HUMAN IN THE LOOP SIMULATION MEASURES OF PILOT RESPONSE DELAY
IN A SELF-SEPARATION CONCEPT OF OPERATIONS

Maria C. Consiglio*, Dr. Sara R. Wilson*, James Sturdy**,
Dr. Jennifer L. Murdoch*, David J. Wing*
*NASA Langley Research Center, **Raytheon Mission Operations & Solutions

Abstract

A human-in-the-loop (HITL) simulation experiment was conducted by the National Aeronautics and Space Administration (NASA) to assess airline transport pilots’ performance and reported acceptance of the use of procedures relying on airborne separation assistance and trajectory management tools. This study was part of a larger effort involving two NASA centers that includes multiple HITL experiments planned over the next few years to evaluate the use of automated separation assurance (SA) tools by both air traffic controllers and pilots.

This paper presents results of measured pilot response delay that subject pilots incurred when interacting with cockpit tools for SA and discusses possible implications for future concept and procedures design.

1 Introduction

The automation of the separation function is expected to reduce the capacity limiting effect of human-based separation control of today’s system. Automation can remove workload bottlenecks, improve prediction of conflicts as traffic patterns become more dense and complex, and provide conflict-free maneuver alternatives to human operators. Future concepts of operations that are being evaluated as part of the Next Generation Air Transportation System (NextGen) [1] include both ground and aircraft based automation for separation. For example, integrated air/ground operational concepts have been proposed in which some aircraft crews exercise separation functions aided by Airborne Separation Assistance System (ASAS) tools on the flight deck, while air traffic controllers exercise ground based separation control for non-ASAS-equipped aircraft and terminal operations [2,3]. ASAS systems integrate advanced decision-aiding automation into aircraft avionics that rely on broadcast surveillance information including aircraft velocity vectors and limited flight plan information through a surveillance capability such as the Automatic Dependent Surveillance – Broadcast (ADS-B).

Automation systems for separation require operators to respond in a timely manner to conflict alerts in accordance with procedures and training.

Delayed responses to conflict alerts may result in late resolutions possibly leading to loss of separation (LOS) as shown in Figure 1.

Fig. 1. Delayed Traffic Conflict Resolutions

There is very limited experience with human performance metrics when interacting with airborne separation automation since to date there are no operational systems and very few experimental studies. These applications are known to be context sensitive, in other
words, human performance is influenced by specific training and procedures as well as other required cockpit tasks. Multiple factors can affect pilot response time and pilot conformance, such as lack of trust as a result of frequent occurrences of false alerts or conflicting training instructions. Also, lack of situation awareness and excessive workload are factors known to affect response delay. The operational procedure within which the alerting system is used is likely to affect the operator’s response time [4].

When pilots are given responsibility for maintaining separation and are trained and expected to react within a certain time period, then a response delayed beyond the specified time range can be considered a pilot error. The Traffic Alert and Collision Avoidance System (TCAS) II procedures specify that pilots are expected to act within 5 seconds of a Resolution Advisory. It is not expected that the procedures for separation provision will require as short a response time as collision avoidance. In research systems, the design of separation tools and procedures generally assume a pilot response time of several tens of seconds.

Pilots operating under Visual Flight Rules (VFR), while not responsible for “separation” in the air traffic control sense, are expected to follow the established rules to see and avoid obstacles and other aircraft. While there are no expected response times specified in the procedures, the measured pilot response times are reported to be between 10 and 12.5 seconds according to the Federal Aviation Administration (FAA) and Department of Defense (DoD) [5].

A recent batch \(^1\) (non-piloted) experiment addressed the potential impact of aircraft crew response delay when interacting with ASAS tools using response delays between 5 and 240 seconds in high density traffic [6]. This “stress test” experiment revealed great resilience of the strategic-only (i.e., flight-plan based) ASAS automation functions being tested, showing performance degradation only at combinations of extreme pilot delays and traffic density levels. These results begin to shed light on future ASAS procedures and reliability requirements of strategic ASAS capability. The next step is to obtain measurements of actual response delays from HITL simulation experiments and flight tests and to assess the impact of observed delays on system performance. In addition, these studies will help to refine pilot procedures, flight deck interfaces and training guidelines.

