BOUNDARY AVOIDANCE TRACKING: CONSEQUENCES (AND USES) OF IMPOSED BOUNDARIES ON PILOT-AIRCRAFT PERFORMANCE

THESIS

Ryan D. Blake, Major, USAF

AFIT/GAE/ENY/09-M03

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GAE/ENY/09-M03

BOUNDARY AVOIDANCE TRACKING: CONSEQUENCES (AND USES) OF IMPOSED BOUNDARIES ON PILOT-AIRCRAFT PERFORMANCE

THESIS

Presented to the Faculty

Department of Aeronautics and Astronautics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Aeronautical Engineering

Ryan D. Blake, BS

Major, USAF

March 2009

AFIT/GAE/ENY/09-M03

BOUNDARY AVOIDANCE TRACKING: CONSEQUENCES (AND USES) OF IMPOSED BOUNDARIES ON PILOT-AIRCRAFT PERFORMANCE

Ryan D. Blake, BS
Major, USAF

Approved:

Paul A. Blue, Maj, USAF (Thesis Advisor)	5 Mar 09 Date
Bradley S. Liebst, Dr. (Member)	5 Mar 09 Date
Richard E. Huffman Jr., Maj, USAF (Member)	5 Mar 09 Date
William R. Gray III (Member)	3 MAR 09 Date

AFIT/GAE/ENY/09-M03

Abstract

The vast majority of aircraft today are *piloted* (whether from inside the cockpit or remotely), and therefore performance and stability are based on a relationship between pilot and aircraft. Understanding this pilot-aircraft relationship is vital for aircraft design and mishap prevention. In 2004 a new concept of the way pilots control aircraft was developed. Termed Boundary Avoidance Tracking (BAT), this concept introduced the idea that often pilots control an aircraft not in an attempt to maintain some condition, but to avoid some real or perceived boundaries. The BAT concept has proven to be useful: 1) in explaining pilot reaction in certain scenarios; 2) in providing objectivity to pilot opinion ratings; and 3) in its use as a flight test technique to artificially increase pilot workload and expose poor handling qualities.

In order to take advantage of the BAT concept, it needs to be fully understood and modeled. This research used ground simulation and flight test to give pilots roll angle tracking tasks, with decreasing boundaries imposed around the task, in order to further understand and refine BAT theory. The experimental data were successfully matched to the BAT model predictions, providing validation of the model. Pilot behavior was studied during the tasks, and the pilot's physical workload was shown to increase as boundaries decreased. These decreasing boundaries were also shown to elicit maximum performance from the pilot at the minimum achievable boundary size. Pilot workload was quantified during these tasks through the use of a secondary task. A new method of combining this workload data with performance data was developed in order to objectively quantify aircraft handling qualities.

To my wife and sons,

who graciously gave me the time I needed to complete this thesis,

and to my God,

who graciously gave me life.

Acknowledgements

This project is the result of concerted effort by many people, without whom this thesis could not have been completed.

I would like to thank Bill Gray from Test Pilot School, who originated the BAT theory which provided the basis for this thesis. He also served as my advisor at TPS, providing guidance throughout all phases of test planning, execution, and data analysis, as well as flying as a safety pilot on many of the test sorties.

Thanks also go to my advisor Paul Blue, who spent much time with me at AFIT developing a direction for this research, and who also agreed to continue as my advisor even after he had moved on from AFIT to another job. He took his personal time to work with me to produce a good thesis, and I am deeply grateful.

I would also like to thank my thesis committee for the time and effort that they spent in reviewing this thesis. In addition to Paul Blue and Bill Gray, the committee included Dr. Brad Liebst and Major Richard Huffman.

The people of the Simulation Branch, Air Vehicles Directorate, Air Force Research Laboratory also provided invaluable help during two sets of simulator tests in two different simulator platforms. Without their help, I wouldn't have two-thirds of the data that I have now. Special thanks go to Jeff Slutz, Amy Linklater, and Mike Coy.

Dan Wheeler and Julie Cagle, two human factors engineers from Edwards AFB, provided all of the support for the eye-tracker equipment used in the LAMARS simulator, and their help is greatly appreciated.

Finally, I completed the last two-thirds of this project as part of a five-person test team at TPS. Much of the work and analysis for this project is due to their efforts, and I look forward to working with them in flight test in the future. They are: Nancy Baccheschi, Glenn Cooley, Chris Keithley, and Chun Hao Lee.

Ryan D. Blake

Table of Contents

Page

Abstract ... **v**

Acknowledgements ... **vii**

Table of Contents ... **viii**

List of Figures ... **xii**

List of Tables ... **xiv**

1. Introduction ... **1**

1.1. Motivation ... 1

1.1.1. Importance of Understanding the Pilot-Aircraft System ... 1

1.1.2. Classical Pilot Modeling and the BAT Model ... 2

1.1.3. Potential BAT Model Uses ... 4

1.2. Research Objectives ... 5

1.3. Research Overview ... 6

1.3.1. Using the Roll Axis ... 6

1.3.2. Characterizing the BAT Model ... 7

1.3.3. Characterizing Pilot Behavior ... 8

1.3.4. Characterizing Pilot Performance ... 8

1.3.5. Characterizing Pilot Opinion Rating in Relationship to a Workload Buildup Task ... 9

1.4. Preview of Results ... 10

1.5. Thesis Overview ... 12

2. BAT Model Details and Experimental Background 13

2.1. Characterizing the BAT Model 13

2.1.1. Mathematical Details 13

2.1.2. Conceptual Demonstration.. 16

2.1.3. Comparison with Experiment 18

2.2. Characterizing Pilot Behavior 21

2.3. Characterizing Pilot Performance... 23

2.4. Characterizing Pilot Opinion Rating in Relationship to a Workload Buildup Task 26

2.5. Summary of Past Research.. 28

3. Experiment Design 30

3.1. Designing for Success 30

3.2. Test Team 31

3.3. Test Platforms... 31

3.4. Test Profiles.. 33

3.4.1. Workload Buildup Profile 35

3.4.2. Workload Buildup plus Secondary Task Profile.................................. 36

3.4.3. Cooper-Harper Profile.. 37

3.5. Test Elements 38

3.5.1. Aircraft Models 38

3.5.2. Tracking Task.. 45

3.5.3. Boundaries 48

3.5.4. Secondary Task 49

3.6. Test Procedure 53

3.7. Test Briefing and Order 54

3.8. Test Subjects 54

3.9. Summary of Experiment Design 55

4. Test Results 56

4.1. Characterizing the BAT Model 56

4.1.1. Model Matching Procedure 57

4.1.2. Numerical Optimization Procedure 57

4.1.3. Modification to the BAT Model 60

4.1.4. Model Matching Example 61

4.1.5. Model Matching Results 64

4.1.6. Summary of BAT Model Characterization Results 68

4.2. Characterizing Pilot Behavior 68

4.3. Characterizing Pilot Performance 72

4.4. Characterizing Pilot Opinion Rating in Relationship to a Workload Buildup Task 75

4.4.1. Cooper-Harper Ratings 76

4.4.2. Objective Performance and Workload Measurements 78

4.4.3. Comparing Subjective and Objective Results 79

4.4.4. Summary of the Relationship Between Pilot Opinion Rating and Performance/Workload in a Workload Buildup Task 86

5. Conclusions 87

5.1. Contributions to Test Methodology 87

5.2. Characterizing the BAT Model 88

5.3. Characterizing Pilot Behavior 89

5.4. Characterizing Pilot Performance 90

5.5. Characterizing Pilot Opinion Rating in Relationship to a Workload Buildup Task 90

5.6. Areas for Further Research 92

Appendix A – Cooper-Harper Scale (Cooper and Harper, 1984) 93

Appendix B – VISTA Flight Test Cards 94

List of Abbreviations, Symbols, and Acronyms 96

Bibliography 98

List of Figures

Page
Figure 1. Block Diagram of the BAT model 13

Figure 2. A Damped System with No Boundary Tracking (Gray, 2005)......................... 16

Figure 3. A Single Instance of Boundary Tracking (Gray, 2005) 16

Figure 4. Two Instances of Boundary Tracking (Gray, 2005).. 17

Figure 5. Onset of Boundary-Driven Oscillation (Gray, 2005).. 17

Figure 6. Boundary Tracking Correlation Case 1 (Gray, 2005) 19

Figure 7. Boundary Tracking Correlation Case 2 (Gray, 2005) 20

Figure 8. Effect of BAT Parameters on Tracking Success (Gray, 2005) 20

Figure 9. Pilot Duty Cycle and Aggressiveness Theory (Gray, 2007) 22

Figure 10. Achieved and Desired Performance Theory (Gray, 2007).............................. 24

Figure 11. BAT DART Tracking Error (Dotter, 2007) 25

Figure 12. Statistical Pilot Opinion Rating Analysis (Dotter, 2007) 27

Figure 13. Qualitative Pilot Opinion Rating Analysis (Dotter, 2007).............................. 28

Figure 14. Infinity Cube Fixed-Base Simulator... 32

Figure 15. LAMARS Motion-Base Simulator.. 32

Figure 16. VISTA Test Aircraft.. 33

Figure 17. Workload Buildup Profile HUD Symbology (VISTA).................................. 35

Figure 18. Workload Buildup plus Secondary Task HUD Symbology (VISTA) 36

Figure 19. Cooper-Harper Profile HUD Symbology (VISTA) 37

Figure 20. Tracking Task Comparison 47

Figure 21. Comparison of Secondary Task Candidates.. 50

Figure 22. Typical Pilot Crosscheck Frequency (LAMARS) 52

Figure 23. Typical Pilot Crosscheck Frequency and Secondary Task Performance (LAMARS) 53

Figure 24. Identifying Areas of Boundary Tracking 62

Figure 25. Fitting the Boundary Tracking Loop of the BAT Model 63

Figure 26. Fitting Both Loops of the BAT Model 64

Figure 27. BAT Model Parameters Sorted by Pilot 65

Figure 28. BAT Model Parameters Sorted by Aircraft Model 66

Figure 29. Example of Increasing Stick Force as Boundaries Decrease 70

Figure 30. Average Stick Force for an 11º Terminal Boundary Size (LAMARS) 70

Figure 31. Average Duty Cycle and Aggressiveness (VISTA) 71

Figure 32. Average Roll Error (VISTA) 73

Figure 33. Contrasting Pilot Performance as Requirements Increase 74

Figure 34. Infinity Cube Cooper-Harper Ratings 76

Figure 35. LAMARS Cooper-Harper Ratings 77

Figure 36. VISTA Cooper-Harper Ratings 77

Figure 37. Determining Critical Boundary Size 79

Figure 38. Workload Buildup and Cooper-Harper Ratings (Infinity Cube) 80

Figure 39. Workload Buildup and Cooper-Harper Ratings (LAMARS) 81

Figure 40. Workload Buildup and Cooper-Harper Ratings (VISTA) 81

Figure 41. Workload Buildup and Aircraft Models (LAMARS) 84

Figure 42. Workload Buildup and Cooper-Harper Ratings for Individual Pilots 86

List of Tables

Page

Table 1. Boundary Tracking Input Equations 15

Table 2. Fixed-Base Simulation Model Parameters 40

Table 3. Motion-Base Simulation Model Parameters 42

Table 4. Flight Test Model Parameters 44

Table 5. Tracking Task Frequency Content 47

BOUNDARY AVOIDANCE TRACKING: CONSEQUENCES (AND USES) OF IMPOSED BOUNDARIES ON PILOT-AIRCRAFT PERFORMANCE

1. Introduction

1.1. Motivation

1.1.1. Importance of Understanding the Pilot-Aircraft System

The vast majority of aircraft today are *piloted* (whether from inside the cockpit or remotely), and therefore performance and stability are a function not just of the characteristics of the aircraft, but of the pilot-aircraft system. It is thus impossible to discuss the suitability of an aircraft to a mission or task without understanding how the pilot will interact with the aircraft in the conduct of that mission or task.

The qualities of this interaction are termed the *handling qualities* of the aircraft. Handling qualities depend on the characteristics of both the pilot and the aircraft. In contrast, the characteristics of an aircraft in response to impulses, steps, ramps, sinusoids, etc. are termed the *flying qualities* of the aircraft. Flying qualities depend only on the characteristics of the airplane, such as the aerodynamics and control systems. The problem is that, unlike engineer-designed control systems, which can be analyzed and predicted because they were designed to be this way, pilots are not easily analyzed or predicted. In other words, flying qualities are fairly easily predicted, but handling qualities are much less easy to predict. Because of this difficulty, evaluating handling qualities is currently a subjective process, based primarily on pilot opinion. In addition, the pilot's lack of predictability forces much of the evaluation to be done after the airplane has been built, when problems are costly and difficult to fix. It is necessary, therefore, to research and understand pilot behavior, and to compile this knowledge into a usable model.

Understanding the pilot-aircraft system is important in design, where money (and perhaps lives) can be saved by ensuring good handling qualities, and identifying and correcting bad handling qualities, before the aircraft is ever built. Despite the best efforts of engineers, however, problems often surface after an aircraft is built. In this case, a good pilot model can explain the deficiency and suggest improvements.

1.1.2. Classical Pilot Modeling and the BAT Model

Attempts were made at developing a useful pilot model as early as World War II. Through the years, research and experiments have developed these early concepts into usable mathematical models, such as the Crossover Model, the Structural-Isomorphic Model, and the Optimal Control Model (McRuer, 1989). All of these models assume that the pilot produces control actions based on observed error between a desired state and an actual state, be it *regulation* (e.g. keep the wings level) or *tracking* (e.g. climb/descend/ turn to the appropriate altitude/heading). This type of control will be labeled *point tracking* in this thesis.

In 2004, a new concept challenged this assumption. Developed by Mr. William Gray at the US Air Force Test Pilot School (TPS), the concept of Boundary Avoidance Tracking (BAT) assumes that a pilot often acts to avoid (rather than maintain) a specific condition. As an illustration, Mr. Gray envisioned bicycling along a foot-wide stripe on a road, which would be an easy task for any bicycler. Suspending this same foot-wide stripe above the Grand Canyon, however, makes the task remarkably more difficult, even though the performance requirements are the same. Classical operator modeling would explain that the fear of failure drives the operator gain to a high level during attempts to maintain the centerline. However, intuition and experience suggest that the bicycler is focused on the *edges* (i.e. boundaries) of the path rather than the centerline, and in fact may lose awareness of the centerline altogether as control is used to avoid the boundaries (Gray, 2004).

Taking an aviation example, a pilot who experiences a strong sideways wind-gust while attempting to land will most likely maneuver to avoid striking a wingtip on the ground, rather than maneuver to return to level flight. The distinction is critical, because the first type of maneuver (ground avoidance) will possibly result in an overshoot of level flight and cause the opposite wingtip to approach the ground. If this continues, a Pilot-In-the-loop Oscillation (PIO) will result. In fact, the origins of PIO were the original motivation for the idea of Boundary Avoidance Tracking.

BAT, then, involves control forces to move *away* from something, and the force decreases as the boundaries move farther away. This is in direct opposition to point tracking, in which control forces are applied to move *toward* something, and the force decreases as the point gets closer.

Boundaries causing BAT could be deadly (as in the case of the path over the Grand Canyon), or they could be damaging but not deadly (as with a structural g-limit), or they could simply be uncomfortable (as with negative g-loads).

The BAT model is not meant to replace the other pilot models, but rather supplement them. The BAT model is a switching model, in which pilots switch between point tracking and boundary tracking as required. When there are no boundaries present, the BAT model collapses to a classical pilot model. In BAT theory, a pilot spends most of the time in point tracking (modeled by any of the classical models mentioned above), and switches to boundary tracking only when necessary.

1.1.3. Potential BAT Model Uses

If the discussion ended here, BAT would simply be an interesting idea. If it can be modeled appropriately and confirmed experimentally, however, it has several potential uses. In fact, some research has already been done to detail the BAT model and confirm it with experiment, and some of the following possibilities are already in use:

- Prediction of aircraft PIO susceptibility (in the presence of boundaries, as in the sideways wind-gust PIO described above) during aircraft design. This requires a validated mathematical pilot model that could be combined with an aircraft model for analysis.
- Exposure of handling qualities deficiencies and PIO tendencies during flight testing. This requires a method to artificially increase pilot workload in order to expose poor handling qualities. Previously, a flight test technique (FTT) known as Handling Qualities During Tracking (HQDT) was used to attempt to force a pilot to use "high-gain" control inputs in a tracking task. HQDT required pilots to track a target as closely as possible, accepting absolutely no error. The problem was that pilots are taught from the beginning that attempting to track with no error

will produce *poor* results, and it was thus very difficult for them to overcome this resistance and force themselves to fly in a manner that they know is unacceptable. Using the BAT concept by requiring tracking in the presence of shrinking boundaries (what has become known as the Workload Buildup FTT) has solved this problem. In the Workload Buildup FTT, the shrinking boundaries in some sense *force* the pilot to track the target with as little error as possible, resulting in maximum performance at the smallest achievable boundary size (Gray, 2007).

- Quantification of handling qualities, supplementing *subjective* pilot comments with *objective* performance measurements in the presence of steadily decreasing artificial boundaries. This assumes that aircraft with good handling qualities would be able to operate in the presence of smaller boundaries than aircraft with poor handling qualities.

1.2. Research Objectives

In just over four years, BAT theory has evolved from an interesting idea into a mathematical model supported by simulator and in-flight experiments, and into a flight test technique designed to aid in the description of aircraft handling qualities and PIO susceptibility. However, there is still work to do. Some of the potential uses mentioned above are yet to be explored, and doing so requires a fuller characterization of the model. In addition, some already demonstrated aspects can be improved or shown in a different light. The research objectives were formulated to address areas for improvement in the BAT theory. The four main objectives for this research were:

- Characterize the mathematical BAT model through comparison with simulator and flight test data
- Characterize pilot behavior (in terms of control stick forces and pilot physical workload) in the presence of decreasing boundaries
- Characterize pilot performance (in terms of tracking error) in the presence of decreasing boundaries
- Characterize the relationship between pilot opinion rating and performance/ workload in a Workload Buildup task

1.3. Research Overview

The research methodology was developed in order to support the objectives listed above. The following sections describe the focus of the research.

1.3.1. Using the Roll Axis

All previous BAT research was in the pitch axis. This research explored the BAT phenomenon in the roll axis for the following reasons:

- Previous BAT flight tests involved large changes in normal acceleration as tracking tasks were followed. These g-loads provided a potential (and unwanted) secondary boundary, because pilot comfort or aircraft limits can in actuality be bigger threats than a simulated boundary. In the presence of unintentional secondary boundaries, pilots may react differently than if the simulated primary boundaries were actual boundaries. This corrupts the data and makes BAT model characterization difficult. The roll axis has no such limits, and can provide "cleaner" data. Cleaner data can result in a more precise mathematical model with better predictive and explanatory ability.

- While pitch PIOs gather a lot of attention, roll PIOs have occurred, often with disastrous results. Several C-17 mishaps have been due to roll PIO during landing, and a lateral-directional PIO was listed as a possible cause of the 2001 crash of an Airbus A300-600 in New York during departure after takeoff (Hess, 2003). Since the BAT model holds the promise of predicting and explaining PIO, it is important to gather experimental data in all axes for which PIO occurs.
- A promising correlation has been shown between pilot opinion ratings and Workload Buildup performance in the longitudinal axis, offering an objective result as a supplement to a subjective rating. It seems obvious that these results should hold in the roll axis, but it still needs to be shown. The use of the Workload Buildup FTT in roll axis handling qualities testing may expose deficiencies before they occur in operational flying, and suggest possible corrective action.

1.3.2. Characterizing the BAT Model

The first potential use of the BAT theory listed above is the prediction of aircraft PIO susceptibility during aircraft design. As mentioned, this requires a *validated mathematical pilot model.* The BAT model development prior to this research has been encouraging. The results make intuitive sense, and they match experimental data fairly well. Still, they don't match perfectly, and it is possible that modifications to some parameters or some assumptions could produce better results.

This research almost doubled the amount of BAT data available, and provided the only available data in the roll axis. It compared the BAT model to a large number of tests in order to determine its suitability with as many different situations, configurations, and pilots as possible. The BAT model was matched to the test data through the use of numerical optimization, which was the first time that optimization had been used for BAT research. In addition, whereas previous research had only exercised portions of the mathematical BAT model, this research compared flight test data to the predictions of the entire model.

1.3.3. Characterizing Pilot Behavior

As listed above, the second use of the BAT theory is the exposure of handling qualities deficiencies and PIO tendencies during flight testing. As mentioned, this requires a method to *artificially increase pilot workload* in order to expose poor handling qualities. Previous research had demonstrated the intuitive result that pilots react to boundaries and change their behavior as necessary to avoid boundaries. This research quantified this changed behavior by examining how often, how fast, and how forcefully the pilot moves the control stick. The relatively new concepts of pilot duty cycle and aggressiveness were used to provide the quantitative measurements. These results were used to validate the Workload Buildup FTT as a method of increasing pilot workload, in order to evaluate aircraft handling qualities during periods of elevated pilot workload.

1.3.4. Characterizing Pilot Performance

In addition to elevating pilot workload, the use of the BAT theory during flight test (in the form of the Workload Buildup FTT) also attempts to *elicit maximum* *performance* from the pilot. Past research had suggested that decreasing boundaries improve pilot performance until the boundaries become too tight and make the task too difficult. Previous flight tests, using a tracking task in the longitudinal axis, suggested that a hidden boundary (in the form of pilot aversion to negative g-loads) caused the pilots to intentionally accept larger tracking error at the beginning of the task in order to decrease the amount of negative g-loads (Dotter, 2007). As boundaries decreased, the pilots were forced to accept these g-loads in order to stay within the boundaries; thus, the tracking performance improved as the boundaries tightened. By using the roll axis, this research eliminated the additional boundary and studied pilot performance to see if decreasing boundaries still improved pilot performance. These results were used to validate the Workload Buildup FTT as a method of eliciting maximum performance from a pilot.

1.3.5. Characterizing Pilot Opinion Rating in Relationship to a Workload Buildup Task

The third use of the BAT theory listed above is the quantification of handling qualities, supplementing *subjective* pilot comments with *objective* performance measurements. A set of previous flight tests attempted to relate pilot opinion ratings to performance during a Workload Buildup task. The performance was defined as the length of time the tracking task was flown before the boundaries were exceeded. This single parameter was compared to the pilot opinion rating. Pilot opinion ratings, however, are based on performance and workload. The uniqueness of the Workload Buildup flight test technique is that, at the smallest achievable boundary size, it forces pilots to their highest workload, and thus results in the best performance achievable with maximum workload. It cannot, however, determine what the pilot workload was at any point other than the end. It is possible that two different aircraft could have the same performance and workload requirements at the end (i.e. their total test length is the same), but one may require a constantly high workload throughout, whereas the other has a workload "spike" in the presence of close boundaries. This is important information that would be lost by reducing ratings to a single parameter.

This research quantified the workload during the test by incorporating a secondary task to be performed as pilots continue to perform the primary tracking task. This workload information was combined with the performance results in order to more fully quantify the relationship between BAT and pilot opinion ratings.

1.4. Preview of Results

The methodology developed in this research provided tasks and techniques which not only provided useful data for this research, but can also be used in future research. The following are the contributions made by this research to BAT test methodology:

- A roll angle tracking task was developed and confirmed to be challenging enough to keep the pilot constantly in the loop, which is important when researching pilot-in-the-loop handling qualities. This is in contrast to the tracking task used in previous flight tests, which was too predictable and allowed the pilot to back out of the loop for long periods of time.

- A new flight test technique was developed to measure performance and workload during a Workload Buildup task, using a primary tracking task and a secondary task.
- Performance of the secondary task was validated as a measure of pilot workload by using eye-tracking equipment during ground simulation to measure pilot attention to the primary and secondary tasks.

The results of this research validated some previous assumptions and suggested new conclusions. The following are the contributions made by this research to the BAT theory and model:

- Numerical optimization was used to choose the BAT model parameters that provided the best match to experimental data, in contrast to previous research in which the matching was done qualitatively.
- The entire BAT model was matched to test data, whereas all previous research had only used the boundary tracking portion of the switching model.
- Based on test results, the BAT model was modified to more accurately model pilot inputs during boundary tracking.
- Pilot physical workload was shown to increase as boundaries decreased. This was the first time that this was demonstrated in the roll axis, and it provided validation for the Workload Buildup FTT as a method to artificially increase pilot workload.
- Pilot performance (in terms of tracking error) was shown, on average, to improve as boundaries tightened. This was a validation of previous findings, but

demonstrated for the first time in the roll axis. This result validated the Workload Buildup FTT as a method of eliciting maximum performance from a pilot.

- A new method of combining performance and workload data from Workload Buildup tasks was developed in order to quantify aircraft handling qualities.

