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

María Zamarreño Suárez, Rosa María Arnaldo Valdés, Francisco Pérez Moreno and Raquel Delgado-Aguilera Jurado

Department of Aerospace Systems, Air Transport and Airports, Universidad Politécnica de Madrid, Madrid, Spain

Patricia María López de Frutos

ATM Research and Development Reference Centre (CRIDA), Madrid, Spain, and

Víctor Fernando Gómez Comendador

Department of Aerospace Systems, Air Transport and Airports, Universidad Politécnica de Madrid, Madrid, Spain

Abstract

Purpose – Air traffic controllers (ATCOs) play a fundamental role in the safe, orderly and efficient management of air traffic. In the interests of improving safety, it would be beneficial to know what the workload thresholds are that permit ATCOs to carry out their functions safely and efficiently. The purpose of this paper is to present the development of a simulation platform to be able to validate an affective-cognitive performance methodology based on neurophysiological factors applied to ATCOs, to define the said thresholds.

Design/methodology/approach – The process followed in setting up the simulation platform is explained, with particular emphasis on the design of the program of exercises. The tools designed to obtain additional information on the actions of ATCOs and how their workload will be evaluated are also explained.

Findings – To establish the desired methodology, a series of exercises has been designed to be simulated. This paper describes the project development framework and validates it, taking preliminary results as a reference. The validation of the framework justifies further study to extend the preliminary results.

Research limitations/implications – This paper describes the first part of the project only, i.e. the definition of the problem and a proposed methodology to arrive at a workable solution. Further work will concentrate on carrying out a program of simulations and subsequent detailed analysis of the data obtained, based on the conclusions drawn from the preliminary results presented.

Originality/value – The methodology will be an important tool from the point of view of safety and the work carried out by ATCOs. This first phase is crucial as it provides a solid foundation for later stages.

Keywords Human factors, Air traffic controllers (ATCOs), Real-time simulations, Electroencephalography (EEG), Eye-tracking

Paper type Research paper

1. Introduction

Air traffic controllers (ATCOs) perform a number of key tasks in the safe, orderly and efficient management of flights. These tasks include monitoring, control and decision-making, which are cognitive activities by nature (Kallus *et al.*, 1999).

Airspace is an invisible infrastructure. ATCOs are under great pressure to manage this infrastructure safely and efficiently (Mogtit *et al.*, 2020).

Due to the very nature of their work, ATCOs are required to communicate with aircraft, coordinate with other control units, issue authorisations and make decisions in a short time and in high-pressure environments.

Furthermore, any modification introduced into the setting or any change in the tools with which controllers work on a day-to-day basis can have a significant impact on their working conditions and the tasks they have to carry out. For this reason,

different studies have been undertaken to model the impact that the future evolution of European air traffic management (ATM) will have on the current Air Traffic Control (ATC) system and, specifically, on ATCOs (by way of example, see Netjasov *et al.*, 2019 and Gómez Comendador *et al.*, 2019).

To understand and mitigate threats to the performance of ATCOs, it is essential to understand the influence of human factors (Edwards *et al.*, 2017).

The work of ATCOs is of particular interest in workload studies because of three relevant characteristics: the potential to experience high levels of mental workload, the need for constant supervision of the aircraft under their responsibility and the high risk of the work performed (Triyanti *et al.*, 2020).

In light of the above, it is obvious that the role played by controllers within ATM is fundamental. Therefore, it makes sense to study their working conditions due to the impact this has on operational safety.

The workload that controllers are subjected to, be it too high or too low, can cause problems when it comes to performing their functions properly. To enhance the safety standards, a systematic and accurate assessment of the workload of ATCOs should be conducted (Socha *et al.*, 2020).

Bearing this in mind, it would be ideal to know the workload thresholds within which ATCOs can perform their work safely and efficiently.

The objective of this project is to develop a methodology to determine these thresholds, using experimental data on affective-cognitive parameters to assess the human factor.

The use of neurophysiological parameters follows the trend established by other studies of interest in the field of human factors (Aricò *et al.*, 2016). The fundamental advantage of these parameters is that they can be recorded continuously during the development of the experiment.

This methodology has a series of characteristics that make it innovative. To obtain data on neurophysiological parameters, the subjects will perform a series of exercises, specifically designed for this project, which are based on simulations. Using this program of exercises, it is possible to know in advance when the controllers will be subjected to greater overload, as well as the moments in which previously designed conflicts will appear.

The intention is to afterwards compare the experimental results with difficulty profiles designed before carrying out the simulations.

In similar studies, where active ATCOs are used as subjects, one of the fundamental limitations of the results obtained is the small sample of participants, which makes it difficult to generalise the conclusions drawn (for example, see Trapsilawati *et al.*, 2017).

To solve this problem and increase the time spent on the simulator, in this study, it is proposed to use students with previous theoretical training in ATC instead of controllers in active service. For the simulation campaign, a sample of 37 participants will be considered, 26 men and 11 women. The age of the participants is 21 ± 3 years. All are students with previous knowledge in aeronautics and ATC but who have not worked with the simulator before. The entire sample is expected to perform the first four exercises of the set in the coming months.

To measure neurophysiological data from ATCOs (in this first stage of the project, university students), two sensors will

be used: a headset to obtain encephalographic data and an eye-tracker to obtain visual tracking data.

In recent decades, research has focussed on the evaluation of user's mental states based on their neurophysiological activity in operating environments (Aricò *et al.*, 2015a, 2015b).

In line with the above, a wireless headset will be used in this study to collect electroencephalographic (EEG) data.

One of the advantages of EEG techniques is that they allow the recording of data with high-resolution times, which is well suited to the tasks performed by ATCOs (Liu *et al.*, 2017).

In reference to visual monitoring, within the literature, there are numerous references to experiments related to the human factor applied to controllers, where these techniques have been used (as an example, see Imants and de Greef, 2011 and Wee *et al.*, 2020).

In the specific case of this project, an eye-tracker installed on a support will be used. This will be placed in front of the controller. Using this sensor, visual data will be recorded, such as the number of blinks, the number of gaze fixations and their duration.

In both cases, i.e. EEG headset and eye-tracker, low-cost equipment has been chosen. This facilitates the extrapolation of the results obtained to other operational scenarios.

In addition, the sensors have been chosen on the basis of the criterion of being as little intrusive as possible. Also, the time required to set up both devices is minimal.

By using two different devices, eliminates an important limitation that the study could otherwise suffer from if the workload was assessed by focusing on a single source of neurophysiological data. (Miyake *et al.*, 2009).

