A TOOL FOR PILOT'S PERFORMANCE AND ENGAGEMENT ASSESSMENT IN HELICOPTER FLIGHT SIMULATOR

Antoni Kopyt akopyt@meil.pw.edu.pl, Warsaw University of Technology (Poland)

Paulina Tomaszewska, paulina.tomaszewska3.stud@pw.edu.pl, Warsaw University of Technology (Poland)

Abstract

The accurate training procedures for future pilots are essential. There is need for objective performance and engagement assessment method. Only the combination of these two factors gives a full picture of pilot's behaviour during the flight. This is the idea behind the general Objective Assessment Tool (gOAT) system. In the paper we present a tool for pilot's performance assessment and the obtained results. One of the main advantages of the proposed method is high resolution enabling detailed, quantitative comparison of different flights. The solution combines many diverse criteria (varying largely in the range of obtained values) in order to generate one, scalar, easy comparable mark. The presented method was evaluated on data collected in helicopter flight simulator while performing the slalom Mission Task Element (MTE) adopted from ADS 33 [7]. The results analysis confirmed that the algorithm can accurately distinguish between the best and worst flights in terms of their handling quality, also appropriately points flights with similar performance. Thus it opened a path to create a fully objective system for detailed training assessment and pilot's progress analysis. The advantage of the system is that its applications are not only limited to helicopter flight simulators. It could be also used in UAVs or mobile robots.

1. INTRODUCTION

1.1. 1.1.Problem outlined

The crucial part of pilot training is performed on flight simulators. They have become integral to the rotorcraft manufacturing, training and research community [8]. Use of flight simulators is cheaper than trainings on real aircraft and enables training of border-line situations [3]. There are many researches aiming to develop the reliability and effectiveness of trainings in flight simulators. As mentioned in [4] the main directions of simulators' modifications is introducing motion platforms and making radio/display communication more realistic. The authors propose another approach of development of simulator training. The proposed general Objective Assessment Tool (gOAT) system has two components – for performance and engagement assessment separately. The performance component is calculated based on flight parameters like: position, velocity, acceleration. Whereas in engagement component biological signals like EEG, ECG could be used conveying information whether pilot is stressed.

The proposed objective rating scale would enhance functionalities of rotorcraft flight simulators, by giving the pilot feedback about his/hers handling quality, therefore leading to better quality of pilot trainings. The feedback is also provided for the instructor, so the pilot's progress is monitored in more details enabling more personalized training responding to trainee's needs.

The main source of information about the performance error is a feedback loop. The error is a measure how real signal (representing a flight) is dissimilar to the reference (target) signal which was established a priori. For this purpose, the set of eight/ten (ten - in case of acceleration signal) metrics were chosen.

1.2. 1.2.Previous works

There are many rating scales available dedicated to assessment of rotorcraft's handling quality, both subjective and objective. One of the most representative examples of subjective approach is Cooper Harper (C-H) scale based on decision tree structure [5]. At the beginning the method was used to give opinion about particular aircraft model (in terms of its manoeuvrability) in order to decide whether it needs any technical modifications. The aim of using C-H has however evaluated and now it also serves to asses helicopters' handling qualities. The following convention was used in C H: the bigger the mark value, the worst the flight was. The set of possible marks in the scale is limited. The modification to the Cooper-Harper scale (MCH-UVD) was later introduced [2]. It took into consideration also the parameters display quality. The objective approach to flight quality assessment is represented in many methods. Some elements can be found in ADS-33 norm, where depending on visibility level different critical parameters were defined in order to classify

Presented at 45th European Rotorcraft Forum, Warsaw, Poland, 17-20 September, 2019

flight of performing specific manoeuvre as desirable or acceptable. The existing popular methods use only one metric to asses error. Also, they enable only quantified scoring which means the set of possible grades is limited to particular number, whereas proposed method has infinite number of permissible values. Currently available methods to assess handling quality are limited and therefore the authors propose new solution. The objective rating scales are fairly easy to be applied. During the test flight, there are some requirements for each maneuver, so the pilot needs to follow them. Those parameters are in most cases fixed values (i.e. constant altitude).

