IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

How much workload is workload? A human neurophysiological and affective - cognitive performance measurement methodology for air traffic controllers

To cite this article: María Zamarreño Suárez *et al* 2022 *IOP Conf. Ser.: Mater. Sci. Eng.* **1226** 012085

IOP Conf. Series: Materials Science and Engineering 1226 (2022) 012085 doi:10.1088/1757-899X/1226/1/012085

How much workload is workload? A human neurophysiological and affective - cognitive performance measurement methodology for air traffic controllers

María Zamarreño Suárez[1*], Rosa María Arnaldo Valdés[1], Francisco Pérez Moreno[1], Raquel Delgado-Aguilera Jurado[1], Patricia María López de Frutos[2] and Víctor Fernando Gómez Comendador[1]

[1]Universidad Politécnica de Madrid (UPM), 28040 Madrid, Spain
[2]ATM Research and Development Reference Centre (CRIDA), 28022 Madrid, Spain

*Corresponding author: maria.zamarreno.suarez@alumnos.upm.es

Abstract. Air traffic controllers play an important role in enabling safe, orderly, and efficient flight management within airspace. By the very nature of their work, they must make critical decisions in a time-critical environment. From a safety point of view, it would be of interest to obtain the workload thresholds (upper and lower limits) in order to define a range in which it would be safe and efficient for air traffic controllers to work. This collaboration project between CRIDA and UPM aims to obtain a methodology to define these workload limits, based on experimental affective-cognitive data and neurophysiological parameters of the performance of air traffic controllers. The workload assessment combines objective and subjective approaches. From the objective point of view, electroencephalography and eye-tracking techniques will be employed. In this paper, the first milestones achieved in the project are presented, as well as the simulation platform to be used in the experiment and the simulation-based set of exercises that has been specially designed for this project.

1. Introduction

Air traffic controllers (ATCOs) play a key role in the safe, orderly, and efficient organisation of air traffic. Their role is central to air traffic control (ATC). This can be seen in the large number of projects developed to identify changes in ATC and, in particular, in air traffic controllers, as a consequence of future changes in the air navigation system (as an example, see [1,2]).

Due to the characteristics of the environment in which they work, controllers have to make critical decisions in a context where time is a critical parameter. To enhance high safety standards, their workload should be within reasonable limits [3].

From a safety point of view, there are two situations to avoid: overload situations for controllers and situations where the workload is too low. Both scenarios can significantly degrade the quality of the actions they perform.

This collaboration project between CRIDA and UPM aims to define the thresholds (upper and lower limits) of workload so that controllers can perform their tasks safely and efficiently. To this end, experimental data on affective-cognitive parameters related to human factors applied to controllers will be used as a basis. This data will be recorded while the controllers develop an exercise programme based on real-time simulations.

Section 2 presents the methodology that is expected to be followed throughout the project. Section 3 gives an overview of the simulation platform that has been configured. Section 4 presents the fundamental characteristics of the exercise set especially tailored for this project. Section 5 then explains the approaches used for the evaluation of the ATCO workload. The preliminary results obtained after the first simulations carried out appear in Section 6. Finally, Section 7 presents the first conclusions drawn from the project.

2. Methodology

The duration of the project is planned to be at least three years. The final result will be the affective-cognitive human factor performance methodology based on the neurophysiological parameters of air traffic controllers. The project is divided into six stages. These stages are represented chronologically in figure 1. This figure shows the stages that have already been completed, the work that is currently being carried out, and the future steps.

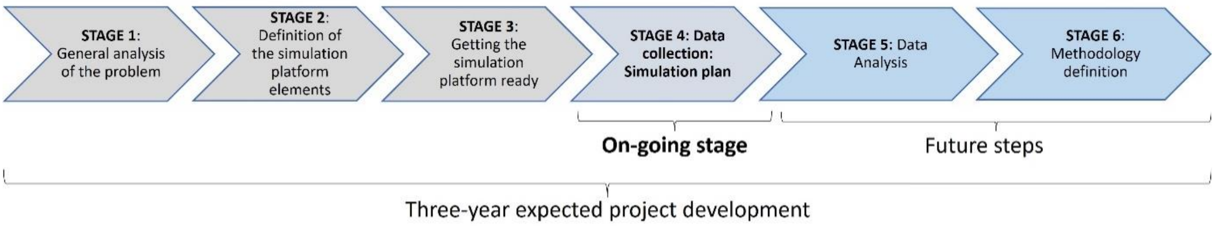


Figure 1. Overview of the project stages: phases completed, work in progress, and future steps.

