Autonomous Flight Testbed (AFT): Designing a Flight Simulation System to Explore Future Human-Machine Teaming Concepts

Ryan Holec, Makenzie Hockensmith, Jessica Broll, Caroline Wittich, Bianca Donadio, Ewart de Visser and Chad Tossell

March 6, 2020

Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 2, 2019
A Regional Conference of the Society for Industrial and Systems Engineering

Autonomous Flight Testbed (AFT): Designing a Flight Simulation System to Explore Future Human-Machine Teaming Concepts

Ryan P. Holec, Makenzie G. Hockensmith, Jessica E. Broll, Caroline Wittich, Bianca Donadio, Ewart de Visser, and Chad C. Tossell

United States Air Force Academy

Corresponding author's Email: C20Ryan.Holec@usafa.edu

Author Note: Ryan Holec, Makenzie Hockensmith, and Jessica Broll are first-class cadets majoring in Systems Engineering with an emphasis in Human Factors at the United States Air Force Academy (USAFA). Caroline Wittich is a 2nd-class cadet majoring in Systems Engineering. Bianca Donadio is a research assistant in the Warfighter Effectiveness Research Center (WERC). Dr. de Visser is the Technical Advisor for the Warfighter Effectiveness Research Center. Lt Col Tossell is an associate professor in the Department of Behavioral Sciences and Leadership at USAFA and directs the WERC. The views expressed in this article are those of the authors and do not represent the United States Air Force, Department of Defense or United States Government.

Abstract: Autonomous flight is becoming more viable. The United States Air Force (USAF) is aiming to leverage autonomous capabilities, such as the QF-16, to team with human pilots in the battlefield. However, there is very little research to assess how autonomous capabilities might integrate with human pilots in real-world warfighting scenarios. Given this limitation, we have designed and developed the Autonomous Flight Testbed (AFT) to explore human-autonomy teaming in a realistic flight simulator. The AFT explores novel concepts of operations for human-piloted F-35 flight with simulated autonomous F-16s. The AFT system, including the architecture, scenarios, and measurement capabilities, are described below. As a proof of concept, we evaluate trust, workload, and situation awareness in scenarios designed with the help of subject matter experts. The results of the study will be reported and discussed along with the AFT as a tool to help the USAF develop capabilities for future autonomous teaming.

Keywords: Systems modeling, Human-autonomy teams, Flight Simulation

1. Introduction

Throughout the United States Air Force's history, pilots have teamed to conduct flight operations in the world to achieve desired effects. Recent operations such as Operation Enduring Freedom, Operation Iraqi Freedom, and Operation Tomodachi would not have been nearly as successful without pilots within the same airframe, and across different airframes, teaming together to complete a variety of missions. In future warfighting environments, autonomous systems will likely be more prevalent. Autonomous systems differ from automated systems in that they are more independent and unpredictable (Kaber, 2018). Given they are becoming more viable, the USAF is exploring the use of autonomous systems as teammates in air combat (Endsley, 2016). This human-machine teaming (HMT) will allow human pilots to oversee and team with a number of autonomous systems across air combat mission sets. The result will likely increase lethality, safety, and efficiency across military services.

Can human pilots effectively supervise these autonomous systems in addition to their own aircraft and still complete their mission? To assess the impact of these future defense capabilities, the Warfighter Effectiveness Research Center (WERC) at the United States Air Force Academy (USAFA) developed the autonomous flight testbed (AFT) to examine the effectiveness of using autonomous wingmen in future Air Force missions. The testbed described in detail below affords HMT research to improve trust, situation awareness (SA), stress, and performance of human-machine teams.

Although there have been some experiments in the past to study trust between humans and robots, there has never been a study about the interactions between human pilots and autonomous wingmen to this degree. This study analyzes the effect of the AFT on pilots with varying degrees of experience, to assess how this technology will impact future Air Force pilots and the validity of this investment. Furthermore, the Autonomous Wingman study is of a higher criticality than past experiments as field testing of this technology will present safety considerations for participants.