In addition to the measurements obtained from the sensors, subjective workload assessments will be recorded by the subjects.

In similar studies in the literature, NASA-TLX questionnaires are commonly used for this purpose (Triyanti *et al.*, 2021). The questionnaire is completed by the participants after the end of the exercise or activity. In this study, the aim is to achieve real-time workload assessments during exercise performance. For this reason, a window has been designed that, at regular intervals, will appear on the radar screen to ask participants to evaluate their workload on a scale of 1–5.

This paper describes the first milestones achieved in a project that aims to establish a methodology for determining workload thresholds for controllers to perform their work safely and efficiently. Compared with other similar projects, this study has a number of innovative features, among them:

- The exercises to be carried out in the simulations have been specially designed for this study. By knowing the difficulty profiles of the exercises before running the simulations, the data obtained can be compared with the expected data.
- From the beginning of the project, it has been decided to combine objective and subjective workload assessment techniques.
- As objective techniques, it has been decided to use EEG and eye-tracking data measurements in parallel. In both cases, low-cost equipment has been used, which would facilitate the idea of extending the results achieved in this methodology to other operational scenarios. The main objective of this paper is to present the framework for data

collection and to justify with preliminary results that this framework is valid for further large-scale data collection and analysis. It is structured as follows. Section 2 sets out the methodology to be used in the project, whereas Section 3 details the simulation platform. Section 4 explains the most important features of the exercise program and its design. Section 5 specifies the objective and subjective methods used for workload assessment. Subsequently, a series of preliminary results are presented in Section 6, following the analysis of the first experimental data recorded. Finally, the general conclusions of the paper and lines of future work appear in Section 7.

2. Methodology

The methodology comprises six stages in total. This paper details the objectives achieved after carrying out the first three stages and some preliminary results after starting with the development of the simulation campaign, the fourth stage.

Figure 1 gives the stages of the project, from the initial analysis to the final definition of the methodology. In addition, it indicates which phases have been completed, what stage is currently ongoing and the future development of the project.

The first step consisted of a general analysis of the problem. This more conceptual stage was followed by a second stage in which all of the elements of the simulation platform were defined. These included the sensors for taking measurements, the conceptual design of the exercises and their main characteristics and the additional functionalities required to complete the basic configuration of the simulator.

Having defined the above, it was time to fine-tune the simulation platform. In this third stage, the exercises were implemented in the simulator. These exercises were then tested to detect areas of improvement which led to new functionalities being identified. These functionalities were incorporated into the platform and tested from the control position screen.

These first three stages made it possible to define the framework with which it is expected to work to obtain valid results. Before being able to proceed to massive data record and analysis, it was necessary to carry out a validation of the entire framework.

For this purpose, the first data recorded during the simulation campaign are used. These preliminary results can be found in Section 7.

After all, components have been designed and are ready for use, the simulation campaign could begin. Participants in the study will use the platform and perform the exercises that have been specially designed for this project. The simulations have already started and will be run over the coming months.

The data obtained in this simulation campaign will be further analysed. The first step will be to study the behaviour of the different variables to decide which of them are of interest when defining the workload thresholds. Because the simulation campaign is still under development, what is presented in this paper is a series of preliminary results. This first analysis is fundamental, as it is the first contact with the data obtained. All further analysis will be conditioned by choice of the reference variables.

3. Simulation platform

In general terms, the simulation platform comprises four elements that must work together in a coordinated manner to obtain the desired results. These are the hardware and corresponding software, the sensors, the exercise plan and finally, the additional features that have been added to the basic configuration of the simulator to provide further information on the actions carried out by the subjects performing the exercises.

The platform consists of five personal computer (PCs) and their corresponding screens. One PC acts as the server and stores the results of the simulations. Another PC, the manager, launches the simulations and coordinates the other computers. A third PC, the editor, uses radar operation simulator and editor

Figure 1 Three-year development foreseen in the project methodology: stages completed, under development and future stages
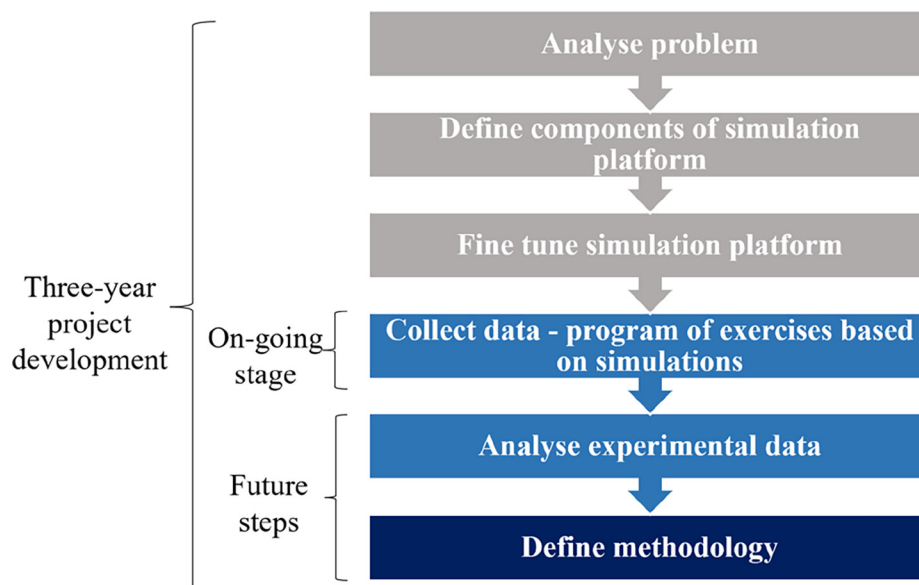
(ROSE) simulation software and is used to design the simulation exercises. Finally, there are two ATM positions where subjects perform ATM tasks. In this case, SkySim software from the Swiss air navigation service provider, Skysoft-ATM, is used. Using this software, participants in the study can see on their radar screen the airspace that controllers working at the area control centres (ACCs) would observe.

The simulations in this project involve different sectors of Madrid ACC, one of the five Spanish ACCs from which (ENAIRE, 2021) provides area control service.

The EMOTIV insight headset, marketed by EMOTIV and the GazePoint (GP3) eye-tracker, marketed by Gazepoint, are used to collect neurophysiological data.

The set of exercises comprises seven totally new exercises specifically designed for this project.

