



24th COBEM - 2017



24th ABCM International Congress of Mechanical Engineering
December 3-8, 2017, Curitiba, PR, Brazil

COBEM-2017-0727

EVOLUTION OF AN EXPERIMENT TO ASSESS PILOT BEHAVIOUR IN FLIGHT SIMULATORS

Diego Hernandez Arjoni

Emília Villani

ITA - Instituto Tecnológico de Aeronáutica Praça Marechal Eduardo Gomes, 50 - Vila das Acácias, São José dos Campos

arjoni@ita.br; evillani@ita.br

Aline Kraemer

ITA - Instituto Tecnológico de Aeronáutica Praça Marechal Eduardo Gomes, 50 - Vila das Acácias, São José dos Campos

aline.kraemer@gmail.com

Abstract.: *The SIVOR Project proposes the development of an anthropomorphic robot based flight simulator. An initial prototype was constructed to develop the motion cue, robot control algorithms and experimental tests based on human factor analysis, capable of verifying the pilot behaviour in the simulation. This work proposes an extension of the bibliographic review concerning flight simulators evaluation and validation, followed by an evolution of the first set of experiments to be applied in the SIVOR prototype simulator, composed by a simplified cockpit built on a robotic platform. The first experimental data analyses have shown some influence of the robotic motion platform in the behaviour of the pilots in comparison to a static simulation platform, to assess this results some methods of data analysis are improved. The first method was based on an adaptation of the FOQA program and defined a set of behavioural parameters to quantitatively assess the motion influence. The current method intends to filter the behavioural variables created based on the tests results and in the bibliographic review aiming to apply them to a more intense flight path based on adaptations of manoeuvres described on QTG and other regulations, considering mathematical model of an Embraer Phenom300, besides the analysis method considers a better and more robust computation of the data. The experiment responses do not follow the same responses of the initial campaigns showing no response on the flight behaviour.*

Keywords: *Pilot Behavior, Flight Simulator, Design of Experiments*

1. INTRODUCTION

In a flight, the aircraft movement is perceived by 3 sensorial systems of a human being (Kleiner; Schlittler; Sánchez-Arias, 2011; Gaerlan, 2010). The input signals are processed by the central nervous system in order to support the definition of the corresponding actuation (VEDA, 2016):

- Visual system: able to perceive motion through the image acquired and processed by the eyes using learned references such as wall, corner and straight lines (Lacquaniti et al., 2014);
- Proprioceptive system: composed of motion sensitive cells in the muscles and skin, able to perceive high frequencies and provide a self-sense of stability and position (Dougherty, 2016);
- Vestibular system: located at the inner ear, capable of perceiving linear accelerations and angular velocities (Vargas, 2009).

The motion platform of a flight simulator is responsible to stimulate primarily the vestibular system, which processes the aircraft accelerations and velocities. However, the simulator cannot just impose the same accelerations and velocities of the aircraft due to its limited workspace. A control routine, known as washout filter, is responsible to convert the aircraft movement in the simulator movement, considering its limitations (Vargas, 2009 apud Telban; Cardullo, 2005).

In order to allow the usage of flight simulators in pilot training, various aspects of the equipment are subjected to certification by qualified governmental institutions such as FAA and ANAC, including the motion cues. To assure the simulator performance, regulations and rules described in the CFR Part60 (FAA, 2016) are adopted. Although some the evaluation of the aircraft dynamic model is described in detail and analysed based on quantitative measures, when it

comes to the motion cues provided by the simulator, evaluation is still mainly performed based on subjective judgments. It depends on pilots' opinion about the realism of the flight in the simulator, when comparing to the flight in a real aircraft.

The study of the effects of the movement in flight simulators are relevant to both academy and industry (Serrand, 2007). Burki-Cohen et. al. (2011) presented an extensive review about feeling of motion in flight simulators, published in 2011. They organized the debate about the role of motion in flight simulators in four main aspects (Burki-Cohen, Sparko, & Bellman, 2011):

- How human perception responds to motion;
- The limitations imposed by the motion platform;
- The subjectivity of the pilot opinion;
- Empirical evidences on the necessity of motion for the transfer of knowledge and experience from the simulator to the airplane.