In the next section, the design of this simulator, including the hardware/software, scenarios, and measurement capabilities, is described. Following this section, the research samples are described, which includes evaluations of visualizations, controls, and concepts of operations. The final sections include an evaluation of the scenarios as well as a short conclusion with future work to develop data-driven methods, concepts, and technologies for future air warfare.

2. The Autonomous Flight Testbed (AFT)

The AFT was designed to be a high-fidelity simulator to allow human pilots to fly with autonomous wingmen (Figure 1). Commercial-off-the-shelf (COTS) technologies were integrated to build the F-35 flight simulator. With the help of subject matter experts (SMEs), we designed the displays, controls, scenarios, and measurement capabilities to provide an immersive experience to study human factors in HMT including trust, situation awareness, and communication. A number of scenarios and measurement capabilities have been developed to assess a wide range of future concepts.

Figure 1: A cadet flying with simulated autonomous F-16s in the Autonomous Flight Testbed (AFT)

2.1 Hardware and Software Capabilities

As shown in Figure 2, the Autonomous Flight Testbed (AFT) was designed to be a high-fidelity simulator. Participants fly the F-35 simulator using the F-35 Hand on Throttle and Stick (HOTAS). The displays found in F-35 simulators used in USAF training are overlaid with new interfaces and models designed by cadets with subject matter experts (SMEs) and faculty at USAFA. For example, one display being tested by cadets was designed to keep track of autonomous wingmen. These wingmen are shown in the left and right monitors in Figure 1. The center monitor displays both the scene of view in addition to the cockpit. All of the display and control configurations were designed with SMEs.

These SMEs also helped us develop a “cheat sheet” that is used to mimic a pilot’s kneeboard. This kneeboard includes the names of the targets at each SAM site, how to make a radio call, the call-signs of each of the autonomous F-16s, and the flight parameters. Kneeboards are used by real Air Force pilots to organize flight-pertinent information, such as weather, area maps, altitude ceilings and floors, and key flight objectives. Actual pilots are continually checking their kneeboard and monitoring the gauges, radar screens, and the Heads-up Display (HUD) that shows their speed and altitude.
A VR system can also be used to display the scenario or for Wizard of Oz (WoZ) teaming.

Prepar3d (P3D) simulates F-35 displays and flights and also is the software. The integration of the four monitors into the testing environment, each with their own information stream, allows for more robust psychological and cognitive fidelity in the experiment, which mimics the heavy workload and strained SA of actual pilots.

Measurement technologies will be described in the final paper.

