in the dependent variable (metric score) explained by the levels of the independent variable (stage of training). Values range from zero to one, a value of one indicating that 100% of the variance is explained by the independent variable. A plot of partial Eta Squared values are presented in figure 13. Whilst the values for the outer loop parameter tracking metrics are generally low, the frequency band measures of control input strategy are high and surpass the values achieved by the Rantanen et al. metrics (2004).

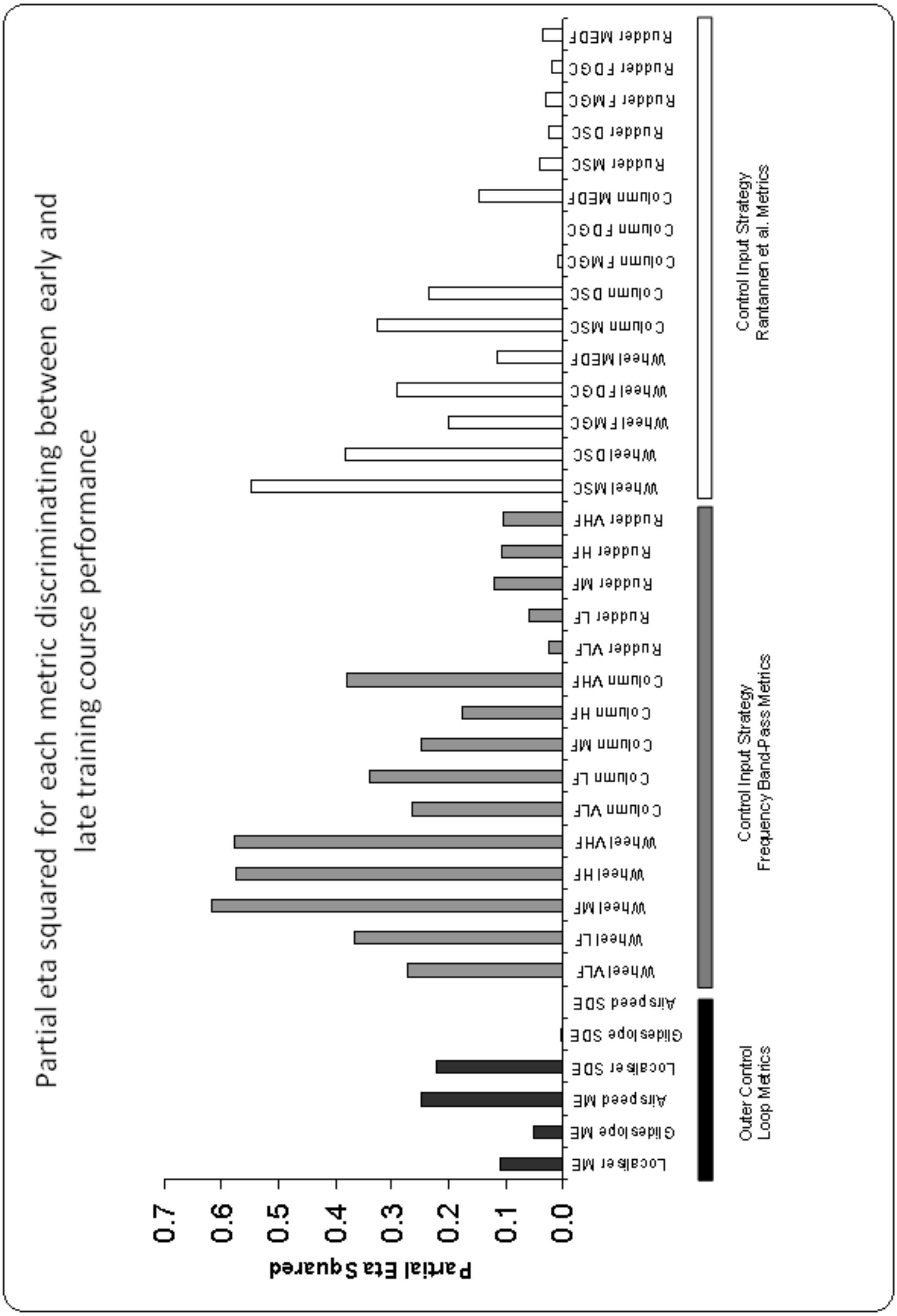


Figure 15 – Plot of partial Eta squared for each performance metric, illustrating how well they discriminated the two levels of student performance.

3.8 Discussion

3.8.1 Measures of outer-loop parameter pursuit tracking

When measuring performance from a traditional perspective (by examining error in outer loop parameters – see McRuer 1982) the results showed only a few small differences in performance over the period of the training course. There were no significant improvements in either the accuracy or smoothness of lateral or vertical flight path tracking on the ILS (see table 2). However, the results indicate that on both measurement occasions the student's demonstrated relatively high levels of tracking smoothness and accuracy and there was only limited scope for improvement.

The results indicate that the smoothness of airspeed tracking improved significantly over the period of the training course (table 2). Students who had gained more manual handling experience on the aircraft type held a more consistent airspeed rather than allowing it to drift and then correct. The results indicate that whilst students at either stage of their training were equally capable of maintaining the aircraft's flight path on the ILS datum, the simultaneous management of the aircrafts energy proved more problematic and benefited considerably from gains in manual flying experience (see table 2).

Students undertaking the course were transitioning to flying a large jet transport aircraft having spent a considerable amount of time flying light twin propeller aircraft. One of the most significant differences between the two aircraft is in the management of their energy, as noted in the introduction. Large jet transport aircraft are generally very aerodynamically 'clean' with highly efficient, low drag wings. They can be reluctant to dissipate energy and slow down. Furthermore their high bypass turbofan engines lag in response to control demands and can generate considerable secondary pitching moments. Consequently the task of maintaining the larger aircraft's energy is far more complex and without adequate control significant airspeed errors can accumulate.

This difference in the complexity of airspeed management appears to show through these results and indicates that ‘outer-loop’ measures did provide some useful information, supporting other studies which have used them to capture differences in manual handling performance (e.g. Rees and Harris, 1995). However, the general sensitivity of such measures appears limited when applied to studies of large transport aircraft, supporting the view that further measures of control input strategy need to be included.

3.8.2 Measures of control input strategy

With regard to control strategy measurements both the frequency banding metrics and the Rantanen metrics successfully discriminated between the performance of the students early and late on the training course on a number of dimensions (see tables 3 and 4).

The frequency band metrics revealed that both early and late in training the majority of the control input power was concentrated at the lower frequency end of the spectrum. This differs somewhat to the results obtained by McDowell (1978) who found control input power to be concentrated at the higher frequencies. However it is expected that the results should differ as this study involved light military jet aircraft which respond rapidly to control inputs. As noted in the introduction the pilots of large transport aircraft must generally adopt a lower frequency control input strategy due to the limitations of the aircrafts response rate. The large spikes in power at the extreme lower end of the spectrum are probably the result of periods of control inactivity i.e. when the inceptors remain static. This control strategy is plausible in a highly stable aircraft such as that used in this study, but would be relatively rare in a more ‘twitchy’ light military jet.

In support of McDowell’s findings the frequency band metrics recorded that over the period of training control input power significantly reduced in both the pitch and roll control axes (see table 3). McDowell found that the low frequency components reduced in power so that the high frequency components of control became more dominant. The results presented in this thesis differ slightly in that the reduction in control power reduction was of the order 60% to 80% across all frequency bands and not specific to any particular one. Again, it is suggested that this is a result of the difference in control system between the aircraft used in these studies.

In general though the results of the frequency band analysis show that there is much less control power input at the lower end of the frequency spectrum. This represents a significant change in the control strategy used to resolve the control problem. It is suggested that as pilots progressed in training they began to make a more varied range of control inputs which suited the errors they observed and needed to correct, rather than using occasional coarse and jerky higher amplitude inputs that would generate the high levels of low frequency power observed in the spectral distributions. In essence the pilots control strategy was refined to the task. These results correspond closely to the observations of Rantanen et al (2004) who found that pilots range of control input frequencies broadened with greater expertise.

In contrast the summary frequency metrics proposed by Rantanen et al. give a coarser measure of the change in control strategy over the period of training. Whilst they similarly indicate a reduction in the total control input power they are less able to quantify how the distribution of this power over frequency was modified (see table 4). The results indicate that the power was more widely distributed over the frequency range but can be no more specific than this. Furthermore the comparison of partial Eta Squared values shows that the Rantanen metrics were generally less sensitive than the frequency band metrics and offered less explanatory power. The reason for this reduced discriminative ability is most likely because these metrics are based on simple averages and thus subtle patterns in the spectral distribution are largely discarded and overlooked.

Neither set of metrics measured any significant change in the yaw control strategy of the pilots over the period of training. This is perhaps a function of the task administered to the pilots since a symmetric thrust ILS generates few yaw control demands in a large swept wing airliner. Without a requirement to demonstrate their expertise it is therefore hard for pilots show that they have improved their performance and ‘stand out from the crowd’. Engine failure events are typically unexpected and challenge the diagnostic and decision making skills of the pilot. Pilots who handle such situations well typically have the capacity and foundation of experience to sense the initial, subtle, variation in the aircrafts response sooner and make appropriate corrective inputs before aircraft state deviations become large and more aggressive corrective action is required. It may therefore be prudent to increase the yaw control demands of tasks used in future studies by including a crosswind component or an asymmetric thrust condition (i.e. engine failure). The latter may be particularly relevant since data from the introduction (CAP 776, 2008) shows many fatal manual flying related accidents to occur post engine failure, where either directional control is lost or the airspeed is allowed to decay excessively.

3.8.3 Overview

Whilst the outer loop tracking metrics indicated that there were only minor differences in a generally high standard of flight path tracking over the period of the training course, the control strategy metrics indicate that the effort required by the pilot to achieve this performance significantly reduced with training and that their control strategy became more refined.

The results support the hypothesis (Baron, 1988) that when measuring manual flying performance in large jet transport aircraft measures of control strategy may be more sensitive to the change in expertise than measures of outer loop parameter tracking. Thus small differences in performance maybe detected using the former metrics before they are detected with the latter.

Clearly it is essential that measures of outer loop tracking performance are included in any study of manual flying since the successful outcome of performance is critical. However, measures of control strategy allow for a separate and complimentary analysis of performance. The results of this study suggest that the two types of performance metric are used in parallel in order to build a more sensitive and broader measurement system.

3.9 Chapter Conclusions

The study suggests that coarse ’event’ type measurement metrics, as employed in FDM/FOQA, are of limited use within the scope of this research study. It also suggests that traditionally employed measures of outer-loop parameter tracking (ME & SDE) are valid in this research setting but may lack sensitivity when applied to large transport aircraft.

Empirically derived data shows that frequency band metrics are able to sensitively measure difference in the control strategy of pilots with different manual flying ability. Pilots with more manual handling experience generally use less control input power to achieve equal levels of tracking performance. It is proposed that these measures are used as an adjunct to the more traditional tracking measures in order to improve overall sensitivity. Alternative measures of control strategy proposed by Rantanen et al are rejected in this instance as they demonstrate poorer sensitivity and explanatory power.

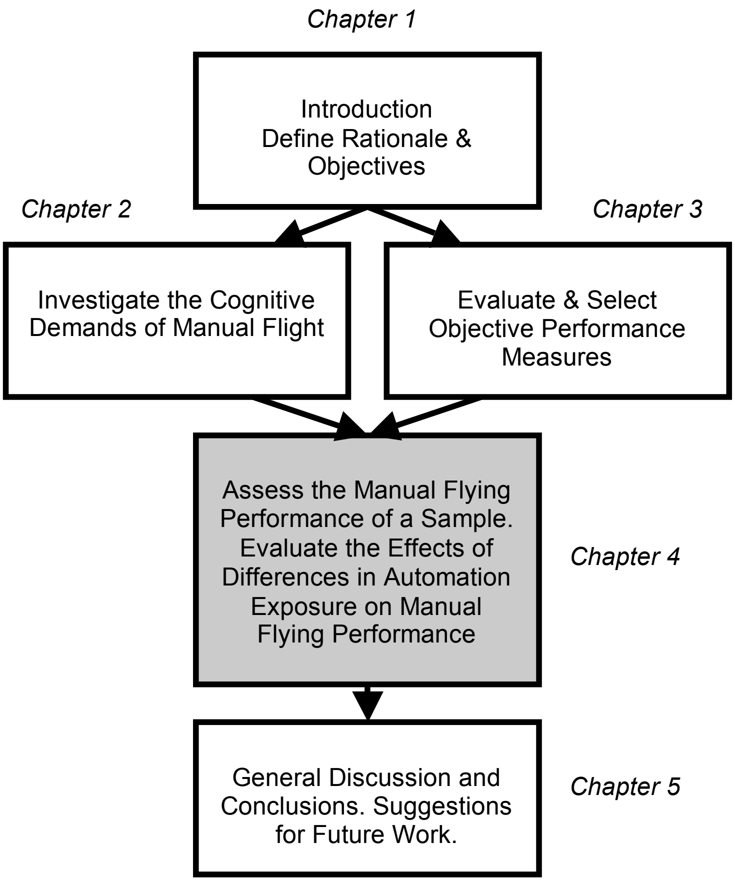
A battery of performance metrics suitable for the fine grained assessment of pilot manual handling ability was chosen from the results of this empirical analysis and is presented in table 5.

Table 5 - Selected battery of performance metrics

Metric	Description
ME	Gives the average accuracy of tracking (dots)
SDE	Gives the variability or smoothness in tracking (dots)
VLF	Control input power in the very low frequency band (0hz to 0.025hz) (deg^2)
LF	Control input power in the low frequency band (0.025hz to 0.075hz) (deg^2)
MF	Control input power in the medium frequency band (0.075hz to 0.125hz) (deg^2)
HF	Control input power in the high frequency band (0.125hz to 0.175hz) (deg^2)
VHF	Control input power in the very high frequency band (0.175hz to 0.225hz) (deg^2)

Chapter 4

Study III: Evaluating the Effects of Differences in Automation Exposure on Manual Flying Performance.



4.1 Introduction

The initial phase of the research programme identified a safety concern which related the reduction in manual flying exposure in air transport operations to a perceived degradation in manual flying ability. Furthermore it was apparent from a review of previous research and publications that there was insufficient objective evidence to properly address this safety concern (see chapter 1). To provide objective evidence an empirical study of professional air transport pilot's manual flying ability was undertaken and is reported in this chapter.

The findings of the previous two studies were used to define the methodology that was adopted, ensuring that it was objective, sensitive and valid. The review of cognitive skills performed in Study I (see chapter 2) helped define the structure of the flying task which was used to elicit pilot's manual flying ability and ensure it was valid from a cognitive perspective. The review and evaluation of numerical performance measures undertaken in Study II (see chapter 3) defined the means by which performance information was extracted from the flight data records collected during the exercises.

4.2 Study Aims and Objectives

- Measure the manual handling ability of a cross section of pilots of highly automated airliners on a representative task

- Validate the chosen numerical performance measures by comparing them with TRE derived measures of performance

- Evaluate the causal relationships between the sample's long term and short term manual handling experience and their manual handling performance.

4.3 Methodology

4.3.1 Research Setting

The third and final research study addressed the practical issue of the loss of manual flying skills in the airline operating environment. Consequently it was vital to preserve ecological validity in the research design so that the results of the study could readily be applied to answering the real world problem presented. This requirement constrained the methodological approach that was adopted. Primarily it demanded that flight crew who were currently operational with an airline were sampled and evaluated in a high fidelity environment that reproduced the demands and context of commercial flight. To provide this environment whilst preserving scientific control and consistency it was decided to evaluate pilot performance in a full flight simulator device during a segment of a Line Orientated Flight Training (LOFT) style scenario. The research was presented primarily as a training opportunity to encourage pilot participation and to reduce the effects of testing induced stress on performance (peak performance). However, participants were fully briefed as to the purpose of the research in accordance with ethical guidelines (see appendix F).

A collaborating airline (a UK based 'low fares' operator with a primarily domestic and European route network) agreed to offer the exercise to its Boeing 737 fleet pilots during a half hour simulator session which immediately followed the completion of the crews annual License Proficiency Check (LPC). This arrangement provided a sizable and diverse sample of pilots within a reasonably short period of time. However, it should be noted that since all sampled pilots had successfully demonstrated fundamental manual handling elements during the preceding proficiency check the range of their performance would likely be somewhat normalised.

Consequently the study aims to discriminate the relatively narrow spread between desirable manual flying performance and merely tolerable manual flying performance (i.e. performance which satisfies the absolute minimum safety criteria but should not serve as an example of good manual flying). However, a methodology that can successfully discriminate this small difference has promise for application to the population where performance spread is likely to be much wider since it demonstrates high sensitivity. It should also be noted that any performance observed during simulator testing is likely to exceed real world performance, especially during abnormal circumstances (Baker and Dismukes, 2002). There are of course obvious ethical considerations which surround assessing the performance of licensed air transport pilots. These and other practicalities precluded an alternative research design in this instance but future studies may benefit by assessing performance just prior to the license proficiency check should this be possible.

4.3.2 Participants

Research participants were 66 professional pilots sampled over a period of four months. All pilots held an Air Transport Pilot License (ATPL) and a Boeing 737-300/400/500 type endorsement. The trial was run concurrently with the airline's annual License Proficiency Check (LPC) programme and participants were recruited as they presented themselves for this simulator session. The airline's Flight Crew Scheduling department allocated crew members in pairs to the various LPC sessions based on their requirement for revalidation which in turn was a function of their initial date of employment and roster availability. Research sessions were administered on those occasions where simulator availability allowed an additional half hour session to be added to the standard four hour LPC session (the simulator was highly utilised by several airlines and not all LPC sessions could be extended to facilitate the research). The criterion for participant selection was therefore convenience based but was not knowingly influenced by the individuals manual flying experience or performance (no participants were undergoing an LPC re-test).

4.3.3 Ethical Considerations

Prior to commencement of data collection all of the airlines pilot's received an email explaining that the research was taking place and inviting them to participate if they wished to do so. It was made clear that participation was voluntary and that all data were made anonymous and would be held securely and in confidence at Cranfield University. They were provided with further details during a meeting in person the day prior to the research exercise. The research exercise was scheduled to run after the completion of the two day LPC items and once the candidate had received their result. Participants were informed that they could withdraw at any stage. All work was approved by the Cranfield University School of Engineering Ethics committee and conformed to the British Psychological Society's guidelines for ethical treatment of participants (see Appendix F for a reproduction of the approved ethics proposal). The work was presented to and supported by several of the major pilots union groups.

4.3.4 Expert Observation

During the research simulator exercises a qualified Boeing 737 TRE occupied the instructor's station. The TRE was responsible for administering the test scenario (including configuring the simulator and data collection equipment via the instructor's operating station) and making observational assessments of the crew's performance. The TRE performance assessment data were collected in order to gauge the convergent validity of the flight data derived performance measures i.e. two different methods of measuring the same performance qualities are compared to check that they are in agreement. Additionally, contrasting the two data sets gave a practical scaling to the flight data derived measures i.e. TRE derived mean score values for tolerable & desirable performance.

