Pilot performance evaluation of simulated flight approach and landing manoeuvres using quantitative assessment tools

P ARCHANA HEBBAR and ABHAY A PASHILKAR*

CSIR-National Aerospace Laboratories, Bangalore 560017, India
e-mail: p_archana@nal.res.in; apash@nal.res.in

MS received 6 January 2016; revised 26 July 2016; accepted 27 September 2016

Abstract. This research work examines the application of different statistical and empirical analysis methods to quantify pilot performance. A realistic approach and landing flight scenario is executed using the reconfigurable flight simulator at National Aerospace Laboratories and both subjective and quantitative measures are applied to the pilot performance data. Simulations were repeated for different difficult landing conditions like landing with degraded visibility, with crosswinds, with degraded aircraft handling qualities and with emergency conditions. Relative assessment of the different applicable metrics is made and significance of task difficulties on pilot performance is investigated. Changes in the pilot's control strategy with respect to primary and secondary tasks are also discussed in detail. Results indicate that analysing pilot's control strategy together with his/her deviations from predetermined flight profile provides a means to quantify pilot performance.

Keywords. Pilot in the loop simulations; cognitive workload; tracking errors; control strategy.

1. Introduction

Any new aircraft technology, be it fly by wire control system design, advanced glass cockpit design or finalizing head up/down display symbologies, is developed with the primary intention to increase aircraft/pilot safety and to ease pilot to accomplish his/her task smoothly. The current practice is to evaluate the usability of the new technology/algorithm through pilot opinion. Although these commonly used subjective methodologies may be fast and are non-intrusive, the judgements rely highly on personal opinions, his/her mood and the surrounding environment at that instant. Hence these methods alone may not be always reliable. Use of quantitative tools and techniques along with the subjective ratings can give more meaningful judgements.

Further to subjective and performance based methods, pilot workload can also be measured through physiological measures. Research on assessment of mental stress using heart rate variability [1], respiration, electroencephalogram signals [2] and eye related signals [3–5] suggests that these physiological parameters do show signatures of cognitive workload when the subject is under stress. However, accuracy of judgement based on these methods basically depends on the measuring instruments. Also, these instruments have to be non-intrusive to pilot's primary flying tasks. In this paper, we restrict our study to subjective and performance based pilot evaluation only.

The realistic study discussed in this paper attempts to prove the concept of quantitative performance assessment. Piloted simulations are carried out using reconfigurable design flight simulator at National Aerospace Laboratories (NAL), Bangalore, India for real life flight task scenarios. The objective of the study is to suggest potential quantitative measures available in the literature to analyse pilot performance. These results are compared with pilot's subjective opinions. The implication of task difficulties on pilot's performance is studied. Also, changes in pilot's control strategy with respect to task priority are examined.

There are many research activities carried out on objective pilot performance evaluation. Ververs and Wickens [6] examined the benefits of display location; display clutter and display intensity on pilot performance in a cruise flight environment using simple statistical analysis methods like mean absolute error, root mean square error (RMSE) and response time delay. Experiments were conducted wherein pilots flew a flight segment while observing changes in heading, airspeed and altitude. Pilots were directed to press a button on the joystick when any randomly appearing external event is detected. Deviations from the flight path and response time to detect sudden events were measured using the statistical methods mentioned above. Experiments were repeated for different display locations, clutter and image contrast ratios.

Smith and Caldwell [7] used RMSE to study pilot fatigue. Their experiments were extensive and the flight profile consisted of many turns and climbs before performing an instrument landing. Each flight task was for 1 h duration wherein

*For correspondence

pilots had to maintain a flight path accuracy of 3° in target heading, 3 knots in target airspeed and 10 ft of the target altitude before each manoeuvre began. The study suggested the usage of objective metrics for measuring pilot's performance.

