Experimental Setup for Air Traffic Control Cognitive Complexity Analysis

Emilio. A. F. Albuquerque F., Luis. G. Trabasso, Alfredo Sandes. S and Marcelo de Araújo B.
Intituto Tecnológico de Aeronáutica (ITA), Brazil

Lishuai Li, R. John Hansman
ICAT, MIT, USA

Abstract — **A simulation environment was designed in order to understand the Complexity factors and to evaluate the influence of the time-based operations on Cognitive Complexity. Three different Complexity metrics were implemented in this experiment. Several participants were exposed to designed scenarios in different orders. Each scenario corresponds to a combination of independent variables, namely the Control Type and the Schedule Type. Preliminary empirical findings of the influence of scenario's order are also presented in this paper.**

Keywords — **Cognitive Complexity, Air Traffic Control, 4DT**

I. INTRODUCTION

The Next Generation Air Transportation System (NGATS) vision is composed of Concepts of Operation (Con Ops) and it focus on increasing the safety, security, and capacity of air transportation operations. For achieving that, NGATS proposes a combination of new procedures and technologies.

As a consequence of these modifications, the controller's workload is very likely to alter. Since controller's workload limits airspace capacity, it must remain in safe limits. Given that one of the key factors that drive the controller's workload is Cognitive Complexity [4], it is necessary to understand the Complexity's behavior under current and future operations. In this paper, it is made the same assumption of [3] about Complexity: that it is related to the cognitive difficulty of controlling the air traffic situation, which in turn is tied to the ability of controllers to maintain safe operations under normal and abnormal conditions.

Among the to-be-implemented key capabilities identified by the NAGTS, this research works with the Four-Dimensional Trajectory (4DT) control, which is the description of an aircraft path both in space and time. Some of the waypoints in a 4DT path may be associated with Control Time of Arrivals (CTAs), which are time windows for the aircraft to cross specific waypoints within a prescribed conformance tolerance [5]. The use of 4DT minimizes the aircraft excess separation resulting from today's control imprecision and lack of predictability, increasing airspace capacity, as a consequence [5].

The structure in Air Traffic Control (ATC) environment and the tasks of projecting and managing the traffic situation will be altered by the 4DT control. It is known that the underlying structures in the ATC environment have a significant influence on cognitive complexity [3]. Therefore, it is expected that additional temporal structures may cause an impact on Cognitive Complexity.

The lack of a clear definition in the literature about Complexity reflects on a wide variety of measures that have been used to evaluate it. This fact brings in difficulties for comparison among researches. Nonetheless, three Cognitive Complexity Metrics were employed in this work. They are:

1. *Modified Cooper Harper Scale* [2]: Proposed by the scientists Cooper and Harper for quantifying how a pilot's workload affects task performance; this scale ranges from 1 to 10. The participant goes through yes or no questions which direct him (her) to the proper rating, reducing diversion. A similar scale for Cognitive Complexity Rating was created from the original scale and implemented.
2. *Modified Aircraft Count* [3]: Instead of simply counting the number of aircraft in the airspace, this metric weighs each aircraft by its contribution to the overall cognitive complexity. In other words, the modified count tries to describe Complexity as an Effective Number of Aircraft (ENA) by the sum of each aircraft contribution. The participants are instructed with a standard aircraft definition and then, during sample times of the simulation, they identify and score aircrafts that are different from this baseline.
3. NASA WAK (*Workload Assessment Keypad*) [8]: This metric was used on NASA experiments. Keypads were installed at each test position as a means of recording complexity ratings. Its scale ranges from 1 to 7 and it is employed in this simulation for comparison with previous researches.

The main objective of this work is to contribute with the understanding of Cognitive Complexity driving factors, especially as regards implementing new tools and procedures of 4DT. It was given a focus on the influence of the scenario's order on the non-skilled participant's performance and perceived Complexity.

II. EXPERIMENT DESIGN

The simulation was built in MATLAB™ and it models arrival routes to Boston's Logan International Airport. It includes arriving traffic and some crossover aircraft. The experiment was designed to simulate the key factors of the current based operations and the time based operations (following the 4D trajectory based system anticipated by the NGATS Con-Ops). Three different types of schedule were used: no schedule (NS), First Come First Serve (FCFS) and Constrain Position Shifting (CPS).