The HITL simulation experiment reported in this paper measured pilots’ response delay when interacting with ASAS automation under high traffic density conditions. Participants were current commercial transport pilots, with no prior experience with self separation concepts. They were given training on cockpit procedures and were expected to respond to traffic conflict alerts without delay. The independent variables of the study were traffic density and arrival time constraints.

The remainder of the paper is organized as follows: Sections 2 and 3 provide a brief description of the simulation platform and experiment respectively. A more in-depth description of this experiment is provided in [7]. Section 4 describes the measures collected and the analysis of results and Section 5 the conclusion.

2. Simulation Platform

The simulation platform utilized in this experiment was the Airspace & Traffic Operations Simulation (ATOS) [8] that runs in the Air Traffic Operations Laboratory (ATOL) at the NASA Langley Research Center. It is comprised of hundreds of real-time, aircraft simulators equipped with ADS-B and self-separation capability. The HITL aircraft simulator is the Aircraft Simulation for Traffic Operations Research (ASTOR). Each ASTOR includes a six degree-of-freedom dynamic airplane model, a Flight Management System (FMS), Mode-S ADS-B datalink capability, and a prototype ASAS called the Autonomous Operations Planner (AOP) [9].

The AOP provides an integrated suite of capabilities for managing traffic conflicts and trajectory changes from the flight deck

\(^1\) In batch experiments aircraft simulators are flown by an automated pilot model.
perspective, including conflict detection, resolution, prevention, and trajectory constraint conformance. These capabilities are developed to a level of fidelity sufficient for research of complete airborne responsibility for self-separation as defined by the Principles of Operation of ASAS Category 4, “Airborne Self-Separation” [10]. In the Joint Planning and Development Office (JPDO) NextGen Concept of Operations, they correspond to the functions of Separation Management and aspects of Trajectory Management [1].

2.1 The AOP Pilot Interface

The AOP conflict detection and resolution logic includes a strategic intent-based function and a tactical intent-based function. (The tactical state-based function was not active in this experiment.) The strategic system is used for FMS-based operations. The tactical system is used for flight using Mode Control Panel (MCP) guidance and serves primarily as a “tactical override” of the strategic system when predicted time to LOS is short. Conflict alerting is modeled as a multi-alert-level system in which the timing of alerts between aircraft are staggered as a method for incorporating a right-of-way rule set (i.e., priority rules), based roughly on the set used in VFR. The AOP provides pilots with two different conflict alert levels depending on the predicted time to LOS. A Level 1 (L1) conflict alert is displayed 6-10 minutes (min) before a predicted LOS. A Level 2 (L2) conflict alert is displayed 0-6 min before a predicted LOS.

Both alerts are displayed on the Navigation Display (ND) with an overlaid white (L1) and yellow (L2) “dog bone” shape on the ownship’s active route as shown in Figure 2. The “dog bone” length indicates the location and duration of the predicted LOS. L1 alerts are accompanied by a chime sound and L2 alerts are also accompanied by an aural warning saying “traffic alert traffic alert”. Furthermore, L2 alerts come with a countdown timer, whereas L1 alerts do not. The countdown timer indicates how much time the pilots have to execute a strategic resolution before “tactical override” compels them to solve the conflict tactically.

Pilots are trained and expected to respond to these types of alerts in a timely manner, according to the severity of the conflict alert.

Fig. 2: AOP Interface: Traffic Conflict Alert

Most conflicts are L1 alerts when first detected and may progress to an L2 alert if a pilot fails to respond on time. On rare occasions, usually due to blunders or severe deviations from the predicted flight path, a conflict may be detected late and displayed as an L2 alert from the beginning. Pilots are trained to perform the same actions for L1 and L2 alerts (i.e., choose and execute a resolution trajectory) but with an increased sense of urgency in the latter case. The conflict detection and resolution (CD&R) timeline is described in the next Section.

