EXPERIMENTAL ANALYSIS OF FLIGHT PERFORMANCE UNDER WORKLOAD VARIATIONS

Diego Hernandez Arjoni*, Emília Villani*, Manuel Rodríguez*, Aline Matheus*, Alex Almeida*, Guilherme Rocha*, Luís Gonzaga Trabasso*, Diego Hidalgo**
***Instituto Tecnológico de Aeronáutica; **Universidad Pontificia Bolivariana**

Keywords: *Flight Simulator, Workload, Pilot Performance*

Abstract

A great amount of aeronautical accidents and incidents in the last decades are associated with human causes, which can be related not only to the human itself, but with the human machine interface associated to a piloting task. This work presents a preliminary experiment that aims at analysing how a set of different tasks increases the workload of the pilot and how pilot's performance is affected by the increasing workload under different flight conditions (normal and abnormal). The experimental procedure considers 3 pilots executing a take-off and stabilization mission, where a group of tasks, based on the MATB-II approach, are systematically presented to the pilot. Variables such as altitude, heading, rate of climb and yaw rate, are measured. The results show that the variables measured near the pilot input command are more affected by the different levels of workload.

1 Introduction

According to the Civil Aviation Panorama published by CENIPA (Centre for Aeronautical Accidents Prevention) in Brazil, human errors are still one of the main causes of aircraft accidents [1]. Other organizations, such as the Boeing aircraft manufacturer, present similar analysis [2], [3].

It is important to notice that human mistakes are not exclusively due to lack of training or inability of the pilot. In many cases, it can be associated to environmental issues, such as loss of situational awareness and confusing Human-Machine Interface (HMI). The excess of displayed information may increase the workload to which the pilot is subjected and lead him/her to erroneous decisions [4].

One approach to tackle this problem is to improve the efficiency of pilot training using high fidelity simulators. Another approach is to improve the design of HMI and aircraft control systems in order to improve the situational awareness of the pilot under different scenarios, particularly in the case of aircraft failures.

In order to contribute to both approaches, the Centre of Competence in Manufacturing (CCM) of the Aeronautics Institute of Technology (ITA) developed the SIVOR Project in partnership with EMBRAER, the Brazilian aircraft manufacturer. SIVOR is a flight simulator that uses a COTS anthropomorphic robot as a moving platform.

The challenges faced for the validation of SIVOR showed us that we should investigate deeply the pilot-simulator interface. Furthermore the need of improving our knowledge about the human factors that affects the piloting activity lead us to propose the Integrated Vehicle Health Management and Human Factor Analysis (IVHM-HFA) Project, a partnership between Brazilian and Sweden academies and enterprises.

This work is part of the IVHM-HFA Project. It describes an experiment designed to analyse the effect of additional tasks on the pilot performance in normal and fault scenarios.

The next sections are organized as following. Section 2 presents a summary of previous experiments, the lessons learned and how they contributed to the proposal of the current experiment. Section 3 describes the experiment design. Section 4 discusses the results. Finally, Section 6 draws some conclusions and discusses future works.

2 Lessons Learned from Previous Work

The first experiments developed in the context of SIVOR simulator tackled the problem of how to validate the motion platform of SIVOR simulator in an objective way, without relying on subjective, and sometimes contradicting, opinions of pilot.

Our first attempt to answer this question consisted of:

a) Proposing three different metrics inspired in the TLX (Task Load Index) from NASA [5]. They are: *response time* to an external event that should triggers the pilot reaction; *precision* achieved when given a reference value for an aircraft variable, such as speed, altitude, and attitude; and *effort* performed by the pilot to follow the reference value, measured by the integration in time of the sidestick command.

b) Proposing different set of flight paths combining take-off, stabilization and climb, landing, off-set landing, stall recovery maneuvers. The flight paths were defined with increasing levels of difficulty.

The results obtained from three different set of experiments are described in [6], [7], and [8]. In all the three cases, analysis was performed using ANOVA and the following model:

$$V_{ij} = \mu + M_i + \beta_j + e_{ij} \qquad (1)$$

where:

V_{ij}: output (precision, effort or response time);
μ: general output mean;
M_i: variance of the motion mode (with or without motion);
β_j: variance of the pilot (blocked factor);
e_{ij}: random error.