The TRE made a global assessment of the handling pilot's manual flying ability by allocating a score on a behaviourally anchored Likert rating scale. The behavioural descriptions which anchored the scale were derived from expert accounts of strong and poor manual handling practice gathered during the cognitive task analysis phase (see Chapter 2). The scoring scale was designed to parallel a rating scale already in use with the candidate airline so that the requirement to train TREs in its administration would be minimised and inter-rater reliability would be maximised. The scoring scale is reproduced at Appendix G. Essentially scores of one or two represented tolerable, but limited manual handling ability i.e. a standard which is not necessarily placing the aircraft in danger but should not serve as an example to others. Scores of three or above all represented varying degrees of desirable manual flying ability, i.e. a standard which the airline is aiming to achieve.

An additional TRE (not associated with the candidate airline) occupied the second simulator observer seat during a random selection of the research exercises. The second TRE was trained in the administration of the observational scoring scale and gave parallel but independent ratings of the participant's manual flying performance. These data were subsequently used to establish the level of inter-rater reliability achieved in administration of the rating scale.

4.3.5 Equipment

Research trials were conducted on a Hughes Rediffusion Simulation Ltd. Boeing 737-300 (see figure 14) full flight simulator equipped with six degrees of freedom hydraulic motion cueing system, hydraulic control loading, 150 degree wide day/night/dusk capable visual display and Honeywell SP300 Auto-flight and Flight Management System (see Appendix H for a reproduction of this aircrafts flight deck layout). The simulator was approved to JAR STD 1A Level D. A customised lesson plan was incorporated into the instructor's operating system (IOS) to manage the research scenario. A simple IOS button selection reconfigured and positioned the simulator to the start of the exercise, initiated the flight data collection procedure and terminated it when the exercise was complete. The data collection procedure ran on a Gould computer integrated into the simulator system. Proprietary format data were recorded in real time and then converted off-line into ARINC

717 format raw bit-stream data. ARINC 717 is the standard data format for flight data and quick access recorders and hence is supported by most flight data analysis software making the collected data portable for analysis.

4.3.6 Manual Handling Experience Measures (Independent Variables)

A pro-forma was developed to gather demographic and flying experience data from the participating pilots (see Appendix I). These data served as the basis for the measures of pilot manual flying exposure (the independent variables of the study). The study aimed to determine the effects of recent and long term manual flying exposure on performance and so correspondingly the pro-forma gathered data relating to recent and long term manual flying experience. Proximal manual flying exposure was estimated from the number of extended manual approaches the pilot had conducted within the previous month.

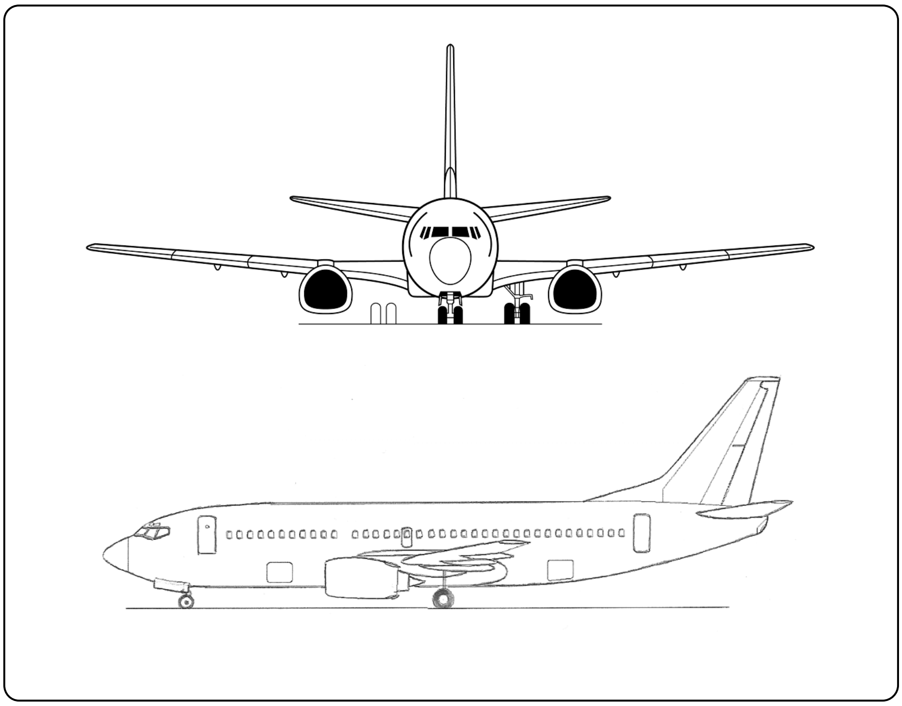


Figure 16 - Front and Profile view of the Boeing 737-300 aircraft which formed the platform for the review of manual flying skills. Notably the aircrafts under-slung engines cause a considerable thrust-pitch coupling. (image adapted from Janes, 2000)

For this purpose an extended manual approach was defined as an approach whereby the autopilot was disconnected prior to or shortly after reaching the final approach point (FAP). Distal manual flying exposure was evaluated from the pilot's accounts of their career history. The list of aircraft types the pilot reported to have flown were divided into highly automated and manual groups (see assumed definition in Chapter 1). The number of hours spent flying manual types was used as a measure of manual flying exposure when set against their total number of commercial flying hours.

In addition to these primary data the pro-forma also collected information about other individual differences which may impact manual flying experience so that their effects could be controlled for or explored in the subsequent analysis. These factors included the non-commercial flying activities of the participant (i.e. general or sport aviation), the training route of the participant and a self evaluation of their own manual flying ability, as well as the standard demographic information.

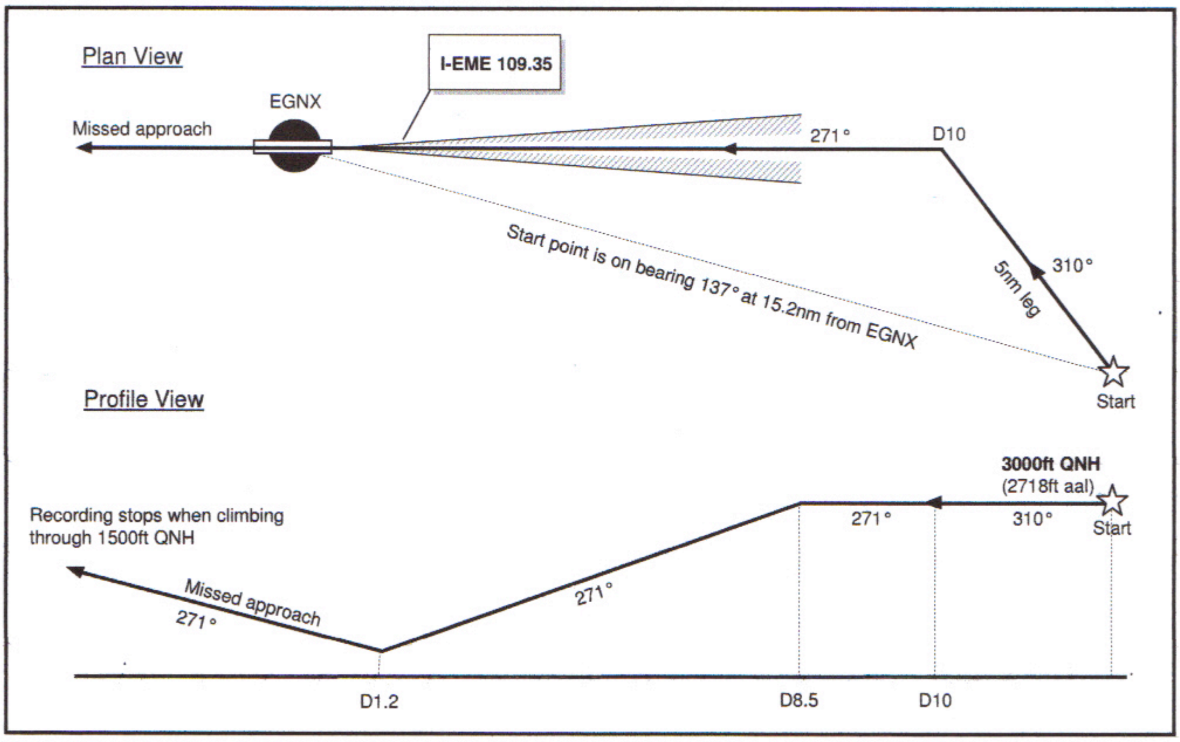
4.3.7 Manual Flying Task

Practicalities precluded the evaluation of a full descent, approach and landing profile and consequently many of the cognitive facets of performance uncovered in Chapter 2 could not be fully tested and evaluated. Instead the research focused on the ILS initial approach, ILS and missed approach procedures. However, information gathered during the CTA was used to incorporate appropriate cognitive complexity into these tasks and therefore challenge this aspect of manual flying expertise, enhancing the validity of the exercise.

Participants were each required to perform a standardised terminal manoeuvring exercise in instrument meteorological conditions and with the aircraft in an asymmetric thrust condition. The simulated failure of an engine increased the yaw control demands of the task which was seen to be lacking in previous studies (see chapter 2). It also made the task more practically relevant since earlier data (see introduction) identified that a large proportion of fatal manual handling related accidents occur following the loss of an engine. To force the participants to adopt manual control the autopilot, flight director and auto throttle systems were all made inoperative (the crew were briefed that these systems would not be available throughout the flight).

The task began from a straight and level condition at platform altitude with the aircraft at relatively high speed in a clean configuration positioned to intercept the ILS localiser. From the information delivered via the standardised ATC brief (see figure 15) participants were required to recognise the need to shed energy relatively rapidly, transitioning to a suitable intercept speed and configuration over a fairly short track distance. Whilst performing the single engine ILS a significant backing crosswind further added to the cognitive demand of the manual flying exercise. Finally, weather conditions at decision height prevented visual acquisition of the runway, mandating a single engine missed approach and placing significant lateral and vertical control demands upon the pilot. Precise management of the aircraft’s performance during the single engine missed approach was necessary to achieve the required climb profile as the aircraft was at a relatively high landing weight.

The exercise incorporated a variety of demanding but operationally relevant manual flight tasks in the short period of time available whilst preserving the continuity and line orientated nature of the simulator session. The ILS tracking task has previously been demonstrated to give good discrimination between weak and strong pilots and has a great deal of validity since it is a likely requirement for manual flight in the real world. Importantly the chosen exercise is well defined spatially and thus can be easily measured using the chosen measurement tools. The EFIS was configured to display only raw flight data, with the electronic horizontal situation indicator (EHSI) set to expanded ILS mode, challenging the pilot’s instrument scanning discipline. The addition of a backing crosswind, asymmetric thrust condition and tight energy constraints were intended to challenge the cognitive processes that would be encountered during real world manual flight operations and are considered by experts to underpin manual flying performance (see chapter 2).



Initial Conditions

Aircraft

Weight ~ 45,000kg
Speed 210kts IAS
Track 310°
Position as indicated above
Straight & level in clean configuration
Engine #2 shut down and safe
Manual control i.e. autopilot, autothrottle & flight director disengaged

Weather

3,000ft wind = 310°/20
Surface wind = 330/15
Surface temp = 5°
QNH 1013mb
Visibility = 400m
Overcast 100ft
Overcast 3500ft

N.B. These will be the actual weather conditions setup in the simulator. The cloud base & visibility conditions reported to the crew when commencing the approach should be above minima i.e. cloud base 300ft and visibility 1,200ft.

Scenario Outline

The flight is positioning for an ILS approach to EGNX runway 27 following a starboard engine failure. The engine has been safely shut down and the associated checklists have been completed.

The autopilot, autothrottle and flight director systems have become unserviceable, necessitating a manually flown approach.

The aircraft is straight and level on a radar heading of 310° at an altitude of 3000ft. ATC have cleared the aircraft to maintain this altitude and heading to intercept the localiser from the left and then to descend with the glideslope. ATC have also cleared the flight to reduce speed at the crews discretion.

At decision height the weather conditions will be below minima necessitating a missed approach. ATC request that the crew fly the published missed approach procedure, climbing straight ahead to 3000ft.

Figure 17 - diagrammatic representation of the task scenario

4.3.8 Experimental Procedure

The TRE who had conducted the participating crew’s LPC prior to the research session also undertook the role of co-ordinator and observer for the research exercise. Following a refreshment break the TRE invited the participating crew into the simulator and delivered a joint research and flight briefing. Participants were asked to fly manually a terminal area exercise (see figure 15) which would commence from straight and level flight at platform altitude and include a raw data ILS approach in IMC. They were informed that the exercise would be undertaken with the starboard engine shut down and secured and without the availability of either the autopilot, flight director or the auto-throttle systems.

Pilots were randomly assigned the duties of pilot flying (PF) and pilot monitoring (PM). The crew were asked to perform the exercise as they would during an operational flight with the PM making the standard company calls, operating the radios and calling the checklists. The TRE initiated the lesson plan through the IOS which configured the simulator into a ‘frozen’ state at the initial point of the research exercise. The crew were given the necessary approach plates, weather briefings and reference data and allowed time to brief and orientate themselves (the scenario assumed that all necessary QRH items had been performed to secure the non-operating engine). With the agreement of the crew the simulation was unfrozen in time but held in geographical position to allow the PF to transition to flying the aircraft in its asymmetric thrust condition. When the PF indicated they were ready the TRE delivered a scripted ATC clearance;

> *“Maintain 3000ft on this heading to intercept the localiser, reduce speed at your discretion, approximately 15 miles to run. Once established cleared to descend with the glideslope”*

When the crews read back had been confirmed the simulation was released from its position freeze and the flight was allowed to progress. Reported weather conditions were marginally above the approach's CAT1 minima allowing the crew to proceed legally with the approach but making the outcome at decision height unpredictable. As the aircraft passed 4nm DME inbound on the ILS the TRE delivered a landing clearance along with an update of the surface wind conditions (moderate with a crosswind component from the right). The simulated overcast cloud base was set slightly lower than the reported height and below decision height, requiring that the crew execute the missed approach procedure (straight ahead to 3,000ft) owing to a lack of visual contact with the runway. The lesson plan automatically froze the simulation and terminated data collection as the aircraft climbed through the acceleration altitude (1,500ft QNH). Following the experimental run the TRE completed the behaviourally anchored rating form, assessing the manual handling performance of the PF on the exercise (see Appendix G).

Figure 18 – Exemplar animation of flight data using FlightScape Insight Animation

The simulation was then reconfigured back to the start of the exercise and the crew were asked to swap PF/PM roles and re-brief themselves for the same approach. The exercise was run again as before, although the crew were informed that the weather conditions may or may not require a go-around (although environmental conditions had in fact not been changed thus necessitating a missed approach). The order in which crew members were assigned the duty of PF was varied randomly over the trials. Following the simulator exercise participants were moved to a de-briefing room and asked to complete the demographic pro-forma which included information about their career history as well as recent and long term manual flying experience. Participants were then thanked for their contribution and given an experimental debrief which included time for feedback and discussion.

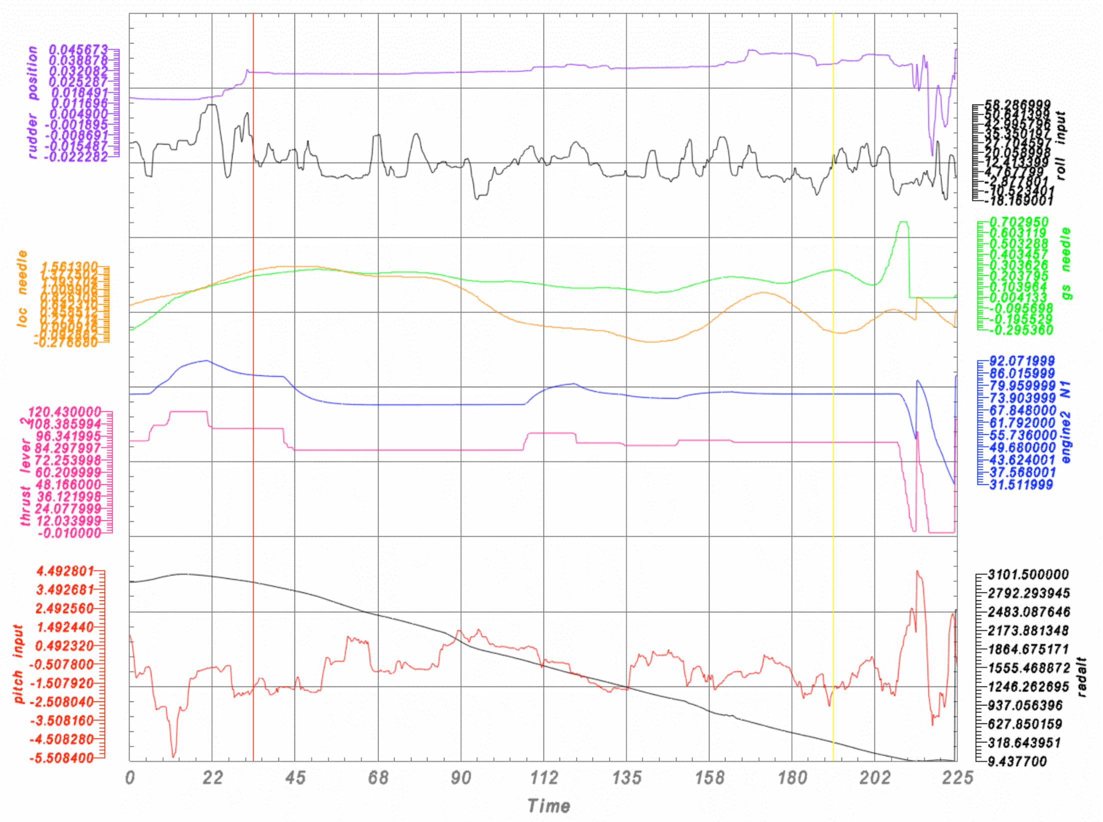


Figure 19 – Exemplar Plots of flight data produce by FlightScape Insight Analysis

4.3.9 Flight Data Derived Performance Measures (Dependant Variables)

Raw ARINC 717 format bit stream data were imported into the Flightscape Insight analysis suite. The data were animated and plotted (see figure 16 and 17) to allow for an initial visual screening of the flights and the detection of any gross anomalies. An event set was defined and run within the analysis software to segment each flight into the three distinct tasks, ‘straight & level’, ‘ILS tracking’ and ‘Missed Approach’. Segmented data files were exported as comma delimited text files (.csv) and passed to the Matlab data analysis suite for further processing. A bespoke M-file (programme script file) was used to compute the battery of performance metrics (see table 5) for each flight segment. The performance metric values were then assembled in an SPSS data file alongside the participants corresponding biographical and career data and the TRE performance rating data for statistical analysis.

