HOPE: An On-Line Piloted Handling Qualities Experiment Data Book

E. Bruce Jackson[*]
NASA Langley Research Center, Hampton, VA 23681

and

Melissa S. Proffitt[†]
Adaptive Aerospace Group, Hampton, VA 23666

A novel on-line database for capturing most of the information obtained during piloted handling qualities experiments (either flight or simulated) is described. The Hyperlinked Overview of Piloted Evaluations (HOPE) web application is based on an open-source object-oriented Web-based front end (Ruby-on-Rails) that can be used with a variety of back-end relational database engines. The hyperlinked, on-line data book approach allows an easily-traversed way of looking at a variety of collected data, including pilot ratings, pilot information, vehicle and configuration characteristics, test maneuvers, and individual flight test cards and repeat runs. It allows for on-line retrieval of pilot comments, both audio and transcribed, as well as time history data retrieval and video playback. Pilot questionnaires are recorded as are pilot biographies. Simple statistics are calculated for each selected group of pilot ratings, allowing multiple ways to aggregate the data set (by pilot, by task, or by vehicle configuration, for example). Any number of per-run or per-task metrics can be captured in the database. The entire run metrics dataset can be downloaded in comma-separated text for further analysis off-line. It is expected that this tool will be made available upon request.

I. Introduction

A novel on-line database, the Hyperlinked Overview of Piloted Evaluations (HOPE), has been developed for use in recent handling qualities simulation experiments for spacecraft operations. This web application was written to be generic in nature and can be applied to any experiment involving human subjects with both objective and subjective data collection. This paper describes the key features of the application and how it can be applied to a typical simulation experiment. HOPE was inspired by previous work in NASA's High Speed Research activity, one of the first team-based piloted evaluations that made use of emerging Web technologies to organize, store and share handling quality experimental data with a widely dispersed team of flight control engineers and pilots.

II. Motivation

During NASA's High Speed Research (HSR) project[1] of the 1990s, both authors were involved in storing simulation data collected from multiple experiments at three different simulation sites into a single Web server for access by the geographically dispersed HSR Guidance and Control integrated development team. Access routines were very crude, consisting of custom-written shell scripts that retrieved time history data and summaries of pilot opinions for various maneuvers, hand-coded into a static hypertext markup language (HTML)-based table.

A series of experiments in spacecraft handling qualities at NASA Langley Research Center,[2,3] in collaboration with NASA Ames Research Center,[4,5,6] were begun in 2007. The previous experience with piloted evaluations of the High-Speed Civil Transport and other conceptual vehicles motivated the authors to take advantage of newly emerging Web technologies to store, retrieve and analyze pilot opinions and task metrics in a user-friendly way.

---

[*] Senior Research Engineer, Dynamic Systems and Control Branch, MS 308, Associate Fellow.
[†] Programmer/Data Analyst, Hampton Operations.

III. Solution Implementation

A. Programming Language

The Ruby programming language[7] was chosen along with the Ruby-on-Rails (RoR) framework[8]. Ruby is an open-source, multi-platform, general purpose scripting language that is strongly object-oriented and includes key features such as open classes (meaning one can redefine or add methods to built-in classes), introspection, and built-in support for unit testing. The Ruby-based Rails framework is an open-source project that adds support for development of Web-based applications; such web applications are designed to access and manipulate an underlying relational database; many open-source and commercial database products (such as MySQL, sqlite, and Oracle) are supported by RoR.

B. Custom code development

Ruby-on-Rails represents a domain-specific application of Ruby to provide a model-view-controller (MVC)[9] Web database application. This meant building a custom Web application for our purposes was straightforward. The HOPE application was realized in approximately 3,000 lines of custom code (excluding comments) in order to provide the current functionality. Ruby and RoR development encourages unit testing of code elements. At this point, approximately 400 lines of unit test code have been written for the HOPE application. (The unit test code, while part of the HOPE project, was not counted in the 3,000 lines of application code.)