The main lessons learned from these experiments are:

1. None of the proposed metrics was able to detect the influence of the motion system in the pilot behavior. We do not know if it is because the pilot behavior is actually not affected by the motion system or if the metric is not sensible enough to detect the influence of the motion system.
2. The pilot was considered as blocked factor in all the analyses. However, it was actually the most influencing factor. We suspect that the pilot variance is masking the effect of any other factor, including the motion.
3. The low number of repetition was a critical factor for the reliability of the experiment, considering the lack of homogeneity of the pilot sample.

From the results of those experiments, it was clear to us that we need to better understand how different pilots are affected by different levels of workload in different conditions.

The relationship between human performance, situational awareness and workload is addressed in several researches. Some works, such as the TLX (Task Load Index) from NASA deal with workload measurements through qualitative questionnaires about mental and physical aspects of the attributed task [5]. Other researches such as Ednsley's, focuses on the study of situational awareness and its direct relation with workload levels (Fig. 1) [9],[10]. In another initiative from NASA, the effect of workload levels on pilot performance is evaluated using the Multi-Attribute Task Battery program, commonly known as MATB-II [11]. The experiment described in the next section is based on the MATB-II proposal.

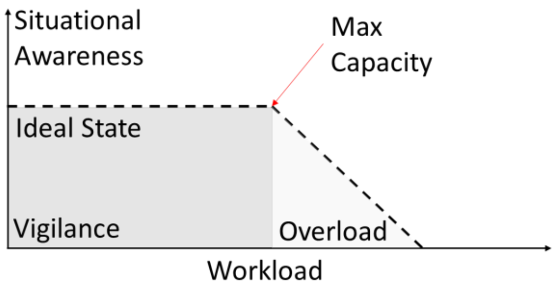

Fig. 1. Performance and Workload Relationship [7].

3 Experimental Procedure

This section present hypotheses, procedure and apparatus used in the experiment.

3.1 The Flight Mission

Briefly, the experiment consisted in having different pilots performing a primary task (take-off and stabilization mission) while being subjected to different workload levels. Each pilot repeated the same mission under different conditions, which could or not include the occurrence of aircraft failures. The different levels of workload were imposed using an approach similar to the one adopted by MATB-II.

The flight path is illustrated in Fig. 2.

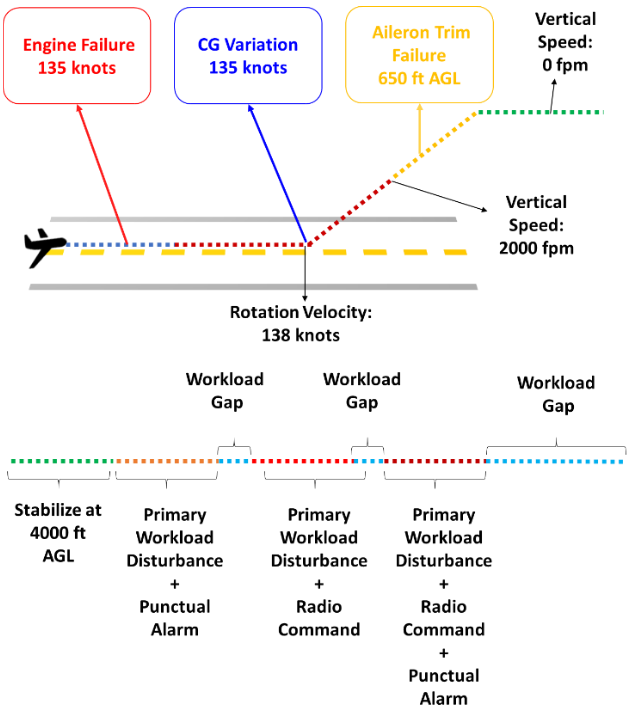

Fig. 2. Flight mission.