4.4 Analysis & Results

4.4.1 Demographic Data

The participating pilots reported widely varied flying experience (see table 6), ranging from young first officers operating on a frozen ATPL, having just completed their initial type rating training, to seasoned training captains holding close to 18,000 flying hours. Generally the samples total commercial flying experience was normally distributed around a mean of 5,887 hours with a standard deviation of 3,839 hours.

The number of hours operating experience of highly automated aircraft (see definition chapter 1) ranged within the sample from 300 hours to 11,500 hours, with a mean of 3,597 hours and a standard deviation of 2,803 hours. Importantly, the proportion of automation to total flying experience varied considerably amongst the sample, being largely dependent upon the career path of the individual prior to taking up their appointment at the host airline. There was thus ample variability in automation exposure with which to contrast against any observed variance in manual flying performance.

Table 6 - Sample Demographic & Career Background

N=66, 32 First Officers (49%), 27 Captains (41%), 7 Training Captains (11%)				
	Min	Max	Mean	Std. Deviation
Age (yrs)	25	56	40	9
Total Commercial Experience (Hrs)	500	17900	5887	3839
Automated Aircraft Experience (Hrs)	300	11500	3597	2803
Boeing 737-3/4/5 Experience (Hrs)	300	8500	2269	1722
Private Flying Experience (Hrs)	0	5000	555	1190
Sectors Flown in Past Month	0	60	25	15
Manual Approaches Flown in past month	0	10	3	3

The amount of non-airline flying conducted by the sampled pilots was also highly variable. The majority of the sample (72%) reported that they had undertaken no significant flying activity outside of their occupational duties. However, amongst the remaining crew there was considerable variation in recreational flying experience up to a maximum of 5,000 hours (mean 555 hours, standard deviation 1,190 hours). This non-airline flying activity included touring and instructing in light aircraft and helicopters, as well as performing sports aerobatics in light powered aircraft and gliders.

When completing the demographic pro-forma pilots were asked to consider if they felt their manual flying ability had been influenced by the experience of operating a highly automated aircraft. In response 77% of the sample indicated that their skills had deteriorated. However 16% of participants felt their skills had not been affected, whilst 7% believed their skills had actually improved, although many of the latter were young pilots who noted that the improvement was probably due to a large proportional increase in their flying experience on large transport types.

4.4.2 Flight Data Performance Measures

The tables in appendix J show the distribution of flight data derived performance measurements for the whole sample, separated by the three flight phases of the research exercise (straight & level, ILS tracking, and missed approach).

4.4.3 TRE Performance Assessment

The TRE observational scoring of general manual handling ability was normally distributed with a slight negative skew (-0.17) resulting in a mean score of 3.41 and standard deviation of 1.03 (see Appendix G for a full description of the scoring scale). However the skew was modest and did not threaten the assumed normality of these data (see figure 18).

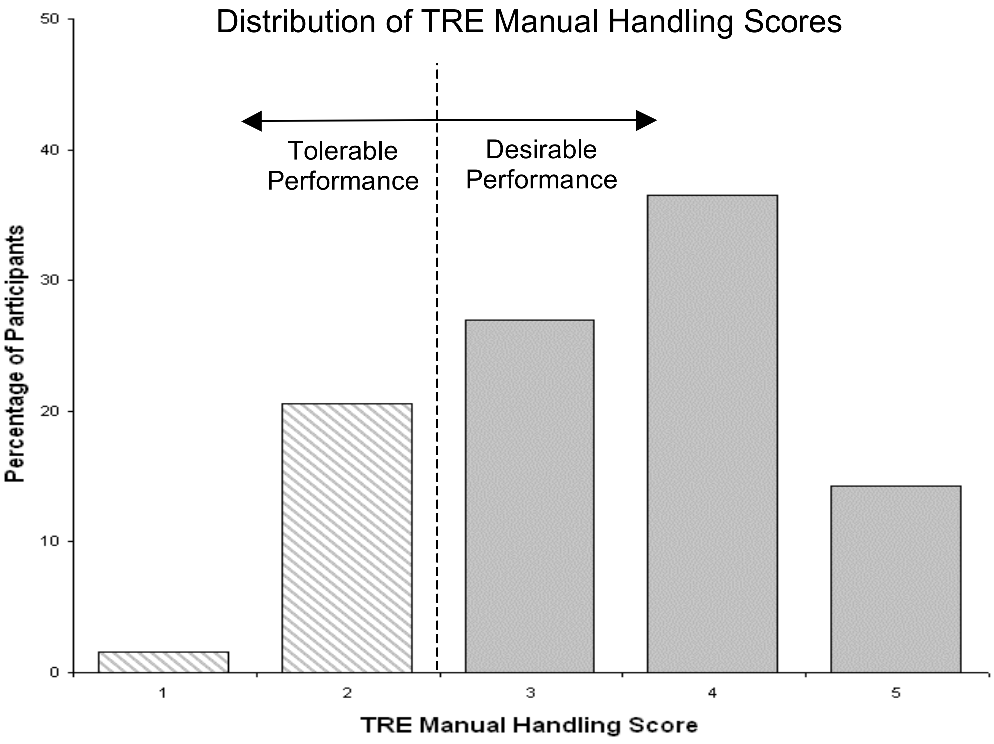


Figure 20 - TRE scoring distribution

Seventy eight percent of the sampled pilots demonstrated desirable manual handling skills whilst 22% demonstrated tolerable but notably weaker manual flying ability. It should be emphasised that during the preceding LPC no participant was awarded an overall fail grade, thus bounding the minimal level of performance observed as tolerable.

Parallel TRE rating was conducted for 24% of the simulated flights (16 cases) to assess the level of inter-rater reliability in scale administration. Cohen's Kappa was computed as a statistic of inter-rater reliability. A value of 0.818 resulted, indicating that the TREs had achieved 'almost perfect agreement' (Landis & Koch, 1977) in discriminating between tolerable and desirable manual flying performance amongst the sample of pilots (see table 7).

Table 7 - Inter-Rater Reliability Assessment of TRE grading using Cohen's Kappa

TRE Inter-Rater Reliability

		Principle TRE		
		Low	High	Total
Secondary TRE	Low	19%	0%	19%
	High	6%	75%	81%
	Total	25%	75%	100%

low = performance approaching lower limits of tolerance (TRE defined)

high = Desirable erformance (TRE defined)

Cohen's K = 0.818

Table 8 - Correlations between performance metric score and TRE score on the straight and level segment of the exercise (those shaded are statistically significant to p<0.05)

TRE-Performance Metric Correlations (Straight & Level)	r	N	p
Spatial Tracking Error			
Altitude ME (ft.)	0.181	50	0.107
Altitude SDE (ft.)	-0.267	50	0.032
Heading ME (deg.)	0.010	50	0.472
Heading SDE (deg.)	-0.183	50	0.104
Control Wheel Power (degs²/hz)			
Very Low Frequency Band	-0.173	50	0.117
Low Frequency Band	-0.113	50	0.220
Mid Frequency Band	-0.269	50	0.031
High Frequency Band	-0.147	50	0.157
Very High Frequency Band	-0.163	50	0.132
Control Column Power (degs²/hz)			
Very Low Frequency Band	-0.076	50	0.302
Low Frequency Band	-0.149	50	0.153
Mid Frequency Band	-0.222	50	0.063
High Frequency Band	-0.043	50	0.384
Very High Frequency Band	-0.201	50	0.084
Rudder Power (degs²/hz)			
Very Low Frequency Band	0.162	50	0.133
Low Frequency Band	0.022	50	0.441
Mid Frequency Band	0.044	50	0.381
High Frequency Band	-0.022	50	0.440
Very High Frequency Band	0.118	50	0.210
Throttle Power (degs²/hz)			
Very Low Frequency Band	0.173	50	0.117
Low Frequency Band	0.107	50	0.232
Mid Frequency Band	0.112	50	0.223
High Frequency Band	0.103	50	0.240
Very High Frequency Band	0.142	50	0.166

Table 9 - Correlations between performance metric score and TRE score on the ILS tracking segment of the exercise (those shaded are statistically significant to $p<0.05$)

TRE-Performance Metric Correlations (ILS)	r	N	p
Spatial Tracking Error			
Localiser ME (dots)	-0.297	50	0.025
Localiser SDE (dots)	-0.625	50	0.000
Glideslope ME (dots)	0.284	50	0.023
Glideslope SDE (dots)	-0.413	50	0.001
Airspeed ME (kts)	-0.279	50	0.025
Airspeed SDE (kts)	-0.350	50	0.006
Control Wheel Power (degs²/hz)			
Very Low Frequency Band	-0.315	50	0.013
Low Frequency Band	-0.265	50	0.031
Mid Frequency Band	-0.331	50	0.010
High Frequency Band	-0.432	50	0.001
Very High Frequency Band	-0.080	50	0.290
Control Column Power (degs²/hz)			
Very Low Frequency Band	-0.088	50	0.272
Low Frequency Band	-0.408	50	0.002
Mid Frequency Band	-0.389	50	0.003
High Frequency Band	-0.212	50	0.070
Very High Frequency Band	-0.072	50	0.310
Rudder Power (degs²/hz)			
Very Low Frequency Band	0.092	50	0.263
Low Frequency Band	-0.068	50	0.319
Mid Frequency Band	-0.112	50	0.220
High Frequency Band	-0.199	50	0.083
Very High Frequency Band	-0.035	50	0.404
Throttle Power (degs²/hz)			
Very Low Frequency Band	-0.008	50	0.479
Low Frequency Band	-0.170	50	0.119
Mid Frequency Band	0.041	50	0.389
High Frequency Band	-0.010	50	0.472
Very High Frequency Band	0.016	50	0.457

Table 10 - Correlations between performance metric score and TRE score on the missed approach segment of the exercise (those shaded are statistically significant to $p<0.05$)

TRE-Performance Metric Correlations (Missed Approach)			
	r	N	p
Spatial Tracking Error			
Track ME (deg.)	-0.092	50	0.265
Track SDE (deg.)	-0.221	50	0.063
Airspeed ME (kts)	-0.038	50	0.397
Airspeed SDE (kts)	-0.118	50	0.209
Control Wheel Power (degs^2/hz)			
Very Low Frequency Band	-0.054	50	0.356
Low Frequency Band	-0.215	50	0.069
Mid Frequency Band	-0.163	50	0.132
High Frequency Band	-0.140	50	0.169
Very High Frequency Band	-0.062	50	0.337
Control Column Power (degs^2/hz)			
Very Low Frequency Band	0.350	50	0.007
Low Frequency Band	-0.098	50	0.252
Mid Frequency Band	-0.262	50	0.035
High Frequency Band	-0.168	50	0.124
Very High Frequency Band	-0.187	50	0.100
Rudder Power (degs^2/hz)			
Very Low Frequency Band	0.042	50	0.387
Low Frequency Band	-0.290	50	0.022
Mid Frequency Band	-0.028	50	0.423
High Frequency Band	-0.134	50	0.179
Very High Frequency Band	-0.210	50	0.074
Throttle Power (degs^2/hz)			
Very Low Frequency Band	0.140	50	0.169
Low Frequency Band	0.041	50	0.390
Mid Frequency Band	0.194	50	0.091
High Frequency Band	0.258	50	0.037
Very High Frequency Band	0.271	50	0.030

4.4.4 Validating the Objective Performance Measures against TRE Scores

Measurement Convergent Validity

Bivariate correlations were performed for each performance metric against the TRE derived manual handling score (see tables 8, 9 and 10). This analysis tested the convergent validity of the measures since theoretically the TRE and numerical assessments of manual handling performance should be convergent.

In general the performance measures computed from the ILS tracking segment of the exercise were highly correlated to the TRE derived manual handling scores (see table 9). All measures of outer-loop performance (localiser, glideslope and airspeed tracking accuracy and smoothness) were negatively correlated with TRE measures of manual handling ability for this flight phase. Thus greater accuracy and smoothness (smaller numerical scores) of tracking was convergent with higher TRE scores.

Performance measures of control strategy were also well correlated to TRE score on the ILS tracking segment of the task (see table 9). Measures of control wheel input power in the very low, mid and high frequency bands were all significantly negatively correlated with the TRE manual handling score. Therefore, lower control input power values converged with higher TRE scores. Furthermore measures of control column input power in the low and mid frequency bands were also negatively correlated with TRE manual handling score. However no measure of rudder or throttle control strategy appeared to be significantly related to the TRE score for manual handling ability.

Performance measures derived from the straight and level segment of the exercise generally did not correlate strongly with the TRE derived measure of manual handling ability (see table 8). However, altitude tracking smoothness and control wheel input power in the mid frequency band were significantly negatively correlated with TRE manual handling score. No further metrics demonstrated significant correlation for this segment.

Similarly, few performance measures computed from the missed approach segment of the task showed significant correlation to the TRE derived manual handling score (see table 10). Those which did show significance were all measures of control input strategy. Rudder input power in the low frequency band and control column input power in the mid frequency bands were both significantly negatively correlated to the TRE manual handling score.

Measurement Sensitivity

The sensitivity of the objective performance measures was assessed by performing independent t-tests between the TRE derived ‘tolerable’ and ‘desirable’ manual handling performance groups (see table 11). Essentially this analysis tests the metrics ability to separate the two different performance groups. Many of the metrics computed for the ILS tracking task were sufficiently sensitive to separate the two groups with a high degree of confidence. Localiser error, Glideslope error and Airspeed error variability (outer-loop performance measures) values were all significantly smaller for the ‘desirable’ group indicating a higher standard of tracking smoothness. However it should be noted that all measures of outer-loop tracking accuracy (Localiser, Glideslope and Airspeed ME) failed to be separated with statistical confidence.

Also on the ILS tracking segment, measures of control wheel input power in the high, mid and low frequency bands, as well as control column input power in the low frequency band, were sufficiently sensitive to successfully separate the TRE assigned performance groups. The only remaining measure which showed such sensitivity was rudder input power in the low frequency band which was computed during the missed approach segment. In this case values of rudder input power were significantly lower for the ‘desirable’ performance group.

Table 11 - Statistically significant (p<0.05) t-test results for each performance metric between TRE assigned performance groups.

Significant t-test between acceptable and desirable performance

	Tolerable		Desirable				
	M	SD	M	SD	t	df	Sig.
Straight & Level							
No Metrics	n/a	n/a	n/a	n/a	n/a	n/a	n/a
ILS Segment							
Localiser SDE	0.055	0.030	0.023	0.013	5.209	48	0.000
Glideslope SDE	0.015	0.015	0.010	0.010	2.906	48	0.006
Airspeed SDE	0.347	0.176	0.241	0.115	2.431	48	0.019
Control Wheel Low Freq	11951	8022	4775	5085	3.680	48	0.001
Control Wheel Medium Freq	9739	12872	3556	2256	2.885	48	0.006
Control Wheel High Freq	13580	9965	6128	4040	3.786	48	0.000
Control Column Low Freq	18	11	10	6	2.901	48	0.006
Missed Approach							
Rudder Low Freq	104	38	75	36	2.378	48	0.022

Table 11 gives the mean group scores for each performance metric which attained a statistically significant t-test value and thus demonstrated strong sensitivity. These scores can assist in setting an appropriate ‘cut score’ which divides ‘tolerable’ from ‘desirable’ manual handling performance on each of the measured dimensions. It is notable that significance was generally limited to those measures derived from the ILS tracking segment. As noted previously (see section 2.1) this flight segment is highly dynamic, with transitions into different states, and of greater cognitive complexity than the other flight phases and as such more demanding of the pilot’s skill.

Multivariate Measurement Sensitivity

The previous analysis demonstrated the sensitivity of each performance metric individually i.e. a univariate analysis. By combining the metrics their sensitivity as a whole may improve. The following analysis aims to build a weighted linear combination of the metrics and assesses whether this has sufficient sensitivity to categorise the pilots into the same performance groups assigned by the TREs. The unique variance explained by each metric will thus be critiqued. The strength and fit of the resulting model will provide a further and more robust validation of the performance metrics convergent validity and sensitivity. Also, such a model may prove an effective way of combining the metrics to provide a useful global measure of manual flying ability, rather than many measures of individual performance dimensions. If the strength and reliability of the resulting model is high then it has potential for application as a multivariate manual handling performance measurement tool for FDM and similar programmes or in future research.

A Binary Logistic Regression (LR) procedure was performed using the SPSS software package. A forward LR stepwise entry method was selected so that performance metrics would be either added to or removed from the model based upon their individual predictive ability. The conventional outer-loop tracking metrics and the control strategy metrics were entered in two separate blocks. The resulting model contained four predictors and a constant. The predictor variables were Localiser SDE (tracking smoothness), Airspeed SDE (tracking smoothness), Control Wheel Input Power in the Low Frequency Band during the ILS, Rudder Input Power in the Low Frequency Band during the missed approach procedure, and a constant (see table 12). The sample to variable ratio was therefore approximately 16:1.

The non-significance of the Hosmer & Lemeshow test indicates that the model is well fitted to the data. The reasonably high Nagelkerke pseudo R^2 value of 0.712 indicates that approximately 70% of the variability in the TRE discrimination is explained by this model and its four predictor variables (see table 12). Therefore the multivariate model agrees highly with the TRE discrimination.

Table 12 - Logistic Regression Model Parameters

Predictor	β	*SE* β	Wald's X^2	*df*	p	*odds ratio*
Localiser tracking SDE	-82.121	30.067	7.460	1.000	0.006	0.000
Airspeed tracking SDE	-8.178	5.390	2.303	1.000	0.129	0.000
ILS Control Wheel Low Freq Power	0.000	0.000	4.269	1.000	0.039	1.000
Missed Approach Rudder Low Freq Power	-0.041	0.018	5.009	1.000	0.025	0.960
Constant	11.209	3.765	8.866	1.000	0.003	n/a
Goodness-of-fit test			X^2	*df*	*p*	
Hosmer & Lemeshow			13.002	*8*	0.112	
Pseudo R^2			R^2			
Cox and Snell			0.478			
Nagelkerke			0.712			

Table 13 - Logistic Regression Model Classification Rate

Logistic regression model classification rate

		Predicted	
		Tolerable	Desirable
Observed	Tolerable	75.00%	25.00%
	Desirable	6.00%	94.60%

Total Model Classification Rate = 89.9%

The total sample size precluded the formation of a hold out sample on which to validate the model. Consequently validation was performed using the general sample. A cut value of 0.6 was set to reduce the number of type 2 errors since these were less tolerable than type 1 errors in the practical setting of the study (this also helped to remove any biasing which the imbalanced group sizes may have generated). The resulting classification rate of 0.898 (see table 13) indicates that overall approximately 90% of the models predicted scores (desirable or tolerable performance) were in agreement with that given by the TRE. In more detail, allocation to the tolerable performance group was 75% accurate whilst allocation to the desirable performance group was 94.6% accurate. The overall classification rate and goodness of fit data indicates that the model is highly reliable and useful as a predictive tool.