The project that is the subject of this study is currently in development. At the moment, the first three stages presented in figure 1 have been completed. Currently, a simulation campaign is being carried out, which corresponds to the fourth stage. Stage 5 and stage 6 will be developed in the future, once all the required data is available.

After the completion of the first three stages of the project, several milestones have been reached.

- The first step consisted of a detailed analysis of the problem and the objectives to be achieved.
- Subsequently, all elements of the simulation platform and their most important characteristics were defined.
- The third stage consisted of setting up the simulation platform for the subsequent execution of the experiment: the equipment was configured, a procedure was established to take measurements with the sensors, a specific set of exercises based on simulations was designed, and additional functionalities were created to appear on the screen of the control position. These functionalities were defined to record more data on the actions taken by the controllers.

As a result of the three phases already completed, an exercise programme has been designed and the ATCO position within the simulator has been improved. Detailed information on the process followed for the creation of the exercises can be found in Section 4 of this article.

Currently, the development of the project is focused on the simulation campaign. Simulations are carried out three days a week, and a first analysis is being performed to obtain preliminary results, which will allow the work to be organised in subsequent stages. These preliminary results and the first conclusions drawn on the basis of the experimental data are presented in Section 6 of this article.

In the coming months, work will continue to obtain more data. Subsequently, a more detailed analysis of all the data obtained will be undertaken, and the conclusions reached will be generalised.

The last step after data processing will be to establish ATCO workload thresholds and define the methodology.

3. Simulation platform

The simulation platform consists of four basic pillars: the IT part, the sensors, the set of exercises, and the new functionalities designed to record data on the actions taken by the ATCOs.

The IT part of the project consists of five PCs: a server, a manager, an editor, and two control positions. The simulator uses SkySim software by Skysoft-ATM. This is a flight simulator for the en-route phase. In this case, the Madrid ACC was the airspace used. Each of the exercises has been designed for different sectors. The editor uses ROSE Simulation software.

From the outset of the project, it was decided to use two different pieces of equipment to record the neurophysiological data measurements of the controllers. A headset will be used to record electroencephalographic (EEG) data and an eye-tracker will be used to register information related to visual parameters.

The set of exercises has been specially designed for this project. Its most relevant features are explained in detail in Section 4.

Finally, in the stage of analysis and identification of the necessary functionalities for the simulator, a series of limitations were identified in the basic configuration of the simulator. Once these gaps were identified, a series of new functionalities were designed to solve them.

These functionalities are auxiliary windows that appear on the radar screen of the control position. The controller can interact with them. In this way, more data can be recorded on the actions taken by the controller during the development of the exercise.

Figure 2 shows an overview of the control position, the visual tracking sensor, and the different tools available for the controllers on the radar screen. Each of the elements is labelled.

