

The influence of automation support on Air Traffic Controller behaviour with a Speed And Route Advisory function

Koen van de Merwe
Training, Simulation and Operator Performance Department
National Aerospace Laboratory NLR
Amsterdam, The Netherlands
merwe@nlr.nl

Esther Oprins^a, Akos van der Plaat^b, Fredrik Eriksson^a
^aResearch & Development, ^bOperations
Air Traffic Control The Netherlands
Schiphol, The Netherlands
e.oprins@lvnl.nl, a.h.vanderplaat@lvnl.nl, f.eriksson@lvnl.nl

The aim of this paper is to describe the influence of automation support on Air Traffic Controller performance, workload and Situation Awareness (SA). Controllers handle traffic through means of tactical control involving heading, speed and altitude instructions. Future Air Traffic Management (ATM) concepts, such as Single European Sky ATM Research (SESAR) and NextGen, promote the use of 4D trajectories, thereby introducing a time-dimension to current control tactics (SESAR, 2007). At present, Amsterdam Area Control (ACC) delivers traffic over to Schiphol Approach control (APP) via three Initial Approach Fixes (IAFs) with a margin of plus or minus 120 seconds between the planned time and the actual time. This variability can make it difficult for APP controllers to merge traffic streams and build a landing sequence, especially during peak periods. In the future, a change in delivery accuracy to less than plus or minus 30 seconds is foreseen thereby aiming to increase the punctuality of flights.

An undesirable increase in workload is expected for Amsterdam ACC controllers to achieve such accuracy without any system support. Therefore, a Speed And Route Advisor (SARA) was developed. This tool provides ACC controllers with a speed and route advise with which the aircraft can meet its planned time over the IAF with the desired accuracy. Whilst such a tool is necessary to keep the workload of ACC controllers within limits, a potential drawback is the possible decrement in controller's SA (e.g. Endsley and Kiris, 1995; Endsley, 1997), which has been shown to be one of the principal competencies of controllers (Oprins et al., 2006).

An experiment was performed to validate SARA in its ability to support the controller in accurately delivering traffic to APP over the IAF and to assess its impact on their workload and SA. Eight Amsterdam ACC controllers and four Maastricht Upper Area Control Center (MUAC) controllers participated in seven scenarios in an Air Traffic Control (ATC) simulator for four days. Seven scenarios were run with varying SARA configurations (e.g. only speed advisories, or speed and route advisories) and baseline configurations.

The results showed a significant improvement in delivery accuracy with the support of SARA. Self-report measures of workload varied significantly between the different scenarios, and appear to be related to familiarity with SARA. Objective measures of workload, as measured through the amount and duration of radiotelephony (R/T) calls and manual inputs, decreased. SA significantly decreased with the use of SARA, although controllers still rated it above average. This was consistent for all SARA conditions. Controllers indicated that they were checking SARA advisories in an effort to understand SARA's 'plan' which often deviated from their own strategy. However, unfamiliarity with SARA may have been an influencing factor.

The results clearly showed the benefit of SARA in supporting controllers in accurately delivering traffic over the IAF without a significant increase in workload. These findings implicate the potential for SARA as a means towards time-based operations around Schiphol Airport. The impact of the use of SARA on SA, however, should be carefully considered in future design and implementation efforts.

INTRODUCTION

ATC performance

The primary aim of Air Traffic Control is to expedite and maintain a safe and orderly flow of air traffic. Similar to other process control tasks in transportation (aviation, shipping, railways) or process industry (e.g., chemical and nuclear plants), the ATC task is considered highly complex and dynamic. The continuous flow of moving aircraft cannot be stopped; timely actions are needed to create safe and most efficient traffic flows before possible collisions become critical. Complex cognitive processes are required to handle the great amount of dynamically changing information in a three-dimensional environment (Garland, Stein & Muller, 1999). Therefore, ATC is also called a complex cognitive or high-performance skill (Schneider, 1990).

Air Traffic Control the Netherlands (Luchtverkeersleiding Nederland; LVNL) has designed the so-called ATC Performance Model (Oprins, Burggraaff and Van Weerdenburg, 2006; Oprins, 2008), which visualizes the complex cognitive processes of air traffic controllers. This model has been applied as a general framework for selection and training design. Since a few years, it is also used at LVNL to assess the impact of new developments in ATM system design on the human role of controllers in a paper study, Human Factor Indication (HFI) and in real-time simulations. The model is the result of a competence analysis performed at LVNL based on literature research and workshops with controllers (see Figure 1).

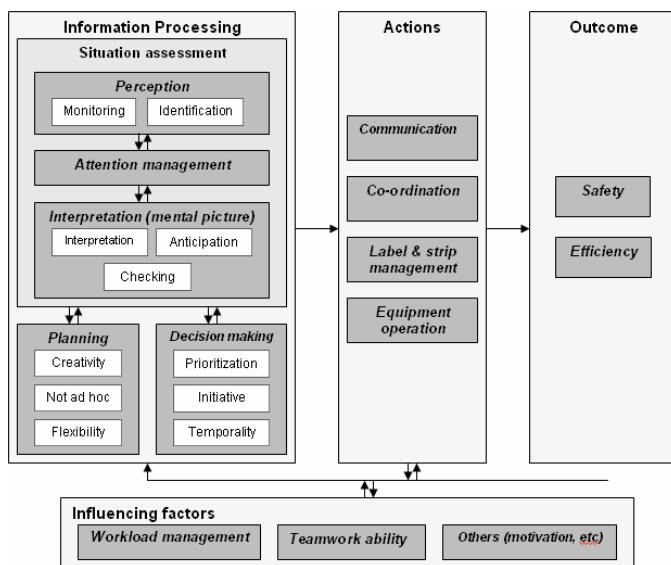


Figure 1. The ATC Performance Model.

The model shows the importance of cognitive processes in which *situation assessment* plays a central role. Information processing guides the actions and this results in safe and efficient traffic handling. One important influencing factor is *workload management*.

Situation awareness. A common assumption is that operators in dynamic and complex tasks such as ATC create a mental representation of the changing environment, which makes it possible to keep the relevant but transient information

in working memory (Garland, Stein & Muller, 1999). Pattern recognition plays a central role; the controller groups aircraft in a certain way to memorize their positions. These patterns help them to create order in seemingly chaotic situations by streaming traffic flows. Much research has been done on how controllers develop the three-dimensional 'mental picture' of the traffic situation. This is usually referred to as *situation assessment*, defined as follows: 'The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future' (Endsley, 1999). Situation Awareness (SA) is considered the product of the process of situation assessment that takes place at three levels: perception (SA1), interpretation (SA2) and anticipation (SA3). Attention management strategies are crucial to keep this ever-changing 'picture' up-to-date (Shebilske, Goetl and Garland, 2000).