Although a great amount of research has been performed to give light to these issues, the benefits of the motion platform are still uncertain. Since the creation of motion flight simulators, the study regarding the need of the motion platforms has evolved from questionnaires and subjective analysis (Reid & Nahon, 1988) to the assessment of the human perceptual systems and the study of human factors and behavior analysis of the flight simulation (Telban; Cardullo & Kelly, 2005).

In this context, this work aims at measuring the influence of the motion system of a flight simulator on the pilot behaviour. For this purpose, we compare the results of a set of flights performed both with and without motion. The experiments were performed using the SIVOR prototype simulator. The SIVOR Project, a partnership between ITA and EMBRAER, consists in the development of a high-fidelity flight simulator based on an anthropomorphic robotic motion platform. The first SIVOR prototype was built to support the development of the motion controller and is illustrated in Fig. 1.

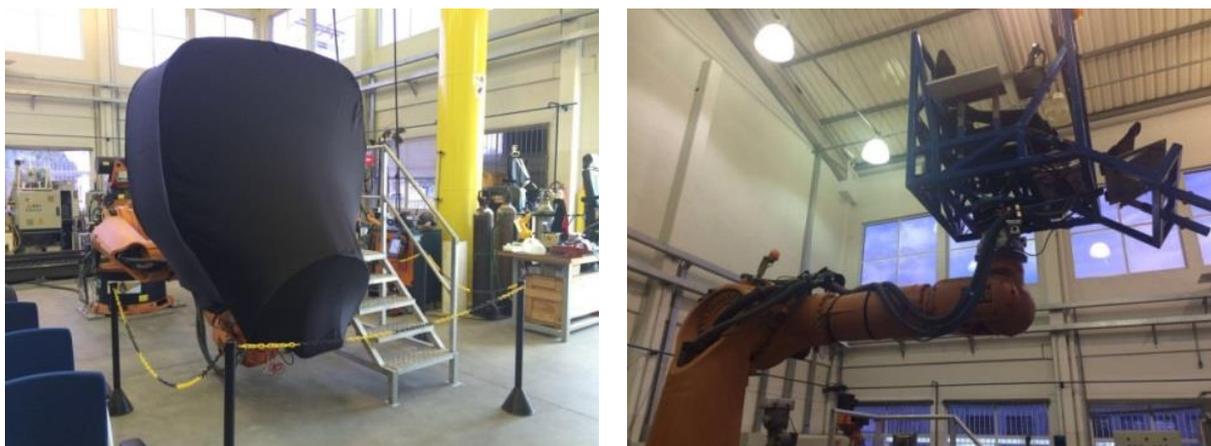


Figure 1. SIVOR prototype simulator.

Our analysis is based on three objective measurements that are inspired in the NASA TLX questionnaire [3]: Workload, Precision and Response Time.

In previous works (Arjoni et. al., 2016), (Moreira et. al., 2016), we used these measurements to analyse the results of two experiments. The first experiment (Arjoni et. al., 2016) focused on a simplified standard civil flight scenario. The FOQA (Flight Operational Quality Assurance) (FAA, 2004) regulation was adopted and modified to build a consistent flight path. Its events (logical association and algebra between collected flight parameters) were used to described by behavioural variables. Some flight segments were subjected to disturbances applied into the control surfaces with the objective of analysing the flight path with and without an increased stimulus. The results of the first campaign show that some of the proposed behavioural parameters, such as the pilot workload, are sensitive to the motion cues.

The results of first experiment lead us to propose a second one considering a different flight path with more complex manoeuvres to test both the aeronautical model limits with a stall recovering, and the behaviour of the pilot when dealing with more difficult procedures, such as landings and offset landings. The results were consistent with those of the first campaign, showing a significant difference between the dynamic simulated flight when analysing the workload measurement in the landing ad offset landing. Both flight campaigns were performed with only two modes of motion: static and dynamic. The purpose was to enhance our knowledge regarding parameters that could be used to evaluate realism of the motion system of a flight simulator, considering the pilot perspective.

