Kling Fanni, Papp Gábor, Rohács Dániel

INTEGRATING HUMAN FACTORS IN REAL-TIME SIMULATIONS

Real-Time Simulation (RTS) is a technique used in Air Traffic Management (ATM) to validate new operational procedures, airspace organisation or support tools. Although the aim is to involve Air Traffic Control Officers (ATCOs) in the simulation, many RTS projects still lack the human factors perspective. This paper presents the benefits of using RTS where the Human Factors are taken into account. First, basic concepts in human cognition are introduced that needs to be covered during a validation. Second, the validation steps are outlined with special attention to their own challenges. Finally, projects that are of key importance in SESAR 2020 (Single European Sky ATM Research 2020) and in Human factors will be elaborated.

Keywords: *Real-Time Simulation, Human factors, Air Traffic Control, Validation, Experimental Psychology*

INTRODUCTION

Successful development of a system, procedure or new airspace organisation depends on having the right methods to ensure that the development meets the given maturity level. The European Operational Concept Validation Methodology is a framework to provide structure and transparency in the validation of ATM operational concepts [1]. E-OCVM meets the criteria outlined above as it considers validation an iterative approach where the selection of validation techniques is based on the concept's maturity. If the scope has been set and the operational concept defined (referred to as V1 stage) significant effort can be placed on experimentation.

The next phase is applied to test the feasibility of the development. At this maturity level (referred to as V2) Fast-Time and Real-Time Simulation are regarded as appropriate research techniques. Fast-Time Simulation (FTS) provides a cost effective way to examine the impact of new aspects on operations and help stakeholders to make informed decisions. The simulation is executed at accelerated speed, which provides opportunity to test various scenarios. In contrast, Real-Time Simulation (RTS) is used to validate new operational concepts with the contribution of air traffic controllers who are proficient in the measured airspace. In order to ensure realistic and high-fidelity environment the human-machine interface can be reproduced to match to the ATCO's own system. Importantly, it is useful to differentiate between small-scale and large-scale simulations. At V2, it is highly feasible that a small-scale simulation is enough to test the concept and apply large-scale RTS only prior implementation, during which mature and well thought-through concepts will be presented to the ATCOs.

The design and execution of a Real-Time Simulation requires experts from a variety of fields. Human Factors experts are involved quite frequently, however, there are still human-in-the-loop simulations where the human factors assessment is either partly present or completely neglected. This paper aims to clarify the role of human factors in ATM projects, more specifically in RTS. Some examples of the human factors methodology are introduced that is currently used at HungaroControl (Hungarian Air Navigation Services).

HUMAN FACTORS

Human factors is a broad term referring to all the aspects that influence a human's capability to accomplish tasks and meet job requirements. Consequently, the experts addressing human factors may represent various fields such as ergonomics, recruitment, safety, experimental psychology or neuroscience.

However, RTS requires a more specialized human factors assessment methodology. The key focus here is very similar, i.e. to evaluate how the new development will impact the operator by investigating each relevant human factors aspect. However, there are certain parts of the validation process in terms of experimental design and metrics selection; data analysis and presentation that need considerable experience in cognitive science or experimental psychology. The important aspects that a human factors expert has to take into account during a RTS are elaborated in the next section, with special attention to human factors constructs and experimental design.

Human Cognition

One of the main goals of Human Factor studies is to ensure good interactions between human capabilities and work environment [2]. Human capabilities are however a broad term that requires further clarification. It encompasses several cognitive function such perception, attention, hearing, decision-making, memory, stress, emotion, executive functions etc. However, human factors and cognitive engineering require viable constructs to gain better understanding and provide prediction on human performance in complex systems [12].