4.4.5 Examining the relationships between Manual Handling Experience and Performance

Bivariate correlations were performed to measure the extent of the relationship between a pilot's general flying experience and their performance on the manual flying task. There was found to be no statistically significant correlation between the total number of flying hours the pilot had accumulated and any dimension of their performance on the assessment (see appendix K)

The previous analysis showed that the total amount of general flying activity did not appear to be a significant influence on the manual flying performance of the pilots. A more detailed analysis was required to investigate whether the composition of that experience did shape their manual flying performance i.e. if that experience was gained predominantly on highly automated types were they likely to perform differently to a pilot who had gained a similar number of hours on manual types?

Although general flying experience did not appear to be a dominant factor in determining manual flying performance, any influence it did have needed to be systematically controlled for when looking in finer detail at the makeup of the pilots flying experience i.e. number of hours on highly automated aircraft, number of recent manual approaches etc. In short, it was necessary to assess the affects of automation exposure fairly across all pilots, regardless of their total operating experience.

Partial correlation analyses were performed, controlling for total flying hours, in order to identify the extent of the relationship that existed purely between manual flying exposure factors and manual flying performance. Partial correlation procedures were performed between the proximal and distal manual flying exposure measures (see section 4.3.6) and the flight data derived manual flying performance measurements.

Table 14 - Statistically significant correlations between number of manual flying hours and performance metrics (controlling for total fixed wing hours)

Measure = Number of Manual Flying Hours Controlling for Total Number of Fixed Wing Hours	Long Term Manual Flying Exposure		
Metric	R	N	p
Straight & Level			
Control Column Input Power in Very Low Freq Band	-0.342	46	0.022
ILS Tracking			
No Significant Correlations	n/a	n/a	n/a
Missed Approach			
Control Wheel Input Power in Mid Freq Band	-0.380	46	0.010
Control Column Input Power in Very Low Freq Band	-0.318	46	0.033
Control Column Input Power in Mid Freq Band	-0.302	46	0.044
Rudder Input Power in Mid Frequency Band	-0.510	46	0.000
Throttle Input Power in Low Freq Band	-0.318	46	0.034

The pilots manual flying performance was somewhat influenced by their long term exposure to automation (see table 14). Those who had spent proportionally more time flying manual types, and thus had lower exposure to automation, generally demonstrated a somewhat different control strategy during the straight and level and missed approach segments of the exercise.

During the straight and level segment long term manual flying exposure was negatively correlated with control column input power in the very low frequency band. During the missed approach segment long term manual flying exposure was negatively correlated with control wheel input power in the mid frequency band, control column input power in the very low and mid frequency bands, rudder input frequency in the mid frequency band and throttle input power in the very low frequency band. However long term manual flying exposure was not significantly correlated with any measures derived from the ILS tracking segment nor any measures of aircraft tracking performance (see table 14).

Table 15 - Statistically significant correlations between recent manual flying exposure and performance metrics (controlling for total fixed wing hours).

Measure = Number of Manual Approaches in the Past Month, Controlling for Total Number of Fixed Wing Hours	Recent Manual Flying Exposure		
Metric	R	N	p
Straight & Level			
Altitude ME	-0.286	46	0.028
Control Wheel Input Power in Mid Freq Band	-0.264	46	0.040
Control Column Input Power in Very Low Frequency Band	-0.263	46	0.040
Control Column Input Power in Mid Frequency Band	-0.381	46	0.005
ILS Tracking			
Airspeed SDE	-0.436	46	0.001
Control Wheel Input Power in Very High Freq Band	-0.249	46	0.047
Control Column Input Power in Mid Freq Band	-0.325	46	0.014
Control Column Input Power in High Freq Band	-0.288	46	0.026
Rudder Input Power in Very Low Frequency Band	-0.275	46	0.032
Missed Approach			
Control Wheel Input Power in Very Low Freq Band	-0.249	46	0.049
Control Wheel Input Power in High Freq Band	-0.381	46	0.005
Control Column Input Power in High Freq Band	-0.344	46	0.010

The number of manual approaches flown in the preceding month was used as an estimate of the pilot's recent manual flying activity. Correlations between this measure and aspects of the pilot's manual flying performance were generally more abundant and stronger than those previously noted, particularly during the ILS tracking segment of the task (see table 15).

During the straight and level segment recent manual flying exposure was significantly negatively correlated with altitude tracking mean error, control wheel input power in the mid frequency band and control column input power in the very low and mid frequency bands. During the ILS tracking segment recent manual flying exposure was significantly negatively correlated with airspeed tracking variability, control wheel input power in the very high frequency band, control column input power in the mid and high frequency bands, and rudder input power in the very low frequency band. During the missed approach segment recent manual flying exposure was significantly negatively correlated with control wheel input power in the very low and high frequency bands, and control column input power in the high frequency band (see table 15). Again it is notable that the more correlated measures occur during the more demanding flight segments, where a greater number of performance elements are subject to change. This point is developed further in the subsequent discussion.

In chapter 3 a table of desirable performance metric attributes was proposed. Table 16 revisits these attributes and identifies how each has been demonstrated, or not, by the selected performance metrics.

Table 16 – Performance metric attributes summarised

Attribute	Descriptive
Reliable	Although every effort has been made to ensure the method by which the metrics are generated is reliable it was not possible to explicitly test their reliability since only one iteration of the experiment was performed.
Valid	The various statistical tests cited show that the measures have demonstrated both uni-variate and multi-variate convergent validity and that the properties that they are measuring are closely aligned with the concept of manual flying performance held by type rating examiners.
Interpretable	The selected metrics each describe an element of performance which is relatively easy to conceptualise, such as control input energy, flight path deviation and as such have high interoperability which facilitated a clear understanding of the observed performance differences.
Sensitive	The performance metrics were able to differentiate between two fundamental levels of performance as judged by type rating examiners and thus was shown to have a useful level of sensitivity.
Applicable	The metrics were computed from data collected during real time simulator evaluations and therefore should be capable of replication in similar simulated environments. However it remains to be seen if such techniques could be adequately applied to data collected from real aircraft and this subject should form significant future research.
General	The metrics have so far been applied to only one aircraft type and operator. Future work is required to evaluate their generalisability.

4.5 Discussion

The results show that within a typical cross section of pilots operating modern highly automated airliners, manual flying ability will vary considerably. Within the sample manual flying ability ranged from that which was only just considered tolerable to that which was considered exceptional. This range of performance was reflected in both the TRE scoring and the numerical performance metrics. This supports the findings of Viellette (1995) who found a similarly broad range of performance in a sample of pilots operating a highly automated type.

It is also apparent from the results that the manual handling exposure of pilots operating these aircraft varies significantly and is not simply a function of overall flying experience as may be expected. Clearly, many of the extremely experienced pilots within the sample had spent a substantial proportion of their career operating ‘manual’ airliners (typically types such as the BAC 1-11, HS Trident and Boeing 707) since they began their flying career prior to the introduction of highly automated types. However, there were also many examples of senior pilots who had spent almost their entire career operating highly automated types. Conversely, there were also many examples of younger pilots who had spent less than 200 hours on highly automated types having spent their initial career operating older manual equipment for smaller airlines or freight carriers. Career paths which lead into the modern automated airliner therefore still vary widely and subsequently generate differences in manual flying experience.

Furthermore the relatively recent manual flying exposure of pilots varies widely (estimated from the number of manual approaches conducted within the past month – see table 6). However, this study cannot directly attribute the causes of this variance. Whilst each pilots recent exposure to manual flight prior to testing would have been influenced by external factors such as their route allocation, seniority, absenteeism or weather conditions, it may also be somewhat shaped by the pilots own attitude towards manual flight. For example, during the cognitive task analysis (see Chapter 2) many pilots reported they felt that manual flying was a ‘catch 22’ situation, implying that if they removed the automatics they were in danger of allowing the aircraft to stray outside of operational tolerances and significantly increase their workload, whilst if they don’t remove the automatics they are in danger of having their manual flying skills decay.

The level of recent exposure to manual flight may be dependent upon the pilot’s attitude to risk, how they perceive the benefits of manual flight and whether they take or seek opportunities to disengage the automatics (see chapter 1). Of course, as highlighted in the introductory review of literature, there are occasions when removing the automatics is perhaps not the most prudent approach, but the circumstances which define when this is appropriate are not rigid. It is thus not surprising that the results show considerable variation between pilots’ exposure to manual flight. It may be prudent to discuss these issues formally in training in order to encourage a measured and consistent practice regarding the deliberate disengagement of the automatics and building of manual handling experience.

4.5.1 Validation and Assessment of Sensitivity of Performance Measures

The correlation analysis aimed to demonstrate that the objective measures were convergent with the TRE measures of manual handling performance and thus that the measures were meaningfully related to the intended theoretical property (manual handling skill).

The correlation analysis showed that the chosen battery of performance metrics was generally very well related to the TRE derived performance scores on the ILS segment (see table 9). This shows that both the outer-loop and the control strategy metrics had high levels of convergent validity. However, as an exception, measures of rudder and throttle control strategy were not significantly correlated to the TRE score.

With the relatively complex demands of the yaw control task (owing to the aircraft’s asymmetric condition) differences in rudder input strategy should have been apparent and it is surprising that this factor did not correlate highly with the TREs' scoring of performance. A possible explanation lies in the simulator's relatively low fidelity reproduction of yawing movements. This may have lessened the TREs' impression of the yaw control strategy during the ILS approach, especially since from the observer's station the view of rudder pedal movements was significantly restricted. Perhaps these elements of the control problem are less critically viewed by the TRE, and may be allowed to deviate with a reasonable degree of freedom.

However, on the missed approach segment measures of rudder control strategy did correlate significantly with the TREs' performance score (see table 10). The application of high levels of asymmetric thrust associated with the missed approach manoeuvre make the adequacy of yaw control strategy more apparent than it is on the ILS segment (see table 9, since poor control results in large amount of swing which is readily observed by the TRE.

The demonstration of convergent validity in this setting supports the findings of the previous metric selection study (see chapter 3) and gives added confidence in the use of the metrics for measuring manual flying skill. It could also be argued that these results demonstrate the validity of the TREs' performance assessment and give more strength to the observations of manual flying variation obtained by Young, Fanjoy and Suckow (2006) which were previously criticised for lacking objectivity.

The sensitivity of the measures was evaluated through their ability to separate pilots into the same performance groups that were defined by the TREs. Independently, on the ILS segment, the control strategy metrics for roll and pitch axes proved to be very sensitive. Roll and pitch metrics discriminated the performance groups with a high degree of confidence (see table 11). However, as before, measures of rudder pedal and throttle movement proved less sensitive on this segment. In contrast, rudder pedal control strategy proved to be a highly sensitive measure on the missed approach segment (see table 11) and again this is attributed the increased lateral control demands of this task.

Whilst measures of outer-loop tracking smoothness were also very sensitive on the ILS segment, measures of tracking accuracy were relatively insensitive. The raw data (see appendix J) shows large variance in SDE scores whilst relatively small variance in ME scores, which indicate that average ILS tracking error was close to zero on most occasions. Effectively, whilst the degree to which pilots deviated from the ILS datums varied considerably, these deviations would generally always be centred around the datum, generating low ME values. The use of separate measures for smoothness and accuracy, rather than a combined RMSE measure, allows for this deeper investigation of performance. This result supports the arguments of Hubbard (1987) in the adoption of SDE and ME over just RMSE.

The sensitivity of the localiser SDE, airspeed SDE, control wheel LF and rudder pedal LF measures in combination was high, indicated by the multivariate models ability to successfully categorise the pilots into the performance groups. (see table 13) The model explains a considerable amount of the total variance (see table 12) and demonstrates that many of the metrics have a high degree of unique variance. Specifically it demonstrates that measures of outer-loop tracking and control strategy can produce a highly sensitive and reliable measure of pilot manual flying skill when used in combination.

In general the sensitivity of the control strategy frequency analysis derived metrics surpassed that of the more traditional outer-loop performance measures. This further supports the hypothesis put forward in chapter 3 and again gives credit to the argument that measures of outer-loop tracking performance should be augmented by measures of control strategy (see Ebbatson, Huddlestone, Harris and Sears, 2006).

4.5.2 Relationships between Manual Handling Experience and Performance

The results indicate that the amount of general flying hours accumulated by a pilot is not a good indication of their ability to manual fly a large transport aircraft in challenging circumstances (see table 14). Similarly the general experience of the pilot on the specific type of aircraft flown and the training route they undertook to obtain their Air Transport Pilots Licence were poor predictors of manual flying performance.

The results show that the long-term accumulation of manual handling experience throughout a pilot's career had only a moderate effect on their manual flying performance on the task (see table 14). There was no impact upon outer-loop tracking performance but moderate effect upon the control strategy applied (see table 14). These differences tended to become apparent during the missed approach manoeuvre where control input power in the pitch, roll and yaw axes were reduced. With significant pitch and yaw effects coupled to the application of asymmetric thrust during the go-around, it is critical that the pilot makes timely and appropriate inputs to the flying controls. The results suggests that pilots with a greater foundation of manual flying experience were able to anticipate these control requirements as they commenced the missed approach and thus could make more refined, lower power control inputs (see table 14). Those with less manual flying experience failed to 'get on top' of the error and needed to make more inputs to control the aircraft, thus demonstrating higher control input powers. This is comparable with the findings of McDowell (1978) and Rantanen et al (2004).

The missed approach segment appeared to be the only portion of the task with sufficient control demand to draw out the differences in manual flying skill attributable to long-term differences in manual flying experience. However, the result confirm Baron's (1988) hypothesis, and the findings from earlier sections of this research program, that control strategy measures are a sensitive indicator of performance.

The effects of recent exposure to manual flying upon manual handling performance were more pronounced (see table 15). Differences in the amount of manual flying the pilot had accumulated within the preceding two months influenced control strategy over all the task segments (see table 14). In all cases the level of control input power reduced with increased experience demonstrating a more refined control strategy. Again, these results compare favourably with those of McDowell (1978). Furthermore the effect on performance was sufficiently strong to cause notable differences in altitude and airspeed tracking performance (see table 14).

The correlation with airspeed tracking smoothness was strongest, showing that pilots who had accumulated more manual handling time within the past two months had better control over the speed of the aircraft on the approach. The result suggests that airspeed control is perhaps more vulnerable to decay than other aspects of manual flying skill. This reflects results obtained in chapter three of this research programme. It also supports evidence from accident data analyses (see Chapter 1, CAP 776) that many fatal accidents and less severe incidents which are attributed to manual handling deficiencies result from a lack of adequate airspeed control. The results help to explain the broad range of performance amongst pilots of highly automated aircraft that Villette (1995) measured but was unable to account for. It is notable that the distribution of performance issues is in line with the distribution of task complexity and demand over the flight phases. As outlined in chapters 1 and 2 the final approach phase is considerably more complex than other phases with many transient control aspects, requiring a more sophisticated mental model structure to resolve. Thus control issues during this phase appear in all axes, rather than just the yaw axis as with the missed approach phase. It highlights the possibility of cognitive capacity failures as a probable explanation for manual flying issues, with poor quality mental models demanding excessive bandwidth to process the complex approach control problem and resulting in aspects being attended to with insufficient frequency.

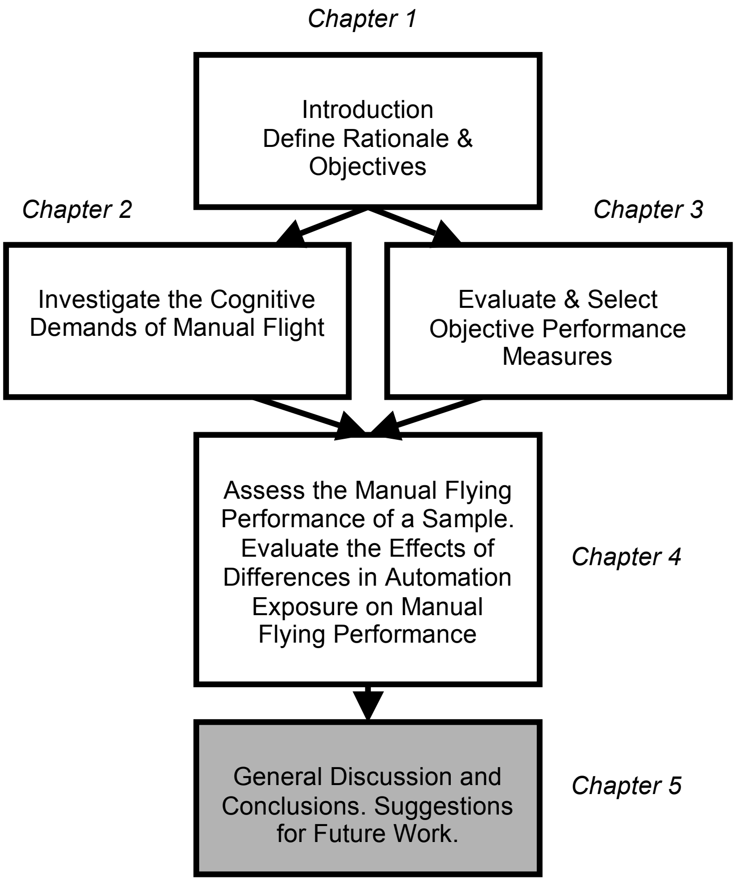
Manual handling performance appears to be influenced more by the amount of manual flying the pilot has undertaken in the few weeks preceding the test than it is by the amount of manual handling experience the pilot has accumulated throughout their career. In essence, the ‘recency’ of the pilot’s manual flying experience is a critical influence upon their manual flying proficiency. It perhaps indicates that manual flying skills decay quite rapidly towards the fringes of ‘tolerable’ performance without relatively frequent practice. Significantly this means that quite a broad range of pilots are susceptible to manual flying skills fade. Previous sections of this work (see chapter 2) have highlighted the importance of anticipatory control over the aircraft, through use of well developed mental models, in order to reduce control input and mental capacity demands during highly transient manoeuvres, such as the ILS and missed approach. The results appear to indicate that increasing the frequency of practice of these highly transient manoeuvres, either in flight or through simulation, could significantly enhance performance, but again a proper training framework needs to be developed to avoid unsafe and inappropriate practice occurring on the line.

The results indicate that for many pilots, even having spent a considerable proportion of their flying career on manual types is not an adequate defence against skills fade if they haven’t managed to practice those skills recently. These results may possibly explain why Viellette (1995) observed such widely varied performance in his sample of pilots of highly automated aircraft.