1.5. Thesis Overview

The current chapter has laid the foundational concepts of pilot modeling in general and Boundary Avoidance Tracking in specific, providing its history and the reasons for moving the theory forward. Chapter 2 goes on to provide the mathematical details to these concepts, describing the specifics of past successes and pointing out the current deficiencies in knowledge. Chapter 3 describes the experiment design, in the buildup process from fixed-base simulation through motion-base simulation and flight test. Chapter 4 discusses the results from simulator and flight tests. Finally, Chapter 5 summarizes the conclusions drawn from the experiments, and gives recommendations for their use, as well as areas for further research.

2. BAT Model Details and Experimental Background

2.1. Characterizing the BAT Model

The BAT concept is straightforward and intuitive, but in order to use it as a predictive tool, a model must be built describing exactly *how* pilots react to boundaries.

2.1.1. Mathematical Details

Mr. Gray hypothesized a model based on switching behavior on the part of the pilot (Gray, 2005). Figure 1 graphically depicts this behavior:

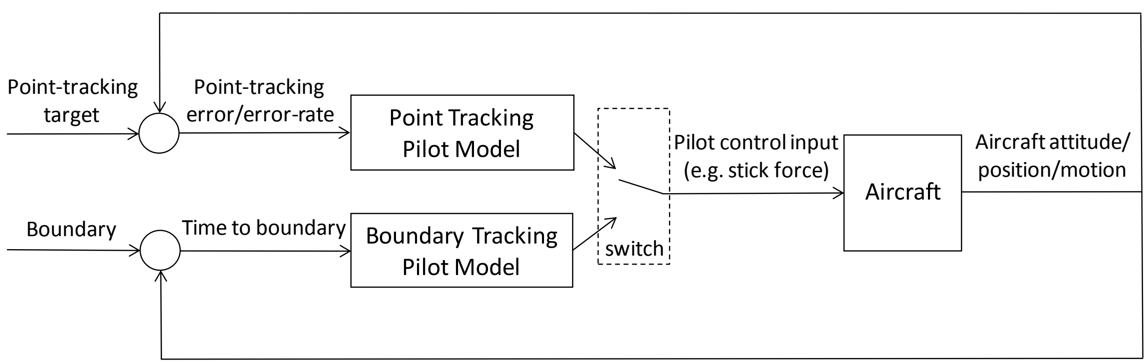


Figure 1. Block Diagram of the BAT model

During point tracking, pilot control input is based on the upper loop, acting on the aircraft position, rate, etc. with respect to the point-tracking target. This loop can be modeled by any classical pilot model. When boundaries become a factor, however, pilot control input is based on the lower loop, acting on the time to boundary (calculated using the displacement from and rate toward the boundary). The time to boundary determines when and with what force the pilot reacts to the boundaries. The pilot selects the larger of the two control forces, switching to the boundary tracking loop as boundaries become a factor, and switching back to point tracking as the boundaries recede. The switch in the diagram represents this decision.

Pilot switching behavior is not a new concept. The Dual-Mode Controller model has been used to explain pilot response to step inputs, and describes different pilot behavior for different phases of compensation (McRuer, 1967). A model in which a pilot monitors numerous error signals and acts on them one at a time has also been proposed (AFFTC, 2002). The proposed BAT model, however, is the first to apply switching behavior to something other than point tracking.

In the hypothesized model, there are four parameters describing pilot characteristics in the presence of boundaries (Gray, 2005):

t_{min} = time to boundary when the pilot begins to react to boundaries (a larger value of t_{min} implies that the pilot begins reacting to boundaries sooner)

t_{max} = time to boundary when the pilot uses maximum gain to avoid boundaries (a larger value of t_{max} implies that the pilot applies maximum force sooner)

K_{bm} = maximum gain used by the pilot to avoid boundaries

τ_b = pilot time-delay during boundary avoidance

The time to boundary (t_b) is computed using the displacement from and rate toward the boundary. The pilot input is hypothesized to increase linearly from zero at t_{min} to K_{bm} at t_{max}, according to equation 1:

$$\frac{t_{min} - (t_b + \tau_b)}{t_{min} - t_{max}} K_{bm} \quad (1)$$

Note that t_{min} is greater than t_{max} (minimum input occurs at a greater time to boundary than maximum input). If the boundaries are exceeded (which implies that they *can* be exceeded without catastrophic results), K_{bm} is held until the position is again within the boundaries. Table 1 shows the details of the boundary tracking input computation.

Table 1. Boundary Tracking Input Equations

Situation	Perceived Threat from Boundaries	Boundary Tracking Input
Displacement inside, moving away from boundary	No threat	0
Displacement inside, moving toward boundary		
$t_b > t_{min}$	No threat	0
$t_b = t_{min}$	Onset of threat	0
$t_{max} < t_b < t_{min}$	Threat increases from no threat to maximum threat	$\frac{t_{min} - (t_b + \tau_b)}{t_{min} - t_{max}} K_{bm}$ where $t_b = \frac{x_b}{(dx_b/dt)}$ $x_b = distance\ to\ boundary$
$t_b = t_{max}$	Onset of maximum threat	K_{bm}
$t_b < t_{max}$	Maximum threat	K_{bm}
Displacement outside boundary	Maximum threat	K_{bm}

These parameters and control laws are used to predict pilot response, and it is important to determine how they vary from pilot to pilot and event to event, or even if they fully represent pilot behavior in the presence of boundaries. They are hypothesized parameters which must be confirmed or modified through experiment. The first objective of this research is to provide this confirmation.

2.1.2. Conceptual Demonstration

After building the model, Mr. Gray demonstrated its performance with computer simulation. Hypothetical boundary-avoidance situations were simulated with the BAT model controlling a moderately-damped 2nd-order plant subject to disturbance inputs in the direction of a boundary. Only the boundary tracking loop of the model was implemented; the point tracking loop was not used. Therefore, the following model responses were due solely to boundary tracking (Gray, 2005):

1. When the boundaries were far enough away such that the time to boundary was never less than t_{min}, no boundary avoidance inputs were made, and the system response damped out naturally (Figure 2).

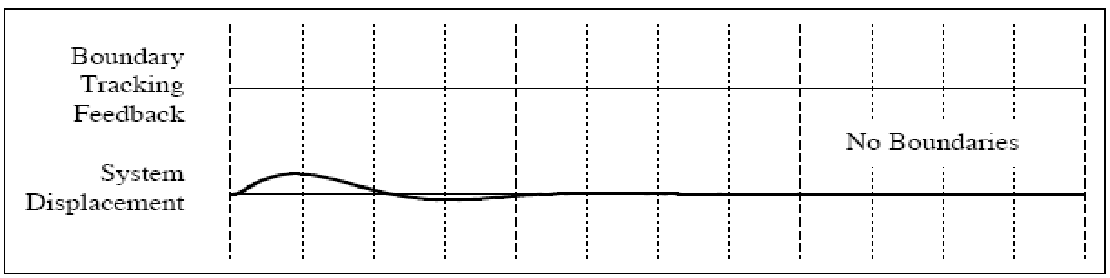


Figure 2. A Damped System with No Boundary Tracking (Gray, 2005)

2. When the boundaries were moved closer such that the disturbance made the time to boundary less than t_{min}, a single boundary avoidance input was made, and the system response then damped out naturally (Figure 3).

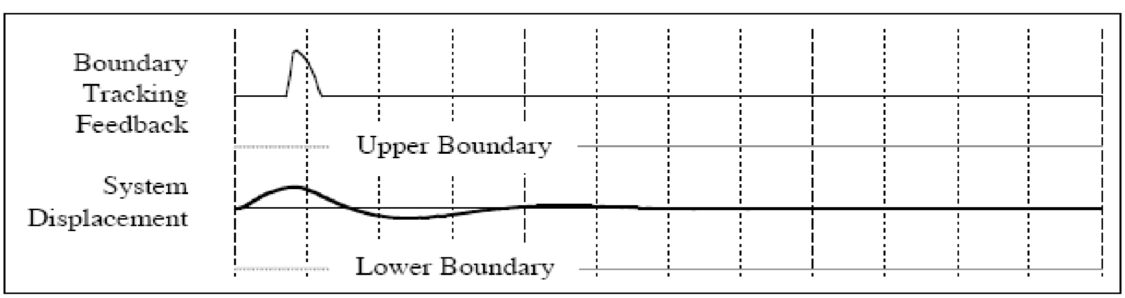


Figure 3. A Single Instance of Boundary Tracking (Gray, 2005)

3. If the boundaries were moved even closer, a boundary avoidance input was made as before, but in this case the input forced the system toward the opposite boundary, and an opposite boundary avoidance input was applied before the system finally damped out naturally (Figure 4).

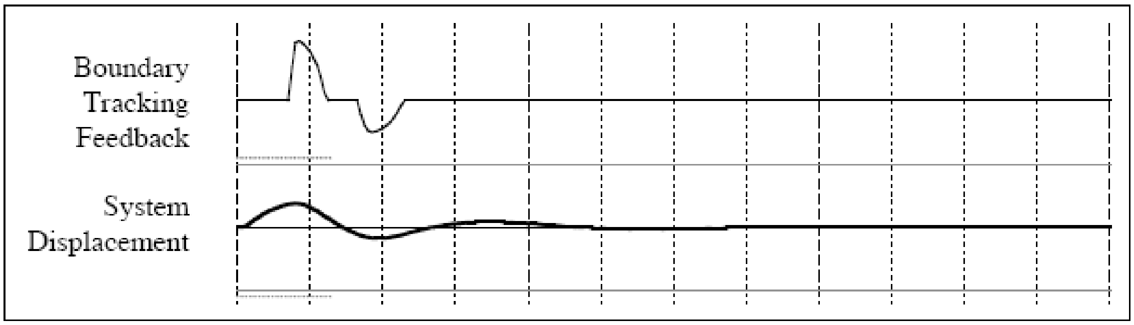


Figure 4. Two Instances of Boundary Tracking (Gray, 2005)

4. As the boundaries moved closer, this input-overshoot-opposite input pattern was repeated, and in fact developed into a PIO (Figure 5).

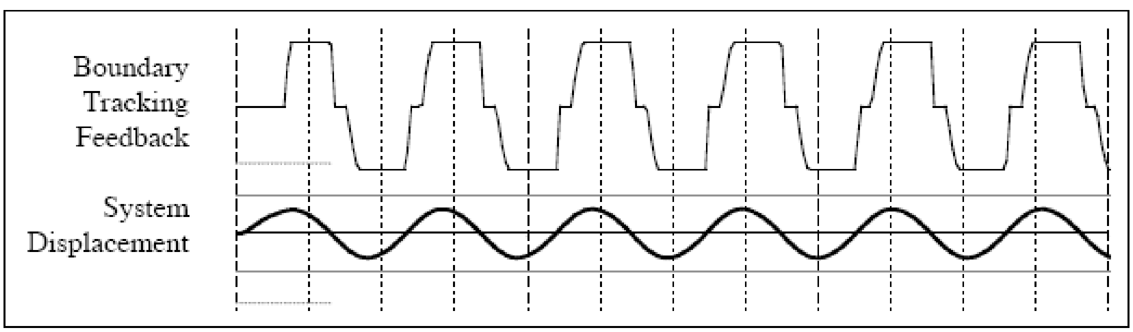


Figure 5. Onset of Boundary-Driven Oscillation (Gray, 2005)

Note that this response seems intuitive, and seems to describe the way pilots would actually act if attempting to avoid boundaries. It does this without a point-tracking model at all; in other words, the predictions are based solely on avoiding boundaries, not the attempt to maintain zero displacement.

2.1.3. Comparison with Experiment

The above demonstrations seem intuitive, and demonstrate that the model is coherent, but they do not demonstrate that actual pilots behave this way. This requires comparison with experimentally-recorded pilot behavior. To make such a comparison, a desktop-computer simulation was created, based on an aerodynamic model of a North American Navion aircraft, with varied time-delay and elevator rate limit (Gray, 2005). The object of the simulation was to follow a lead aircraft as it climbed and descended. The simulation used a simple compensatory display in which only the *difference* between the altitude of the lead aircraft and the test aircraft was displayed, along with error boundaries that decreased by 25% every 60 seconds. Subjects were instructed to treat the boundaries as deadly, and in fact the task stopped when a boundary was exceeded. Time histories of altitude error, error rate, and elevator displacement were recorded.

The predictions of the boundary tracking pilot model (the lower loop of Figure 1) were compared to the experimental data. Note that no point tracking model was used. The data were studied to identify possible BAT events, and these time-segments were evaluated assuming that the pilot had already switched to boundary avoidance inputs, and therefore *all* pilot inputs were boundary avoidance inputs. Thus, the full extent of the pilot switching behavior was not explored. The model depicted in Figure 1 was evaluated only from the *time to boundary* signal to the *pilot control input* signal.

The experimentally-recorded error and error rate were used to compute a time history of the time to boundary, which was used in the boundary tracking pilot model to predict pilot control input (as in the demonstrations of the previous section). The only input to the model was the actual t_b time history, and the output was the BAT predicted pilot control deflection. The four BAT parameters were adjusted for each case so that the BAT model predictions matched the experimental data as close as possible. The parameters were adjusted using an iterative process in which the parameters were changed and the subsequent results were qualitatively evaluated. Each test case was therefore described by the values of the four parameters.

The following two figures show examples of the degree to which the BAT model correctly predicted pilot input, based on the assumption that the entire time-segments studied were BAT events. The predictions agree quite well with the data, but there were actual pilot inputs in the data which were not predicted by the model. Since only the boundary tracking loop of the model was used, these inputs were theorized to be point tracking inputs.

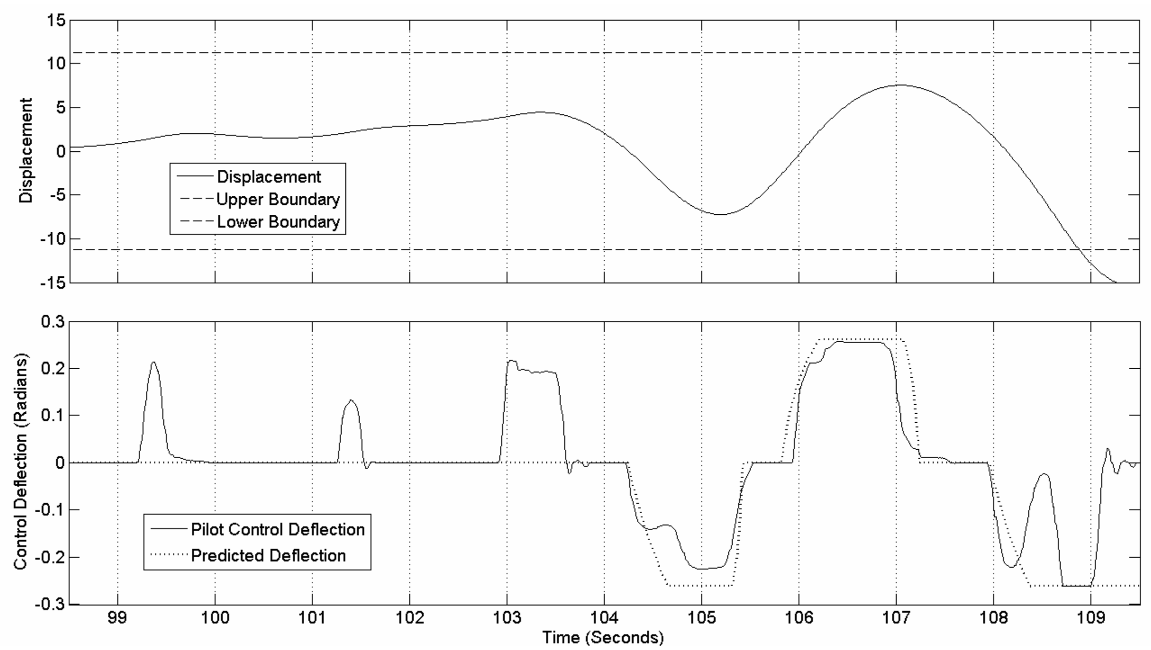


Figure 6. Boundary Tracking Correlation Case 1 (Gray, 2005)

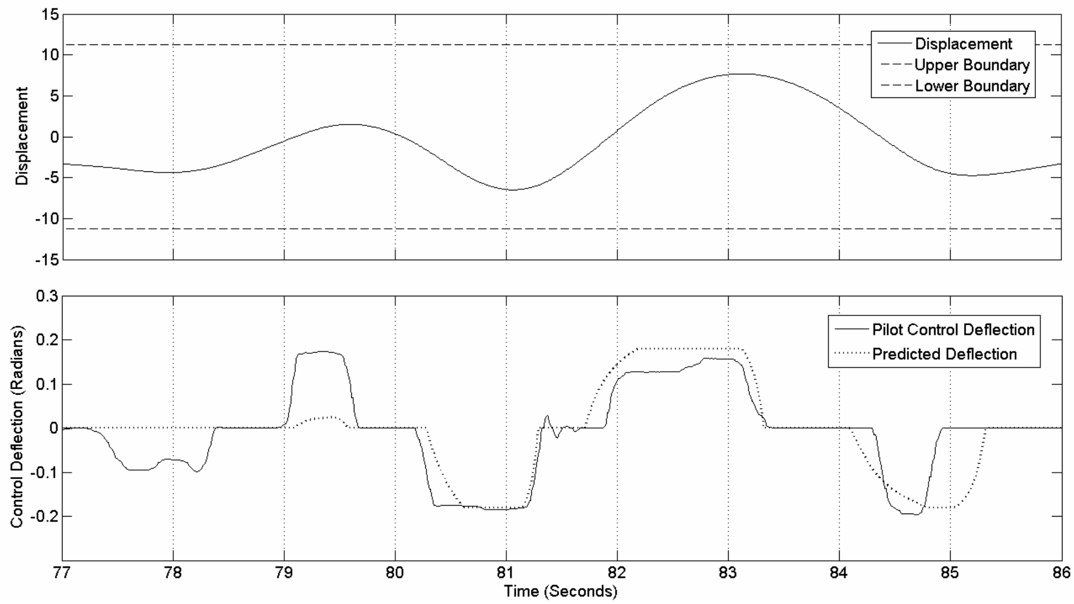


Figure 7. Boundary Tracking Correlation Case 2 (Gray, 2005)

A total of 27 tests were conducted, and the BAT model parameters were fit to each test. The resulting values of t_{min} and t_{max} were then plotted against total test length (i.e. the length of time before the boundaries were exceeded) in an attempt to find a correlation between BAT parameters and task performance (Figure 8).

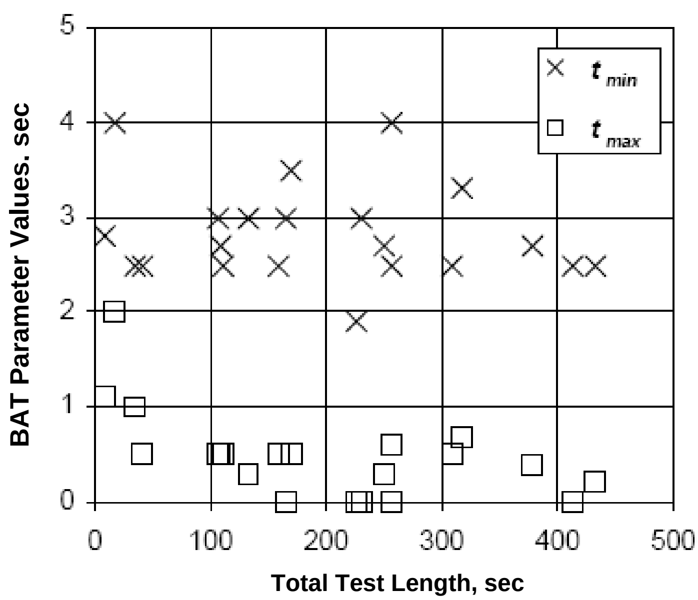


Figure 8. Effect of BAT Parameters on Tracking Success (Gray, 2005)

No strong correlation was found between the individual parameters and total tracking time; however, subjects with a high value of t_{max} tended to do worse, indicating that they selected maximum gain too soon, which resulted in over-controlling and an early boundary excursion.

2.2. Characterizing Pilot Behavior

In addition to describing pilot behavior at particular instances of time, with a particular time to boundary, the BAT theory also describes how pilot behavior changes over time, as boundaries are decreased. The effect of boundaries on pilot behavior can be seen by studying pilot control stick (inceptor) movement. If the average difficulty of the tracking task remains constant with time (by repeating the task to coincide with each change in boundary size), then any change in pilot inceptor movement should be due to the changing boundaries. To adequately describe this change, Mr. Gray proposed describing pilot inceptor movement with the new concepts of pilot duty cycle and aggressiveness (Gray, 2007). These two parameters are collectively called pilot inceptor workload.

Duty cycle describes how often the pilot is moving the controls. It is defined as the percentage of time that the pilot is moving the controls (above a pre-defined noise threshold). Aggressiveness describes how fast the pilot is moving the controls. It is defined as the root-mean-square of the control force (or displacement) rate. These two concepts sum up pilot inceptor workload by describing how often and how fast the pilot is moving the controls. Figure 9 shows the relationship between the two parameters.

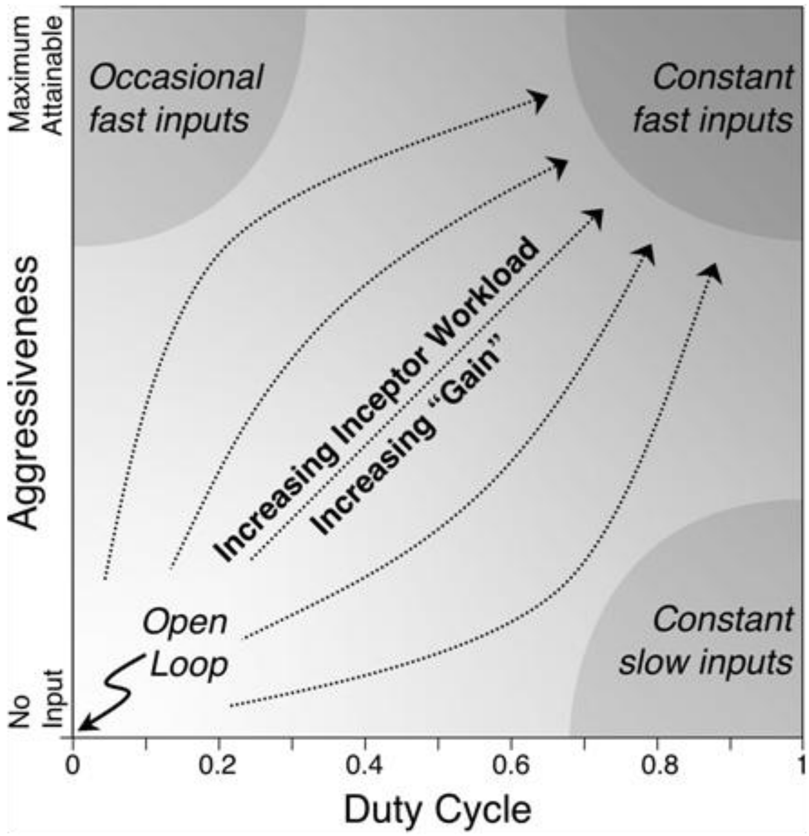


Figure 9. Pilot Duty Cycle and Aggressiveness Theory (Gray, 2007)

The Workload Buildup FTT (developed from BAT theory) is based on the assumption that decreasing boundaries will increase pilot workload, and the technique is therefore assumed to be useful in order to evaluate aircraft handling qualities during periods of elevated pilot workload. The concepts of duty cycle and aggressiveness provide a way to measure the increased pilot workload and validate the Workload Buildup FTT.

The HAVE BAT flight tests at TPS (Warren, 2006) provided the first flight test data used for BAT research. These tests used the Workload Buildup FTT in order to investigate pilot behavior in the presence of boundaries. The test aircraft (a T-38) was required to fly close formation with a target aircraft (another T-38 which maneuvered longitudinally), while maintaining position within certain visual references. An aircraft-mounted camera was placed in the test aircraft to record these visual references for use in data analysis. The data proved very difficult to extract from the video, and only limited BAT model characterization was possible. In addition, problems in data recording did not allow calculation of duty cycle and aggressiveness. The tests were successful, however, in showing that a PIO often occurred when the boundaries were tightened sufficiently, suggesting that the pilot workload had, in fact, increased.

The BAT DART tests (Dotter, 2007) also used the Workload Buildup FTT, but improvements in data recording allowed control stick forces to be accurately recorded. These tests used the Variable-stability In-flight Simulator Test Aircraft (VISTA), a variable-stability F-16. The VISTA solved the problems inherent in the T-38 camera system by displaying a tracking task and boundaries in the Heads-Up Display (HUD), thereby eliminating the need for a target aircraft and video cameras, replacing them with on-board software and data recording devices. It also allowed the simulation of four different aircraft-response models in one aircraft during the same flight. The data from the BAT DART tests were analyzed by Mr. Gray using duty cycle and aggressiveness (Gray, 2007). The data showed a general trend of increasing duty cycle and aggressiveness as boundaries tightened. This result was not true in all cases, but it did show some promise, and so this research further examined the concept.