Therefore, in most of the ATM studies the three common human cognition and performance constructs referred to are workload, situational awareness and trust in automation. It has been demonstrated that workload is significantly impacted by ATC (Air Traffic Control) Complexity factors, such as aircraft count, horizontal proximity or sensitivity of conflict [7]. High workload in turn can lead to deficient decision making and errors [3] [14]. In spite of the disagreement about its nature and definition "workload remains an important and measurable entity" [4]. Several methods exist to evaluate and predict workload such as i) performance-based measurements, ii) subjective measurements, iii) physiological measurements [5] [6]. Performance-based measurements are used on the presumption that the high workload will be visible in higher frequency occupation time [7], slower reaction times, errors. Physiological evaluations are based on the recording of operator's neurophysiological signals measured with EEG, fNIRS, GSR, eye-tracking [8]. Subjective measurements are based on the assumption that the operator is able to report on the experienced workload [5]. Each approach has its advantages and disadvantages. Due to the relatively new technology, neurophysiology-based metrics have not yet been implemented at many simulators. The remaining two are however broadly used in aviation psychology.

With regards to the subjective measurement, many argue that workload is not a clear concept [2], and means different concepts to individuals. Some may define it as the time pressure under which the task is performed, others as the level of effort needed to execute the task, success in performance, the mental demand and physical movements it requires or the combination of these elements [4]. In contrast to the multidimensional property, the spare capacity or the lack thereof can be an additional consideration. Was the operator able to perform the task? Was there a possibility to execute additional tasks at the same time? If this is the main evaluation criteria, then multidimensional rating scales may be unnecessary, and a more concrete, decision-tree based approach can be applied which minimizes the workload reporting differences across participants [9] [24]. The questionnaires that are most widely used in aviation-related simulations are presented in Figure 1.