It is worth noting that an additional tool, the so-called flight strip bay, was designed to be displayed on the screen of the ATCO position. This houses the flight progress strips of the different aircraft.

This allows the designer to introduce a series of discrepancies between the flight plans programmed in the editor and the actions that the controller is intended to carry out. For example, for some flights in the exercises, certain waypoints or the flight level have been modified. The aim is to identify the level of attention of the controller, who must check each of the flights before they enter the sector, during their stay and when they leave the sector.

Finally, additional windows with specific functionalities have been designed to appear on the controller screen to record further information about the actions performed by the participants in the experiment.

From the recorded data, it is possible to determine the precise moments in which the participants in the experiment identify a conflict, assume an aircraft under their responsibility or transfer an aircraft that leaves their sector. It is also possible to know if they carry out these two previous actions in accordance with the information on the flight progress strip of the aircraft in question.

Figure 2 gives an overview of the control position, whereas one of the simulations is being carried out. The person performing the role of controller is wearing the headset, which records encephalographic data. The eye-tracker is placed directly in front of them to track their gaze. The screen shows a representation of the Zamora (ZM) sector within Madrid ACC and the aircraft populating the sector.

The flight progress strip that is currently selected appears in the top righthand corner of the screen. The small windows on the radar screen are those that have been added to the basic configuration of the simulator to provide additional data when analysing the actions carried out by the controller.

Furthermore, an additional window appears at regular intervals on the left side of the screen that asks the controller to subjectively evaluate the workload they are under.

4. Design of the exercise program

4.1 General features

The program of exercises used to validate the methodology has been specifically designed for this project.

Figure 3 shows the general features of the exercise program.

Figure 2 Overview of the control position and sensors

Each of the participants in the experiment must perform seven exercises on the simulator, each lasting 45 min. The first minute and a half of each exercise have been programmed to be free of events so that each participant can adjust the position of the radar screen to best suit themselves. Before beginning the program of exercises, there is an initial dummy exercise to allow the subjects to familiarise themselves with the set-up.

The exercises have been designed so that they can be simulated in parallel by two controllers. Therefore, each of the exercises has been programmed to be simulated in two different sectors simultaneously.

The seven exercises are grouped into three phases, depending on their difficulty: basic, intermediate and advanced. The exercises have been designed in such a way that the difficulty increases progressively. The easiest exercise is the first (with a score of 140 points), and the most difficult is the last (with a score of 320 points).

The fact of structuring the exercise programme in phases of difficulty is recurrent in similar projects. Specifically, in this study, the naming of the different phases has been carried out, taking the SESAR AUTOPACE project as a reference (SESAR JU, 2017). The number of exercises and their duration have been defined considering the optimisation of previous work with the simulation platform and the subsequent organisation of the simulation campaign.

In the design of the exercises, a series of events have been defined, to which a score has been assigned.

The scoring system was designed following an iterative process. It takes the prior experience of colleagues and other expert users with the simulator into account as well as information gleaned from detailed interviews with the subjects after having carried out each of the simulations.

When designing the exercises, the different events were programmed at specific moments to achieve pre-set, so-called difficulty profiles. From these profiles, it is possible to estimate the workload to which the controllers will be subjected during the simulations. Having carried out the simulations, it will be

Figure 3 General features of the exercise program
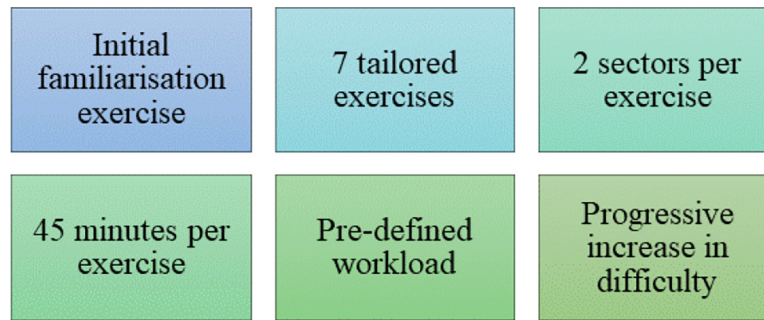
possible to analyse the data in detail and see which parameters comply with the expected outcome and which deviate from it.

4.2 Event scores
The first step in designing the exercises was to consider a series of basic events to be included in the simulations. These were chosen based on previous experience with the simulator.

Table 1 gives the events, their duration and the score assigned to each.

The program of exercises encompasses twelve different events. It is necessary to know the score that has been assigned to each as well as its average duration on the simulator.

This is required because, in the preliminary design of the exercises, the difficulty profiles are calculated for each of the minutes of the simulation.

The first event in the table, identification, refers to the moment when an aircraft appears on the screen and the controller detects it.

When accepting an aircraft, the flight plan programmed in the editor may or may not coincide with that set out in the flight progress strip. Depending on the case, the controller will either accept the aircraft or act on it before accepting it.

Similarly, when the aircraft is about to leave the sector, the controller may transfer it directly if it meets the conditions set out in its flight plan, or it may be necessary to take some action on it before transferring.

It was decided to consider five types of conflict in the simulations. In cruise-cruise conflicts, the two aircraft involved are at a certain flight level when the five nautical miles (NM)

separation between them is violated. In cruise-climb, cruise-descent and climb-descent conflicts, the principle is the same, except that one or both aircraft are in evolution at the time the separation is expected to be violated.

The conflict is designed so that, at a given moment, the speed of the second aircraft (the aircraft behind) increases progressively, approaching its predecessor. This causes the 5 NM minimum separation to be infringed and the conflict warning to be triggered. During the vectoring event, the controller gives vector guidance to the aircraft.

Of the 12 events listed in Table 1, the first 11 are absolute events. That is, the score characterises the occurrence of the event.

A particular case is the last event, the monitoring event. In this case, the associated score is expressed per minute. That is, the absolute score of the event will depend on the time the aircraft remains within the controller's sector of responsibility. The controller monitors the aircraft under his/her responsibility on a constant basis. Therefore, the longer an aircraft is in the sector, the greater the absolute workload associated with its monitoring.

Based on prior experience on the simulator, it was found that one parameter which significantly affects the difficulty of the events perceived by the controllers is the number of aircraft present in the sector at any given time. The greater the number of aircraft, the greater the perceived difficulty of the events.

Therefore, three levels of traffic density have been defined: low, medium and high. The scores that appear in Table 1 are the low scores, awarded when the number of aircraft in the sector is less than five.