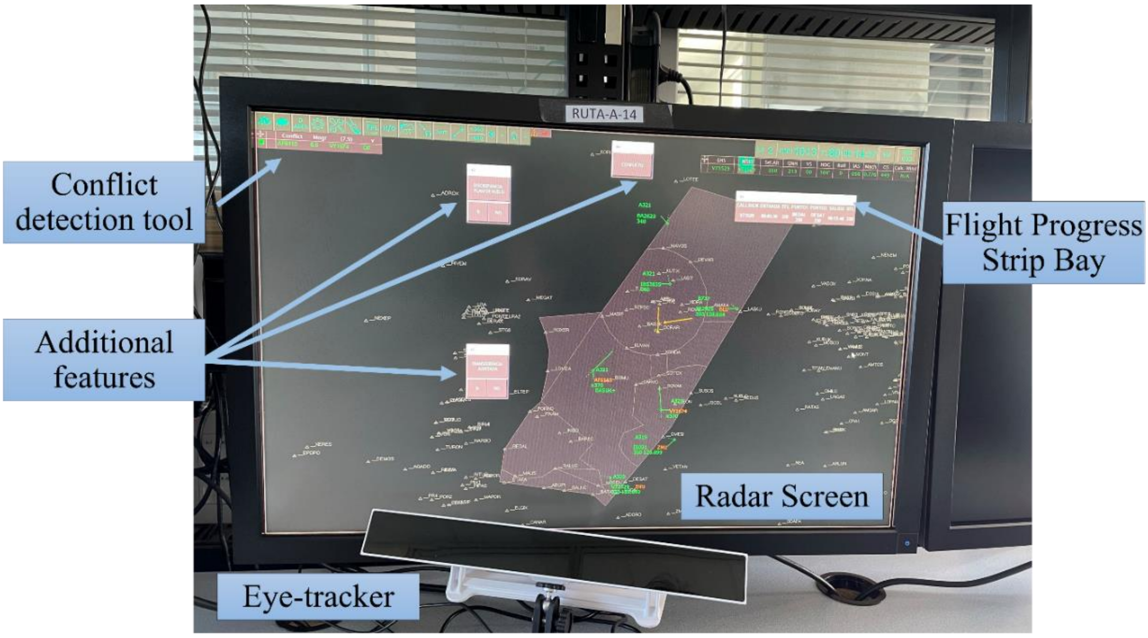


Figure 2. Overview of the ATCO position.

4. Design of the exercise plan

4.1. General characteristics

The set of exercises has the following characteristics:

- Before starting with the exercises used to record neurophysiological parameters, the participants in the experiment will perform an initial training exercise to learn the basic commands of the simulator.

- A total of seven exercises have been designed, each lasting 45 minutes.
- To allow two subjects to perform the exercise in parallel in the two control positions, each exercise has been designed for two different sectors. For each of the exercises, its difficulty is the same in both sectors.
- Before developing the simulations, it is possible to anticipate to some extent the results to be obtained. This is because the exercises have been designed considering a series of events distributed over the total duration of the exercise.
- Each exercise has been created on the basis of a previously defined theoretical difficulty profile. The difficulty score of the exercises increases progressively throughout the set.

4.2. Scoring of the events

One of the first steps in the exercise design programme was to define the scores of the events. Previous experience in the simulator and the comments and impressions of the people who had acted as controllers were considered when awarding the scores.

A total of 12 events were considered that covered the basic actions of the ATCOs. The identification event (with a basic score of 1 point) was defined to consider the moment when an aircraft appears on the screen until the controller identifies it.

At the time of acceptance of an aircraft, two events were defined: standard acceptance (3 points) and acceptance that requires action (5 points). Similarly, at the moment of transferring the aircraft to the adjacent sector, two different events were defined: transfer (3 points) and transfer with action (5 points).

Five different types of conflicts have been considered. The conflict that occurs between two aircraft at cruise flight level has been assessed with 7 points of difficulty. The overtaking conflict, assessed with 8 points, occurs when two aircraft are on the same trajectory and the one behind is faster than its predecessor, causing the faster aircraft to approach the other one in a progressive way. The combination of conflicts in which one or both aircrafts are in evolution has been assessed with a score of 9 points. Associated with each of the conflicts, a vectoring event is defined, with a score of 3 points. This event considers the vector guidance provided by the controller to the aircraft to resolve the conflict.

All the above events have an absolute score. There is an additional one, the monitoring event, which considers the controller's workload in supervising the aircraft under his/her responsibility. In this case, the score awarded is 0.1 for each minute the aircraft is within the sector.

These scores refer to the low traffic density situation (less than 5 simultaneous aircraft in the sector). In the case of medium traffic density (between 5 and 9 aircraft in the sector, respectively), these base scores increase by 5%. Similarly, when the density is high (more than 10 simultaneous aircraft), the base score increases by 10%.

4.3. Design

The preliminary design of the exercises is carried out with a specially designed spreadsheet. Once this preliminary design is completed, information on the different aircraft, their entry and exit times is available, and work can begin with the simulator editor to enter all the aircraft which appear in the exercise.

The preliminary design of the exercises has been structured in the four steps shown in figure 3.