Ruby-on-Rails must be hosted on a web server; Rails includes a web server for a standalone application or Rails can be installed under an existing Apache[10] server. The HOPE application includes support for authorized user accounts, so the resulting HOPE databook can be open, closed, or read-only to Web visitors.

C. Problem space decomposition

In order to collect data from handling qualities and other simulation- or flight-based piloted experiments, investigators usually encounter (and have broken the problem space down into) the following highlighted **objects**:

> An evaluation **pilot** participates in a piloted **experiment**‡. During that experiment, she participates in one or more **sessions** and performs **task cards** meant to evaluate various **vehicle configurations** performing certain maneuvers or **tasks**. In order to form an opinion, she performs multiple **run cards** of the same task in the same vehicle configuration, some runs for practice and other runs for data. During the runs, the performance of the vehicle is recorded and certain summary **metrics** are collected (such as touchdown descent rate or docking attitude) that are compared against pre-specified **task criteria** for desired or adequate (or not adequate) performance. At the end of the runs, based on her performance and perceived amount of compensation and workload, she provides **evaluations** in the form of ratings or pilot opinions, and verbal comments, of that configuration and task. She may provide additional summary **evaluations** in response to questionnaires in the course of the experiment.

The objects indicated above correspond to tables within the underlying relational database, where each row of the table represents a single object instance. The relationship between the database tables is shown in Fig. 1. Each column of a table contains information about a specific object; columns ending in _id indicate keys that refer to an object (row) in another class (table) as shown in the figure.

---

‡Information about the experiment is stored in the Exp Info table as shown in Figure 1.

Hyperlinked Overview of Piloted Evaluations (HOPE) Database Schema
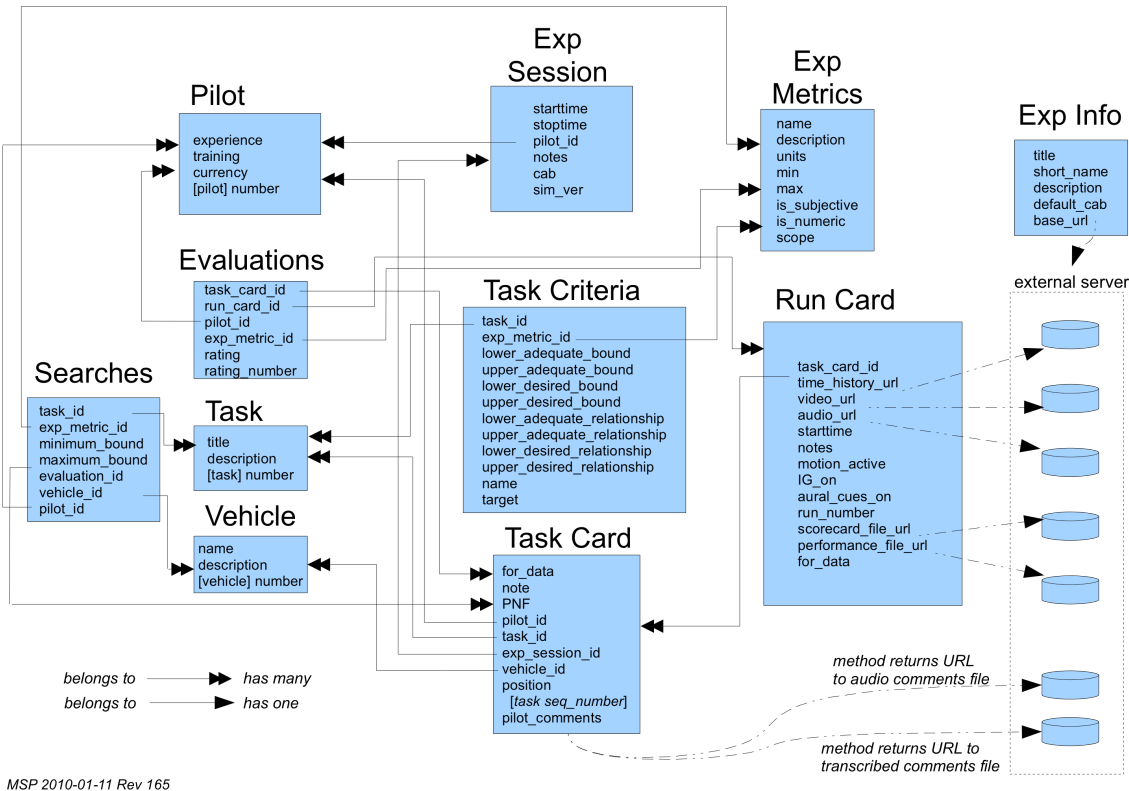
Figure 1. Database schema relationships.

