JOURNAL OF AEROSPACE COMPUTING, INFORMATION, AND COMMUNICATION
Vol. 8, June 2011

Quantifying Pilot Performance During Landing Point Redesignation for System Design

Zarrin K. Chua* and Karen M. Feigh†
Georgia Institute of Technology, Atlanta, Georgia 30332

DOI: 10.2514/1.51712

The requirements for the next generation of lunar landers include being capable of landing in terrain that is both poorly lit and hazardous, increasing the importance and difficulty of an already critical mission task known as landing point redesignation. During this task, the crew interacts with an on-board automated flight system to finalize the touchdown point. This paper presents the results of a study conducted to quantify pilot performance and task completion time during landing point redesignation; the effect of environmental and mission parameters on those measures; and a design space exploration using human performance contours for different definitions of performance. Additionally, this paper presents the analysis of a comparison between a reference automated system and the pilots' site selection performance. The pilots completed the task in an average of 20 s. The number of identifiable terrain markers, landing points of interest, and the pilot's expectation of terrain features had effect sizes too small to be detected in this experiment. Human performance was found to be best when the importance of fuel consumption was weighted less than system safety or proximity to points of interest. In these same conditions, the human pilot was also observed to perform better than a reference automated system. Fuel penalty associated with prolonged decision time remains the largest detriment to human landing point redesignation task performance.

Nomenclature

$D_{(\mathrm{H,POI})}$	Distance from hazards, points of interest
f_w	Fuel score based on time to complete
H_{LA}	Hazard score for the landing area
P_{LA}	Point of interest score for the landing area
R_{LA}	Roughness score for the landing area
S_{LA}	lope score for the landing area
$w_{()}$	Weighting for safety, proximity of interest, fuel consumption

I. Introduction

THE next generation of lunar landers are expected to significantly improve upon the Apollo Lunar Excursion Module (LEM)—they must be capable of global surface accessibility and landing in areas of hazardous, low-lit terrain. To achieve these performance requirements, a significant amount of computational power and automation

---

Received 21 July 2010; accepted for publication 5 February 2011.

* Graduate Research Assistant, Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, AIAA Student Member, zarrin@gatech.edu
† Assistant Professor, Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, AIAA Member.

will be included in the lander system design. Unlike the Apollo missions, where automation and computational power were driven by technological capabilities, modern system designers will need to use alternative metrics to determine what capabilities to provide to the astronauts, i.e. what tasks/functions to automate. Accordingly, lander designers will need to consider human–system collaboration, and mission designers will need to decide how to trade off the robustness of human decision-making capabilities with the need for accurate, precise, and expedited decisions achievable by an automated flight system. The role of automation will likely be decided on a task-by-task basis. Landing point redesignation (LPR) is one such task that would significantly benefit from a joint human–automation approach. This task is defined as an opportunity for on-board crew to refine a final touchdown point within a pre-designated landing zone.

Despite previous investigations [1–4], human–system interaction during the LPR task is still not well characterized, prompting further examination of human behavior in a representative simulated environment. Specifically, the contribution of the crew during the LPR is not well understood and arguments for and against the inclusion of human operations can be made. To address this issue, this study examines human behavior during LPR within a three-dimensional performance envelope. This study utilized a human–subject experiment designed to measure the performance of pilots at the LPR task in terms of task completion time as well as the accuracy of the site selection and to examine the impact of select environmental and mission factors on crew performance. Specifically, this study had three goals:

1. To establish quantitative estimates of time to complete and site selection accuracy for a simulated LPR task using a representative LPR display design;
2. To determine the impact of the number of points of interest, number of identifiable terrain markers, and the similarity of terrain observed to terrain expected on time to complete on site selection accuracy; and
3. To establish performance contours for different definitions of system success for use by system designers.

This paper reviews current research regarding human involvement during LPR and outlines the motivation for detailed, quantitative analysis on human performance during the task. Additionally, this paper provides an overview of the experiment design, defines the LPR human performance envelope, and presents a comparative analysis of human performance to a reference automated landing system. Lastly, the authors discuss the contributing factors to human performance during LPR and possible adjustments for improved behavior.

II. Background

The LPR task is complex, high-risk, and time-critical, as evidenced in previous lunar landings. Known then as "landing point *designation*," Apollo astronauts accomplished this task using a combination of their perception of the landing terrain, and a reticle-etched window known as the landing point designator. The crew viewed the targeted landing site through the landing point designator by using coordinates provided by the flight computer. The reticle markings conveyed the necessary guidance to reach a chosen landing site, wherein the crew manually piloted the vehicle to their intended destination. This task occurred during powered descent initiation (PDI), shortly after high gate (Program 64), approximately at an altitude of 7515 ft. (2291 m), a descent rate of 145.4 fps (44.3 m/s), and an inertial velocity of 514 fps (157 m/s) [5].

While this procedure worked for the Apollo mission, landing point designation was dependent on landing conditions such as the lighting and viewing angles conducive to human vision capabilities. Landing in flat and level terrain was further necessary to avoid taxing the capabilities of the vehicle and to simplify the landing site selection task. These constraints restricted the Apollo mission to the lunar equatorial region, reaching no farther than 26.7° from the equator. The design of the next generation lander should provide global lunar access and eliminate limitations due to landing conditions and terrain features by providing the crew with the support needed to safely land the vehicle. This support is envisioned as having two new features: advanced sensor data and smart landing algorithms. LPR, and its inherent reliance on the crew to finalize the touchdown point, generally requires a non-fuel-optimal trajectory to orient the spacecraft appropriately to view the landing zone and to allow the crew time to make a decision. Prolonged decision-making is costly as there is a high correspondence between time and fuel consumption. The decision between fully automating the LPR task versus allowing the crew to remain in-command is unresolved. Quantitative data are required to better characterize the tradeoff between the range of function allocation choices available.

With the increased expectations for next generation lunar landers, recent studies related to crewed lunar landing have focused on the development of support systems (i.e. displays and landing algorithms) to aid LPR. One such effort is the development of an autonomous flight manager (AFM) by NASA's Autonomous Landing and Hazard Avoidance Technology (ALHAT) team. The AFM is analogous to a flight management system found in commercial aircraft—providing guidance, navigation, and control cues [6], monitoring system health [7], and interacting with the crew, including prompts for supervisory commands [8]. With respect to the LPR task, the AFM serves two purposes: (1) processing raw sensor data into a form comprehensible to the crew and (2) suggesting alternative landing sites to the a priori baseline site using the sensor data provided. This paper has assumed that the primary sensor used by the AFM will be a LIght Detection And Ranging (LIDAR) of equal capability as that employed by the ALHAT project. Until the AFM completes its assessment of the terrain, the only source of terrain information available to the crew is likely to be either a window or a camera view. Unfortunately, the lighting conditions of some desired landing areas, such as the far side of the Moon, may obscure the crew's ability to acquire terrain information unaided. With the additional information provided by the LIDAR in combination with the AFM, the crew can evaluate the alternative landing sites, find a site that satisfies its specified criteria (e.g. safety, required fuel, or nearness to point of interest), and designate the final landing site.