1.3. Current training methods

Among the existing training methods still the instructor's opinion is the most important factor. Regarding the tasks and maneuvers where the result is clearly 0/1 choice (landed/not landed, acknowledge the malfunction, etc.) the majority of maneuvers are assessed by a subjective instructor's rate. Clearly there are some maneuvers with some flight parameters being assessed but if so, its number is very limited. The training procedures extorts the exchange between the instructors to avoid the one-man rate. Despite the professional skills of the instructors such a way of assessing the pilot's skills can lead to false conclusions. The mentioned style of rating relies on too many factors that may affect final pilot's result. The nowadays computers and operational capabilities allows to implement much more advanced algorithms that works online. Thus, developing of a system that would provide a detailed, objective and automated result was a motivation for the research. By applying an objective method, the spectrum of the measured parameters could be than extended, which clearly widen the range of information about the pilot's performance..

2. METHODOLOGY

The presented algorithm to asses handling quality is a new approach that takes advantage of multivariate (multicriterial) analysis that is a key to maintain objectiveness. System bases on comparing the real registered signal to the reference one defined within the manoeuvre. The flight parameters can be defined in the tool settings. It is worth noticing that some parameters like target velocity or flight height doesn't lead to significant changes in the algorithm. However, choosing as the reference trajectory another MTE requires major changes in formulas. The proposed tool Is modular, so it is possible to combine it biofeedback analysis from various sensors. The proposed method of comparing real signal to the reference signal consists of multiple steps.

2.1. Set of criteria

In the first part of the algorithm eight diverse criteria values for each pair real-target signal are computed. The choice of criteria was driven by necessity to gather methods that would be sensitive to different types of mismatch between real and reference signal (example presented later in the paper). The criteria varied also in terms of magnitude order allowed. The following assessing criteria were selected for the purpose of this work:

- integral test of convergence (ISE, ITAE, ITSE, ISEDT)

The first three criteria are widely used in control systems whose aim is to reduce error between real system output and target output with the time passing. The ISEDT criterion was proposed by the authors in order to penalise more errors occurring at the beginning of the flight when a pilot is less tired.

- similarity measures ("anticorrelation", cosine distance)

Anticorrelation term was introduced meaning complementary value to 1-Pearson correlation coefficient whose maximum value is 1. In this way it was ensured that every criterion has a feature that the lower its value, the better pilot performance. The cosine distance is calculated based on cosine probability that is often applied in the field of Natural Language Processing (NLP).

- Chebyshev metric (maximum metric)

This measure is sensitive to unexpected peak in signal that would not be detected by integral criteria if it doesn't last long.

- Discomfort measures (mean signal derivative, time domain difference measure, ride diagram)

Presented at 45th European Rotorcraft Forum, Warsaw, Poland, 17-20 September, 2019

All the measures evaluate smoothness of the ride. The physical interpretations of the last two criteria exists only in case of acceleration signal. Analysis was conducted only in time domain as in [6] researchers claim that changes of acceleration in time domain impact passengers more than changes in frequency domain. The time domain difference measure criterion [6] matches the observation of Giulianno and Ugo [1] that people are more vulnerable to transient than stationary accelerations. That is the reason why only the parts of signal that are greater than the sum of RMS of two analyzed signal is considered. The ride diagram criterion introduced in [6] analyses separately transient and stationary accelerations similarly to previous criterion. In order to obtain scalar value throughout this criterion the following formula was introduced:

$$discomfort = \frac{M_{stationary}}{M_{transient}} \quad (1)$$

2.2. Analysis of criteria sensitivity

In our work the substantial need for using multicriteria method was shown. We analyzed values of ten chosen criteria for the representative modifications of reference signal (described in previous section). It turned out that using two different criteria can lead to completely contrary quality assessment. Taking as an example a signal with constant value added, we observe that 'anticorrelation' criterion results in zero value (meaning there is no error), whereas any other criteria leads to nonzero values, which is a correct assessment.