The flight mission is composed of the following sequence of events and manoeuvres:

1. The aircraft starts on the ground, with the parking brakes applied, the engines running, and the flaps set to minimum.
2. When commanded to start, the pilot must set the aircraft power to 100%, flaps to second level, and wait for stabilization of the engines;
3. Once stabilized the pilot may release the parking breaks and start the take-off run.
4. When reaching an altitude of 3000 ft the pilot must retreat the flaps and landing gear;
5. Once the VR = 138 knots is reached, the pilot must execute a rotation of the aircraft and maintain a vertical velocity of 2000 fpm until reach 3900ft;
6. At the point of 3900 ft the pilot must start to stabilize the aircraft at 4000±100ft and vertical velocity at 0 fpm
7. At the entire time, the pilot must keep the heading and the bank angle of the aircraft at 0±5° and control the aircraft only by the side stick, without changing the power of the aircraft.
8. When commanded the pilot must change the COMM standby frequency and set it to the selected frequency as fast as possible;
9. If the alarm turns on, the pilot must turn it off as fast as possible;
10. After submitting the pilot to the three levels of workload, the flight is interrupted.

The flight mission has 4 variations obtained from the different failure situations introduced in the first part of the path:

- **No failure**;
- **Altered CG**: the aircraft CG is moved aft during take-off rotation to a point where the aircraft longitudinal stability decreases;
- **Engine failure**: the right engine turns off suddenly at the speed of 135 knots;
- **Aileron trim failure**: the aileron trim suddenly blocks at 50% of its maximum value at an altitude of 2400ft (approximately 650ft from the ground).

Only one or none failure is introduced in each flight. Each pilot repeated the flight 12 times, corresponding to 3 repetitions of the 4 failure situations.

In the second part of the path, four workload conditions are imposed:

- **W1:** Primary Workload (PW) with no additional task (workload gap);
- **W2:** Primary Workload (PW) + Punctual Alarm (PA) task;
- **W3:** Primary Workload (PW) + Radio Command (RC) task;
- **W4:** Primary Workload (PW) + Punctual Alarm (PA) task + Radio Command (RC) task;

The punctual alarm is characterized by an uncomfortable sound and a visual change of a HUD instrument which turns a green “lamp” to red. This punctual alarm must be turned off by the pilot as soon as possible. The radio command is an instruction to change the radio standby frequency to the informed value and swap it with the selected comm frequency.

3.2 Experiment Hypotheses and Model

The following hypothesis were defined for the experiment:

- The pilot's previous flying experience is a significant factor that affects the pilot performance;
- The level of workload to which the pilot is submitted is a significant factor that affects the pilot performance.

The pilot performance is defined as the ability of the pilot to impose and maintain a predefined value to some aircraft flight variables, such as altitude.

A set of 4 flight sections were extracted from each flight, corresponding to the different workload levels. For each path, a set of 4 flight variables are analysed: altitude, rate of climb, heading, and yaw rate. In all the cases, the pilot was instructed to maintain the variable at a certain level. For each flight variable, the associated pilot performance is calculated as the absolute deviation from the set point. An additional variable is associated to the command intensity imposed by the pilot (Command Intensity Indicator – CII). It was calculated as the mean of the absolute deviation from zero of the detrended commanded input to the elevator.

In total, each flight mission produced 4 values for each of the 5 output variables that measure the pilot performance.

In order to verify the outlined hypotheses, a set of ANOVAs (Analysis of Variance) are performed [12]. All the 5 output variables were described by the statistical model (2):

$$V_{ij} = \mu + W_i + P_j + (WP)_{ij} + e_{ij} \qquad (2)$$

where:

V_{ij}: output value;
μ: general output mean;
W_i: variance of the Workload Factor;
P_j: variance of the Pilot Factor;
WP_{ij}: variance of the interaction between the Workload Factor and the Pilot Factor;
e_{ij}: random error.

3.3 Experiment Apparatus and Configuration

The experiment was performed in a static simulator based on Flight Gear 2018, which run on a desktop workstation. The aircraft used in the experiment is a Boeing 737-300 available on the Flight Gear database. The aircraft model was modified to inject faults. The pilot commands the aircraft through a Saitek 52 PRO sidestick and power lever set, as illustrated in Fig. 3.