Workload management. Controllers regularly switch between low and high mental workload, depending on the traffic situations (e.g., number of aircraft, complexity). This is called workload management. But mental workload has also a strong subjective component (Averty et al., 2004).

Controllers continuously apply strategies, which are individually different, to keep safety (conflict detection), efficiency (traffic delay) and their own mental workload ('personal efficiency') in optimal balance (Oprins & Burggraaff, 2006). SA is needed to identify and enact the most safe and efficient solution to solve specific (conflict) situations. In addition, controllers keep their own mental workload under control by adjusting their strategies towards less effortful if needed. If possible they revert to routine actions, standard procedures and 'simple' solutions that need less attention and that gain time, for instance, by a lower load of radiotelephony. Depending on the evolving situation (routine – non-routine), they switch between low and high workload.

Reduction of work complexity in ATC.

Internally, LVNL is coping with a shortage of controllers. This is not uncommon in many busy and complex airports. Due to the complex cognitive nature of the ATC task only a small number of people are able to acquire the required competences within a reasonable period of training (Schneider, 1990). LVNL is attempting to solve this problem by improving selection and training, and by designing new ATM systems that make the work less complex. Research on training performance of all trainees between 2003 and 2006, using the ATC Performance model, has shown that ineffective situation assessment and workload management are the two most important reasons for failing (Oprins, 2008). This suggests that these competences are more difficult to learn than others and require extra attention in designing less complex ATM systems.

Previous research has shown that increasing automation, as expected in future ATM systems, could make work less complex. A possible risk of more automation is often referred to as the 'out-of-the-loop' performance problem (Endsley & Kiris, 1995). In case of automation failures system operators may have diminished ability to perform tasks manually, due to lesser awareness of the state and processes of the system, i.e. SA. There are three reasons why this happens. First, monitoring tasks may lead to vigilance problems. Alertness decreases as controllers usually have much trust in the

equipment. Second, passive information processing seems to be inferior to active information processing in detecting the need for manual intervention and reorientation to the state of the system. Third, without any feedback, people are really out of the loop and they cannot assess the effectiveness of their requests and actions.

More automation can also increase SA (Endsley & Kiris, 1995). In a more monitoring role, controllers are better able to distribute their attention, especially when the system provides superior, integrated information to the controllers. In addition, SA may be improved by a strong reduction of workload. A partial automation strategy should keep the negative and positive effects in balance. It is usually argued that routine tasks should be fully automated to reduce workload, while automation should support SA by offering better and more integrated information to the controllers.

These issues have been addressed in research on ATM system design (Endsley, 1997). ATM is also moving towards more monitoring (cf. 'supervisory control'; SESAR, 2007). Human-centred design in ATM suggests that routine tasks such as radiotelephony should be automated (cf. datalink), that the presentation of information to controllers should be improved for supporting SA, and that decision support tools are needed to choose the right solutions. However, ATM system designers are still searching for the right balance in automation, also in relation to fallback systems (machine or human).

At Schiphol Airport, work complexity for controllers is particularly high because of the large numbers of air traffic and bunching associated with peaks. The ATM strategy of LVNL focuses on accommodating a growth of air traffic as expected in the future while making work less complex. For this purpose, support tools for controllers will be introduced at LVNL. The main question is how these tools can be designed in such a way that situation awareness will be improved and workload will be reduced for the benefit of the controllers.

Current situation at Schiphol Airport

Amsterdam Airport Schiphol (EHAM) is the busiest of the Netherlands, using a maximum of 3 out of 6 runways to balance capacity and demand. It is one of the two main hubs of Air France/KLM, mainly used by Royal Dutch Airlines (KLM). KLM runs a hub-and-spoke operation, therewith providing short connection times between flights, which results in four arrival peaks and four departure peaks per day and off-peak periods in between.

Departure peaks will normally be managed by the use of two take-off runways and one landing runway, resulting in a capacity of up to 80 departures and 38 arrivals per hour. Two landing runways are used during arrival peaks in addition to one take-off runway, and the hourly capacity for this combination is 65 arrivals and 40 departures. Only one take-off and one landing runway are used in off-peak periods, with an hourly capacity of 38 arrivals and 40 departures. The total number of movements for Schiphol is circa 1100 (February 2009) per day.

Air traffic control service in the Netherlands is provided by MUAC, the Ministry of Defence, and the LVNL. In terms of the arrival streams for the Schiphol Terminal Manoeuvring Area (TMA), traffic is fed by Amsterdam ACC via three entry points called Initial Approach Fixes (IAF). These three individual traffic flows are subsequently merged to either in

two or one stream for the landing runway(s) in use. This merging after the IAF is done by Schiphol Approach and currently is done by the use of individual manoeuvring instructions (i.e. radar vectoring). System support is provided to avoid overloads in the TMA. This system support is an arrival management system called LVNL Inbound Planner (IBP), and assigns arriving aircraft a landingslot 14 minutes prior to passing the IAF. This landingslot is the basis on which an Expected Approach Time (EAT) for the IAF is calculated. ACC is subsequently required to deliver the arriving aircraft for Schiphol to APP via the IAF within plus or minus 120 seconds from the EAT.

As mentioned in the previous section, the bunching associated with the arrival peaks is a frequent and undesirable disturbance of the desired stability of the Schiphol ATM system. These bunches often lead to a temporary increase in workload for the ACC controllers, and has a negative impact on the complexity and therefore also on the efficiency of the operation. The APP controllers often have difficulties with the way the traffic is presented to them at the IAFs and the subsequent merging of the streams. Amongst other reasons, some of these difficulties can be ascribed to the 120 seconds margin that can be applied when deviating from the EAT by ACC. Steps are foreseen that can potentially alleviate some of these difficulties and reduce the task complexity of the controllers working position.