Hanson *et al* [8] assessed the usability of different adaptive control techniques by using pilot's handling qualities ratings for an air-to-air tracking manoeuvre. The ratings were computed for pitch stick deflections from two pilots while using different adaptive controllers and with simulated aircraft failures such as failed stabilator and reduced pitch damping failure. The metrics indicated that adaptive controllers indeed reduce pilot's workload in failure conditions. Field and Giese [9] applied RMSE and power spectral density (PSD) of pitch control column activity during handling qualities research investigations for a lateral offset approach and landing task. The emphasis in this study was to evaluate the changes in the handling qualities for the same aircraft configuration in three different simulators.

In a nutshell, it is seen that although many statistical measures are available in the literature, it is difficult to specify one common metric to measure performance because they are mainly context specific. Hence, a battery of metrics has to be evolved to assess the different performance dimensions. In this study, we attempt to apply all the above-mentioned metrics to evaluate pilot performance for a manual approach and landing task.

The study discussed in this paper deals with manual control alone and not with strategic or decision making levels of performance. The study addresses the effect of changes in the task difficulties on pilot's performance. Also, all the experiments conducted are with a single pilot flying different simulated flight conditions. This is considered to be sufficient because the results dwell on the applicability of the methodology and pilot's action towards threats and no conclusions are made regarding any pilot dependent comparisons. The assumption made in this analysis is that increase in task difficulty tends to reduce pilot's performance [10]. Task difficulty may be defined as the perception that the pilot has regarding the task (like how hard or easy it is to perform the task). This, in combination with personal characteristics such as experience, capabilities and his/her physical or psychological state and aircraft design determines pilot workload.

This paper begins with introducing the task under test and the simulation setup in section 2. Section 3 presents the workload score through NASA Task Load index (TLX). Section 4 assesses the validity of the quantitative metrics used in the study. Section 5 discusses the analysis results. Section 6 summarises the research, identifies the key contributions and suggests future directions.

2. Task scenario and simulation setup

It is known that take-off and landing phases are the highest accident prone regimes of any flight [11]. One such landing scenario is considered in this study. The flight task

commences on approach at 7 NM from the runway threshold, at an altitude of 5000 ft above mean sea level (MSL) and towards the runway bearing 88° . The initial approach speed is 130 knots. The total approach task is split into four stages as shown in figure 1.

- Maintain level flight up to a distance of 10 km from runway threshold with speeds close to 125 knots.
- Commence descent at 10 km away from runway threshold, maintaining speed close to 120 knots.
- At 5 km from touchdown, select flaps to 20° and continue for landing with speed maintained close to 115 knots.
- Continue approach with 20° flaps (reducing speed closer to 115 knots) and touchdown as planned on the dumbbell, with speeds close to 105 knots.

The task is repeated under different conditions set as

- Normal (basic setting with no changes introduced).
- Changed visibility conditions: The visibility range is restricted to below 1000 ft above ground level with the same approach sequence and conditions as detailed above. Pilot has to carry out manual instrument flying using display indicators for distance from runway, runway bearing, altitude, rate of descent with aircraft attitude and bank displayed on the primary flight display (PFD). Runway location was indicated on PFD by the ADF pointer set to the coordinates for the touchdown point.
- Changed wind conditions: Crosswinds from 360° heading and from 180° heading are simulated with a magnitude of 18 knots.
- Changed control input delays: A continuous delay of 0.1, 0.2 and 0.5 s is introduced in the elevator and aileron control input columns at the start of the simulation. Hence there is a delay in the pilot input and the response.
- Engine on fire emergency: One of the engines is simulated to be on fire at ~ 8 km from touchdown. Pilot needs to monitor the warning, immediately take actions as per emergency procedures – cut off power to that engine, shut down the engine, fire shot 1, monitor the warning, if the warning is still present fire shot 2, while continuing to control the aircraft for a safe landing.

Figure 1. Task scenario under test.

Figure 2. Simulation setup.

NAL's reconfigurable design flight simulator is used for conducting the pilot in the loop simulations [12]. The simulator design is based on model-based approach. Hence, it has the advantage that any additions or modifications in the flight dynamic model, force feel characteristics of the pilot controls or updates in the various displays can be made quickly. Flight dynamic model of a twin-engine light transport aircraft is used in the study. Figure 2 shows the simulator setup.