Currently, the LPR task is expected to occur during the Powered Descent Phase. Just prior to LPR, the vehicle performs a pitch-up maneuver, placing the vehicle in an orientation suitable for LIDAR sensor operation [6]. This maneuver is expected to occur at approximately 1 km in altitude, at a velocity of 100 m/s (nominal trajectory) [9]. The time required to complete the LPR task impacts fuel consumption both directly (e.g. every second spent deciding where to land requires an additional second of fuel) and indirectly as it requires an appropriate descent trajectory. As with the Apollo missions, the landing trajectory will need to be designed to provide the crew a visual of the landing area (either via a window, camera, or LIDAR). Unfortunately, these trajectories are unlikely to be fuel optimal. Preliminary analyses estimate 30s are needed for LPR task completion [1]. In this short period of time, the astronauts must absorb terrain information, perform tradeoffs of safety and fuel consumption, and select a site. Likewise, the crew must adapt to any unanticipated terrain features or events.

To support the crew's expeditious and accurate completion of the LPR task, previous studies [1–4] have studied the effect of automation on crew performance, developed assisted decision-making algorithms, and identified potential task bottlenecks. Forest et al. developed a landing site algorithm that suggests alternative sites based on the crew's objective function preference. The study also designed a candidate LPR display design, but did not model human interaction with such a system [2]. Needham investigated the impact of varying levels of automation on human performance during LPR, concluding that higher levels of automation allowed for shorter completion time [3]. In addition, Needham developed a set of icons that could overlay landing site terrain characteristics on a top-down sensor scan map. Lastly, Chua and Major derived a task model and used this model to identify likely LPR bottlenecks [4]. These bottlenecks were addressed by redesigning the LPR display to simplify the information layout and to utilize new symbolism to represent site characteristics. One aspect of the redesigned display is the incorporation of a series of buttons, called hot keys, which allow the crew to choose a predesignated weighting distribution for the objective function. The hot keys replaced previously suggested slider mechanisms and selecting one returns the top three sites matching the criteria specified. Five hot keys corresponding to safety priority, fuel efficiency priority, proximity to POIs priority, balanced (between the first three hot keys), and an a priori baseline were used in this experiment. Chua's LPR task model also incorporates expert decision-making theory [10] to account for specialized astronaut behavior [1]. However, the estimates of the time required for LPR task completion derived from this task model have yet to be validated. Measures such as task completion time and strategy are available from the Apollo missions, but their applicability to LPR on next generation lunar landers is limited. Apollo operated under levels of autonomy well below what is generally considered acceptable today. Direct comparison with Apollo—while possible—is not appropriate here, because there is not a common basis for comparison.

III. Methods

A. Objective

The main purpose of this study is to characterize human behavior during LPR across three dimensions of performance: safety, proximity to POI, and fuel consumption. These three dimensions provide the boundaries for an overall performance envelope that can be used to quantify the impact of crew input during the LPR task on system performance. Defining this envelope requires quantitative human performance data specific to LPR, data which currently exist in limited and disparate forms [3,4]. The study's three goals: first, to provide quantitative estimates of task completion time and site selection accuracy during LPR; second, to measure the impact of environmental and mission parameters on task completion time and site selection accuracy; and third, to outline performance contours for varying definitions of system performance success across the selected performance dimensions are based on the results of a human-in-the-loop experiment. The following sections detail the experimental design methodology that drives the first and second goals.

B. Experimental Design

Crew performance and decision-making are likely influenced by several factors. Previous lunar landings, analogous studies, and cognitive theory illuminated three likely influences on crew performance: the number of *Identifiable Terrain Markers* (ITMs) [11], or formations of hazards; the number of *Points of Interest* (POIs) [11, 12] or geographical targets that drive the purpose of the mission; and *terrain expectancy* (ε) [10], or differences between terrain presented during the crew mission briefing and actual terrain encountered. These three influences were used as within-subject independent variables. Based on consultation with subject matter experts, the following levels were examined: the number of ITM (1, 2, 3, and 4), the number of POI (1 and 2), and terrain expectancy (expected, unexpected).

Two measurable dependent variables were chosen to assess pilot performance: time to complete the task and a parameterized pilot performance. *Time to complete* (T_c) is the time in seconds from the beginning of the LPR task, when the annotated terrain image is first available to the pilot, to when the pilot selects his or her landing site. *Pilot Performance Score* (PPS) represents the pilot's ability to perform the LPR task given the options available in a given scenario. PPS is a weighted combination of the three dimensions: safety, proximity to POI, and fuel consumption, and is calculated using Eq. (1). Additionally, two covariates were measured but not controlled for: *flight experience*, defined by the number of flight hours, and *top pilot certification* (TPC), visual flight rating (VFR), instrumental flight rating (IFR), commercial pilot license (CPL)), which represents the most advanced flight rating owned by the pilot.

A $4 \times 2 \times 2$ mixed design of experiments was used to represent the 16 possible landing scenarios. The number of ITMs is a blocked, between-subjects variable based on a low/high scheme—participants viewed maps with either 1, 3 or 2, 4 ITMs. POI and terrain expectancy are within-subjects variables. Thus, pilots experienced only 8 of the 16 scenarios. The order of the runs within- and between-subjects was balanced to reduce any potential bias in run order.