Expected future situation at Schiphol Airport

The strategy of LVNL is laid down in the ATM System Vision and Strategy and lists all the strategic developments in the short- and medium term. These strategic steps are necessary to accommodate an increased traffic demand with improved safety performance, and to improve the environmental performance as a consequence of more demanding targets. Amongst a number of developments focussing on the entire operation, the following specific developments focus on the improvement of the management of arrival traffic at Schiphol (see also SESAR, 2007):

- Introduction of Trajectory Based Operations (TBO);
- Continuous Descent Approaches (CDAs) during daytime operations to enhance environmental performance;
- Development of conflict-free routes in the TMA to reduce task complexity for APP controllers;
- Noise-friendly approach procedures;
- More accurate inbound planning; and

LVNL ATM system goal

The current LVNL ATM System is insufficiently able to cope with fixed, noise friendly and conflict-free routes in the TMA, as these cannot be implemented under the assumption that current capacity levels must be met. The main reason for this is the inaccuracy following from the fact that the traffic is delivered by ACC to APP within a plus or minus 120 seconds time window. This inaccuracy leads to the requirement for APP to manually optimize the handling of the arrival streams within the TMA. It is assumed that improved accuracy at the delivery from ACC to APP will result in more stable and predictable traffic streams in the TMA. These more predictable and stable traffic streams will result in better service to the customer, but

should also lead to a reduction of work complexity for controllers.

The night-time-operations currently in use at Schiphol resemble CDA procedures in which aircraft are able to fly their own descent profile and accompanying speeds within the active constraints. This increases flight efficiency and reduces CO₂ emissions. However, these procedures, consisting of fixed arrival routes from 7000 feet to the runway threshold, result in a capacity of 24 arrivals per hour, well below the daytime number of 35 arrivals per hour. This means that the required capacity levels cannot be met with this type of operations in daytime situations under current circumstances. Additionally, the current inaccuracy is also a limiting factor in this perspective. An improved accuracy in the delivery from ACC to APP is a requirement for the introduction of high capacity CDAs during daytime operations.

A second aspiration for improved accuracy is the increased workload in case of bunching. If aircraft are not transferred in a bunch, but longitudinally separated, then the workload is much better manageable resulting in stable and predictable traffic streams. As a result the task to manage these flows becomes also easier for the controllers.

The last aspiration for improved accuracy is predictability and transparency to the airspace users. Pilots will be better able to manage the most efficient flight profile if their flight is planned well ahead. As improved accuracy will ultimately lead to improved transfers and therefore a better planning, it is assumed that there are major benefits on the airspace users' side.

As mentioned before, the current ATM System does not allow for additional workload for the controller by requiring them to increase their delivery accuracy of arrival traffic over the IAF from plus or minus 120 seconds to plus or minus 30 seconds. Previous studies have shown that estimations of future aircraft positions by controllers become increasingly inaccurate the further in time the prediction is made (Boudes and Cellier, 2000). Since in the future ATM System Strategy an increased accuracy is required earlier in time, this would place a too great a burden on the capabilities of controllers. In order to mitigate this potential increase in workload, system support is foreseen to enable the increased accuracy performance target to less than plus or minus 30 seconds. The Speed And Route Advisor (SARA) tool was designed to support the controller to meet this target.

Basic SARA functioning

The SARA tool operates by providing controllers a speed and route combination for every inbound flight. The speed and/or route combination is displayed to the controller and it will allow the controller to give a single speed and route clearance to the aircraft for the entire descent. A single clearance will have the potential advantage that it will decrease the workload for the controller and aircrew. It will also allow the aircrew to use the Flight Management Computer (FMC) in the descent, thereby optimizing the descent profile as much as possible within the active constraints.

The SARA tool relies on several functions in the ATC system: IBP, surveillance data, and a Trajectory Predictor. The performance of these support functions determines the performance of SARA. The SARA logic processes a flight in seven steps. These steps are described below:

1. The flight appears to the ATC system and is entered in the IBP.
2. Once the planning is considered stable, the SARA process starts.
3. SARA uses the Expected Approach Time (EAT) for the flight.
4. SARA interacts with the TP and collects the flights current position and plan. It also uses the TP to calculate the flights Estimated Time Over (ETO) the IAF. For this calculation SARA assumes that the route entered into the ATC system will be the route flown.
5. SARA compares the EAT and ETO. If the difference is outside a set threshold (less than plus or minus 30 seconds) it will initiate the process to find a new speed/route combination to match the requirements following from the threshold.
6. An iterative process is started where SARA uses the TP to calculate a speed/route combination that will bring the aircraft to the IAF such that the EAT and ETO is below the threshold value.
7. Once a speed and route combination is found that is within the threshold it is communicated to the controller in an integrated manner on the radar screen. For this experiment, the advisories were integrated in the aircraft label.

When developing new concepts it is difficult to determine the requirements in enough detail when the system is only described at a conceptual level. In the development of SARA this was approached by the use of iterative development cycles. Each cycle consisted of three steps: requirement definitions, technical development and system testing. The test results from one cycle feed into the requirements of the next. The SARA functionality was developed using five such development cycles. It allowed the controllers to try out a range of possible solutions. In each cycle the most promising where selected for further development. In addition to being a flexible development process it also kept the development costs down because ideas that proved difficult, expensive or not helpful during development could be detected early.

In accordance with earlier recommendations, controllers remain in control and are fully responsible for separation of the traffic (Prevot, Lee, Callantine, 2003). SARA only supports with calculating the speed and route combination best suited to meet the planning. In future developments of SARA a conflict detection and resolution step could be added to the process.

The impact of SARA on air traffic controllers

With SARA, the operation at Schiphol will gradually change from a tactical first-come-first-serve operation towards a time-based operation. These operations might have a quite large impact on the controller's SA, and hence his subsequent capacity to act. However, the degree to which SA is affected depends on the specific operational design and task allocation between humans and systems. The SARA tool could help controllers to instruct the right speeds and routes to aircraft in order to meet a specific point on time. This might decrease their workload as once the instruction is given the controller mostly needs to monitor the follow up. Only in case of a conflict he would need to give an updated instruction.

With SARA, controllers will have to incorporate time as a fourth dimension in their mental picture in order to plan,

prioritize and sequence flows, as well as to assure separation. This requires more anticipation and strategic thinking than nowadays. In their current way of working, their decisions are based on certain three-dimensional patterns of aircraft on a certain moment of time. Being in time on a waypoint within small margins changes the controllers SA because more 'thinking-in-time' is required than they are used to. Currently the controllers are more 'thinking-in-distance' and this determines how they sequence the arrival traffic. Consequently, with SARA tactical control will move towards more strategic control with a larger planning horizon (Oprins, Zwaaf, Eriksson, Van de Merwe and Roe, 2009).