Based on the results of the previous experiments, this work proposes a third experimental campaign, composed of high gain manoeuvres inspired on the QTG (Qualification Test Guide) (FAA, 2016), (London: Royal Aeronautical Society, 2005). The flight includes systemic failures of flight systems, such as engines and flaps. The tests are performed with 4 pilots using the same motion platform and a similar but rather improved washout filter. An improvement is adopted

in the behavioural variables calculation to avoid false results and to discard some limiting hypothesis such as the data consistency and the non-homogenous sampling of the first campaign. The results obtained in this experiment do not confirm the sensitiveness of workload measurements to the motion platform, but can be used to direct future studies towards human factors in flight simulation.

This paper is organized as follows. We present the description of the testing methodology, statistical models and data processing used in the current experiment. We, then discuss the results of the statistical analysis. Finally, we draw some conclusions and discuss future work.

2. THE EXPERIMENT

This chapter presents the design of experiment adopted in this paper. It describes the testing procedure, the statistical model and the data processing and analysis rules.

2.1 The flight path of the experiment

As discussed in the introduction, this work has been proposed based on the results obtained from two previous experiments. These experiments consisted of performing a standard flight (first experiment) and a set of maneuvers of great difficult (second experiment). The flights were performed both with and without operation of the motion system, called static and dynamic operation modes of the simulator. In both experiments, three different behavioral parameters were measured and compared: workload, precision and response time. The variables related to workload in the cockpit and precision of flight were found to be sufficiently sensitive to the operation mode of the simulator (static or dynamic).

In order to deepen the current research, in this work two maneuvers are proposed based on the QTG rules of testing. Both maneuvers are related to systemic failures, meaning that the pilot must deal with sudden and unwarned problems, creating a substantial and instantaneous amount of workload and stress.

Based on the analysis of Table A1B —Table of Tasks vs. Simulator Level of the 14 CFR Part 60 (FAA, 2016) two subjective assessments are chosen to evaluate the parameters: an engine failure during takeoff and a hydraulic system failure. The engine failure at takeoff was accomplished in a very straight forward way, with a controlled speed of fail. On the other hand, the hydraulic failure was chosen to be applied on only one flight surface, avoiding couplings. The flaps were chosen, once it has discrete values of operation, and has no continuously interaction with the pilot during flight, differently from control of the elevators, ailerons and rudders. These characteristics allow the pilot to manipulate the other surfaces having full control of them and without having to deal with external disturbances.

The chosen maneuvers are adapted to be included in a single mission and are illustrated by the Fig. 2