2.3. Characterizing Pilot Performance

Prior to the advent of the BAT theory, Mr. Gray developed a theory of the relationship between required and achieved performance during a piloting task. This theorized relationship (shown in Figure 10) suggests that as performance tolerances become tighter, pilots will respond with increased performance, until the tolerances become so tight that the pilot is unable to perform as required. A PIO often occurs at this point.

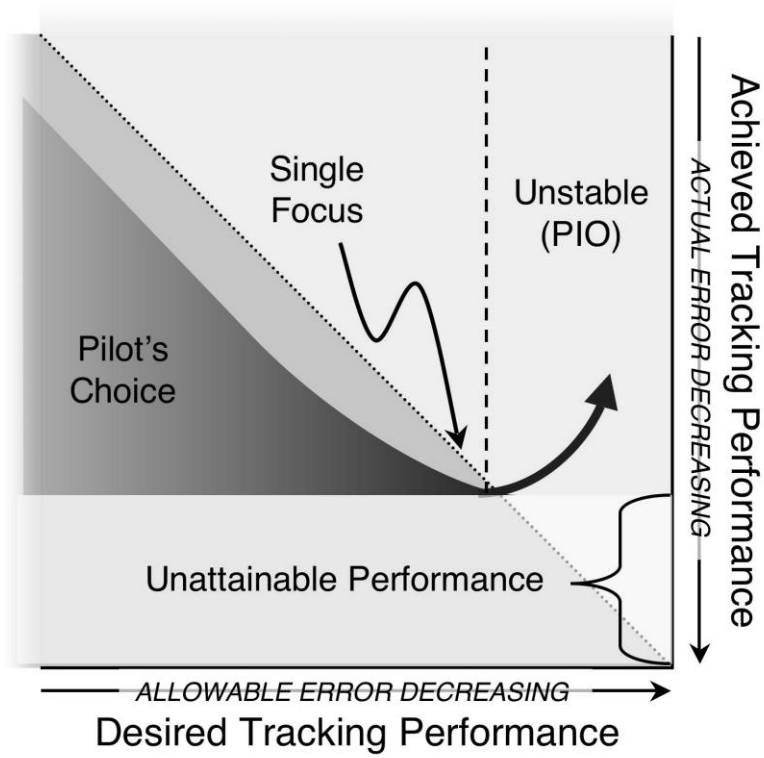


Figure 10. Achieved and Desired Performance Theory (Gray, 2007)

The Workload Buildup FTT provides a method to test the relationship between required and achieved performance. In the Workload Buildup FTT, the shrinking boundaries represent changing performance requirements. According to the theory, pilot tracking error should decrease as the boundaries tighten, until the pilot can no longer remain within the boundaries.

As mentioned in the previous section, the HAVE BAT flight tests were successful in showing that a PIO often occurred when the boundaries were tightened sufficiently.

However, data recording problems prevented the measurement of tracking error as a function of boundary size, and so the relationship between required and achieved performance could not be quantified.

Improvements in test design and data recording during the BAT DART tests allowed performance to be recorded as a function of boundary size. Figure 11 shows the average tracking error as a function of boundary size for three of the aircraft models used during the flight tests. The data support the theoretical relationship shown in Figure 10, suggesting that achieved tracking performance was being altered by the required tracking performance. The data do not show the dramatic rise in error that is suggested in Figure 10, but this is likely due to the fact that the task stopped when a boundary was exceeded. This did not leave time for increased error to have much of an effect on the average error for the final boundary size.

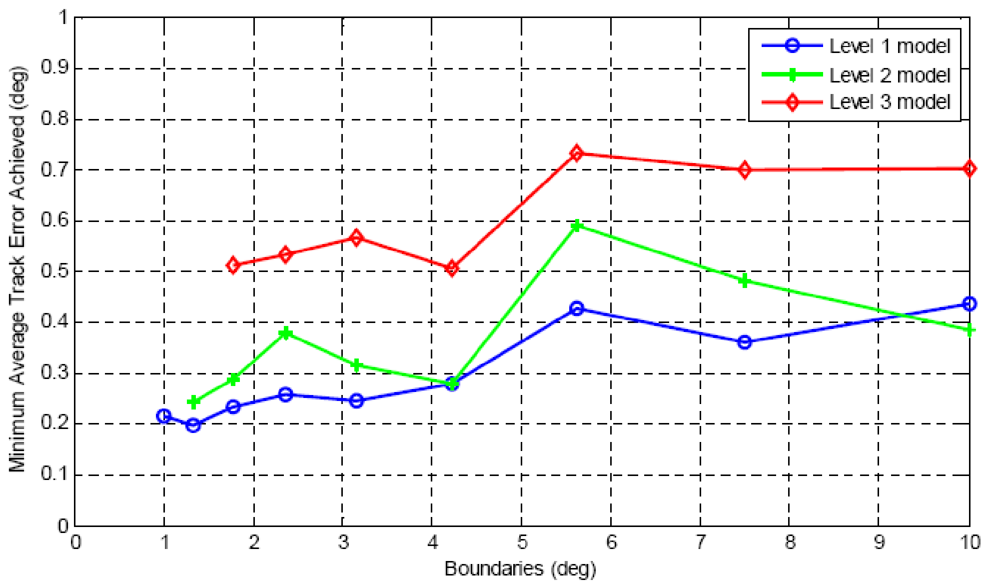


Figure 11. BAT DART Tracking Error (Dotter, 2007)

It was noted, however, that this trend was clearer in flight test than in ground simulation. It was theorized that during flight test, pilots altered their behavior at the beginning of the profile (when the boundaries were wide) in order to minimize the required negative g-loads. This aversion to negative g-loads, then, served as a hidden boundary. The ground simulator was unable to faithfully reproduce the negative g-loads experienced during flight, and therefore pilots did not react as they did in flight test.

2.4. Characterizing Pilot Opinion Rating in Relationship to a Workload Buildup Task

As discussed above, the Workload Buildup FTT is used to artificially increase pilot workload (by shrinking boundaries) up to the point of instability or unacceptable performance. This point is the performance limit for a particular pilot-vehicle combination. Intuitively, some aircraft (among a large spread of pilots) should be able to operate successfully in the presence of smaller boundaries than others, and should probably also receive higher pilot opinion ratings. If a definite correlation between performance during a Workload Buildup FTT (defined as the length of time before boundaries are exceeded or, alternatively, the minimum achievable boundary size) and pilot opinion rating could be found, this technique could be used to make objective measurements in what is currently a subjective aircraft rating process.

This was the focus of the BAT DART tests. In a succession of tests from desktop simulation, through fixed-base and motion-base simulation, to flight test, aircraft models with different handling qualities were flown during identical Workload Buildup tasks. The task was a longitudinal tracking task with shrinking boundaries on both sides of the target. As in the desktop simulations conducted by Mr. Gray, the pilots were instructed to treat the boundaries as deadly, and the task stopped when a boundary was exceeded. Pilot-opinion ratings (using the test community standard Cooper-Harper ratings (Appendix A)) and corresponding performance (defined as the length of time before boundaries were exceeded) were analyzed, and were shown to have a qualitative correlation.

Figure 12 shows an attempt to find a statistical relationship between Cooper-Harper rating and BAT performance. Based on the limited amount of data, there appeared to be a correlation. However, the Cooper-Harper scale is an ordinal scale, and therefore statistical curve fitting is not technically appropriate. Figure 13 shows the same data presented in a qualitative way, which does not require curve-fitting. A qualitative correlation between Cooper-Harper ratings and Workload Buildup performance is evident.

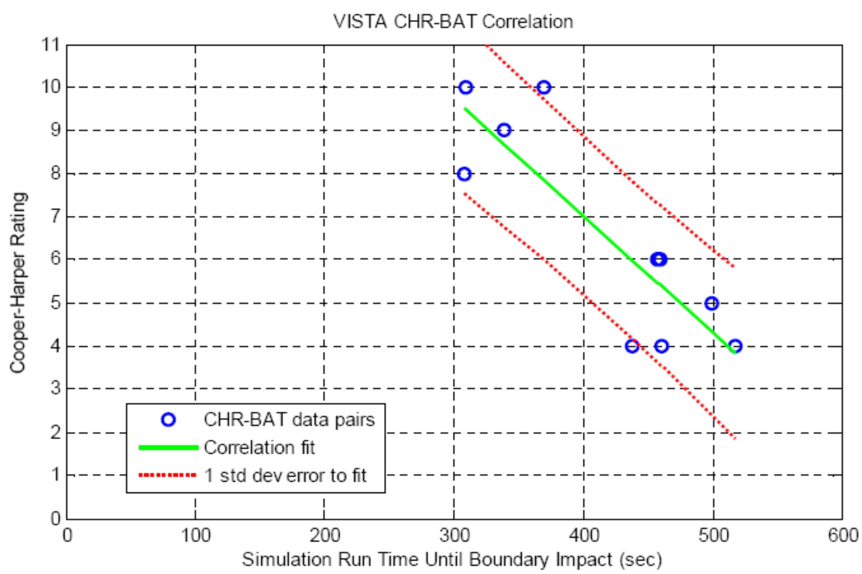


Figure 12. Statistical Pilot Opinion Rating Analysis (Dotter, 2007)

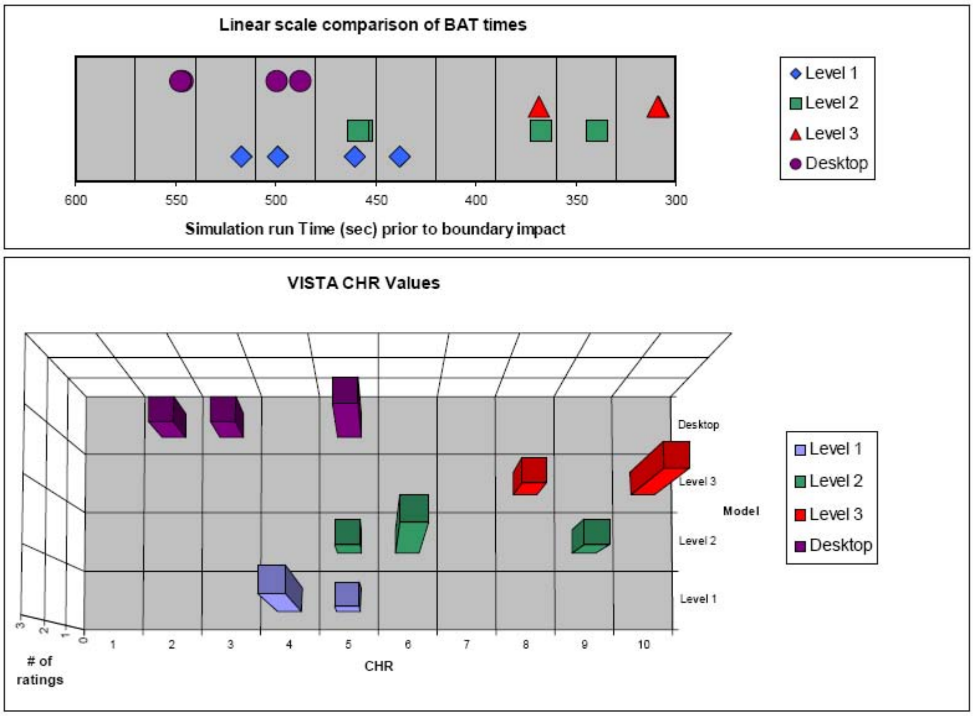


Figure 13. Qualitative Pilot Opinion Rating Analysis (Dotter, 2007)

2.5. Summary of Past Research

The BAT theory has been advanced significantly in the four years since it was first introduced. The initial idea that pilots often act to avoid, rather than maintain, a certain condition has been developed into a usable mathematical model. The model has been demonstrated through batch computer simulations and real-time simulations with a pilot in the loop. The model predictions have been shown to agree well with the experimental data.

The Workload Buildup FTT, based on artificially increasing pilot workload by imposing boundaries on pilot performance, was developed using BAT theory in order to evaluate aircraft handling qualities during periods of elevated pilot workload. Pilot workload has been measured using the new concepts of pilot duty cycle and aggressiveness, collectively called pilot inceptor workload. Flight test has demonstrated that inceptor workload tends to increase as boundaries are tightened.

Pilot performance has also been shown to increase as boundaries are tightened. Flight tests using the Workload Buildup FTT demonstrated that decreasing boundaries tended to elicit maximum performance from pilots. Results from these tests showed that in cases where pilots are prone to accept large errors (for example, in order to avoid physical discomfort), the shrinking boundaries forced the pilots to accept less error in order to remain within the boundaries.

Finally, a promising qualitative correlation has been shown between pilot opinion ratings and Workload Buildup performance. The Workload Buildup FTT offers a method to gather objective data to supplement the subjective ratings and comments that are currently the basis of handling qualities evaluations.

3. Experiment Design

3.1. Designing for Success

As described in Chapter 1, the following were the four main objectives for this thesis:

- Characterize the mathematical BAT model through comparison with simulator and flight test data
- Characterize pilot behavior (in terms of control stick forces and pilot physical workload) in the presence of decreasing boundaries
- Characterize pilot performance (in terms of tracking error) in the presence of decreasing boundaries
- Characterize the relationship between pilot opinion rating and performance/ workload in a Workload Buildup task

The experiment was designed to meet these objectives.

A buildup approach was used to develop the appropriate test techniques and data analysis procedures, in order to maximize the success during flight test. Initial parameters and techniques were evaluated using desktop simulation without test subjects, solely as a means of evaluating the suitability of the aircraft models, tracking task, flight test techniques, and data analysis procedures. These observations were used to design the experiment for the first simulator tests. The first set of simulator tests was conducted in a fixed-base simulator. Lessons learned from this set of tests were used to modify the procedures for the motion-base simulator tests. Likewise, lessons learned from the motion-base simulator tests were used to modify the procedures for the flight tests. Even though the simulator tests were used as a buildup to flight test, the simulator tests were still designed to give usable data. In this way, three sets of data were gathered during the research, which were then analyzed and used for advancing BAT theory.

3.2. Test Team

Initial research, test development, and fixed-base simulation were done by the author at the Air Force Institute of Technology (AFIT) at Wright-Patterson Air Force Base, Ohio. Further test development and execution, including the motion-base simulation and flight test, were accomplished by a five-member test team (including the author) at the Air Force Test Pilot School (TPS) at Edwards Air Force Base, California. The TPS test program was known as AT BAT (Blake and others, 2008).

3.3. Test Platforms

Both simulator tests were conducted at the Air Force Research Laboratories (AFRL) at Wright-Patterson Air Force Base. The Infinity Cube simulator was used for the fixed-base simulation (Figure 14). This simulator featured out-the-window visual displays (200° horizontal by 120° vertical field-of-view), a HUD, a Heads-Down Display (HDD) and a fixed side stick. The Large Amplitude Multi-Mode Aerospace Research Simulator (LAMARS) was used for the motion-base simulation (Figure 15). This simulator was a five degree-of-freedom simulator. (The LAMARS was not capable of translation in the longitudinal axis, but this capability was not considered important for these roll axis tests.) The LAMARS also featured out-the-window visual displays (120° horizontal by 40° vertical field-of-view), a HUD, and a fixed side stick.

Figure 14. Infinity Cube Fixed-Base Simulator

Figure 15. LAMARS Motion-Base Simulator

The NF-16D VISTA was used for flight test (Figure 16). The VISTA was a modified Block 30 F-16D with a custom Digital Flight Control Computer (DFLCC) installed. The Variable Stability System (VSS) in the VISTA was a five degree of freedom system used to simulate various types of aircraft handling qualities, within the control authority and bandwidth of the actuators and flight control surfaces. The front seat was the simulation cockpit and the rear seat had control of the VSS. This project used the fixed side-stick in the VISTA.

Figure 16. VISTA Test Aircraft

3.4. Test Profiles

There were no existing techniques for BAT research in the roll axis. Therefore, considerable time and effort were spent in developing appropriate test techniques. Initial desktop simulations were conducted using a variety of profiles and tracking scenarios, and the most appropriate profiles were chosen after many hours of comparison testing.

The method of measuring pilot workload in support of the fourth objective (characterizing the relationship between pilot opinion rating and performance/workload in a Workload Buildup task) was researched considerably. Pilot inceptor workload (i.e. duty cycle and aggressiveness) was discussed in Chapter 2 as a way to measure pilot workload. This method is useful in showing how a pilot's physical workload increases as boundaries decrease, and was explored in this research when characterizing pilot behavior in the presence of boundaries. However, it does not take mental workload into account. Mental workload is an important component of pilot opinion ratings. Previous research had used secondary tasks to measure operator mental workload (Jex, 1987). The secondary task method was chosen for this research.

Jex's work used secondary tasks to decrease primary task performance to the point of instability. However, the profiles for this research were developed so that pilots were required to sacrifice secondary task performance when necessary to avoid boundary excursions. In this way, decreasing secondary task performance was used to identify increasing workload. The type of secondary task used for these profiles was developed after much research into workload measurement and after much initial computer simulation, and the secondary task refinement continued throughout each test phase in order to produce the most suitable task.

As a result of the test technique development efforts, three test profiles were created in order to support the research objectives. All three profiles were based on a roll tracking task, in which a roll angle target (moving in a random-appearing manner) was displayed as a dotted line in the HUD. The pilot attempted to keep the target aligned with a caret at the top of the HUD. This tracking task was identical in all profiles. These profiles represent a significant contribution to BAT research, because in addition to providing data for this research, they can be used for future BAT tests. The specifics of each profile are described in the following sections.

3.4.1. Workload Buildup Profile

The purpose of the Workload Buildup profile was to observe pilot behavior and performance in the presence of boundaries, and to support the objective of further refining the BAT model. In this profile, boundary lines were displayed on both sides of the target line, and the boundaries decreased every 30 seconds. Pilots were instructed to treat the boundaries as life-threatening, and maneuver as required to prevent exceeding them. The profile stopped automatically when the boundaries were exceeded for 0.5 second. The purpose of the 0.5 second delay was to gather data for pilot behavior outside given boundaries.

Figure 17 shows an example of the flight test HUD symbology for the Workload Buildup profile. Because of lessons learned in ground simulator tests, and because of differences in software implementation, the HUD symbology was slightly different among the test platforms.

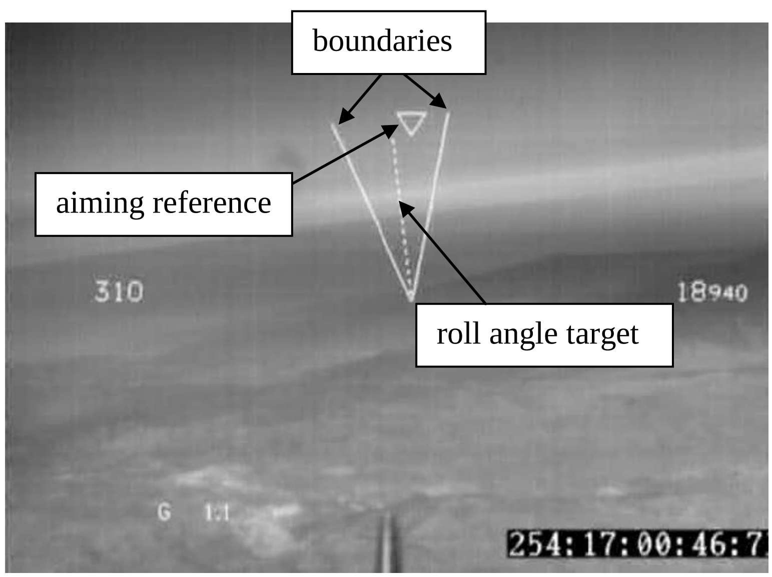


Figure 17. Workload Buildup Profile HUD Symbology (VISTA)

3.4.2. Workload Buildup plus Secondary Task Profile

The purpose of the Workload Buildup plus Secondary Task profile was to objectively measure performance and workload during a tracking task in the presence of boundaries. This profile, coupled with the Cooper-Harper profile discussed next, was designed to explore the relationship between pilot opinion rating and Workload Buildup performance/workload. The profile was identical to the Workload Buildup task, with the addition of a secondary task. The secondary task consisted of four different arrow types which were displayed to the pilot one at a time in a random-appearing sequence. The pilot was required to deflect a switch on the throttle in the direction of the displayed arrow. The pilot was instructed to treat staying within the boundaries as the primary task, and to sacrifice secondary task performance as required to remain within the boundaries. Figure 18 shows an example of the VISTA HUD symbology for the Workload Buildup plus Secondary Task profile. As with the Workload Buildup profile, the HUD symbology was slightly different among the test platforms.

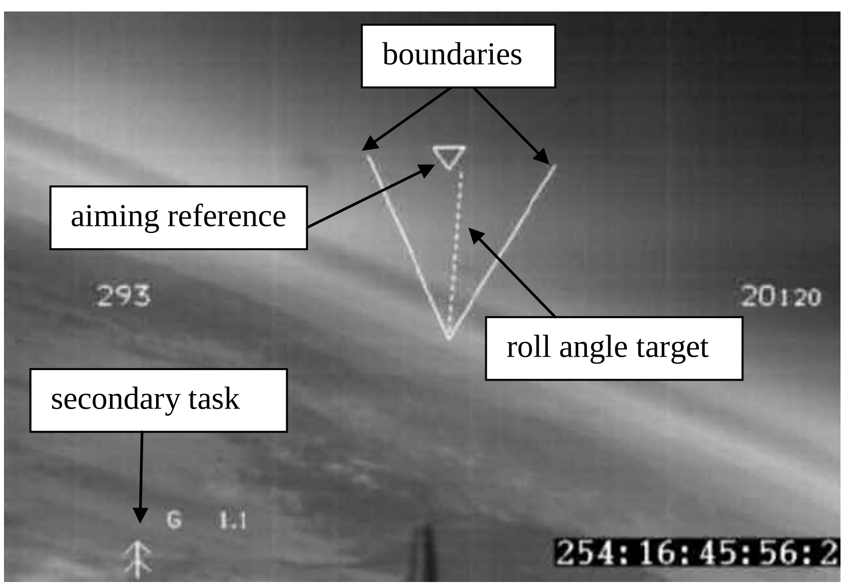


Figure 18. Workload Buildup plus Secondary Task HUD Symbology (VISTA)

3.4.3. Cooper-Harper Profile

The purpose of the Cooper-Harper profile was to gather pilot opinion ratings from each pilot for each aircraft model. This profile, coupled with the Workload Buildup plus Secondary Task profile discussed above, was designed to explore the relationship between pilot opinion rating and Workload Buildup performance/workload. The profile lasted for a fixed period of two-minutes. There were no boundaries in this profile. Instead, two sets of reference lines were displayed on each side of the roll target, identified as the "desired" and "adequate" performance regions. The amount of time spent within each region was recorded and displayed to the pilot at the end of the profile, and the pilot could compare his or her performance to given performance requirements. The pilot then assigned a Cooper-Harper rating based on this performance and a subjective assessment of his or her workload. Figure 19 shows an example of the VISTA HUD symbology for the Cooper-Harper profile.

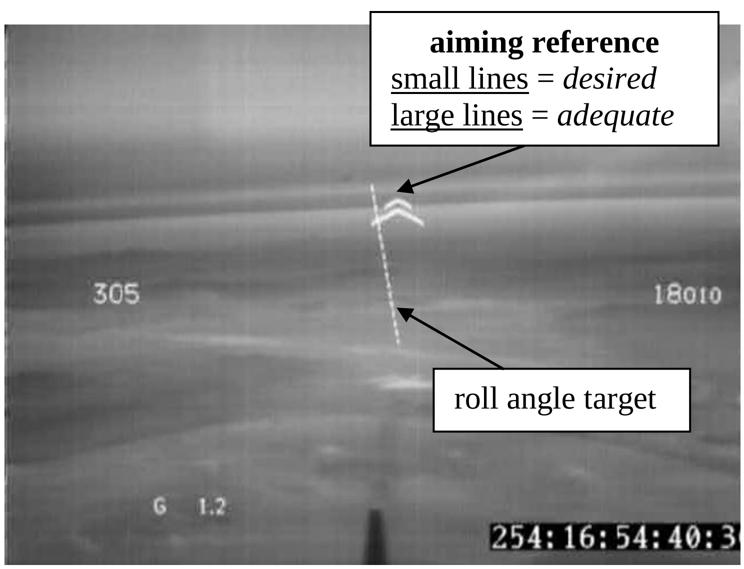


Figure 19. Cooper-Harper Profile HUD Symbology (VISTA)

3.5. Test Elements

In order to implement the test profiles described above, the following elements were needed:

- Aircraft models
- A tracking task
- A set of decreasing boundaries
- A secondary task

As testing progressed from fixed-base simulation to motion-base simulation to flight test, changes were made to these test elements as required to meet the challenges and differences of each new test platform. Changes were also made in order to further refine the flight test techniques and provide the best data. These changes were not detrimental to data analysis. No attempt was made, for example, to match simulator data to flight data, or to make the VISTA fly like the simulator. The tests were treated as separate data sets, all of which provided different examples of and insight into pilot behavior in the presence of boundaries.