The aircraft enters in the airspace following the arriving structure to an airport. The airspace structure is composed of four main streams converging to the right edge of the screen. There are two kinds of aircraft: the normal one that follows given routes and the crossover one that crosses the traffic. The task of the participant is to deliver the normal arriving traffic to the second last waypoint (the CTA – Control Time of Arrival) and then to the last waypoint on the right edge of the screen. The crossover aircraft have to be safely delivered to the opposite side of the airspace that they came from.

The participant has to manage the arriving aircraft as safely and quickly as possible. He/she has three primary goals, in decreasing order of importance:

1. *Guarantee the airspace safety;*
2. *Avoid losing aircraft from the airspace;*
3. *Manage traffic efficiently.*

Independent Variables

In experiment described herein, the independent variables were chosen in such a way that the whole process was not too long and the key aspects could be represented. As the intent was to analyze the Cognitive Complexity under certain automations features, these were first depicted in some levels or graduations.

Under the NGATS' trajectory based operations concepts, only the first level of a four-dimensional trajectory was chosen, namely the presence of a Control Time of Arrival (CTA) in the airspace main route.

Under four-dimensional trajectories, it is reasonable to assume the presence of a sequence scheduler. Then, it was decided to implement certain levels of schedule. The first one represents an operation without planned schedule; the second one is First Come, First Serve schedule and the third level is the optimized schedule, similar to the optimized schedules that might be used in a Trajectory-Based Operations environment. The last two levels are very compact schedules, obeying the wake vortex constrains as tightly as possible.

The optimized scheduler is named Constrain Position Shifting (CPS). The algorithm implemented on this work was developed by [1] and [6]. It consists of an optimized schedule which minimizes both the average delay and the total time span.

The independent variables can be divided as follows:

- Command Structure (two levels)

1. Current Operations: position-based operations – direction and speed commands;

2. First level of 4DT: time-based operations – time, direction and speed commands.

- Schedule Type (three levels)

1. None

2. First Come, First Served

3. Constrain Position Shifting

Scenarios

The route structure consists on the Norwich Three Arrival Routes. In order to minimize the learning and memory effects during the whole experiment, it was decided to:

- Invert the route structure in some scenarios;
- Divide the traffic in chunks of traffic and reorder them.

The participant has the impression of dealing with a new situation when the route structure is inverted, but actually he/she is dealing with the very same route. It is expected that this is enough for the participant not to use the same strategies that were used in previous scenarios and not to wait for the aircraft to come in specific times and attitude.

Each chunk of traffic is a design of the entering traffic, defined by the *traffic entering rate* (entering aircrafts per minute). Three chunks were created. Chunks A and C are easy and they may come at the beginning or at the end of the experiment. Chunk B is the hardest one and it comes only on the middle. Since each chunk defines an aircraft entering rate, the transition between chunks defines the medium traffic load.

Once defined the independent variables, six different possible scenarios are possible, as shown in Table 1.

Table 1 The Scenarios and the Independent Variables [7]

		Control Type	
		Position Based Heading, Speed	**Time Based** Heading, Speed, Time
Schedule Type	**None**	Scenario 1 Structure: Normal Traffic: A-B-C	Scenario 4 Structure: Normal Traffic: C-B-A
	FCFS	Scenario 2 Structure: Flipped Traffic: C-B-A	Scenario 5 Structure: Flipped Traffic: A-B-C
	CPS	Scenario 3 Structure: Normal Traffic: C-B-A	Scenario 6 Structure: Normal Traffic: A-B-C

The surveys carried out with the participants were done at the moments that the traffic load is most relevant. These moments are: (1) easy at the beginning; (2) medium and increasing difficulty; (3) hard; (4) medium and decreasing difficulty and (5) easy at the end.

Figure 1 illustrates the entering traffic and the sample times.