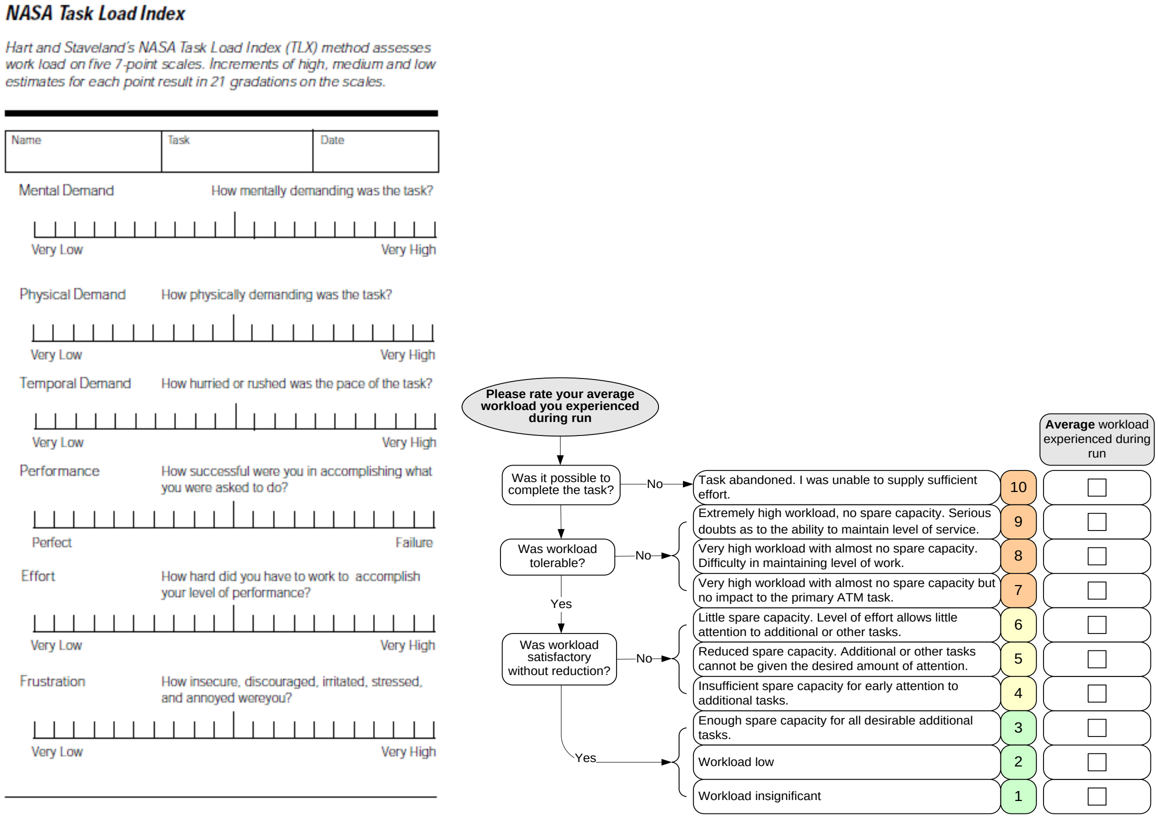

Figure 1. Two of the most broadly used questionnaires for workload estimation in aviation-related simulations (NASA-TLX on the left and the Bedford Workload Scale on the right)

NASA-TLX (see in Figure 1) entails six dimensions that are considered important for the total workload [4]. The Bedford Workload Scale (see in Figure 1) is a uni-dimensional rating scale designed to identify operator's spare mental capacity while completing a task. The single dimension is assessed using a hierarchical decision tree that guides the operator through a ten-point rating scale, each point of which is accompanied by a descriptor of the associated level of workload. It has been originally developed for the assessment of pilots [24]. At HungaroControl, both are used depending on the tested concept, however, they have been digitalized and modified in order to ensure that each participant understands the difference between the statements in Bedford at face value. The associated colours and numbers have been eliminated to avoid that participants only mark the statements according to the number, and do not take into account the statement.

Beside workload, situation awareness (SA) is also an essential aspect of human factors. In brief, SA refers to the ATCO's mental picture of the constantly changing traffic situation. Defined by [10], SA is constituted by three elements: i) perception of the elements in the environment within a volume of time and space, ii) the comprehension of their meaning, iii) projection of their status in the near future. Consequently, in the ATM domain SA involves the continuous awareness of the location of each aircraft with its own parameters and their predicted future location relative to each other. Several aspects can contribute to the degradation of SA, such as stress, fatigue, memory failure, unavailability of data [10] [11]. The SASHA-Q (Situation Awareness for SHAPE Questionnaire) developed by [25] is frequently used at HungaroControl. The degradation of situation awareness is addressed by several operational situations that may have happened in the last run, like being surprised by something, chance of forgetting something important or starting to focus on a single problem or specific area. However, the planning, prioritization of tasks are also part of the questionnaire.

The importance of measuring trust in automation is continuously increasing as many new concepts in ATM utilize automated system support [20]. Research conducted by [13] showed that the operator's trust in the automation directly influenced his use of the system in a simulated task. Low levels of trust may lead to disuse [12]. In turn, high level of trust can lead to overreliance and failure to monitor the situation (referred to as a complacency issue).

These constructs are often tested through empirical studies such as simulations. The next section will present how the knowledge about human factors and data science can be integrated in the design of a real-time simulation and then in the evaluation of the data.

Real-Time Simulation

Real-Time simulation is an essential technique to test new and mature concepts and assess their impact on human operators. Real-Time simulation provides a relatively controlled and repeatable environment [1] that is fundamental for validation. Some may argue that the disadvantages of RTS are the learning-effect and its expense in contrast to what it offers. Although it has to be accepted that RTS is not an experiment and a balanced experimental design is difficult to achieve [15], the integration of human factors may account for these claims and help to create a valuable simulation.

The development of a successful Real-Time Simulation certainly requires excellent teamwork. Adequate project and time management; correct hence realistic simulation environment; reliable simulator platform and voice-communication system are all inherent part of the simulation. If any of the listed elements are deficient or neglected, the simulation may not reach its intended goal and result, let alone customer and ATCO satisfaction. Human factors experience brings an additional value to the simulation as well. First, a human factors expert has to ensure a good interaction between operator and the simulator environment. Second, accurate experimental design will ensure that the impact of the tested development is measurable hence demonstrable.

Preparation

The first phase where human factors or experimental psychology comes into play is the definition of simulation hypotheses. Due to its characteristics (i.e. relatively controlled and repeatable) RTS is designed to compare various alternatives of the new concept [15] and/or based on the counterfactual theory of causation, compare the impact of new and current system on operators with each other [17]. It is of great importance to agree on hypotheses that are testable with metrics and provide meaning for the management.