C. Participants

Twenty pilots participated in this experiment, representing a wide variety of flight experience and pilot training. Participation in the experiment was open to individuals holding a private pilot license (PPL), a CPL, or a military pilot rating. At least 80 h of flying experience in either fixed-wing aircraft or helicopters was required. This stipulation ensured enough familiarity with standard aircraft landing procedures and the process of selecting a landing site while ensuring adequate participation. Because of the unavailability of astronauts, single-engine, fixed-wing pilots were deemed sufficient for this preliminary study due to their familiarity with the in-flight emergency known as *engine out*, an event similar to LPR in that the mental task of selecting a suitable landing site that balances safety and achievability are required. *Engine out* occurs when a single-engine aircraft experiences an engine failure. Helicopter *auto-rotations* present similar mental tasks with even greater time constraints. At this time, the pilot must make an emergency landing [12], which involves stabilizing the aircraft and finding a suitable place to land. The pilot must quickly select and maneuver the aircraft to a suitable landing site, i.e. one that consists of level and relatively smooth terrain. His available decision-making time is limited and he must ensure that he can maneuver to the desired landing location. While safety is an immediate key criterion for a landing site, the pilot must also be mindful of the plane's potential energy (i.e. altitude) and possible landing sites. These attributes translate directly to cockpit behavior by monitoring remaining fuel resources and recognizing possible sources of ground assistance. A demonstration of capability in assessing and reacting to engine out is a mandatory aspect of obtaining a PPL.

Twelve pilots held a VFR rating, seven held an additional IFR rating, and one pilot held a CPL. The pilots have flown single- and multi-engine aircraft both for personal and for commercial use. The mean was 277 h of flying ($\sigma = 307$ h). No military pilots participated, and only one pilot had experience flying helicopters. The pilots were randomly separated into two groups for ITM frequency blocking: 8 participants (six VFR, two IFR) in the ITM(2,4) group and 12 (six VFR, five IFR, one CPL) participants in the ITM(1,3) group. Figure 1 illustrates the distribution of the hours of flight experience. The flight experience mean was 181 h in the (2,4) group ($\sigma_{(2,4)} = 140.5$ h) and 340 h in the (1,3) group ($\sigma_{(1,3)} = 373.1$ h). The discrepancy in flight hours and inequality of VFR/IFR pilots was

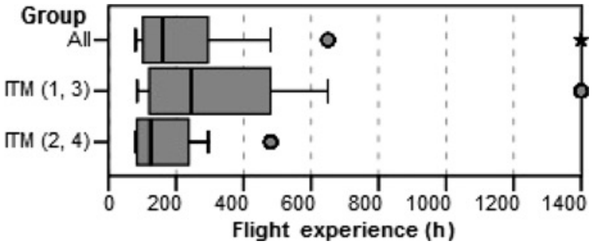

Fig. 1 Distribution of flight experience (h).

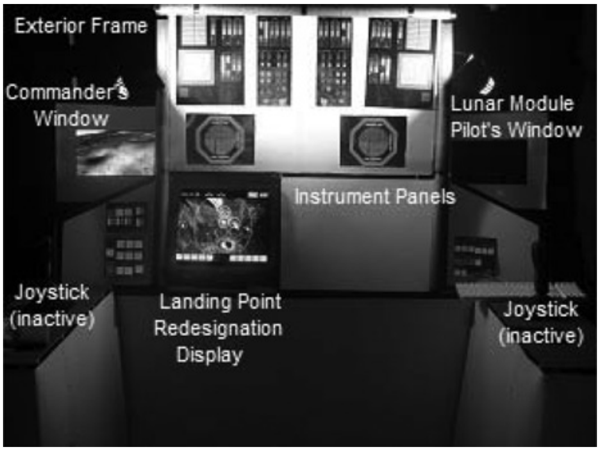

Fig. 2 Full mock lunar lander view.

not determined until after the experiment and was heavily influenced by a single pilot with over 1000 flight hours. As such, flight hours and pilot certification are included as covariates in the data analysis. Likewise, the experiment began with an equal distribution of pilots for both groups, but two pilots were unable to achieve proficiency in the LPR task after the allotted training time leading to exclusion of their data in the analysis.

D. Apparatus

A lunar lander simulator including a physical mock up and LPR reference displays and AFM guidance algorithms was created, as seen in Fig. 2. The overall goal was to create an environment that sufficiently emulates the true scenario experienced by the crew during LPR. In particular, the software simulation is driven to model the LPR reference display, AFM guidance algorithm, and test the experiment hypotheses; the hardware simulation is geared toward exhibiting the realism of next-generation lunar landers.

The physical structure of the lander simulator consists of an exterior black cloth-covered PVC frame and an interior cockpit support structure. The exterior frame mimics the encapsulated nature of lunar landers and focuses the user's attention to the landing displays. The interior frame supports pictorial instrument panels similar to the Apollo LEM and static images of primary flight displays (PFDs) illustrated vis-à-vis Multi-function Electronic Display System [13] (i.e. "glass cockpit"). The PFDs were kept static to prevent pilot distraction and avoid confounding decision-making time; the pilots were told that the AFM was flying the lander to permit full pilot concentration on the LPR task. Additionally, the interior frame supports three computer monitors (one LPR display, two simulated windows) and additional static instrument panels. Two joysticks are used to mimic the role of commander and lunar module pilot. As this experiment did not focus on manually controlling a lunar lander, the joystick input capabilities were disabled. Comparisons to the projected interior of Altair, a conceptual lunar lander design, confirmed the validity of this experimental setup [14]. Figure 2 illustrates the final form of this mock lunar lander.

The software developed for this experiment is broken into two major components: an out-the-window perspective to further the lunar lander mockup hardware and a Pseudo-AFM LPR algorithM (PALM) that emulates the AFM algorithm of alternative landing site selection [2]. The LPR display was based on the design proposed by Chua et al. [1]. This display design was not chosen based on superiority over other candidate designs. Rather, this design provided a publicly available symbology template for use in this investigation.

Simulating a realistic lunar descent required presenting the user an out-the-window perspective. Ideally, this perspective would offer dynamic and unique lunar terrain to mimic lunar vehicle traversal. Although environment simulation software exists, as demonstrated by the variety of COTS and government-sponsored software (Orbiter [15], EDGE [16], X-Plane [17], EagleLander3D [18], the current state of art still lacks the full breadth of lunar terrain accessibility necessary to uniquely pair window views to final landing site areas. This experiment used the EagleLander3D Apollo flight simulator software to record out-the-window perspectives from three stages of flight: vehicle pitch-up, approach, and terminal descent. The length of video record is fixed to 30, 45, 15 s, respectively. The videos are played in accordance with the pilot's actions—once the LPR task is completed, the approach video is interrupted to show terminal descent sequence, effectively signaling the end of the LPR task. Despite the low-fidelity nature of this aspect of the simulation, the pilots found the out-the-window display added to the realism of the simulation.

