COGNITIVE CONSTRAINTS FOR AUTOMATION ON FLIGHT TESTING

Diogo Silva Castilho
Massachusetts Institute of Technology
Boston – MA

The first flight of a new aircraft is still a dangerous event. Despite all simulations and software predictions, test pilots face many unknowns when a prototype leaves the ground for the first time. The cultural celebration of first flights masks the concerns of many stakeholders about the technical challenges of the new equipment. The pilot extensively prepares to react properly to unexpected situations and often bring a new story to tell, but in a time when remotely piloted and autonomous aircraft fly every day, the question about how to use their technologies to save a test pilot life arises. This study investigates the technical advantages of using specific autopilot modes and remote or autonomous controls. It also discusses the disadvantages of relying on airborne sensors instead of using pilots cognitive capabilities and judgment. The analysis on the data collected by students in a Flight Testing Course supports that there are clear advantages of the suggested new approach. The control stick input technique to explore the longitudinal stability of an aircraft is used as an example of human limitations on measuring quantitative variables. The results are extended to the critical phases of the campaign and the analysis points to new safety constraints that cannot be ignored.

Humans and machines progressively share the control of numerous vehicles and this symbiosis has reached a level of maturity that makes it a main topic of research. Aviation pioneered the use of automation for decades, but human pilots are still manually controlling aircraft in dangerous and precise situations.

Despite the fact that machines execute many tasks more precisely and faster than humans, according to Fitts List, humans are more versatile, innovative and better for error correction and judgment (de Winter and Dodou, 2014). Automation is not able to autonomously support strategies to manage complexity, anticipate the dynamics of cross-adaptive processes or to deal with tradeoffs and dilemmas (Woods and Sarter, 2000). However, as algorithms become more reliable and versatile, increases on the levels of automation in flight controls happen through improvements on autopilots modes.

The activity of Flight Testing (FT) must safely verify the accomplishment of requirements and validate the product for certification. If automation can reduce costs, time of development, or make it safer, then manufacturers must explore related technologies using new devices and techniques.

FT campaigns of fixed-wing aircraft designed to be piloted by humans can improve the use of automation in two different ways. The first is a human-machine partnership providing a more efficient investigation of aircraft characteristics with inputs precise in timing, frequency, and amplitude. The second is the remote operation of the aircraft to execute dangerous test events.

The Benefit of Precision

Handling qualities (HQ) events test for longitudinal and latero-directional stabilities. The FT crew measures the natural frequency of an aircraft oscillation and the damping after an input on that axis (IPEV, 2015). Test pilots are trained to investigate many different frequencies and amplitudes on each axis trying to induce a pilot-aircraft coupling. If it ever happens, it causes changes in the design or reduced operational limits.

An analysis of data collected during the 2016 Brazilian FT Course, focusing on the technique applied to the excitation of the Short Period mode, proved that automation would provide a more efficient investigation of the handling qualities than traditional methods. There are several different techniques to investigate the Short Period. All of them have an input followed by the observation of the aircraft's reaction.

Figure 1 shows the stick input followed by the natural response of the aircraft in angle of attack (AoA) and pitch angle. The first input that the pilot provided is a frequency sweep followed by multiple short duration inputs to investigate the Short Period. In this sweep, the pilot starts cycling the control stick longitudinally in a low frequency while maintaining a relatively constant altitude (+/-20ft) and air speed (+/-2kt). For doublet applications, the pilot memorizes the frequency in which reactions have more amplitude. The two inputs before 100s in the x-axis of figure 1 are longitudinal doublets.