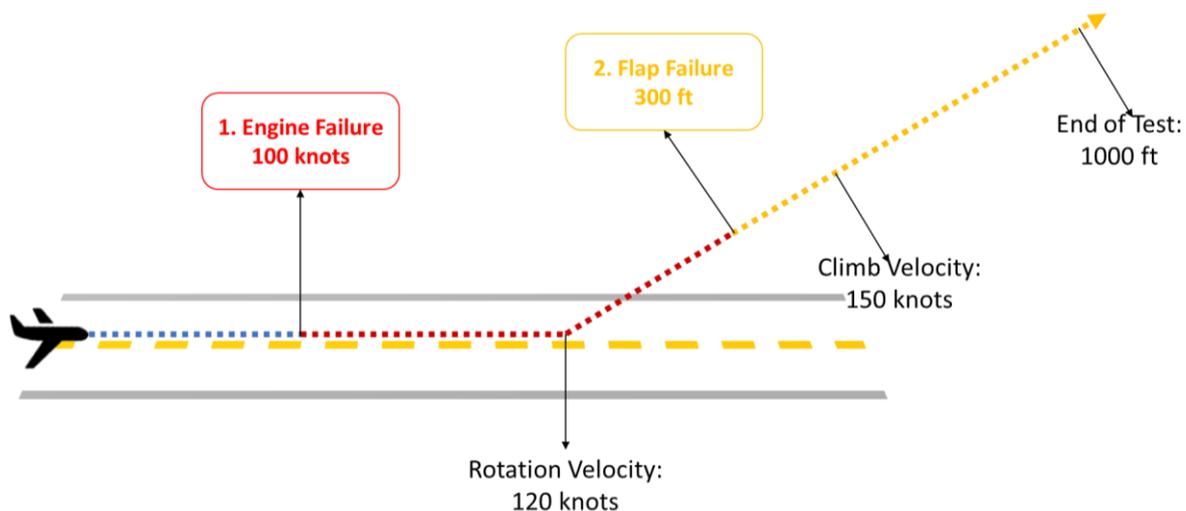


Figure 2. Flight mission.

The flow chart of Fig. 3 describes the instruction provided to the pilot for this experiment.

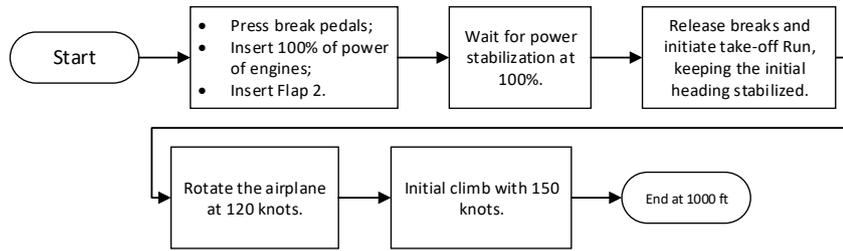


Figure 3. Flowchart of the mission.

At the specified points indicated in the Fig. 2, the failures 1 and 2 may or may not happen. In every test only one failure can happen, despite its occurrence, the mission is keep unaltered. Each type of test (no failure, engine failure and flap failure) must be repeated 3 times by each pilot at the two operation modes of the simulator (static and dynamic).

The failures injected in the simulator have the following behavior:

- Engine failure: at 100 knots, the right engine turns off and its power decrease to 0%;
- Flap failure: at 300 ft. the right flap retracts and the left flap keep locked at flap 2 without any control.

In both cases the pilot is instructed to never decrease the engine power nor retract flaps and landing gear.

The simulator runs using a Phenom 300 aerodynamic model available in public domain, as in the previous experimental campaigns. Even though, the sampling frequency is increased once the present maneuvers involve sudden change of behavior.

2.2 Statistical modeling and data collection

The proposed behavioral parameters are calculated through a series of variables collected in a data frame during the flights. Each data frame is composed of variables of attitude, altitude, speeds, joystick inputs, engine power, among others. Based on the collected data, a total of 6 parameters are calculated:

- Precision (mean and standard deviation) in engine failure: calculated as the heading mean and standard deviation, considering the data collected from the failure moment until it reaches 2000 samples at 80Hz sampling;
- Precision (mean and standard deviation) in flap failure: calculated as the heading mean and standard deviation, collected from the failure moment until it reaches 1200 samples at 80Hz sampling;
- Workload in engine failure: calculated as the sum of the absolute values of the aileron, rudder and elevator commands given by the pilot and collected from the failure moment until it reaches of 2000 samples at 80Hz sampling;
- Workload in flap failure: calculated as the sum of the absolute values of the aileron, rudder and elevator commands given by the pilot and collected from the failure moment until it reaches of 1200 samples at 80Hz sampling.

The flap failure sample is smaller than the engine failure to keep the consistency of the mission without enlarging the total time of one flight.

The data collection intervals are illustrated by the Figs. 4 and Fig. 5.