2.3. Dimensionality reduction

The previous algorithm step resulted in 8 (or 10 values – in case of acceleration signal). In order to reduce space dimensionality machine learning techniques were applied. The overall final mark of handling quality is an arithmetic mean of marks referring to trajectory, height, velocity and acceleration signals. helicopter and the vessel are known and available for controller. It is assumed that this data is collected using onboard integrated INS/GPS systems (from the vessel data is transmitted to the helicopter using data link).

Operation of the LQR depends on selection of the weighting matrices values (Q – state weight matrix, R – control weight matrix). Here, selection of these values was made using iterative expert method till the answers of the helicopter model were adequate to fulfill the task of automatic landing on a moving confined vessel's deck.

3. EXPERIMENTAL DATAFLOW

3.1. 3.1.Hardware and Software

In order to verify reliability of proposed method dataset was collected by performing flights in rotorcraft simulator SW4. The simulator consisted of rotorcraft fixed cabin of real size with indicators and flight instruments imitation. As the part of device there is three-channel display set with spherical screen (180x45). The attached software enables registration of over 150 diverse flight parameters. The simulated flight was conducted in virtual environment that has many flight paths specified in ADS 33 norm implemented, such as slalom, pirouette etc. The presented algorithm for assessing handling qualities was implemented in Matlab environment. The program was compatible with simulator software.

3.2. Dataset

It consisted of 31 flights done in rotorcraft simulator SW4. The pilot was asked to conduct slalom maneuver (described in ADS- 33) maintaining constant height (20 m) and constant velocity (25 m/s). During all of the simulated flights the visibility was good. Realization of the maneuver was analyzed in terms of trajectory signal (latitude and longitude), height, forward velocity and acceleration. The reference trajectory may be seen at figure 1.

4. RESULTS

Signals corresponding to flights from the collected dataset were used as input to the algorithm.

4.1. Edge cases

It was observed that the proposed tool successfully pointed trajectories that were insufficiently mimicking reference signal as well as the one with the smallest error. Below we present the comparison of the trajectories that got the highest and lowest mark, which means the best and the worst among the others (see figure 2).

In the above example it was observed that almost every criterion pointed correctly the trajectory representing better performance by giving it smaller values. Nevertheless, comparing values of mean derivative criterion would slightly suggest

Presented at 45th European Rotorcraft Forum, Warsaw, Poland, 17-20 September, 2019

contrary conclusion about performance of the two analyzed signals.

4.2. 4.2Trajectories with similar mark

On the figure 3 and the tab;le 2 the comparison of trajectories with similar mark generated by program is presented.

By analyzing criteria value for each of the three trajectories above it has been noted that values got only via ITAE criteria present the same trend as shown on Fig. 2. There were even two criteria (Chebyshev distance and ITSE) that resulted in values representing inverse trend.

5. CONCLUSIONS

The proposed method proves that multicriterial character of analysis is crucial as using only one criterion can be misleading resulting in conclusions contrary to reality. The implemented algorithm has good accuracy in terms of comparing signals by generating quantitative marks. It can distinguish between “good” and “bad” flights in terms of performance.

6. FUTURE WORK

The next area of development is conversion of presented techniques used in performance assessment tool to engagement assessment component. Due to the complexity of the biofeedback signals (EEG, EMG, GSR and others) it is possible that the process may not be straightforward and may require modifications specific to field of biological signal analysis. Nevertheless, the approach of developing two objective rates: performance and engagement will definitely provide much wider information about the pilot/operator

1. REFERENCES

[1] Giulianno F., Ugo A., Experimental Assessment of Ride Comfort: A Statistical Approach, 2nd International Conference on Vehicle Comfort Ergonomic, Vibrational, Noise and Thermal Aspects, SAE 92A280.