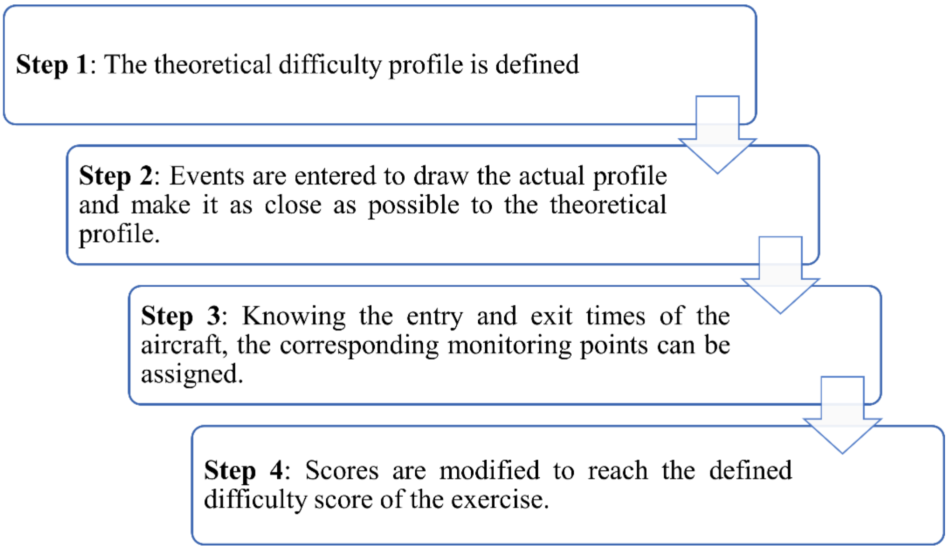


Figure 3. Stages of the design of the exercises.

The starting point was the total difficulty score of the exercise. Similarly, the theoretical difficulty profile to be achieved was defined in advance. This profile is characteristic of each of the phases of the set of exercises.

Taking into account the theoretical difficulty profile, events with an absolute score were included throughout the exercise, trying to ensure that the temporal distribution of difficulty was as similar as possible to the theoretical difficulty.

In the third step, the monitoring event is included, the only one that has a different score depending on its duration. Knowing the times of entry and exit of the aircraft in the sector, the contribution of this event can be included in the diagram defined above.

Finally, the iterative process of error correction begins, modifying the position of the events as necessary so that the graph of the temporal distribution of the actual difficulty and the theoretical difficulty coincide as much as possible.

4.4. Designed exercise set

After following the design process discussed above, the seven exercises of the set have been obtained. The choice of this number of exercises is due to the optimisation of previous work done with the simulation platform. It has been decided that this number is the best balance between the time available for the design of the exercises and the objectives of the project.

In a first phase of the simulations currently under development, the study participants will perform the first four exercises of the set, together with an introductory exercise, in order to familiarise themselves with the simulation platform.

Each of the seven exercises is characterised on the basis of six parameters, which can be seen in figure 4.

The exercises are organised into three **phases of difficulty**: basic, intermediate, and advanced. The basic phase consists of two exercises, the intermediate phase consists of three exercises, and the advanced phase consists of two exercises.

The idea of structuring the exercises in phases of difficulty is recurrent in other similar studies. In particular, for the naming of the phases considered, the fundamental reference has been the SESAR AUTOPACE project [4]. The number of exercises in each of the phases has been decided on the basis of the schedule of the simulations to be carried out.

The difficulty increases progressively, with the basic phase being the easiest and the advanced phase the most complex. This increase in difficulty is achieved by increasing the number of aircraft, increasing the difficulty of the valley zones of the difficulty profile, and progressively introducing conflicts to be resolved with greater demands on the controller.