When the number of aircraft is between five and nine inclusive, the traffic density is deemed to be medium and the scores that appear in the column "Base Score" are increased by 5%. Similarly, when the number of aircraft is 10 or greater, the traffic density is considered to be high and the scores in the table are increased by 10%.

In this way, the scores assigned to each of the events vary throughout the exercise depending on the number of aircraft in the sector at that time.

4.3 Design
When designing the exercises, a spreadsheet specially designed for this purpose was used. When the field for the name of the event is populated, the spreadsheet automatically assigns scores based on the event in question and the number of aircraft in the sector at that time.

Table 1 List of events, duration and scores

Event	Duration (s)	Base score
Identification	10	1
Acceptance	30	3
Acceptance with action	60	5
Transfer	30	3
Transfer with action	60	5
Conflict cruise-cruise	30	7
Overtaking	30	8
Conflict cruise – climb	30	9
Conflict cruise – descent	30	9
Conflict climb – descent	30	9
Vectoring	30	3
Monitoring (per minute)	60	0.1

Based on previous work carried out, a theoretical difficulty profile was defined for each phase of the exercises. These theoretical profiles are the starting point. Another input is the exercise design difficulty value that should be reached after including all events.

In view of the abovementioned, the second step consists of distributing the events throughout the entire duration of the simulation.

The spreadsheet automatically spreads the difficulty associated with each event over the corresponding minutes that comprise the total duration of the event.

The monitoring event needs to be discussed in detail. As stated above, all prior events are absolute.

To achieve a score for monitoring, it is necessary to know the times of entry and exit of the aircraft from the sector. Part of the spreadsheet is specially designed to handle this.

The total difficulty of the exercise (D_t) is given in equation (1) as the sum of the difficulty of each of the events (D_i), where i is each of the events included in the design of the exercise. $i = 1$ is the first of the events entered in the simulation and $i = n$ is the last event for each exercise. The value of "n" varies for each of the exercises because not all have the same number of events:

$$Dt = \sum_{i=1}^{n} D_i \qquad (1)$$

The design process is iterative and continues until the real difficulty profile is as close as possible to the theoretical one, and the total difficulty is as close as possible to the design one.

As an example, Figure 4 shows the difficulty profiles obtained after the design of Exercise 1, the first in the exercise program.

In the figure, the blue bars represent the theoretical difficulty profile, and the orange bars represent the difficulty profile of the exercise on completion of the design. As can be seen, the difficulty profile comprises two symmetrical cycles with a valley in the middle, where the workload is very low.

The evolution over time (in minutes) of the exercise is shown on the X-axis, and the difficulty values or scores are given on the Y-axis.

As previously mentioned, no event is introduced until one and a half minutes to allow the controller to adjust the position of the radar screen to best suit themselves. This means that the

first minute of the exercise has no events, as can clearly be seen in the graph above.

Looking at the Y-axis, it can be seen that, for Exercise 1, all the difficulty values are below 8.

By way of comparison, Figure 5 presents comparable profiles for Exercise 7: the most difficult of the exercises. In this case, the theoretical profile comprises two cycles, the second having slightly higher difficulty values than the first. There is also a valley; however, in this case, the difficulty does not decrease significantly.

It can be seen from the graph there are times when the difficulty (Y-axis) is above 10 in this exercise.

After designing the spreadsheet, the different events and aircraft were introduced to the ROSE Simulation software.

For each of the exercises, there are tables with a record of all of the programmed aircraft and events. These tables can be cross-checked with the records of the actions taken by the controller after each of the simulation exercises.

4.4 Designed exercise set

There are seven exercises in total. The first two comprise the basic phase in which the difficulty progresses from 140 to 160 points. Exercise 3, Exercise 4 and Exercise 5 comprise the intermediate phase with scores of 190, 220 and 260 points, respectively. Finally, the advanced phase is made up of Exercise 6 and Exercise 7, with scores of 280 and 320 points, respectively.

Table 2 lists the different exercises, the sectors in which they have been designed, and the overall difficulty score for each exercise.

Each exercise has been programmed for two sectors, hence the letters "A" and "B". One subject will carry out all the exercises denominated A, the other will perform the B exercises.

The specific sectors that the controllers are responsible for will change depending on the exercise. In this way, having completed the entire set of exercises, each controller will have been responsible for different sectors of Madrid ACC.

In the basic and intermediate phases, each of the subjects will be responsible for a single control sector. In the advanced phase, to introduce longer routes and a greater number of conflicts. The controllers are responsible for two sectors.

Note that due to its reduced size, Teruel (TER) sector has always been considered to be an adjunct to the Castejón (CJ) sector. These two sectors comprise a single block.

Figure 4 Theoretical and actual difficulty profiles of Exercise 1
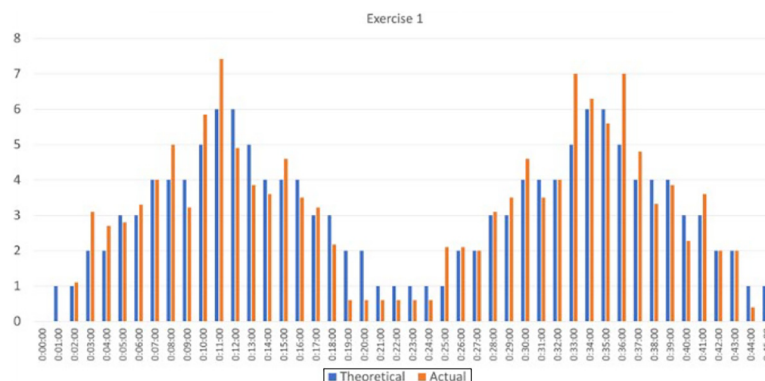
Figure 5 Theoretical and actual difficulty profiles of Exercise 7
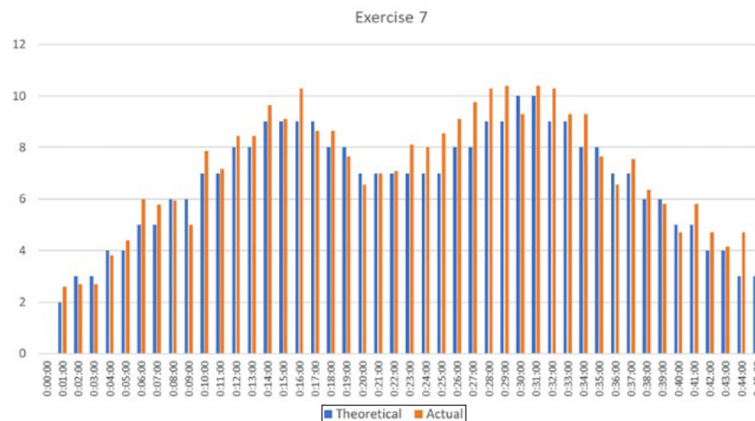
Table 2 List of exercises, sectors in which they have been designed and corresponding difficulty scores