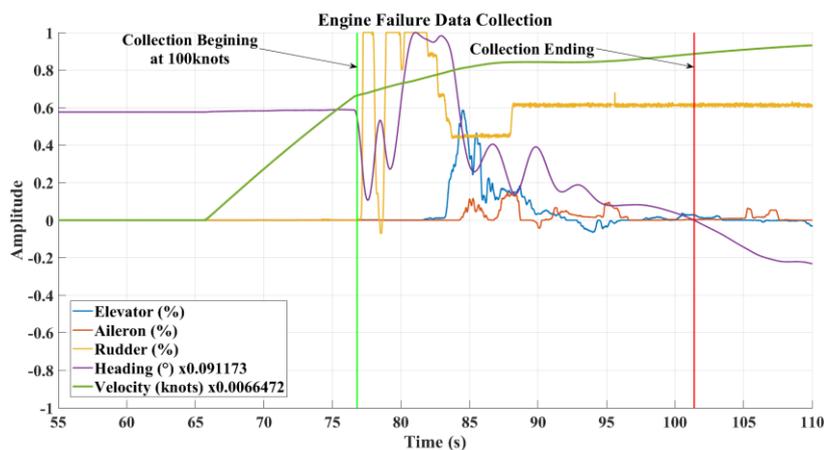


Figure 4. Example of data collect from a flight with flap failure.

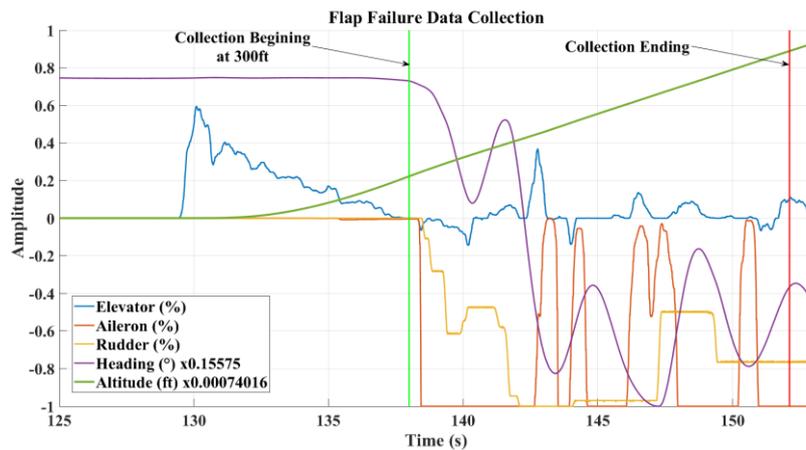


Figure 5. Example of data collect from a flight with engine failure.

The experiment was performed by 4 professional pilots: 3 of them were flight test pilots from the Brazilian Air Force, and 1 was a commercial airline pilot. All the pilots have more than 3000 hours of experience and are trained in more than 10 types of aircraft.

Each pilot performed each type of the three flight missions (no failure, flap failure or engine failure) six times: thrice in the dynamic mode and thrice in the static mode, totalizing 18 test runs. The flights without failure were introduced only with the purpose of preserving the surprise effect of the failures. Although alerted by the possibility of failures during the flight, the pilots did not know in advance when or whether or not the current flight would have any failure.

Each parameter is subjected to an Analysis of Variance (ANOVA) statistical test with a Randomized Complete Block Design (RCBD) composed of one factor, namely, “motion type” and one blocked factor, the pilot, so that its variance can be disconnected from the main factor under study. Equation 1 describes the statistical model.

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

where:

- V_{ij} : output value: mean, standard deviation or workload;
- μ : general output mean;
- M_i : variance of the motion mode;
- β_j : variance of the pilot block;
- e_{ij} : random error.

3. ANALYSIS AND RESULTS

This section presents the method used in the statistical analysis and the ANOVA test results (Montgomery, 2013).

3.1 Statistical analysis method

The statistical analysis is performed with 10% of significance once the test has a small sample of pilots. To verify the consistency of the data, it is subjected to two prior verifications:

- Shapiro-Wilk test: verifies the normality of the sample with a 5% significance. The null hypothesis states for the normality of the data, and the alternative hypothesis states for the non-normality;
- Bartlett test: verifies the equality of variances between the analysed samples with 5% of significance. The null hypothesis states for the equality of the variances (homoscedasticity), while the alternative hypothesis states for the non-equality of the variances.