[2] Graham H.D., Cummings M., Donmez B., Brzezinski A., Modified Cooper Harper Scales for Assessing Unmanned Vehicle Displays, 2010.

[3] Kalavsky P., Rozenberg R., Socha L., Socha V., Gazda J., Kimlickova M., Methodology of Pilot Performance Measurements. MAD-Magazine of Aviation Development, 5(2), 25-30, 2017.

[4] Longridge T., Bürki-Cohen J., Go T.H. , Kendra A.J., Simulator fidelity considerations for training and evaluation of today's airline pilots. Proc. 11th Internat. Sympos. Aviation Psych., Ohio State University, Columbus, 2001.

[5] Miller J., The Cooper-Harper Aircraft Handling Qualities Scale, from the Viewpoint of a Human Factors Flight Test Engineer, 1998.

[6] Strandemar K., On Objective Measures for Ride Comfort Evaluation, Automatic Control Department Of Signals, Sensors And Systems, ISBN 91-7178-204-4, Royal Institute Of Technology (KTH) Stockholm, 2005.

[7] United States Army Aviation And Missile Command Aviation Engineering Directorate Redstone Arsenal, Alabama, ADS-33E-PRF, Aeronautical Design Standard, Performance Specification, Handling Qualities Requirements for Military Rotorcraft, 2000.

[8] White M.D., Perfect P. , Padfield G.D., Gubbels A.W., Berrayman A.C., Acceptance Testing of a Rotorcraft Flight Simulator for Research and Teaching: the Importance of Unified Metrics, 2009.

Presented at 45th European Rotorcraft Forum, Warsaw, Poland, 17-20 September, 2019

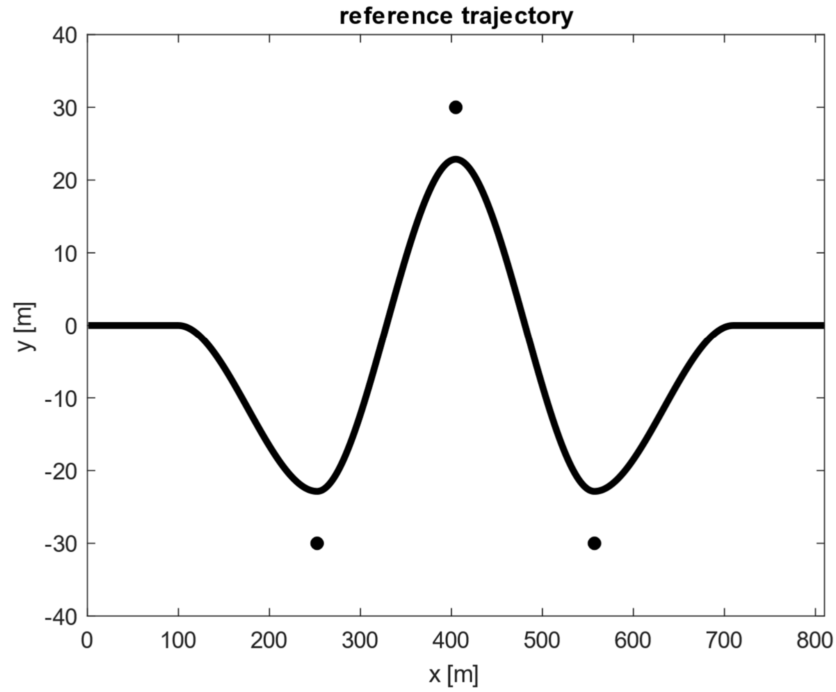

Figure 1. Reference trajectory and pylons positions (circles)