3.5.1. Aircraft Models

In order to observe pilot behavior in the presence of boundaries in a variety of situations, and in order to search for a correlation between Workload Buildup performance/workload and pilot opinion rating, multiple models were needed to simulate aircraft with different handling qualities. No specific models were required, only that the models should be different from each other. Three models were developed for each test, with an attempt to design aircraft models with Level 1, Level 2, and Level 3 handling qualities. Information on handling qualities Levels can be found in the Department of Defense Military Standard Flying Qualities of Piloted Aircraft, MIL-STD-1797B (DOD, 2006).

Due to changes made between simulator tests and flight tests, each set of aircraft models was different than the others. This resulted in nine different aircraft models (three each for fixed-base simulation, motion-base simulation, and flight test). The models were labeled A through I.

3.5.1.1.Fixed-Base Simulator Aircraft Models

Since no previous BAT research had been done in the roll axis, much effort was spent in developing appropriate lateral-directional aircraft models which would be useful in support of the research objectives. Desktop simulations were used to evaluate potential models, and the most appropriate models were chosen. After considerable time and effort and many candidate aircraft models, the best models were shown to be those that restricted variables to as few as possible. Minimizing variables simplified analysis and allowed tighter control of handling qualities. Therefore, the fixed-base simulation models were developed to respond only in the roll axis, with neutral spiral stability and a completely canceled Dutch Roll mode. Thus, the aircraft response was reduced to the roll subsidence mode, and it was modeled as a simple transfer function between stick force and roll rate:

$$p = \frac{{}^{1}/_{\tau_r}}{s + {}^{1}/_{\tau_r}} F \qquad (2)$$

where

p = roll rate

F = stick force

τ_r= roll mode time constant

s = Laplace variable

This model choice reduced the model design variables to a single parameter. MIL-STD-1797B provides ranges of roll mode time constants which are predicted to give a range of handling quality levels. Based on these predictions, the following values were used to create three different aircraft models:

Table 2. Fixed-Base Simulation Model Parameters

	τ_r	Predicted Handling Qualities Level
Aircraft A	0.5	1
Aircraft B	4	2
Aircraft C	8	2

The models were further evaluated in the simulator prior to actual testing, and the time constant for Aircraft C was reduced from an original value of 10 (which is the predicted cutoff in MIL-STD-1797B between Levels 2 and 3), because the original value proved unflyable for the given task. In addition, the pitch and throttle responses were set to zero, simulating a perfect autopilot and autothrottle, in order to focus pilot attention only on the given tracking tasks.

3.5.1.2.Motion-Base Simulator Aircraft Models

The motion-base simulator tests were used to test and evaluate the proposed aircraft models for flight test. The model development was done by Calspan Corporation, the operators of the VISTA, based on the requirements of this research. These models were given additional degrees of freedom, in an attempt to model a more realistic airplane response than a first-order transfer function. Therefore, a fourth-order lateral-directional model was used. The roll mode time constant, spiral mode time constant, and Dutch Roll frequency and damping were the available design parameters.

As was done with the fixed-base models, MIL-STD-1797B was used to select flying qualities parameters which would provide a range of handling quality levels. Each of the parameters was chosen using the MIL-STD predictions to correspond to the desired handling qualities level (Level 1, 2, and 3). The aircraft model transfer functions are shown in the following equations:

Aircraft D: $$\frac{p}{\delta_a} = \frac{-29.15s^3 - 47.07s^2 - 242.5s + 0.4576}{s^4 + 3.81s^3 + 12.64s^2 + 18.13s + 0.181} \quad (3)$$

Aircraft E: $$\frac{p}{\delta_a} = \frac{-11.72s^3 - 2.406s^2 - 10.44s + 0.01966}{s^4 + 1.119s^3 + 1.193s^2 + 0.9209s + 0.009254} \quad (4)$$

Aircraft F: $$\frac{p}{\delta_a} = \frac{-6.152s^3 - 0.2133s^2 - 0.8743s + 0.001645}{s^4 + 0.526s^3 + 0.1732s^2 + 0.08168s + 0.0008033} \quad (5)$$

where: δ_a = aileron deflection (degrees)

The characteristic equation of the aircraft model transfer functions can be factored into the following form:

$$(s^2 + 2\zeta_{dr}\omega_{dr} + \omega_{dr}^2)\left(s + {}^{1}/_{\tau_r}\right)\left(s + {}^{1}/_{\tau_s}\right) = 0 \qquad (6)$$

where:

ζ_{dr} = Dutch roll damping ratio

ω_{dr} = Dutch roll natural frequency

τ_r = roll mode time constant

τ_s = spiral mode time constant

The four parameters in the factored characteristic equation were the design parameters, and they had the following values:

Table 3. Motion-Base Simulation Model Parameters

	ω_{dr} $(radians/second)$	ζ_{dr}	$\tau_r(seconds)$	$\tau_s(seconds)$	Predicted Handling Qualities Level
Aircraft D	3	0.3	0.5	99	1
Aircraft E	1	0.1	1.1	98	2
Aircraft F	0.4	0.02	2.5	100	3

Due to the continuously closed-loop nature of the roll tracking task in this project, the spiral mode was not considered a factor, and was not changed significantly among the models. A standard, previously verified longitudinal model (resulting in Level 1 handling qualities in the pitch axis) was used for all three models. The intent was to ensure that the longitudinal handling qualities were not a distraction during the tests.

During the simulator tests, none of the pilots complained about the longitudinal response, and so the longitudinal handling qualities were considered suitable for the test. The longitudinal transfer function is shown in Equation 7:

$$\frac{q}{\delta_e} = \frac{-0.1241s^3 - 12.88s^2 - 0.1884s - 0.04623}{s^4 + 7.48s^3 + 12.77s^2 + 0.1882s + 0.03586} \tag{7}$$

where:

q = pitch rate

δ_e = elevator deflection (degrees)

During the simulator tests, the achieved lateral-directional handling qualities of the models were not as predicted. The three models had noticeably different handling qualities, but were each (in their own way) approximately Level 2 aircraft. As a result, the models were modified for flight test.

3.5.1.3.Flight Test Aircraft Models

In order to provide a larger spread among the handling qualities of the three models, the flying qualities parameters were changed from those in the motion-base simulation. These models were analyzed using the predicted VISTA aerodynamics and were tested with hardware-in-the-loop simulations. The analysis and simulations suggested that the models would be acceptable for flight test. However, during the calibration sorties, these models proved to be uncontrollable. Due to inaccuracies in the VISTA aerodynamic model, the programmed transfer functions did not provide the predicted handling qualities. Therefore, in order to reduce the effects of aerodynamic uncertainty, the models were all programmed to have the same Dutch Roll characteristics (Level 1), and the roll mode time constant was varied to provide the different handling qualities (as in the fixed-base simulations). After multiple iterations, the following transfer functions were programmed in the VISTA:

Aircraft G
$$\frac{p}{\delta_a} = \frac{-41.02s^3 - 73.38s^2 - 403.7s + 0.7464}{s^4 + 5.157s^3 + 15.13s^2 + 30.49s + 0.9795} \quad (8)$$

Aircraft H
$$\frac{p}{\delta_a} = \frac{-13.05s^3 - 26.96s^2 - 122.1s + 0.2252}{s^4 + 2.722s^3 + 10.68s^2 + 8.424s + 0.1631} \quad (9)$$

Aircraft I
$$\frac{p}{\delta_a} = \frac{-10.31s^3 - 17.06s^2 - 95.53s + 0.1773}{s^4 + 2.213s^3 + 9.753s^2 + 3.804s - 0.00801} \quad (10)$$

These transfer functions corresponded to the following flying qualities parameters:

Table 4. Flight Test Model Parameters

	ω_{dr} $(radians/second)$	ζ_{dr}	$\tau_r(seconds)$	$\tau_s(seconds)$	Predicted Handling Qualities Level
Aircraft G	3	0.3	0.3	31	1
Aircraft H	3	0.3	1.1	50	2
Aircraft I	3	0.3	2.4	t_d = 380	3

where: t_d = time to double amplitude.

The spiral mode time constant was not intentionally changed, but it was allowed to change as necessary to appropriately model the roll mode in the VISTA. The longitudinal model from the LAMARS simulations was used in the VISTA, and the pilots confirmed that the longitudinal handling qualities were not objectionable.

3.5.2. Tracking Task

The tracking task for this research was designed as a random-appearing, moving roll angle target, displayed as a dotted line in the HUD. The task was designed to repeat every 30 seconds, to correspond with the boundary decrease interval. This allowed for evaluating performance criteria against boundary size without adding tracking task differences as a variable. In other words, if the task did not repeat regularly, there would be no way to tell if total tracking time was determined by boundary size or by a singularly difficult tracking segment which occurred at an inopportune time.

The tracking task used in the BAT DART tests was designed beginning with a sum of three sine waves, which was used as the pitch input for an F-16 model; the resulting flight-path angle of the F-16 was the displayed tracking task. Because of the arrangement, and limited number, of sine waves, as well as the damping supplied by the aircraft model, the tracking task became a series of fairly long climbs and descents, followed by oscillations at the top and bottom. As a result, the tracking task was fairly predictable and allowed the pilot to back out of the loop for long periods of time. Since the intent is to evaluate pilot-in-the-loop handling qualities, a more challenging task was needed.

Previous research has been conducted in building suitable random-appearing tracking tasks in human response studies, with recommendations given for the spread of the frequency content, the amplitude distribution, etc. (McRuer, 1965). However, the recommended spread of frequencies (2 decades) included frequencies that were too high to be used in this study, because they moved the boundaries faster than the aircraft could respond, resulting in an impossible task and almost immediate boundary excursion. As a result, a spread of only 1 decade was used. Some of the other recommendations were incorporated, such as including frequencies below 0.5 radian per second and using at least 5 sine waves.

In addition to the recommendations above, the tracking task was tailored to the aircraft models. Desktop simulations were conducted using a variety of tracking tasks, and the most appropriate task was chosen after many hours of comparison testing. Tracking task refinement also continued throughout each test phase in order to produce the most suitable task. Simulation showed that some of the proposed tracking tasks resulted in almost identical performance among all three aircraft models, while others were extremely difficult even for the best model. To understand the difference between the tasks, the root-mean-square (rms) roll rate and acceleration of the candidate tasks were calculated, and a balance was found between values that were too low (resulting in no difference between aircraft model performance) or too high (resulting in poor performance for even the best model). Figure 20 compares the final tracking task with the task used for the BAT DART tests. The rms values show that the task designed for this research was, in fact, more challenging than the BAT DART task, as desired.

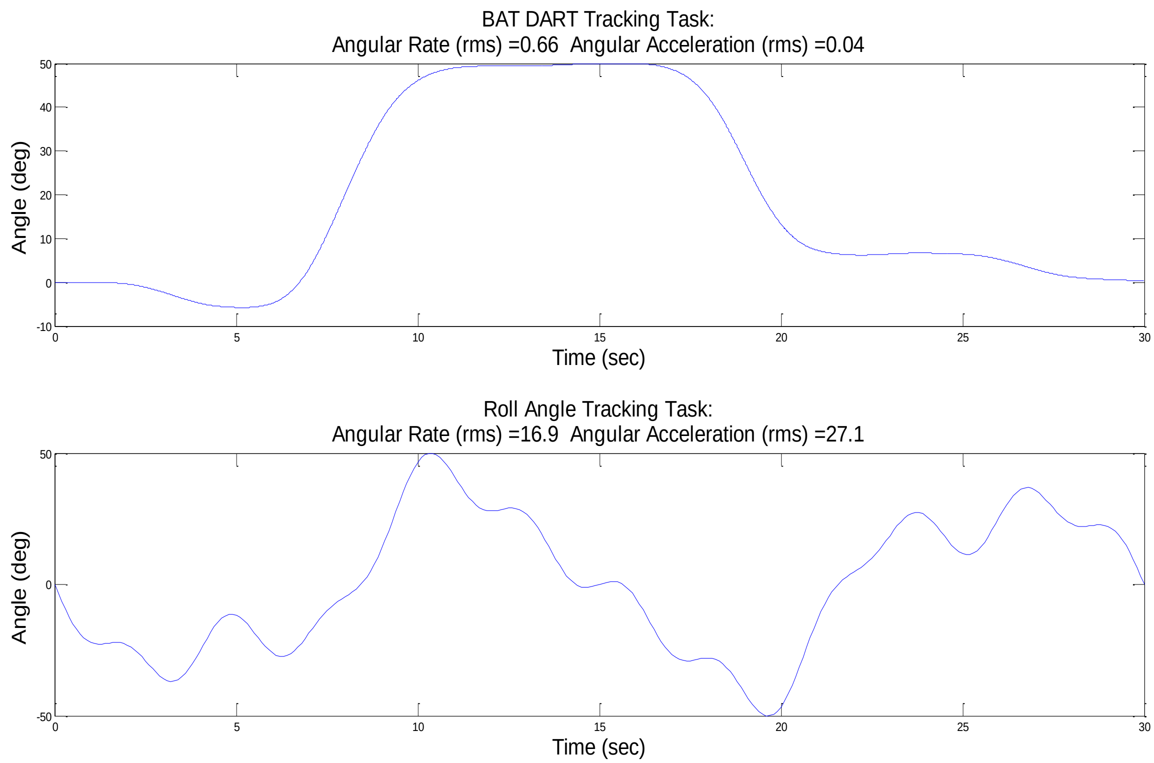


Figure 20. Tracking Task Comparison

Table 5 lists the frequency and amplitude content of the tracking task. The listed amplitudes were scaled by 33.57 to give a maximum bank angle of 50º.

Table 5. Tracking Task Frequency Content

Cycles per 30 seconds	Frequency (rad/sec)	Amplitude / 33.57
2	.4188	-1
3	.62832	0.1
5	1.0472	-0.3
7	1.4661	0.1
11	2.3038	-0.2

As with the aircraft models, changes were required to the tracking task as the testing progressed to motion-base simulation and flight test. In the LAMARS simulator, the sine wave amplitudes proved too high. This moved the roll target faster than was suitable for the aircraft models. As a result, all of the amplitudes were scaled by a factor of 0.67. Furthermore, even this scaled-down task proved too difficult during flight test. The final tracking task used in the VISTA included only the first three sine waves from Table 5.

Throughout all three series of tests, pilots consistently commented that the tracking task was challenging and unpredictable, and so the tracking tasks were considered suitable. This tracking task has already proved useful in other flight tests as well, and has been incorporated into handling qualities simulations used in the TPS syllabus.

3.5.3. Boundaries

The boundaries were displayed as solid lines to the left and right of the bank angle target which moved with the target, thus providing bounds on the error. In the fixed-base simulator, the boundaries started ± 30º from the target, and steadily moved closer as the test progressed, with a total decrease of 20% every 30 seconds. Data analysis from these fixed-base simulator tests suggested that the boundaries started too close to the target, and so the boundaries for the motion-base simulation were started ± 40º from the target. Additionally, the boundaries for all further tests were programmed to shrink step-wise every 30 seconds instead of gradually fading in, as in the fixed-base simulation. This was done in order to correspond with the tracking task repetition every 30 seconds, and to provide a constant difficulty level (due to boundaries) for each 30 second period.

Due to difficulties in programming the VISTA displays, the boundaries were implemented differently in the aircraft than in the motion-base simulation. For flight test, the boundaries began at ± 30º from the target (as in the fixed-base tests) and shrank by approximately 8º every 30 seconds (instead of the 20% decrease used in the motion-base tests).

3.5.4. Secondary Task

Since this was the first time that secondary tasks had been used in BAT research, it was not clear at the start how the task should be designed. So, in the first set of tests (the fixed-base simulations), two types of secondary tasks were explored. One of them consisted of the arrows described in the profile descriptions above, displayed on the HDD. The other task consisted of a small square displayed at random-appearing times in the corner of the HUD airspeed display. When the square was displayed, the pilot was required to press and hold a button on the throttle. When the square was removed, the pilot was required to release the button.

Analysis of the data from these tests showed that the square task was not challenging enough, and performance of this task remained relatively constant throughout the profiles. This made the task unsuitable as a measure of pilot workload. The arrow task was challenging enough, and performance of this task degraded appropriately for a measure of workload. However, the crosscheck required from the HUD down to the HDD caused several pilots to exceed the boundaries as they were looking down at the secondary task. This was an undesired effect. Figure 21 compares the two secondary tasks. After the fixed-base simulations, the arrow task was chosen, but was moved to the lower left corner of the HUD, in order to relieve the time required for crosscheck, and prevent exceeding boundaries due to secondary task crosscheck. Both the motion-base simulator tests and the flight tests used the new HUD arrow task as the secondary task.

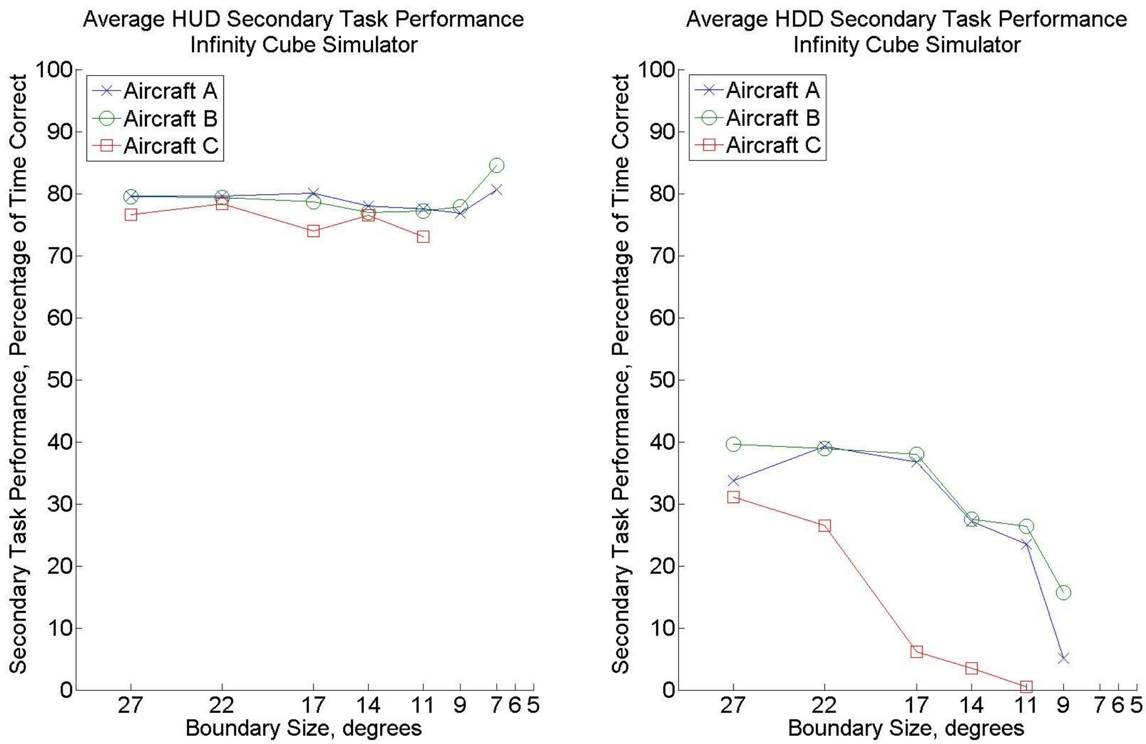


Figure 21. Comparison of Secondary Task Candidates

The secondary task was used as a measure of pilot workload, under the assumption that as boundaries decreased, pilots would spend more time focusing on the roll tracking task and less time focusing on the secondary task. Thus it was assumed that degraded secondary task performance meant that pilots were looking at the secondary task less often. In order to confirm this assumption, and to validate the secondary task as an appropriate measure of pilot workload, eye-tracking equipment was used during the motion-base simulator tests. This equipment was not cleared for flight test, so it was only used in the simulator.

The eye-tracker system recorded the focal point of the pilot's eye when performing the Workload Buildup plus Secondary Task profile. The data included a series of 24 frames-per-second, MPEG-4 digital video in 752x480 pixel resolution. The eye-tracker used a Charge-Coupled Device (CCD) to track the center of the pilot's right pupil and another CCD to record video in the direction the pilot's head was pointed. The equipment then superimposed a red cross (representing the center of the pupil) on the front view video data. Eye tracker performance was limited by its lens field-of-view (roughly ±15 degrees), CCD sensitivity, and bore-sight alignment. Additionally, pilot blinking would often cause the eye tracker to temporarily lose track. Even with these limitations, the team still gathered representative data to validate the secondary task as an appropriate measure of pilot workload.

Figure 22 shows a typical example of a pilot's secondary task crosscheck for each aircraft model. A crosscheck was defined as a movement of the pilot's eye from the primary task to the secondary task. The crosscheck frequency (measured in Hertz) was averaged for each 30 second boundary interval. Decreasing crosscheck meant that the pilot was looking at the secondary task less often. As expected, the crosscheck ability of the pilot decreased as the boundaries decreased, and also decreased in proportion to decreasing handling qualities.

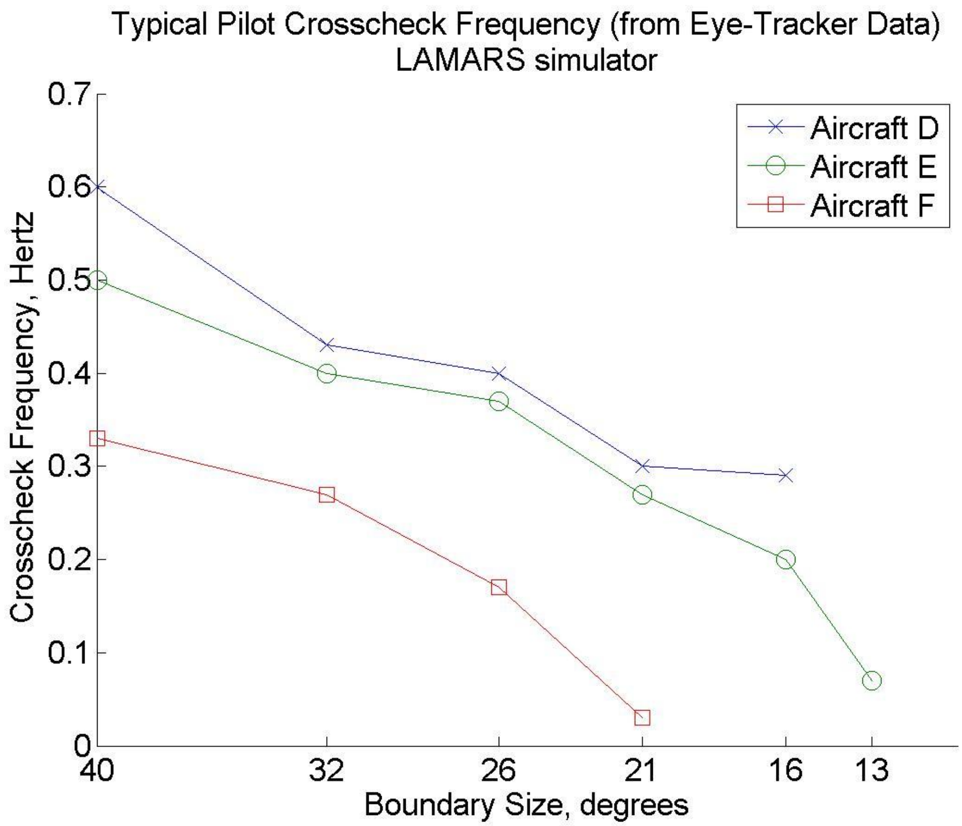


Figure 22. Typical Pilot Crosscheck Frequency (LAMARS)

Figure 23 shows a typical example of a pilot's secondary task crosscheck compared to his secondary task performance on a single Workload Buildup plus Secondary Task profile. The trends of the two sets of data followed each other closely, which validated the secondary task as an appropriate measure of pilot workload. This was an important result, and a significant contribution to BAT test methodology. The secondary task developed for this research provided important workload data during these tests, and can be used in BAT research in the future in order to measure pilot workload.