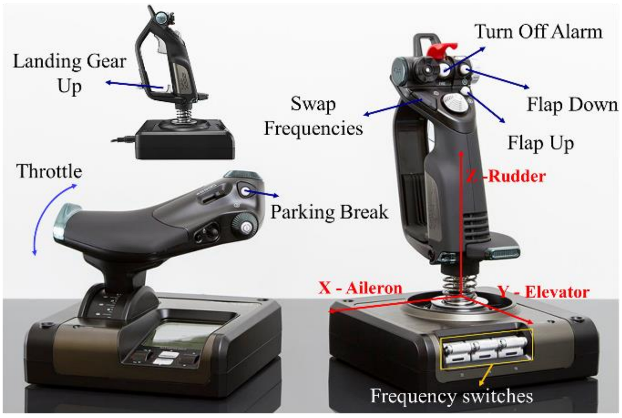

Fig. 3. Joystick commands.

The implementation of a blended scenario that combined the real flight information and the additional workload management tasks was achieved using both XML scripts and programming scripts interpreted by the simulator. Some additional features such as a HUD (Head Up Display) was built to accommodate virtual objects, making the disposal of information to all the pilots constant and intuitive. Fig. 4 displays the aircraft cockpit and HUD used on the experiment.

Fig. 4. Cockpit used in the simulation.

The HUD had two main functions: display necessary flight data and display the workload tasks that should be managed, as illustrated in the Fig. 5.

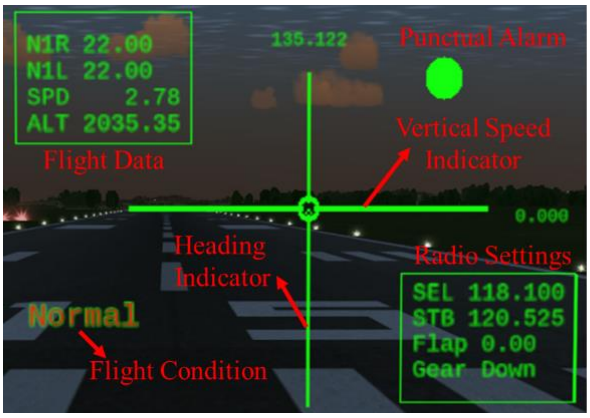

Fig. 5. HUD Functions.

3.4 Additional Considerations

Considering that the aircraft failures were introduced with no prior notice, the pilots could eventually lose the control of the aircraft. In order to assure a reliable data sample from each pilot, the following rules were adopted:

- Each flight condition can be repeated 2 times additionally to the 3 normal trials in case of control loss;
- If the pilot has more than 3 control losses, the data from the runs must be discarded;
- The pilot must have at least 2 data samples for each flight condition;
- In case of difficulties to stabilize, the run is classified as failure, but cannot be repeated once it is characterized as a piloting problem and not an aircraft loss;

The Table 1 summarizes the number of flights classified as valid for the three pilots that took part in the experiment.

Table 1. Data collected from pilots.

Pilot	Normal	Modified CG	Engine Failure	Aileron trim failure
1	3	3	2	3
2	3	2	3	3
3	3	3	3	3

As introduced in Section 3.2, this work considers five pilot performance indicators, calculated from the following variables of the recorded flight data:

- Altitude (ft);
- Rate of Climb (ft/s);
- Heading (°);
- Yaw Rate (°/s);
- Joystick position.

An example of the record altitude variable is presented in Fig. 6. The pilot performance indicator associated to the altitude variable is calculated as the mean of the absolute value deviation from the set-point (4000 ft ALS).

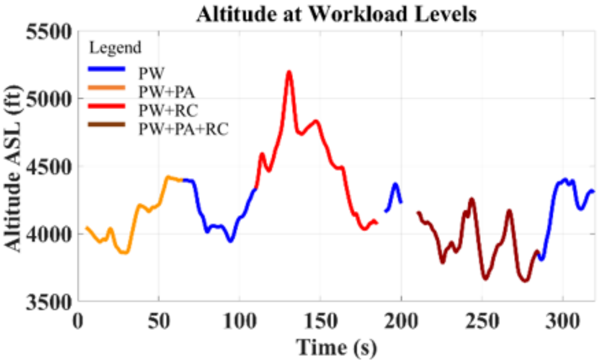

Fig. 6. Altitude data.