D. Detailed object descriptions

The main objects shown in Fig. 1 are discussed in more detail below:

1. Pilots

Each pilot object contains information regarding her background, training and experience. This anonymized information is usually published in handling quality studies for transparency in the experiment. Each evaluation pilot is assigned an evaluation pilot number, which uniquely identifies that pilot for the experiment.

2. Vehicles/Configurations

Handling qualities experiments normally evaluate more than one vehicle or control configuration, which may include various inceptor or effector characteristics, guidance, control or display variations, or flight conditions. In HOPE, the vehicle table is populated with a description of each of these variations in the experimental test matrix.

3. Tasks

It is not unusual to have the pilot evaluate more than one task, such as takeoff, turns, weapons delivery, target tracking, landing, or for spacecraft, docking. Each different task is assigned a unique number in the HOPE database.

4. Task cards

Each evaluation of a specific task by an individual pilot with a given vehicle configuration that is performed is part of an object called a task card, which may or may not correspond with an equivalent (physical) flight card. Each task card has a unique sequence number for each participating pilot, which usually corresponds with the sequence in which the task card is performed. There are normally multiple "runs" of the given task within a single flight card.

American Institute of Aeronautics and Astronautics

5. Run cards

Each run associated with the experiment has a run card, which contains information about the task, vehicle configuration, and pilot. One or more run cards make up a task card, and each run card will have per-run metrics associated with it through an evaluation record.

6. Evaluations

There are three types of evaluations. First is the pilot summary, which are the overall summary comments or answers to a questionnaire. Second are the per-task evaluations, which are the pilot subjective answers and ratings such as Cooper-Harper ratings. Third are the per-run evaluations, which are the performance metrics for the given task.

Figure 2. Adding a new Pilot.

7. Experimental Metrics

Each experimental metric record represents a single question being answered by the experiment. The metric can be pilot opinion per task or a measured performance of a certain parameter per run. The experimental metric record represents the question (e.g., "touchdown sink rate in ft/sec;" an evaluation record represents the response to the question (e.g. "3.42").

IV. Interaction with HOPE during the experiment

HOPE allows for the recording of experimental data in real-time as the experiment runs. A user with editing rights can log into the database via a web interface. After logging in, the user would follow the steps listed below to add a pilot and her sessions to the experiment database.

1. Adding a New Pilot

The user would select the "Pilots" link from the side menu, which will take him to the list of evaluation pilots. At the bottom of the page the user would click on the "Add pilot" link, which will take him to the page where he should add the information about the evaluation pilot as shown in Fig. 2. This information includes a unique pilot number for identification purposes. When the form is completed, the user should click on the "Create" button at the bottom of the page.

2. Recording a New Session

Next, a new test session (e.g. a test flight, or simulation entry) record will need to be created. There may be more than one session per pilot, depending on the experiment. To mark the beginning of a new session in the aircraft or simulator, the user should select the "Sessions" link from the side menu and then select the "Add a new session" at the bottom of the page. This

Figure 3. Recording a New Session.

takes the user to the "Recording a new session" page as shown in Fig. 3. The user would then record the start time, pilot number, and perhaps simulation software version or other notes. He can edit the session later to record the stop time and add additional notes. When the page is complete, the user should click on the "Create" button at the bottom of the page.