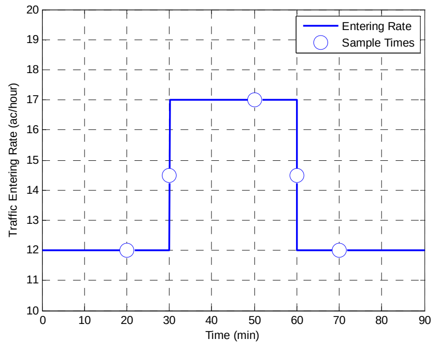

Figure 1 Taffic and Sample Time configurations (adapted from [7])

Method

It is described the adopted heuristics that was considered adequate for the definition of the experiments set, as a subset of the full factorial experiment that would require 720 experiments. It was decided to split the participants in two types: (1) those who execute first the position-based and only then the time-based scenarios; and (2) those who execute first the time-based and only then the position-based scenarios. With this arrangement, the participants explore the key factors of each command structure at a time. With this role, the whole set of experiment orders is reduced to 72. As the behavior of Cognitive Complexity is barely known, it was also decided to repeat each experiment order at least 4 times, what would yield to 288 experiments. Because of time restrictions, the total of 72 experiment orders was further reduced to 12. Each one of these orders would be repeated 6 six times, totalizing 72 experiments.

Up to now, the whole planned set of 72 participants couldn't be done. The analysis in this article was done over a set of 48 participants – each order was executed by 4 different participants. Table 2 illustrates the experiment orders.

Table 2 Listing of Scenario Orders

Experiment Order	Experiment Type	1st scenario	2nd scenario	3rd scenario	4th scenario	5th scenario	6th scenario
1	1	1	2	3	5	4	6
2	1	1	3	2	5	6	4
3	1	2	1	3	4	5	6
4	1	2	3	1	4	6	5
5	1	3	1	2	6	4	5
6	1	3	2	1	6	5	4
7	2	4	5	6	3	1	2
8	2	4	6	5	3	2	1
9	2	5	4	6	2	1	3
10	2	5	6	4	2	3	1
11	2	6	4	5	1	2	3
12	2	6	5	4	1	3	2

Other very important issue tackled by this research was to find out ways of minimizing memory, learning and cognitive strategies developed by the participants. Figure 2 illustrates how the experiment plan was organized.

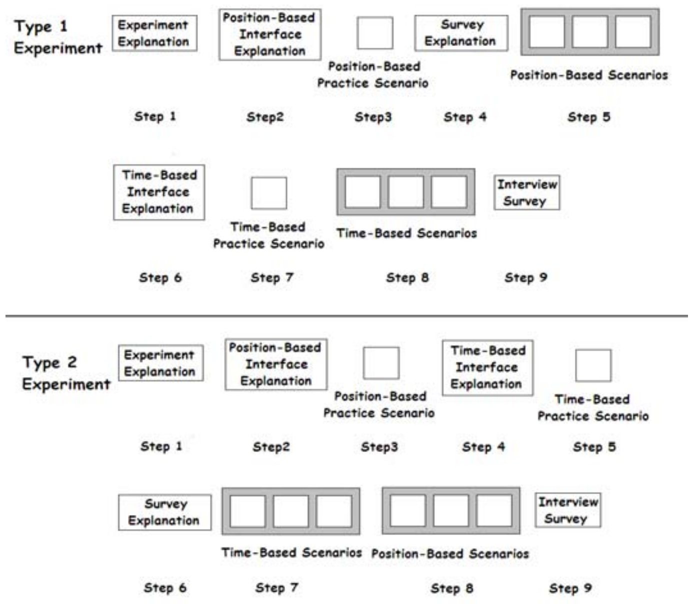

Figure 2 Experiment plan according to the order's type

III. RESULTS

Given the way the Effective Number of Aircraft (ENA) was previously described, each aircraft can receive a score that ranges from 0 to infinite (one aircraft with score below 1 is less complex than the standard aircraft; and an aircraft with its score higher than 1 is more complex than the standard aircraft). This is not a proper rating, if there is an intention of comparison of this metric with the other ones. It was decided to correct the ENA results to the same range of the Cooper Harper (CH) scale from 1 to 10. The following calculation procedure was done for each participant:

1. All of the aircraft ratings were normalized to the interval [0, 10], based on the highest aircraft individual score given by this participant.
2. With this normalization, each aircraft contribution was summed, for the calculation of the instantaneous Cognitive Complexity.
3. All the instantaneous Cognitive Complexity were normalized to the interval [0, 10], now based on the highest single instantaneous Complexity previously calculated.
4. The [0, 10] interval was transformed in a [1, 10] interval, as the CH scale.