Once the experiment was performed with humans, the data is subjected to great uncontrollable variances and may result in outliers. To deal with this problem, some rules are established to allow the correct removal of these points without harming the experiment reliability when the normality or the homoscedasticity of the data is not inside the purposed significance levels:

- Execution of a standardized residuals test as described by the Eq. 2

$$d_{ij} = \frac{e_{ij}}{\sqrt{MS_E}} \quad (2)$$

where:

d_{ij} : standardized residual;

e_{ij} : residual;

MS_E : Mean square error.

- Retreat of any value equal or higher than 3 standard deviations;
- Repeat the Shapiro-Wilk and the Bartlett test;
- If the significances are satisfied, apply the ANOVA, otherwise reapply the standardized residuals;
- Retreat on residual equal or great than 2 and repeat the Shapiro-Wilk and the Bartlett test;
- If the significances are satisfied present the results with and without the “outlier” otherwise the parameter is subjected to too much uncertainty and should be removed.

3.2 Results

We observe that for the flap failure, two data sets of Pilot 3 were lost due to a too lack of data samples. One of the lost data set was in static mode, while the other was in dynamic mode. We consider that the loss of this data does not affect significantly the results of the experiment.

The data boxplots are presented in the Fig. 6 for the engine failure and Fig. 7 for the flap failure.

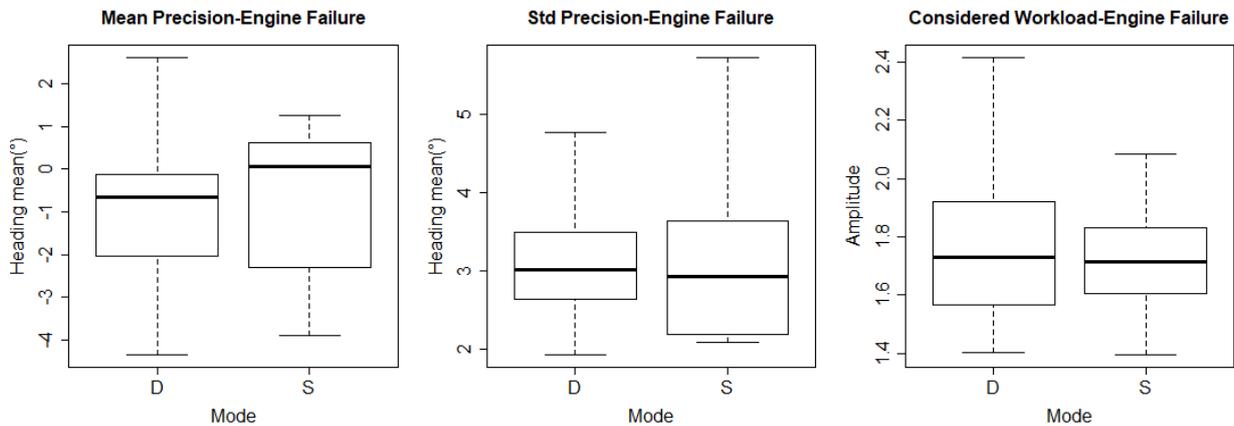


Figure 6. Boxplots of the engine failure results.