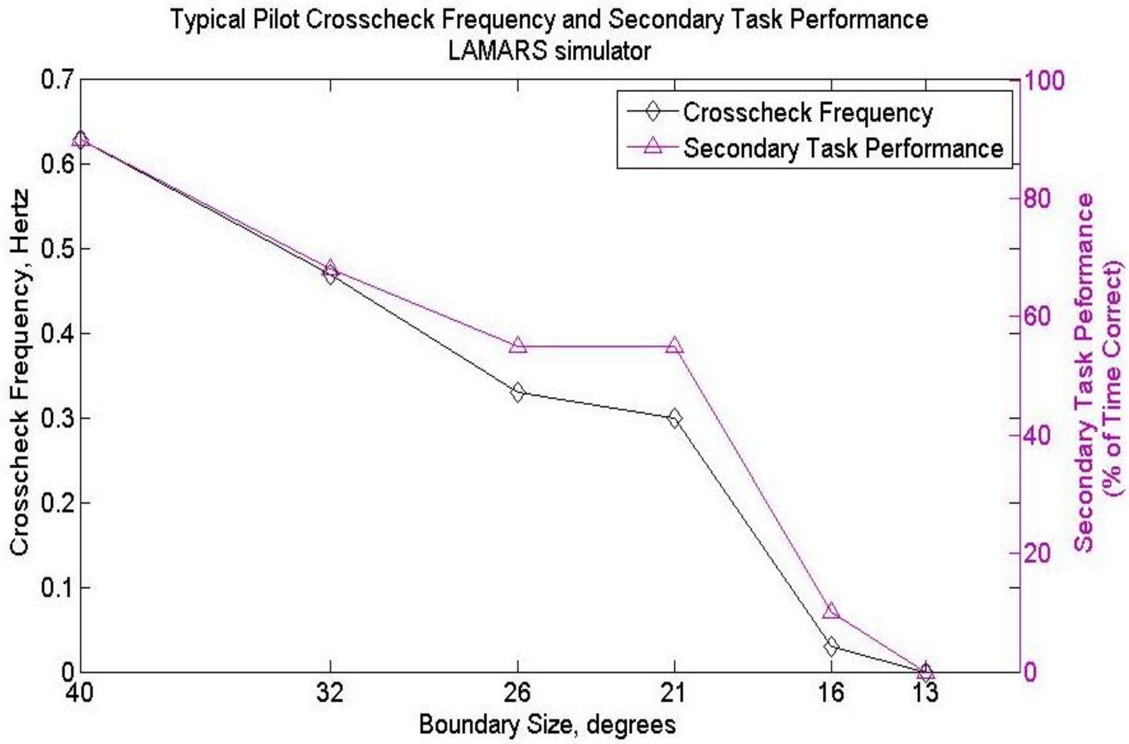


Figure 23. Typical Pilot Crosscheck Frequency and Secondary Task Performance (LAMARS)

3.6. Test Procedure

Combining all of the profiles listed above resulted in a total of nine tests per pilot (three profiles for each of the three aircraft models), plus three practice tests to ensure the pilots understood the testing procedures. Common elements to all tests included:

- Roll-angle tracking, with the same task in all tests
- A 15 second "warm-up", during which the pilot could fly the aircraft, but the task remained stationary and the boundaries were non-functional.

The test cards used for the VISTA flight tests are include in Appendix B.

3.7. Test Briefing and Order

The testing procedures were explained fully to each pilot before any of the tests were conducted. Pilots were told that they would be evaluating different types of aircraft, but no specifics were given. Testing order was randomized for each pilot, in order to reduce the effects of learning or fatigue.

3.8. Test Subjects

Twenty-two subjects participated in the fixed-base simulator tests. All subjects were rated military pilots; the least experienced pilot had 800 hours of military flight time. Over 60% of the pilots were fighter pilots, 10% were bomber pilots, and around 15% each were transport or helicopter pilots.

Fourteen subjects participated in the motion-base simulator tests, including the five members of the TPS test team. Four of the subjects had participated in the previous fixed-base simulator tests. Eleven of the subjects were rated military pilots, one subject was a military navigator, and two of the subjects (from the TPS test team) were engineers. Of the pilots, 46% were fighter pilots, and 27% each were transport or helicopter pilots. The least experienced pilot had 750 hours of military flight time.

The five TPS test team members were the only participants in the flight tests, but each member flew two sorties. All three of the test team pilots were fighter pilots; the least experienced pilot had over 1500 hours of military flight time.

Thus, in all there were 32 different test subjects and 46 test sets. Each set included 9 profiles, for a total of 414 test events. Over 60 total hours of ground simulation and flight time were accomplished.

3.9. Summary of Experiment Design

Because there were no techniques for BAT research in the roll axis prior to this research, considerable time and effort were spent in developing appropriate test techniques to support the research objectives. After research and preliminary simulation, three profiles were developed which would collectively address all four objectives. In addition to the structure of the profiles, the individual elements of the profiles were researched, simulated, and developed to gather the most pertinent data.

All three profiles were based on a roll angle tracking task that was developed to eliminate unwanted boundaries that were found in previous tests, and to keep the pilot in the loop at all times, in contrast to tracking tasks used in previous research. Three lateral-directional aircraft models were created, after much iteration, in order to observe pilot behavior in the presence of boundaries in a variety of situations, and in order to search for a correlation between Workload Buildup performance/workload and pilot opinion rating.

A secondary task was also developed for one of the profiles in order to explicitly measure pilot workload. This type of workload measurement in the presence of boundaries had not been attempted before, and so many hours of research and simulation were conducted in order to design the most effective secondary task. After the task was designed, it was validated as a measure of pilot workload through the use of eye-tracker equipment. The eye-tracker data and the secondary task data showed remarkable correlation.

The development of the flight test techniques and test elements present a significant contribution to BAT research methodology. These test techniques were useful not only for this research, but can be used in future BAT research as well.

4. Test Results

As described in Chapter 1, the following were the four main objectives for this thesis:

- Characterize the mathematical BAT model through comparison with simulator and flight test data
- Characterize pilot behavior (in terms of control stick forces and pilot physical workload) in the presence of decreasing boundaries
- Characterize pilot performance (in terms of tracking error) in the presence of decreasing boundaries
- Characterize the relationship between pilot opinion rating and performance/ workload in a Workload Buildup task

The test data were analyzed in order to support these objectives. There were over 400 test profiles available for analysis, and the amount of data provided a solid base for validating assumptions and suggesting new conclusions.

4.1. Characterizing the BAT Model

A total of 123 Workload Buildup profiles (one for each pilot-aircraft pair) were analyzed and compared to the BAT model predictions. Previously, all attempts to fit the BAT model to experimental data had used only the boundary tracking loop of the model. The switching function of the model had not been confirmed. Therefore, the data gathered in this research were used to fully characterize the entire BAT model.

4.1.1. Model Matching Procedure

The basic steps of the model matching procedure were as follows:

1. Search the data in each profile for an area of suspected boundary tracking.
2. Vary the four BAT parameters so that the predictions of the boundary tracking loop matched the test data for the suspected boundary tracking area.
3. Keep the four BAT parameters fixed at their values from Step 2, and vary parameters from the point tracking loop so that the predictions of the whole model match the data from the entire profile.

These steps are explained fully in the following sections.

4.1.2. Numerical Optimization Procedure

Previous BAT model matching had been done using an iterative process in which the parameters were changed, and the subsequent results were qualitatively evaluated. For this research, the BAT model was matched to test data using numerical optimization techniques. The *goodness* of the model was the extent to which the predicted control forces matched the actual stick forces from the test data. (All of the simulator and flight tests used a fixed side-stick, and so the control stick inputs were in terms of force rather than displacement.) Therefore, the error between the model predictions and the test data was minimized.

For the boundary tracking loop optimization, the predicted stick force from the model was calculated using the actual time-to-boundary from the test data. The predicted stick force was a function of t_b, which was fixed by the test data, and of the four BAT model parameters (t_{min}, t_{max}, K_{bm}, and τ_b), which were varied during the optimization.

These BAT model parameters were the design variables. A time-history of the error between the actual stick force and the predicted stick force was created, as shown in the following equations:

$$\boldsymbol{e}(\boldsymbol{t}, \boldsymbol{t_b}, \boldsymbol{x}_{BAT}) = \boldsymbol{F_{actual}}(\boldsymbol{t}) - \boldsymbol{F_{predicted}}(\boldsymbol{t_b}, \boldsymbol{x}_{BAT}) \tag{11}$$

$$\boldsymbol{x}_{BAT} = [t_{min} \quad t_{max} \quad K_{bm} \quad \tau_b] \tag{12}$$

where

$\boldsymbol{e}$	= model prediction error vector
$\boldsymbol{t}$	= time vector
$\boldsymbol{F_{actual}}$	= actual stick force vector from test data
$\boldsymbol{F_{predicted}}$	= predicted stick force vector from BAT model
$\boldsymbol{x}_{BAT}$	= vector of design variables

The error vector was then multiplied with its transpose in order to create a scalar cost function in terms of the BAT model parameters, as shown below:

$$c(\boldsymbol{x}_{BAT}) = \boldsymbol{e}(\boldsymbol{t}, \boldsymbol{t_b}, \boldsymbol{x}_{BAT})^T \boldsymbol{e}(\boldsymbol{t}, \boldsymbol{t_b}, \boldsymbol{x}_{BAT}) \tag{13}$$

The model parameters were all required to be positive, since pilots were assumed to always begin reacting to boundaries before the boundaries were exceeded. In addition, t_{min} was required to be greater than t_{max}, since minimum input occurs at a greater time to boundary than maximum input. With these constraints, the optimization problem for the boundary tracking model was formulated as follows:

$$minimize \quad c(\boldsymbol{x}_{BAT}) = \boldsymbol{e}(\boldsymbol{t}, \boldsymbol{t_b}, \boldsymbol{x}_{BAT})^T \boldsymbol{e}(\boldsymbol{t}, \boldsymbol{t_b}, \boldsymbol{x}_{BAT}) \tag{14}$$

subject to

$$\boldsymbol{x}_{BAT} > 0 \tag{15}$$

$$t_{min} > t_{max} \tag{16}$$

The optimization for the point tracking model was formulated in a similar way. The BAT model does not specify what point tracking model to use. A proportional-plus-derivative control model was chosen because it could be easily used in the time domain. (Previous research showed that operator behavior can be modeled using proportional-plus-derivative control (Arif and Inooka, 1999)). The point tracking stick force was calculated according to the following equation:

$$F_{PT} = K_p e_{pt} + K_d \dot{e}_{pt} \tag{17}$$

where:

F_{PT} = point tracking stick force

K_p = proportional control gain

e_{pt} = point tracking roll error

K_d = derivative control gain

$\dot{e}_{pt}$ = point tracking roll error rate

For the point tracking model optimization, the predicted stick force from the model was calculated using the actual tracking error and error rate from the test data. The predicted stick force was a function of e_{pt} and $\dot{e}_{pt}$, which were fixed by the test data, and of the two point tracking model parameters (K_p and K_d), which were varied during the optimization. These point tracking model parameters were the design variables. The point tracking error equations are as follows:

$$\boldsymbol{e}\left(\boldsymbol{t}, \boldsymbol{e_{pt}}, \dot{\boldsymbol{e}}_{\boldsymbol{pt}}, \boldsymbol{x}_{PT}\right) = \boldsymbol{F_{actual}}(\boldsymbol{t}) - \boldsymbol{F_{predicted}}\left(\boldsymbol{e_{pt}}, \dot{\boldsymbol{e}}_{\boldsymbol{pt}}, \boldsymbol{x}_{PT}\right) \tag{18}$$

$$\boldsymbol{x}_{PT} = [K_p \quad K_d] \tag{19}$$

The point tracking parameters were also required to be positive. The optimization problem for the point tracking model was therefore formulated as follows:

$$minimize \quad c(\boldsymbol{x}_{PT}) = \boldsymbol{e}\left(\boldsymbol{t}, \boldsymbol{e_{pt}}, \dot{\boldsymbol{e}}_{\boldsymbol{pt}}, \boldsymbol{x}_{PT}\right)^T \boldsymbol{e}\left(\boldsymbol{t}, \boldsymbol{e_{pt}}, \dot{\boldsymbol{e}}_{\boldsymbol{pt}}, \boldsymbol{x}_{BAT}\right) \tag{20}$$

subject to

$$\boldsymbol{x}_{PT} > 0 \tag{21}$$

During the optimization, the cost function was minimized using a line-search method (the *fmincon* routine in MATLAB®). In each case, the results were also evaluated qualitatively to ensure that the match was reasonable.

4.1.3. Modification to the BAT Model

During attempts to achieve the best fit to the boundary tracking data, it appeared that pilot boundary tracking inputs did not increase linearly from t_{min} to t_{max} (as was theorized in Equation 1). Instead, the rate of change seemed to increase as the time to boundary approached t_{max}. In an attempt to model this behavior, the BAT model was modified. In the new model, the stick force increased quadratically from t_{min} to t_{max}. The new stick force equation is as follows:

$$\left(\frac{t_{min} - (t_b + \tau_b)}{t_{min} - t_{max}}\right)^2 K_{bm} \tag{22}$$

Optimizations were performed with the new and old model, and over 80% of the time the minimized error was smaller when using the new model, by an average of 2%. Thus, the new model was used for all data analysis.

4.1.4. Model Matching Example

As in the desktop simulator studies performed by Mr. Gray, the Workload Buildup profiles were searched for areas of boundary tracking (based on pilot comments, obvious areas of boundary-induced PIO, or areas with very small time to boundary). The boundary tracking loop of the BAT model was then fit to the test data for these limited time periods. As in Mr. Gray's research, it was assumed that all inputs during these periods were due solely to boundary tracking. The reason for starting with these areas was to be sure that the BAT parameters were only being used to describe boundary tracking. Attempting to fit BAT parameters to areas of point tracking would produce erroneous results. Once the BAT parameters were fixed, the point tracking parameters could be fit to the rest of the data.

Figure 24 shows data from a typical Workload Buildup Task. The circled area on the middle plot (a plot of the tracking error) shows a time period where the boundaries were closely approached and in fact exceeded (with the 0.5 second tolerance). Coincident with this was a large change in stick force, and the shape of the stick force during this period is a classic sign of PIO. Since the PIO occurred simultaneously with the boundary excursion, it was assumed that this was a period of boundary tracking. A numerical optimization *of only the boundary tracking model* was then performed over this limited time period.

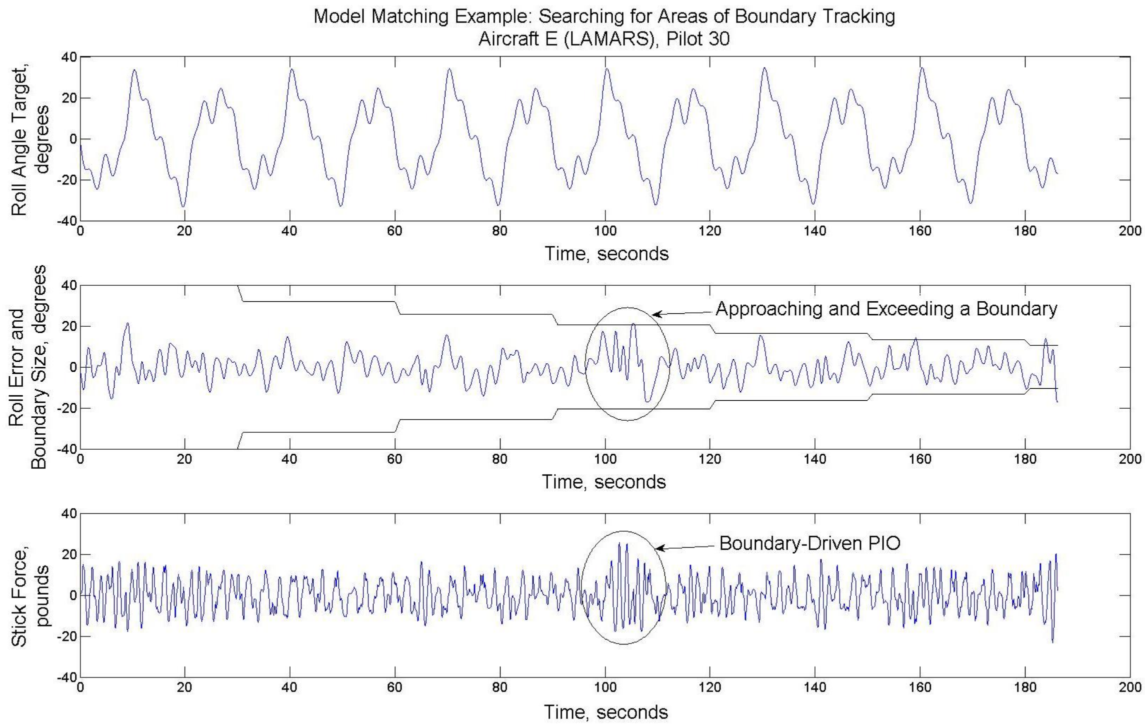


Figure 24. Identifying Areas of Boundary Tracking

The results of the boundary tracking model optimization for this example are shown in Figure 25. While the predictions do not match the actual values exactly, they do approximate the general shape, height, and location in time. Importantly, the parameters have been *optimized*, so that this represents the best prediction that the BAT model can give. Once the BAT parameters were optimized over a period of known boundary tracking, they were fixed for the rest of the model matching. Thus, a time history of predicted boundary tracking inputs could be created for the entire profile. The next step was to calculate predicted point tracking inputs.

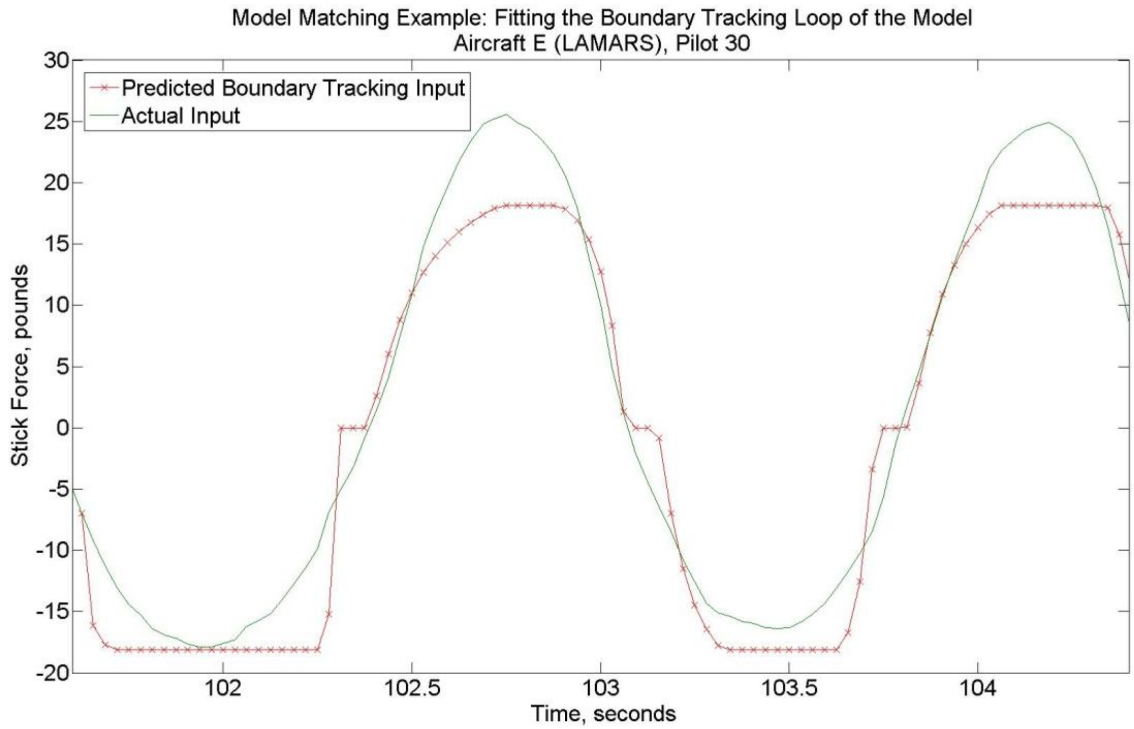


Figure 25. Fitting the Boundary Tracking Loop of the BAT Model

With the BAT parameters fixed, the point tracking optimization was performed over the whole time history, using the full switching model. At each step of the optimization, the point tracking pilot input was calculated and compared to the boundary tracking input (which was fixed from the previous optimization). For each time step, the larger input was chosen, as prescribed by the BAT model. Thus, for each step of the optimization, a time history of predicted input was created which was a combination of point tracking and boundary tracking.

Figure 26 shows the optimization results for the current example. Only a portion of the profile is shown for the sake of clarity. The predicted stick force from the model is divided into periods of point tracking and periods of boundary tracking. Notice that the hypothetical inputs switch between the two control strategies according to the model. Once again, the predictions do not exactly match the actual values, but the general shape and size are correct.

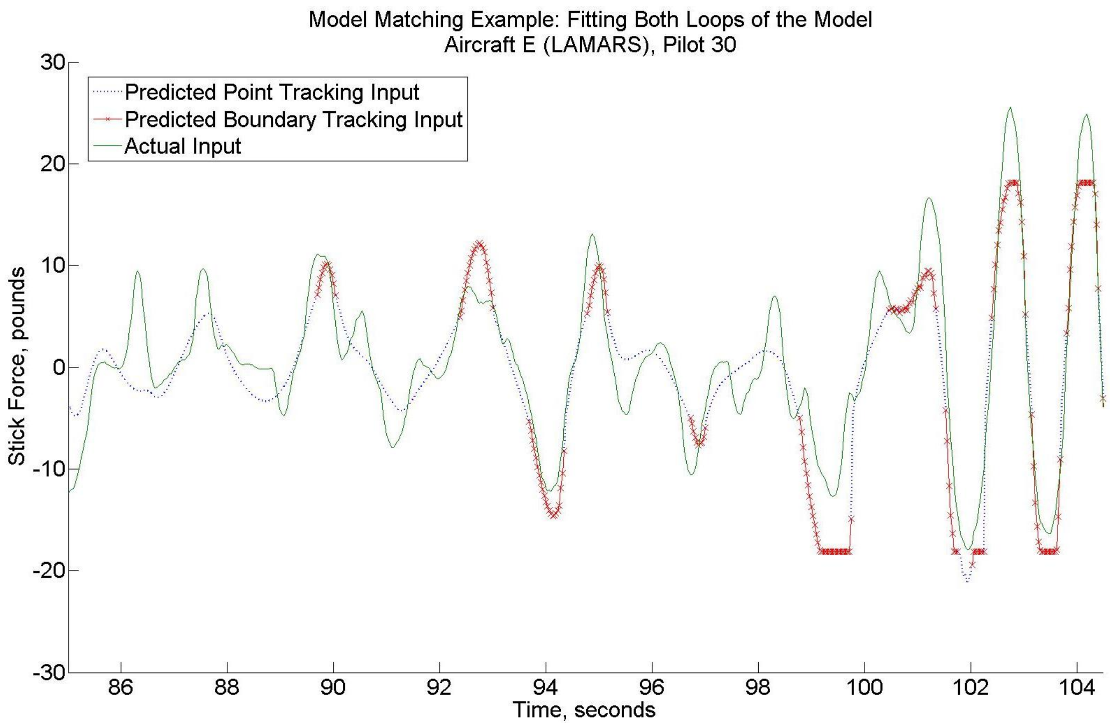


Figure 26. Fitting Both Loops of the BAT Model

4.1.5. Model Matching Results

The optimization procedure was followed for all 123 pilot-aircraft combinations. The results are summarized on the next two pages.