The data collection during the whole simulation was not restricted to the proposed Cognitive Complexity Metrics: participants' performance data were collected as well, namely the number of collision and lost aircraft.

Subjective Complexity Rating

Figure 3 shows the results obtained with 48 ITA (Instituto Tecnológico de Aeronáutica) students, which indicate that the subjective complexity ratings under time-based control were slightly lower than those under position-based, by both the modified CH and WAK scales, at all sample times. No statistically significant difference was found though. The same results were obtained by [7] in their experiment with 22 participants, all of them with some kind of air traffic control experience.

The aim of the analysis presented in Figure 3 is to have a solid comparison of our experiment results with the results obtained at MIT (Massachusetts Institute of Technology) [7]. The first two charts in Figure 3 shows the same analysis and overall results obtained by them. The third chart in this Figure introduces a normalized ENA analysis. No statistically significant difference was found between position and time-based operations.

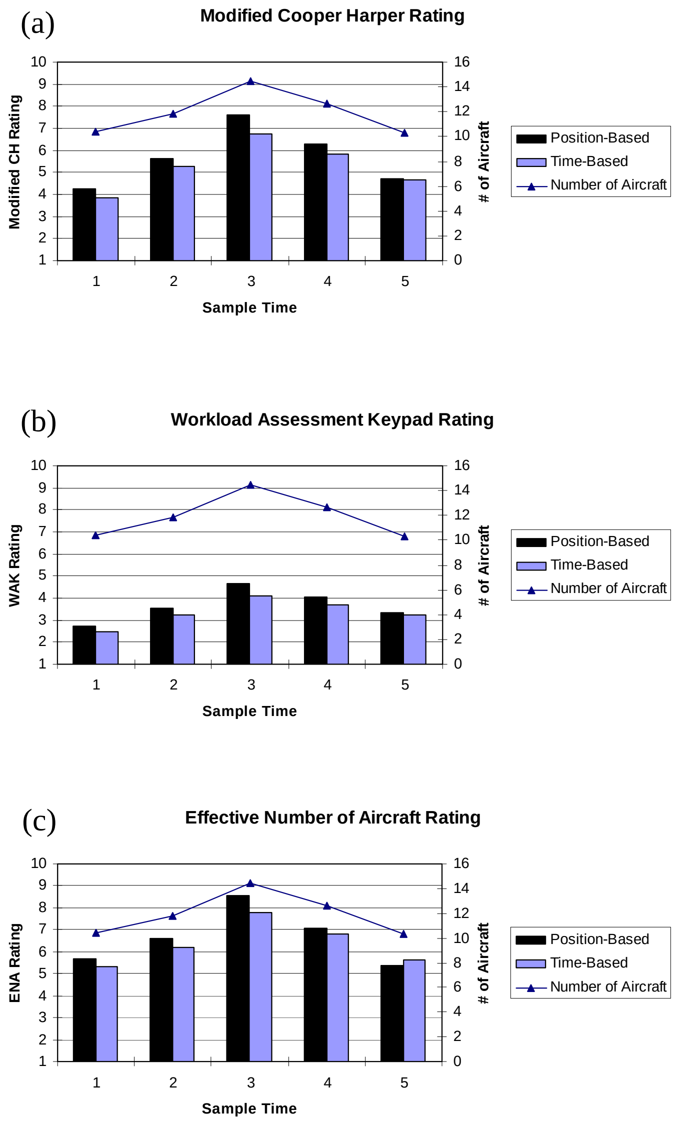

Figure 3 Subjective Complexity Ratings by (a) modified CH (adapted from [7]), (b) WAK (adapted from [7]) and (c) normalized ENA

The three complexity metrics show the same general behavior with respect to the traffic load, which is the complexity rating correlating with the traffic load. The position-based operations tends to be evaluated slightly more complex than the time-based operations. The only exception for that occurs with the ENA measure on the last survey. This result may be explained as the effort of the participants to correct the timing errors of the remaining aircraft. These errors can be quite big, just based on the fact that they can have been accumulated during the last 30 simulated minutes or so. The need of these further corrections may introduce additional complexity to the sector.