In addition, SARA implies that certain tasks of controllers are moved to the system. Currently, controllers determine the speeds and routes for aircraft by themselves. SARA will help them in the decision making process by providing speed and route advisories. Controllers might lose their feeling of control when their work moves too much towards supervisory control. They might have difficulty to trust the system when solutions are in conflict with their own plan and their SA might be undermined. In other words, they cannot use their own strategies for traffic handling anymore. Dependent on the specific design of SARA, controllers could have less insight into the specific flight paths of aircraft. This will definitely decrease their SA. Consequently, it might make it difficult for them to renew their SA if manual interventions are needed in case of system failures and other circumstances (e.g., weather) in which SARA may not work. Switching between these automated (routine) and manual (non-routine) operations can substantially increase their workload. It depends on the frequency of using conventional methods to which extent the controllers can act as the fallback.

METHOD

To understand the impact of SARA on the behaviour of controllers an experiment was devised that investigates the influence of SARA on controller's delivery accuracy, workload and Situation Awareness. Furthermore, the experiment aimed to gain insight into potential improvements that could be made to the tool in order to optimize its effectiveness.

Experimental design

The Real Time Simulations with SARA were performed at NLR's ATC Research SIMulator (NARSIM). The experiment was conducted during four days that were spread out over two weeks. During the first two days, eight LVNL controllers participated (N=8). In the second week four MUAC controllers joined four LVNL controllers (N=2x4). The design was such that in the first week, the SARA concept within LVNL airspace could be investigated. During the second week, the influence of MUAC controllers could be researched. For consistency and comparison purposes, the data presented in this paper is data derived from LVNL controllers only.

A single simulation run involved two controllers and two pseudo-pilots working in tandem for parts of the LVNL managed airspace (Amsterdam ACC sector 1 and sector 2). The pseudo-pilots had radio contact with the controller for the specific sector. In week 2, the LVNL controllers and pseudo-pilots were joined by two MUAC controllers and two pseudo-pilots that controlled aircraft in specific MUAC sectors

(Coastal and Munster upper area sectors). Because NARSIM has eight controllers working positions, four identical runs could be executed simultaneously in the first week versus two identical simultaneous runs in the second week.

Two familiarization runs were executed for each pair of controllers to familiarize them with the simulator and the SARA Human Machine Interface (HMI). Next, the pairs executed four experimental runs in the first week and three runs in the second week. For comparison purposes, the same traffic sample was used for all runs. However, to avoid familiarization with the traffic sample, the callsigns were shuffled between each run. Furthermore controllers switched working positions to also avoid effects resulting from the familiarity of the controllers with the traffic for a specific sector and inter-controller working strategies. The measured traffic sample contained 18 flights with destination Schiphol. In the first week the four experimental runs consisted of two baseline runs and two SARA runs. Run 1 resembled current operations and functioned as a baseline in which controllers had standard system support and delivered aircraft at the IAF with an accuracy of plus or minus 120 seconds or less compared to the EAT. Run 2 functioned as a second baseline in which controllers had a stricter time target similar to the SARA runs (less than plus or minus 30 seconds) and limited system support. The support consisted of a delta time (ΔT ; EAT - ETO) presented in the aircraft label. In runs 3 and 4, SARA provided speed-only advisories, and speed and route combinations respectively. During the second week MUAC controllers joined the four remaining LVNL controllers. A baseline run with a target of less than plus or minus 30 seconds, a SARA speed run and a SARA speed & route runs were performed (run 5, 6 and 7 respectively). Within MUAC airspace the controllers issued speed and route advisories to the aircraft. However, no route options were available within MUAC airspace. Therefore route advisories that were issued by SARA in MUAC airspace were only applicable in LVNL airspace. By providing the route instructions as early as possible the pseudo pilot was able to let the aircraft fly a more optimized descent profile. The properties of the simulation runs are depicted in Table 1.

| Run | IAF target time (sec) | System support | Participating controllers |
|-----|-----------------------|--------------------|---------------------------|
| 1 | Within +/- 120 | Standard | LVNL |
| 2 | Within +/- 30 | Delta T in label | LVNL |
| 3 | Within +/- 30 | SARA speed | LVNL |
| 4 | Within +/- 30 | SARA speed & route | LVNL |
| 5 | Within +/- 30 | Delta T in label | LVNL & MUAC |
| 6 | Within +/- 30 | SARA speed | LVNL & MUAC |
| 7 | Within +/- 30 | SARA speed & route | LVNL & MUAC |

Table 1. Properties of the simulation runs. Run 1, 2 and 5 are baseline runs, the others are the SARA runs.

Quantitative and qualitative data was gathered during and after each simulation run. First, the accuracy with which the controllers managed to meet the EAT for each aircraft was measured. This measurement was called 'EAT adherence'. As a subjective measure of workload the Instantaneous Self Assessment (ISA) was used. Controllers were prompted for input every three minutes. Objective measures of workload consisted of calculating the total number of R/T calls (i.e. radiotelephony; the verbal instruction administered to the aircrew), the average time spent on R/T by each controller, and the number of instructions entered into the system through the

Touch Input Devices (TID; i.e. after instructions are administered to the aircrew the controller enters them into the system via a TID). Directly after each simulator run, the controllers filled in an adapted version of the SASHA-Q Situation Awareness questionnaire (Dehn, 2008). Additionally, these questionnaires also contained open questions regarding workload, usability and acceptance. Interviews were held after each run to obtain in-depth information regarding their experiences with SARA. During the runs, human factor observers were taking notes.

RESULTS

Repeated measures analyses of Variance (ANOVA) were used for statistical comparisons. Partial eta-squared (η^2) is given as a measure of effect size. Pairwise comparisons were performed where appropriate with Bonferroni corrections. For each analysis an $\alpha < .05$ was used. All analyses were performed using SPSS 15.0.1. Only the results of the first week were used in the statistical analyses (run 1 to 4). The number of participants in the second week (run 5 to 7) was too few to perform meaningful statistical analyses. Therefore, these results are presented as an illustration and addition to the results of the first week.

EAT adherence

Data was obtained for 18 flights in the four experimental runs of the first week (run 1 to 4) and was analyzed for missing values and outliers. Data was gathered for four pairs of controllers. The results showed a significant delivery accuracy improvement when SARA was used, $F(3,63) = 40.918$, $p < .001$, $\eta_p^2 = .661$. The means and standard deviations for each run are depicted in Figure 2.