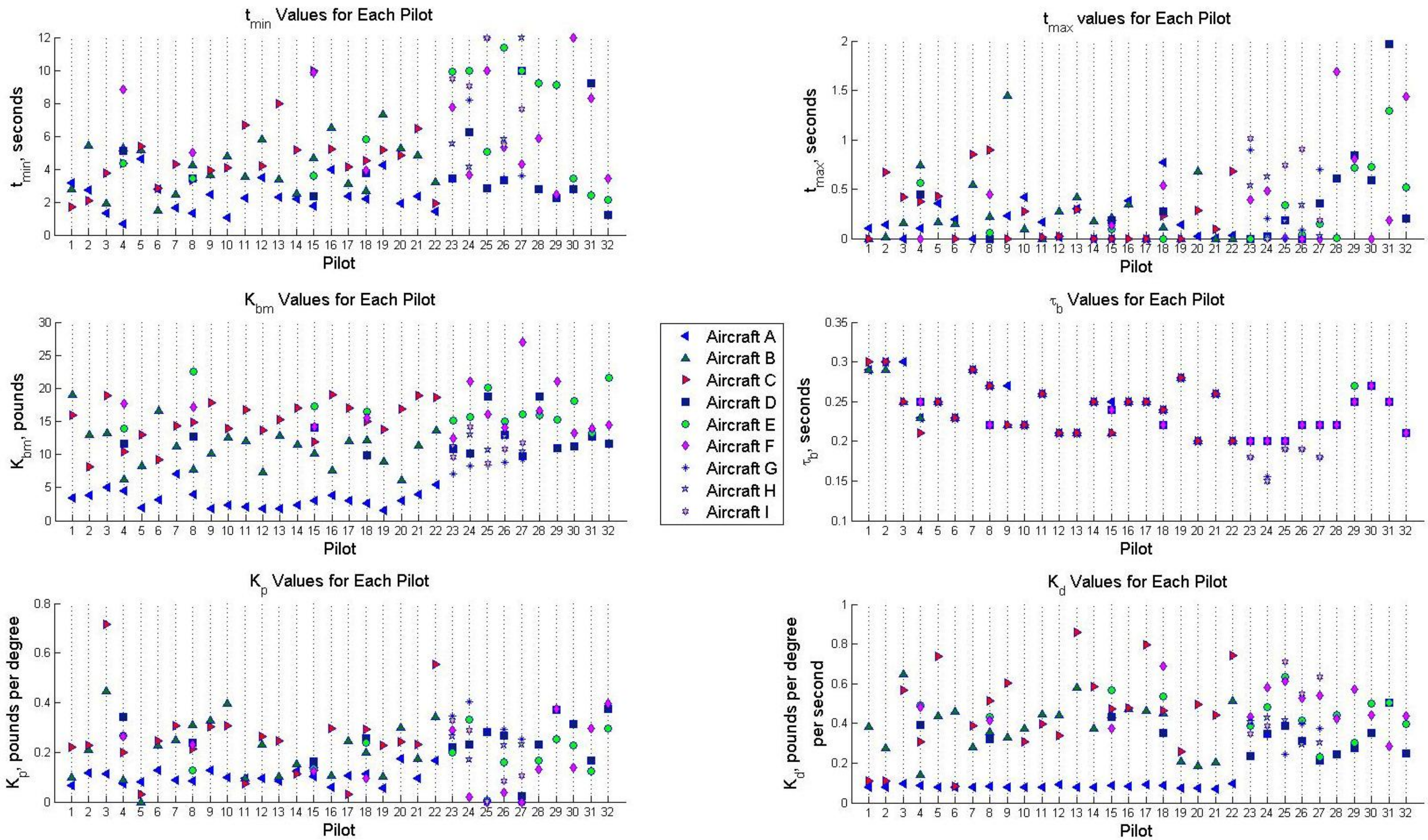


Figure 27. BAT Model Parameters Sorted by Pilot

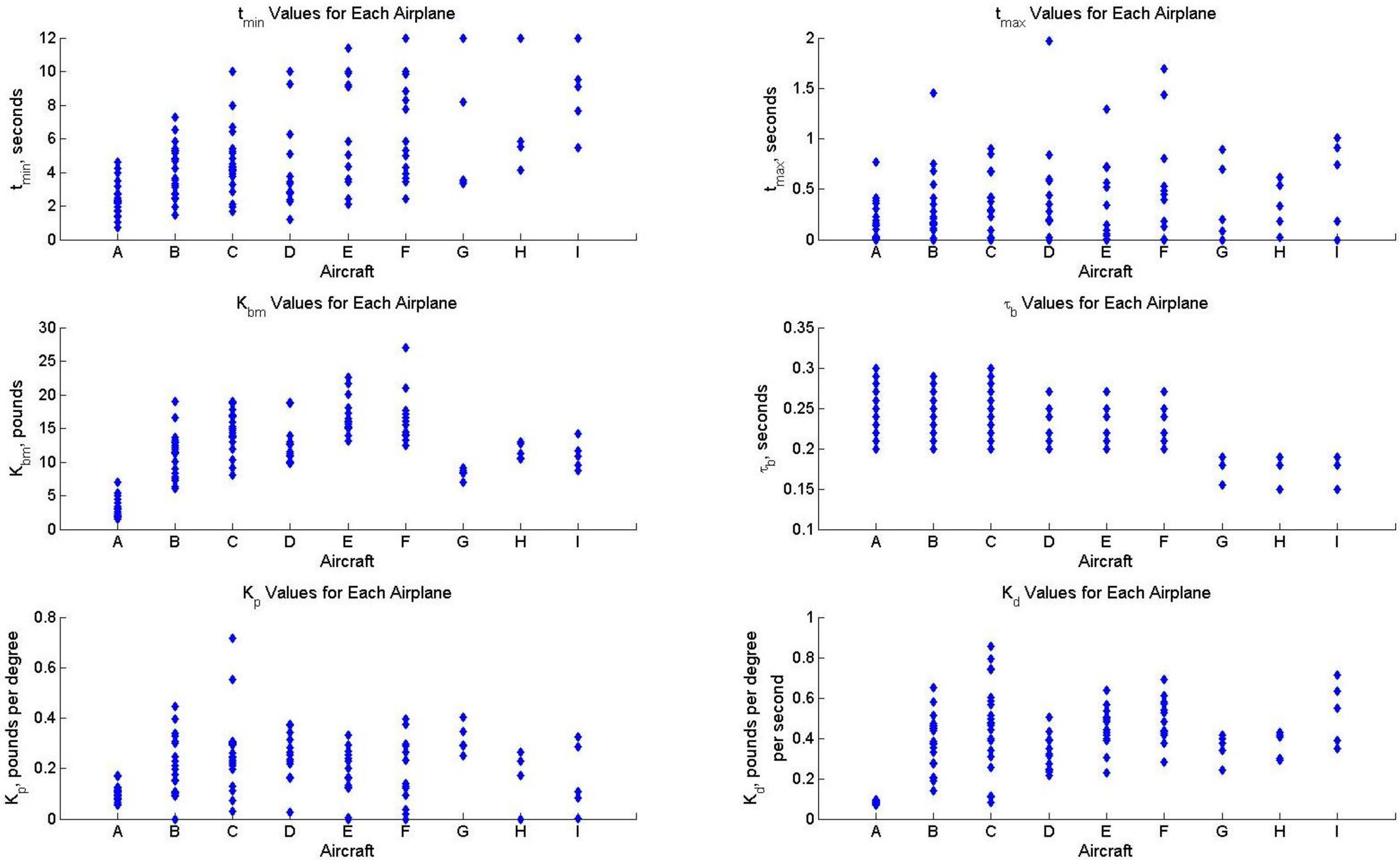


Figure 28. BAT Model Parameters Sorted by Aircraft Model

The first conclusion that can be drawn from these results is that the BAT parameters vary among pilots, as shown in Figure 27. For a given aircraft, each pilot has a unique set of BAT parameters, representing that pilot's control strategy. The HAVE BAT flight tests and the desktop simulations by Gray focused on the parameter variations among pilots, and the data from this research provide confirmation of the conclusions from these previous tests.

The second conclusion is that there are trends in the BAT parameter variation among aircraft, as shown in Figure 28. For example, there appears to be a general upward trend in t_{min} values among aircraft models of the same platforms (i.e. from A to C in the Infinity Cube simulator, from D to F in the LAMARS simulator, and from G to I in the VISTA aircraft). This makes intuitive sense, because pilots adapt their control strategies to fit the airplane they are controlling. It makes sense that a pilot would begin reacting to boundaries sooner in an aircraft that he or she knows is sluggish or unpredictable. These trends of BAT parameter variations among aircraft had not been demonstrated in previous research.

In addition to characterizing pilots, the parameters therefore also characterize airplanes. The average K_{bm}, for example, was much less in Aircraft A than in Aircraft C. Just as pilots compare a Level 1 and a Level 2 airplane, they can now also compare a "K_{bm} 4" and a "K_{bm} 15" airplane. This is an important result, because it demonstrates that the BAT parameters describe the change in pilot control strategies among aircraft in much the same was as frequency, damping, and time constants describe flying qualities.

The fourth objective of this research (discussed in a later section) attempts to quantify handling qualities, supplementing *subjective* pilot comments with *objective* performance measurements. The trend in parameter variation among aircraft shown here provides an additional method to quantify aircraft handling qualities, by describing a given aircraft according to the average BAT parameters used by pilots to control that aircraft.

4.1.6. Summary of BAT Model Characterization Results

The BAT model was matched to the test data from 123 different profiles through the use of numerical optimization. BAT model matching had never before been done on this scale, and had never used optimization. Optimization not only ensured the best fit to the test data, but also suggested a change to the boundary tracking stick force equation which resulted in better matches. In addition, whereas previous research had only exercised portions of the mathematical BAT model, this research compared flight test data to the predictions of both the boundary tracking loop and the point tracking loop of the BAT model.

The results validate the conclusion from previous research that BAT parameters vary among pilots. The results also demonstrated for the first time a trend of changing parameters among aircraft. Since the parameters are based on pilot control strategies (i.e. they are based on the pilot-aircraft system), they can therefore be used to provide an objective measure to aircraft handling qualities.

4.2. Characterizing Pilot Behavior

A total of 276 profiles with boundaries were performed during this research. Analysis of these profiles confirms that pilots do, in fact, react to boundaries. This is intuitive, and is a confirmation of the results from all previous BAT research, but it is the first time that it has been demonstrated in the roll axis. The PIOs and boundary tracking inputs analyzed in the previous section are one way to describe exactly *how* pilots react to boundaries.

Another way to describe how pilots react to boundaries is to look at how their behavior changes, on average, as boundaries are tightened. The tracking task in this research was designed to repeat every 30 seconds, coincident with the boundary changes. This ensured that each 30 second period had the same average tracking difficulty, and the only changing parameter was the boundary size. Thus, any change in a pilot's average behavior from one period to the next should be due solely to changes in boundary size. (The fixed-base simulations used constantly decreasing boundaries, but this procedure was changed for the motion-base simulator and flight tests. The fixed-base simulator results were analyzed using the average boundary size during each 30 second period.) One behavior parameter that can be analyzed is a pilot's average stick force. In over 80% of the Workload Buildup profiles analyzed, the average stick force during the smallest achieved boundary size was higher than the average stick force at the largest boundary size. Figure 29 shows an example of this type of increase. Once again, nothing changed from period to period except the boundaries. Therefore, this increase can be logically attributed to the changing boundaries.

This monotonically increasing average stick force is not typical of all pilots, but it did occur fairly often. When the data are averaged among pilots, the results are less dramatic, but they still show a slight increase in average stick force as boundaries decrease. Figure 30 is one example, in which the results were averaged for those pilots whose final boundary size during the Workload Buildup task was 11 degrees. Averaging only among pilots with the same terminal boundary size prevented skewing the data, because pilots with different terminal boundary sizes likely have different control strategies and baseline average stick forces.

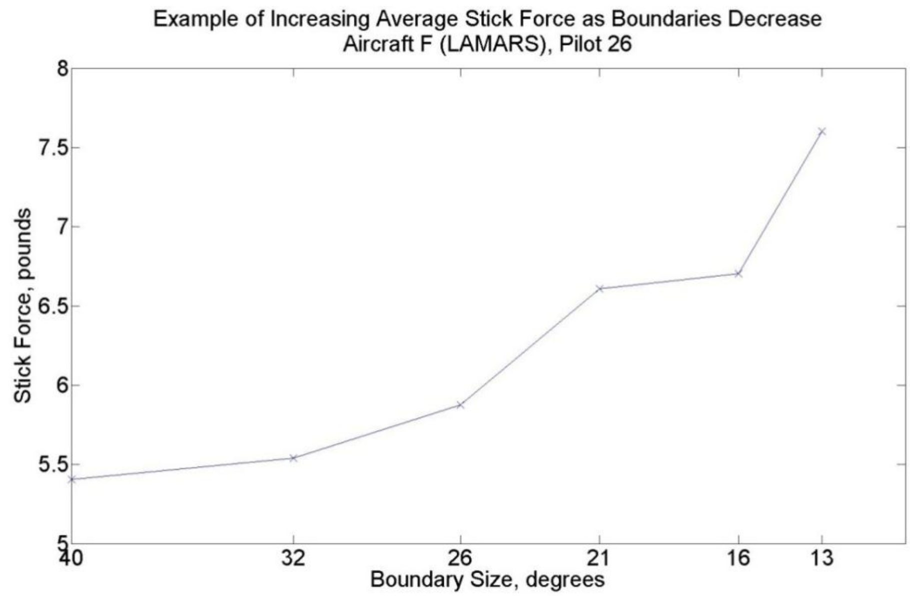


Figure 29. Example of Increasing Stick Force as Boundaries Decrease

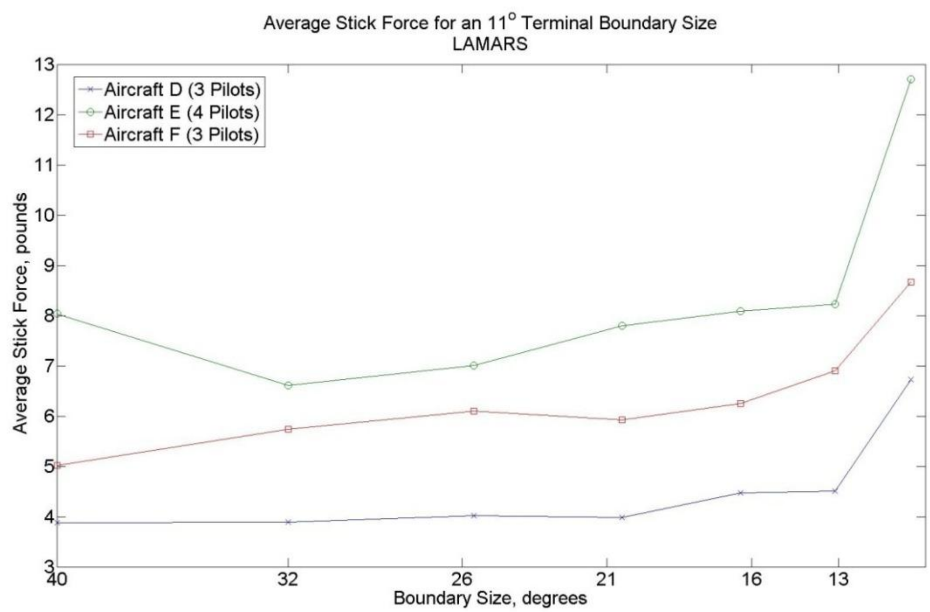


Figure 30. Average Stick Force for an 11º Terminal Boundary Size (LAMARS)

Another way to analyze changing pilot behavior is with the concept of duty cycle and aggressiveness, described in Chapter 2. As a reminder, duty cycle describes how often the pilot is making control stick inputs. Aggressiveness describes how fast these inputs are made. Together, these parameters describe pilot inceptor workload. Increasing duty cycle and aggressiveness corresponds to higher physical workload.

Figure 31 shows the results of a duty cycle and aggressiveness analysis on the VISTA flight test data. (Results were similar for the ground simulator tests.) As with the average stick force, the results were averaged among pilots with the same terminal boundary size. The data show a general increase in both duty cycle and aggressiveness as boundaries tighten, with the noticeable exception of the 23 degree terminal boundary

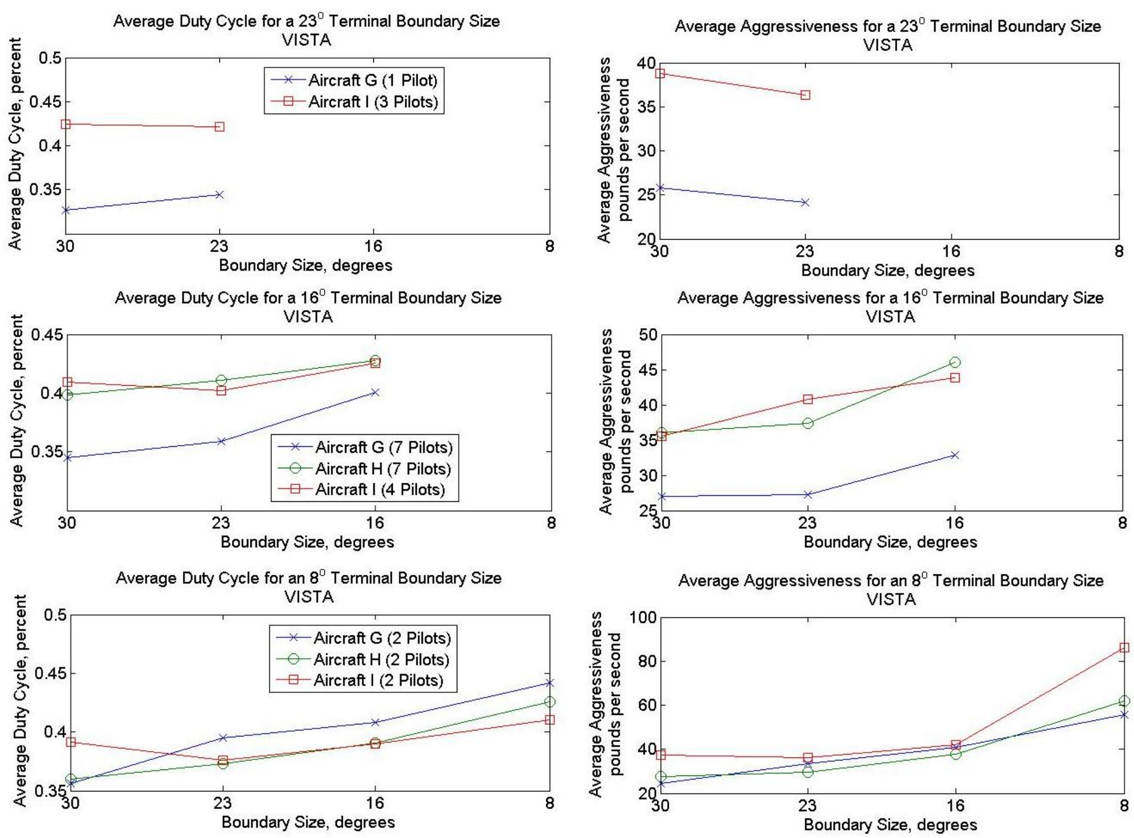


Figure 31. Average Duty Cycle and Aggressiveness (VISTA)

size. A possible explanation for this is that perhaps these pilots exceeded the boundaries at such a large terminal boundary size precisely because they failed to increase their inceptor workload as required.

From the stick force and inceptor workload data, then, it appears that decreasing boundaries make pilots work harder. They use more force, more often, more quickly. This is an important conclusion, because it validates the Workload Buildup FTT as a method of increasing pilot workload in order to expose handling qualities deficiencies and PIO susceptibilities.

4.3. Characterizing Pilot Performance

The previous section demonstrated that decreasing boundaries increase pilot physical workload. If workload goes up as boundaries decrease, the next question is: what happens to performance? As theorized in Chapter 2 (Figure 10), tighter tolerances drive increased performance, until the tolerances become impossible to work within. The Workload Buildup profiles in this research provide data to test this theory.

The data in Figure 32 suggest that the increasing performance requirements (in the form of boundaries) did, in fact, increase performance. There is a trend toward lower average error for all terminal boundary sizes, with the following exceptions: there is an increase in average error for Aircraft H and I in the 16 degree terminal boundary, and for Aircraft I in the 8 degree terminal boundary. However, the performance theory also states that at some point requirements can become too stringent, and performance will suffer. This is most likely what occurred during these periods of increasing error.

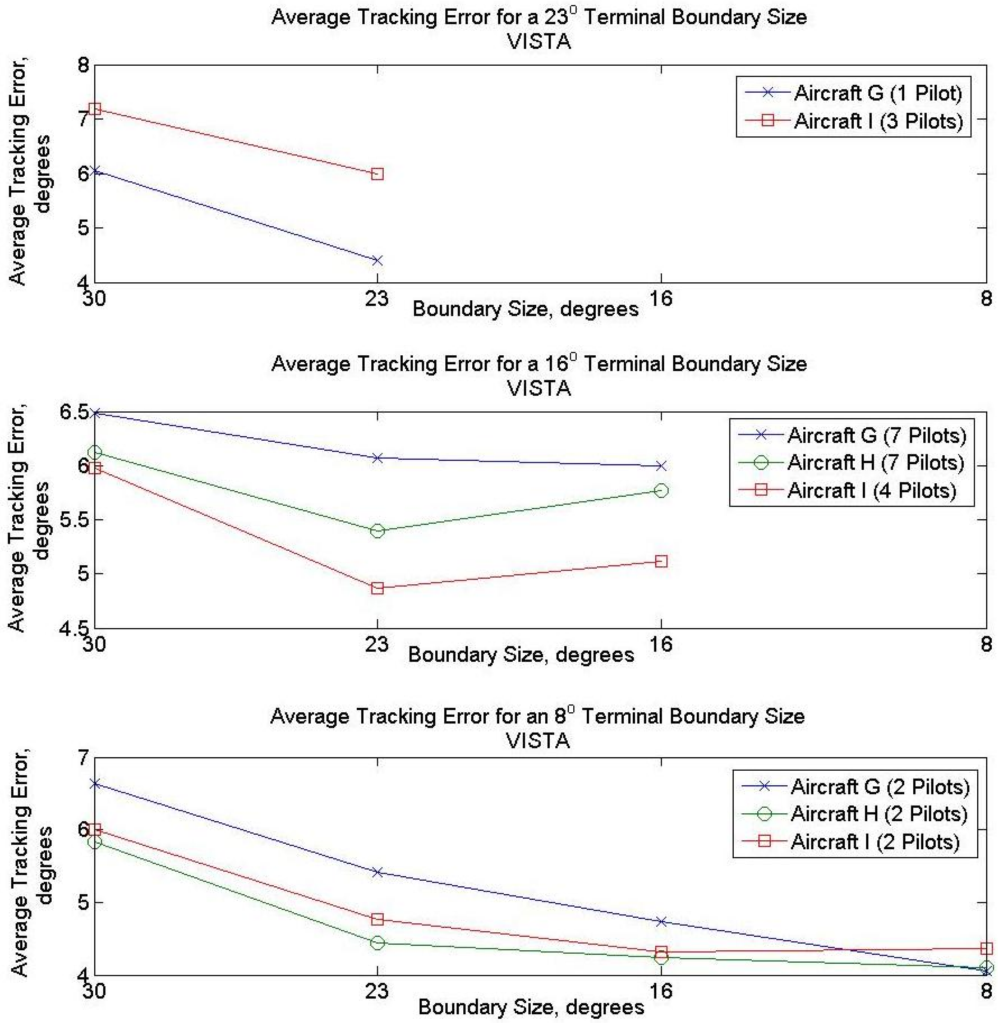


Figure 32. Average Roll Error (VISTA)

The BAT DART flight tests also demonstrated a decrease in average error as boundaries decreased. It was theorized that pilots altered their behavior at the beginning of the profile (when the boundaries were wide) in order to minimize the required negative g-loads. This aversion to negative g-loads, then, served as a hidden boundary. The current research does not have this additional boundary, so it is interesting that the results from these tests are similar to the results from the BAT DART tests.

The similarity in results can be explained by looking at performance from individual profiles. By studying the individual profiles, it is evident that the demonstrated decrease in roll error is true on average, but it is not always true in particular. Figure 33 contrasts the behavior of two different pilots attempting the same task in the same aircraft. Pilot 27 discovered that he could remain within the boundaries

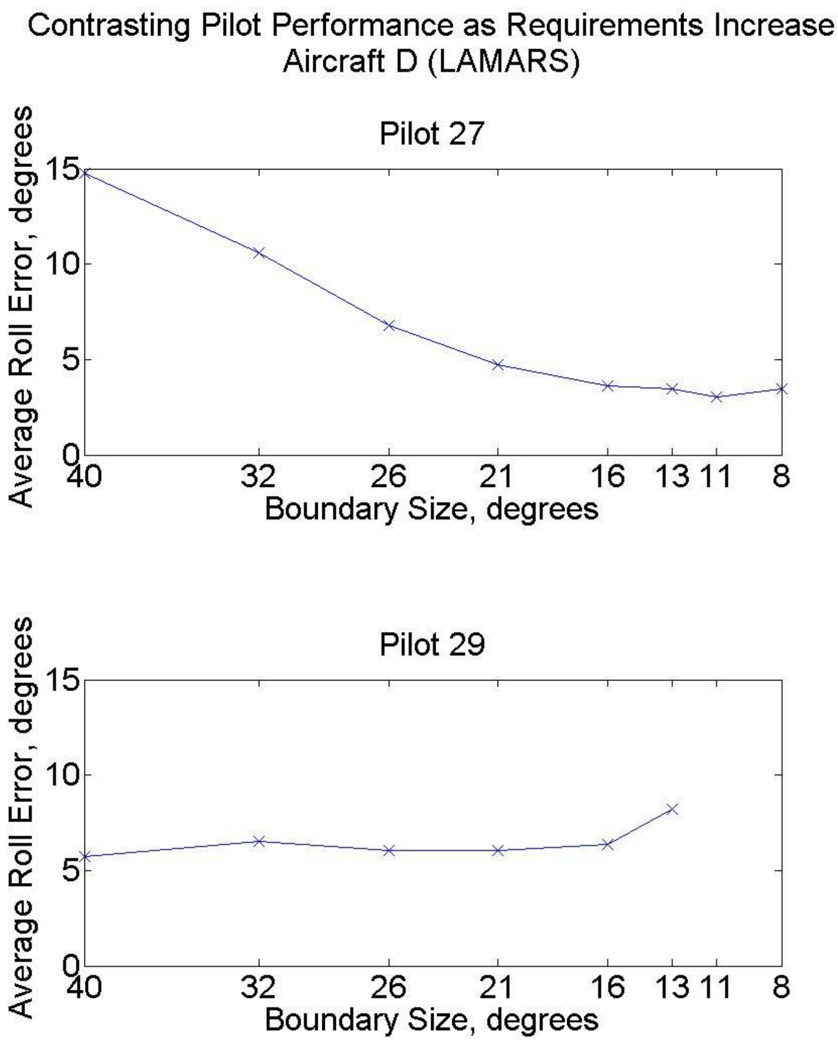


Figure 33. Contrasting Pilot Performance as Requirements Increase

during the first 30 second period without making any control inputs at all. As the boundaries decreased, he began to make control inputs, but only as little as required. Finally, at about the 16 degree boundary size, his performance leveled off. He had reached his maximum performance level. This perfectly supports the theory that increasing performance requirements will increase performance until maximum attainable performance is reached.