By defining the hypotheses the human factors experts and the customer have to have a preliminary plan on the number of variables and scenarios that will be tested. Variables may be sector configuration, traffic samples, procedures etc. Each variable can have more levels: two types of sectorization (e.g. current and new), two types of traffic sample (e.g. low and high traffic level), two types of procedure (with automation tool or without). It is essential for the experimental design to understand that if a variable has more levels (even two, let alone three) it can inevitably lead to learning effect. The reason is the following. Real-Time Simulation often uses within subject experimental design. It means that each participant is exposed to all conditions of the experiment and therefore serves as his/her own control. The provided benefit is considerable because the influence of the manipulated variable is separable and results in high statistical power. To put it briefly, it can be evaluated how individual behavior changes when the circumstances of the simulation change [16].

However, if the operator is exposed to the same environment and only slight modification occurs (seeing the same traffic sample but with three different procedures), the learning effect will influence his/her performance. Consequently, close attention must be paid on the number of levels a variable may have. The main issue in large scale RTS compared to lab experiment is that there is less chance to randomize the conditions. In lab experiments one participant will start with the first condition, while another participant starts with the second, etc. In large scale RTS, all ATCOs are participating at the same time, thus randomization is not viable. A plausible solution for randomization would be to have one more traffic sample with similar occupancy. However, adding more variables will influence the length of the simulation, which is again a critical factor in RTS from resource and cost perspective. Therefore, it is recommended to select the most intriguing variables that have only two levels and have at least two traffic samples minimizing the learning effect.

Figure 2. Real-Time Simulator platform at HungaroControl. The platform encompasses 34 controller and 27 pilot working positions with advanced ATM tools and applications, System Coordination and Data-Link environment.

Rostering is also a key factor that ensures the mitigation of the learning effect and the gathering of data. Rostering refers to the seating principle where the ATCOs will be rotated to different controller working positions (sector's position in sectorized environment), hence can provide feedback on various positions. In order to reach statistical significance and avoid biased results due to individual differences, some scenario repetitions with different rostering are considered necessary (three or four at least). Too many repetitions are however superfluous, as it may happen that controllers are rostered to the same controller working position twice in the same scenario.

Still in the preparation phase, the human factors expert selects and/or develops metrics that will provide evidence to confirm or reject the hypotheses. The metrics are mostly performance-related and questionnaire-based for subjective evaluation. The key concepts to be addressed are workload and situation awareness. There are certain questionnaires that are widely accepted (e.g. NASA-TLX, Bedford Workload Scale, SASHA-Q), but every simulated concept requires a more specialized questionnaire to investigate the impact. The addressed fields are automation, adequacy of operating methods, hotspots, teamwork, etc. (for a thorough assessment of these fields see [23]). The joint development of such questionnaire with subject matter experts is strongly recommended.

At last, human factors experts have to ensure that the participating ATCOs received the necessary training for the simulation. Not only the proper human-machine interaction has to be ensured, but that the ATCOs understand the goal of the simulation and the concepts to be tested in the RTS. This acceptance test usually happens a few weeks prior the RTS.

Execution and Analysis

During the RTS the human factors expert ensures that the simulation runs according to the Validation plan. In order to obtain reliable data the experimental design and variables to be tested should not be modified during the execution. However, it has to be realized that if a concept does not seem viable from the beginning, balanced experimental design might have to be abandoned for the sake of the success [15]. In addition, as Wise and colleagues [15] point out, ATCOs are not merely experimental subjects but subject matter experts whose opinion counts a great deal during the simulation.

Observing the RTS provides vital input for the debriefing sessions and the report. The performance indicators may serve as the basis of the debriefing sessions and can support the facilitation of the discussions.

Once the simulation has been completed and all relevant data (objective, subjective) have been collected, the human factors analyst will first verify the quality of the data, then analyze it with the aid of statistical tools. After the statistical calculation the human factors expert has to interpret the results, connect results from different sources to obtain a holistic picture. Coordination with other project members is essential at this phase. For instance, discussions with the OPS (operational) team may shed some lights on the concerns not known for the human factors expert.