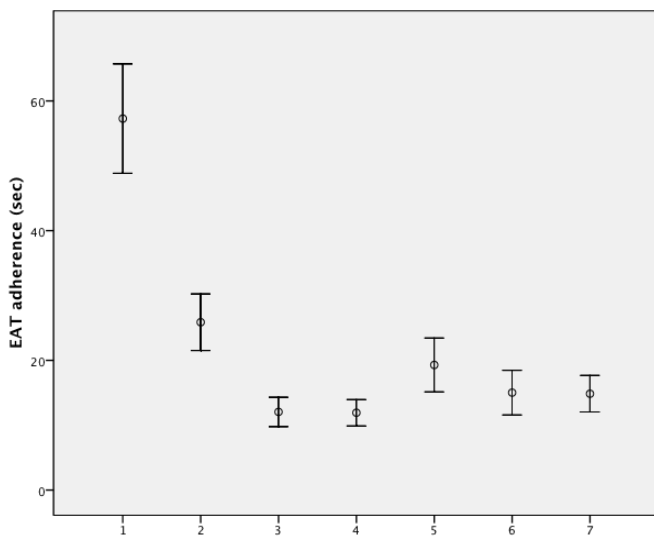


Figure 2. EAT adherence

The average absolute EAT adherence improved from the two baseline runs (run 1 and 2) to the two SARA runs (around 57 and 25 seconds accuracy to around 12 seconds accuracy; run 3 and 4). No significant differences were found between the speed only and the speed & route variants (run 3 and 4). Interestingly, setting the target at less than 30 seconds and

providing the controllers with limited system support (a delta T in the aircraft label; run 2) already significantly improved the accuracy to approximately 25 seconds.

The data for the second week show similar results (run 5 to 7). Delivery accuracy lies around 15 to 20 seconds with limited system support (run 5) and the SARA runs (run 6 and 7). There seems to be little difference between the first and the second week.

Workload

ISA. Eight LVNL controllers produced nine ISA scores each during each run during the first week (run 1 to 4). A significant effect was found between the four runs, $F(3,68) = 17.256$, $p < .001$, $\eta_p^2 = .432$. Workload in the SARA runs (run 3 and 4) was rated lower than the second baseline (run 2). Run 2 imposed a significantly higher workload on the controllers compared to the average of their ratings of the other runs, $p < .01$ (run 2 vs. run 1, 3 and 4). Run 4 (speed and route) was rated to be as equally demanding as run 3 (speed-only), $p = .701$.

The results from the second week show similar results. The second baseline (run 5) appears to be imposing a higher workload on the controllers compared to the SARA runs for week 2 (run 6 and 7). The results are depicted in Figure 3.

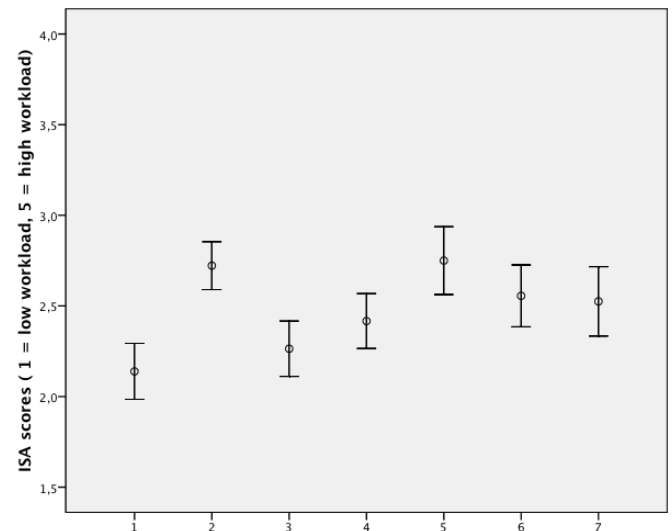


Figure 3. ISA scores

R/T calls. After removing one outlier from the dataset of the first week seven measurements were obtained for the total number of R/T calls for eight LVNL controllers. A significant effect was found for this type of workload measure, $F(3,3) = 21.985$, $p < .05$, $\eta_p^2 = .956$. The SARA speed and routes run (run 4) required the lowest number of calls. The number of calls in this run was found to be less than baseline run 2 and the SARA speed-only run (run 3). A potential difference was found between run 4 and baseline run 1 ($p = .067$). SARA run 3 did not differ from the two baseline runs (run 1 and 2). The two baseline runs did not differ from each other.

A large spread in the data was found for the baseline in week 2. This is likely due to the few participants in week 2 (4

LVNL controllers) who also showed large individual differences in the number of R/T calls. The SARA runs in week 2 (run 6 and 7) appear to be in line with the SARA outcomes in the first week (run 3 and 4). The results are depicted in Figure 4.

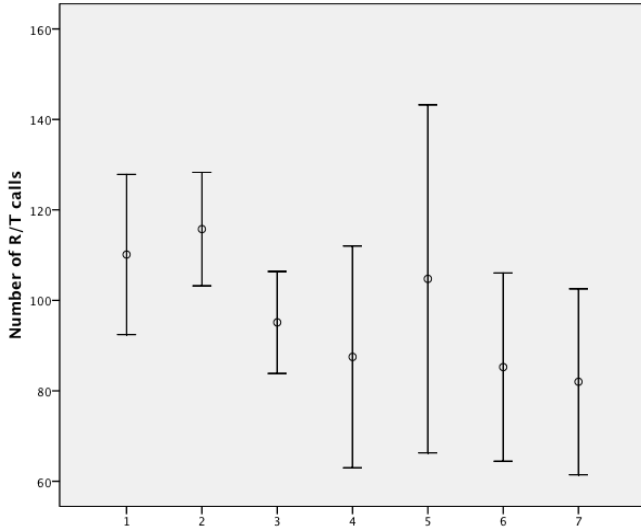


Figure 4. Number of R/T calls

R/T Time. Eight measurements were obtained for the total time spent on R/T calls (in seconds) for the four simulation runs in the first week. An ANOVA showed significant differences between the four runs, $F(3,4) = 28.951, p < .01, \eta_p^2 = .956$. The lowest amount of time spent on R/T was found in the SARA speed & route run (run 4). There were no differences found between the first baseline (run 1) and the SARA speed only run (run 3). However, these two runs showed a reduced amount of R/T time compared to the second baseline (run 1 and 3 vs. run 2). No differences were found between the two baseline runs 1 and 2.

Again, the data from the second week does not appear to be much different from the data captured during the first week. See Figure 5 for the means and standard deviations for time spent on R/T.