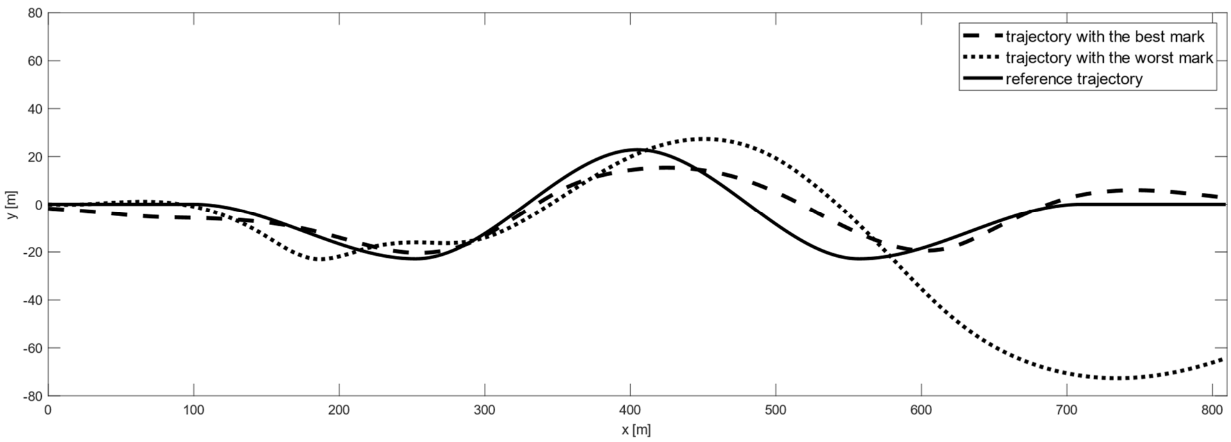

Figure 2 Comparison of reference and real trajectory signals

(with the best (0,992E+06) and the worst mark (22,190E+06) generated by the tool)

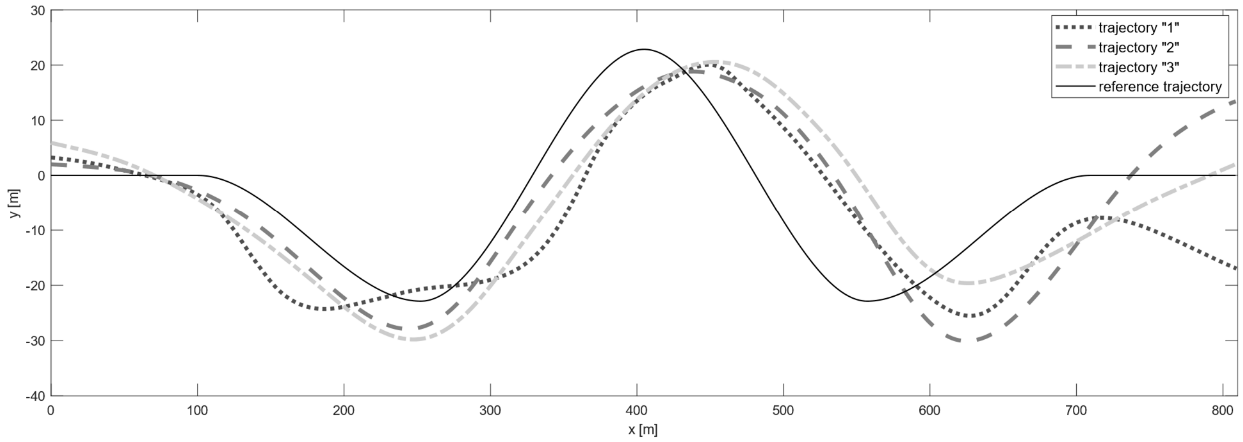

Figure 3 Three registered trajectories with similar performance mark and reference signal (solid line)
The brighter line colour, the lower mark (better performance)