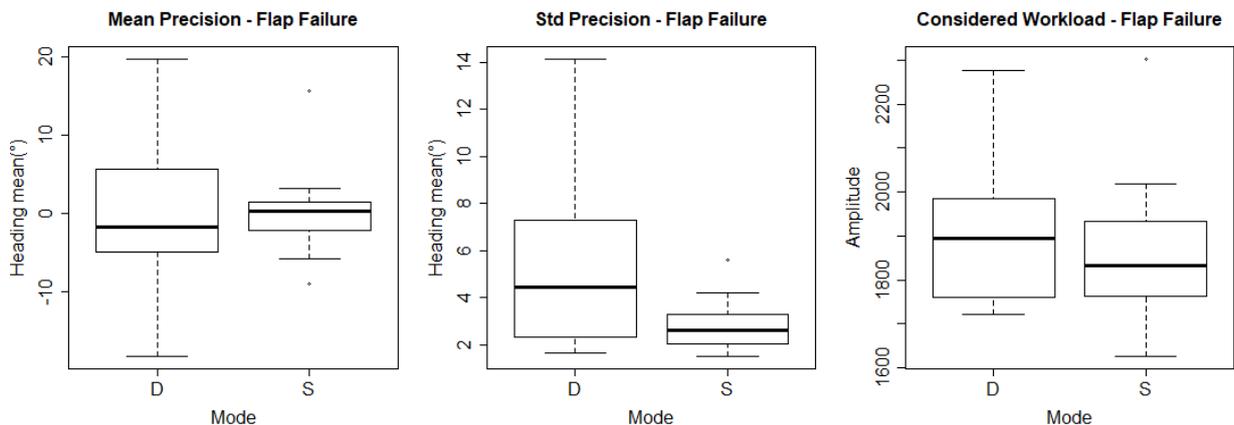


Figure 7. Boxplots of the flap failure results.

While analysing the standard deviation of the precision variable and the considered workload for the flights with flap failure, the normality and homoscedasticity hypothesis were not fulfilled even after the outlier retreat rules were applied, resulting in the exclusion of the parameters from the final results.

Tab. 1 resumes the final results obtained from the ANOVA analysis, accomplished with the software RStudio®.

Table 1. Results of the ANOVA analysis

Manoeuvre	Parameter	Outliers	ANOVA factor mode P - Value	Shapiro - Wilk P-Value	Bartlett for factor mode P-Value	Bartlett for block pilot P-Value	Mode Factor Influence
Engine Failure	Precision - Mean	0	0.68	0.31	0.70	0.20	Not Confirmed
	Precision - Std	0	0.79	0.37	0.49	0.09	Not Confirmed
	Considered Workload	0	0.47	0.67	0.21	0.78	Not Confirmed
Flap Failure	Precision - Mean	0	0.89	0.87	0.09	0.09	Not Confirmed
	Precision - Std	-	-	-	-	-	Excluded
	Considered Workload	-	-	-	-	-	Excluded

The ANOVA results do not confirm the influence of the motion platform in the workload variable. Additionally, it pointed that the precision parameters are not influenced by the motion system too, even in the case of difficult situations, such as under the occurrence of failures.

It is important to mention that the engine failure has a more realistic characteristic than the flap failure once this type of system defect has a low probability of occurrence and the asymmetry between the flaps is usually smaller than the near 20° imposed to the pilots. The unfamiliarity with the problem was intended to enhance the workload of the pilots once their attention and experience must be used to recover from an unknown and untrained danger.

4. CONCLUSIONS

This work contributes to the discussion of how to evaluate the motion platform of flight simulators in an objective way. It present the results of an experiment that evaluates the influence of the motion system in three proposed variables: workload, precision mean and precision standard deviation. The experiment consists on asking four pilots to perform a flight mission under three conditions: with no failure, with engine failure and with flap failures. The three flights were repeated with and without the activation of motion platform.

The results obtained were compared to those of two previous experiments, that were based on a normal flight and a flight with difficult manoeuvres. The previous experiments had showed sensitivity of the workload variable to the motion mode, and, in some cases, also the sensitivity of the precision variable.

This work does not confirm the previous results. This opposite effect can be directed to some enhancements in the measurements in order to improve robustness of the data and some variation on the data analysis to avoid data collection problems, besides, some limiting hypothesis were loosen, namely, the pilots sample had a better homogeneity.