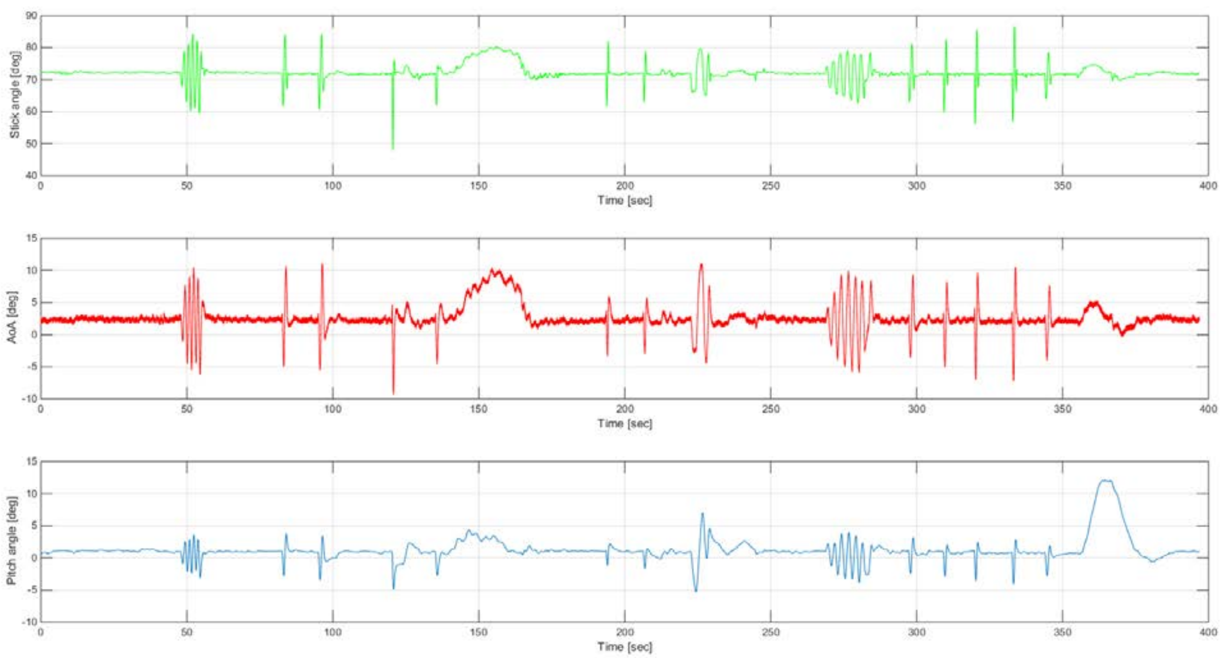

Figure 1. Recorded flight data of short period test events

The pilot performs the first doublet at a small amplitude to avoid exceeding load factor limits. If the response in pitch is safe, i.e., not resulting in pilot-aircraft coupling, the next must have more amplitude, but with the same frequency.

Flight test engineers analyze the relation between the input on the flight stick and the reaction of the aircraft to build aerodynamic models of the aircraft considering the technique applied by the pilot as standard. The same issues observed in this analysis happen directionally for the investigation of the Dutch Roll.

The data was collected from the flights of three pilots. As a result of training, in theory, there should be only a small variation in input frequency among them. However, pilots applied doublets with frequencies varying from 0.45 to 1.25Hz and this variability on the input, just after the frequency sweep, leads to uncertain conclusions made by engineers. The analysis of pilot's inputs showed a cognitive pattern as they all have a similar but wide spread in input frequencies. This proved that it is incorrect to assume that the test pilot input frequency is always precise. When the precision of input is critical, a specific autopilot mode could provide inputs on a flying stick or yoke that are precise in amplitude and frequency. It would eliminate the uncertainties and characterize the response of the aircraft more efficiently. Other phases of FT campaigns could also take advantage of the precision provided by automation, including fixed decelerations for stall (e.g.: 1 kt/s) and windup turns (e.g.: dg/dt = 1g every 5s).

Merging the natural advantages of humans and machines in a cockpit, in 2016, Aurora Flight Sciences developed for DARPA a concept program called ALIAS (Aircrew Labor In-Cockpit Automation System). This system has portable hardware and software that can be configured in less than one month to operate any different type of aircraft. The goal of this system is to reduce crew requirements by the robot replacing the copilot in the right seat. The machine reads the gauges using machine vision and its arms operate the yoke and the throttle levers. The combination of the strengths of humans and robots in the cockpit supposedly provide less workload for the pilot and enforce the execution of all procedures. The system has no Artificial Intelligence to be predictable and reliable for the pilot.

Similar systems might be developed to test aircraft that have mechanical flight controls, including the hydraulically boosted ones. The idea of applying the inputs on the yoke and pedals is to include all the looseness and inflections of the control system on the analysis. For aircraft with fly-by-wire controls and autopilot, a simple autopilot mode might be implemented exclusively for testing.

All of these experimental solutions would be applied before certification. Thus, the risk analysis of the FT campaign must address malfunctions and systemic issues. Woods and Sarter (2000) named the following new problems caused by automation: over automation, human error and bad man-machine coordination. The solution to deal with all of these is training because the use of intermediate levels of automation with inaccurate mental models are a source of new unsafe control actions (Leveson, 2012). All mode confusion in regular operation characterize scenarios that were not sufficiently explored during FT and other development phases.

Risk in FT

Robots are easier to replace than humans. Cost and time to train a professional limit the replacement of highly qualified human operators. After manufacturing new hardware and loading the latest version of the software, the robot is ready to face the unknown again. Moreover, technology is getting cheaper and more accessible. Thus, the replacement of human operators by automation for dangerous tasks becomes more attractive.