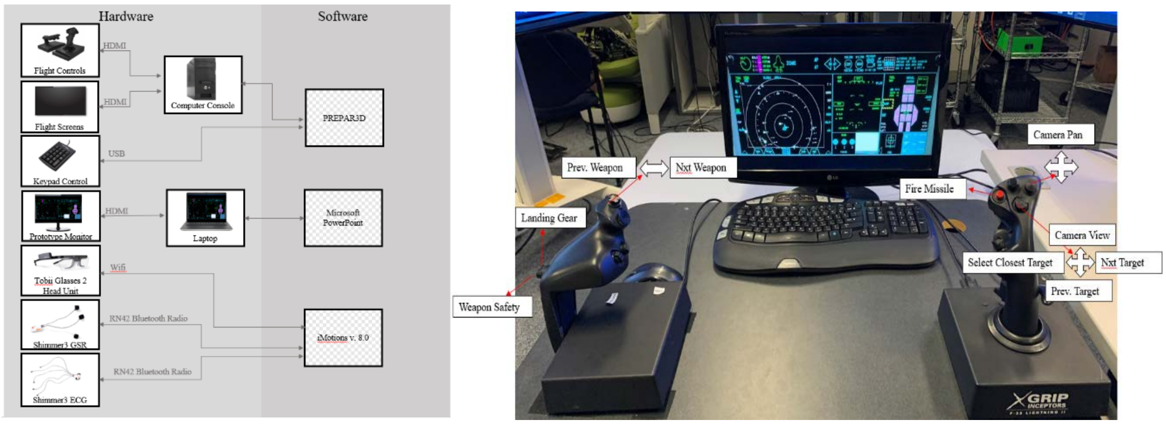

Figure 2: Hardware-software Integration along with display and controls

2.2 Concept of Operations for Autonomous Flight Scenarios

The participant flies the F-35 using the F-35 HOTAS which includes both the throttle and the stick. The center monitor is a cockpit view for the F-35. The monitor on the left displays a back view of the F-35 and then weapons gauge. The right monitor displays the fly-by view of the F-35, which allows the participant to view his/her F-16s form up into formation. Integrating the four monitors into the testing environment added to the inherent workload presented to the participants as they are overloaded with information, requiring greater situational awareness and attention. This allowed for a higher fidelity experience for the pilots, resembling the workload experienced by actual pilots. The team has further increased fidelity by adding additional conditions that the participant must abide by and monitor throughout the scenario, to include: maintaining a speed between 300-500 knots while engaging enemy targets, stay below 12,000 ft AGL, and pull no more than 6 G's throughout the scenario. There are two scenarios that have been created based on modern day Air Force SEAD (suppression of enemy air defenses) missions. The low task-load scenario consists of three enemy sites that each contain SAMs (surface to air missiles) that must be destroyed to complete the mission. The participant is in charge of deciding how to destroy the targets, whether it be deploy all three of the F-16s and perform a supervisory control mission, engage each of the targets individually, or a combination of F-16s and using the F-35's weapons. The test is concluded when all of the three targets are neutralized and two out of three F-16s are formed up on the F-25.

The second scenario is the high task-load scenario. The design of this scenario is based off of the first scenario; however, it incorporates an additional SAM site designated for the F-35 pilot to destroy and two aerial adversaries that must be destroyed. The scenario forces the participants to utilize their autonomous wingmen in order to complete the mission. The scenario concludes once the targets are all destroyed and two of the three F-16s form up on the F-35. The high task-load scenario increases workload by increasing the amount of targets that must be destroyed and the inherent difference in targeting moving aerial adversaries as compared to stationary ground targets. The flow of the experiment is detailed below in Figure 2 and the controls for the AFT test bed is also shown in Figure 4.

3. Proof-of-Concept Study

The goal of the proof-of-concept study is to evaluate the design and fidelity of the AFT. We expected high workload scenarios to be more difficult and elicit more errors. The physiological measures should correlate with the increased activities within the scenario. Likewise, the low workload scenarios were expected to be elicit better performance.

3.1 Method

In the initial stages of testing, six participants, consisting of five seniors and one sophomore, were put through the majority of the scenario. The cadet participants completing the experiment had little to no flight hours, ranging from 0 to 17