Exercise	Sector	Difficulty score
Basic phase		
1A	TL	140
1B	AS	140
2A	ZM	160
2B	CJ + TER	160
Intermediate phase		
3A	AS	190
3B	TL	190
4A	CJ + TER	220
4B	ZM	220
5A	TL	260
5B	AS	260
Advanced phase		
6A	AS + BL	280
6B	CJ + TL + TER	280
7A	CJ + TL + TER	320
7B	AS + BL	320

Table 3 Number of aircraft, number of discrepant flights and conflicts in each exercise

Exercise	No. of aircraft	No. of discrepancies	No. of conflicts
Basic phase			
1A	14	0	3
1B	14	0	3
2A	14	4	4
2B	14	4	4
Intermediate phase			
3A	18	4	4*
3B	18	4	4*
4A	21	10	4*
4B	21	10	4*
5A	22	8	6**
5B	22	8	6**
Advanced phase			
6A	28	6	5*
6B	28	6	5*
7A	30	5	6*
7B	30	5	6*

The other features of the set of exercises are given in Table 3.

As can be seen in the table, the number of aircraft progressively increases as the exercises progress.

On a number of flights, some of the crossing points or flight levels in the flight progress strip have been deliberately modified to force the controller to act on the aircraft. In cases where there are discrepancies between the data on the strip and the flight plan programmed in the simulator, the information on the strip must prevail.

The number of discrepancies indicates the number of flights that have been designed in such a way that ATCO is required to perform certain actions on them to match the flight plan on the flight progress strip.

As can be seen in Table 3, the number of conflicts also progressively increases within each of the phases.

The values containing an asterisk identify a programmed conflict, where at least one of the aircraft is in evolution.

5. Assessment of workload

Mental workload may be assessed in two different ways: objective and subjective.

Subjective assessment of mental workload is usually carried out by asking the participants in the experiment to complete a questionnaire after they have finished performing a task. Objective assessment involves the measurement of neurophysiological parameters (Martin *et al.*, 2011).

In this study, the workload will be assessed both objectively and subjectively.

5.1 Objective assessment of workload

For the objective assessment of workload, it has been decided to measure neurophysiological data. The two sensors used are shown in Figure 6.

On the left, the EMOTIV insight headset for recording EEG data is presented, and on the right, the Gazepoint GP3 eye-tracker for visual monitoring.

Figure 6 Sensors for recording neurophysiological data

From the point of view of EEG data analysis, the first step will be to study the performance metrics recorded using the EMOTIV Insight EEG headset. Headsets from this manufacturer have previously been used in several human factor studies applied to ATCOs (Kouba *et al.*, 2019).

The insight model has five hydrophilic semidry polymer electrodes (AF3, AF4, T7, T8, Pz) and two CMS/DRL references on the left mastoid process. Its sampling rate is 128 samples per second per channel. The battery lasts up to 8 h using the universal serial bus receiver and up to 4 h using Bluetooth low energy, allowing several simulations to be carried out in a single day. This wireless model has previously been used in experiments to measure the performance of ATCOs (Fitri Trapsilawati, 2020).

The GP3 eye-tracker uses a 60 Hz machine-vision camera in its processing system. It has 0.5–1 degrees of visual angle accuracy and ±15 cm range of depth movement. For its positioning, it is recommended to be placed approximately 65 cm from the user, with an upward angle pointing toward the face.

When analysing and interpreting the data recorded by the sensors, it will be of interest to have the time markers of the moments when the events that cause an increase in the workload of the controllers occur.

During the exercise design process, the time of the simulation at which the design events occur is available. However, it may be that each of the events is perceived by the participant in the experiment with some delay or that, due to the development of the exercise, other events are generated that were not initially planned.

To make these time references available, a series of buttons have been designed to appear on the radar screen at the control position. The participant in the experiment will press each of them to indicate certain actions.

This approach has been followed as it is not too intrusive and is similar to that used in other ATC experiments to collect information about the participants while they are performing the exercises. The buttons designed are as follows.

The first is a button that the ATCOs must press every time they detect a conflict. The simulator does have a conflict detection tool, and the aircraft involved in a conflict change colour when the separation minimum is violated; however, it may take time for the controller to detect a conflict, depending on the workload at the time.

Based on this same principle, two additional windows have been created that remain continuously on the controller's radar screen.

The first of these identifies aircraft with which there is a discrepancy between the flight plan programmed in the simulator and the one appearing in the flight progress strip.

Before allowing an aircraft to come under their responsibility, controllers must press the "YES" button if they observe a discrepancy. Otherwise, they press "NO". Once the exercise has been completed, and if carried out correctly, the results should show as many Yes/No replies as the number of aircraft that have entered the sector.

Because information is available on which aircraft have been programmed with discrepancies and which have not, the results can be reviewed afterwards to identify deviations. In this way, it is possible to identify potential workload peaks which do not allow controllers to adequately fulfil all the simultaneous tasks required of them.

There is a similar window to deal with the transfer of aircraft to the adjacent sector. In this case, the controller must press "YES" if the conditions under which the transfer was made coincide with the information given in the flight progress strip. Alternatively, the controller presses "NO".

The minute of the simulation in which the controller presses each of the buttons is recorded in a file that is saved on the PC. Using these time references, the data recorded by the sensors can be analysed, considering the actual actions performed by the participants in the experiment and comparing them with the design actions.

5.2 Subjective assessment of workload

To complement the objective assessment, a decision was taken to include a subjective evaluation of the workload by the controllers themselves when carrying out the exercises.

To achieve this, a window was designed that appears on the left of the ATCO radar screen at regular intervals throughout the exercises. It is an adaptation of the instantaneous self-assessment measurement method, one of the most frequently used measures of mental workload in real-time simulations (SESAR_HP Repository, 2012).

Every two and a half minutes, controllers are asked to rate their workload on a scale of 1–5, with 1 being the smallest workload value and 5 being the highest.

As has been shown in previous studies, asking controllers to assess their workload at these intervals does not interfere with the main task performed under normal conditions (López de Frutos *et al.*, 2019).