At last, the human factors expert can provide conclusions and recommendations for the management based on the outcome of the simulation.

Relevant projects

Human Factors Assessment is also an integral part of the project in Single European Sky ATM Research (SESAR 2020), which is the technological pillar of the European Union’s Single European Sky (SES) initiative [18]. SESAR 2020’s Industrial Research and Validation Programme focuses on four key features (High Performing Airport Operations, Optimised ATM Network Services, Advanced Air Traffic Services, Enabling Aviation Infrastructure). Several projects have been dedicated to assess these areas. In this section two of those will be highlighted, namely High Performing Airport Operations - Remote Tower for Multiple Airports and Advanced Air Traffic Services – Separation Management En-Route and TMA (terminal manoeuvring area) [22] because these innovative technologies have special importance from human factors perspective.

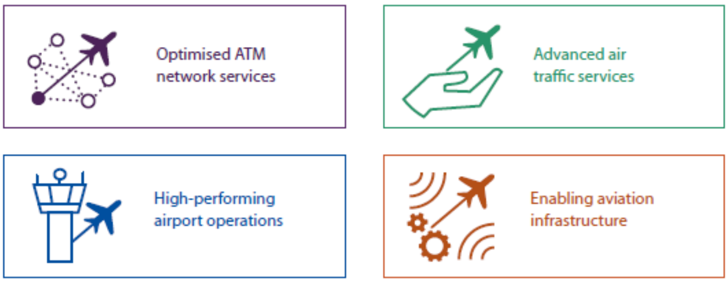

Figure 3. The four areas in ATM (Key Features) defined in the European ATM Master Plan

Therefore, the above mentioned projects use Human Performance Assessment methodology to address the human factors related aspects (human performance bears a very similar meaning to human factors in this case) that may be impacted by the new concepts. Importantly, the feasibility of both solutions can be addressed with Real-Time Simulation.

Multiple Remote Tower operation is a considerably new concept that envisages one operator providing ATS services to more than one aerodrome in parallel. This topic is of key importance for human factors experts for several reasons. First, the number of tasks an ATCO will have to perform and the working methods will change. The Human-Machine Interface (HMI) will be configured in a way that two or more aerodromes can be monitored. Hence the number of information displayed will increase. Human factors experts have to address how this change will impact workload and situation awareness, taking into consideration the different procedures and characteristics of the aerodromes (geographical area, weather conditions etc.) and the usability of the human-machine interface.

Flight Centric ATC (FCA) is one of the key solutions within Advanced Air Traffic Services – Separation Management En-Route and TMA. The aim of the concept also known as Sectorless ATM is to dissolve sector boundaries and to have one controller in charge for a single flight to guide it through a large airspace [19]. As a basic principle of Flight Centric ATC, a controller is no longer in charge of managing the entire traffic within a given sector. Instead, he/she is now responsible for a certain number of aircraft which he/she controls from the entry into the given airspace to the exit, whereas other controllers are responsible for a certain number of different aircraft within the same airspace [20]. ATC has still the responsibility for separation provision and the controller has to ensure a conflict-free flight. According to the plan, advanced automation will support the executive controller in conflict detection and resolution and could perform the traffic assignment based on specified strategies. Therefore, significant changes to controllers' tasks and teamwork are expected [21]. For instance, the executive controller will be able to take on more tasks such as monitoring and within-FCA coordination, while the planner controller's role will be rather coordination with adjacent sectors. However, the radical change will influence several fields such as communication and HMI, which in turn may have important implications on human performance. For instance, considering the possibility to allocate different frequencies to the executives [20], not only ATCO's but also pilot's situation awareness may be impacted by not perceiving everything that happens around them. The adequate presentation of concerned flights is also a factor that requires a thorough assessment as it can directly impact ATCO's situation awareness and workload. Most importantly however, the reliability of the advanced automation tools has to be ensured, since without trust in their functionality the whole concept is put at risk.