A highly experienced military pilot carried out the experiments. The pilot has experience in flying twenty five different types of aircraft and has over 3400 flight hours of flying. After initial briefing to the pilot and readiness of the data acquisition setup inside cockpit, pilot flew around several nominal approaches to get acquainted with the simulator. Twenty simulations were executed including all the test scenarios discussed above. Eight nominal cases, two with low visibility conditions, six with windy conditions, four with control input delays, four with engine on fire, and four with low visibility conditions along with control input delay. All simulations were conducted over a span of ten days.

Even though simulator is for two-pilot (pilot, co-pilot) aircraft, simulations are carried out only by one main pilot. Hence, the situation creates more demand on the pilot as he/she is responsible for performing both primary flying task of aircraft control and secondary tasks like monitoring and decision making.

3. Subjective workload assessment

According to NASA TLX, the main factors or sources contributing to task load are physical demand, mental demand, temporal demand, frustration, effort and performance [13]. It is a two-part evaluation procedure consisting of weights and ratings. Procedure to compute NASA TLX workload score is

- Evaluate the contribution of each factor to the workload of a specific task.

- Next, obtain numerical ratings for each scale (from 0 to 100) that reflects the magnitude of that factor in a given task.

Hence, with six sub-scales, there are 15 possible pairwise comparisons. Workload is then computed as

$$Workload = \frac{\sum_{i=1}^6 w_i r_i}{15} \quad (1)$$

where w_i is the weight and r_i is the rating for the i th task load source.

NASA TLX version 1.0 questionnaire was administered to pilot after every simulation. Weights and ratings for each task load component were provided by the pilot for all the test scenarios. Workload score is then computed correspondingly. Table 1 shows the subjective judgements by the pilot for all the test scenarios, along with his comments. As per pilot's opinion, landing in nominal conditions was the easiest with a workload score 19. On the other hand, one engine on fire task caused maximum workload (workload score of 35.6).

4. Quantitative analysis

The following analysis methods are used in this study:

4.1 Measurement of pilot control strategy

At the lowest level, analysing the time history of aircraft control surfaces like aileron, elevator, rudder and engine power provides a basic insight into the way pilot reacts to the situation.

Plotting pilot's aggressiveness levels on the inceptor indicates the rate at which pilot changes his inputs on the stick. One such available metric is called the pilot inceptor workload (PIW) plot [14]. PIW plot is a plot of aggressiveness versus duty cycle. Aggressiveness is defined as the rate of change of movements on the control columns. Aggressiveness is computed as

$$Aggressiveness = \sqrt{\frac{1}{n-1} \sum_{i=2}^n \left(\frac{\delta_i - \delta_{i-1}}{t_i - t_{i-1}} \right)^2} \quad (2)$$

Duty cycle is the percentage of time pilot changes his input on the control column. Duty cycle is given as

$$Duty Cycle = 100\% * \frac{1}{t_n - t_2} \sum_{i=2}^n x_i$$

$$\text{where } x_i = \begin{cases} 0 & \text{for } \frac{\delta_i - \delta_{i-1}}{t_i - t_{i-1}} < \text{noise threshold and } |\delta_i| < \delta_{max} \\ 1 & \text{otherwise} \end{cases} \quad (3)$$

t_2 is the start time +1 and t_n is the end time of the data set; n is the number of data points; δ_i are the discrete values of

Table 1. Results of NASA TLX questionnaire.