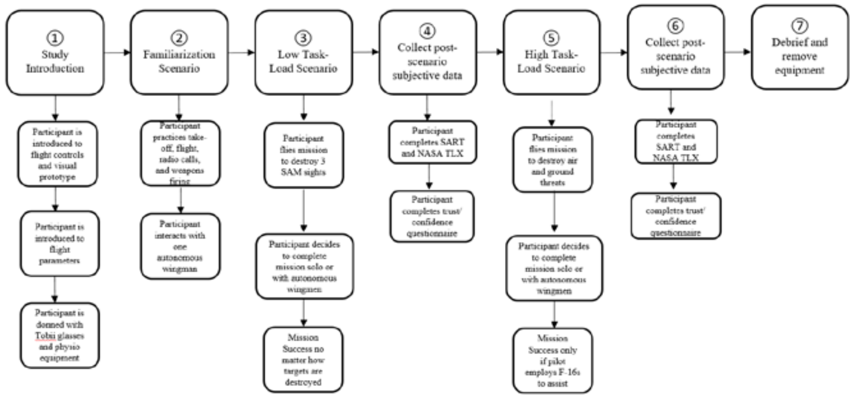

Figure 3: Experiment Flow Chart

hours of flight. Given that these participants were run in the initial stages of testing, they did not use the visual prototype and there were no radio commands given by the experimenter. The dry run participants only completed the familiarization flight scenario and the low task-load scenario. These participants were solely asked for their feedback on the difficulty of the scenario, based on the controls and their ability to perform the basic tasks of flying the F-35, utilizing the cameras, and the ability to select weapons and target the enemy SAM sites. The feedback from the dry runs allowed the team to update the controls on the HOTAS and add further explanation in the experimental script on how to use the weapons gauge and target SAMs. The dry run participants showed the team that there were too many initial targets in each of the targeted areas which led to targeting confusion and increased target selection frustration.

The experimenter solicited feedback from the participants based upon the study's validity and clarity; updates were made to the run book reflecting the changes suggested that would allow for a higher fidelity flight experiment. Unlike the initial pilot studies, the visual prototype was used and audio commands were given during the second round of testing.

2.2 Results

Results will be described in the final paper. The data from this experiment will capture the capabilities of the participant in both a low and high task-load scenario and see how the situational awareness of the participant, the trust in the autonomous wingmen, overall performance, and workload vary across the scenarios and participants (Air Force pilots vs. non-Air Force pilots).

4. Conclusions

A novel flight simulator has been presented as a testbed to conduct human-machine teaming research. The immediate goal of this project was to build a realistic flight platform for experimentation. As such, we presented both the design of the system as well as an initial experiment to explore the scenarios designed to elicit a range of human behaviors in flight with autonomous wingmen.

The experimental design has its limitations and flying in a simulator is always different from actual combat data. The high stress of actual combat is not tested in these scenarios and further research in more stressful situations and scenarios should be conducted prior to applying autonomous wingmen to the war zone.

Still, the end goal of this capstone is to be able to provide the stakeholders and the U.S. Air Force with viable and accurate results that can be internalized and lay the groundwork for future human-machine teams in the U.S. military. Furthermore, the methods used to take quantitative measurements of all factors of interest, especially trust, can be used for future studies. Before any air asset is put into commission, it is necessary to have a deep understanding of the way that the human interacts with the system. This experiment is a stepping stone toward the development of data-driven methods, concepts and technologies that can be applied to future studies on air warfare.

5. References

Chen, J. Y, & Barnes, M. J. (2014). Human-Agent Teaming for Multirobot Control: A Review of Human Factors Issues. *IEEE Transactions on Human - Machine Systems, 44*(1), 13-29. doi:10.1109/thms,2013.2293535

Cummings, M. L., & Guerlain, S. (2007). Developing Operator Capacity Estimates for Supervisory Control of Autonomous Vehicles. *Human Factors*, *49*(1), 1–15. https://doi.org/10.1518/001872007779598109

Donath, D., & Schulte, A., (2015). Behavior Based Task and High Workload Determination of Pilots Guiding Multiple UAVs. *Procedia Manufacturing,* 3, 990-997. https://doi.org/10.1016/j.promfg.2015.07.156

Kaber, D. B. (2018). Issues in human–automation interaction modeling: Presumptive aspects of frameworks of types and levels of automation. Journal of Cognitive Engineering and Decision Making, 12(1), 7-24.

Mirchi, T., Vu, K., Miles, J., Sturre, L., Curtis, S., Strybel, T., (2015). Air Traffic Controller Trust in Automation in NextGen. *Procedia Manufacturing*, 3, 2482-2488. https://doi.org/10.1016/j.promfg.2015.07.509.

Wickens, Christopher D. "Situation Awareness and Workload in Aviation." *Current Directions in Psychological Science*, vol. 11, no. 4, 2002, pp. 128–133. *JSTOR*, www.jstor.org/stable/20182789.