Following the same pattern, the heading indicator uses the constant set point of 135.1° (the runway heading). An example of the heading variable is presented in Fig. 7.

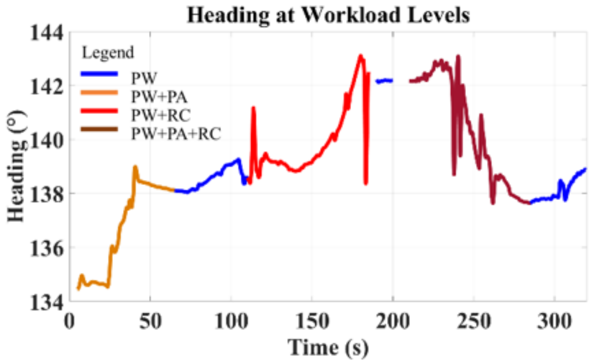

Fig. 7. Heading data.

In order to allow a more detailed analysis, the rate of climb and the yaw rate are considered as performance indicators as well, both are calculated by taking the mean of their absolute deviation from zero. Examples of rate of climb and yaw rate are displayed in Fig. 8 and Fig. 9.

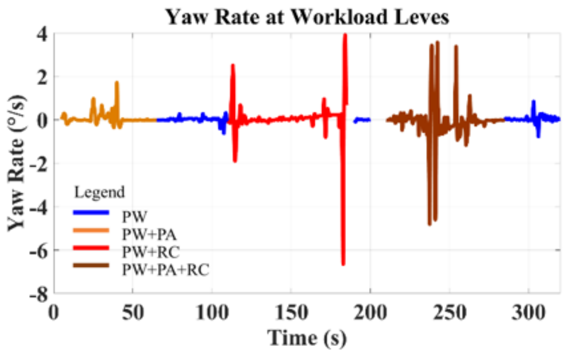

Fig. 8. Yaw rate data.

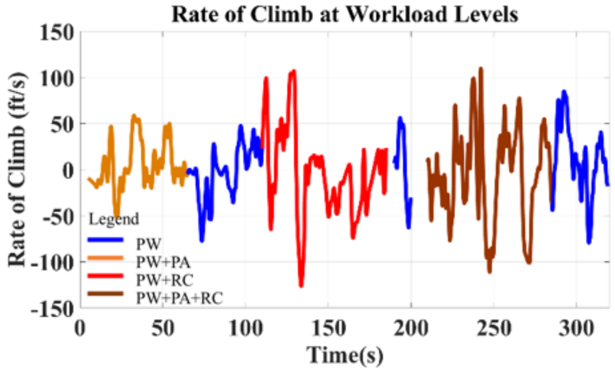

Fig. 9. Rate of climb data.

The command intensity indicator (CII) is calculated using the mean of the absolute detrended value of the joystick command at each workload sector. An example of measured joystick position is presented in the Fig. 10.

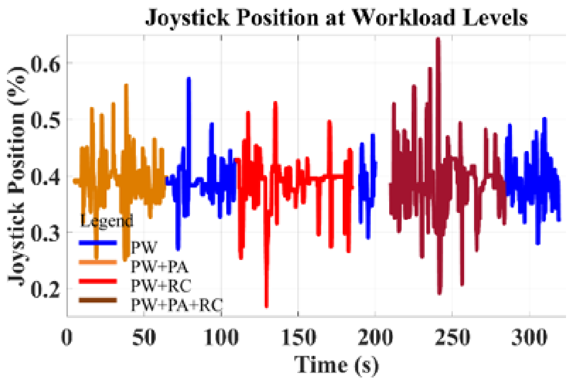

Fig. 10. Joystick position.

4 STATISTICAL ANALYSIS

This section presents a threshold analysis to verify the difficult imposed by the experiment, followed by the complete ANOVA [12] analysis and a discussion of the experiment results.