As previously discussed, the lander is expected to be equipped with an AFM that offers alternative landing sites based on an objective function that is set by the crew using the hot keys. The PALM developed for this investigation takes an input package (lunar satellite photography map, hazard identification, POI location), scans the map for alternative landing sites, and outputs sites based on each of the following objective functions [1]: safety, fuel efficiency, proximity to POIs, balanced (equal weighting of the first three), and a priori (weighting based on data available at launch). The landing site selection algorithm is modified from the work of Cohanim and Collins [19]. The output map image for each objective function contains hazardous area highlights, the POI(s), the baseline point and three alternative sites, and symbols for the relative goodness of slope and roughness of the expected landing area. These maps are generated prior to the experiment, and the map display corresponds to the actions of the user simulating a real-time calculation of alternative landing sites without the computational cost or increased risk of simulation failure.

The PALM also computes the PPS. The PPS is calculated using Eq. (1). Although the mission objective is to place the lander in an area free of major hazards and of preferable terrain characteristics (flat and level terrain), there exists a region within each metric that constitutes as sufficient performance. This region must be factored into the performance formula to account for more realistic figures of merit.

$$P_{\text{score}} = w_{\text{f}} f_{\text{w}} F_{\text{LA}} + w_{\text{p}} D_{\text{POI}} + w_{\text{s}}(S_{\text{LA}} + R_{\text{LA}} + D_{\text{h}})/3 \tag{1}$$

$$f_{\text{w}} = \frac{45\,\text{s} - \text{time to complete}}{45\,\text{s}} \qquad f_{\text{w}} \in [0, 1]$$

$$P_{\text{LA}} = \begin{cases} 0 & \text{if } D_{\text{POI}} < 1.5\text{ m} \\ \dfrac{20\text{ m}}{D_{\text{POI}}} & \text{if } D_{\text{POI}} > 20\text{ m} \\ 1 & \text{if } 1.5\text{ m} \leqslant D_{\text{POI}} \leqslant 20\text{ m} \end{cases}$$

$$H_{\text{LA}} = \begin{cases} 0 & \text{if } D_{\text{H}} < 0\text{ m} \\ \dfrac{D_{\text{H}}}{30\text{ m}} & \text{if } 0\text{ m} < D_{\text{H}} < 30\text{ m} \\ 1 & \text{if } D_{\text{H}} \geqslant 30\text{ m} \end{cases}$$

$$S_{\text{LA}}, R_{\text{LA}} \in [0, 1]$$

where D_{POI} and D_{H} are distances from the POIs and hazards, respectively; S_{LA}, R_{LA}, H_{LA}, P_{LA} are the slope, roughness, hazards, and POI scores, respectively, of each landing area. These scores are absolute values—no perfect sites were included in this experiment and therefore a maximum score was not achievable. For purposes of this experiment, all weights (w_{f}, w_{s}, w_{p}) were set equal to 0.33. Following calculation of the PPS, all of the possible landing sites can be ranked. This pilot performance ranking is more easily generalized across all scenarios as compared to PPS and is used in limited cases. All analyses involving main effect modeling use PPS.

E. Procedure

Each testing session lasted 2 h. The initial briefing introduced the LPR task, the simulator, and performance considerations that were explicitly included in a pilot's score for this experiment (i.e. fuel consumption, safety, and proximity to POI). The pilots practiced the LPR task for 45 min in the simulator, where they received feedback on their performance after each run. This exercise allowed the pilots to become comfortable with the simulator. The pilots' proficiency with the simulator was evaluated at the end of the training session. Pilots passing the evaluation were permitted to continue with the experiment. The testing session was comprised of eight runs of preselected specified landing scenarios. Prior to each scenario, the pilot was briefed on the specifics of the scenario using two satellite images of the expected landing area, the location of the one or two POI (with a note that there is no preference between the two sites). Each run included a different scientific mission objective. These science objectives were written to minimize possible bias on task strategy. The pilots were also informed that the lander was equipped with a rover, but the capabilities of the rover were limited.

IV. Results

This section presents the results of the quantitative pilot performance measures during LPR and an analysis of human performance under the full breadth of mission success definitions. Except in the instance of time to complete, interactions between main effects were not found to be statistically significant. Unless otherwise stated, the statistical analysis was performed using both parametric (repeated-measures ANOVA, linear regression) and non-parametric (Spearman's correlation (ρ), Kendall Tau correlation (τ), or Friedman's test) as appropriate. Significance for all tests was set at $\alpha = 0.05$. Unless otherwise noted, flight experience and TPC are included as covariates in the repeated-measures ANOVA.

A. Estimates of LPR Task Completion Time and Pilot Performance

The pilots completed the LPR task in less than 42 s, and no pilot aborted a run. On average, the LPR task was completed in 20.39 s ($\sigma = 9.05$ s), with minimum and maximum times of 4.08 and 41.53 s, respectively. In 54% of the cases, the pilot chose one of the top three sites by selection score, whereas 7% of the cases resulted in pilot selecting the poorest three sites by selection score. While these poorer site selections resulted in feasible landing locations, better sites were available at the selection time. Figure 3 illustrates the distribution of TPC by final PPS.

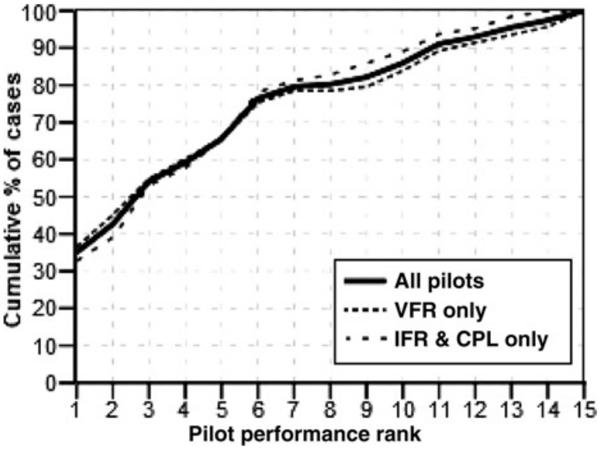

Fig. 3 Distribution of final PPS.