2.2 The Conflict Detection and Resolution Timeline

As shown in Figure 3, the timeline for the life span of a strategically resolved conflict by the AOP, a conflict begins at first detection, but for the pilot, it begins when it is first displayed on the ND. The earliest display time for aircraft with higher priority is 7 min to a predicted LOS. The “give-way” aircraft gets the alert as soon as the conflict is detected.
The pilot response delay $d_1$, or “realization time,” is the time elapsed from the first display of the conflict alert to the pilot’s reaction of requesting a resolution from the AOP. A distracted or busy pilot that fails to notice a conflict may exhibit longer $d_1$ delay. Next, the resolution cycle time is the time it takes AOP to compute one or more conflict-free trajectory alternatives. This time is variable and can be affected by traffic density and complexity [11]. The cycle ends when the resolution trajectory is displayed to the pilot. The time elapsed from resolution display to the pilot's action of uploading a resolution route to the FMS is the response delay $d_2$. At this point, pilots are already aware of the conflict and are considering the resolution route offered by AOP. During this “selection time” the pilot carefully evaluates the proposed vertical and or lateral resolution routes by examining both on the ND. Once the resolution route is uploaded to the FMS, it is displayed as a white dashed line and needs to be executed to become the new active route. The pilot response delay $d_3$ is the time elapsed from the pilot’s upload action to the action of executing the new FMS route. This “decision time” gives the pilot the option to reconsider the new route before committing to the change.

It must be noted that the chart in Figure 3 represents a nominal timeline in which there are no intervening events. Frequently, other events such as new traffic conflicts can occur during any of the phases described. Those cases need to be treated individually to identify the timing of the pilot actions.

Figure 3. Conflict Detection and Resolution Timeline

3 Experiment Description

The study reported in this paper was part of a larger effort involving NASA Ames and Langley Research Centers. It comprises multiple HITL experiments planned over the next few years to evaluate the use of automated separation assurance (SA) tools by both air traffic controllers and pilots.

The primary goal of this first experiment was to assess the degree of comparability achievable with a companion HITL experiment conducted to investigate ground-based automated separation [12].

An additional goal of the experiment was to assess the agility of self-separation operations in managing trajectory-changing events in high traffic density, en-route operations with arrival time constraints. Secondary goals included assessing pilot acceptability of the concept, tools, and procedures; collecting pilots’ subjective workload ratings; and objectively measuring safety, efficiency, and pilot performance. This paper presents results of measured pilot response delay that subject pilots incurred when interacting with cockpit tools for SA.
3.1 Experiment Design

The experiment included two sets of scenarios: one set of four 30-min scenarios, and one set of three 15-min scenarios.

A 2x2 within-subjects design was used to collect data during the 30-min scenarios that involved either the presence or absence of scheduling assignments provided to aircraft operating within an airspace having a sustained traffic density level either 1.5 times (1.5x) or 2 times (2.0x) greater than approximate current day airspace capacity. Two replicates of each 30-min scenario were performed for a total of eight simulation runs.

A separate 3x1 within-subject design was used to collect data during three 15-min experiment scenarios. The three 15-min scenarios had a 2.0x traffic density level, and two of the 15-min scenarios included scripted events that manipulated the timing of aircraft trajectory changes. These changes were induced by a data link message containing an updated required time of arrival (RTA) necessitating a delay maneuver. Two replicates of each 15-min scenario were performed for a total of six simulation runs.

3.2 Participants and Training Approach

Data were analyzed from forty-eight commercial transport pilots. Two of these pilots had not flown a transport category aircraft within the last year. However, the remaining 46 pilot participants consisted of 38 commercial transport pilots employed by U.S. air carriers or aircraft manufacturers and eight commercial transport pilots employed by European air carriers.

The 48 subject pilots participated as groups of 12 in four separate three-day experiment sessions conducted in March 2010. Each experiment session consisted of a series of training exercises involving classroom instruction and hands-on simulated flight, the completion of 14 data collection scenarios, and a final debrief session involving the pilot participants and members of the research team.

