MVL 13.08

PILOT DETECTION OF SYSTEM FAILURES DURING A LUNAR LANDING TASK IN A MOTION SIMULATOR

J.D. Kaderka[1,2*] and K.R. Duda[2]

[1]Man-Vehicle Laboratory, Department of Aeronautics and Astronautics, MIT, *jkaderka@mit.edu

[2]The Charles Stark Draper Laboratory, Inc.

Introduction

Future complex systems, such as those found in piloted aircraft and spacecraft, will undoubtedly utilize significant automation to enhance pilot capabilities and enable novel mission scenarios. Off-nominal conditions may arise in the vehicle, such as automation or hardware failures, and the ability of the pilot to correctly synthesize the presented state or system status information to detect failures might ultimately influence pilot safety and mission success. Small deviations of failure detection in time-critical tasks can have a large impact on the outcome of a mission. A series of experiments were performed in the NASA Ames Research Center Vertical Motion Simulator (VMS) to investigate the effect of vehicle control mode, motion cues, and failure type on human failure detection performance.

Experiment Apparatus

A 6 DOF lunar landing simulator was used at the NASA VMS (Figure 1) for this experiment. Terminal descent was simulated; initial conditions were 100 seconds from touchdown. The simulator's interior resembles a potential layout that would be used for lunar landing (Figure 2). Pilots were standing in the cockpit, had three displays [Primary Flight Display (PFD), Horizontal Situation Display (HSD), and a Digital Elevation Map (DEM) / Hazard Map (Figure 3)], and had an out-the-window view of the lunar surface through Apollo-style windows. Apollo-style dynamics were used in the simulation. An attitude joystick and translational hand controller were used for pilot inputs.

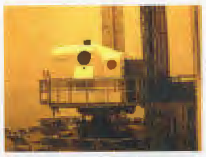

Figure 1. NASA Ames Research Center Vertical Motion Simulator.

Figure 2. Interior view of the lunar lander cab with the three displays and out-the-window view.

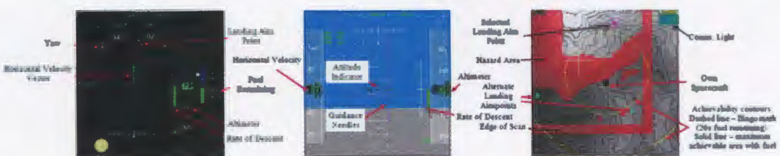

Figure 3. Three displays used for lunar landing

Experiment Design

A within subjects, full factorial design was employed for this experiment. Pilots were initially briefed on the experimental tasks, trained to a specific flying proficiency level, and then performed two data collection sessions in an experiment that lasted approximately five hours. Forty-four data trials (two replicates) were conducted with independent variables of motion cues, control mode, and failure type pseudo-randomized. Each trial lasted 70 seconds; subjects initially performed a landing point designation in autopilot for the first 20 seconds, transitioned into a pre-assigned control mode [(1) autopilot; (2) pitch, roll, and yaw command (TA); or (3) pitch, roll, yaw, and rate of descent command (TA-ROD)], and then flew the vehicle using guidance cues while monitoring system states for possible system failures (Figure 4). Three failures – (1) thruster failed on, (2) noise in the descent radar, and (3) a fuel leak (Table 2) – were incorporated in the experiment. Only one failure was possible for a given trial and trials with no failures were incorporated (18% rate) to reduce the expectancy of failures. Mental workload was assessed by response times to a secondary task while situational awareness was assessed by pilot callouts of system states during the trials. Both of these measures have been previously been utilized in pilot-in-the-loop lunar landing simulations [1, 2].

Table 1. Subject tasks during trials

Goals of Subject	
Primary Task	Null guidance errors in manual control
	Detect and diagnosis system failures
Secondary Task	Respond to a communications light
Situation Awareness	Make verbal callouts (altitude, fuel remaining, DEM landmarks) throughout the trial

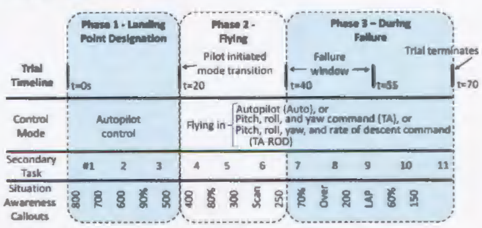

Figure 4. Timeline of each trial and the placement of all tasks of the subject

Table 2. Possible system failures and their detection strategy

Failure	Description	Primary Cue	Secondary Cues
Thruster Stuck On	One thruster stuck on resulting in a pitch down / roll left vehicle movement	Unexpected movement in attitude indicator (PFD); Motion Cues	Increase number of necessary control inputs
Descent Radar Failure	Increased level of noise in altimeter, rate of descent, and horizontal velocity	Unexpected movement in guidance needles (PFD)	Increased level of noise in altimeter, rate of descent, and horizontal velocity
Fuel Leak	Increased rate of fuel consumption	Achievability contours shrink at faster rate (DEM)	Fuel gauge decreases at faster rate (PFD, HSD)

Hypotheses

Based on the literature, five hypotheses were formed.