Specifically, the controller is asked to consider the current workload, regardless of the exercise they are performing or its position within the exercise program.

The window appears on the screen for 20 s. After that time, if none of the numbers has been selected, the option "NOT PRESSED" is recorded.

The window will not appear in the initial minutes of the simulation because during that time, it is assumed that the subject will be adjusting the radar screen to best suit themselves and could evaluate the workload due to this initial configuration rather than that directly related to the occurrence of events.

All the responses are recorded in a file that indicates the option selected and the moment of the simulation in which it was chosen. These responses will allow a comparison of the controllers' subjective perception of workload with the difficulty profile of the exercise design.

6. Preliminary results

In this section, the first preliminary results are presented regarding the recorded neurophysiological data and the subjective assessment of workload by the participants in the study.

To produce the graphs in this section, data from the first five participants in the experiment have been considered. In this first stage of the simulation campaign, the participants performed the first four exercises of the plan, i.e. the two exercises of the basic phase and the first two exercises of the intermediate phase.

The objectives to be achieved with these preliminary results were the following:

- In relation to the measurements recorded by the eye-tracker, the idea was to study the behaviour of certain parameters and to define, on the basis of these first results, which ones would be of interest when establishing future workload thresholds, as they present a logical behaviour as the difficulty of the exercises progresses.
- With regard to the EEG data, the objective was similar. Of the six performance metrics, the aim was to try to identify some kind of correlation between the recorded data and the expected trends to present certain parameters as candidates for establishing workload thresholds based on them.
- In the case of the subjective workload assessments, the aim was to compare the results obtained and their graphical representation with the exercise design difficulty profile to validate that the perceived workload matches the expected results.

In the following subsections, the conclusions drawn on the basis of the previous objectives are presented. Similarly, the first graphs and summary tables produced are included.

6.1 Objective assessment of workload: visual parameters

After studying different visual parameters, it has been decided that the parameter of greatest interest for establishing workload thresholds is the number of blinks. As it has already been demonstrated in previous studies, the number of blinks is a valid feature for assessing workload (Aricò *et al.*, 2015a, 2015b).

According to the literature, it would be expected that the blink rate would decrease as the workload to which the controller is subjected increases.

To study this parameter, a graphical representation was made for each of the exercises, considering the samples of the first five participants.

In each of the graphs, the *Y*-axis represents the number of blinks accumulated in an interval of 60 s. On the *X*-axis, the time from the start of each of the exercises is represented.

To clearly represent the trend of the parameter for each of the participants, the representation of an exponential trend line associated with each of the participants has been included.

Figure 7 shows the graph corresponding to Exercise 1 (top) and that corresponding to Exercise 2 (bottom).

These representations have been made for all four exercises. It was considered important to include the graph for the first exercise to highlight the difference in trends for the different participants.

Figure 7 Number of blinks from the first five participants in the study for Exercise 1 (top) and Exercise 2 (bottom)
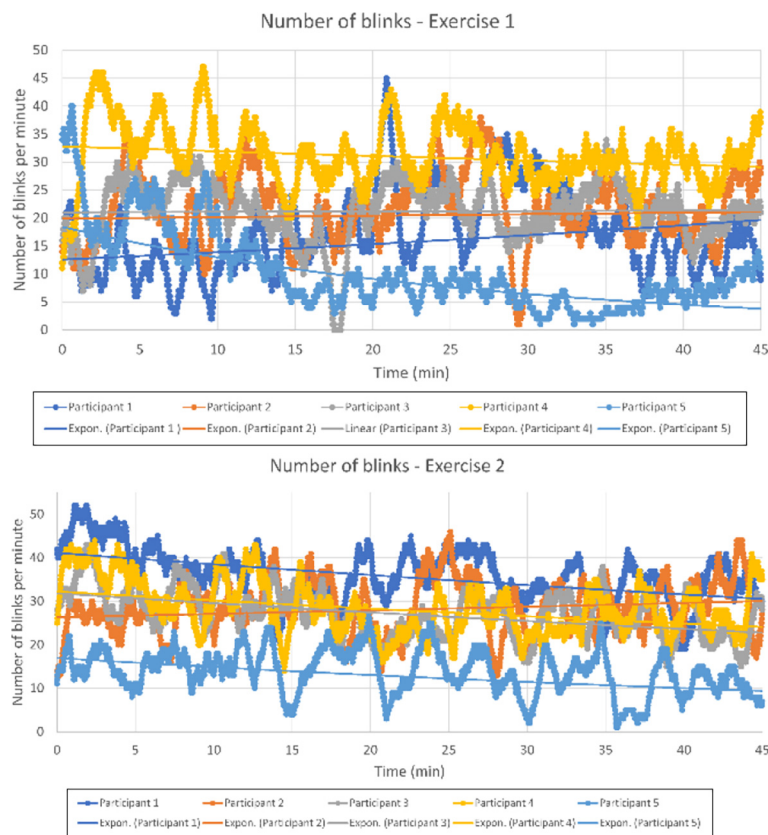
As this is the first exercise performed by them on the simulator, each of the participants will have faced a different set of difficulties.

In the case of the data from Exercise 2, greater homogeneity in the trends can already be seen. For most of the participants (participants 1, 3, 4 and 5), a downward trend can be observed.

Comparing the two graphs, it can be seen that the number of blinks decreases with increasing difficulty and, therefore, the concentration required. This correlation has also been identified by comparing the graphs of Exercise 3 and Exercise 4.

Based on these preliminary results, the expected correlation between the number of blinks and the difficulty of the exercises is fulfilled. For this reason, the number of blinks would be a variable chosen to express the workload thresholds in the later stages of the study.

6.2 Objective assessment of workload: electroencephalography data

In the case of EEG data, in this first analysis, a relationship was established between the values obtained for the six performance parameters provided by the headset manufacturer and the evolution of the simulations.

Each of the parameters can take values between 0 and 100. As with the visual parameters, graphical representations were made to study the behaviour of each of the six parameters.

Table 4 summarises the first results obtained. It also indicates those parameters with which it has not been possible to establish a clear correlation. In the future, the study will be repeated with a larger data sample.

Based on preliminary results, it has been possible to detect a certain relationship between the parameters of engagement, stress and focus and the evolution of the simulations.

At the moment, these three parameters would be candidates to be used in defining workload thresholds. However, the size of this first sample is too small to draw definitive conclusions.