During their training exercises, pilots received instruction regarding the self-separation operational concept and associated cockpit technology and procedures. Timeliness of responses to traffic alerts was emphasized, and pilots were instructed that, when notified of a traffic conflict, they were expected to resolve the conflict without delay and to execute one of the offered resolutions in a timely fashion [7]. However, the pilots were not given a specific response time requirement (e.g., 5 seconds for TCAS). It was expected that the pilots would take the time necessary following the alert to understand the conflict situation before acting.

4. Results and Discussion

Pilot response times recorded during the 30-min and 15-min scenarios were analyzed separately to determine if the duration and other substantial scenario differences had an effect on response time. Differences included the number of observed conflicts (more conflicts were observed during the 30-min scenarios), proximity of the top-of-descent (often reached in the 30-min scenarios but never in the 15-min scenarios), and inclusion of the scripted data link required time of arrival (RTA) message in the 30-min scenarios.

Data from the 30-min scenarios are presented first. There were a total of 242 strategically resolved conflicts during the 30-min scenarios. Of these conflicts, 230 were encountered by all 48 pilots and first appeared as L1 alerts. The remaining 12 conflicts first appeared as “pop-up” L2 alerts and were encountered by 11 different pilots. For each pilot, the mean values of $d_1$, $d_2$, and $d_3$ were computed over all eight of the experiment’s 30-min scenario runs.

Response times to L1 and L2 conflicts were analyzed separately to assess the effect of alert level and its associated urgency on pilot response time.

4.1 L1 Alerts for the 30-min Scenarios

4.1.1 Pilot Response Delays $d_1$

A histogram of $d_1$ for the 48 pilots is shown in Figure 4. The data are right skewed.

2 Conflicts resolved tactically (using the MCP rather than the FMS) were excluded from this initial analysis.
and there is an outlier at 84.20 seconds (sec). The mean pilot response delay $d_1$ for the 48 pilots is 15.37 sec, and the standard deviation is 11.58 sec, with a minimum of 5.38 and a maximum of 84.20 sec.

From the histogram in Figure 4, the pilot response delay $d_1$ for L1 conflict alerts appears to approximately follow a lognormal distribution with estimated parameters $\hat{\mu} = 2.60$ and $\hat{\sigma} = 0.46$.

4.1.2 Pilot Response Delays $d_2$

A histogram of the pilot response delay $d_2$ for the 48 pilots is shown in Figure 5, and we can see that the data are right-skewed. The mean pilot response delay $d_2$ for the 48 pilots is 9.91 sec and the standard deviation is 3.94 sec, with a minimum of 5.35 and a maximum of 25.50 sec. From the histogram in Figure 5 the pilot response delay $d_2$ appears to approximately follow a lognormal distribution with estimated parameters $\hat{\mu} = 2.23$ and $\hat{\sigma} = 0.35$. A 95% confidence interval on the mean pilot response delay $d_2$ is (8.41, 7.51).

4.1.3 Pilot Response Delays $d_3$

A histogram of the pilot response delay $d_3$ for the 48 pilots is shown in Figure 6, and once again the data are right skewed. The mean pilot response delay $d_3$ for the 48 pilots is 7.24 sec and the standard deviation is 3.17 sec, with a minimum of 1.85 and a maximum of 16.37 sec. From the histogram in Figure 6, the pilot

![Fig. 4. Histogram of $d_1$ with lognormal distribution fit (30-min scenarios)](image)

![Fig. 5. Histogram of $d_2$ with lognormal distribution fit (30-min scenarios)](image)

![Fig. 6. Histogram of $d_3$ with lognormal distribution fit (30-min scenarios)](image)

A summary of the descriptive statistics for response delays $d_1$, $d_2$, and $d_3$ to L1 conflict alerts is shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>StDev</th>
<th>Min</th>
<th>Med</th>
<th>Max</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_1$</td>
<td>15.37</td>
<td>11.58</td>
<td>5.38</td>
<td>12.53</td>
<td>84.20</td>
<td>(11.77, 15.38)</td>
</tr>
<tr>
<td>$d_2$</td>
<td>9.91</td>
<td>3.94</td>
<td>5.35</td>
<td>8.62</td>
<td>25.50</td>
<td>(8.41, 10.29)</td>
</tr>
<tr>
<td>$d_3$</td>
<td>7.24</td>
<td>3.17</td>
<td>1.85</td>
<td>5.45</td>
<td>16.37</td>
<td>(5.80, 7.51)</td>
</tr>
</tbody>
</table>
All three response times seem to be similarly distributed. The mean response time to a conflict’s first detection is approximately 16 sec and the mean time to evaluate the resolution routes, d2 is about 10 sec. This is the time during which pilots were evaluating and comparing the resolution routes offered by AOP (vertical and lateral maneuvers) before uploading a particular route.