3. Adding a new Task Card

At this point the first task card would probably be evaluated in the experiment. There will be multiple task cards for each session; task cards are added from the Sessions page. The user should select the "Sessions" link from the side menu and then select the "Show" button next to the current session. On this page the user will find a link to "Add a new task card" which will open the page shown in Fig. 4. The user would then complete the form which includes three pull down menus for pilot, task, and vehicle configuration. After completing the page the user would select the "Create" button at the bottom of the page.

4. Adding Run Cards

The user would next select "Task Cards" from the side menu and then select "Show" for the current task card. On this page the user would add the run cards that complete the evaluation of the vehicle configuration (forming a task card). The user would click on "Add a run" link and enter the run information as shown in Fig. 5 and then choose "Create".

5. Adding post-run opinions

Following a complete set of runs to evaluate a particular vehicle configuration, an evaluation pilot is usually asked to provide quantitative ratings in the form of Cooper-Harper ratings[11], task load assessments[12], Likert scale assessments[13], and other quantitative opinions. These should be entered on the associated task card (which is the parent of the associated run cards) as shown in Fig 6.

Figure 4. Adding a new Task Card.

Figure 5. Add a new Run Card.

American Institute of Aeronautics and Astronautics

Figure 6. Task Card with pilot ratings and a transcript of Pilot Comments.

6. Performing post-session data upload and adding transcriptions and other links

A necessary step after each simulation session is to invoke a custom import script to load the summary data from post-processed time histories to populate the per-run metrics table. These are automatically associated with the appropriate run cards through the experiment-unique run number. Another (optional) post-session step involves transcribing any verbal comments that might have been recorded; these are added to the existing task card (e.g. Fig. 6) to back up the pilot's quantitative ratings entered earlier. Finally, links to any external recordings such as audio or video may be added if they are not in a consistent labeling scheme. If a consistent labeling scheme is employed, the HOPE source code may be modified to automatically provide a URL pointing to the external data.

7. Pilot questionnaires

Often the experimental protocol requires the pilot, at the completion a block or at the end of her participation in the experiment, to complete an oral or written questionnaire. The pilot page previously created is edited to capture the answer to these questions, shown in Fig 7.

American Institute of Aeronautics and Astronautics

LaRC-3 DATA BOOK

EP number: 5

Experience: 3600 hrs TT in variety of aircraft over 30 years including 20-30 hrs in helo. IP in T-38, T-2, F-18. 2x NTPS instructor.

Training: NTPS once as student and 2x as IP

Currency: Flew C-17 within two weeks of test. Flew Apache with accel cue on head-up display.

Sessions: Mon, Nov 23 14:46 Tue, Nov 24 08:30

Task cards & runs: There have been 21 task cards and 65 runs performed by this pilot.

Summary questionnaire:

Question	Answer
Were the cockpit displays adequate for the task?	
	Yes.
What influence did the displays have on the task?	
	Seeing through the HWD and the symbology on the head down display (specifically, the FLIR) was disconcerting.
Comment on any desirable (if any) display improvements you'd like to see	
	For the head down displays, swap the position of the ADI and the FLIR display. During the free flight run (at end of day, not collected for data), the roll and yaw needles were confused. For the HWD, the monochrome display made it hard to see different symbols especially when against the moon terrain. The HWD seemed blurry/unfocused both days. The HWD acceleration ball is too sensitive. The HWD ADI is too small to adequate show the ships attitude. Needs more precision.
How adequate were the out-the-window visuals for the task, and how much did you use the outside view?	
	The out the window was used a lot especially for the low control power cases. Would like to have more look down to see more of the landing zone.
Did the training session adequately prepare you for this evaluation?	
	Yes. Liked having split session (afternoon then morning) because it allowed time to think on the training/experience of the afternoon session.
Did you notice any significant learning curve effects after training?	
	No
What could be improved in the training session?	
	none
With respect to different trajectories, what did you consider to be the most significant effect of the different trajectories (glidepath angles) flown?	
	With high control power, no difference in trajectories. For lower control power runs, liked the steeper (30 deg) trajectories.
Any other comments by evaluation pilot	

Edit | Back

Figure 7. Pilot Information with answers to an end-of-test questionnaire.