B. Impact of Environment and Mission Parameters on LPR Completion Time and Pilot Performance

The LPR task completion time is a critical value, impacting elements of lunar landing such as the fuel consumption and the design of the descent trajectory. The success of a mission is heavily dependent on the pilot performance during this task. Understanding the major influences on both measures is important for characterizing human-driven LPR. In this experiment, there were two major hypotheses regarding task completion time and pilot performance. First, as the number of ITMs and POIs increase, and the terrain is unexpected, the crew will require more time to select a landing site. A repeated-measures ANOVA was utilized to test these hypotheses. POI was not a significant effect on T_{c} for either the ITM(1,3) or the ITM(2,4) groups. Similarly, ε did not significantly affect T_{c} for either ITM(1,3) or ITM(2,4). The number of ITMs does not significantly affect T_c for ITM(1,3) or ITM(2,4). The interaction between ε and ITMs had significant effect on T_{c} for the ITM(2,4) group only, $F(1,3) = 10.349$, $P = 0.049$. The pilots in this ITM group performed the LPR task slower as the number of ITMs increased in cases of unexpected terrain. The covariates were not significantly related to any of the main effects for either ITM group.

Second, the PPS was hypothesized to become worse with an increase in the number of POIs and ITMs, and when the terrain was unexpected. The effect of POI on PPS was not significant for either ITM group. Similarly, the effect of terrain expectancy on PPS was not significant for either ITM group. The effect of the number of ITMs on POI is marginally significant for the ITM(1,3) group, $F(1, 6) = 5.926$, $P = 0.051$, but is not significant for the ITM(2,4) group, $F(1, 3) = 0.265$, $P = 0.642$. Thus, none of the independent variables included in this investigation were measured to have significant effects on PPS.

The covariates were also evaluated for possible correlation with the task completion time and site selection accuracy. Flight experience had a significant effect on time to complete, $F(15, 134) = 8.692$, $P = 1 \times 10^{-13}$. As expected, flight experience is inversely related to time to complete, $r = 0.260$, $P < 0.020$, meaning that pilots with more flight hours completed the task more quickly than pilots with less experience. Flight experience and TPC were not found to have a significant relationship with site selection accuracy. The authors acknowledge that other reasons could have caused this distribution in the range of time to complete. To rule out this possibility, analyses of run order and blocking variable were conducted and did not determine any statistical differences. Likely, this distribution is due to individual differences.

C. Investigation of the Human Performance Envelope during LPR

Given the framework of the simulated task, the experimental data were extrapolated to provide initial insight on human capabilities during a simulated LPR task. Specifically, the data collected in this experiment were used to speculate about the likely capabilities of an astronaut crew and the conditions under which system designers might choose crew control over automatic control or vice versa.

The pilots' site selections and completion times measured here are assumed to be a conservative, but representative sample of astronaut behavior, or human control, during lunar landing. This assumption is best analogized to a Monte Carlo analysis, where the pilots' behaviors in the experiment are assumed to be the behavior of one astronaut; the differences in landing scenarios and pilot experience is similar to the uncertainties associated with the inputs to a system; and the landing site selection and completion time are the result of performing a lunar landing under the prescribed inputs. As such, the actual site selection data are a sampling of the design space between the experiment independent variables and covariates (POI, ITM, ε, TPC, flight experience) and dependents (site selection, completion time). For purposes of this analysis, these site selections are further assumed to be independent of time (i.e. the same site selection distribution is achievable regardless of the amount of time available). However, the site selection itself means little without assigning some definition of quality.

To account for the variability in weighting distribution on the measures of interest, the PPS for the 56 cases was determined using Eq. (1). Starting from a weight of 0, one weight w_i was incremented by 0.1 while the other two weights were equivalent and equal to $0.5(1 - w_i)$. This schema results in 33 unique weighting combinations, with combinations such as fuel consumption dominated (w_{f}, w_{p}, $w_{\mathrm{s}} = 0.9, 0.05, 0.05$) and non-fuel consumption dominated (w_{f}, w_{p}, $w_{\mathrm{s}} = 0.1, 0.45, 0.45$). Each line in Fig. 4 represents the change in PPS with respect to LPR task time to complete. As expected, the score decreases as completion time grows large—this trend is indicative of the growing fuel penalty, i.e. the longer it takes to make the decision, the more fuel is consumed. However, depending on the operationalization of performance, i.e. one combination of metric weights, the slope magnitude varies. The variable impact of fuel penalty is best illustrated in the leftmost contour of Fig. 4; the plot focused strictly on the

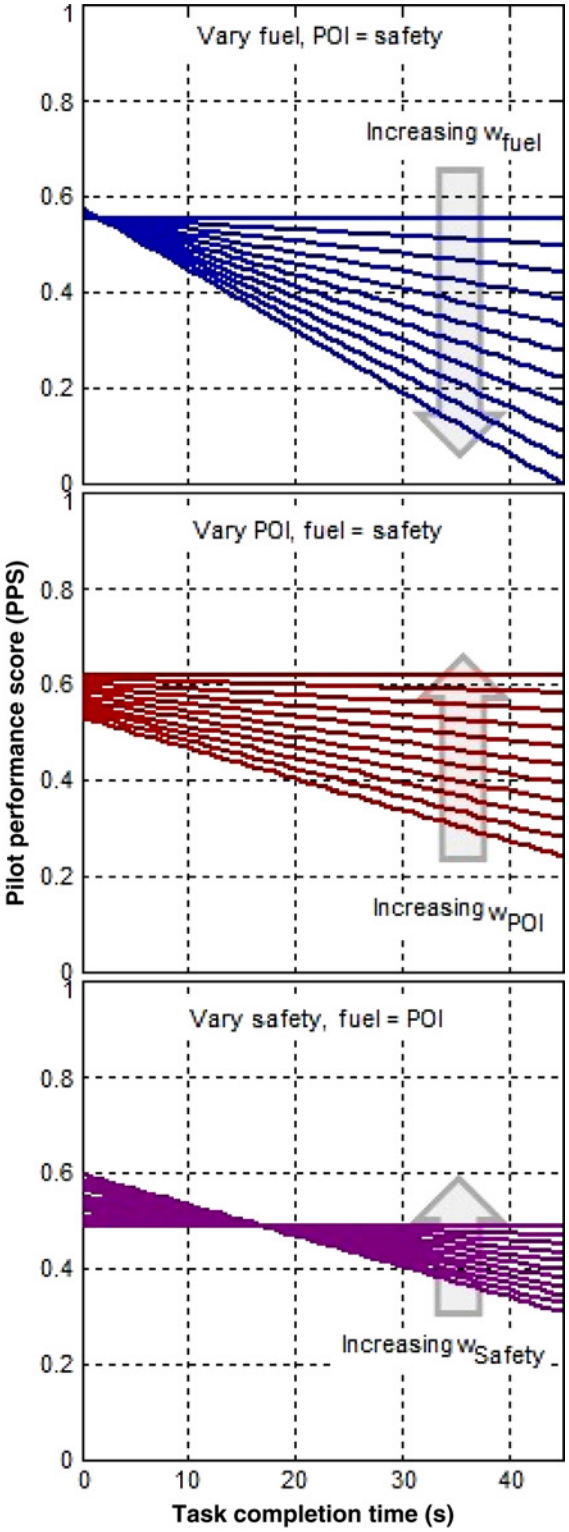

Fig. 4 Theoretical PPS under variable weighting combinations.

impact of fuel consumption criterion. Here the top line represents a w_f of 0 (where performance is operationalized as independent of fuel burn), the contour on the plot is constant, a straight line. The safety and proximity of POI score of each site is independent of time. However, at a w_f of 1 (a scenario where fuel is of singular importance), the PPS is strictly a function of time and subsequently results in a score of 0 if the maximum scenario time is used. Similar patterns are noticeable in the plots of proximity to POI and safety, the center and bottom plots, respectively.