It has also been observed that the ENA results showed greater values than the CH results and also a smaller standard deviation. A possible conclusion for that is the improvement of the exploration of the cognitive complexity mechanism by the ENA rating, especially when it calls to the impacts of the airspace structure. In other words, the participant tends to describe and explore better the cognitive complexity behavior, even unconsciously. The similarity of these results with those obtained at MIT [7] show that this research is very well grounded to be continued with further analysis.

Scenario's Order Analysis

As showed in Figure 2, at the end of the experiment, a survey is done with the participant. Many people who executed Type 2 Experiment commented that it would have been much easier if they have done the position-base scenarios first. Similarly, the Type 1 participants commented that no inversion would make the experiment easier or with a smoother learning process. Such feedbacks motivated the effect analysis of the exposure to a certain scenario order.

In such analysis, a behavior that would be tried to capture is the evolution in time of factors that, at least hypothetically, would influence the Cognitive Complexity. These factors would be effects of learning (mastering of the participant about the simulation and its interface), memory (remembering of the solutions employed previously to the same chunk, for example) and cognitive strategies development (they may evolve as a result of the mental pictures and representation of the traffic situation, which, by its turn, may evolve with practicing and exposure to different command structures, for example). A preliminary result that came from this analysis is shown in Figure 4.

The green line shown in Figure 4 with an angular slope of 45 degrees was plotted. Red lines divide the chart in two regions: one with ENA and CH ratings higher than 8 and other one with ENA and CH ratings higher than 6. Each point corresponds to the mean results for ENA and CH ratings for each scenario. One mean standard deviation from the mean is plotted for both ENA and CH ratings. These results were collected from the 3rd survey (when the participant is exposed to the hardest traffic load). The bigger the circle around each point, the larger the number of collisions until the 3rd survey. Each point has a label: the subscripted number refers to the order when the scenario was executed and the number in parenthesis is the scenario number itself. For example, consider the point with the label $O4_1(2)$, it refers to the first scenario executed in the 4th order, which is the scenario number 2.

Figure 4 shows that the results for the ENA rating tend to be higher than the ones for the CH rating, which, by its turn, is more spread out than the ENA Rating. This observation agrees with the comparison of charts in Figure 3. On the chart for the 4th order, an interesting and expected result is that the first three scenarios were indicated as more complex than the last three, by a large difference. Furthermore, they can be separated in two different clusters, one being more complex than the other in both metrics. There might be at least two hypotheses for that: (1) the evidence of the effects of learning, memory and cognitive strategies development evolving with respect to time, and (2), on the last three scenarios, the participants were exposed to the time-based control, which tends to reduce complexity, as indicated in Figure 3.