The results of this study support anecdotal and subjective evidence of the loss of manual flying skills put forward in the introduction to this thesis (e.g. Curry, 1985, Weiner, 1989). It is prudent to recall that those studies, conducted from the late 1970s through the early 1990s, showed pilots to be concerned over potential skill fade but confident enough to still practice manual flight relatively frequently. The results of this thesis indicate that such practice was likely sufficient to prevent any significant skill decrement. In contrast, Wood (2004) notes that the modern air transport environment often limits the opportunity for pilots to disengage the automatics and exercise their manual flying skill (see also the introductory chapter to this thesis) drawing added emphasis to the findings of this work.

Chapter 5

General Discussion & Conclusions



5.1 General Discussion

Subjective data and anecdotal evidence suggested that pilots of highly automated airliners may be vulnerable to the loss of their manual flying skills. However, there was insufficient objective data to support this safety concern and guide any remedial action. This thesis forms part of a response to that safety concern and provides more substantive evidence of the extent and causes of the degradation of manual flying skills. The following discussion summarises the findings of the research.

The cognitive task analysis study (see chapter 2) revealed the dominant role of cognition in manual flying skill. Expert pilots reported using highly refined mental model structures and heuristics in order to predict the performance of their aircraft in its dynamic environment. The study found that the level of refinement of these models is closely linked to the performance achieved in manual flight. Pilots reported using advanced meta-cognitive skills to isolate elements of the control problem, reducing its complexity, and narrow their information gathering scan and reducing cognitive workload. The results support the work of Moray (1999) and Sarter et al (2003) who found that expertise was closely linked to mental model structure. The development of effective cognitive mechanisms which place little demand on the pilots mental capacity may therefore be as important as the development of robust motor-schema in achieving manual control of the aircraft. The study suggests that when measuring manual flying performance careful consideration should be given to designing a task which challenges the cognitive aspects of performance as well as the physical aspects.

The second study evaluated performance measurement techniques and selected a battery of metrics suitable for analysing manual flying skill. The relatively coarse ‘event’ type measures used within the flight data monitoring environment were considered to be insufficiently sensitive for this purpose. The study recognised that owing to lags and slow response rates in the control systems of large transport aircraft the pilot’s control input strategy may not be well reflected in the resultant behaviour of the aircraft. Importantly, prior research (Baron, 1988) suggested that the level of input energy employed to manoeuvre the aircraft is an important indicator of skill. Thus, whilst conventional ‘outer-loop’ tracking metrics could measure the ‘product’ of performance (the aircraft’s behaviour), further measures needed to be applied to measure the refinement of the pilots control input strategy. The study evaluated two frequency analysis based methods of quantifying control strategy. Whilst one method (McDowell, 1978) used frequency banding to give a raw interpretation of the pilot’s control input power spectra the other performed a more complex analysis to produce summary measures of the same power spectra (Rantanen et al., 2004). The results of an empirical evaluation of the sensitivity of the two techniques justified selecting the frequency banding technique as a measure of the pilot’s control input strategy and confirmed the hypothesis that more skilled performance was shown by reductions in control input power across all frequency bands.

The third study used the measures from chapter 3 to evaluate the manual flying performance of a sample of air transport pilots. The results of this study indicated that the manual flying performance of a significant proportion of that sample (see 18) was very low and approaching the limits of acceptability defined by type rating examiners. Furthermore, by comparing the performance measures with demographic data and the amount of manual flying undertaken in the weeks preceding the study, it was identified that a lack of recent manual flying skills practise was likely to cause a substantial degradation of manual flying ability, particularly with respect to airspeed management on the approach which was identified as a factor in manual handling related accidents (CAP 776, 2008). The ‘recency’ of the pilots manual handling experience appeared to outweigh any benefit of long term manual flying experience. Generic skills research (see chapter 2) has identified that cognitive skills are more vulnerable to decay than psychomotor skills. Although the study cannot directly provide evidence to show that the pilot’s cognitive abilities have faded, this could perhaps be inferred given that the results indicate that manual flying skills overall are relatively vulnerable to decay. The findings of this study support earlier anecdotal and subjective concern relating to the loss of manual flying skills (e.g. Curry, 1985; Wiener, 1989; Tenney, Rogers and Pew, 1998).

5.2 Conclusions

The research identifies that manual flying skills are vulnerable to decay through a lack of experience, supporting earlier anecdotal and subjective evidence (e.g. Tenney, Rogers and Pew, 1998). The research also furthers other empirical work (Viellette (2005); Young, Fanjoy and Suckow (2006)) by finding that subtle differences in the operational experience of pilots may have pronounced effects on their manual flying ability.

The results suggest that a strong foundation of manual handling experience may be insufficient to guard against manual flying skills decay. In contrast, pilot’s with relatively high levels of recent manual flying experience performed better than other pilots on a manual handling exercise regardless of their longer-term manual flying background. The benefits of manual handling ‘recency’ thus appear to be considerable, drawing emphasis to observations (Wood, 2004) that such practise is becoming increasingly difficult to accommodate in the modern air transport environment.

More specifically, the results of this work demonstrate that airspeed control on the approach is significantly improved in pilots who have gained more recent manual flying experience. Given the dominant role of poor airspeed control in manual handling related accidents (CAP 776, 2005) this research advocates frequency of manual handling practise as an effective measure against such events. However, as in all such cases a holistic view must be taken. It is the task of regulators and operators to consider the balance of risk and consider whether other areas of safety may be jeopardised by advocating increased manual flying practise.

The research also shows that a sensitive and reliable measurement technique is required to properly evaluate manual flying performance. The results demonstrate that frequency based metrics can provide a sensitive measure of pilot performance in air transport aircraft when applied directly to control input data, supporting the views of Baron (1988). Such metrics enable a broader analysis of pilot performance, reporting on the ‘process’ by which the pilot achieved control of the aircraft, and complimenting more traditional metrics which report the ‘product’ of performance. Furthermore, the results of this research show that measures of control strategy can be more sensitive to changes in manual handling performance than the more traditional ‘outer-loop’ tracking measures, and when acting in combination the result is more sensitive still.

Finally, it is worth noting that although this research has primarily focussed on the plight of manual flying skills, there is reason to be encouraged. At the time of writing an unusual incident had caught the attention of the aviation community and the wider media. A highly automated Airbus A330 airliner experienced a flight control computer anomaly during cruising flight which caused it to adopt a highly unusual nose down attitude. The crew’s timely response was to disconnect the auto-pilot and manually recover the aircrafts attitude and trajectory, minimising the altitude excursion. An interim investigator briefing (ATSB, 2008) praised the crew’s exemplary manual handling ability. This event demonstrates that it is possible for pilots to operate a highly automated airliner and preserve their manual flying ability. Furthermore, it highlights that such skills are still important in a modern air transport environment.

5.3 Suggested Future Research

This thesis has identified a relationship between the manual flying experience and performance of pilot’s operating a highly automated aircraft. More specifically, those pilots were employed by an airline that operated a predominantly short-haul route network. These pilots typically flew upwards of four sectors a day, often to airfields which demanded a manual approach due to a lack of ground based navigational aids. Contrastingly, pilots of long-haul aircraft typically fly considerably fewer sectors, often only six or eight a month (the pilot acting as ‘pilot flying’ for only a fraction of those approaches). Consequently, pilots of long haul aircraft are likely to experience considerably less manual flying exposure than the pilots sampled within this research. Given the observation of a sizeable effect within the current study it may therefore be prudent to replicate the work with a sample of long-haul pilots since the effects may be even more pronounced.

The current study is effectively a cross-sectional analysis of pilot manual flying skill i.e. the analysis was conducted at a fixed point in time. It may be worthwhile adopting the sensitive measurement methodology developed within this study to perform a longitudinal analysis of pilot manual flying skill. Periodic samples of manual handling proficiency could be made so as to capture a pilot transitioning from manual aircraft to highly automated aircraft. This profile of performance data could be used to study in more detail how advanced flight deck technology impacts manual handling skill. Clearly this research would be resource intensive and would require adequate protection against the effects of sample attrition over the extended data collection period.

A further development may be to integrate the research with a Flight Data Monitoring programme and look for evidence of manual skills attrition in real world derived flight data. The analysis techniques adopted within this thesis may also be of benefit to the Flight Data Monitoring community and it would be beneficial to research how these techniques could be shared.

References

Air Accident Investigation Branch (2006). *Airbus A320-232 HA-LPB, AAIB Bulletin No. 4/2007*. Air Accident Investigation Branch, Farnborough.

Alamberti, R. (1998). Automation in Aviation: A Human Factors Perspective. In: Garland, D.; Wise, J.; and Hopkin, D. (eds.) (1998). *Aviation Human Factors*. Lawrence Erlbaum Associates, New Jersey.

Arthur, W.; Bennett, W.; Stanush, P.; and McNelly, T. L. (1998). Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Human Performance.* Vol. 11, No. 1, pp. 57-101.

Australian Transport Safety Bureau (2007). *Human Factors Analysis of Australian Aviation Accidents and Comparison with the United States.* B2004/0321. Australian Transport Safety Bureau, Canberra.

Australian Transport Safety Bureau (2008). Qantas A330 Accident Media Conference. 14 October 2008. Australian Transport Safety Bureau, Canberra.

Baddeley, A. (1986). *Working Memory.* Oxford University Press, New York.

Baker, D. and Dismukes, R.K. (2002). A Framework for Understanding Crew Performance Assessment Issues. The International Journal of Aviation Psychology. Vol. 12, No. 3, pp. 205 – 222.

Baron, S. (1988), Pilot Control. In Wiener, E. and Nagel, D. (eds.) (1998). *Human Factors in Aviation.* Academic Press, London.

Beilock, S.; Kulp, C.; Holt, L.; and Carr, T. (2004). More on the Fragility of Performance: Choking Under Pressure in Mathematical Problem Solving. *Journal of Experimental Psychology: General.* Vol. 133, pp. 584-600.

Billings, C. (1997). *Aviation Automation: The Search for a Human-Centered Approach.* Lawrence Erlbaum, Philadelphia.

Boeing (2008). *Statistical Summary of Commercial Jet Airplane Accidents, Worldwide Operations 1959-2007.* Boeing Commercial Airplanes, Seattle, Washington.

Carley, K.; and Palmquist, M. (1992). Extracting, Representing, and Analyzing Mental Models. *Social Forces.* Vol. 70, No. 3, pp. 601 – 636.

Chidester, T. (2003). Understanding Normal and Atypical Operations Through Analysis of Flight Data. *Proceedings of the 12th International Symposium on Aviation Psychology;* April 14 – 17, 2003. Dayton, Ohio.

Childs, J. and Spears, W. (1986). Flight-Skill Decay and Recurrent Training. *Perceptual and Motor Skills.* Vol. 62, pp. 235-242.

Civil Aviation Authority (2003). *CAP 739: Flight Data Monitoring*. Civil Aviation Authority, Gatwick.

Civil Aviation Authority (2005). *Standards Document 24. Version 05: Guidance to Examiners: Multi-pilot Aeroplanes (MPA) Type Rating skill tests and proficiency checks.* Civil Aviation Authority Safety Regulation Group, Gatwick.

Civil Aviation Authority (2008). *CAP 776: Global Fatal Accident Review 1997 – 2006*. Civil Aviation Authority, Gatwick.

Curry, R. (1985). *The Introduction of New Cockpit Technology: A Human Factors Study.* NASA-TM-86659; REPT-86659; NAS 1.15:86659; Pagination 71P.

Damos, D.; John, R.; and Lyall, E. (1999). The Effect of Level on Automation on Time Spent Looking Out of the Cockpit. *International Journal of Aviation Psychology.* Vol. 9, no. 3, pp. 303-314.

Damos, D.; John, R.; and Lyall, E. (2005). Pilot Activities and the Level of Cockpit Automation. *International Journal of Aviation Psychology.* Vol. 15, no. 3, pp. 251-268.

Davenport, M. and Harris, D. (1992). The Effect of Low Blood Alcohol Levels on Pilot Performance in a Series of Simulated Approach and Landing Trials. *International Journal of Aviation Psychology*. Vol. 2., pp. 271 – 280.

Delzell, S.; Johnson, W.; and Liao, M. (1998). Pilots' Spatial Mental Models for Memory of Heading and Altitude. *Proceedings of the 17th Digital Avionic Systems Conference*. Bellevue, WA.

Ebbatson, M. (2006). Practice Makes Imperfect: Common Factors in Recent Manual Approach Incidents. *Human Factors and Aerospace Safety.* Vol. 6, no. 3, pp. 275-278.

Ebbatson, M.; Huddlestone, J.; Harris, D.; and Sears, R. (2006). The Applicatiion of Frequency Analysis Based Performance Measures as an Adjunct to Flight Path Derived Measures of Pilot Performance. *Human Factors and Aerospace Safety*. Vol. 6, No. 4, pp. 383 – 394.

Endsley, M. (2006). *Situational Awareness.* John Wiley & Sons, New Jersey.

Flach, J. and Jaques, P. (2003). A Search For Meaning: A Case Study of the Approach-To-Landing. In Hollnagel, E. (ed.) (2003). *Handbook of Cognitive Task Design.* Lawrence Erlbaum, Oxford. pp. 171-191.

Flight Safety Foundation (2000); *Approach and Landing Accident Reduction Tool Kit.* Flight Safety Foundation, Virginia.

Goteman, Ö. and Dekker, S. (2003). Flight Crew and Aircraft Performance During RNAV Approaches: Studying the Effects of Throwing New Technology at an Old Problem. *Human Factors and Aerospace Safety.* Vol. 3, No. 2, pp. 147-164.

Harris, W.; Hancock, P.; Arthur, E.; and Caird, J. (1995). Performance, Workload, and Fatigue Changes Associated with Automation. *International Journal of Aviation Psychology.* Vol. 5, No. 2, pp. 169-185.

Hawkins, F. (1998). *Human Factors in Flight*. Ashgate, Aldershot.

Hegarty, M. (2004), Mechanical Reasoning by Mental Simulation. *Trends in Cognitive Sciences.* Vol. 8, no. 6, pp. 280-285.

Heinrich, H. (1959). *Industrial Accident Prevention: A Scientific Approach*. 4th Ed. McGraw-Hill.

Holder, B. and Hutchins, E. (2001). What Pilots Learn About Auto-Flight While Flying on the Line. *11th International Symposium on Aviation Psychology;* March 5-8, 2001. Ohio State University, Columbus, Ohio.

Hubbard, D. (1987). Inadequacy of Root Mean Square Error as a Performance Measure. In Jensen, R. (ed.) (1987). *Proceedings of the Fourth International Symposium on Aviation Psychology*. Ohio State University, Columbus, Ohio, pp. 698-704.

Hutchings, E. (1995). How a Cockpit Remembers Its Speed. *Cognitive Science.* Vol. 19, pp. 265-288.

Jane's Information Group (2000). *Jane's All the World's Aircraft*. 91st Edition.

Johnson, N. and Rantanen, E. (2005). Objective Pilot Performance Measurement: A Literature Review and Taxonomy of Metrics. *Proceedings of the 13th International Symposium on Aviation Psychology,* Dayton, Ohio.

Johnson, N.; Rantanen, E.; and Talleur, D. (2004). Time Series Based Objective Pilot Performance Measures. *International Journal of Applied Aviation Studies.* Vol. 4, No. 1, pp. 13-29.

Joint Airworthiness Authorities (2006). *JAR-FCL Section 1 - Aeroplane*. Joint Airworthiness Authorities, Cologne.

Kirwan, B. (1992). *A Guide to Task Analysis*. Taylor and Francis, Oxford.

Klein, G. and Crandall, B. (1999). The Role of Mental Simulation in Problem Solving and Decision Making. In Hancock, S. and Flach, J. (eds.)(1999)

Applications of the ecological approach to human-machine systems. pp. 324-358.

Latorella, K.; Pliske, R.; Hutton, R.; and Chrenka, J. (2001). *Cognitive Task Analysis of Business Jet Pilots' Weather Flying Behaviours: Preliminary Results.* NASA TM 211034, NASA Langley.

Landis, J. and Koch, G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*. Vol. 33, pp. 159 – 174.

McDowell, E. (1978). *The Development and Evaluation of Objective Frequency Domain Based Pilot Performance Measures in ASUPT [Final Report, Mar. 1977 - Feb. 1978].* AD-A059477; AFOSR-78-1239TR; Pagination 41P.

McRuer, D. and Jex, H. (1967). A Review of Quasi-Linear Pilot Models. *IEEE Transactions on Human Factors in Electronics*. Vol. HFE 8, No. 3, pp. 231 – 249.

McRuer, D. (1982). *Pitfalls and Progress in Advanced Flight Control Systems.* AGARD CP-321, AGARD/NATO, Neulliy-sur-Seine.

Militello, L.; Hutton, R.; Pliske, R.; Knight, B.; Klein, G.; and Randel, J. (1997). *Applied Cognitive Task Analysis (ACTA) Methodology.* NPRDCTN-98-4, Navy Personnel Research and Development Centre, San Diego, CA.

Moray, N. (1990), A Lattice Approach to the Structure of Mental Models. In Broadbent, D. and Reason, J. (eds.) *Human Factors in Hazardous Situations,* Oxford University Press, Oxford , pp. 129-135.

Moray, N. (1999). Mental Models in Theory and Practice. In: Gopher, D. and Koriat, A. (1999). Attention and Performance XVII: Cognitive Regulation of Performance: Interaction of Theory and Application. The MIT Press, Boston.

Nunes, A. (2003) The Impact of Automation Use on the Mental Model: Findings from the Air Traffic Control Domain. *Proceedings of the 47th annual Meeting of the Human Factors and Ergonomics Society;* Santa Monica, California. Human Factors and Ergonomics Society.

O'Hare, D.; Wiggins, M.; Batt, R.; and Morrison, D. (1994). Cognitive Failure Analysis for Aircraft Accident Investigation. *Ergonomics*. Vol 37, pp. 1855 – 1869.

Owen, G. and Funk, K. (2007)*, Flight Deck Automation Issues: Incident Report Analysis*. http://www.flightdeckautomation.com/incidentstudy/incident-analysis.aspx . Accessed 24 September 2008.

Parasuraman, R., Molloy, R. and Singh, I. (1993), Performance Consequences of Automation-Induced "Complacency". *International Journal of Aviation Psychology.* Vol. 3, No. 1, pp. 1-23.

Parasuraman, R.; Sheridan, T.; and Wickens, C. (2000), "A Model for Types and Levels of Human Interaction with Automation", *IEEE Transactions on Systems, Man and Cybernetics.* Vol. 30, No. 3, pp. 286-297.

Press, W.; Flannery, B.; Teukolsky, S.; and Vetterling, W. (1989). Fourier Transform of Discretely Sampled Data. *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd ed. Cambridge University Press, Cambridge.

Rantanen, E.; Talleur, D.; Taylor, H.; Bradshaw, G.; Emanuel Jr., T.; Lendrum, L.; and Hulin, C. (2001). Derivation of Pilot Performance Measures from Flight Data Recorder Information. *Proceedings of the 11th International Symposium on Aviation Psychology; Columbus, Ohio*; Mar. 5-8, 2001, Ohio State University, Columbus, Ohio.