Figure 4 provides valuable insight on the range of theoretical human performance. No PPS exceeds 0.7, even in instances of negligible fuel weight. This score is not necessarily indicative of poor performance, as PPS is an absolute score and perfect sites were not introduced in this experiment. The best PPS varied by scenario. However, the pilots were capable of receiving no scores for PPS (absolute failure) in the unlikely scenario that fuel is the only metric of interest, with no regard to safety or proximity to POI. The effect of POI illustrates that the pilots are stronger at finding sites near the POIs than finding sites where safety is the only concern. Interestingly, the pilot's site selection reflects the tendency to select more fuel conservative sites near the POI. The bottom plot, which highlights the impact of safety, provides a visualization of this decision.

D. Comparison to a Reference Automated System

This generalization of astronaut crew capabilities was used to provide preliminary quantitative reasoning regarding the assignment of human control during LPR instead of an automatic landing system. As previously noted, human-directed LPR invokes an increase in fuel consumption. The fuel consumption requirement associated with an automated system can be significantly less due to both faster decision time and more optimal trajectory, thus increasing the overall performance score. The results of this study show that private pilots require 12–28 s to absorb the information and make the LPR decisions. However, an automated system may not have accurate information, be capable of adapting to scenarios differing from its reference design, or may experience failure. Therefore, a comparative analysis is useful in analyzing how different operationalizations of performance might provide useful insight to system designers as they choose between human control and automated control.

This analysis of the tradeoff between human and automated system control during the LPR task requires three assumptions. First, an assumption is made that an automated landing system capable of autonomously guiding the lander to any point on the lunar surface is available. Secondly, to ensure an equitable comparison, it is assumed that the same trajectory is followed as used in this experiment, which is notionally based on the ALHAT trajectory [9]. The comparison of fuel consumption is thereby simplified to task completion time. While it may seem that this assumption is not likely to hold, as a fully automated vehicle would likely fly a more fuel optimal trajectory which does not permit a clear view of the landing site, there are compelling reasons for choosing such a trajectory. For robustness, it is likely that the automation may be designed to complete the LPR task in case of crew incapacitation or that the crew may serve as backup to automated systems. Either case will require the use of a human-centric trajectory. Additionally, using this reference trajectory is useful for this level of comparison as the actual trajectory planned is still under development. The final assumption assumes that the choice of landing site is static for the automation and corresponds to the a priori site used in the experiment, which had a mid-ranking between 5 and 10 (out of a possible 15), providing the pilots an equal chance of choosing a site which with a higher or lower score. Fifty-six out of 157 runs fall into this category and were used for this analysis. Automated control therefore implies selection of the baseline site, completed in 0 s. This assumption is the most controversial. It represents the case where the site is selected prior to launch and no additional information is included in the final landing point decision. Although this basis for comparison may at seem first to bias the comparison away from the automation, it represents a reasonable basis for comparison as any other basis would be confounded with a seemingly endless array of possible automation capabilities which are neither available at this point, nor proven in their effectiveness. The same definition of performance quality is used to compare human control and automated control. This analysis uses the same schema as the one used for examining human performance, again providing 33 weighting combinations of the three metrics of interest, fuel consumption, proximity to POI, and safety. A 10% uncertainty is applied to the comparison of PPS and automated performance score (APS)—if the PPS is within 10% of the APS score, then the PPS score is considered equivalent.

The results of this analysis are summarized in Fig. 5. Using these definitions and assumptions, the experiment results show that the pilots performed better than or equal to the automated system in 18% and worse than a basic automatic landing system in 82% of the selected cases. These results are heavily influenced by decision time

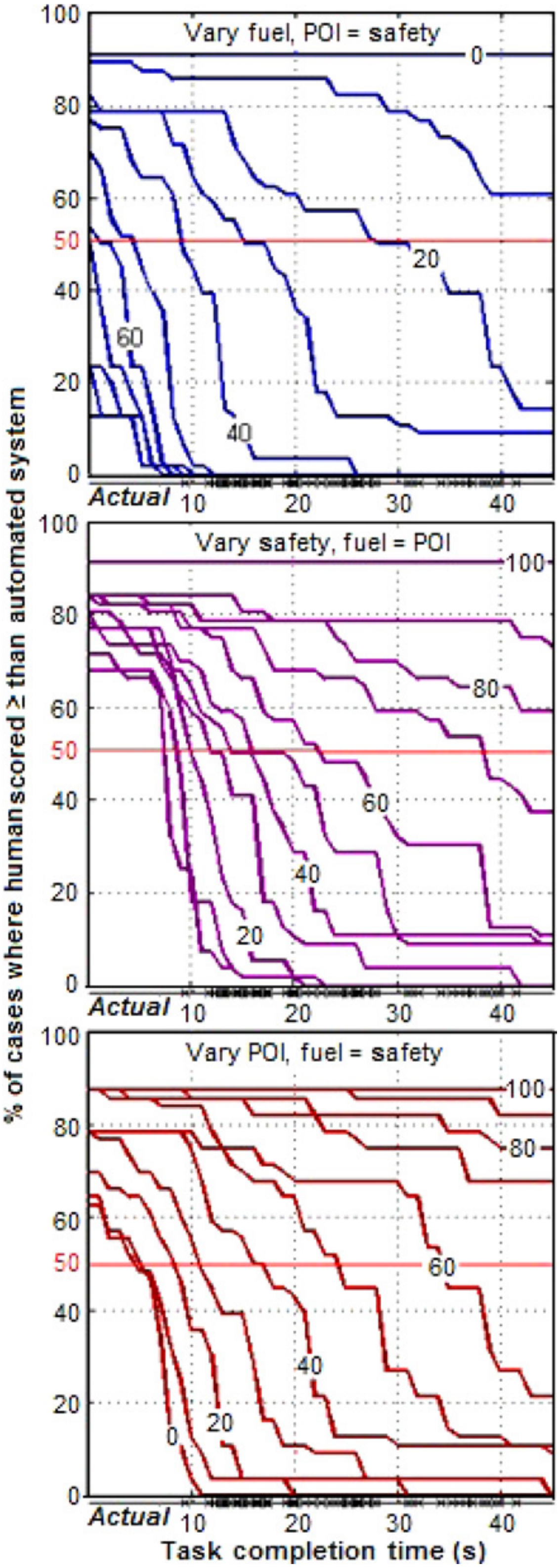

Fig. 5 Comparison of automated and human performance scores under variable definitions of mission success.