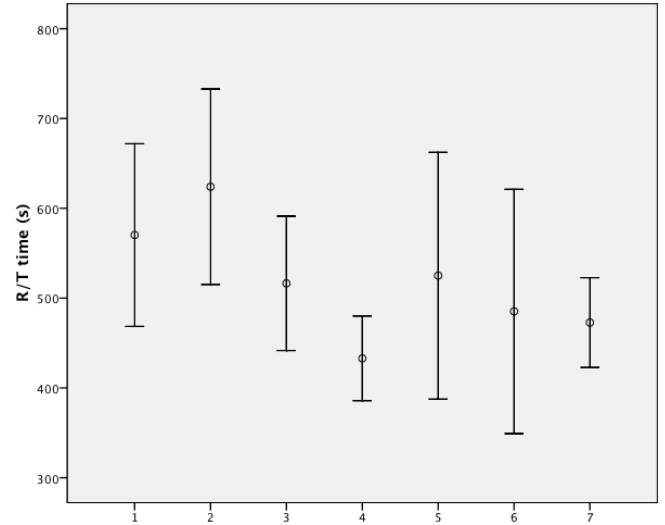


Figure 5. Time spent on R/T (s)

TID inputs. Eight measurements were obtained for the number of TID inputs for the four simulation runs in the first week. An ANOVA showed significant effects for the number of TID inputs, $F(3,4) = 11.091, p < .05, \eta_p^2 = .893$. The lowest number of inputs was found in the SARA speed & route run (run 4) compared to baseline run 2 and SARA run 3. A potential difference was visible between baseline run 1 and SARA run 4, $p = .051$. The highest number of inputs was found in baseline run 2 and potentially with baseline run 1, $p = .081$.

The results for week 2 seem to be similar compared to week 1. Again, large individual differences were found for run 5 consistent with the number of R/T calls of run 5. The three runs performed in week 2 (run 5 to 7) do not show large differences in terms of TID input amongst themselves and between the first week and the second week (except compared to run 4). The results are depicted in Figure 6.

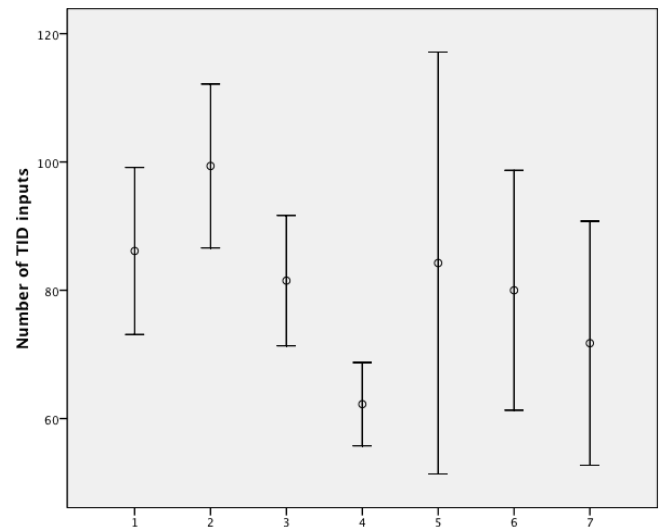


Figure 6. Number of TID inputs

Situation Awareness

The questions from the SASHA-Q questionnaire were averaged to serve as a total SA score for each controller (N=8). Four questions were used that were applicable to both the SARA runs (run 3 and 4) and the baseline runs (run 1 and 2). The Repeated Measures ANOVA showed a significant difference in SA scores between the four runs in the first week, $F(3,29) = 37.304$, $p < .001$, $\eta_p^2 = .794$. SARA runs 3 and 4 showed lower SA ratings compared to the two baseline runs 1 and 2. No significant differences were found between the two SARA runs (run 3 and 4) as well as between the two baseline runs (run 1 and 2).

Data from the second week appear to be similar to the data found in week 1. It seems that the baseline run in week 2 (run 5) received higher average SA scores compared to the SA scores for the two SARA runs (run 6 and 7). The results are depicted in Figure 7.

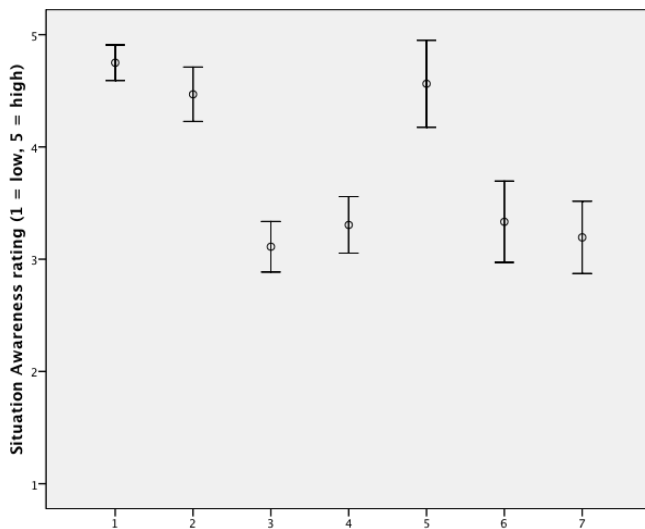


Figure 7. Situation Awareness ratings

Qualitative results

Workload. A reduction in workload was experienced by the controllers with the use of SARA, especially with route-options enabled. This was especially noticeable in terms of R/T load. Some controllers however, mentioned that they felt that, with SARA active, other activities needed to be performed and not necessarily more or less. That is, a change in working method was experienced by some controllers that may have resulted in a lower physical task load, but a similar mental workload.

With the current implementation of SARA only arriving aircraft are provided with speed and/or route advisories. This led some controllers to mention that potential difficulty of working with two working methods: the arrival traffic stream under SARA advisories and the departure traffic stream under 'normal' control which could potentially add to their workload.

Situation Awareness. Controllers mentioned that with SARA they felt 'less engaged' in the traffic situation compared to the baseline runs. It was mentioned that with SARA they followed an advice and monitored its progress. Some controllers mentioned that they felt that, because SARA produces an advice at the FIR entry to meet the time over the

IAF, this would mean a solution for more than meeting the time alone, i.e. a conflict-free advice. This sometimes lead to controllers solve conflicts late rather than early.

Controllers also mentioned that when SARA was active, they felt that they spent time to understand SARA's 'plan' as part of an effort to create a mental picture of the traffic situation in contrast to generating their own plan.