Some activities could already use robots as the preliminary tester. However for social purposes, the presence of the human is still essential, despite the implicit danger. For example, sending an astronaut to explore Mars is more socially relevant than doing the same with robots receiving orders from earth. Similarly, facing the unknown behavior of a new aircraft, test pilots face adversities and human presence on the first flight of a prototype is still the target for historical pictures and the cover of magazines.

To open the flight envelope of an experimental piloted aircraft means flying at extreme speeds, altitudes and load factors to explore and enhance the performance of the aircraft. The pilot must work together with the FT engineer to determine the operational limits that will be followed during the entire life of the product. The exploration of these limits often provides amazing stories about things that went wrong, including losing the control of the aircraft or losing its parts. Most test pilots learn about mishaps that marked the development of important aircraft. In each of these stories, there was one common fact: the life of the pilot was in danger.

There is a big expectation for the first flights of new aircraft. When test pilots and FT engineers execute a long campaign plan, they are so concerned about not adding complexity to the first flight that many times they don't even retract the landing gear or change the flaps position. That happens not only because it is a technical milestone. First flights are also a media event loaded with cultural celebration.

On the other hand, the subsequent flights explore critical features, such as the aerodynamic flow during the stall, handling characteristics or the aircraft controllability when aborting a takeoff run. On many of these flights, the pilots know from the risk analysis that the probability of finding undesired vibrations or controllability issues is higher during specific events. The pilot prepares his mindset to react to surprises using emergency procedures that might include ejecting from the prototype. He flies because he was taught to do it. He flies because the adrenaline of facing the unknown makes him feel good. He flies because he seeks personal glory (O'Mara, 2011) by being the main character of stories to be told.

Each test has a piloting technique and its execution has a series of cognitive demands; all of which have safety impacts. The first type of investigation deals with finding operational performance limits, including stalling and maximum speed, the maximum load factors, and the maximum altitude. The second type relates to the handling qualities, e.g. sources of pilot-aircraft coupling and spins. Finally, system testing also has critical events, like the weapons separation from pylons, launch rails, and bomb bays.

The Performance phase extends the flight envelope using the build-up approach[1] to investigate unknown behaviors. The exploration of high **speeds** might find a buffet on the structure with potential loss of parts. At the other end, lower speeds explored during stall investigations might cause the pilot to lose control of the aircraft. Both situations require complex sensing, diagnosis and judgement of the test pilot to determine operational limits.

While chance of structural issues due to excess **load** is very small, if it happens there is no time to react. The signals that the structure is about to collapse are cognitively perceived by the pilot as noises and vibrations different than usual. If the positive or negative limit is high, as in fighter aircraft, the senses of the pilot are affected by the g-force and his or her judgment is compromised. Automation would not suffer such restrictions and the combination of acceleration and vibration sensors, and microphones would provide the recording of the phenomena and a basic reaction.

For high **altitudes**, risk is related to pressurization issues, such as noises and the increase in cabin altitude. Pilots have a better judgment than automation about diagnosing off-nominal situations with the structure or sub-systems when the aircraft struggles with huge differential pressures. But for first climbs, the effects of hypoxia and decompression explosions are

[1] Build-up approach means that the event starts at a safe initial condition and the parameter is increased gradually and at a constant rate. This rate depends on how unpredictable the behavior of the system is to that extreme condition.

extremely dangerous to humans. In this case, a first flight to the maximum altitude without a human on board would reduce drastically the severity of this test event on the risk analysis.

Adding external payloads in pylons require an investigation of the effects of **flutter**, a resonant vibration of wing tips and stabilizers. If the oscillation is divergent, the test might be catastrophic. Even after using software and wind tunnels, this test is still important. Devices designed to produce these oscillations are installed on the trailing edge of the surfaces and the aircraft take off with a chase aircraft to record videos of the test. Accidents caused by flutter are not common, but their severity is often high. Thus, automation would be welcome for the same reasons as in load factor.

For handling qualities, the **spin** is one of the most critical maneuvers on a FT campaign of training and combat aircraft. The pilot must explain the behavior of the aircraft while reading speed, attitude and altitude. After the recovery, the pilot classifies the spin according to a metric chosen for the test. The maneuver itself take less than one minute, but the workload and dizziness are close to the human limitations. The remote control of a spin would be challenging because its implicit delay in communication interfere with the successful exit from spinning.