and the associated fuel penalty. However, the results of this study found no correlation between performance and decision time.

The most distinctive feature of these contours in Fig. 5 is that no definition exists where the human always outperforms the automated system. The maximum level (91%) of human performance occurs when safety is weighted 100% and fuel is completely ignored. A similar situation occurs when proximity to POI is the primary objective function. At this definition, the human pilot has selected a site closer to the POI in 88% of the cases. The human pilot fares well when both POI proximity and safety are equally critical and fuel consumption is of no importance, selecting better sites than the automated landing system in 91% of all cases. The same general pattern and relationship with fuel consumption is seen in all three graphs of Fig. 5. As task completion time increases, the fuel penalty grows proportionally until a point where the delay in human pilot decisions reduce the performance to below that achievable with a basic automated landing system. The contours in Fig. 5 illustrate that the more heavily fuel consumption is weighed ($w_{\mathrm{f}} > 0.5$), the stronger the case against including human involvement in LPR as the human pilot cannot generally make decisions fast enough to outscore the automated landing system.

In many ways, these contours are analogous to cumulative distribution functions (CDFs). If the 56 cases performed in this experiment are representative of typical site selection choices during LPR, then these contours can assist in determining the conditions which support or detract from human control. This graph provides three types of information to the mission design: the appropriate amount of human decision time, usable performance definitions, and expected human performance compared to the automated system. For example, if the lunar vehicle was constrained to hold enough fuel for 20 s of LPR, then the mission designer can compare human and automated performance for different combinations of weighting distributions. These contours show that if the primary driving factors are safety, proximity to POI, or both, at weighting distributions of at least $\geqslant$0.7, 0.6, or 0.45 each, respectively, then the human is likely to choose a better landing site than one chosen a priori in at least 60% of the cases. These contours are useful in providing initial estimations of region of human pilot performance inclusion to the LPR task despite their limitations. Specifically, the authors acknowledge that the modeled human performance may not be accurately representative of actual astronaut behavior and that an automated lander would most likely not operate under the same conditions as a crewed lander. Additionally, the automatic landing system was assumed to have 100% reliability. These contours can be made more accurate by using trained astronauts and accurate unmanned trajectories.

V. Discussion

Data collected in this study contribute to a maturing assemblage of data regarding the performance of humans during an LPR task. Overall, the experiment results illustrate that in a simulated environment humans can complete the LPR task in less than 42 s and generally select the top-ranked sites when available. This task may be completed as quickly as 4 s, but results indicate that pilots are likely to finish the task within 12–28 s. This decision-making process requires additional fuel consumption, which can evoke substantial performance penalties. However, the significance of this penalty is dependent on the assigned mission criteria. Regardless of the criteria used, humans generally select safe sites and sites near the POI(s).

The study results also showed that no significant effects from the independent variables investigated here (POI, terrain expectancy, and ITM) were found to affect the quantitatively measured dependent variables (time to complete, PPS). The absence of significant effects has several plausible explanations, the most likely of which is lack of statistical power. More samples would have granted a stronger capability to measure small changes in the dependent measures. Additionally, the lack of results could be attributed to a poor correlation between the performance measurement used in this evaluation and true performance. As no consensus yet exists as to the optimal objective function to use when evaluating LPR performance, the formulation used for this evaluation was constructed with the help of subject matter experts. The authors acknowledge that changes to this formulation may affect the outcome of this experimental data analysis, as seen in the data extrapolation investigation. Furthermore, the use of non-astronaut participants may have skewed the results. An argument could be made that aircraft flight experience is not analogous to a lunar landing simulation, and thus, the participants lacked the experience with the LPR task necessary to more closely resembling the behavior of experts. Expert pilots and astronauts are more likely to have better performance and the results of this study therefore mark a conservative estimate on LPR performance. Finally, limited scenario briefing—minutes compared to months of actual training—in combination with visuals used to simulate terrain elements (i.e. number of hazards, expectancy) may not have been sufficiently contrastive to cause an appreciable time delay.

The data collected in this study were used to examine several theoretical situations in order to begin quantifying LPR performance under human control. The results presented here provide mission designers the data needed for conceptual design as well as aid in determining the human pilots' role during LPR. From this analysis, it is clear that the human pilot is capable of finding landing sites that are congenial to vehicle safety and mission success. However, given any realistic mission scenario, this landing site decision-making process must occur quickly, otherwise an extensive fuel consumption penalty is invoked.

There are several possible strategies mission designers can employ to reach maximized LPR task performance. First, the reference automated system used in this analysis assumed the system that did not receive any real-time data. The baseline point is formulated on a priori, pre-launch mission data. Therefore, the trends presented would shift significantly if a real-time automated decision-making algorithm was used. Implementing an improved algorithm, however, may introduce complexity and additional costs, such processing time and memory requirements. Second, astronauts could be trained to complete the LPR task in a fixed period of time. This method has been used successfully during the Apollo missions and will most likely be employed on future missions. While this study illustrated no significant correlation between time to complete and site selection (with respect to safety and proximity to POI only), quicker decisions will minimize the impact of the fuel consumption penalty. Consequently, the possibility exists that through training and personnel selection, better performance could be achieved in shorter time periods. Thus, a more useful solution is to determine the time period necessary for well-trained individuals to make an LPR decision and to train astronauts under those time constraints to determine their likely performance. Lastly, the specific role of the astronaut may need to be adjusted. This experiment scenario was designed such that the pilot would be responsible for evaluating a number of automation-suggested landing sites and making a final decision. The pilot was also told there was no "fail-safe" mode—the absence of a decision would result in a mission abort. A different level of responsibility may result in improved LPR performance. For example, improved performance may occur if multiple redesignation opportunities are available throughout the landing sequence.