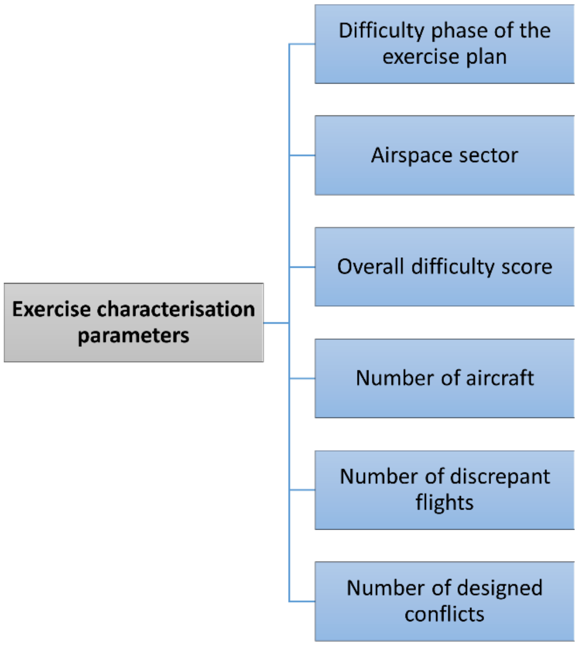


Figure 4. Overview of the characterisation parameters of the exercise.

The exercises have been designed in **different sectors of the Madrid ACC airspace**. The idea is that the subjects of the experiment perform ATC tasks in different sectors throughout the exercise programme. In the first two phases of difficulty, each controller will be responsible for a single control sector. In contrast, in the advanced phase, their responsibility will increase to two adjacent sectors to be able to include longer routes and a larger number of conflicts during the design phase.

The **overall difficulty score** is one of the most important parameters in the characterisation of exercises. It is calculated as the sum of the partial difficulties of the events considered in the design. The exercises of the basic phase have a difficulty between 140 and 190 points. Exercises in the intermediate phase are in the range of 220 to 260 points. Finally, the advanced phase exercises have a difficulty of 280 and 320 points, respectively.

The **number of aircraft** refers to the total number of aircraft entering the sector during the exercise. This number also increases progressively throughout the exercise programme. In exercise 1, 14 aircraft appear in the ATCO's airspace of responsibility. The maximum number of aircraft, a total of 30, is reached in exercise 7, the last exercise of the set.

The number of discrepant flights refers to aircraft with a flight plan in the simulator and a different flight plan in the flight progress strip bay. The controller shall identify these flights and act on them so that their trajectory matches that defined in the flight progress strip. These flights have been designed to appear specifically at certain times during the exercises to increase their difficulty.

Both discrepant flights and designed conflicts are two tools widely used during the design phase to complicate the exercises. For this reason, in each exercise, these two complexity factors have been balanced against each other. This is the reason why neither of these two parameters increases gradually throughout the set of exercises, but rather they vary in each of the phases and in each of the exercises, so that the total difficulty increases progressively.

The number of designed conflicts contributes significantly to the difficulty score of the exercise, as conflicts are the events with the highest associated scores.

As discussed above, 5 of the 12 basic events considered in the design of the exercises are different types of conflicts. Scoring varies according to the type of conflict, considering the resolution of some of them easier than others.

Conflicts have been introduced at specific moments of the exercises to increase their difficulty. Both the number of conflicts and the time at which they are introduced depend on the exercise in question.

As an example, exercise 1 has the lowest number of conflicts with a total of 3, whereas exercise 7 has been designed including 6 conflicts.

5. ATCOs' workload assessment

In general, the ATCO workload assessment can be performed using two techniques: objective and subjective. In this project, it has been decided to use both techniques in combination to compare the results obtained. Figure 5 shows schematically the approaches used, the techniques used, and the sensors to be used in the objective assessment.

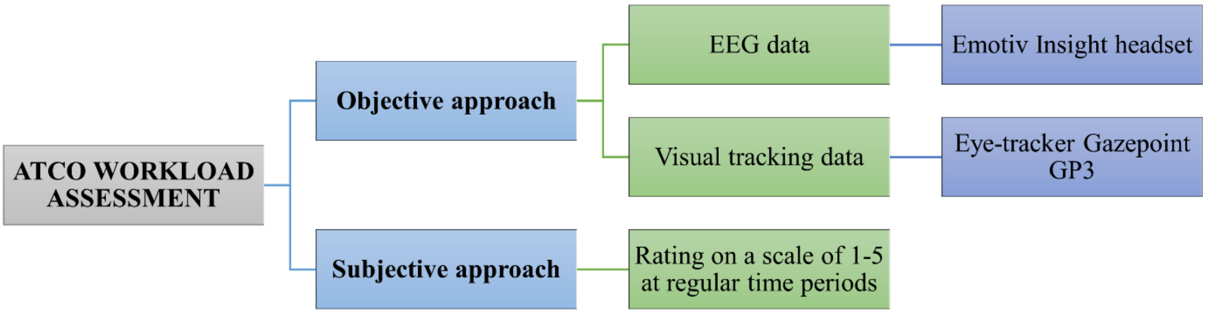


Figure 5. Workload assessment of ATCOs: approaches, techniques, and sensors used.