4.1 Threshold analysis

The threshold analysis aims at verifying if the pilots respected the thresholds imposed on each variables, and informed to them at the beginning of the experiment.

The analysis consists of calculating the mean and maximum value of the performance variables for the normal flight condition and for the primary workload at the normal flight condition.

The means for altitude and heading were subjected to the hypothesis test described in (3):

$$\begin{cases} H_0: \mu \leq TH \\ H_1: \mu > TH \end{cases} \tag{3}$$

where:

μ: Mean of the absolute difference between the data and the established set-point.
TH: Established Threshold for the variable.

The μ value is calculated by (4).

$$\mu = \frac{\sum |x_i - A|}{n} \tag{4}$$

where:

x_i: Sample of the considered variable (4000 ft for Altitude or 135.1° for Heading);
A: Set-point;
n: Sample length.

The means for rate of climb and yaw rate subjected to the alternative hypothesis test (5):

$$\begin{cases} H_0: \mu = 0 \\ H_1: \mu \neq 0 \end{cases} \tag{5}$$

where:

μ: mean of the sample.

The results are presented in the Table 2.

Table 2. Normal flight condition deviations.

	Max Deviation	Means	Accepted Hypothesis	P-Values
Altitude (ft)	2244.47	368.79	1	0.0008
Rate of Climb (ft/s)	342.46	-1.01	1	0.0045
Heading (°)	18.82	3.47	1	0.0001
Yaw Rate (°/s)	4.62	0.01	1	0.0000

The results of the hypothesis tests for the Normal flight in the primary workload condition are presented in the Table 3.

Table 3. Normal flight condition at primary workload section deviations.

	Max Deviation	Mean	Accepted Hypothesis	P - Values
Altitude (ft)	2244.47	393.74	1	0.0013
Rate of Climb (ft/s)	342.46	-1.06	0	0.0986
Heading (°)	18.82	3.53	1	0.0002
Yaw Rate (°/s)	4.62	0.02	0	0.1001

The hypothesis test shows a difficulty in maintaining the specified mission thresholds even in the flight with the lower level of difficulty for the variables of altitude and heading. It can be verified that the rate of climb and yaw rate had their thresholds respected at normal flight, in the primary workload section. On the other hand, the pilots were not able to keep the flight into the limited boundaries for the altitude and heading variables.

These tests show a baseline of the difficulty imposed to the pilot in the flight, and additionally, they exhibit that variables which are directly controlled by the pilot have more chances to be followed and respected.

4.2 ANOVA Analysis

The performance indicators of each flight path were subjected to an ANOVA procedure to verify the influence of the workload level on the pilot's performance and the influence of the piloting attitudes on its own performance.

Initially, the performance data were tested to confirm the normality hypothesis (Shapiro-Wilk test) and the homogeneity of variances hypothesis (Bartlett test).

Once the experiment is performed with human beings, some variables are not controllable and unpredicted attitudes can happen. As a result, outliers are common in the dataset. In order to deal with this issue, a standardized residual verification (6) is accomplished, and points which residual variance exceeds 3 standard deviations were removed from the data set if not compromising the sampling redundancies.

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}} \tag{6}$$

Each one of the 4 considered flight conditions results are displayed in the Table 4 to Table 7. The confidence level used is 95%, and the P-Values of the testing data is analysed. The null hypothesis of the Shapiro-Wilk test considers the residuals sample a normal distribution. Similarly, the null hypothesis of the Bartlett test endorses the homogeneity of variances.

When the normality assumption is not verified, the corresponding data was not submitted to the ANOVA analysis.

Table 4. Normal flights (P-values).

Condition	Altitude	Rate of Climb	Heading	Yaw Rate	CII
Shapiro-Wilk	0.0481	0.1542	0.1135	0.0695	0.6559
Bartlett W	0.0683	0.8517	0.0773	0.7585	0.7102
Bartlett P	0.2361	0.1618	0.0001	0.0447	0.1519
Workload Influence	-	0.1230	0.2106	0.4169	0.7055
Pilot Influence	-	0.0000	0.0003	0.0147	0.0002
Interaction influence	-	0.3290	0.1409	0.3506	0.8272

Table 5. CG modified flights (P-values).