VI. Conclusion

The LPR task permits on-board crew to evaluate and select alternative sites prior to terminal descent and touchdown. During this task, the crew must balance their safety and the goals of the mission without violating fuel constraints. An experiment was performed to (1) obtain quantitative estimations of performance and completion time; (2) evaluate variables (terrain expectancy, number of points of interest, and number of identifiable terrain markers) believed likely to impact task completion time and pilot performance; and (3) characterize human performance during LPR for different definitions of mission success. This experiment demonstrated that the LPR task could be completed accurately in a time frame of 42 s or less. The effect of the independent variables on the dependent variables was too small to be detectable given the statistical power of the experiment.

Based on the results of this experiment, humans tend to select better landing sites than a reference automated system when safety and proximity to points of interest are the most critical criteria. However, the decision-making time required for humans incurs significant fuel consumption costs. Thus, in landing scenarios when reserving fuel is of greater priority, mission designers may opt to limit human control during landing site selection or require humans to demonstrate the ability to decide in a very short period of time. Assuming a balanced weighting between performance dimensions, this study demonstrated that human pilots were able to match the performance of a perfect automated system for more than half of the examined cases. Adjustments to astronaut training or improvements to on-board decision-making aids would enhance the synthesized site-selection performance.

The push for increased manned presence to difficult space environments prompts additional capabilities beyond previous and current vehicles. Performing crewed spaceflight in time-critical, complex mission scenarios in lunar, asteroid, and low earth orbit environments will likely set the standard of vehicle performance for other landing targets, such as Mars. However, the vehicle capability required for crewed Mars missions will be more difficult to achieve due to the current limitations in entry, descent, and landing technology for heavy payloads [20]. Until these technological limitations are addressed, detailed analyses of crewed landing vehicle performance will be limited to feasible landing scenarios.

Acknowledgments

The authors thank James Tolbert and Vidhur Vohra for their assistance in developing the experiment equipment and executing the experiment, members of the Georgia Tech Space Systems Design Laboratory and Cognitive Engineering Center for their support and expertise, and Robert Braun for his continued guidance and advisement. This work is sponsored in part by the NASA Graduate Student Researchers Program (GSRP).

References

[1] Chua, Z. K., Major, L. M., and Feigh, K. M., "Modeling Cockpit Interface Usage during Lunar Landing Redesignation," *Proceedings of the 15th International Symposium of Aviation Psychology*, Dayton, OH, April 2009.

[2] Forest, L., Cohanim, B., and Brady, T., "Human Interactive Landing Point Redesignation for Lunar Landing," *Proceedings of the 2008 IEEE Aerospace Conference*, IEEEAC 1109, Big Sky, MT, March 2008.
doi:10.1109/AERO.2008.4526299

[3] Needham, J., "Human–Automation Interaction for Lunar Landing Aimpoint Redesignation," Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2008.

[4] Chua, Z., and Major, L., "Task Modeling for Lunar Landing Redesignation," *Proceedings of the AIAA InfoTech@Aerospace Conference*, AIAA 2009-2059, Seattle, WA, April 2009.

[5] Riché, L. J., Colton, G. M., and Guillory, T. A., "Apollo 11 Flight Plan," NASA AS-506/CSM-107/LM-5AS-506/CSM-107/LM-5, July 1969.

[6] Brady, T., Schwartz, J., and Straube, T., "Operational Concept Description for the Autonomous Landing and Hazard Avoidance Technology (ALHAT)," NASA, ALHAT 2.0-002, 2006.

[7] Kessler, L. J., and Forest, L. M., "Human-Interactive Autonomous Flight Manager for Precision Lunar Landing," Presentation, *Proceedings of the AAAI Fall Symposium on Spacecraft Autonomy*, Fall 2006.

[8] Forest, L., Kessler, L., and Homer, M., "Design of a Human-Interactive Autonomous Flight Manager (AFM) for Crewed Lunar Landing," *Proceedings of the AIAA Infotech@Aerospace*, AIAA-2007-2708, Rohnert Park, CA, May 2007.

[9] Davis, J., Striepe, S., Maddock, R., Hines, G., II, S. P., Cohanim, B., Fill, T., Johnson, M., Bishop, R., DeMars, K., Sostaric, R., and Johnson, A., "Advances in POST2 End-to-End Descent and Landing Simulation for the ALHAT Project," *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, AIAA 2008-6938, Honolulu, HI, August 2008.

[10] Klein, G., *Sources of Power: How People Make Decisions*, The MIT press, Cambridge, MA, 1998.

[11] Godwin, R., *Apollo 11: The NASA Mission Reports*, Vol. 1, Apogee Books, Burlington, Ontario, Canada, 1999.

[12] Marsh, A. K., "In-Flight Emergencies: Engine Out," Aircraft Owners and Pilots Association, http://www.aopa.org/pilot/features/inflight9901.html [retrieved February 2009].

[13] McCandless, J., Hilty, B., and McCann, R. S., "New Displays for the Space Shuttle Cockpit," *Ergonomics in Design: The Quarterly of Human Factors Applications*, Vol. 13, No. 4, 2005, pp. 15–20.

[14] "Take a Spin around the Inside of Altair," NASA, http://blogs.nasa.gov/cm/blog/Constellation.blog/posts/post_1233778865603.html [retrieved February 2009].

[15] Schweiger, M., "Orbiter Simulator," http://orbit.medphys.ucl.ac.uk [retrieved February 2009].

[16] Homan, D., "EDGE (Engineering Doug Graphics for Exploration)," NASA, http://graphics.jsc.nasa.gov [retrieved March 2009].

[17] Meyer, A., "X-Plane," *Laminar Research*, http://www.x-plane.com [retrieved February 2009].

[18] Monsen, R., "EagleLander3D," http://eaglelander3d.com [retrieved January 2009].

[19] Cohanim, B., and Collins, B., "Landing Point Designation Algorithm for Lunar Landing," *Journal of Spacecraft and Rockets*, Vol. 46, No. 4, 2009, pp. 858–864.
http://dx.doi.org/10.2514/1.42002

[20] Braun, R. D., and Manning, R. M., " Mars Exploration Entry, Descent, and Landing Challenges," *Journal of Spacecraft and Rockets*, Vol. 44, No. 2, March–April 2007, pp. 310–323.
http://dx.doi.org/10.2514/1.25116

Kelly Cohen
Associate Editor