Rantanen, E.; Johnson, N.; and Talleur, D. (2004). *The Effectiveness of a Personal Computer Aviation Training Device, a Flight Training Device, and An Airplane in Conducting Instrument Proficiency Checks. Volume 2: Objective Pilot Performance Measures.* Final technical report AHFD-04-16/FAA-04-6.

Reason, J. (1990). *Human Error*. Cambridge University Press. Cambridge.

Rees, D. and Harris, D. (1995). The Effectiveness of ab-initio Flight Training Using Either Linked or Unlinked Primary Axis Flight Controls. *International Journal of Aviation Psychology*. Vol. 5., pp. 291 – 304.

Sarter, N.; Woods, D.; and Billings, C. (1997). Automation Surprises. In Salvendy, G. (ed.) (1997) *Handbook of Human Factors & Ergonomics.* Second ed., Wiley, Chichester.

Sarter, N.; Wickens, C.; Mumaw, R.; Kimball, S.; Marsh, R.; Nikolic, M.; and Xu, W. (2003). Modern Flight Deck Automation: Pilots Mental Model and Monitoring Patterns and Performance. *Proceedings of the 12th International Symposium on Aviation Psychology*; April 14-17, 2003, Dayton, Ohio.

Shannon, E. (1998). Communication in the Presence of Noise. *Proceedings of the IEEE.* Vol. 86, No. 2, pp. 447 – 457.

Tenney, Y., Rogers, W. and Pew, R. (1998). Pilot Opinion on Cockpit Automation Issues. *International Journal of Aviation Psychology.* Vol. 8, No. 2, pp. 103-120.

Terelak, J. (1993), Anxiety and Eye-Hands-Legs Coordination in Young Pilots. Proceedings of the *7th International Symposium on Aviation Psychology:* Apr 26-29 1993, Ohio State University, Columbus, Ohio.

Van Es., G. (2002). *Advanced flight data analysis*> NLR-TP-2002-091, NLR, Netherlands.

Veillette, P. (1995). Differences in Aircrew Manual Skill in Automated and Conventional Flight Decks. *Transportation Research Record.* Vol. 1480, pp. 43-50.

Vivian, M. (2004). Flight Operations Department Communication (FODCOM) 24/2004. Civil Aviation Authority Safety Regulation Group, Gatwick.

Walley, S. (1995), Mental Rehearsal Techniques for Reducing Skill Decay of Unemployed Commercial Pilots. *Aviation Psychology: Training and selection; Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP), Trinity College, Dublin, Ireland, 1994.* Vol. 2. Avebury Aviation, Vermont.

Wickens, C. D. (2000). *Aviation Psychology,* Oxford University Press. Oxford.

Wiegmann, D. and Shappell, S. (2001). Human Error Analysis of Commercial Aviation Accidents: Application of the Human Factors Analysis Classification System. *Aviation, Space, and Environmental Medicine*. Vol. 72, pp. 1006 – 1016.

Wiegmann, D. and Shappell, S. (2003). *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System*. Ashgate, Aldershot.

Wiener, E. (1988). Cockpit Automation. In Wiener, E. L. and Nagel, D. C. (eds.) (1988). *Human factors in Aviation.* Academic press, London.

Wiener, E. (1989). *Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft.* NASA CR 177528, NASA Ames, CA.

Wiener, E.; Chute, R.; and Moses, J. (1999). *Transition to Glass: Pilot Training for High-Technology Transport Aircraft.* NASA/CR-1999-208784, National Aeronautics and Space Administration, Washington, DC.

Wood, S. (2004). *Flight Crew Reliance on Automation.* CAA 2004/10, Civil Aviation Authority, Gatwick.

Young, J. P., Fanjoy, R. O. and Suckow, M. W. (2006), Impact Of Glass Cockpit Experience On Manual Flight Skills. *Journal of Aviation Aerospace Education and Research.* Vol. 15, No. 2, pp. 27-32.

Appendix A

Cognitive Interview Protocol

ACTA Interview Schedule

Introduction

1. What do you think are the individual skills that we refer to when we talk about manual flying skills?
2. In your experience do you think the level of manual flying skill is changing?
3. What do you think are the reasons for this change?

Task Diagram

1. I'd like to think now about a specific element of your job that requires cognitive skill to perform.
2. Introduce the task and identify the segment to be analysed.
3. What do you do to manage the approach? How do you ensure that you will arrive at the appropriate points at the appropriate time, speed and configuration?
4. I'd like you to think about what you do when you perform this task without the automation (elaborate if necessary). Can you describe it in a number of linked steps; let's say no more than six?
5. Which of these steps involve complex cognitive skills, and by cognitive skills I mean decisions, judgements - thinking skills?
6. Would this diagram be different if you had the automation available? How?

Knowledge Audit

List of cognitive probes to investigate each task step

1. What is the goal of this task?
   What are you trying to achieve by carrying out this task?
2. What set of conditions prompts you to start performing this task?
3. What set of conditions tells you to stop performing this task?
   What are you looking for to stop doing this task?
4. Can you give examples of the decisions you have to make when performing this task?
5. How do you know that at this point?
6. What is the important information you need to know to carry out this task? Where would you find this information?
7. How would you calculate that information?
   What strategy would you use to work that out?

8. Do you have to make any assumptions to perform this task?

9. Are there any rules of thumb/tricks of the trade that you use to make this task easier?
   Can you think of an example to explain how you would use this rule of thumb?

10. Can you think of an example of how you would work smart at this task i.e. not cut corners but work efficiently and achieve more with less?

11. How does the FMS assist you with this task?
    Would the absence of the FMS affect how you perform this task?

12. How does the auto throttle assist you with this task?
    Would the absence of the auto throttle affect how you perform this task?

13. How does the flight director assist you with this task?
    Would the absence of the flight director affect how you perform this task?

14. What would you have to do if ATC suddenly changed the plan?
    i.e. gave you a different crossing altitude or changed the route/runway

15. How do you know where you are at this point?

16. What is your next goal at this point?

17. How do you know where you are in relation to your goal?

18. How do you decide if you will achieve your goal?
    How would you judge if you weren’t going to make your goal?

19. What would you do if you decided you couldn’t achieve your goal?
    What corrective action would you take at this point?
    How would you decide what corrective action to take?

20. Can you give me an example of when you realised that the way you were doing this wasn’t going to work and you would have to do it differently?

21. Can you think of a time when you were performing this task that you suddenly noticed something that appeared obvious, but the other pilot did not?

22. How may a less experienced person struggle with this task?
    What specifically causes them to struggle?

Simulation Scenario

Please read through this scenario as if you were the pilot flying. Take time to read through and imagine the scene, and please use the supporting information. Think about what actions, decisions, judgements or general thoughts you may have at each step as you read through. When you have finished I will ask you to list what you thought were the major events in the scenario, these can be actions, decision points, judgements or any points in the timeline you feel were important.

Please list the major events in the sequence they occurred. At this stage I don’t want to go into too much detail, just list what the event was and I’ll write them down the side of this table. After we shall come back to each in turn and explore them in more detail.

When the major events have been listed use selective probes from above list to elicit details of the cognitive demands of that event and strategies employed etc.

Appendix B

Cognitive Demand Tables

Planning Demands

Cognitive Demand	Why Difficult	Cues/Factors	Strategies/Actions	Novice Errors	Automation
How is the plan likely to change and how should I respond?	Requires the evaluation of multiple options from uncertain information. resource intensive and ill defined.	Traffic intensity, weather trends and forecast, location idiosyncrasy	Plan for most likely changes, i.e. runway change, minima change	Don't know likely changes and therefore can't anticipate or plan for them.	HUMAN
Select speed for initial descent					
Select strategic points to assess approach		ATC restrictions, approach plate, weather	Use ATC waypoints as check points, also assign intermediate points to evaluate performance. Use 3 in 1 to determine their position and altitude.	Insufficient checks between major waypoints - allow errors to accumulate, fail to meet waypoint crossing restrictions	FMS provides a continuous data of profile error against the planned profile
How could the plan change and when?	Requires knowledge of local operations, awareness of the influence of many factors.	Traffic, weather, ATC	Anticipate if typical route alterations may occur and how you would respond	Novices don't consider the influence of enough factors. Don't have fall back plans if things change. Consequently they are likely to stick with the current plan if things do change.	HUMAN
Anticipate approach routing	Many unknown variables, using historical data for prediction	Weather reports, weather radar	Mentally simulate what you expect the approach to look like and consider factors that may change	Don't take the time to do this, get flustered if things do change	HUMAN
Mentally simulate the approach to commit to memory	Requires considerable time and cognitive resources	Approach plates, knowledge of aircraft performance, crossing restrictions		Fail to review approach and commit salient features to memory. Frequently need to review printed approach information	FMS stores waypoints and displays for instant review - provides pictorial representation of the approach (in plan only)
Which navigation aids should I use?			Select nav aids and standbys so that minimal effort is required to interpret your position	Don't have strategic standby nav aids in case of re-route. Stick with inappropriate beacons and spend more effort computing their position.	FMS computes a position automatically from several aids. No need to select individual aids.
What is my fuel duration	No single source of information, must be integrated and predicted from multple information sources.	Fuel flow rate, anticipated average fuel flow, fuel remaining, time to destination	Compute time to destination based upon track and average airspeed, multiply by average fuel flow and subtract fuel usage from that remaining		HUMAN
Where are the critical obstacles that effect me?	Must be incorporated into the mental image from a number of sources and held.	Approach charts, terrain, radar, traffic, TCAS, ATC restrictions			
What will be my likely routing?	May be influenced by many factors, large degree of uncertainty	Knowledge of local idiosyncrasy, weather, traffic, time of day	Draw on knowledge of previous operations to that airfield, expect full procedure if traffic is heavy, expect possible short cuts when quiet	Do not consider changes to the standard plan, get caught out if plan is then changed	HUMAN
What will be my landing weight?	Need to calculate landing weight. requires interpretation and extrapolation of disparate information sources	Current aircraft load, expected avaerage fuel flow, remaining track to target	Estimate fuel usage for remaining track and subtract from current aircraft load		FMC generates predicted fuel loads and estimated landing weight
Which approach will I be required to use?	Many influencing factors, some may be peculiar to the airfield			Lack of knowledge of the airfield	HUMAN
Calculate when to start initial descent	Need to know aircrafts typical idle profile	Speed, altitude, aircraft performance, track to target, winds, temperature	Idle descent at fixed airspeed, work out total track miles for descent using 1 in 3 rule		FMS determines optimal TOD point
Where should I begin my descent			Idle descent profile, use 3 in 1 rule to determine track miles for altitude.		FMS calculates optimal descent point
How should we transition to the landing configuration	Must reduce speed + rate of descent. Aircraft descent characteristics change with config change. External demands change - more ATC requirements close to airfield	Distance to FAP, altitude, speed, configuration, descent rate	Aim to be in landing configuration a couple of miles before FAP. Use rules for changing aircraft speed and consider aircraft limitations.	Concentrate on flying the profile but forget to configure the aircraft - configure late, rushed flap scheduling, braking with flaps	FMS provides flap scheduling information on ASI and VNAV achieves path to lose airspeed

Planning Demands Continued

What landing configuration should I use?	Requires consideration of a number of factors. Balance between expediancy and efficiency.	Runway length, aircraft weight, wind, other weather, aircraft status, traffic, airfield idiosyncrasy	Aim for lower landing speed if turnaround time short due to lower brake cooling times. Higher speed if gate at the end of the runway. Assess situation and consult reference card for config and speeds.	Always use the same landing configuration which may be safe but not necessarily the most efficient	HUMAN
Store targets and constraints for recall	Requires information to be comitted to memory, resource intensive.	Charts, altitudes, speeds, track	Mentally simulate approach which aids memorisation of salient features. Make written notes of important points.	Novices often do not have the capacity to undertake this task and instead revert to information gathering during the approach	FMS stores route information and presents it for rapid recall.
Retain a reference picture of the planned approach and targets	The integrated approach plan and aspects not printed on the individual approach charts must be held in memory.		Make written notes of critical flight information. Mentally simulate planned approach to commit to memory	Will have insufficient capacity to make notes or commit approach to memory. Will have to regenerate and search for information frequently during the approach.	FMS records waypoint information and presents plan pictorially on navigation display for reference.
What targets should I aim to achieve on the approach?	Between major waypoints there may be no stipulated targets	ATC restrictions, approach charts	Aim to achieve ATC stipulated waypoints but also insert targets to divide those segments into smaller sub-segments. Insert targets where route changes may be considered. Memorise the targets position (track) and altitude.	Use only the basic waypoints as their target structure and therefore are more vulnerable to estimation errors when computing position. Allow larger deviations to build up before correcting.	FMS can generate a continuous target, extrapolating between waypoints, and give constant feedback of relative error
How many miles does the approach demand?	Information gathered from disparate sources, not easy if not a published approach	Chart distances, ATC information	Add up sector distances, estimate complex turns based on knowledge	Add too many or too few miles for unknown distances i.e. turns.	FMS generates precise information about route length
What will the weather situation be?	Large degree of uncertaintly in information, often requires extrapolation of historical data or estimation between two data points.	Weather forecasts, current reports, knowledge of weather behaviour, local variations		Novices may not anticipate unusual weather behaviour at certain destinations as they have no experience of it	HUMAN
Plan how to lose energy on profile	Aircraft energy is a function of multiple variables and their interaction must be understood. There may be many options of how to loose energy and the most appropriate means must be selected.	Aircraft performance, profile restrictions, altitude, speed, commercial pressure	Have a standard model. Plan for a basic 3 degree approach at idle and allow suffcient track with config changes at standard points. Mentally simulate and identify constraints that may require a change to the standard model. Modify config changes, speeds or profile to meet differences.	Insufficient knowledge of aircraft performance and therefore do not recognise that energy management plan is inappropriate.	FMS will plan a suitably managed approach profile. Manually edited approaches however may not have achievable energy profiles.

Execution Demands

Cognitive Demand	Why Difficult	Cues/Factors	Strategies/Actions	Novice Errors	Automation
When do I need to start this turn?	Difficult geometry for mental calculation, no source of information provided	Airspeed, turn rate	For 90 degree turn at 250kts and standard rate, lead distance is groundspeed over 100	Have not got a similar rule of thumb	FMS displays predicted turn radius and lead in distance on ND
How long will it take to reach a target?	No direct indication, needs interpretation	Airspeed, track, wind	Track miles divided by groundspeed	Don’t calculate time, tasks managament suffers, become rushed	FMS offers ETA for waypoints
What is the state of the wind?	Requires capacity for mental arithmetic. May be working with historical data and therefore need to estimate or compensate	Actual wind report, wind forecasts, groundspeed data	Compute headwind by looking at difference between groundspeed indication (from dme) and airspeed. Compute headwind and crosswind components from actual wind reports. Compute average wind over altitude change by looking at forecasts/reports. Extrapolate wind at flight level forcasts to determine wind at current level.	Base computations on forecast wind information and fail to seek current wind report.	INS gives constant indication of current wind conditions. The need to assess deminishes to some extent as wind compensation is built into the auto-flight system.
Choose a descent rate to reduce speed for next gate	Must slow but also remain on profile. Speed and altitude are tightly coupled via energy. Must meet gate but inefficient to slow too early.	Speed, descent rate, gate altitude, current altitude	In idle descent, reduce rate of descent to slow. Predict track miles for descent to target altitude at that rate (rate = time, time = distance) adjust rate depending on track error		FMS alerts if drag is required to achieve gate, VNAV manages profile and chooses appropriate descent rate
How much altitude do I need to lose?	Mental arithmetic which demands capacity	Approach charts, target altitude, current altitude, pressure adjustment			HUMAN
How many miles do I have to go?	Can rarely be determined precisely as charts contain limited information - changes if you're not exactly flying the published procedure	DME, rules of thumb	Select navigation aid at strategic location to provide distance information easily. Minimal interpretation. Add sector lengths and estimate unknowns.	will stick with FMS value of track following an ATC change for longer than expert. Will not recompute track value or seek that information	FMS porvides very accurate value for track miles to run, although this is also invalid if route is changed.
What are the headwind and crosswind components?	Needs to be determined from historical information.	Airspeed, groundspeed, time, wind reports	Derive from wind reports or use DME and timing against airspeed to determine headwind factor. Allow error margin.	Use historical information without checking if it is accurate.	FMS provides current wind vector information based on INS
How far have I got to go?	Needs to be interpretend from numerous sources. Difficult to calculate precise track of complex paths due to geometry	DME, approach charts	Knowledge of approximate radius of turns at various speeds. Add up approach segments on charts		FMS performs complex geometry and gives precise indication of track on planned flightpath

Monitoring Demands

Cognitive Demand	Why Difficult	Cues/Factors	Strategies/Actions	Novice Errors	Automation
When do I need to start this turn?	Difficult geometry for mental calculation, no source of information provided	Airspeed, turn rate	For 90 degree turn at 250kts and standard rate, lead distance is groundspeed over 100	Have not got a similar rule of thumb	FMS displays predicted turn radius and lead in distance on ND
How long will it take to reach a target?	No direct indication, needs interpretation	Airspeed, track, wind	Track miles divided by groundspeed	Don't calculate time, tasks managament suffers, become rushed	FMS offers ETA for waypoints
What is the state of the wind?	Requires capacity for mental arithmetic. May be working with historical data and therefore need to estimate or compensate	Actual wind report, wind forecasts, groundspeed data	Compute headwind by looking at difference between groundspeed indication (from dme) and airspeed. Compute headwind and crosswind components from actual wind reports. Compute average wind over altitude change by looking at forecasts/reports. Extrapolate wind at flight level forcasts to determine wind at current level.	Base computations on forecast wind information and fail to seek current wind report.	INS gives constant indication of current wind conditions. The need to assess deminishes to some extent as wind compensation is built into the auto-flight system.
Choose a descent rate to reduce speed for next gate	Must slow but also remain on profile. Speed and altitude are tightly coupled via energy. Must meet gate but inefficient to slow too early.	Speed, descent rate, gate altitude, current altitude	In idle descent, reduce rate of descent to slow. Predict track miles for descent to target altitude at that rate (rate = time, time = distance) adjust rate depending on track error		FMS alerts if drag is required to achieve gate, VNAV manages profile and chooses appropriate descent rate
How much altitude do I need to lose?	Mental arithmetic which demands capacity	Approach charts, target altitude, current altitude, pressure adjustment			HUMAN
How many miles do I have to go?	Can rarely be determined precisely as charts contain limited information - changes if you're not exactly flying the published procedure	DME, rules of thumb	Select navigation aid at strategic location to provide distance information easily. Minimal interpretation. Add sector lengths and estimate unknowns.	will stick with FMS value of track following an ATC change for longer than expert. Will not recompute track value or seek that information	FMS porvides very accurate value for track miles to run, although this is also invalid if route is changed.
What are the headwind and crosswind components?	Needs to be determined from historical information.	Airspeed, groundspeed, time, wind reports	Derive from wind reports or use DME and timing against airspeed to determine headwind factor. Allow error margin.	Use historical information without checking if it is accurate.	FMS provides current wind vector information based on INS
How far have I got to go?	Needs to be interpretend from numerous sources. Difficult to calculate precise track of complex paths due to geometry	DME, approach charts	Knowledge of approximate radius of turns at various speeds. Add up approach segments on charts		FMS performs complex geometry and gives precise indication of track on planned flightpath