To carry out a subjective evaluation, a window has been designed to appear on the controller's screen at regular intervals while the ATCO is developing the simulations.

The text in the window asks the controller to evaluate his/her workload according to a scale of 1-5. He/she would choose 1 in the case where the workload is very low and 5 in the case where the ATCO is very overloaded.

20 seconds after appearing, the window disappears automatically. If the controller has not responded to any of the options, this is also recorded.

On the other hand, it has been decided to use neurophysiological measurements as an objective measurement technique. Two different sensors have been chosen.

For EEG data collection, the Emotiv Insight headset will be used, which has the advantages of being easy to connect, low-cost and wireless. This headset has already been used in previous studies focused on measuring EEG data applied to ATCOs (as an example, see [4]).

Eye-tracking techniques have been used in numerous experiments in the field of human factors applied to air traffic controllers before (as an example, see [5]). In this case, the GP-3 eye-tracker from the manufacturer Gazepoint will be used. This sensor is placed just in front of the controller, which makes it ideal, as it is non-intrusive and allows the participants complete freedom of movement.

6. Preliminary results

The preliminary results presented in this section correspond to the first analysis carried out after the collection of experimental data after the first weeks of simulations.

In this analysis, the first five participants in the simulations have been considered. All of them are students with theoretical knowledge of ATC. During the first part of the simulation campaign, the first four exercises, i.e., the two exercises of the basic phase and the first two exercises of the intermediate phase, were performed on the simulation platform by the first five participants in the experiment.

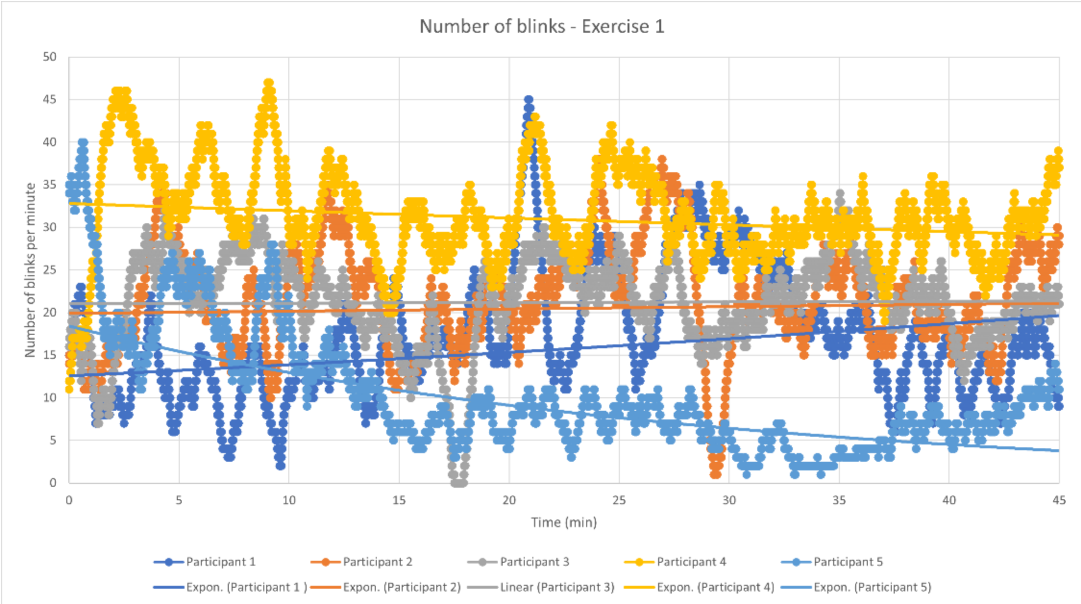
The following are the first conclusions drawn from the analysis of experimental eye-tracking and EEG data.