Upcoming researches should improve the measurement of the workload by adopting other processing technics or even measuring physiological parameters related to stress, attention and workload, allowing the study to be more accurate and enhancing the reliability of the experimental data collected. Additionally, the importance of the experience of the pilots on the workload measurements should be took into account, experimental campaigns considering homogenous group samples of pilots and non-pilots with different degrees of experience will be able to show clearer results towards the behaviour changes in the different simulation modes

5. ACKNOWLEDGMENTS

This work was supported by FAPESP (Processes 2012/51085-3), CNPq (Processes 132801/2015-9 and 306408/2014-7) and CAPES.

6. REFERENCES

- Aeroplane Flight Simulator Evaluation Handbook: International Standards for the Qualification of Aeroplane Flight Simulators. (2005). London: Royal Aeronautical Society.
- Arjoni, D. H., Rocha, G. C., Moreira, A. H., Nicola, R. M., Oliveira, W. R., Silva, A. V. S., ... Trabasso, L. G. (2016). *Experimental Evaluation of the Human Performance on a RoboticFlight Simulator based on FOQA Parameters*. In Proceedings of the Aerospace Technology Congress (Vol. 1, pp. 1–11). Stockholm. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-132931>

- Burki-Cohen, J., Sparko, A., & Bellman, M. (2011). Flight Simulator Motion Literature Pertinent to Airline-Pilot Recurrent Training and Evaluation. *AIAA Modeling and Simulation Technologies Conference*. doi:10.2514/6.2011-6320
- Dougherty, P., Ph.D. (n.d.). Neuroscience Online. Retrieved May 16, 2016, from <http://neuroscience.uth.tmc.edu/s2/chapter02.html>
- FAA, Flight Simulation Training Device Initial and Continuing Qualification and Use 14 C.F.R. § 60, 2016
- Gaerlan, M. G. (2010). *The role of visual, vestibular, and somatosensory systems in postural balance* (Unpublished master's thesis). UNLV.
- Hart, S. G., Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index) Results of Empirical and Theoretical Research. *Human Mental Workload*, p. 139–183.
- Kleiner, A. R., Schlittler, D. X., & Sánchez-Arias, M. D. (2011). The role of visual, vestibular, somatosensory and auditory systems for the postural control. *Revista Neurociências*, 349-357.
- Lacquaniti, F., Bosco, G., Gravano, S., et al., “*Multisensory Integration and Internal Models for Sensing Gravity Effects in Primates*,” *BioMed Research International*, vol. 2014, Article ID 615854, 10 pages, 2014. doi:10.1155/2014/615854
- Montgomery, D. C., *Design and Analysis of Experiments*, 8th ed., John Wiley & Sons, Inc., Hoboken, New Jersey, USA, 2013, p.730.
- Moreira, A. H., Arjoni, D. H., Nicola, R. M., Silva, E. T., Villani, E., & Trabasso, L. G. (2016). *Experimental evaluation of the contribution of adding a motion system to an EDS*. In Proceedings of the Aerospace Technology Congress (Vol. 1, pp. 1–10). Stockholm. Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-132932>
- Reid, L. D., & Nahon, M. A. (1988). Response of airline pilots to variations in flight simulator motion algorithms. *Journal of Aircraft*, 25(7), 639-646. doi:10.2514/3.45635
- Serrand, A., 38., 2007, St. Louis. A380 Aircraft 0 Test Campaign. St. Louis: Airbus, 2007. 9 p.
- Telban, R.J., Cardullo, F.M., and Kelly, L.C. “Motion Cueing Algorithm Development: Piloted Performance Testing of the Cueing Algorithms”, NASA/CR-2005-213748, 2005.
- US DOT (2004): Advisory Circular: Flight Operations Quality Assurance (FOQA), AC No: 120-82.
- Vargas, M. B. (2009). *Controle de uma plataforma de movimento de um simulador de vôo* (Master's thesis, UFSCAR) (pp. 1-172). UFSCAR.
- VEDA (Ed.). (2016). *The Human Balance System - A Complex Coordination of Central and Peripheral Systems*.

7. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.