Sl. no.	Task details	Workload score	Pilot remarks
1	Normal approach and landing	19	–
2	Approach and landing tasks with crosswinds	28	–
3	Approach and landing under low visibility conditions	29	Manual instrument flying using limited navigation aids.
4	Approach and landing task with varying control input delays	29.667	Delay causes the pilot to respond more than that is required and hence give inputs more number of times to compensate for it.
5	Approach and landing task with one engine fire simulated at 8 km from touchdown	35.667	Practise, training and familiarity with the cockpit environment will impact smooth execution of the task and could impact the workload assessment.

the stick deflections in mm and δ_{max} is the maximum stick deflection. Noise threshold is computed based on pilot's stick movements falling in the 5th percentile of rate of change of inceptor's displacement. This is to assure that threshold value taken is such that very small signals from inceptor that can be regarded as involuntary stick movements are avoided. Hence x_i is 1 whenever the rate of change of inceptor deflection is within the top 95 percentile of the stick movements and stick deflection is within its maximum range.

Figure 3 shows the PIW plot. Increasing duty cycle and aggressiveness increases workload.

Hence PIW is a measure of pilot's effort. Workload increases if both the parameters increase.

Pilot's actions on the inceptor can also be analysed using frequency domain metrics such as PSD and autocorrelation. Dominant frequencies of control stick power distribution are computed using Fast Fourier transform ($X(n)$) on time series data $x(m)$.

$$X(n) = \sum_{m=1}^N x(m) e^{-\left(\frac{2\pi i}{N}\right)(m-1)(n-1)} \quad \text{where } i = 1\sqrt{-1}. \quad (4)$$

PSD is then given by $\frac{|X(n)|^2}{N}$.

Low frequency, low amplitude inputs imply that pilot uses less energy to achieve the desired results [15].

Figure 3. PIW plot.

Figure 4 shows the PSD plots for pitch and roll control columns for one nominal case. Visual inspection of the spectral plots of pitch and roll control columns show 0.25 Hz to be the common cut off frequency. This frequency range is further divided into five equally spaced frequency bands – 0–0.05 Hz, 0.05–0.1 Hz, 0.1–0.15 Hz, 0.15–0.1 Hz and 0.2–0.25 Hz. Frequencies of operation of pitch and roll pilot control sticks for nominal test cases are analysed with respect to these bands. Analysis results using PSD metric is discussed in section 5.2.

The additional information gained using these metrics when compared to PIW is an understanding of the power applied on the inceptor to accomplish the task. For example if pilot needs to hold the aircraft attitude (in pitch or roll) for long duration, he may not apply frequent stick movement, but he uses sufficient power to maintain minimum errors. The theory behind using this metric is although it is desirable that the pilot should achieve key flight path with accuracy, it is undesirable that the pilot exhausts more physical and cognitive energy to accomplish the task.

The basic drawback of both of the metrics discussed above is that they do not show the distribution of control stick deflections with respect to flight time. As the task progresses, pilot's control input behaviour may change due to sudden disturbances or due to any system malfunctions. This cannot be captured if we compute power versus frequency relationship averaged over the entire time history. In such situations, a three-dimensional metric needs to be used wherein pilot control power versus both frequency and time is assessed. This objective can be achieved by using power frequency metric [16]. Power frequency is computed as follows:

- Control stick position data is divided into small enough segments where these segments are assumed to be stationary. Short period Fourier transform (STFT) is computed for this data $x(m)$ as

Figure 4. PSD plots for pitch and roll control column deflections.

$$X(n, \omega) = \sum_{m=1}^N x(m)W(n-m)e^{-\left(\frac{2\pi i}{N}\right)(m-1)(n-1)} \quad (5)$$

where $W(n-m)$ is the windowing function and hence $x(m)W(n-m)$ represent a short time section of $x(m)$.

- Half power frequency of the absolute power of the above data $\left(\frac{|X(n, \omega)|^2}{N}\right)$ is calculated for each time segment. This shall be the time varying cut off frequency. As the time varying cut off frequency gives the power at a given time instant, the magnitude of power at that time relative to power at all other times is lost. To solve this, cut off frequency at time t_i is multiplied by the maximum of the power over the entire frequency range at time t_i .

$$PF(t_i) = \omega_{\text{cutoff}}(t_i) * \max(\text{PSD}(t_i)). \quad (6)$$

As can be seen from the explanation above, STFT is used to compute the time frequency representation for the control input data. A window length of 1024 is chosen based on assumption that the input signal is stationary during this period. The basic constraint with this transformation is the selection of the windowing function. Too narrow window results in good time resolution, but poor frequency resolution and vice versa. As a result, an alternative approach of using continuous wavelet transform is also attempted and the results are compared.