With a larger volume of data, the analysis will be repeated to try to establish additional relationships with the other parameters for which it has so far not been possible to do so.

6.3 Subjective assessment of workload

For each of the exercises, the responses recorded by each of the participants as the simulations progressed were represented graphically.

The *Y*-axis of these graphs represents the workload value assessed by each of the participants, on a scale of 1–5. It was detected that some participants did not press any of the buttons the first time they were asked to do so and also at the moments of greatest workload. The absence of response in the first minutes was assumed to be "1". Following the same reasoning, the absence of response in the intervals with higher workload values has been interpreted as "5".

In the graph, the mean of the results obtained for the five participants is shown as a dashed line. Using the graph, the values recorded were compared with the difficulty profile of the exercise design, and it was found that, in general, the workload perceived by the participants coincided with what would be expected, taking the theoretical difficulty profiles as a reference.

As an example, Figure 8 shows the graph for Exercise 3. In the representation, the workload values assessed by each of the five participants and the trend line of the mean values can be seen.

There are a number of parameters that allow the results to be validated:

- In the theoretical design profile, the exercise was set up with two workload cycles, the second one being more difficult than the first one. It can be seen that the workload values collected are higher in the second half of the exercise than in the first.
- During the first cycle, the most difficult events (especially conflicts) were programmed around minute 12. Participants will need a few minutes to be able to resolve them, which translates into high expected workload values in the minutes after minute 12. This can be confirmed in the graphical representation shown in Figure 8.
- In the second cycle, the most difficult events were expected around minute 29, with the total score of the events entered being higher than in the case of the first cycle. High workload values are expected to be recorded around minute 29. The hypothesis can be verified by the results shown in Figure 8. All participants selected high workload values in the minutes after minute 29, where they have to manage the increase of tasks to be performed to cope with the introduced events.

A similar study has been conducted for the remaining three exercises. In view of these preliminary data, it can be concluded that, in general, the workload values perceived by the participants are consistent with the theoretical difficulty profiles used in the exercise design.

Table 4 Preliminary results on the behaviour of the performance metrics during the progress of the simulations

Parameter	Behaviour observed
Engagement	After comparing the trends of the different exercises, it is found that there is a directly proportional relationship between engagement and difficulty. The highest values of this parameter are found at the times when the most difficult events occur
Stress	There is a relationship between the upward trend of the stress parameter and exercises with higher levels of difficulty
Excitement	On the basis of these initial data samples, it has not been possible to draw any clear
Relaxation	conclusions on the evolution of these parameters
Interest	
Focus	A relationship has been observed between the increase in the value of this parameter at times when the most difficult events occur

Figure 8 Subjective evaluation of the workload in Exercise 3: values assessed by each participant and average values
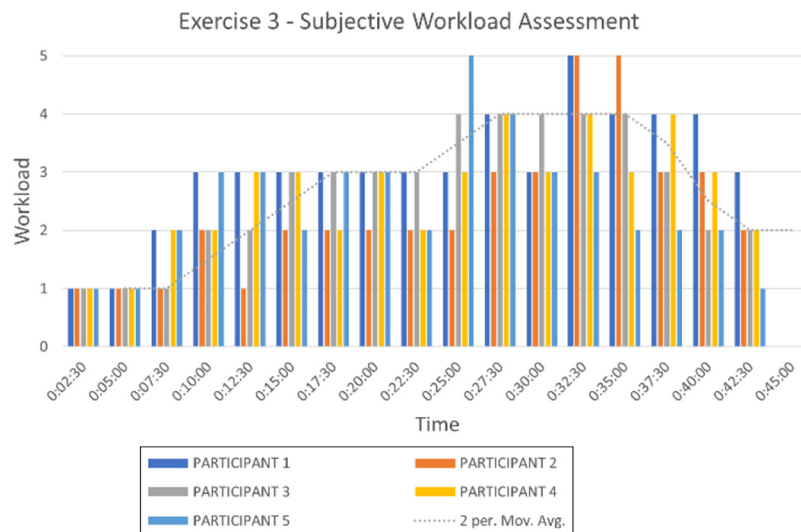
7. Conclusions and further work

Compared with other high-hazard industries, ATM is still "human-centred" (Cokorilo, 2013). Due to the importance of the tasks performed by ATCOs, it is of interest to assess their workload to determine the optimal thresholds of this. The first necessary part of the study is to establish a framework and validate it to ensure the validity of the subsequent conclusions.

This paper describes the process of creating and validating such a framework based on a series of preliminary results.

An *ad hoc* simulation platform has been set up for this study, and it has been decided to use objective and subjective methods for the evaluation of the ATCO workload from the outset.

A series of seven exercises have been created to be simulated by the participants in the study. The innovative aspect of these exercises is that they have been created from pre-set difficulty profiles. In this way, certain expected results can be predicted before the simulations are carried out. The first experimental subjective workload assessment data largely confirm that the assumptions initially made regarding the perceived workload of the participants were correct.

From the point of view of objective workload measurement, it has been decided to use low-cost sensors to record EEG and visual tracking data. The use of these sensors is also considered innovative as it will allow the methodology to be easily extrapolated in the future to real operating scenarios, where additional factors to those studied in a laboratory can be considered.

In this paper, the preliminary results obtained from the first five participants in the simulation have been analysed. With such a small sample size, it is not possible to draw general conclusions about the parameters. The aim in this state was, after analysing the behaviour of the variables, to propose certain parameters that could serve as a reference for establishing workload thresholds in the future.

With reference to visual data, the parameter of greatest interest was the number of blinks. Based on these preliminary results, it would be of interest to use this variable when setting workload limits.

On the other hand, in relation to EEG parameters, initial correlations have been detected between increases in exercise difficulty and performance metrics related to engagement, stress and focus. These variables would be the first candidates to be considered when establishing thresholds related to brain activity.

To draw more general conclusions, once the designed framework has been validated, it is time to move on to massive data collection. To this end, 37 participants, all of them students with previous notions in ATC, will take part in the simulation campaign that will be extended over the next months.

Future work will focus on the detailed analysis of all the data to be recorded. Once the data have been analysed in detail, it will be possible to move on to the definition of workload thresholds. In future simulation campaigns, it could be interesting to include in the study other parameters to be analysed, such as the pre-state of the subjects. Once the entire methodology will be defined, this project could significantly improve the activity carried out by ATCOs, as operational decisions can be taken at different horizons taking their expected workload levels as a reference. Similarly, another possible application would be the determination of the optimal number of aircraft that controllers could safely manage according to their expected workload.