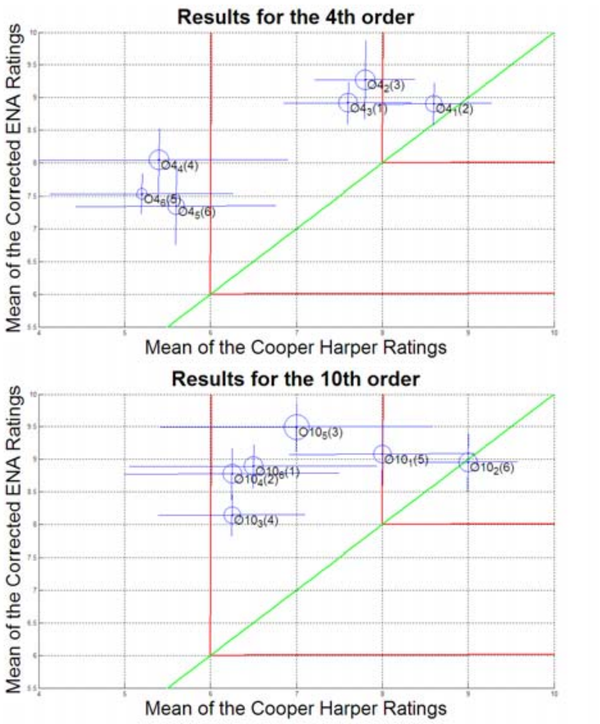

Figure 4 Results for the 4th and 10th order

However, more detailed analysis has shown that the gathered data do not yield sufficient statistical evidence of differences between these two groups at a significance level of 10% or smaller. The same lack of evidence is observed for all the 11 other orders, especially because they look much more like the chart for the 10th order, which present the scenarios results very close to each other and with high standard deviations.

Based on these findings, one might question if the Complexity metrics can capture Complexity variations regarding the factors previously mentioned (learning, memory and cognitive strategies development). If they can somehow capture some factors (as seen in Figure 3), but not specifically these factors, it becomes evident why these effects turn to be unnoticed. Anyhow, such preliminary results do not indicate, for now, new tools that might measure these factors or how these factors would compute in the overall Cognitive Complexity.

On the other hand, if there are very good reasons to believe that effects such as learning, memory and developed cognitive strategies in fact influence the perceived difficulty and the participant's performance, maybe the not detection of these effects simply show that good choices for the experiment design were made. In other words, this implies that decisions related to the minimization of effects of memory (reordering of the traffic chunks and airspace structure inversion in some scenarios), of learning (detailed concepts, interface and overall experiment explanation, followed by practice scenarios) and of cognitive strategies development (moments and duration of the practice scenarios) were in fact effective.

IV. CONCLUSION

Results show that this work is very well grounded with other similar works. So, there was the intention of making further and complementary analysis, especially as regards evaluating effects of scenarios sequencing. But this analysis is pretty much inconclusive in this sense. Cognitive Complexity has proven again to be composed of a large number of elements whose interactions and effects challenge traditional analysis.

Further literature research and analysis of the collected data is being done, in order to better identify the effects of experiment order on the participant's committed errors and perceived Complexity. The focus is being conducted on factors such as learning, memory and cognitive strategies: how they affect Complexity and how they could be measured. Comparison between the employed metrics in this work is also promising and this is being done.

For now, a new set of participants is going to be tested. These new participants are going to be skilled professionals on Air Traffic Control operations.

V. REFERENCES

[1] BALAKRISHNAN, H. and B. CHANDRAN. S*cheduling* Aircraft Landings under Constrained Position Shifting. In AIAA Guidance, Navigation, and Control Conference and Exhibit. Keystone, Colorado, August 21 2006.

[2] CUMMINGS, M.L., K. MEYERS, and S.D. SCOTT, Modified Cooper Harper Evaluation Tool for Unmanned Vehicle Displays, presented at the UVS Canada Conference, 2006.

[3] HILBURN, B. Cognitive complexity in air traffic control - A literature review. EUROCONTROL EXPERIMENTAL CENTRE. EEC Note No. April, 2004.

[4] HISTON, Jonathan M., R. John HANSMAN, Guillaume Aigoin, Daniel Delahaye, and Stephane Puechmorel. 2002. Introducing Structural Considerations into Complexity Metrics. Air Traffic Control Quarterly Vol.10, No. 2, 2002.

[5] JOINT PLANNING AND DEVELOPMENT OFFICE. Concept of Operations for the Next Generation Air Transportation System. June 13, 2007.

[6] LEE, Hanbong. Tradeoff Evaluation of Scheduling Algorithms for Terminal-Area Air Traffic Control. Master Thesis, MIT, June 2008.

[7] LISHUAI Li, R. John HANSMAN and Albuquerque, EMILIO. Part Task Simulation of Air Traffic Control Complexity with Time-based Control. MIT, july 2009 (In print).

[8] KOPARDEKAR, P., and S. MAGYARITS. 2002. Dynamic Density: Measuring and Predicting Sector Complexity [ATC]. In Digital Avionics Systems Conference,

2002. Proceedings. The 21st, 1:2C4-1-2C4-9 vol.1. doi:10.1109 /DASC.2002.1067920.