V. Use of HOPE during analysis of experimental results

A. Statistical measures

At this point, the HOPE application provides limited analysis of data, e.g., statistics (mean, minimum, and maximum), for any collection of task cards. Most number-crunching should be performed on the complete set of run summary data. A link is provided at the beginning of any collection of runs to download a comma-separated values file of that data set to an external application.

B. Search capability

HOPE includes a search capability for combination of metric values, which allows for additional analysis. Figures 8-13 show an example of how HOPE can be used for analysis. Figure 8 shows a search being performed to find all of the task cards where a Cooper-Harper Rating of 9 or 10 was given.

Figure 8. Database Search.

C. Data hyperlinking

A primary benefit of an on-line, hyperlinked databook is the ability to quickly maneuver around the dataset through links. For example in Fig. 9 the result shows seven task cards with a CHR of 9 or 10. It is noticed that tasks 600 and 800 are given 9s and 10s by more than one pilot. However task 900 is given a 9 by only one pilot. The user can look in more detail at the underlying rational for why the pilot assigned a CHR of 9 ("controllability in question") by clicking on the task number (900) to view a page describing task 900 (Fig. 10).

American Institute of Aeronautics and Astronautics

LaRC-3 Data Book

Home
Pilots
Vehicles
Tasks
Metrics
Sessions
Task Cards
Runs

Search

Logout
mproffitt

Search Results

Number of search results found: 7

Pilot:

Task:

Vehicle:

Experimental Metric: CHR

Minimum Bound: 9.0

Maximum Bound: 10.0

Task Cards:

Hover over the task or vehicle/config column for a description of the task or configuration

Session+Seq	Pilot	Task	Vehicle Config	Runs	CHR	Note	Obsrvr	Actions		
Wed, Oct 28 10:00	003	3	800	3	3	10	PIO prone	Jackson	Show	Edit	Destroy
Wed, Oct 28 10:00	010	3	600	3	4	10	One abort and one crashed into habitats. "Got behind with PIO which blossomed."	Jackson	Show	Edit	Destroy
Tue, Dec 15 08:45	018	7	600	3	3	9	small hdot to give time to capture and correct lateral/longitudinal positions	Bailey	Show	Edit	Destroy
Wed, Dec 16 14:26	001	8	900	4	3	9	Below 100 ft, needles don't give desired performance. The large lag made it uncontrollable. Up and away is controllable. Control response and guidance major factors in CHR rating. Control reversal obvious.	Jackson	Show	Edit	Destroy
Wed, Dec 16 14:26	005	8	800	3	3	10	Unrecoverable PIO. Sees jumpiness and unpredictability of the needles. Good video of crash from run 689.	Jackson	Show	Edit	Destroy
Wed, Dec 16 14:26	006	8	600	3	3	10	After final landing sound, cut all sounds generate by software. Can't correct drift rates because of low control power. Displays need symbology to help compensate for the sluggishness.	Jackson	Show	Edit	Destroy
Thu, Dec 17 08:36	015	8	600	4	3	10		Jackson	Show	Edit	Destroy

Pilot rating statistics for these task cards

Rating	N	Min	Max	Avg	Median	Std Dev
CHR	7	9.0	10.0	9.714	10.0	0.452
TLX M	7	90.0	100.0	92.857	90.0	4.518
TLX Ph	7	20.0	100.0	71.429	90.0	32.701
TLX Te	7	80.0	100.0	92.143	90.0	6.468
TLX Pf	7	0.0	50.0	10.000	0.0	16.903
TLX E	7	90.0	100.0	94.286	90.0	4.949
TLX F	7	70.0	90.0	84.286	90.0	7.284
Approach opinion	7	1.0	2.0	1.143	1.0	0.350
RHC opinion	7	1.0	2.0	1.286	1.0	0.452

Figure 9. Search Results.