Failure detection will be: (1) fastest in the full manual control mode,
(2) fastest with motion cues, and
(3) fastest for failures occurring in the primary flight display versus a secondary display.

In addition, mental workload and situation awareness will:
(4) have a decrement following the mode transition, and
(6) have a further decrement when a failure occurs.

Results

Fourteen instrument rated pilots (22-32 years old, 13M/1F) with a range of 190-2500 hours total time (40-1500 total instrument time) were recruited from the San Francisco Bay area. All subjects were trained to fly the vehicle with less than 6° RMSE in both pitch and roll and less than 2 ft/s in rate of descent. Latency of failure detection was analyzed with a mixed hierarchical regression while secondary task response time and situation awareness callouts were analyzed with a Friedman test.

Failure detection (key findings):
- Difference between control modes
  - Fastest detection in TA
  - Slowest detection in TA-ROD
- No difference with motion cues
- Difference between failure types
  - Thruster failure detected fastest
  - Radar failure detected slowest
- Cross effects exist between failure type and control mode

Table 3. Failure detection matrix

	Detection	No Detection
Failure Present	457	47
Failure Absent	11	101

Table 4. Failure diagnosis

Correct Diagnosis	418
Incorrect Diagnosis	39

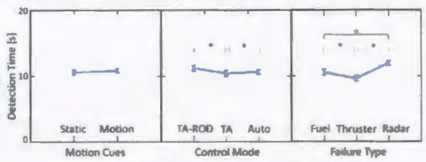

Figure 5. Failure detection main effects (* p<0.05)

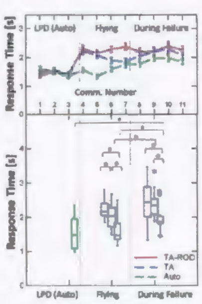

Figure 6. Secondary task response time (* p<0.002, Bonferroni correction)

Mental workload and situation awareness (key findings):
- Difference between trial phases
  - Decrement after mode transition
  - Further decrement during a failure
- No difference with motion cues
- Difference between control modes in flying phase
  - Best performance in Auto
  - Worst performance in TA-ROD
- Difference between control modes during the failure phase
  - Best performance in Auto
  - Worst performance in TA-ROD

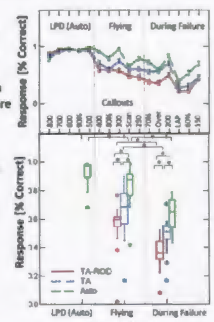

Figure 7. Situation awareness callouts (* p<0.002, Bonferroni correction)

Discussion

As expected there was a significant difference in detection time between failure types; the thruster failure, which provided a strong deviation in attitude indicator and motion cues was detected fastest. However, the radar failure, which also manifests on the PFD resulted in the slowest detection. This slow detection could be due to subjects (while in manual control) initially interpreting errors in guidance needles as their own poor performance instead of immediately attributing these errors to a failure. Another surprising result was that no significance was found in turning on/off motion cues. When asked about the motion cues post-experiment, severel pilots commented that they tried to ignore motion cues and only fly by instrument flight rules. Finally, another surprising result is that the TA control mode resulted in the fastest detection time overall. While it was hypothesized that TA-ROD would result in fastest detection due to subjects being in the control loop, subjects appeared to be highly worked in this control mode at the expense of their ability to detect failures.

Future Work

Results of this experiment give insight into which factors play a role in detecting system failures by experienced pilots. An additional experiment at the Draper fixed-base simulator will utilize an eye tracker to determine (1) pilot resampling and dwell times on failed instruments just prior to failure detection and (2) how the pilot's scan pattern changes with different control mode. These results, along with the VMS results, will be used to inform a human performance model of a lunar lander touchdown task, which will be parametrically analyzed to determine cases most sensitive to failure detection.

Acknowledgements

Charles M. Oman (MIT), Jessica J. Marquez (NASA), Alan Natapoff (MIT), Aaron Johnson (MIT), Scott Reardon (NASA), Maggie Taula (SJSU), and Steve Beard (NASA)

References

[1] Hainley, C.J. (2011) *S.M. in Aeronautics and Astronautics*. MIT. Cambridge, MA. 1-389.
[2] Stimpson, A.J. et al. (2011) *IEEE Aerospace Conference*. 1-11.

This work was supported by the National Space Biomedical Research Institute through NASA NCC 9-58 Project HFP02001