6.1. Eye-tracking data analysis

After the first analysis carried out, the visual monitoring parameter of greatest interest for the future establishment of workload thresholds was the number of blinks.

To study this parameter, a series of graphical representations were made. For each exercise, the curve relates the number of blinks accumulated at 60-second intervals (Y-axis) to the associated timeline represented in minutes (X-axis). An exponential trend line associated with each participant is included to study the evolution of the parameter over the course of each exercise.

As an example of the representations obtained, figure 6 shows the evolution of the number of blinks for exercise 1 (top) and exercise 2 (bottom) grouped for the 5 participants in the experiment.



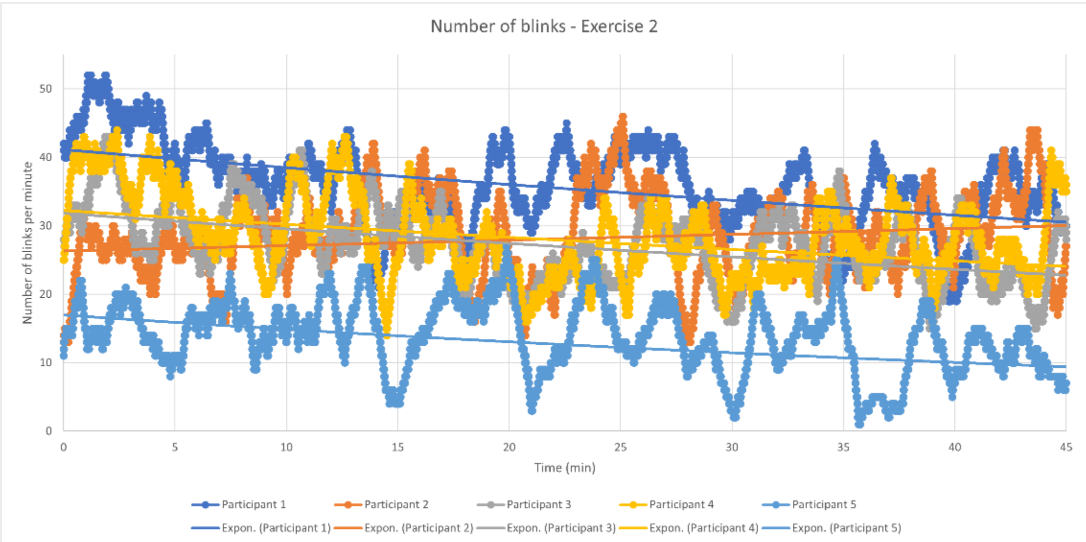


Figure 6. Number of blinks results: exercise 1 (top) and exercise 2 (bottom).

The representation of the data from the first exercise has been included because the difference between the trends of the different subjects is remarkable. This divergence in the results is associated with the fact that for the five participants, it was the first exercise performed on the simulator. The different tendency of each of the subjects is conditioned by the particular difficulties encountered by each of them.

In the graph for exercise 2, a substantial change can be seen in the trends described in the case of each participant, showing a clear greater homogeneity between the participants with respect to exercise 1. The trends for participants 1, 3, 4, and 5 are downward, while for participant 2 the trend is still slightly upward. In the case of the first four, the comparison of the two exercises would show a clear evolution of the concentration demanded in exercise 2 with respect to exercise 1.

According to theory, what would be expected to consider the number of blinks as a parameter of interest for establishing the workload limits would be an inversely proportional relationship between the value of the blink frequency and the level of concentration demanded. This criterion is met by comparing the graphs of exercises 1 and 2. The existence of such a correlation has also been verified by considering exercises 3 and 4.

Based on the first results analysed, the number of blinks is considered to be the most interesting parameter studied in relation to visual tracking, and, in this first analysis, its use to establish the workload limits of the controllers is valid.

Other parameters analysed were the number of fixations per minute and the diameter of the pupil. With this first sample, it was not possible to establish any correlation between their values and the workload. In the future, when a larger sample of data is recorded, the study will continue.