Condition	Altitude	Rate of Climb	Heading	Yaw Rate	CII
Shapiro-Wilk	0.2150	0.9659	0.1986	0.5768	0.3397
Bartlett W	0.3301	0.1870	0.0033	0.2879	0.8474
Bartlett P	0.1022	0.0019	0.0002	0.2095	0.2469
Workload Influence	0.3538	0.0466	0.0503	0.0771	0.6770
Pilot Influence	0.0828	0.0332	0.0001	0.0333	0.0000
Interaction influence	0.3898	0.4979	0.3388	0.1559	0.8190

Table 6. Engine failure flights (P-values).

Condition	Altitude	Rate of Climb	Heading	Yaw Rate	CII
Shapiro-Wilk	0.3103	0.1616	0.0001	0.0038	0.8232
Bartlett W	0.0070	0.6404	0.3538	0.0080	0.2265
Bartlett P	0.0392	0.0002	0.0000	0.3647	0.0015
Workload Influence	0.0988	0.0127	-	-	0.6717
Pilot Influence	0.1247	0.0000	-	-	0.0996
Interaction influence	0.1163	0.3439	-	-	0.7686

Table 7. Aileron trim failure flight (P-values).

Condition	Altitude	Rate of Climb	Heading	Yaw Rate	CII
Shapiro-Wilk	0.6264	0.7567	0.0307	0.0076	0.3264
Bartlett W	0.0511	0.3509	0.3353	0.0062	0.9270
Bartlett P	0.1205	0.2329	0.0000	0.0001	0.9138
Workload Influence	0.1740	0.0006	-	-	0.4349
Pilot Influence	0.1250	0.0000	-	-	0.0477
Interaction influence	0.3270	0.0172	-	-	0.7537

The Table 8 shows the group of outliers removed from the data set.

Table 8. Removed outliers.

	Rate of Climb	24
Normal	Yaw Rate	16
	Command Intensity Indicator	14; 34
CG	Yaw Rate	4

An analysis of the obtained results shows that the pilot factor is significant in many cases, which means that the obtained performance varies according to the pilot. Furthermore, the workload factor influences the pilot rate of climb performance for all the flights but the normal situation. This result can also be observed in the boxplots of Fig. 11. It indicates that for missions performed using mostly the longitudinal dynamic at continuous flight paths of the aircraft, the rate of climb can be used for detecting changes of workload rather than the altitude indicator.

Finally, the latero-directional dynamics induced by the engine failure and the aileron trim failure, created a non-normal aspect of the yaw rate and heading indicators, preventing the analysis from being trustable.