Modify Plan Demands

Cognitive Demand	Why Difficult	Cues/Factors	Strategies/Actions	Novice Errors	Automation
Is that a recoverable error?	No firm yes or no answer, needs experience to understand what is and is not recoverable.	Track miles to next gate, distance from threshold	Error tolerance reduces as you get closer to your stabilised approach point		Autoflight will attempt recovery but may not succeed
Can I accept this plan change?	Requires rapid assesment of many variables	Terrain, type of ATC service, radar coverage, sufficient track to lose altitude	3 to 1 rule to determine track required for altitude change. As rule of thumb during early stages of the approach reject if the change leaves you more than 2-3000ft above profile.	Accept plan changes without proper asesment. Reluctance to 'say no' to atc.	The FMS will advise if approach exceeds capability
How should we respond to this error in profile?	Requires option generation and evaluation in time pressured environment, knowledge of aircraft performance abilities.	Magnitude of altitude error or track error, aircraft config	Have reference rules of thumb for acceptable error "thresholds" - first level is just adjust rate or descent - second level needs config change - third level is more track or go around	Do not have rough guidelines and have to revert to first principles to determine what is acceptable or unacceptable. Requires lots of capacity so often causes bottleneck	Autoflight automatically adjusts profile to achieve next gate.
How should I correct this profile error?	Multiple alternate options	Weight, speed, height	Sacrifice energy with speedbrake or accept an increase in speed and dive off altitude - this depends on constraints. Convert profile into track error. Use rules of thumb for acceptable track errors i.e. 5 miles is recoverable, 10 will require speedbrake. always use all of the speedbrake for this	Don't recognise the need to manage speed. Don't make a positive correction and allow aircraft to 'drift' back onto profile, rushed speed correction close to target. Reluctant to request track miles from ATC. Reliance on controller to prompt altitude.	Auto-flight automatically adjusts profile to achieve next gate.
Should I execute go-around or continue to land?	Time pressured decision	Visual scene, aircraft status, weather, altitude, stability, speed, configuration			HUMAN
Can we correct the profile error?	Depends on many varaiables, with complex interaction. Requires knowledge of aircraft performance, often outside of typical operating range.	Track miles, airpseed, aircraft weight, altitude error, speed error, configuration	Have good knowledge of the aircrafts performance, i.e. how long it takes to slow down at different descent rates and configurations. Estimate track required to correct and compare to what is available.	Become overloaded, show good control of aircraft's position but fail to consider energy and the extended implications of their actions. May dive to recover profile but fail to anticipate speed increase for next gate.	The FMS will advise if approach exceeds capability
Can I accept this track change?	Track change likely to occur at high workload time and requires significant re-planning. Need to rapidly assess situation		Determine how many track miles will be added/subtracted and have rules for what is or isn't achievable	Must go back to first principles to determine if track change can be accepted or rejected, consumes capacity. Other parameters suffer. Communication suffers. Novice may allow himslef very large margins due to uncertainty.	The FMS will advise if approach exceeds capability

General Demands

Cognitive Demand	Why Difficult	Cues/Factors	Strategies/Actions	Novice Errors	Automation
How can I my minimise workload?	Time pressures reduce that available for planning. Lots of information needs to be interpreted and memorised		segment task into sub-segments, consider only the parameters applicable to that phase	Fail to segment task and carry out tasks concurrently	HUMAN
What flight information do I need to attend to and where can I find it?	There are a multitude of information sources which must be attended to and integrated, demanding capacity	Flight phase, task goal	Develop and use effective scan patterns. Identify which scan pattern should be used for each phase of flight to attend only to relavent information. Be disciplined in frequently executing the scan.	Use the same scan for all phases of flight, inefficiently attending to irrelavent information or missing important information. Scan infrequently or in a random fashion, looking at the wrong things at the wrong times.	Auto-flight reads flight parameters at very high frequency and process information to generate actions.
What flight information do I need?	Need to simultaneously attend to many variables - assign priority to information - changes depening upon flight phase		Develop a disciplined and frequent scan pattern. Develop and use different scan patterns depending on phase of flight, attending only to pertinent information.	Tend to include irrelavent information in scan and fail to scan frequently	Auto-flight system samples flight parameters very frequently
Am I on top of the situation? - metacognition	Requires skill to assess your own cognitive performance	Stress levels, number and frequency of errors		Do not consider their own performance	HUMAN

Appendix C

ACTA Paper Simulation

Manchester Approach Scenario

You are PF operating a medium sized commercial twin jet transport aircraft on a scheduled passenger service from Shannon to Manchester. Weather at Manchester is reported as above company minimums, with cloud overcast at 2,800ft, wind 320/11kt and temperature/dew point 05/03.

The aircraft is not equipped with a Flight Management Computer and has been dispatched with no significant malfunctions.

You are tracking inbound to the Wallasey VOR, R278, on the L975 Airway. At D52 from the VOR, Manchester clear you for the MIRSI 1B STAR, crossing WAL above FL130, to be 6000ft by MIRSI, and expect vectors for the ILS DME approach runway 24R.

At D3 from the WAL VOR Manchester advise that the landing runway has switched to 06R. Manchester offers you a vectored approach onto the ILS for a 9 mile final, and questions if you can accept. Surface wind is reported as 340/13kt. The MCT VOR is tuned on the second box and DME currently indicates 33nm.

Appendix D

CAA Standards Document 24 Flight Parameter Tolerances

Tolerance

Altitude or Height

Normal Flight ± 100 ft
With simulated engine failure ± 100 ft
Starting go-around at decision alt/ht + 50 ft/-0 ft
Minimum descent alt/ht + 50 ft/-0 ft

Tracking

All except precision approach ± 5°
Precision approach half scale deflection azimuth and glidepath

Heading

All engines operating ± 5°
With simulated engine failure ± 10°

Speed

All engines ± 5kts
Asymmetric +10 /-5kts and never below V2

Further Guidance

1 Height Accuracy
The candidate need not be failed if an error of more than 100ft occurs 2/3 times. However, the examiner should seriously consider awarding an individual fail if:-
a) Height error of more than 200ft occurs.
b) An error of 100ft or more is uncorrected for an unreasonable period of time.
2 Approach Minima
a) On a non-precision approach when constant descent profile is flown care must be taken not to descend below MDH/A when a missed approach is being conducted.
b) RVR must be checked against airfield minima prior to commencing an approach to land.
3 Tracking Accuracy
a) A failure should be awarded at any time during the test/check if there is an inability to settle within +/- 5° of the specified track or correcting track the wrong way and maintaining the error for an unreasonable period.
4 Speed Accuracy
The 5 kts limit in climb, cruise and approach should be extended to 10 kts in the case of jet aircraft and an airspeed error of 15 kts at any time.

NOTE: When making an assessment, handling qualities and aircraft performance should be taken into account.
If the test/check is conducted in an aircraft, the examiner should make allowance for turbulent conditions.

Appendix E

Matlab M-File for the Computation of Frequency Band Metrics (adapted from Johnson, Rantanen & Talleur (2004))

```
function spectral=frequency(parameter)



% Purpose:

%

% estimates the power spectral density function of the time series data for

% the parameter specified and compute the power in each frequency band.

%

%

% Record of revisions:

%

%   Date       Programmer     Description of change

%   ====       ==========     =====================

%  5 Dec 06     Ebbatson       Function adapted

%

%

% Define variables:

%

% SOURCEDATA  --  Array containing all the raw flight data

% FLIGHT   --  Index of rawdata file in focus (ex file)

% FS --  Parameter sampling rate

% BASERATE --  Highest sampling rate in rawdata file (ex fs)
```

```
global FS
global BASERATE
global FLIGHT
global METRICINDEX
global OUTPUT
global SEGMENT


% Calculate the power spectral density function upto the Nyquist frequency using the fft


data=getsample(parameter);
BASERATE=FS/2;
NPO2=2.^(ceil(log(length(data))/log(2)));
NumUniquePts=ceil((NPO2+1)/2);
spectrum=fft(data,NPO2);
spectrum=spectrum(1:NumUniquePts);
MX=abs(spectrum).^2;
MX(1)=MX(1)/2;
if ~rem(NPO2,2)
   MX(length(MX))=MX(length(MX))/2;
end
MX=MX*2; MX=MX/length(data);
f=(0:NumUniquePts-1)*2*BASERATE/NPO2;
```

```
PSD=MX;

Freq=f';



%Calculate cumulative sum of power spectral density


y=cumsum(MX);

PSD_total=sum(MX);

norm_y=y/PSD_total;


% create a matrix containing the freqency band limits


bands=5;

bpf1=0.05;

bpf2=0.10;

bpf3=0.15;

bpf4=0.20;

bpf5=0.25;

bp_freq=[bpf1, bpf2, bpf3, bpf4, bpf5];


%find matrix index which corresponds to the frequency band limit


for k=1:bands

   d=find(f>=bp_freq(k));

   bp_index(k)=d(1);
```

```
end


%determine power in the signal up to frequency band limit


bp1=y(bp_index(1));

bp2=y(bp_index(2));

bp3=y(bp_index(3));

bp4=y(bp_index(4));

bp5=y(bp_index(5));


%determine power just in the frequency band


VLF=bp1;

LF=bp2-bp1;

MF=bp3-bp2;

HF=bp4-bp3;

VHF=bp5-bp4;
```

Appendix F

Ethics Proposal for Airline Study

Loss of Manual Flight Skills in Air Transport Pilots
Ethics Proposal for an Experimental Study

Introduction

Modern jet transport aircraft typically employ automated systems to perform the basic flying duties. The pilots, rather than manipulating the primary flying controls, command the aircraft by entering targets into the auto-flight system and then monitoring its performance. In this manner many of the psycho-motor and cognitive skills required for traditional manual flight are redundant. There is therefore a credible concern within the aviation industry that the infrequently exercised manual flight skills may be decaying, perhaps towards the limit of acceptable standards.

This research study aims to measure the manual flying performance of a broad sample of current, UK licensed, jet transport pilots, exploring the general level of proficiency and any sources of variation within that sample.

Methodology

Both technical and non-technical aspects of pilot performance will be investigated using flight data records and observer assessments. The data is to be collected whilst pilots perform part of their bi-annual simulator based proficiency check. This check is a standard licensing requirement for all UK air transport pilots and incorporates a number of manual flight elements. Flight data will be recovered from the simulator device and the participating license examiners will be asked to provide the observational assessments. Participating pilots will also be asked to provide demographic and career background information via an anonymous pro-forma following the simulator session.

Informed Consent

All air transport pilots are required to complete a proficiency check within six months of their previous validation. Consent to approach individual pilots to request their participation will be sought from both the British Airline Pilots Association (BALPA) and from the appropriate managers of the participating airlines.

As the individual pilots report for their training session, and at a convenient time which does not interfere with their preparation activities, the examiner in charge (an airline staff member who also acts as an appropriately trained member of the research team) will introduce the research and provide a written explanation for the crew to consult, requesting their participation. At the end of this document crews will be asked to provide written consent if they wish to participate in the research. It will be emphasised that non-participation will not be viewed negatively.

Deception

There is no requirement or intention to deceive the participants. Prior to the trial the participants will be informed that the general purpose of the research is to investigate the pilots handling performance. The researcher will not detail the specific phases of flight to be assessed in order to avoid influencing the participant’s response. These details will be given during the debrief session.

Debriefing

At the close of the study the participants will be fully debriefed as to the exact purpose of the research. They will be provided with a written explanation of the research, including the contact details of the research team should they later have any questions. Participants will also be asked to avoid discussing the research with other potential participants to avoid biasing their performance.

Right to Withdraw

As part of the pre-study briefing it will be stressed that all participants have the right to withdraw from the study at any time and that any data contributed up to that point will subsequently be destroyed. However, it will also be explained that it will be impossible to remove their data following the days exercise as it will have been de-identified and aggregated with other data.

Confidentiality

All data will be collected anonymously. Data sets will be identified and collated according to the time and date when they were recorded. The researcher will hold the raw data securely and confidentially. It will be explained that any published data will also be anonymous and as part of an aggregated set. No individual data record will ever be presented.

Risk to Participants

The study will not involve any risk of physical or psychological harm and duress. The research study uses the existing license proficiency check process and therefore does not alter the standard by which participating pilots are measured.

Protection of Participants

The participants will be assured of confidentiality, and briefed that any data downloaded from the simulator will be de-identified and stored securely at Cranfield University. They will also be encouraged to contact the researcher if they wished to discuss any other concerns.

Observational Work

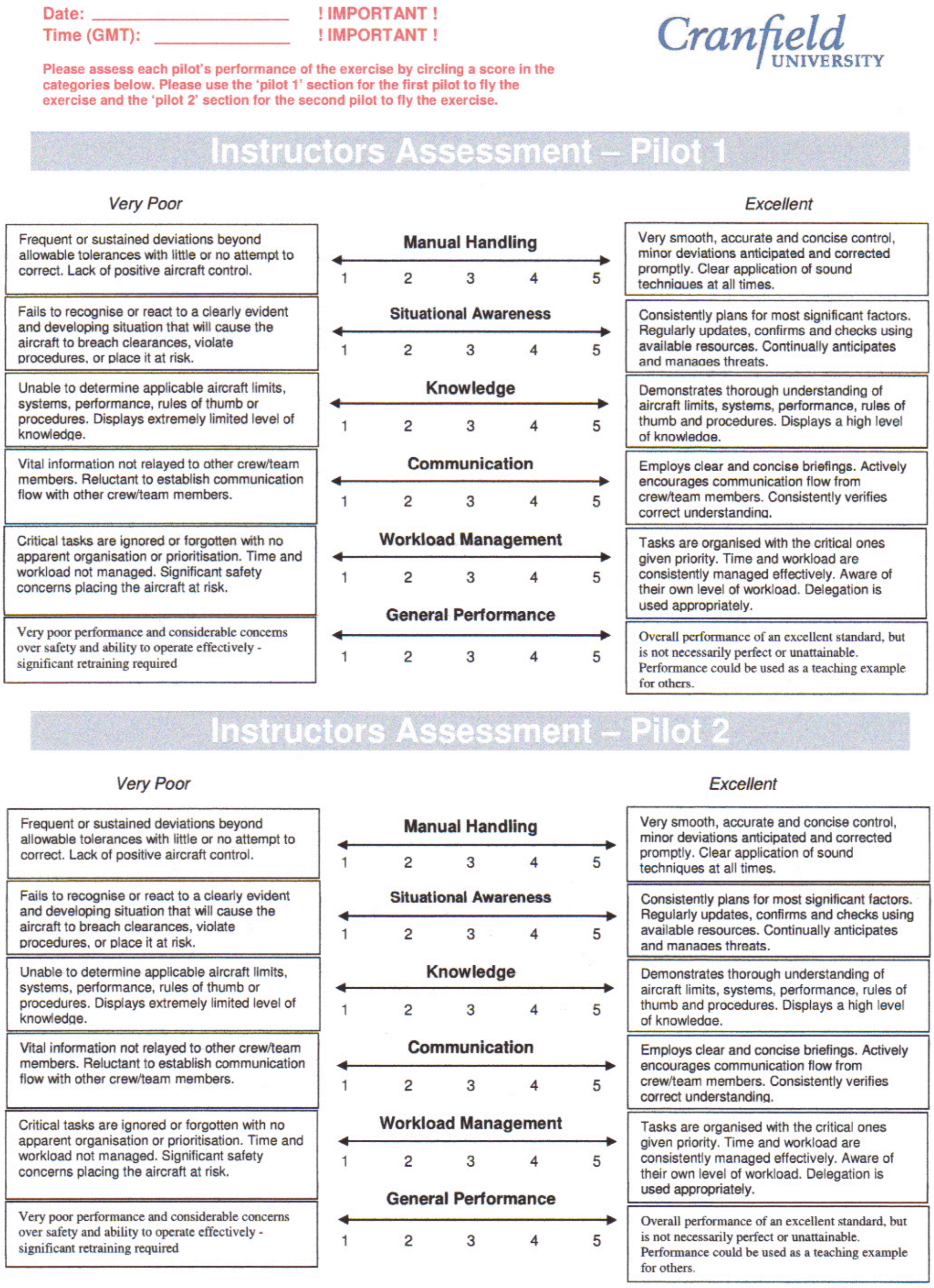
Observational work will be carried out with participants during their proficiency check programme. Consequently, the participant will be aware they are being observed and why. However, the observations will take place during normal assessment sessions when the participants expect to be observed by the examiner.

Professional Conduct

The research will be carried out in such a manner as to uphold the continued reputation of the university. The researcher will also ensure that the research will be conducted in a professional manner to ensure the continued support of the public for similar work. Final approval for the research will be obtained from the Cranfield University Ethics Committee.

Appendix G

TRE Scoring Scale



Date: ____________________ ! IMPORTANT !

Time (GMT): ________________ ! IMPORTANT !

Please assess each pilot's performance of the exercise by circling a score in the categories below. Please use the 'pilot 1' section for the first pilot to fly the exercise and the 'pilot 2' section for the second pilot to fly the exercise.

Cranfield UNIVERSITY

Instructors Assessment – Pilot 1

Very Poor		*Excellent*
Frequent or sustained deviations beyond allowable tolerances with little or no attempt to correct. Lack of positive aircraft control.	**Manual Handling** ↔ 1 2 3 4 5	Very smooth, accurate and concise control, minor deviations anticipated and corrected promptly. Clear application of sound techniques at all times.
Fails to recognise or react to a clearly evident and developing situation that will cause the aircraft to breach clearances, violate procedures, or place it at risk.	**Situational Awareness** ↔ 1 2 3 4 5	Consistently plans for most significant factors. Regularly updates, confirms and checks using available resources. Continually anticipates and manages threats.
Unable to determine applicable aircraft limits, systems, performance, rules of thumb or procedures. Displays extremely limited level of knowledge.	**Knowledge** ↔ 1 2 3 4 5	Demonstrates thorough understanding of aircraft limits, systems, performance, rules of thumb and procedures. Displays a high level of knowledge.
Vital information not relayed to other crew/team members. Reluctant to establish communication flow with other crew/team members.	**Communication** ↔ 1 2 3 4 5	Employs clear and concise briefings. Actively encourages communication flow from crew/team members. Consistently verifies correct understanding.
Critical tasks are ignored or forgotten with no apparent organisation or prioritisation. Time and workload not managed. Significant safety concerns placing the aircraft at risk.	**Workload Management** ↔ 1 2 3 4 5	Tasks are organised with the critical ones given priority. Time and workload are consistently managed effectively. Aware of their own level of workload. Delegation is used appropriately.
Very poor performance and considerable concerns over safety and ability to operate effectively - significant retraining required	**General Performance** ↔ 1 2 3 4 5	Overall performance of an excellent standard, but is not necessarily perfect or unattainable. Performance could be used as a teaching example for others.