Pilot 29, however, had relatively constant roll error throughout the entire profile. From the beginning, he was attempting to faithfully track the target. In his mind, he had given himself performance requirements that were more stringent than the requirements from the boundaries. He apparently continued to use these internal requirements until the boundaries became as tight as his own tolerances and he exceeded the boundary.

So, it does appear to be true that increasing performance requirements result in increasing performance. But these performance requirements may often reside in a pilot's mind. The important result is that for those pilots who are willing to accept large errors, the Workload Buildup FTT forces them to accept less error as boundaries are decreased, and in the end it forces them to perform at their maximum capacity. This is an important validation of the Workload Buildup FTT.

4.4. Characterizing Pilot Opinion Rating in Relationship to a Workload Buildup Task

Results from the previous sections have provided performance and workload measurements which provide objective data to describe how a pilot controls an aircraft, and these measurements therefore provide quantitative measures of aircraft handling qualities. However, since Cooper-Harper ratings are the accepted standard in the flight test community for describing handling qualities, it is important to compare these ratings to the objective measurements obtained from Workload Buildup tasks.

4.4.1. Cooper-Harper Ratings

The ratings given during the Cooper-Harper profiles are presented in the following figures. Note that according to current practice, these ratings (along with pilot comments) are the only available data for handling qualities evaluations.

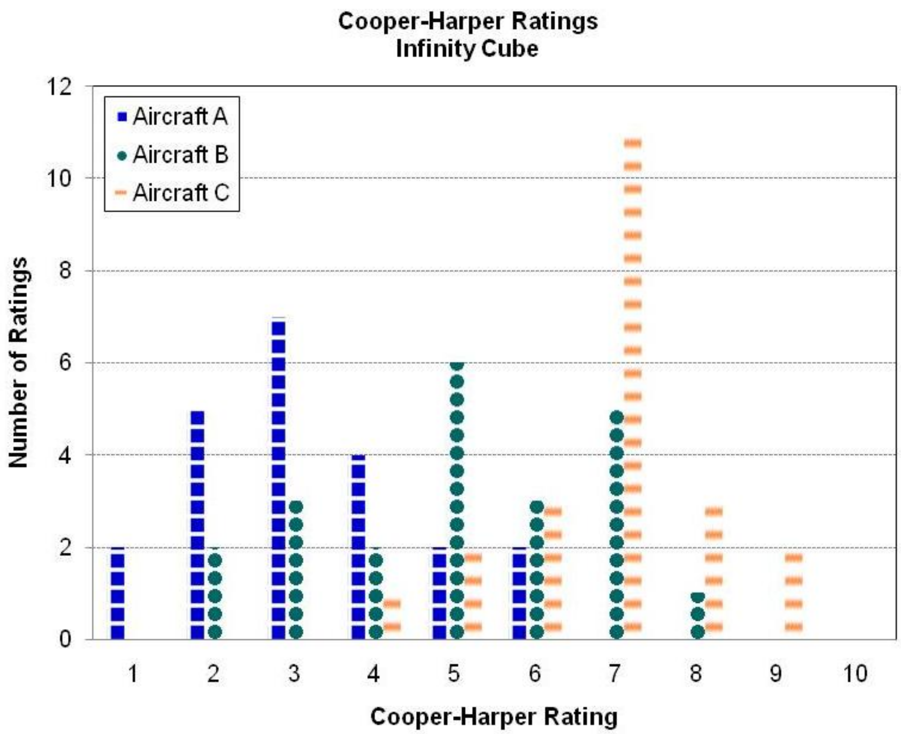


Figure 34. Infinity Cube Cooper-Harper Ratings

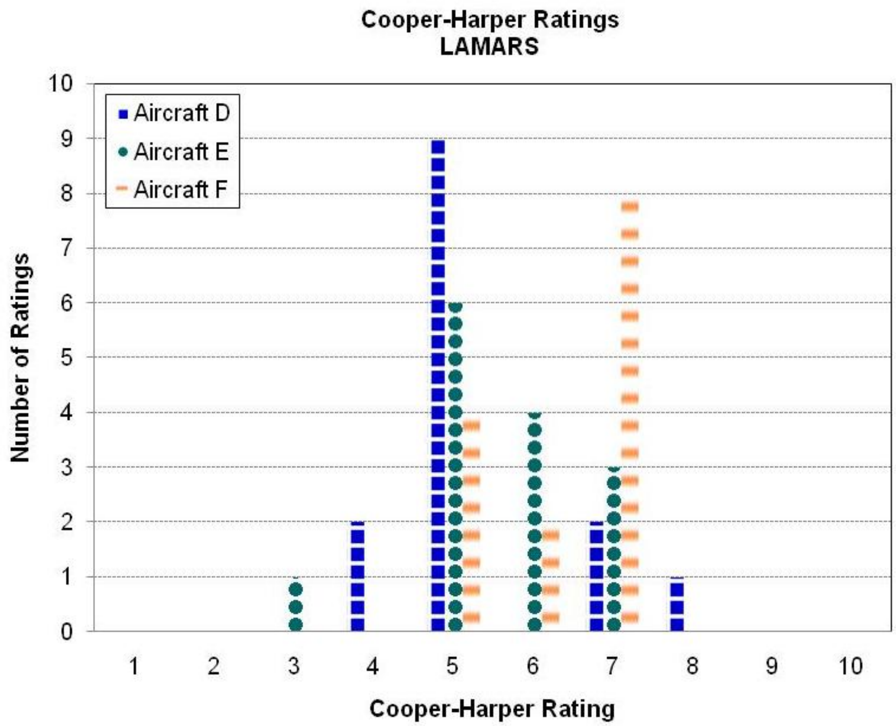


Figure 35. LAMARS Cooper-Harper Ratings

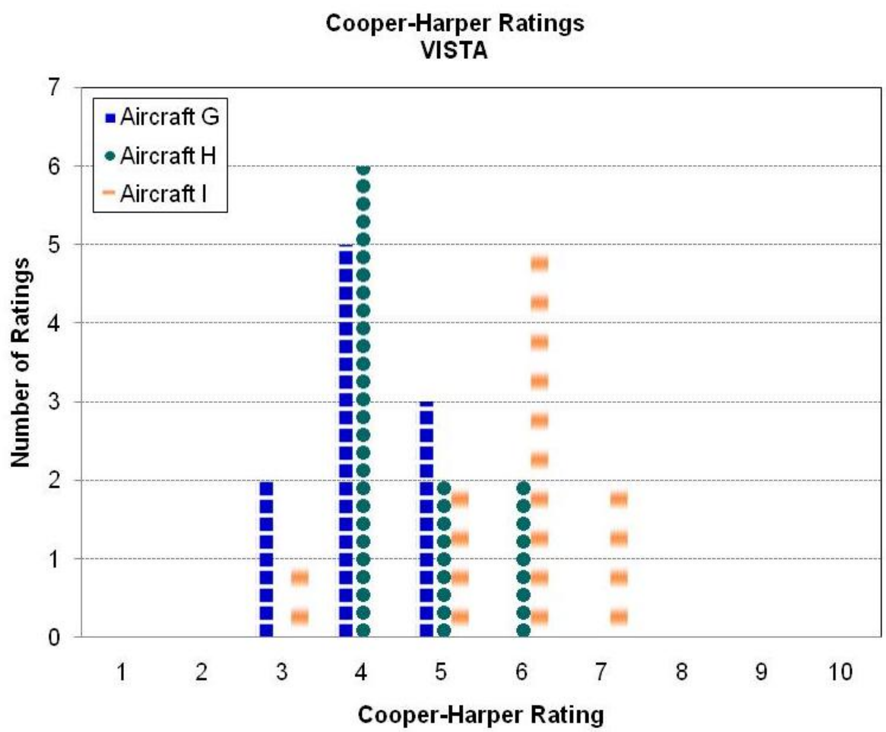


Figure 36. VISTA Cooper-Harper Ratings

The aircraft models in the Infinity Cube simulator had overlapping Cooper-Harper ratings, but there is still evident separation among the models. The median and mode ratings are 3, 5 and 7 for Aircraft A, B and C, respectively. This makes the models roughly Level 1, Level 2, and Level 3. The ratings in the LAMARS simulator show even more overlap, particularly between Aircraft D and E. Both Aircraft D and E are roughly Level 2 aircraft. Aircraft F appears to be on the border between Level 2 and Level 3. The VISTA ratings are similar.

4.4.2. Objective Performance and Workload Measurements

The data from the Workload Buildup plus Secondary Task profiles were analyzed in order to provide performance and workload measurements. Performance was defined as the minimum achievable boundary size. Workload was defined as the last boundary size for which the pilot could perform the secondary task correctly 50% of the time. This boundary size was labeled the critical boundary size. Thus, the performance and workload during the profile were summed up by two numbers. These numbers are objective measures that can supplement the subjective opinions of the pilot. (This is in addition to the objective measurements provided by the BAT model parameters.)

Figure 37 shows an example of how these measurements were made. When Pilot 30 flew aircraft F, he achieved greater than 50% secondary task performance during the first two boundary sizes (40º and 32º), but fell below 50% secondary task performance for the remainder of the profile. He finally exceeded the boundaries at a 21º boundary size. Thus, for Pilot 30, Aircraft F had a critical boundary size of 32º and a minimum achievable boundary size of 21º. Similarly, Aircraft D had a critical boundary size of 21º and a minimum achievable boundary size of 13º. With Aircraft E, Pilot 30 maintained greater than 50% secondary task performance for the entire profile. Thus, the critical boundary size and minimum achievable boundary size were both equal to 13º for Aircraft E.

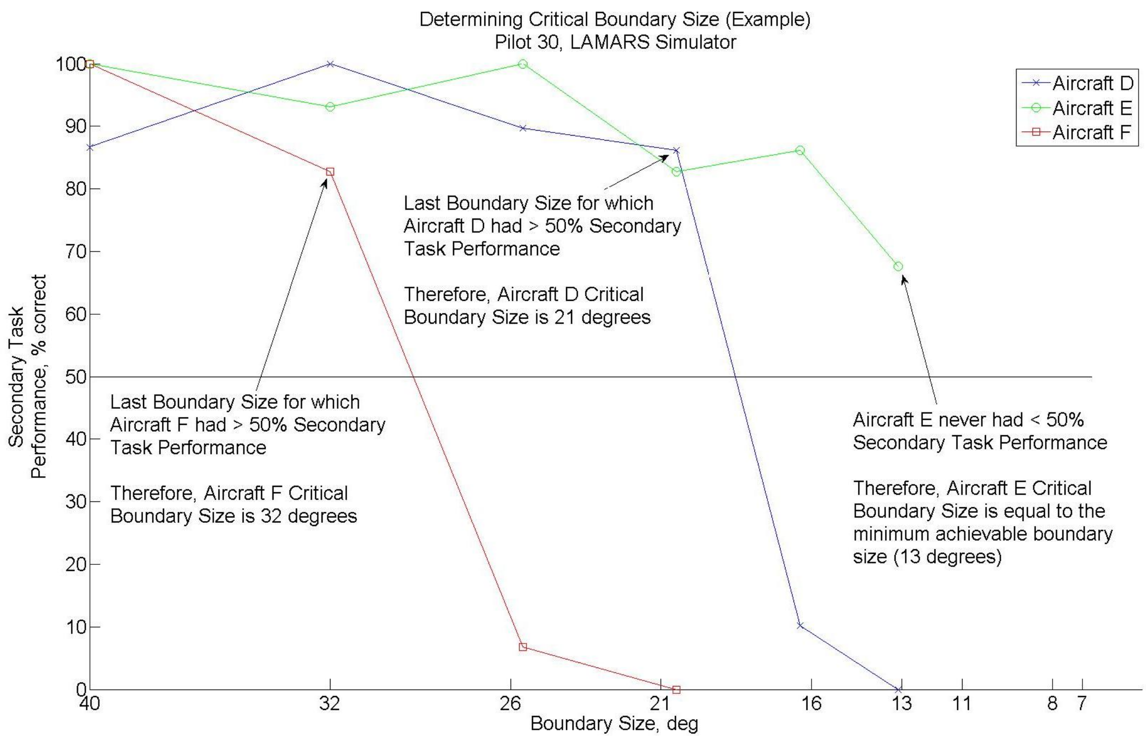


Figure 37. Determining Critical Boundary Size

4.4.3. Comparing Subjective and Objective Results

Combining the data from the Workload Buildup plus Secondary Task profile and the Cooper-Harper profile, there are three parameters describing an airplane: minimum achievable boundary size, critical boundary size, and Cooper-Harper rating. Figure 38 through Figure 40 present critical boundary size versus minimum achievable boundary size, with the Cooper-Harper rating placed as a marker.

Performance increases from left to right along the chart. Data points on the left side of the chart indicate pilot-aircraft pairs that had a high minimum achievable boundary size (i.e. boundaries were exceeded early, at a large boundary size, and thus performance is poor).

Workload decreases from top to bottom along the chart. Data points on the top side of the chart indicate pilot-aircraft pairs that had a high critical boundary size (i.e. secondary task performance dropped below 50% early, at a large boundary size, and thus workload is high).

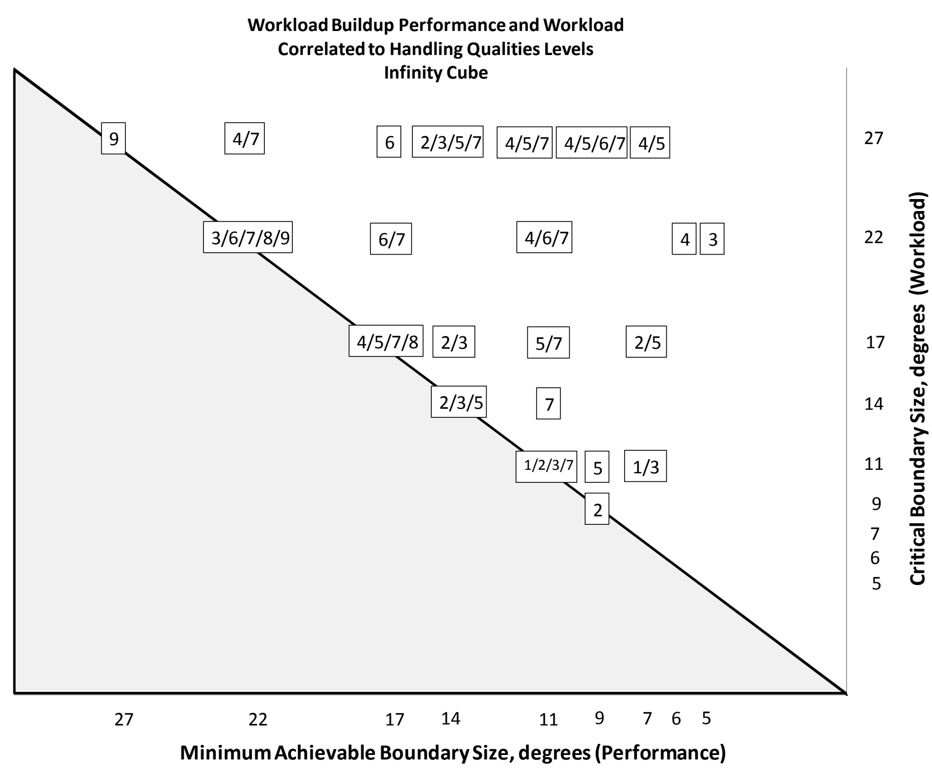


Figure 38. Workload Buildup and Cooper-Harper Ratings (Infinity Cube)

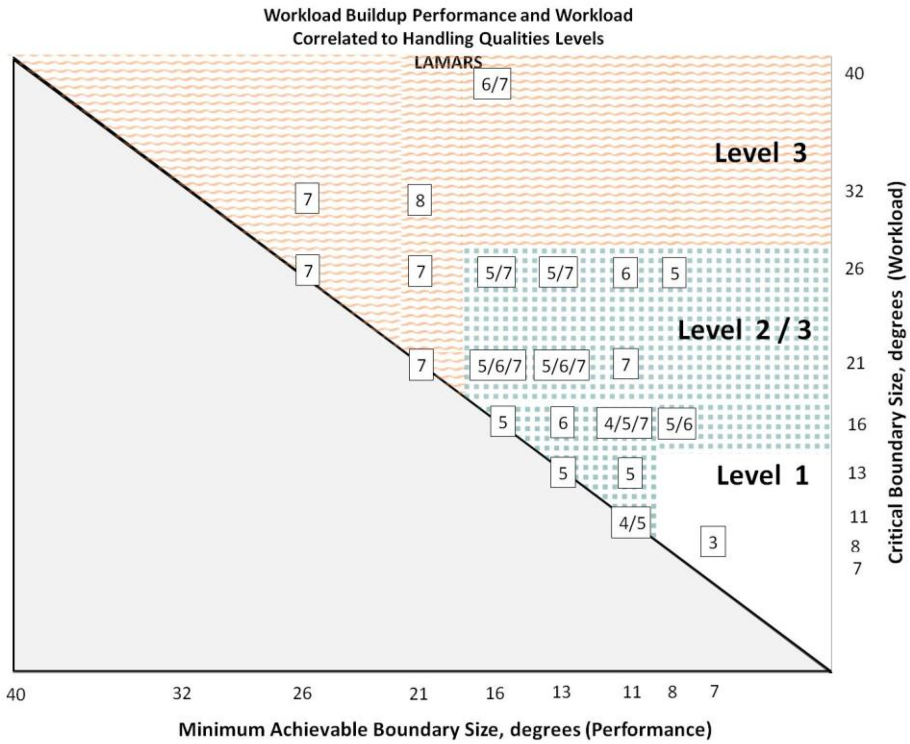


Figure 39. Workload Buildup and Cooper-Harper Ratings (LAMARS)

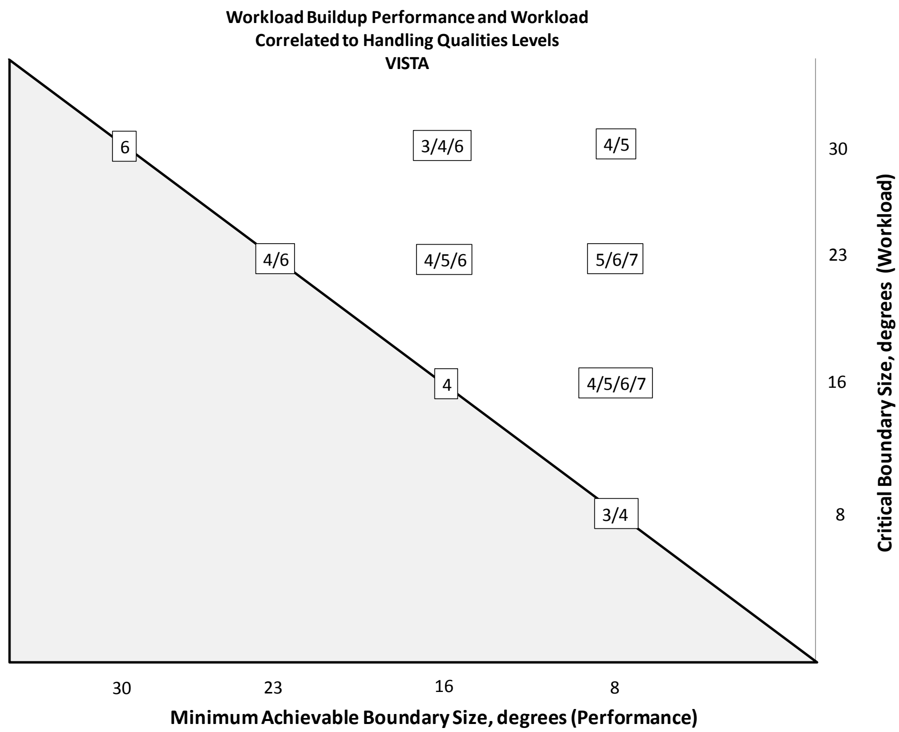


Figure 40. Workload Buildup and Cooper-Harper Ratings (VISTA)

The solid black line, running from the top left to the bottom right of the chart, represents the situation where a pilot maintains spare capacity all the way to the minimum achievable boundary size. The bottom left half of the chart is unattainable (spare capacity has no meaning after the boundaries have been exceeded). The top right half of the chart represents the normal state of affairs: primary task saturation is reached before minimum achievable boundary size is reached.

Intuitively, better Cooper-Harper ratings should be plotted near the bottom right corner of the chart (i.e. aircraft receiving these ratings should have smaller minimum achievable and critical boundary sizes). Worse Cooper-Harper ratings should be plotted toward the top and left of the chart. A Cooper-Harper rating plotted near the top right of the chart would theoretically represent an aircraft which performed well (i.e. had a small minimum achievable boundary size) but had a high workload (i.e. lost spare capacity early).

The data from the fixed-base simulations (Figure 38) show Cooper-Harper ratings scattered across the chart, with no obvious dividing lines between ratings or even handling quality Levels. However, most of the pilots in the Infinity Cube tests were not trained in the use of the Cooper-Harper scale. Even before these data were analyzed, it appeared that many of the pilots were using the scale incorrectly. There were instances of Level 1 ratings given to aircraft which could not attain even adequate performance in the Cooper-Harper profile. Therefore, the fact that the Cooper-Harper ratings are scattered across the chart may be due primarily to the misuse of the Cooper-Harper scale. The misuse of the Cooper-Harper scale is actually a good reason to adopt the use of critical boundary size and minimum achievable boundary size as handling qualities measurements, because these measurements are objective and are not subject to misuse.

Likewise, the VISTA data in Figure 40 show quite a bit of scatter. There are two explanations for this. One explanation is that, as can be seen in the Cooper-Harper rating histogram (Figure 36), the aircraft models were similar, and were all roughly Level 2 aircraft. It is possible that the entire plotted area in the chart is a Level 2 area. If an actual Level 1 aircraft was tested with these profiles, it is possible that its performance and workload would put it at the bottom left corner of the chart, away from the Level 2 points. The second explanation is that, as mentioned in Chapter 3, the boundaries during the VISTA tests decreased linearly (rather than as a percentage). The step from ±16º to ±8º was a large one, and no pilot operated within the ±8º boundaries for more than a few seconds. A finer decrease increment may have exposed difference in performance and workload in the boundary regions between 16º and 8º. This might have separated the ratings on the chart and allowed some division between ratings.

The LAMARS data in Figure 39 actually do show a correlation between Cooper-Harper ratings and critical boundary size/minimum achievable boundary size. The handling qualities of the three aircraft models were similar, as in the VISTA tests, but there appears to be some spread among the ratings in the chart. The only aircraft receiving a Level 1 rating is plotted in the bottom right corner of the chart, away from the others. The only aircraft receiving a Cooper-Harper rating of 8 is plotted in the upper left portion of the chart, fairly well separated from the Level 2 ratings. Much of the chart in the middle is a mix of Level 2 and Level 3 airplanes. But a line can be drawn through a minimum achievable boundary size of 16º, and another through a critical boundary size of 26º. To the top and left of this line, all ratings are Level 3 (with the exception of one outlier, a Cooper-Harper rating of 6).