Figure 5(a) shows the normalised pitch and roll control column power versus frequency versus time graph for a nominal landing scenario computed using STFT. Figure 5(b) shows the same when computed using wavelet transform. It can be seen that the peaks in the amplitudes with respect to time and frequency correlate with each other in both the cases. However, wavelet transform gives better resolution in frequency. As a result, the power frequency will also be more accurate. Hence power frequency analysis is performed using wavelet transform only (section 5.1). The inference from figure 5 is that the control column power in pitch axis shows a peak in the control stick usage

between 80 and 110 s of simulation time. This is the descent phase wherein flaps are lowered to reduce airspeed.

It is noted here that all the above-mentioned metrics should be analysed with respect to the errors in terms of deviations from the specified flight path.

4.2 Flight path tracking accuracy metrics

The main flight parameters under consideration in this exercise are airspeed, altitude, rate of descent, heading, latitude and longitude at the point of touchdown. The allowable tolerances from Director General of Civil aviation (DGCA) advisory circular [17], from CAA standards [18], and from pilot comments are tabulated in table 2. Some of the statistical tools such as RMSE, percentage of time spent outside tolerance intervals and number of deviations outside tolerance intervals are used to determine the deviations from the specified tolerance limits.

One way analysis of variance (ANOVA) with 0.05 level of significance is used as the statistical analysis tool for comparing the variance of the following:

- Dependency on pilot's errors related to different flight conditions: statistical significance of the variance of different parameters like airspeed and heading with respect to different flight conditions.
- To find whether there is any dependency related to pilot's control strategy with respect to the control column power in different frequency bands.

5. Analysis results

Main emphasis of this study was to determine the changes in pilot's performance levels with task difficulties introduced in a simulated environment. Degradation in pilot's control strategy due to distractions caused by secondary tasks is also analysed.

Figure 5. Contour plots of power v/s frequency with respect to time for one case. (a) Using short period Fourier transform. (b) Using wavelet transform

Table 2. Tolerance limits for different flight segments.

Flight segment	Tolerance limits	Remarks
Straight and level flight	Altitude – 1500 m AGL $\pm$ 30:28 m ($\pm$ 100 ft) Heading – 88 ± 5 deg Airspeed – 125 ± 10 knots	CAA standards Document 24
Initial descent	Bank angle – < 30 deg Approach speed = 120 ± 10 knots Rate of descent < 720 ft/min or 3.6728 m/s.	DGCA advisory circular dated April 2009 and pilot recommendations
Final descent	Bank angle – < 30 deg Approach speed = 115 ± 10 knots Rate of descent < 720 ft/min or 3.6728 m/s	DGCA advisory circular dated April 2009 and pilot recommendations
Landing phase	Sink rate at touchdown < 7 ft/s (2.134 m/s) Touchdown zone: longitudinal touchdown earlier than a point on the runway 200 ft or beyond 2700 ft from threshold. Lateral touchdown with the outboard landing gear more than 70 ft. from runway centreline.	FAA advisory circular [19] and DGCA circular

Basic analysis of simulated flight segments for all the scenarios shows that rate of descent, sink rate at touchdown and touchdown zone are maintained within tolerance limits for all the simulations. Hence these parameters are not considered for further discussions and comparisons.

5.1 Changes in pilot performance with task difficulty

Data from all the scenarios mentioned in section 2 are analysed to determine the level of difficulty. Segment P1–P5 are nominal cases, P6–P9 are with crosswinds, P10–P12

Figure 6. Tracking accuracy plots for landing phase.

are with low visibility conditions, P13–P15 are with control column delay, P16–P18 are engine on fire and P19–P20 are low visibility with delay in control column response.