Table 2. Lognormal distribution estimated parameter values for pilot response delays d1, d2, and d3 for L1 alerts (30-min Scenarios)

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \hat{\mu} )</th>
<th>( \hat{\sigma} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 )</td>
<td>2.60</td>
<td>0.46</td>
</tr>
<tr>
<td>( d_2 )</td>
<td>2.23</td>
<td>0.35</td>
</tr>
<tr>
<td>( d_3 )</td>
<td>1.89</td>
<td>0.44</td>
</tr>
</tbody>
</table>

After uploading the route it took pilots approximately 8 sec to execute the route. Many factors are likely to have affected those times; among them new procedures and responsibilities, training, trust in the system, interface design, experiment modeling limitations, etc.

![Pilot Response Delay Distribution](image)

Fig. 7. Lognormal distribution fit of \( d_1 \), \( d_2 \), and \( d_3 \) for L1 alerts (30-min scenarios)

At the end of the experiment, pilot participants provided abundant feedback to that effect which will be assimilated in new design guidelines for the AOP interface and pilot procedures. In future batch\(^3\) simulation experiments, these pilot response delays can be simulated as constant values by using the means in Table 1. They can also be randomly generated using the lognormal distribution with the estimated parameters in Table 2, which would have the probability density functions in Figure 7.

4.2 L2 Alerts for the 30-min Scenarios

Of the 242 strategically resolved conflicts, only 12 were first displayed as L2 alerts. These unusual “pop-up” conflicts were encountered by 11 different pilots. The mean values of \( d_1 \), \( d_2 \), and \( d_3 \) were computed for each pilot and then aggregated over all the pilots. It should be noted that this analysis is based on a small dataset. Descriptive statistics are summarized in Table 3.

Table 3: Descriptive statistics for pilot response delays \( d_1 \), \( d_2 \), and \( d_3 \) for L2 conflict alerts and 11 pilots (30-min scenarios)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>StDev</th>
<th>Min</th>
<th>Med</th>
<th>Max</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d_1 )</td>
<td>7.65</td>
<td>5.49</td>
<td>3.15</td>
<td>6.80</td>
<td>23.00</td>
<td>(4.46, 9.48)</td>
</tr>
<tr>
<td>( d_2 )</td>
<td>7.87</td>
<td>3.86</td>
<td>4.25</td>
<td>6.60</td>
<td>17.55</td>
<td>(5.48, 9.52)</td>
</tr>
<tr>
<td>( d_3 )</td>
<td>5.10</td>
<td>3.56</td>
<td>1.50</td>
<td>4.60</td>
<td>14.30</td>
<td>(2.78, 6.46)</td>
</tr>
</tbody>
</table>

As shown in Tables 1 and 3, the response delays for the “pop-up” L2 alerts are much smaller than for the L1 alerts, particularly \( d_1 \). This finding is consistent with the expected higher sense of urgency intended by these conflict alerts. The mean initial reaction time \( d_1 \) for L2 alerts was approximately half of the corresponding time for L1 alerts.

There are multiple reasons to explain this behavior but clearly the interface design for L1 and L2 alert levels combined with the yellow symbology, appear to have had the intended effect of shortening the pilot response time. It is also possible that pilots waited longer to act on L1 alerts to see if the conflict was

---

\(^3\) In batch experiments aircraft simulators are flown by an automated pilot model.
resolved by the other aircraft or was a false alert.