Instructors Assessment – Pilot 2

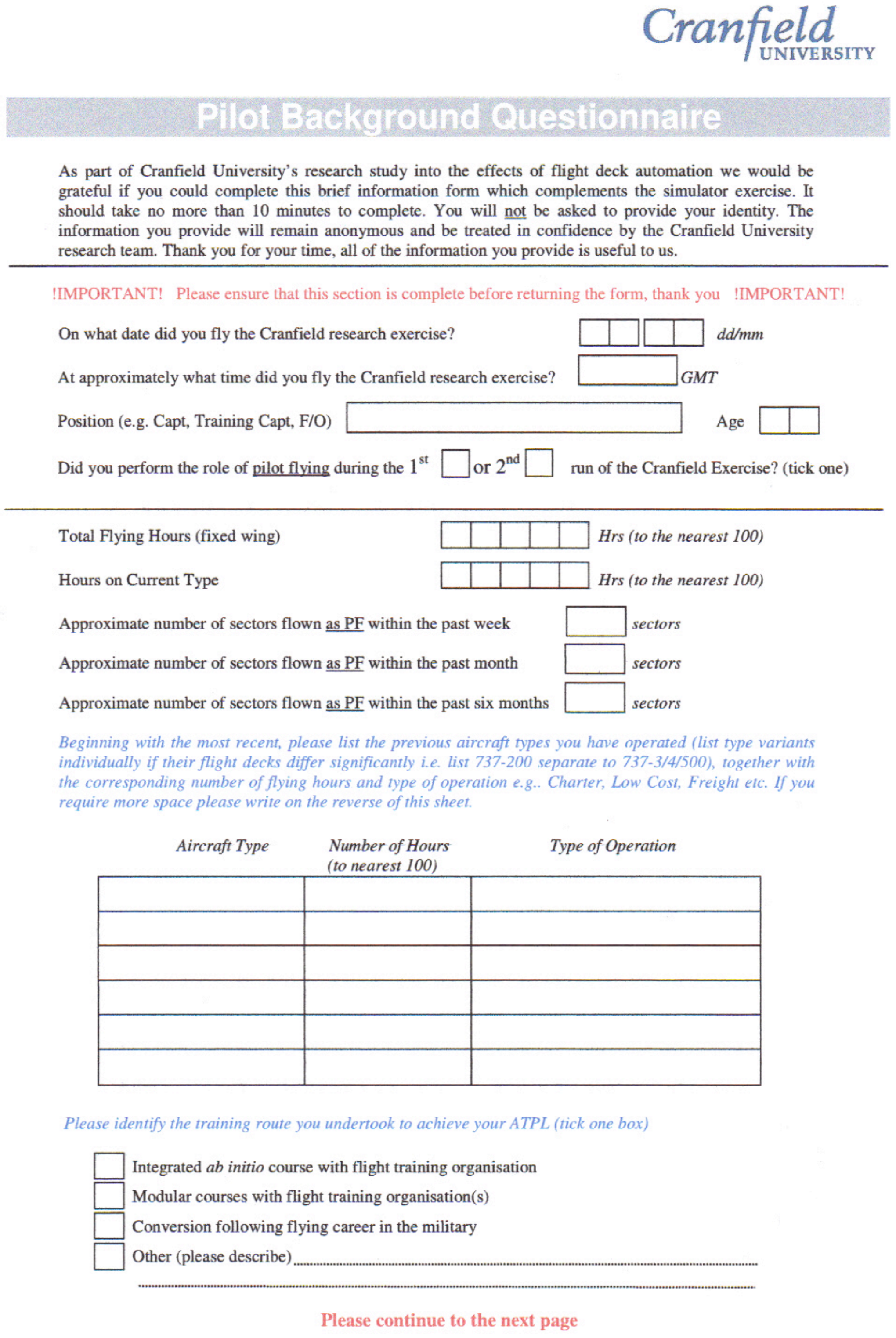
Very Poor		*Excellent*
Frequent or sustained deviations beyond allowable tolerances with little or no attempt to correct. Lack of positive aircraft control.	**Manual Handling** ↔ 1 2 3 4 5	Very smooth, accurate and concise control, minor deviations anticipated and corrected promptly. Clear application of sound techniques at all times.
Fails to recognise or react to a clearly evident and developing situation that will cause the aircraft to breach clearances, violate procedures, or place it at risk.	**Situational Awareness** ↔ 1 2 3 4 5	Consistently plans for most significant factors. Regularly updates, confirms and checks using available resources. Continually anticipates and manages threats.
Unable to determine applicable aircraft limits, systems, performance, rules of thumb or procedures. Displays extremely limited level of knowledge.	**Knowledge** ↔ 1 2 3 4 5	Demonstrates thorough understanding of aircraft limits, systems, performance, rules of thumb and procedures. Displays a high level of knowledge.
Vital information not relayed to other crew/team members. Reluctant to establish communication flow with other crew/team members.	**Communication** ↔ 1 2 3 4 5	Employs clear and concise briefings. Actively encourages communication flow from crew/team members. Consistently verifies correct understanding.
Critical tasks are ignored or forgotten with no apparent organisation or prioritisation. Time and workload not managed. Significant safety concerns placing the aircraft at risk.	**Workload Management** ↔ 1 2 3 4 5	Tasks are organised with the critical ones given priority. Time and workload are consistently managed effectively. Aware of their own level of workload. Delegation is used appropriately.
Very poor performance and considerable concerns over safety and ability to operate effectively - significant retraining required	**General Performance** ↔ 1 2 3 4 5	Overall performance of an excellent standard, but is not necessarily perfect or unattainable. Performance could be used as a teaching example for others.

Appendix H

Boeing 737 Classic Series - Main Panel

Appendix I

Demographic Pro-Forma



Cranfield UNIVERSITY

Pilot Background Questionnaire

As part of Cranfield University's research study into the effects of flight deck automation we would be grateful if you could complete this brief information form which complements the simulator exercise. It should take no more than 10 minutes to complete. You will not be asked to provide your identity. The information you provide will remain anonymous and be treated in confidence by the Cranfield University research team. Thank you for your time, all of the information you provide is useful to us.

!IMPORTANT! Please ensure that this section is complete before returning the form, thank you !IMPORTANT!

On what date did you fly the Cranfield research exercise? ☐☐ ☐☐ *dd/mm*

At approximately what time did you fly the Cranfield research exercise? ☐ *GMT*

Position (e.g. Capt, Training Capt, F/O) ☐ Age ☐☐

Did you perform the role of pilot flying during the 1st ☐ or 2nd ☐ run of the Cranfield Exercise? (tick one)

Total Flying Hours (fixed wing) ☐☐☐☐☐ *Hrs (to the nearest 100)*

Hours on Current Type ☐☐☐☐☐ *Hrs (to the nearest 100)*

Approximate number of sectors flown as PF within the past week ☐ *sectors*

Approximate number of sectors flown as PF within the past month ☐ *sectors*

Approximate number of sectors flown as PF within the past six months ☐ *sectors*

Beginning with the most recent, please list the previous aircraft types you have operated (list type variants individually if their flight decks differ significantly i.e. list 737-200 separate to 737-3/4/500), together with the corresponding number of flying hours and type of operation e.g.. Charter, Low Cost, Freight etc. If you require more space please write on the reverse of this sheet.

Aircraft Type	*Number of Hours (to nearest 100)*	*Type of Operation*

Please identify the training route you undertook to achieve your ATPL (tick one box)

- ☐ Integrated *ab initio* course with flight training organisation
- ☐ Modular courses with flight training organisation(s)
- ☐ Conversion following flying career in the military
- ☐ Other (please describe)..

....

Please continue to the next page

Recent Flying Experience

Please consider your flying experiences whilst acting as the PF over the previous 6 months

At what altitude did you typically engage the autopilot following takeoff? ☐☐☐☐☐ *ft*

At what altitude did you typically disengage the autopilot prior to landing? ☐☐☐☐☐ *ft*

Approximately what percentage of the approaches that you flew during this period were classified as Precision Approaches? ☐ %

Please consider the definition of a 'manual approach' to be an approach where the autopilot is disengaged either before or upon commencing the final approach (i.e. requiring a prolonged period of manual flight). The approach may be flown with or without the use of autothrottle or flight director systems.

Approximately how many days have passed since you last flew a manual approach? ☐ *days*

Approximately how many manual approaches have you flown in the past week? ☐ *approaches*

Approximately how many manual approaches have you flown in the past 2 weeks? ☐ *approaches*

Approximately how many manual approaches have you flown in the past month? ☐ *approaches*

Approximately how many manual approaches have you flown in the past 6 months? ☐ *approaches*

Do you feel that your manual flying skills have been affected as a result of operating a highly automated aircraft? (Please tick one box)

☐ *Yes, they've improved*
☐ *No, they've not changed*
☐ *Yes, they've deteriorated*

Do you regularly participate (or have you previously participated) in flying activities outside of airline work or initial training? (please tick the appropriate boxes and indicate how frequently/when you participated)

	How many hours? (approx)	When? (e.g. 1997-present)
☐ Gliding (non aerobatic)		
☐ GA powered, fixed wing (non aerobatic)		
☐ GA rotary wing		
☐ Aerobatics (powered or gliding)		
☐ Flight instruction (powered or gliding)		
☐ Other (please describe)..		

The following questions relate specifically to the Cranfield simulator exercise. Please mark the appropriate box.

How much mental effort did the task require?	*Insignificant*	1 2 3 4 5	*Excessive*
How much physical effort did the task require?	*Insignificant*	1 2 3 4 5	*Excessive*
How much time pressure did you feel?	*Insignificant*	1 2 3 4 5	*Excessive*

**Thank you for taking the time to provide this information.
If you have any further comments regarding this questionnaire or the research exercise, please feel free to write them on the reverse of this sheet.**

Please return this form to the member of staff who provided it or the deposit box in the administration office.

Appendix J

General Performance Metric Results

Straight & Level Segment Performance Metrics – description of distribution

Straight and Level Tracking Performance				
	Min	Max	Mean	SD
Spatial Tracking Error				
Altitude ME (ft.)	-328.066	171.787	4.178	79.678
Altitude SDE (ft.)	0.724	18.441	5.277	3.501
Heading ME (deg.)	-12.822	8.103	-4.497	3.764
Heading SDE (deg.)	0.055	0.586	0.239	0.132
Control Wheel Power (degs²/hz)				
Very Low Frequency Band	1254	74627	15478	14551
Low Frequency Band	162	14437	4345	2849
Mid Frequency Band	335	12288	2900	2375
High Frequency Band	413	11617	3355	2398
Very High Frequency Band	200	4471	1524	1081
Control Column Power (degs²/hz)				
Very Low Frequency Band	4	590	120	128
Low Frequency Band	1	59	13	14
Mid Frequency Band	1	42	7	7
High Frequency Band	0	10	4	2
Very High Frequency Band	0	9	2	2
Rudder Power (degs²/hz)				
Very Low Frequency Band	242	1786	597	291
Low Frequency Band	4	51	16	9
Mid Frequency Band	2	18	6	4
High Frequency Band	1	9	3	2
Very High Frequency Band	0	4	1	1
Throttle Power (degs²/hz)				
Very Low Frequency Band	67924	210268	101144	27887
Low Frequency Band	1104	6058	2513	871
Mid Frequency Band	366	1671	741	270
High Frequency Band	195	883	404	143
Very High Frequency Band	113	362	201	55

ILS Segment Performance Metrics – description of distribution

ILS Tracking Performance				
	Min	Max	Mean	SD
Spatial Tracking Error				
Localiser ME (dots)	-0.542	2.179	0.146	0.354
Localiser SDE (dots)	0.006	0.103	0.030	0.022
Glideslope ME (dots)	-1.098	0.203	-0.024	0.177
Glideslope SDE (dots)	0.003	0.032	0.011	0.006
Airspeed ME (kts)	-10.502	14.115	0.458	5.529
Airspeed SDE (kts)	0.083	0.666	0.268	0.140
Control Wheel Power (degs^2/hz)				
Very Low Frequency Band	1253	67281	13020	11138
Low Frequency Band	339	30891	6341	6451
Mid Frequency Band	775	48763	5018	6760
High Frequency Band	1381	32733	8371	7065
Very High Frequency Band	656	23725	4599	4096
Control Column Power (degs^2/hz)				
Very Low Frequency Band	3	260	62	58
Low Frequency Band	2	47	12	9
Mid Frequency Band	1	25	8	6
High Frequency Band	1	16	5	3
Very High Frequency Band	1	17	4	3
Rudder Power (degs^2/hz)				
Very Low Frequency Band	476	1834	1042	321
Low Frequency Band	10	81	28	12
Mid Frequency Band	1	63	7	10
High Frequency Band	1	25	5	6
Very High Frequency Band	1	11	3	2
Throttle Power (degs^2/hz)				
Very Low Frequency Band	61730	112806	82965	10523
Low Frequency Band	1209	4129	2170	553
Mid Frequency Band	179	1554	418	331
High Frequency Band	123	1076	285	219
Very High Frequency Band	95	804	235	162

Missed Approach Performance Metrics – description of distribution

Missed Approach Tracking Performance				
	Min	Max	Mean	SD
Spatial Tracking Error				
Track ME (deg.)	-16.169	16.158	-2.897	4.559
Track SDE (deg.)	0.137	3.005	0.635	0.486
Airspeed ME (kts)	-4.516	21.766	5.479	6.859
Airspeed SDE (kts)	0.152	1.492	0.758	0.352
Control Wheel Power (degs²/hz)				
Very Low Frequency Band	900	110807	17219	20156
Low Frequency Band	1597	60983	10843	10651
Mid Frequency Band	889	22272	6648	5433
High Frequency Band	525	32395	6124	5821
Very High Frequency Band	117	10201	1967	2079
Control Column Power (degs²/hz)				
Very Low Frequency Band	3	459	73	88
Low Frequency Band	2	84	17	16
Mid Frequency Band	1	20	7	5
High Frequency Band	1	24	7	6
Very High Frequency Band	0	21	4	4
Rudder Power (degs²/hz)				
Very Low Frequency Band	1370	6357	3303	1013
Low Frequency Band	32	194	83	41
Mid Frequency Band	8	157	45	35
High Frequency Band	8	140	33	27
Very High Frequency Band	1	27	9	6
Throttle Power (degs²/hz)				
Very Low Frequency Band	126383	247849	169576	26460
Low Frequency Band	1821	5215	3350	768
Mid Frequency Band	678	2205	1258	342
High Frequency Band	474	1685	994	306
Very High Frequency Band	170	545	309	94

Appendix K

Correlation of General Flying Experience with Performance

Correlations between performance metric score and number of fixed wing hours on the straight and level segment of the exercise.

TRE-Performance Metric Correlations (Straight & Level)

	r	N	p
Spatial Tracking Error			
Altitude ME (ft.)	-.010	49	.949
Altitude SDE (ft.)	-.103	49	.489
Heading ME (deg.)	-.029	49	.844
Heading SDE (deg.)	-.026	49	.861
Control Wheel Power (degs²/hz)			
Very Low Frequency Band	-.046	49	.760
Low Frequency Band	-.141	49	.923
Mid Frequency Band	-.053	49	.388
High Frequency Band	.143	49	.534
Very High Frequency Band	-.126	49	.978
Control Column Power (degs²/hz)			
Very Low Frequency Band	-.090	49	.545
Low Frequency Band	.070	49	.641
Mid Frequency Band	.163	49	.272
High Frequency Band	.214	49	.149
Very High Frequency Band	.101	49	.500
Rudder Power (degs²/hz)			
Very Low Frequency Band	.275	49	.062
Low Frequency Band	.162	49	.277
Mid Frequency Band	.075	49	.615
High Frequency Band	.161	49	.278
Very High Frequency Band	-.02	49	.892

Correlations between performance metric score and total fixed wing hours on the ILS tracking segment of the exercise.

TRE-Performance Metric Correlations (ILS)	r	N	p
Spatial Tracking Error			
Localiser ME (dots)	-.187	49	.199
Localiser SDE (dots)	-.072	49	.623
Glideslope ME (dots)	-.017	49	.910
Glideslope SDE (dots)	-.005	49	.970
Airspeed ME (kts)	.053	49	.715
Airspeed SDE (kts)	.080	49	.582
Control Wheel Power (degs²/hz)			
Very Low Frequency Band	-.226	49	.119
Low Frequency Band	-.183	49	.207
Mid Frequency Band	-.062	49	.672
High Frequency Band	.275	49	.056
Very High Frequency Band	-.054	49	.715
Control Column Power (degs²/hz)			
Very Low Frequency Band	.195	49	.179
Low Frequency Band	-.133	49	.362
Mid Frequency Band	-.058	49	.694
High Frequency Band	-.219	49	.131
Very High Frequency Band	-.062	49	.674
Rudder Power (degs²/hz)			
Very Low Frequency Band	-.164	49	.261
Low Frequency Band	-.237	49	.101
Mid Frequency Band	-.207	49	.154
High Frequency Band	-.106	49	.467
Very High Frequency Band	-.192	49	.187

Correlations between performance metric score and total fixed wing hours on the missed approach segment of the exercise.

TRE-Performance Metric Correlations (Missed Approach)			
	r	N	p
Spatial Tracking Error			
Track ME (deg.)	.010	49	.947
Track SDE (deg.)	.086	49	.564
Airspeed ME (kts)	-.281	49	.055
Airspeed SDE (kts)	-.109	49	.465
Control Wheel Power (degs^2/hz)			
Very Low Frequency Band	-.252	49	.088
Low Frequency Band	-.208	49	.161
Mid Frequency Band	-.080	49	.591
High Frequency Band	.249	49	.092
Very High Frequency Band	-.077	49	.608
Control Column Power (degs^2/hz)			
Very Low Frequency Band	.216	49	.144
Low Frequency Band	-.124	49	.408
Mid Frequency Band	-.099	49	.510
High Frequency Band	-.274	49	.063
Very High Frequency Band	-.127	49	.394
Rudder Power (degs^2/hz)			
Very Low Frequency Band	-.224	49	.131
Low Frequency Band	-.259	49	.079
Mid Frequency Band	-.202	49	.173
High Frequency Band	-.102	49	.496
Very High Frequency Band	-.158	49	.288