American Institute of Aeronautics and Astronautics

The user could then click on the underlined "11" in the "11 task cards" phrase to retrieve all of the task cards in which that task was evaluated. The summary of results (Fig. 11) show 11 task cards where task 900 was evaluated, and that 4 different pilots evaluated the task. Three of the pilots gave it a CHR level one or two. Only one pilot gave the task a level 3 rating.

The user can select the sequence number of the second-to-last task card (001) and view the task card, Fig. 12., along with notes and pilot comments and links to audio files. From inspection of the pilot comments, it appears this pilot was still coming up on the learning curve for the task. The sequence number in Fig. 12 confirms that this was the first evaluation performed by pilot 8.

Figure 10. Description of task "900".

Figure 11. All Task Cards for a given task.

LaRC-3 Data Book

Home
Pilots
Vehicles
Tasks
Metrics
Sessions
Task Cards
Runs

Search

Logout
mproffitt

LaRC-3 Task Card 2009-12-16-1426-001

Task: 45 deg approach with constant decel guidance (task 900)

Vehicle Configuration: 4 - Mid-control power Altair

Session: Wed, Dec 16 14:26 with **pilot** 8

Card sequence for this pilot: 1

Observer: Jackson

Notes: Below 100 ft, needles don't give desired performance. The large lag made it uncontrollable. Up and away is controllable. Control response and guidance major factors in CHR rating. Control reversal obvious.

Run cards:

Hover over the task or vehicle/config column for a description of the task or configuration

Run #	Pilot	Time	Task	Obsrvr	Notes	For data	Scores	Actions		
674	8	14:28	900	Jackson	config 2904. Desired performance.	N	--	View	Edit	Destroy
675	8	14:31	900	Jackson	config 3904. Not adequate position .	Y	66	View	Edit	Destroy
676	8	14:34	900	Jackson	config 4904. Not adequate on position.	Y	66	View	Edit	Destroy

Add a run to this task card

Task Ratings: **Comments:** PC (.wav) Mac (.aif) Transcript

Metric	Value
CHR	9
TLX M	90
TLX Ph	90
TLX Te	80
TLX Pf	10
TLX E	90
TLX F	80
Approach opinion	2
RHC opinion	2

Pilot 8, Sequence 1, December 16, 2009

Cooper-Harper Rating:
Okay, this is for card number 1, one practice approach, two attempted landings in the landing site, with the task being landing within certain parameters and the criteria being 15 feet for desired and within 25 feet for adequate performance. The pilot decisions, was it controllable? Yes, barely. Is adequate performance attainable with a tolerable pilot workload? No, deficiencies definitely require improvement. I would rate it a 9, major deficiencies, intense pilot compensation is required to retain control. This is a grade for the overall plant here, the whole system. What I found was up to about 100 feet or so you could make corrections with the needles. Below about 100 feet on the practice I did okay, on the two real ones the needles did not give me the performance I wanted and so I attempted to go manual. But it's very difficult to anticipate when you go manual I shifted my scan to what I call a NAV display and tried to interpret the velocity vector of the little predictor. But that's so far such a tremendous lag that I grossly over controlled it and got translational rates that were uncontrollable as far as stopping them which put me outside of the box. So basically up and away it's a controllable configuration, but the combination of bad guidance from the needles in my opinion one thing trying to follow the guidance with the control laws that are with this particular configuration, with the apparent lags in response and then the ability to over control the configuration. Even though I was fairly within tolerances coming to about 100 feet, I rapidly got outside tolerances inside of 100 feet as the gain went up in the task. So bottom line the very slow responsiveness of the overall system led to what I would call very low frequency PIO but a very high amplitude PIO where I got way, way outside the bounds of being able to bring it onto a spot landing. So the issues to me on this one is not only the control responses but also the guidance and the ability to feel like I was forced to try and manually take over and use symbology on the NAV display which did not work in that particular case because by focusing on the NAV display and not looking at the ADI I was unable to anticipate the rates that were developing.

Figure 12. Task Card detail.