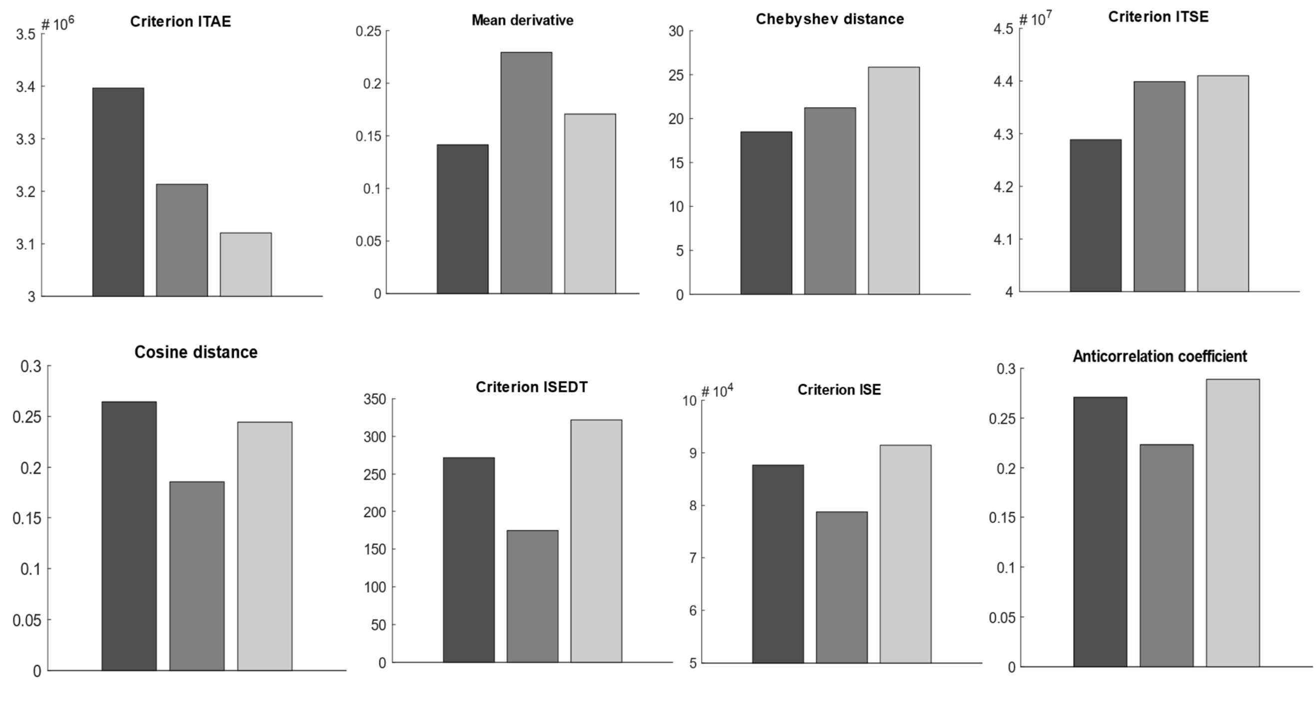

Figure 4 *Criteria values corresponding to trajectories shown on Fig. 3 (note: in order to ensure better visibility of mutual relations, on some charts the 'y' axis does not start from zero)*

Table 1 Criteria values corresponding to trajectories shown at Fig. 2

criteria	ISE	ITAE	ITSE	ISEDT	anticorrelation	d_{cos}	d_{Ch}	mean derivative
trajectory with worst mark	883233	10997219	6.2E+08	1300.332	0.709	0.533	72.652	0.125
trajectory with best mark	27480	1643237	1.3E+08	103.220	0.110	0.095	15.171	0.178

*Tab. 1. Performance mark of trajectories shown on Fig. 3 *expressed in millions*

Trajectory number	Performance mark*
"1"	2.513
"2"	2.477
"3"	2.446

Presented at 45th European Rotorcraft Forum, Warsaw, Poland, 17-20 September, 2019