References

Aricò, P., Borghini, G., Graziani, I., Imbert, J.P., Granger, G. and Benhacene, R. (2015a), "ATCO: neurophysiological analysis of the training and of the workload", *Ital. J. Aerosp. Med*, Vol. 12, pp. 18-34.

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Graziani, I., Imbert, J.-P. and Granger, G. (2015b), "Reliability over time of EEG-based mental workload evaluation during air traffic management (ATM) tasks", *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Milan.

Aricò, P., Borghini, G., D., Flumeri, G., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R., (2016), "Adaptive automation triggered by EEG-Based mental workload index: a passive Brain-Computer interface application in realistic air traffic control environment", *Frontiers in Human Neuroscience*, Vol. 10 No. 539, and Babiloni, F.

Cokorilo, O. (2013), "Human factor modelling for fast-time simulations in aviation", *Aircraft Engineering and Aerospace Technology*, Vol. 85/ No. 5, pp. 389-405.

Edwards, T., Homola, J., Mercer, J. and Claudatos, L. (2017), "Multifactor interactions and the air traffic controller: the interaction of situation awareness and workload in association with automation", *Cognition, Technology & Work*, Vol. 19 No. 4, pp. 687-698.

ENAIRE (2021), "Control de tráfico aéreo (ATC)", available at: www.enaire.es/servicios/atm/servicios_de_transito_aereo_ats/control_de_trafico_aereo_atc (accessed 29 Agosto 2021).

Fitri Trapsilawati, M.K. (2020), "EEG-Based analysis of air traffic conflict: investigating controllers' situation awareness, stress level and brain activity during conflict resolution)", "The", *Journal of Navigation*, Vol. 73 No. 3, pp. 678-696.

Gómez Comendador, V.F., Arnaldo Valdés, R.M., Vidosavljevic, A., Sánchez Cidoncha, M. and Zheng, S. (2019), "Impact of trajectories' uncertainty in existing ATC complexity methodologies and metrics for DAC and FCA SESAR concepts", *Energies*, Vol. 12.

Imants, P. and de Greef, T. (2011), "Using eye tracker data in air traffic control", 29th annual conference of the European Association of Cognitive Ergonomics. Rostock.

Kallus, K. Van Damme, D. and Dittman, A. (1999), "Integrated job and task analysis of air traffic controllers: phase 2, task analysis of en-route controllers", European Air Traffic Management Programme Rep No HUM. ET1. ST01. 1000-REP-04: EUROCONTROL. Brussels.

Kouba, P., Šmotek, M. and Tichý, T. (2019), "Identification and monitoring of traffic operators' fatigue level", *Smart Cities Symposium*. Prague.

Liu, Y., Trapsilawati, F., Hou, X., Sourina, O., Chen, C.-H., Kiranraj, P., ... Ang, W.T. (2017), *EEG-Based Mental Workload Recognition in Human Factors Evaluation of Future Air Traffic Control Systems*, Trandisciplinary Engineering: A Paradigm Shift.

López de Frutos, P.M., Rodríguez Rodríguez, R., Zheng Zhang, D., Zheng, S., Cañas, J.J. and Muñoz de Escalona, E. (2019), "COMETA: an air traffic controller's mental workload model for calculating and predicting demand and capacity balancing", Longo L., Leva M. (Eds) *Human Mental Workload: Models and Applications. H-WORKLOAD 2019. Communications in Computer and Information Science*, 1107 Springer, New York, NY, pp. 85-104.

Martin, C., Cegarra, J. and Averty, P. (2011), "Analysis of mental workload during En-route air traffic control task execution based on Eye-Tracking technique", *Engin. Psychol. and Cog. Ergonomics, HCII 2011. LNAI*, Vol. 6781, pp. 592-597.

Miyake, S., Yamada, S., Shoji, T., Kuge, T. and Yamamura, N. (2009), "Physiological responses to workload change: a test/retest examination", *Applied Ergonomics*, Vol. 40 No. 6, pp. 987-996.

Mogtit, A., Aribi, N., Lebbah, Y. and Lagha, M. (2020), "Equitable optimized airspace sectorisation based on constraint programming and OWA aggregation", *Aircraft Engineering and Aerospace Technology*, Vol. 92 No. 8.

Netjasov, F. Mirkovic, B. Simic, T.K. Babic, O. Gómez Comendador, V.F. Arnaldo Valdés, R.M. and Domínguez Pérez, D. (2019), "Identification of safety critical hazards to support future air traffic controller training. 23rd ATRS world conference", Amsterdam, 2-5 July.

SESAR_HP Repository (2012), "Instantaneous self assessment of workload (ISA)", available at: https://ext.eurocontrol.int/ehp/?q=node/1585 (accessed 15 Agosto 2021).

SESAR JU (2017), "D3.2 competence and training requirements", AUTOPACE Project. H2020-SESAR-2015-1.

Socha, V., Hanáková, L., Valenta, V., L., Sochaa, U., Ábela, R., Kušmírek, S., ... Tecl, J. (2020), "Workload assessment of air traffic controllers", *Transportation Research Procedia*, Vol. 51, pp. 243-251.

Trapsilawati, F. Liu, Y. Wee, H.J. Subramaniam, H. Sourina, O. Pushparaj, K. ... Lye, S.W. (2017), "Perceived and physiological mental workload and emotion assessment in En-Route ATC environment: a case study", Transdiciplinary Engineering: A Paradigm Shift.

Triyanti, V., Azis, H.A., Prasetyawan, Y., Iridiastadi, H. and Yassierli, A. (2020), "Workload and fatigue assessment on air traffic controller", *IOP Conf. Ser.: Mater. Sci. Eng*, Vol. 847.

Triyanti, V., Azis, H.A., Prasetyawan, Y., Iridiastadi, H. and Yassierli, A. (2021), "Individual factors related to mental workload in air traffic controller", in Gutierrez, A.M.J., Goonetilleke, R.S. and Robielos, R.A.C. (Eds), *Convergence of Ergonomics and Design. ACEDSEANES 2020. Advances in Intelligent Systems and Computing, AISC 1298*, 271-278.

Wee, H.J., Lye, S., J. and Pinheiro, e-P. (2020), "Monitoring performance measures for radar air traffic controller using eye tracking techniques", *AISC*, Vol. 964, pp. 727-738.

Corresponding author

María Zamarreño Suárez can be contacted at: maria.zamarreno.suarez@alumnos.upm.es