VI. Concluding remarks

Having an on-line, always up-to-date, centrally-served databook is a benefit to a flight test or simulator-based handling qualities research effort. Multiple test conductors can record data into the central repository; it is possible for multiple simultaneous simulations or flight experiments to be conducted and recorded in parallel.

Hyperlinks within the databook allow quick browsing of high-level overviews and drilling down to specific details of a particular run. The analyst can read transcribed (or listen to recorded) pilot commentary and watch the recorded video of a particular run from a single website. Configuration control is enhanced in that the software and hardware configurations of the experimental environment are recorded and all team members view the same, immediately available, data.

At any time during the weeks of an experiment, the latest set of data from the experiment is available for review and analysis. Preliminary data trends can be observed to keep the test matrix within reasonable scope and to ensure adequate parameter boundaries are being tested.

The ability to easily view and navigate through the data in several ways has proven helpful in analysis and documentation of a handling qualities experiment. It is expected that this tool will be made available upon request to other researchers so that they may benefit and improve upon it.

References

[1]Jackson, E. Bruce, Raney, David L., Glaab, Louis J., and Derry, Stephen D., "Piloted Simuation Assessment of a High-Speed Civil Transport Configuration," NASA TP-2002-211441, 2002.

[2]Bailey, Randall E., Jackson, E. Bruce, Goodrich, Kenneth H., Ragsdale, W. Al, Neuhaus, Jason, and Barnes, Jim, "Initial Investigation of Reaction Control System Design on Spacecraft Handling Qualities for Earth Orbit Docking," *J. Guidance, Control and Dynamics*, Vol. 32, No. 6, November-December 2009, pp. 1733-1735.

[3]Jackson, E. Bruce, Goodrich, Kenneth H., Bailey, Randall E., Barnes, Jim, Ragsdale, W. A. and Neuhaus, Jason, "Investigation of Control System an Display Variations on Spacecraft Handling Qualities for Docking with Stationary and Rotating Targets," NASA TM-2010-216194, 2010.

[4]Bilimoria, Karl D., "Effects of Control Power and Guidance Cues on Lunar Lander Handling Qualities," *J. Spacecraft and Rockets*, Vol 46, No 6, November-December 2009, pp. 1266-1271.

[5]Bilimoria, Karl D, Mueller, Eric, and Frost, Chad, "Handling Qualities Evaluation of Pilot Tools for Spacecraft Docking in Earth Orbit," AIAA 2009-5665, AIAA Guidance, Navigation and Control Conference, August 2009

[6]Mueller, Eric, Bilimoria, Karl D., and Frost, Chad, "Dynamic Coupling and Control Response Effects on Spacecraft Handling Qualities During Docking," *J. Spacecraft and Rockets*, Vol 46, No 6, November-December 2009, pp. 1288-1297.

[7]Flanagan, David, and Matsumoto, Yukihiro, "The Ruby Programming Language," O'Reilly & Associates, Sebastapol, CA, 2008.

[8]Ruby, Sam, Thomas, Dave, and Hansson, David Heinemeier, "Agile Web Development with Rails, Third Edition," The Pragmatic Programmers, Raleigh, 2009.

[9]Reenskaug, Trygve, "Models – Views – Controllers," Personal archive of the author (http://heim.ifi.uio.no/~trygver/1979/mvc-2/1979-12-MVC.pdf), December 1979, cited January 2010.

[10]The Apache Software Foundation, "The Apache HTTP Server," http://httpd.apache.org, cited January 2010.

[11]Harper, R P, Jr. and Cooper, G E., "Handling Qualities and Pilot Evaluation," AIAA Journal of Guidance, Control, and Dynamics, Vol. 9, Sept.-Oct. 1986, pp. 515-529.

[12]Hart, S.G. and Staveland, L.E., "Development of NASA-TLX (Task Load Index): results of empirical and theoretical research," Human Mental Workload, North-Holland, Amsterdam, 1988, pp. 139-183

[13]Likert, Rensis, "A Technique for the Measurement of Attitudes," Archives of Psychology, Vol 140, 1932, pp. 1-55.