Controllers changed their interaction with the SARA tool during the course of the simulations. In the second week it was observed that controllers regained some of their SA by not adhering to advisories all the time. It was observed that sometimes advisories were used as a 'general guidance' to give an aircraft a speed that would more or less be adequate to meet the time over the IAF. When SARA provided subsequent advisories, these would first be evaluated by the controllers for their usefulness before they were instructed. This was contrary to the controller's behaviour early in the simulation where every given advice was accepted and instructed.

Working method. Controllers mentioned they had difficulty with building a traffic sequence whilst using SARA. In normal operations (as mimicked by runs 1 and 5) sequences were made with a separation of 5 nm and an EAT adherence of less than plus or minus 120 seconds. With SARA the target was reduced to less than plus or minus 30 seconds. This clearly resulted in the controller focusing more on time (meeting the time over the IAF) than on distance (maintaining 5 nm separation).

DISCUSSION

The aim of the experiment was understand the impact of SARA on its ability to support the controller in a more accurate delivery of traffic at the IAF. Furthermore, the experiment aimed to understand the impact of this support tool on controller functioning. Specific emphasis was laid on the controller's workload and SA. The results showed that with the support of SARA the controller was able to deliver arrival traffic more accurately to the IAF. An initial gain in delivery accuracy was seen even when the controller had minimal system support (only a delta T in the aircraft label; run 2). However, with the aid of SARA this accuracy was further improved (run 3 and 4). Similar results were found in the second week (run 5, and 6 and 7 respectively).

With SARA subjective workload (ISA scores) did not increase compared to the baseline. However, Metzger and Parasuraman (2006) showed that communication and coordination tasks also can be a considerable source of workload to the controller, especially under high-traffic conditions. The results from this experiment showed that objective workload (number of R/T calls and TID inputs) reduced compared to the baseline, especially for SARA with speed and route options. This indicates that at an important part of controller workload can be reduced when using SARA. The results also showed that workload (subjectively as well as objectively measured) was highest for the run with minimal system support (run 2).

Interestingly, SA was highest in the two baseline runs (run 1 and 2) and dropped significantly when SARA was used (although still rated as above average). The ATC Performance Model suggests that SA is one of the prime information processing components of a controller (Oprins, Burggraaff and Van Weerdenburg, 2006; Oprins, 2008). Without SARA a controller builds up a mental picture by perceiving, interpreting

and anticipating on the traffic stream. Based on this continuous process the controller decides on the required instructions for aircraft in the traffic stream. With SARA part of this activity is transferred to SARA since it provides the controller with advisories that have not been part of the mental processes of the controller. Alternatively, the mental picture created by the controllers and the resulting instructions may differ from the solutions provided by SARA. Therefore, it is understandable that controllers rated their SA as lower compared to the baseline scenarios.

SARA was specifically designed to aid the controller in an increased delivery accuracy of traffic over the IAF. It was reasoned that without SARA controllers would experience an unacceptable increase in their workload whilst aiming to reach a target of plus or minus less than 30 seconds. It is interesting to note that with minimal system support the controller is able to deliver traffic more accurately at the IAF. As expected, SA is maintained under these circumstances. However, it was also shown that workload is highest in this scenario (subjectively as well as objectively measured). This finding seems to support the notion that controllers need a support tool to improve their delivery performance. In this current setup however, this delivery improvement may come at the cost of a reduced, albeit still acceptable, level of SA.

Controllers' interaction with SARA changed over the course of the experiment. During the first week (run 1 to 4) controllers adhered to almost every advice provided by SARA. During the second week some controllers regained some of their SA by assessing every advice for its applicability. Also controllers were better able to anticipate on SARA's advisories. This helped controllers to better manage their workload and SA. This suggests that sufficient familiarization with SARA is required before it is used in an operational setting. In this experiment, controllers became more used to SARA after a few hours. This may be an indicative time for training purposes when implementing SARA in an operational setting.

The stricter focus on time (30 seconds vs. 120 seconds) had a large influence on the working strategies of the controllers. Nowadays time is of lesser importance since controllers focus on creating 5 nm sequences. When doing so they most often meet the required time over the IAF. Therefore, in present operations time is of less importance. With a target of within plus or minus 30 seconds controllers will have to invest more effort to meet the target and requires a different mind set by the controllers. This may mean that generating sequences of 5 nm may not be enough to meet the target, but that more precise actions are required. In SESAR and NextGen there is a large focus on stricter time-based operations in which the entire trajectory of an aircraft is planned from gate to gate together with strict fixed times over waypoints (SESAR, 2007). The experiences in this study may shed some light on the expected future working methods for controllers.

In this experiment two versions of SARA were tested: a speed-only options and a speed and route option. A third option is foreseen that will incorporate conflict management (CM) to provide the controller with conflict-free speed and/or route advisories. The implementation of the third phase of SARA is foreseen around the implementation time of SESAR (i.e. 2020). This particular version of SARA was out of the scope of this experiment and it was therefore not possible to investigate its consequences for the controller. However previous studies on controller performance and workload under mature Free Flight

may hint at considerations for design and implementation of this version.

Free Flight is a concept that aims to shift most of the separation responsibility from the controller to the aircrew. In this situation controllers monitor the flow of traffic and only intercept to ensure separation, to preclude exceeding airport capacity, to prohibit unauthorized flight through special airspaces and to ensure safety (RTCA, 1995). This means that most of the time controllers only monitor the flow of traffic and are not actively controlling whilst still being responsible for the separation of aircraft.

Galster, Duley, Masalonis and Parasuraman (2001) found that in a study simulating Free Flight controllers' performance decreased in terms of speed of conflict detection and the number of detected self-separations (movements when two aircraft ensure separation without controller intervention). They concluded that controllers may become vulnerable under such a scheme in which separation decisions are ceded to the aircraft but controllers are still responsible for separation.

A similar situation may occur in the case of SARA with a CM function. In the present study a decrease in SA was found with the use of SARA with speed and/or route options. This finding is similar to the out-of-the-loop-performance problem with an increased level of automation (Endsley and Kiris, 1995). A further impact on SA is expected when SARA's advisories are conflict free. Checking SARA's advisories for potential conflicts becomes unnecessary since they are intended to be conflict free. With controllers partially out of the loop, they may not be up to the challenge, due to complacency issues, in case conflicts are not resolvable by SARA (Galster et al., 2001; Wickens, Mavor, Parasuraman and McGee, 1998x).