The data were also analyzed by plotting critical boundary size versus minimum achievable boundary size as before, but with the aircraft model (instead of the Cooper-Harper rating) placed as a marker. The results for the LAMARS simulator are shown in Figure 41. The worst performance (26º minimum achievable boundary size) and the highest workload (40º critical boundary size) were achieved by the aircraft with the worst predicted handling qualities (Aircraft F). However, Aircraft F also had performance and workload in the bottom right of the chart, equal to the performance and workload of the

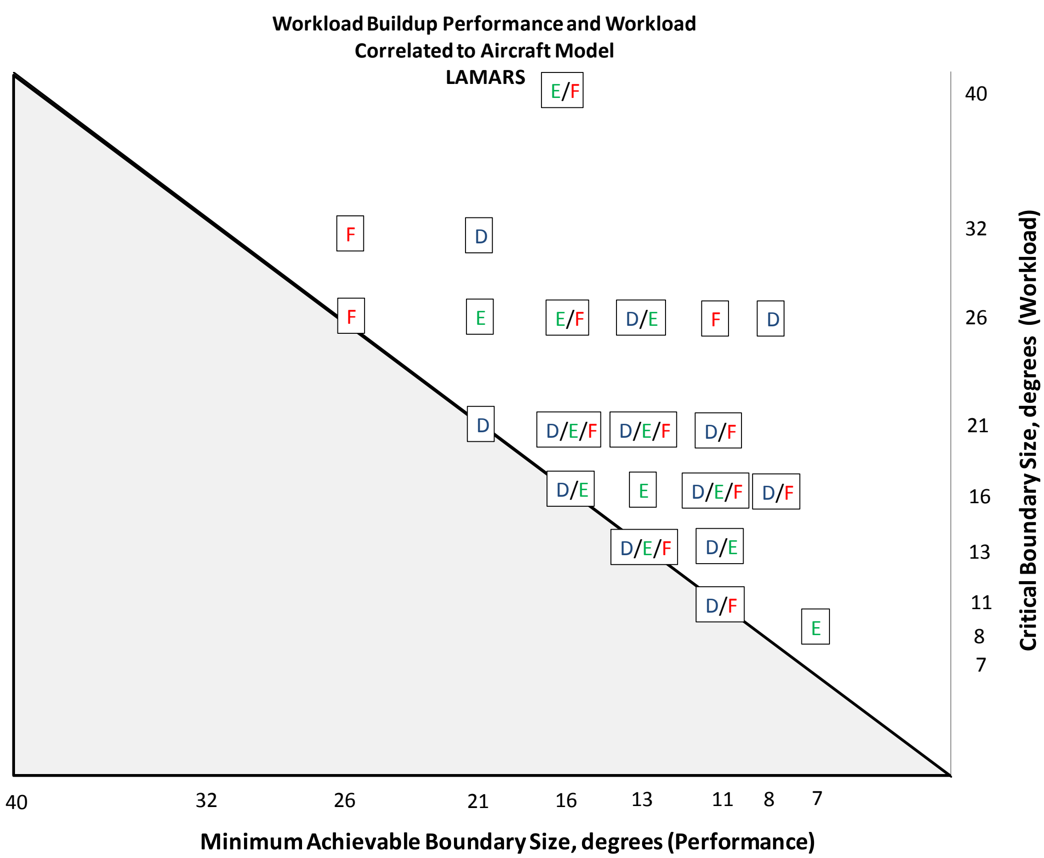


Figure 41. Workload Buildup and Aircraft Models (LAMARS)

aircraft with the best predicted handling qualities (Aircraft D). Many of the plotted data points include all three aircraft models. In general, there appears to be a better correlation between Workload Buildup measurements and Cooper-Harper rating than between Workload Buildup measurements and aircraft model. The probable reason is that some pilots adapt better to certain aircraft than others, and can therefore achieve good performance and low workload in an aircraft that another pilot might have difficulty with. This is one reason why Cooper-Harper ratings often show so much scatter; different pilots sometimes like different airplanes. The fact that there is no obvious correlation between Workload Buildup measurements and aircraft model is still an important observation, because it shows that the Workload Buildup measurements behave in much the same way as Cooper-Harper ratings.

An important result from the test data is that the theorized relationship between Cooper-Harper ratings and critical boundary size/minimum achievable boundary size appears to be coherent for individual pilots. In other words, in most situations pilots rated aircraft at the top left of the chart worse than aircraft at the bottom right of the chart. Two examples are shown in Figure 42. In these two examples, the ratings generally improve from left to right and top to bottom. This means that pilots were aware of their own performance and workload and assigned Cooper-Harper ratings accordingly. Thus, for a given pilot, the combination of critical boundary size and minimum achievable boundary size appear to sum up a pilot's *relative* impressions of the handling qualities of different aircraft. This is exactly how the pilot opinion rating process is supposed to work, and it validates Workload Buildup measurements as objective measures of handling qualities.

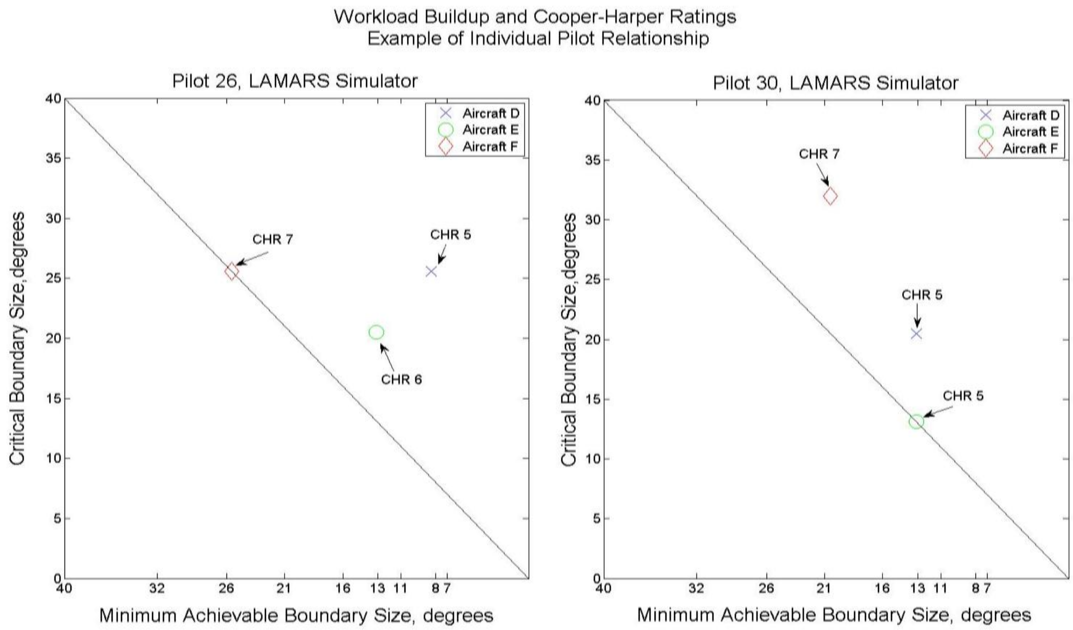


Figure 42. Workload Buildup and Cooper-Harper Ratings for Individual Pilots

4.4.4. Summary of the Relationship Between Pilot Opinion Rating and Performance/Workload in a Workload Buildup Task

The data from the Workload Buildup plus Secondary Task profiles were analyzed in order to provide objective performance and workload measurements. These measurements were combined in a unique way in order to characterize pilot opinion rating in quantitative terms. As pilot performance (in terms of minimum achievable boundary size) improved, and as workload (in terms of critical boundary size) decreased, pilots tended to assign better Cooper-Harper ratings. This trend generally held true for individual pilots, and was even shown to be true in aggregate for the LAMARS test data. The new method of measuring and presenting objective performance and workload measurements is a useful tool which can be used to quantify handling qualities, and which should be researched further.

5. Conclusions

In just over four years, the BAT model has evolved from an interesting idea into a mathematical model supported by simulator and in-flight experiments, and into a flight test technique designed to aid in the description of aircraft handling qualities and PIO susceptibility. Much has been learned about the way pilots behave in the presence of boundaries, and these results have been put to use in the flight test community.

This research focused on the following four objectives:

- Characterize the mathematical BAT model through comparison with simulator and flight test data
- Characterize pilot behavior (in terms of control stick forces and pilot physical workload) in the presence of decreasing boundaries
- Characterize pilot performance (in terms of tracking error) in the presence of decreasing boundaries
- Characterize the relationship between pilot opinion rating and performance/ workload in a Workload Buildup task

5.1. Contributions to Test Methodology

Because there were no techniques for BAT research in the roll axis prior to this research, considerable time and effort were spent in developing appropriate test techniques to support the research objectives. The methodology developed in this research provided data, such as quantitative workload measurement, which had not been previously available. In addition, the tasks and techniques which were developed have been validated and can be used in future research.

A new tracking task was developed for this research which was shown to be an improvement on tracking tasks used in previous BAT research. This task not only provided much useful data for this research, but has also been incorporated into handling qualities simulations used in the TPS syllabus.

A new flight technique was also developed to measure performance and workload during a Workload Buildup task, using a primary tracking task and a secondary task. This type of workload measurement in the presence of boundaries had not been attempted before, and many hours of research and simulation were conducted in order to design the most effective secondary task. After the task was designed, it was validated as a measure of pilot workload through the use of eye-tracker equipment. The eye-tracker data and the secondary task data showed remarkable correlation.

5.2. Characterizing the BAT Model

The mathematical BAT model had been demonstrated in theoretical situations and computer simulations in previous research, but it had never before been tested with flight data. In fact, combining both loops of the model and demonstrating its switching behavior had never been attempted at all. In this research, 123 pilot-vehicle systems performing a Workload Buildup profile were analyzed using numerical optimization techniques. BAT model matching had never before been done on this scale, and had never used optimization. Optimization not only ensured the best fit to the test data, but also suggested a change to the boundary tracking stick force equation which resulted in better matches. In addition, whereas previous research had only exercised portions of the mathematical BAT model, this research compared flight test data to the predictions of the entire model.

The results from this analysis suggested new conclusions. First, the assumption that the boundary tracking feedback increases linearly from t_{min} to t_{max} was challenged. A proposed change to the model suggested that the feedback increases quadratically from t_{min} to t_{max}, and this new model produced a lower minimized cost during optimization over 80% of the time.

The parameters from all 123 profiles were collected, and the results suggest that the parameters are not constant for a given pilot. The pilot adapts his or her control strategy to fit the airplane, and thus selects different BAT parameters for different airplanes. This also suggests something about the airplanes themselves. The data showed groupings of parameters that varied between airplanes, suggesting that different pilots adopt similar boundary tracking strategies for a given airplane. These BAT parameters, then, can be useful not only in describing a pilot, but also in describing an aircraft. The trend in parameter variation among aircraft is a new result, and provides a method to quantify aircraft handling qualities, by describing a given aircraft according to the average BAT parameters used by pilots to control that aircraft.

5.3. Characterizing Pilot Behavior

The data from the BAT model characterization process demonstrated that boundaries affect pilot behavior in measurable ways. It also possible to describe a pilot's behavior in the presence of boundaries by looking at his or her behavior as boundaries are tightened. The data from this research show that as boundaries are tightened, a pilot's average stick force tends to increase. A pilot's duty cycle and aggressiveness (collectively known as inceptor workload) also increase as boundaries are tightened. From the stick force and inceptor workload data, then, it appears that decreasing boundaries make pilots work harder. They use more force, more often, more quickly. This is an important conclusion, because it validates the Workload Buildup FTT as a method of increasing pilot workload in order to expose handling qualities deficiencies and PIO susceptibilities.

5.4. Characterizing Pilot Performance

The data also indicate that, on average, decreasing boundaries improve pilot performance until the pilot is performing at maximum capacity. This is a validation of previous research results. Pilots tend to accept a certain amount of error when the performance demands are not too stringent, but will tighten control when large error could cause exceeding potentially fatal boundaries. It is important to note that this error acceptance on the part of the pilot is internal, but can be forced lower by the imposition of boundaries. (However, as described above, this comes at the expense of higher workload.) Thus a pilot's performance can actually be improved through the imposition of boundaries. This is an important validation of the Workload Buildup FTT, which was designed to elicit the maximum possible performance from a pilot.

5.5. Characterizing Pilot Opinion Rating in Relationship to a Workload Buildup Task

Previous research had shown a qualitative relationship between pilot opinion rating and performance in a Workload Buildup task. Since workload is just as important as performance when determining the handling qualities of an aircraft, this research attempted to incorporate a quantitative measure of pilot workload in the form of a secondary task. This workload measurement was then combined with performance measurements in a Workload Buildup task in order to provide objectivity to what is currently a subjective pilot opinion rating process.

The performance measure during the Workload Buildup plus Secondary Task profile was chosen to be minimum achievable boundary size. An aircraft that can track a target in the presence of tight boundaries is a well performing airplane. The workload measure was chosen to be the boundary size at which the secondary task could no longer be completed 50% of the time. This point was labeled the critical boundary size. An airplane that allows the pilot to maintain spare capacity in the presence of tight boundaries is an airplane that requires a low workload. The two parameters were plotted against one another and compared to the Cooper-Harper rating given to the airplane in a separate task.

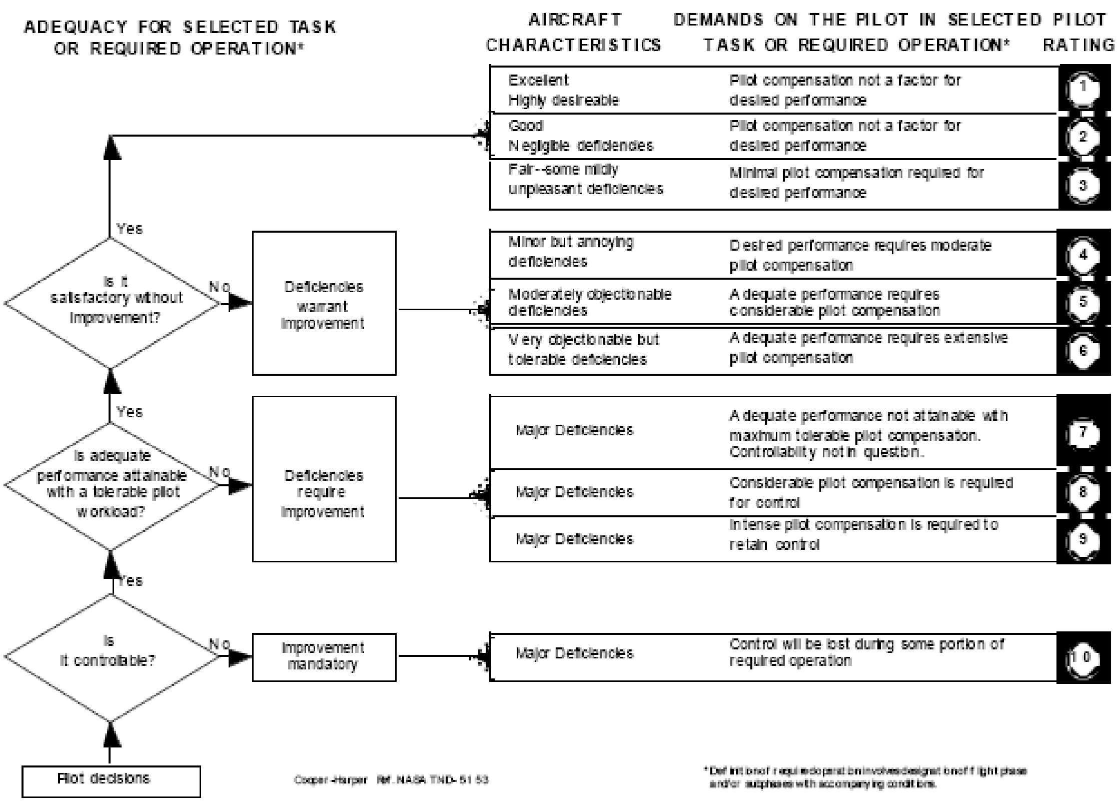
As pilot performance (in terms of minimum achievable boundary size) improved, and as workload (in terms of critical boundary size) decreased, pilots tended to assign better Cooper-Harper ratings. This trend generally held true for individual pilots, and was even shown to be true in aggregate for the LAMARS test data. This new method of measuring and presenting objective performance and workload measurements is a useful tool which can be used to quantify handling qualities, and which should be researched further.

5.6. Areas for Further Research

The BAT model has now been exercised in its entirety and shown to produce reasonable results. The parameters grouped themselves in interesting ways, providing a quantitative measure of aircraft handling qualities, but there is not yet an explanation for why the BAT parameters take on certain values. Future research should attempt to discover the basis for these parameters. The ultimate goal is to develop a BAT model so that it does not generally vary among pilots or aircraft. Such a model could then be used in the aircraft design phase in order to predict handling qualities.

More research needs to be done into the correlation between pilot opinion rating and performance in a Workload Buildup task. This research was hindered somewhat by aircraft models which had very similar handling qualities, thus making the task of differentiating between models difficult. Future research should focus on obtaining models with dramatically different handling qualities and widely separated Cooper-Harper ratings.

Appendix A – Cooper-Harper Scale (Cooper and Harper, 1984)



Appendix B – VISTA Flight Test Cards

	AT BAT Data Sortie #	

DATE	JON/MSN MT080500	FREQUENCY 286.8 MHz	STEP A/R	START A/R	TAKEOFF 1000
TEMP/PA	T/O GW 28899	ZERO FUEL GW 21004	T/O CG 35.7%	JOKER 2.0	BINGO 1.5

CALLSIGN	AIRCREW	OPS #	TAIL#	A/A	M1/M3	T/O

RESTRICTIONS/LIMITATIONS

	AIRSPEED		ACCEL (g)				
CONFIG	KCAS	MACH	SYM ±	ASYM ±	AOA	ROLL	BANK
Centerline	440 (VSS) 550	0.9 (VSS) 1.2	-2.4 / +6.8		CAT I		

<u>MISSION PROFILE:</u>
Backseat Takeoff
Roll Frequency Sweep Y and Z Aircraft
Profiles per test matrix

Task	Profile	Config	D1S	AT BAT	Run	Notes
1	BAT Plus	8 1 Fuel	016	200		
2	BAT	8 2 Fuel	016	000		
3	BAT Plus	8 3 Fuel	016	200		
4	CHR	8 2 Fuel	015	000		
5	BAT	8 1 Fuel	016	000		
6	CHR	8 3 Fuel	015	000		
7	BAT Plus	8 2 Fuel	016	200		
8	BAT	8 3 Fuel	016	000		
9	CHR	8 1 Fuel	015	000		

Backseat approach / landing

ACCEL	ROTATE	T/O SPEED	DISTANCE	IMMED APP	DIST	APP SPEED + FUEL

SFO (H/W)	/ /	SFO (PRAC)	FUEL ( )	/ /

Configuration **CR**			**RUN 1**	**1**
	Condition	Databand	Tolerance	Limits
ALT	**20,000ft**	**±5,000ft**	N/A	
A/S	**300KCAS**	**±50KCAS**	N/A	

Procedure

Profile:
Aircraft Model:
Primary Task:
Secondary Task:

Data

Time of Flight:
CHR (N/A If BAT PLUS or BAT):
Time to Drop Secondary Task (N/A If CHR or BAT):
Pilot Comments:

Note: Nine test run cards like this one were taken on each sortie

Next Card Setup Conditions

	RUN 2	2

List of Abbreviations, Symbols, and Acronyms

Symbol	Definition	Units
AFRL	Air Force Research Laboratories	--
BAT	Boundary Avoidance Tracking	--
CCD	Charge-Coupled Device	
FTT	Flight Test Technique	--
HDD	Heads-Down Display	--
HQDT	Handling Qualities During Tracking	--
HUD	Heads-Up Display	--
LAMARS	Large Amplitude Multi-Mode Aerospace Research Simulator	--
PIO	Pilot-In-the-loop Oscillation	--
TPS	Test Pilot School	--
VISTA	Variable-stability In-flight Simulator Test Aircraft	--
VSS	Variable Stability System	--
δ_a	aileron deflection	deg
δ_e	elevator deflection	deg
ζ_{dr}	Dutch Roll damping ratio	--
τ_b	pilot time-delay during boundary avoidance	sec
τ_r	roll mode time constant	sec
τ_s	spiral mode time constant	sec
ω_{dr}	Dutch Roll natural frequency	rad/sec
c	cost function for BAT model optimization	lbs^2
e	pilot model error	lbs
e_{pt}	point tracking roll error	deg

$\dot{e}_{pt}$	point tracking roll error rate	deg/sec
F	stick force	lbs
F_{actual}	actual stick force from test data	lbs
$F_{predicted}$	predicted stick force from test data	lbs
F_{PT}	point tracking stick force	lbs
g	normal load factor	--
K_{bm}	maximum gain used by the pilot to avoid boundaries	A/R
K_d	derivative control gain	lbs/deg/sec
K_p	proportional control gain	lbs/deg
p	roll rate	deg/sec
q	pitch rate	deg/sec
rms	root-mean-square	A/R
s	Laplace variable	1/sec
t	time	sec
t_b	time to boundary	sec
t_d	time to double amplitude	sec
t_{min}	time to boundary at which pilot begins to react to boundaries	sec
t_{max}	time to boundary at which pilot uses maximum gain to avoid boundaries	sec
x_b	distance to boundary	A/R
x_{BAT}	vector of design variables for boundary tracking model optimization	A/R
x_{PT}	vector of design variables for point tracking model optimization	A/R

Bibliography

Air Force Flight Test Center (AFFTC). *Flying Qualities Testing*. Edwards AFB, CA: 20 February 2002.

Arif, Muhammed and Inooka, Hikaru. "Experimental Study of Human Performance in High Precision Tracking Task Using Iterative Learning Technique", *Interdisciplinary Information Sciences*, 5:73-78 (1999).

Blake, Ryan; Lee, Chun-Hao; Keithley, Chris; Baccheschi, Nancy; Cooley, William. *Limited Investigation of Auxiliary Tasking During Boundary Avoidance Tracking (AT BAT).* AFFTC-TIM-08-09, 2008.

Cooper, George and Harper, Robert. *Handling Qualities and Pilot Evaluation*. Wright Brothers Lectureship in Aeronautics, 1984.

Department of Defense. *MIL-STD-1797B, Flying Qualities of Piloted Aircraft*. Wright Patterson AFB, OH: 15 February 2006.

Dotter, Jason. *An Analysis of Aircraft Handling Quality Data Obtained from Boundary Avoidance Tracking flight Test Techniques*. Air Force Institute of Technology, 2007.

Hess, Ronald. *An Inquiry into Whether a Pilot-Induced Oscillation was a Factor in the Crash of American Airlines flight 587*. National Transportation Safety Board, 2003.

Gray, William. *Boundary-Escape Tracking: A New Conception of Hazardous PIO*. Society of Experimental Test Pilots, 2004.

Gray, William. *Boundary-Avoidance Tracking: A New Pilot Tracking Model*. American Institute of Aeronautics and Astronautics, 2005.

Gray, William. A *Boundary Avoidance Tracking Flight Test Technique for Performance and Workload Assessment*. Society of Experimental Test Pilots, 2007.

Jex, Henry. *Measuring Mental Workload: Problems, Progress and Promises*. Systems Technology, Inc. 1987.

McRuer, Duane; Graham, Dunstan; Krendel, Ezra; Reisener, William Jr. *Human Pilot Dynamics in Compensatory Systems*. AFFDL-TR-65-15, 1965.

McRuer, D.T.; Hoffman, L.G.; Jex, H.R.; Moore, G.P.; Phatak, A.V.; Weir, D.H.; Wolkovitch, J. *New Approaches to Human-Pilot/Vehicle Dynamic Analysis*. Systems Technology, Inc. 1967.

McRuer, Duane; Clement, Warren; Thompson, Peter; Magdaleno, Raymond. *Minimum Flying Qualities Volume II: Pilot Modeling for Flying Qualities Applications*. Systems Technology, Inc. 1989.

Warren, Randy. *An Investigation of the Effects of Boundary Avoidance on Pilot Tracking*. Air Force Institute of Technology, 2006.

REPORT DOCUMENTATION PAGE		*Form Approved* *OMB No. 0704–0188*

1. REPORT DATE (DD–MM–YYYY)	2. REPORT TYPE	3. DATES COVERED (*From — To*)
26-03-2009	Master's Thesis	June 2007 – March 2009

4. TITLE AND SUBTITLE
Boundary Avoidance Tracking: Consequences (and Uses) Of Imposed Boundaries On Pilot-Aircraft Performance

5a. CONTRACT NUMBER
5b. GRANT NUMBER
5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Ryan Blake, Maj, USAF

5d. PROJECT NUMBER
5e. TASK NUMBER
5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/ENY)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT/GAE/ENY/09-M03

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
United States Air Force Test Pilot School

Mr. William Gray
220 South Wolfe Ave
Edwards AFB, CA 93524
(661) 527-2761
william.gray@edwards.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
USAF TPS

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This thesis presents the results of research into the Boundary Avoidance Tracking (BAT) theory, which states that often pilots control an aircraft not in an attempt to maintain some condition, but to avoid some real or perceived boundaries. This pilot modeling concept was studied using over 30 pilots in simulator and flight tests. The pilot-aircraft system was evaluated with 3 different lateral-directional control models. Pilots were given a roll angle tracking task in the presence of shrinking boundaries. Pilots were also given a secondary task in some of the profiles in order to measure workload. Approximately 42 hours of simulation and 10 test flights were accomplished. The simulations were conducted in the Infinity Cube simulator and the Large Amplitude Multi-mode Aerospace Research Simulator (LAMARS) at the Air Force Research Laboratory. Flight test sorties were flown on the NF-16D Variable Stability In-Flight Simulator Test Aircraft (VISTA). The mathematical BAT model was compared to the flight test data in order to confirm the theory and validate the model. In addition, a correlation between pilot performance during these tasks and pilot opinion of the aircraft was studied.

15. SUBJECT TERMS
Test Management Project (TMP) Test Pilot School (TPS) Flight Testing NF-16D Aircraft Variable Stability In-flight Simulator Test Aircraft (VISTA) Pilot Induced Oscillation (PIO) Boundary Avoidance Tracking (BAT) Cooper-Harper Rating Scale

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Paul Blue, Maj, USAF
					19b. TELEPHONE NUMBER (Include Area Code)
U	U	U	UU	114	(402) 294-1497 paul.blue@stratcom.mil

Standard Form 298 (Rev. 8–98)
Prescribed by ANSI Std. Z39.18