Figure 6 shows the pilot's tracking accuracies for all the segments during descent phase of flight. It can be seen from the RMS errors of airspeed (ANOVA: $[F(4,10) = 1.32, \rho < 0.3274, \eta^2 = 0.3456]$) that there is no significant difference in maintaining desirable speeds even in adverse conditions. However, significant effects are achieved for error in heading $[F(4,10) = 5.86, \rho < 0.0108, \eta^2 = 0.7011]$ at 0.05 level of significance. Further comparisons to know which pairs of means are different (using multiple comparison tests) reveal that means of simulations with engine on fire condition are significantly different from others. Errors in all other conditions are comparable. Rate of descent and the bank angle are also well within tolerance. Hence it can be stated that pilot has tried with his utmost capabilities to maintain flight path even in high workload conditions.

Figure 7 shows the PIW plots for all test cases. It can be seen that pilot uses least levels of aggressiveness along with low duty cycles for nominal case. Nevertheless, in simulations in low visibility conditions and with control column input delays, pilot is initially aggressive on his inputs; but commands with finer inputs subsequently. Engine on fire cases show highest workload score. This is in agreement with pilot's subjective feedback as stated in table 1.

Figure 7 shows that engine on fire case shows highest level of aggressiveness in both pitch and roll control

Figure 7. PIW plots for all test cases.

columns. It can be seen from figure 8 that power frequencies on control stick for both the axes increase after pilot recognises the engine fire warning on the head down display. Here P1–P4 are nominal landing scenario and P5–P8 are with engine on fire. The subsequent actions taken by pilot like cut off the corresponding engine, fire shot, monitor, and so on are the secondary tasks which utilises pilot's mental resources. This influences the physical activity on the pitch control column. This can be improved if pilot is more acquainted with the cockpit environment.

Figure 8. Control stick responses for descent segment with engine on fire condition. (a) Power frequency plots computed using wavelet scalograms. (b) Time instants of engine on fire.

Hence it can be concluded from the analysis that even though pilots maintain good tracking performance even in difficult situations, their workload levels can be deduced through the control column deflections.

5.2 Difference in control strategy with respect to priority of task

Intent of a landing scenario is to maintain desired altitude and velocity during different phases of flight and to

touch down at the designated zone with minimum error. Also, pilot needs to fly wings level during the entire flight task. Hence it is a dual axis control task wherein both longitudinal and lateral axes have to be monitored and maintained. In this section, we analyse as to how the task is prioritized by the pilot to achieve the desired objective.

Figure 9 shows the error plots in the straight and level segment for nominal case simulations. Airspeed, altitude and heading deviations illustrated in the figure shows that there is mostly a compromise in the tracking performance in either longitudinal or lateral axis. If P4 has maintained airspeed, his deviations with respect to altitude and heading are more. On the same lines, P7 shows good tracking accuracy in airspeed and heading, but tracking in altitude is poor.

Analysis of pilot's control strategy for pitch and roll control columns can give more insight into pilot's control behaviour. Figure 10 shows that the dominant frequencies of operation of control stick for pitch and roll columns are different. There is significant difference in the distribution of power used for pitch control column deflections with respect to the different frequency bands (table 3). Power in lowest frequency band is significantly higher than those of the other four groups. More power is given for pitch control at very slower rate. Hence anticipation of control in pitch axis is good.

However, power used for roll control stick deflection is distributed throughout the frequencies of interest. It is seen that roll stick deflections are more prominent in the frequency ranges of 0.1–0.2 Hz. That is, small deflections are given more frequently. This is because although the task demands more concentration on pitch axis, constant roll corrections are given periodically to maintain the aircraft in wings level condition.

Hence longitudinal tasks are more demanding than the roll manoeuvres. Quantitative analysis shows that:

Figure 9. Tracking errors-straight and level segment.

Figure 10. PSD plots for all nominal simulation data.