6.2. EEG data

The processing of EEG data is more complex than that of visual tracking data. In this first analysis, an attempt was made to establish a relationship between the six performance parameters provided by the headset manufacturer and the progress of the simulations. The parameters studied are engagement, stress, excitement, relaxation, interest, and focus. Each of these parameters can take values between 0 and 100.

The preliminary conclusions drawn from this first sample of data are as follows.

- In relation to the engagement parameter, a comparison of trends between different exercises shows a directly proportional relationship with the workload. The highest values of this parameter are found at the moments when the most difficult events occur.
- Similarly, there is a relationship between the upward trend of the stress parameter and exercises with higher levels of difficulty.
- In the case of focus, a certain relationship has been detected with the most difficult events. When these events (conflicts) occur, maximum values in the level of focus are observed.
- For the remaining three parameters, that is, excitement, relaxation, and interest, with the first data sample, it was not possible to establish any correlation with workload.

7. Conclusions and future works

Compared to other high-hazard industries, air traffic management is still human-centred [6]. This is why ATCOs are essential. The workload to which they are subjected can potentially affect the safety of operations. For this reason, it is of interest to study the workload of ATCOs and to define the limits within which it should fall to ensure that they perform their tasks safely and efficiently.

As a preliminary step before starting the recording of experimental data to establish the workload thresholds, this project succeeded in creating and fine-tuning a simulation platform specially designed for this experiment. This platform consists of an IT part, sensors, an exercise programme based on simulations, and functionalities that complete the basic configuration of the simulator.

The set of exercises and the methodology followed for their design have been a major milestone in the project. Likewise, the basic configuration of the radar screen of the ATCO positions has been completed to record all the necessary data during the simulations.

In the first tests carried out, the simulation platform was validated for its use in establishing the workload limits of the controllers using experimental human factor performance data as a reference.

The simulation campaign is currently in development. With a first sample of data, the first analysis of EEG and visual tracking data was performed.

In relation to visual tracking data, the preliminary results establish that there is a clear relationship between the decrease in the number of blinks as the difficulty of the exercises increases. This parameter is a candidate for setting workload limits.

With respect to the EEG data, a first correlation has been established between the values of the engagement, stress, and focus parameters with the moments of greatest difficulty of the exercises. After the first analysis, these three parameters would be of interest in establishing the workload limits.

However, these initial conclusions are limited by the limited size of the sample analysed, as only data from the first five participants after having performed the first four exercises is available. Work will continue to establish relationships after recording a larger sample of data and having analysed the entire sample. Once a more in-depth analysis has been achieved, it will be possible to establish workload thresholds.

References

[1] Gómez Comendador VF, Arnaldo Valdés RM, Vidosavljevic A, Cidoncha MS, and Zheng S. 2019 Apr 24. Impact of trajectories' uncertainty in existing ATC complexity methodologies and metrics for DAC and FCA SESAR concepts. *Energies* **12**(8)

[2] Arnaldo Valdés RM, Iglesias J, Gómez Comendador VF, Crespo J, Pérez L, Alonso JF and Rodríguez Sanz Á. 2018. Computerized brain interfaces for adaptive learning and assessment. International Conference on Intelligent Human Systems Integration. *Intelligent Human Systems Integration* 237-41

[3] Socha V, Hanáková L, Valenta V, Socha L, Ábela R, Kusmírek S, et al. 2020. Workload assessment of air traffic controllers. In: *Transportation Research Procedia*. (Elsevier B.V) 243–51

[4] SESAR JU. 2017. D3.2 Competence and training requirements. AUTOPACE Project. H2020-SESAR-2015-1.

[5] Trapsilawati F, Herliansyah MK, Nugraheni ASANS, Fatikasari MP and Tissamodie G. 2020 May 1. EEG-based analysis of air traffic conflict: investigating controllers' situation awareness, stress level and brain activity during conflict resolution. *Journal of Navigation*.**73**(3):678–96

[6] Wee HJ, Lye SW and Pinheiro JP. 2020. Monitoring performance measures for radar air traffic controllers using eye tracking techniques. *Advances in Intelligent Systems and Computing*. Springer Verlag 727–38

[7] Cokorilo O. 2013. Human factor modelling for fast-time simulations in aviation. *Aircraft Engineering and Aerospace Technology* **85**(5):389–405