At their current maturity levels both concepts can be assessed with Real-Time Simulations. As described in the first section, Real-Time Simulation provides valuable insight into the magnitude and quality of the change, be a small or large scale validation. Useful experiences can be gained with the contribution of experts from various fields (ATC, Technical, OPS, Safety, Human Factors and Project Management).

CONCLUSION

In this article the importance of human factors assessment in Real-Time Simulations has been emphasized. First, key concepts in the field of human factors and cognitive engineering have been introduced, with special attention to workload, situation awareness and trust in automation. These constructs are the main interest in Real-Time Simulations, but human factors knowledge also ensures the application of the right experimental design for the tested concept in order to achieve reliable results.

In case of SESAR 2020 where innovative R&D solutions are tested, human factors expertise will help to identify the proposed concept's benefits and issues and support the formalization of mitigation strategies. These aspects can be addressed with Real-Time Simulation to demonstrate the concept and gather considerable feedback from the participants and subject matter experts. This way, ATM experts and decision makers will be able to foresee potential changes in human performance and to prevent safety issues related to workload and situation awareness.

REFERENCES

[1] EUROCONTROL. European Operational Concept Validation Methodology E-OCVM, 3rd Edition, February 2010

[2] WICKENS, Christopher D., et al. Engineering psychology & human performance. Psychology Press, 2015.

[3] REASON, James. "Human error: models and management." BMJ: British Medical Journal 320.7237 (2000): 768.

[4] HART, Sandra G., and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research." Advances in psychology 52 (1988): 139-183

[5] RUBIO, Susana, et al. "Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods." Applied Psychology 53.1 (2004): 61-86.

[6] MESHKATI, Najmedin, et al. "Techniques in mental workload assessment." In J. Wilson & E. Corlett (Eds.), Evaluation of human work. A practical ergonomics methodology (pp. 605–627). London: Taylor & Francis (1995).

[7] DJOKIC, Jelena, Bernd LORENZ, and Hartmut FRICKE. "ATC Complexity as workload and safety driver." convergence 7.11 (2008): 12.

[8] ARICÒ, Pietro, et al. "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 10 (2016).

[9] ELLIS, Kyle Kent Edward. Eye tracking metrics for workload estimation in flight deck operations. The University of Iowa, 2009.

[10] ENDSLEY, Mica R., and Mark D. RODGERS. Distribution of attention, situation awareness, and workload in a passive air traffic control task: Implications for operational errors and automation. No. DOT/FAA/AM-97/13. Federal aviation administration Washington DC office of aviation medicine, 1997.

[11] SNEDDON, Anne, Kathryn MEARNS, and Rhona FLIN. "Stress, fatigue, situation awareness and safety in offshore drilling crews." Safety Science 56 (2013): 80-88.

[12] PARASURAMAN, Raja, Thomas B. SHERIDAN, and Christopher D. WICKENS. "Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs." Journal of Cognitive Engineering and Decision Making2.2 (2008): 140-160.

[13] LEE, John, and Neville MORAY. "Trust, control strategies and allocation of function in human-machine systems." Ergonomics35.10 (1992): 1243-1270.

[14] KEINAN, Giora. "Decision making under stress: Scanning of alternatives under controllable and uncontrollable threats." Journal of personality and social psychology 52.3 (1987): 639.

[15] WISE, John A., V. David HOPKIN, and Paul STAGER, eds. Verification and validation of complex systems: Human factors issues. Vol. 110. Springer Science & Business Media, 2013.

[16] CHARNESS, Gary, Uri GNEEZY, and Michael A. KUHN. "Experimental methods: Between-subject and within-subject design." Journal of Economic Behavior & Organization 81.1 (2012): 1-8.

[17] BYIERS, Breanne J., Joe REICHLE, and Frank J. SYMONS. "Single-subject experimental design for evidence-based practice." American Journal of Speech-Language Pathology 21.4 (2012): 397-414.