- Pilot adaptively distinguishes the higher priority task based on his experience.
- Pilot has good anticipation of control on the higher priority tasks. The lower priority tasks are generally controlled through finer control commands.

6. Conclusions and future work

The study investigated the various means to explore quantitative methods for evaluating pilot performance. We analysed five different scenarios as mentioned namely nominal manual approach and landing, landing in low visibility conditions, landing with crosswinds, with delay introduced in elevator and aileron control input columns and an emergency landing scenario with engine on fire condition. The focus was on analysing changes in pilot's performance levels with variations in task difficulty. Subjective ratings by pilot that were computed using the NASA TLX workload score were extremely useful in this regard.

Analysis was primarily based on quantifying pilot's control strategy and comparing with his flight path maintenance. Standard statistical metrics such as RMSE were used to summarize the overall error and specify the deviations in the intended flight path. Such metrics can be applied for event analysis. That is, they are fixed based on the scenario. Also, it is proven in section 5.1 that major percentage of pilot's cognitive energy during demanding situations is expended in controlling the pilot control columns. In this regard, metrics such as PIW, PSD and power frequency plots are extremely informative. All these metrics have their own limitations and hence have to be selected carefully based on the analysis requirements. For example, PIW plot quantifies pilot's effort levels on the control columns, but fails to provide an understanding of the frequency and amplitude levels of power applied on the control columns. The amount of power applied on the control column is an indication of the cognitive energy exhausted by the pilot to accomplish the task. This is provided by the PSD plots. Also, both these metrics fail to show changes in the pilot's control strategy with respect to time. This is given by the power frequency metric. Hence,

Table 3. ANOVA results for different frequency bands.

Control column	ANOVA results	Remarks	Comments
Pitch control column	$F(4,6) = 11.56$, $\rho < 8.47e^{-6}$, $\eta^2 = 0.6065$	Mean of power in 0–0.05 Hz is significantly higher than the others.	Good anticipation of control. But higher workload.
Roll control column	$F(4,6) = 4.95$, $\rho < 0.0035$, $\eta^2 = 0.3975$	Significant difference in the means of all the 5 frequency bands. Pair wise comparisons show that stick control power is significantly higher in between 0.1 and 0.2 Hz than in other bands.	As the task is felt secondary, roll corrections are made at a higher rate.
Throttle	$F(4,6) = 53.92$, $\rho < 2.83 e^{-13}$, $\eta^2 = 0.8779$	Distribution of power is significantly higher in 0–0.05 Hz band.	Good anticipation of control. But higher workload.

power frequency metric is useful whenever changes in pilot's relative effort as the task progresses need to be determined. Case studies discussed in this paper show the emphasis of the choice of different metrics.

The analysis results of this study also point to the following conclusions:

- It is found that the conclusion derived through performance metrics is in line with the pilot's subjective ratings. Both the results suggest that tasks to be accomplished with one engine on fire during landing caused maximum workload when compared with rest of the scenarios.
- Pilot's performance in terms of tracking accuracy does not change abruptly with task difficulty. Pilots try, as long as possible, to maintain their performance by increasing their effort. However, the effort required to achieve this objective increases significantly with the perceived task difficulty. Hence control strategy measurement, together with tracking errors can provide a meaningful indication of the task demands.
- Pilot has good anticipation of control on the higher priority tasks. The lower priority tasks are generally controlled through finer control commands.

Initial analysis results show a direct proportional relationship of task difficulty with pilot's control behaviour. However, the study presented herein is restricted to a single pilot conducting all the tasks. More simulation exercises in different task conditions and with more number of pilots need to be conducted in future to investigate workload features in depth. Also, physiological measurement methods such as eye gaze, EEG and heart rate variability need to be explored.

Acknowledgements

The work is carried out under a project sanctioned by Aeronautical Research and Development Board (ARDB). The authors sincerely thank the test pilot, Air Cmdr (Retd) Jose Mathappan, for supporting this study with his valuable comments and support for data collection.