Finally, among all sub-systems tests, first-time **weapon separations** on military aircraft is critical because it might cause damage that interferes with the aircraft's controllability[2] and demands a fast decision about ejection. Remote operation would provide better chances of recovering the prototype, but as with spins, the delay would limit a proper reaction on controls.

The technology necessary to remotely control an aircraft with a seat and controls for a human pilot already exists. The QF-16 is an adaptation on the flight controls of a regular F-16 that enables it to be controlled remotely. The system has been flying since 2013 and reached operational capability as aerial target in 2016.

The challenge of building a machine to autonomously react properly and timely to all of these dangerous situations is the core of a cognitive paradigm, because the sum of methodological and theoretical approaches to all aspects of human psychology such as instincts, motor skills, memory, speech, values, personality, and problem-solving is too complex to be reproduced by software.

The safety improvement on aviation statistics with automation leads us to believe that, little by little, Intelligence Augmentation (IA) will reduce workload and increase autonomous properties up to the moment that Artificial Intelligence (AI) will take over and reduce the remote control to emergency modes of operation. In the light of the hexagon of cognitive sciences (Miller, 2003), the use of IA and the safe substitution of the human by AI is a multidisciplinary endeavor. This evolution must respect social phenomena and consider user's behavior when reacting to scenarios of automation failures and mode confusion.

In October 27th, 2016, Uber Technologies Inc. released a white paper picturing an aerial vision for urban transportation, envisioning that "pilot aids will evolve over time into full autonomy, which will likely have a marked positive impact on flight safety" (Uber, 2016). The initial certification process and operation of these new machines will be as piloted aircraft. Thus,

---

[2] When a bomb is launched from a pylon, one or two explosive charges are used to initiate the movement of the bomb away from the aircraft. Depending on many aspects, such as the charges sizing, sideslip, angle of attack, speed, attitude, and altitude, the bomb might present unstable separation and collide with the aircraft. Self-propelled weapons, such as missiles and rockets, use launch rails or launchers, but they are equally susceptible to separation issues like limited time for reaction.

the first generation of this urban VTOL (Vertical Takeoff and Landing) will pass on a regular FT campaign with the same performance and HQ issues discussed on this paper. It it will be an opportunity to prove the value of using higher levels of automation building statistical proof for users and regulators.

Conclusion

This paper discusses the main advantages and challenges of using different levels of automation on piloted and remote/autonomous control on fixed-wing aircraft designed to be piloted by humans. The use of machines acting on flight controls or devoted autopilot modes for precision on FT techniques along with the remote operation on dangerous events are new applications of automation with potential to make FT campaigns safer and more efficient. These applications are restricted to the FT events that do not require complex perception or judgment.

Intermediary levels of automation as a new autopilot mode require the adjustment of the test pilot’s mental models to the limitations of the system. This means that more preparation is necessary to avoid surprises. When the aircraft is fully autonomous or remotely operated for dangerous events, those sources of confusion diminish and the risk comes from the incompleteness of software or delayed communications link. The development of such systems will bring unforeseen accidents that must be addressed in the FT campaign risk analysis.

References

de Winter, J.C.F., Dodou, D., 2014. Why the Fitts list has persisted throughout the history of function allocation. Cogn. Technol. Work 16, 1–11. doi:10.1007/s10111-011-0188-1

Elevate, U., 2016. Fast-Forwarding to a Future of On-Demand Urban Air Transportation.

Fitts, P. M. (1954). "The information capacity of the human motor system in controlling the amplitude of movement". Journal of Experimental Psychology 47(6), 381–391.

Instituto de Pesquisas e Ensaios em Voo (IPEV), Divisão de Formação em Ensaios em Voo. Manual de Técnicas de Ensaios em Voo, Estabilidade Estática Longitunal EQ-02-TEV. São José dos Campos, 2015.

N. G. Leveson, Engineering a Safer World: Systems Thinking Applied to Safety. The MIT Press, 2012.

Miller, G.A., 2003. The cognitive revolution: A historical perspective. Trends Cogn. Sci. 7, 141–144. doi:10.1016/S1364-6613(03)00029-9

O’Mara, Raymond P. 2011. The Sociotechnical Construction of Precision Bombing : a Study of Shared Control and Cognition by Humans, Machines, and Doctrine During World War II / by Raymond P. O’Mara. c2011.

N. B. Sarter, D. D. Woods, and C. E. Billings, “Automation surprises,” in Handbook of Human Factors and Ergonomics, vol. 2, John Wiley and Sons, 1997, pp. 1926– 1943

Acknowledgements

This research was possible due to the scholarship provided by CNPq/Brazil.