Several studies have provided recommendations on the out-of-the-loop-performance problem. It has been argued that to keep controller in the loop, they should retain some of their responsibilities and automation should support them in their decision-making (Wickens et al., 1998). Previous research also showed that controllers more easily accept automation support if they are in command (Prevot et al., 2003). That is, they found that solutions over which controllers had a choice were more readily accepted compared to solutions that appeared automatically. For SARA, this may mean that controllers should be able to have a choice about which speed and/or route combination they want to issue rather than having SARA presenting a single solution only. This may be especially relevant for SARA with CM. One solution for keeping the controller in the loop may be to present various conflict-free solutions which the controller can choose from, possibly with visual support. This way, controllers remain in control of the traffic, whilst SARA is able to support the controller in his decision-making.

A further design issue that was proven to be useful is the use of instant feedback. When controllers issued instructions to the aircraft controllers immediately perceived the consequences of their instructions through the change in delta T for that aircraft on the radar screen. This is in contrast to current operations in which, due to technical reasons, the change in delta T only slowly changes closer to the IAF. Instant feedback enhances the controller's ability to stay on top the traffic situation by 'scanning the traffic, identifying the need for an action, and issuing a proactive instruction' (Prevot et al., 2003, p. 8). Any delay in such feedback may cause the controller to become behind in handling other traffic with reactive

controlling behaviour as a result. The behaviour of the delta T function with SARA speed and/or route was shown to be a successful implementation of instant system feedback.

The iterative design-cycle used in this study may be useful in overcoming such issues in further developing SARA for implementation. In such a cycle controllers and designers can test various alternatives in more detail. When a version is considered stable, Real-Time Simulations such as the one presented in this study can subsequently be used to investigate the impact on controller functioning (workload, SA and working method) in more depth. Based on these findings changes can be made to the design before it is implemented into an operational system. By using such a cycle (design, evaluation, simulation, implementation) with the users at the center of the developments, the potential for a successful system that can meet stricter future requirements whilst keeping controllers in the loop can be increased.

Conclusion

It can be concluded that SARA was successful in aiding the controller in reducing variability of arrival traffic over the IAF. A reduction of objective workload in combination with an increase in accuracy was observed without an increase in mental workload. SA was affected, although it was still rated above average. Familiarity with SARA may have played a role here as controllers changed their interaction with SARA in the second week and used it more as an advice tool, the way it was intended. The stricter focus on time rather than distance caused a larger impact than anticipated. This may mean that controllers need to change their way of thinking from 'thinking-in-distance' to 'thinking-in-time' (Oprins, Zwaaf, Eriksson, Van de Merwe and Roe, 2009).

This experiment has provided some insights into the future of ATM and its consequences for the controller. SESAR and NextGen are aiming for stricter time-based-operations with subsequent automation support for the controllers. At present it is not immediately clear which impact this may have on controller functioning, especially with decision support tools utilizing conflict probes. A conflict-free SARA should be able to fully support the controller if strict design principles are taken in to account. The iterative design-cycle used in this study forms an integral part of the future developments of SARA such that controllers are supported in their tasks and that increases in efficiency, safety and environmental performance can be achieved.

ACKNOWLEDGEMENTS

This project is sponsored by the Knowledge and Development Centre (KDC); a cooperation of Dutch ANSPs, airlines, airports and research centres. We would like to thank the SARA project team for their efforts into the development of SARA and their support in this experiment.

REFERENCES

Averty, P., Collet, C., Dittmar, A., Athènes, S. & Vernet-Maury, E. (2004). Mental workload in air traffic control: an index constructed from field tests. *Aviation, space, and environmental medicine*, 75(4), 333-341.

- Boudes, N. & Cellier, J. (2000). Accuracy of Estimations Made by Air Traffic Controllers. *The International Journal of Aviation Psychology*, 10(2), 207-225.
- Dehn, D. M. (2008). Assessing the impact of automation on the air traffic controllers: the SHAPE questionnaires. *Air Traffic Control Quarterly*, 16(2), 127-146.
- Endsley, M.R., & Kiris, E.O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors*, 37(2), 381-394.
- Endsley, M.R. (1997). *Situation awareness, automation and free flight*. Paper presented at the FAA/Eurocontrol Air Traffic Management R&D seminar. Saclay, France.
- Galster, S.M., Duley, J.A., Masalonis, A.J. & Parasuraman, R. (2001). Air Traffic Controller Performance and Workload Under Mature Free Flight: Conflict Detection and Resolution of Aircraft Self-Separation. *The International Journal of Aviation Psychology*, 11(1), 71-93.
- Garland, D.J., Stein, E.s. & Muller, J.K. (1999). Air traffic controller memory: capabilities, limitations and volatility. In: Garland, D.J., Wise, J.A. & Hopkin, V.D. (Eds.). *Handbook of aviation human factors*, (455-496). Mahwah, NJ; Erlbaum.
- Metzger, U. & Parasuraman, R. (2006). Effects of Automated Conflict Cuing and Traffic Density on Air Traffic Controller Performance and Visual Attention in a Datalink Environment. *The International Journal of Aviation Psychology*, 16(4), 343-362.
- Oprins, E. (2008). *Design of a competence-based assessment system for ATC training*. Doctoral dissertation, Maastricht University.
- Oprins, E., Burggraaff, E. & Weerdenburg, H. van (2006). Design of a competence-based assessment system for ATC training. *The International Journal of Aviation Psychology*, 16(3), 297-320.
- Oprins, E., Zwaaf, D., Eriksson, E., Van de Merwe, G.K., Roe, R. (2009). Impact of future time-based operations on Situation Awareness of air traffic controllers. Submitted to the ATM R&D seminar 2009.
- Prevot, T., Lee, P., Callantine, T. & Smith, N. (2003) Trajectory-oriented time-based arrival operations: results and recommendations. *Proceedings of the 5th ATM Seminar Budapest*.
- RTCA (1995). *Report of the RTCA Board of Director's Select Committee on free flight*. Washington, DC: Author.
- SESAR (2007). Deliverable 3: The ATM Target Concept (DLM-0612-001-02-00a). SESAR consortium.
- Shebilske, W.L., Goettl, B.P. & Garland, D.J. (2000). Situation awareness, automation and training. In: Endsley, M.R. & Garland, D.J. (Eds.). *Situation awareness: analysis and measurement* (pp. 303-323). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schneider, W. (1990). Training high-performance skills: fallacies and guidelines. In: Venturino, M. (Ed.). *Selected readings in human factors* (pp. 297-311). Santa Monica, CA: The human factors society.
- Wickens, C.D., Mavor, A., Parasuraman, R., & McGee, J. (1998). *The future of air traffic control: human operator and automation*. Washington, DC: National Academy Press.