Nomenclature

δ_i	stick deflection
ω_{cutoff}	cut off frequency
ρ	level of significance
t_n	flight time
$X(n)$	Fourier coefficients of stick data
$X(m)$	time series of stick data
PF	power frequency
η^2	effect size

References

- [1] Bonner M A and Wilson G F 2001 Heart rate measures of flight test and evaluation. *Int. J. Aviat. Psychol.* 12(1): 63–77
- [2] Petkar H, Dande S and Yadav R 2009 A pilot study to assess designer's mental stress using eye gaze system and electroencephalogram. In: *Proceedings of ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 899–901
- [3] Anders G 2001 Pilot's attention allocation during approach and landing: eye and head tracking research in an A330 full flight simulator. In: *11th International Symposium on Aviation psychology (ISAP)*, Columbus, OH
- [4] Biswas P and Langdon P 2015 Multimodal intelligent eye-gaze tracking system. *Int. J. Human Comput. Interact.* 31(4): 277–294, doi:[10.1080/10447318.2014.1001301](https://doi.org/10.1080/10447318.2014.1001301)
- [5] Kramer A F 1990 Physiological metrics of mental workload: a review of recent progress. *Navy Personnel Research and Development Centre*
- [6] Ververs P May and Wickens C D 1998 Head up displays: effect of clutter, display intensity and display location of pilot performance. *Int. J. Aviat. Psychol.* 8(4): 377–403
- [7] Smith J K and Caldwell J A 2004 Methodology for evaluating the simulator flight performance of pilots. Brooks City Base, TX: Air Force Research Laboratory. Report No, AFRL-HE-BR-TR-2004-0118
- [8] Hanson C, Schaefer J, Burken J J, Larson D and Johnson M 2014 Complexity and pilot workload metrics for the evaluation of adaptive flight controls on a full scale piloted aircraft. *NASA Dryden Flight Research Center*; Edwards, CA United States. Document ID. 20140005730
- [9] Field E J and Giese S E D 2005 Appraisal of several pilot control activity measures. In: *Proceedings of AIAA Atmospheric Flight mechanics conference and exhibit*, 15–18 August 2005, San Francisco, California
- [10] Watson R W, Ntuen C and Park E 1996 Effects of task difficulty on pilot workload. In: *Proceedings of the 19th International Conference on Computers and Industrial Engineering*. 31(1–2): 487–490
- [11] Boeing 2014 *Statistical summary of commercial jet airplane accidents (1959–2014)*. Washington, USA: Aviation Safety
- [12] Pashilkar A A 2014 Trends in simulation technologies for aircraft design. *J. Aerosp. Sci. Technol.* 22(1): 1–10
- [13] Hart S G 1986 NASA Task load Index (TLX). Volume 1.0; Paper and pencil package. *Human performance research group, NASA Ames Research Centre*. Document ID. 20000021488
- [14] Shepherd M J, MacDonald A, Gray W R and Cobb R G 2009 Limited simulator aircraft handling qualities evaluation of an adaptive controller. In: *Proceedings of IEEE Aerospace Conference Paper 1292*. doi:[10.1109/AERO.2009.4839613](https://doi.org/10.1109/AERO.2009.4839613)
- [15] Ebbatson M, Huddleston J, Harris D and Sears R 2007 The application of frequency analysis based performance measures as an adjunct to flight path derived measures of pilot performance. *Hum. Factors Aerosp. Saf.* 6(4): 383–394
- [16] Lampton A and Klyde D H 2012 Power frequency: a metric for analyzing pilot in the loop flying tasks. *J. Guidance Control Dyn.* 35(5): 1526–1537

- [17] Operations circular 9. 2011 Pilot workload, AV.22024/8/2010-FSD. *Technical report*, Director General of Civil Aviation
- [18] Version 05 Standards document 24 2005 Guidance to examiners: multi-pilot aeroplanes type rating skill tests and proficiency checks. *Technical report*. Civil Aviation Authority
- [19] FAA Advisory circular - AC120-28D. 1999. *Technical report*, Federal Aviation Administration