[18] SESAR Joint Undertaking. "SESAR 2020: developing the next generation of European Air Traffic Management", (2014) downloaded from: http://ec.europa.eu/research/press/jti/factsheet_sesar-web.pdf

[19] BIRKMEIER, Bettina, Bernd KORN, and Dirk KÜGLER. "Sectorless ATM and advanced SESAR concepts: Complement not contradiction." Digital Avionics Systems Conference (DASC), 2010 IEEE/AIAA 29th. IEEE, 2010.

[20] BIRKMEIER, Bettina. Feasibility analysis of sectorless and partially automated air traffic management. Diss. Technische Universität Carolo-Wilhelmina zu Braunschweig, 2015.

[21] BIRKMEIER, Bettina, et al. "Change of controller tasks in a sectorless ATM concept-first results." Integrated Communications, Navigation and Surveillance Conference (ICNS), 2012. IEEE, 2012.

[22] SESAR Joint Undertaking. "SESAR 2020 Multi-annual Work Programme", (2015) downloaded from http://ec.europa.eu/research/participants/data/ref/h2020/other/wp/jtis/h2020-wp-multi-annual-sesar-ju_en.pdf

[23] EUROCONTROL. "HF Case Workarea descriptors". (2008) downloaded from http://www.eurocontrol.int/articles/human-factors-case

[24] ROSCOE, Alan H., and Georges A. ELLIS. A subjective rating scale for assessing pilot workload in flight: A decade of practical use. No. Rae-tr-90019. Royal aerospace establishment Farnborough (United Kingdom), 1990.

[25] DEHN, Doris M. "Assessing the impact of automation on the air traffic controller: the SHAPE questionnaires." Air traffic control quarterly16.4 (2008): 127.

AZ EMBERI TÉNYEZŐK ELEMZÉSÉNEK INTEGRÁLÁSA VALÓS IDEJŰ SZIMULÁCIÓKBA

A valós idejű szimuláció célja új eljárások, komplex légtérstuktúrák vagy irányítói támogatóeszközök validálása az ATM iparágban. Bár a légiforgalmi irányítók bevonása jelentős előnyt jelent validálás során, sok szimuláció még mindig elhanyagolja az emberi tényezők vizsgálatát. A cikk ezért bemutatja az emberi teljesítményvizsgálat valós idejű szimulációkba való integrálásának előnyeit. Röviden összefoglalja az emberi kogníció alapjait, amelyek relevánsak lehetnek szimuláció során, majd szisztematikusan kifejti a validációs lépéseket azok sajátosságaival; előnyeivel és kihívásaival. Végül a SESAR 2020 és humánfaktor kutatások szempontjából fontos projektek kerülnek részletezésre.

Kulcsszavak: *valós idejű szimuláció, emberi tényezők, légiforgalmi irányítás, validáció, kísérleti pszichológia*

Kling Fanni Szimulációs adatelemző HungaroControl Magyar Légiforgalmi Szolgálat Zrt. fanni.kling@hungarocontrol.hu orcid.org/0000-0001-7379-9069	Kling Fanni Data Scientist HungaroControl Hungarian Air Navigation Services fanni.kling@hungarocontrol.hu orcid.org/0000-0001-7379-9069
Papp Gábor Szimuláció és Validáció csoportvezető HungaroControl Magyar Légiforgalmi Szolgálat Zrt. gabor.papp@hungarocontrol.hu orcid.org/0000-0002-5676-9711	Papp Gábor Head of Simulation and Validation Unit HungaroControl Hungarian Air Navigation Services gabor.papp@hungarocontrol.hu orcid.org/0000-0002-5676-9711
Dr. Rohács Dániel, PhD Szakmai fejlesztési osztályvezető HungaroControl Magyar Légiforgalmi Szolgálat Zrt. daniel.rohacs@hungarocontrol.hu orcid.org/0000-0002-4629-4417	Rohács Dániel, PhD Head of Research, Development & Simulation Dept. HungaroControl Hungarian Air Navigation Services daniel.rohacs@hungarocontrol.hu orcid.org/0000-0002-4629-4417

http://www.repulestudomany.hu/folyoirat/2017_3/2017-3-18-0433_Kling_F-Papp_G-Rohacs_D.pdf