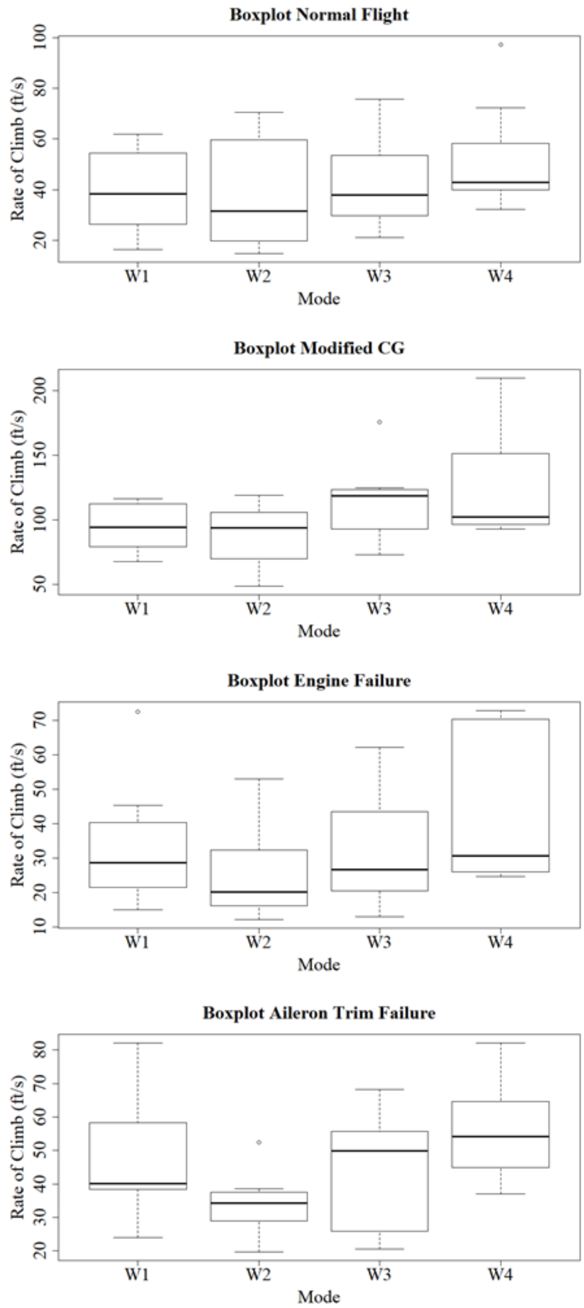

Fig. 11 - Rate of climb boxplots.

5 Conclusion

The results of the experiment shown an influence of the different levels of workload on the flight performance when variables which are closest to the piloting command are assessed.

The Command Intensity Indicator has not shown a significant result, in opposition to the initial idea.

Despite dealing with reduced sampling and simplified hardware, the experiment was able to contribute to the definition of future studies. It gives some hints on how the workload affects the variables registered in the aircraft FDR (Flight Data Record). It also confirmed that the pilot is a significant variable that affects flight performance, meaning that each pilot was able to keep the aircraft safe while having distinct actions.

Future studies regarding human factors and the workload influence shall use Physiological data collected with specialized sensors in addition to FDR data logs, enhancing the capability and efficiency of the experimentations, as well as the reliability of the experimental campaigns.

References

[1] CENIPA. (2015). *Ocorrências Aeronáuticas: Panorama Estatístico da viação Brasileira - 2006 -2015*(Rep.). Comando da Aeronáutica.

[2] Boeing (Ed.). (2016). *Statistical Summary of Commercial Jet Airplane Accidents: Worldwide Operations | 1959–2015*. Seattle, Washington: Aviation Safety Boeing Commercial Airplanes.

[3] Boeing. (2007). MEDA Investigation Process. *AERO*, 14-21.

[4] Dekker, S. (2017). *The Field Guide to Human Error Investigations*. Taylor & Francis.

[5] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology Human Mental Workload*,139-183. doi:10.1016/s0166-4115(08)62386-9

[6] Arjoni D. H. et al. Experimental evaluation of the human performance on a robotic flight simulator based on FOQA parameters. Proceedings of the Aerospace Technology Conference, 2016, Stockholm, v. 1. p. 1-11.

[7] Moreira, A. H. et al. Experimental evaluation of the contribution of adding a motion system to an EDS. Proceedings of the Aerospace Technology Conference, 2016, Stockholm, v. 1. p. 1-11.

[8] Arjoni D. H. et al. Artigo ICAS. Experimental analysis of flight performance under workload variations. The 31st Congress of the International Council of the Aeronautical Sciences, 2018, Belo Horizonte, v.1 p. 1-10.

[9] Endsley, M. R. (2013). Situation Awareness. *Oxford Handbooks Online*. doi:10.1093/oxfordhb/9780199757183.013.0006

[10] Endsley, M., & Jones, W. (1997). *Situation Awareness Information Dominance & Information Warfare*(pp. 1-94, Rep.). Springfield, Virginia.

[11] Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R., Jr. (2011). *The Multi-Attribute Task Battery II (MATB-II) Software for Human Perfomance and Workload Research: A User's Guide*(pp. 1-269, Rep.). NASA.

[12] Montgomery, D. C. (2013). *Design and analysis of experiments*. Hoboken: John Wiley & Sons.

Contact Author Email Address

Mail to: arjoni@ita.br; evillani@ita.br

Acknowledgements

This work was supported by CAPES, FAPESP (Process 2012/51085-3), FINEP/VINNOVA (Process 01.17.0038. 00), CNPq (Process 303271/2017-5).

Copyright Statement