4.3 L1 and L2 Alerts for the 15-min Scenarios

There were a total of 57 strategically resolved conflicts during the experiment’s 15-min scenarios. Of these conflicts, 56 were L1 conflicts, and only one was a “pop-up” L2 conflict. Conflicts were encountered during the 15-min scenarios by only 29 of the 48 pilots. For each pilot, the mean values of d1, d2, and d3 were computed over all strategically resolved conflicts that he encountered during his six runs.

Table 4: Descriptive statistics for pilot response delays d1, d2, and d3 for all conflict alerts and 29 pilots (15-min Scenarios)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>StDev</th>
<th>Min</th>
<th>Med</th>
<th>Max</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td>14.15</td>
<td>7.49</td>
<td>6.70</td>
<td>11.35</td>
<td>35.45</td>
<td>(10.75, 15.06)</td>
</tr>
<tr>
<td>d2</td>
<td>10.81</td>
<td>5.53</td>
<td>3.60</td>
<td>9.30</td>
<td>26.90</td>
<td>(8.20, 11.56)</td>
</tr>
<tr>
<td>d3</td>
<td>7.26</td>
<td>5.63</td>
<td>2.40</td>
<td>5.15</td>
<td>27.95</td>
<td>(4.58, 7.46)</td>
</tr>
</tbody>
</table>

Table 4 presents a summary of the descriptive statistics for pilot response delays d1, d2, and d3, and confidence intervals on the mean delays. A comparison of these data with those shown in Tables 1 and 3 indicates that for the 29 pilots that encountered strategic conflicts, the mean pilot response delays recorded during the 15-min scenarios are similar to those recorded during the 30-min scenarios for the 29 pilots that encountered strategic conflicts.

Table 5. Lognormal distribution estimated parameter values for pilot response delays d1, d2, and d3 (15-min scenarios)

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \hat{\mu} )</th>
<th>( \hat{\sigma} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1</td>
<td>2.54</td>
<td>0.44</td>
</tr>
<tr>
<td>d2</td>
<td>2.28</td>
<td>0.45</td>
</tr>
<tr>
<td>d3</td>
<td>1.77</td>
<td>0.63</td>
</tr>
</tbody>
</table>

In fact, 29 of the 48 pilots encountered conflicts in both the 15-min and 30-min scenarios. For these pilots, the scenario differences had no significant effect on mean pilot response delay d1, d2, or d3 using alpha = 0.05. A total of five outliers were encountered ranging between 27.95 and 36.90 seconds.

In future non-piloted simulation experiments, these pilot response delays can be simulated as constant values by using the means in Table 4. They can also be randomly simulated using the lognormal distribution as shown in Figure 8 with the estimated parameters in Table 5.

5 Conclusion

Separation assurance automation must be designed to account for, and mitigate the impact of varying human response. This must be accomplished through extensive testing and evaluation of new technologies involving HITL experimentation and batch “stress” tests. The former allows the collection of measurements of human operator performance variability while the latter is needed to address the resilience of the new systems using performance models based on the experimental measures. This process will enable the design of preventive procedures and more robust automation.

This paper provided an in-depth analysis of the pilot response delay measures obtained during a recent HITL experiment that
involved forty eight commercial transport pilots.

A lognormal distribution was found to be a good fit for the pilot response delays and appropriate coefficients for the individual delay components $d_1$, $d_2$, and $d_3$ were provided. These results can be applied in future experiments to implement pilot performance models for similar applications. Response delay times can be modeled using the values provided in Tables 1 and 3 or randomly generated using the lognormal distribution with the coefficients provided in Tables 2 and 5.

Additionally, it was determined that the total pilot response times were well within ranges verified during prior stress tests and no safety impact in terms of LOS was observed.

The interface design methods intended to reduce pilot response in more urgent situations (L2 alerts) had the desired effect of getting the attention of the pilots and reducing the overall CD&R time.

Finally, results from this study will contribute to the understanding of aircraft crews’ subjective perception of their new role and responsibility in a self separation concept of operations and provide design guidelines for new procedures and tools.

References


Contact Author Email Address

Maria.C.Consiglio@NASA.gov

Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS2010 proceedings or as individual off-prints from the proceedings.