The influence of task load on situation awareness and control strategy in the ATC tower environment.

Maik Friedrich[1] Maresa Biermann[2] Patrick Gontar[2] Marcus Biella[1] Klaus Bengler[2]

June December 11, 2017

Words 6659

[1] German Aerospace Center, Braunschweig

[2] Technische Universität München

Author:

Maik Friedrich,

ORCID 0000-0003-3742-2322

maik.friedrich@dlr.de

+49 (531) 295 - 2598

German Aerospace Center

Lilienthalplatz 7, Braunschweig, NI D-38108

Acknowledgments

This research received funding from the German Aerospace Center institutional funding mechanism dedicated to the human center approach for automation, which is sponsored by the Federal Ministry for Economic Affairs and Energy.

Abstract

The safe and efficient operation of air traffic is highly dependent on the performance of the Air Traffic Control Officer (ATCO). The Air Traffic Control Officers control the traffic within defined areas by monitoring the traffic and granting clearances. A key element in analyzing the ATCOs is their interaction with the environment through their workplace. Especially the influence of task load on their situation awareness (SA) and applied control strategy provides information on the quality of the workplace. As task load increases, controllers are able to maintain performance by using different management or compensation strategies. This article supports the evaluation of ATCO's workplaces by focusing on whether probe techniques for assessing SA are applicable for tower control operation and for measuring the influences of increased task load on the control strategy. An experiment with nine ATCOs was conducted in a simulated real-time air traffic control environment. Different measurements for SA were applied and compared regarding their efficiency and validity. The manipulation of task load and visibility influenced the SA and control strategy at the same time. Performance metrics were selected in advance to evaluate the participant's efficiency. SA was measured with a probe technique and an offline self-assessment method. Findings suggest that probe techniques increase the insight into the understanding of SA in comparison to self-assessment and that they are applicable to the air traffic control environment. Control strategies were derived from the information-gathering process via the eye-movement behavior and connected to task load. The results imply that SA is part of the individual performance and that increasing demand through task load is handled with an adaptation of the control strategy.

Keywords

Task load; Situation Awareness; Strategy identification; Eye Tracking; Air Traffic Control

1 INTRODUCTION

Air traffic control (ATC) is a system that relies highly on the capabilities of human operators for the safe processing of flights. With European air traffic estimated to double by the year 2020 (EUROCONTROL/FAA 2010), the tasks of Air Traffic Control Officers (ATCOs) are becoming more challenging. ATC covers the surveillance and control of the movements in airspace and on the runways of airports to ensure a safe, organized, and fluent processing of air traffic (Mensen 2014). Kontogiannis and Malakis (2013) characterize the ATC task with “time pressure, multiple goals, interconnected tasks and high consequences of error”. The task takes place in a dynamic environment, in which information from different navigation, communication, and surveillance systems must be continuously processed. New technologies are currently being developed to support operational personnel, such as assistant systems in the case of low visibility (Teutsch and Postma-Kurlanc 2014) or completely new workplaces in the case of remote tower operations (Fürstenau 2016). The introduction of new systems, if performed without the proper knowledge of related effects (e.g. situational awareness and workload), always increases the probability of errors (Baumann and Krems 2007; Endsley 2000). When incidents or accidents in ATC occur, the consequences are often severe, especially in civil aviation. These events are frequently attributed to human error (Redding 1992). This is why a thorough understanding of human performance and influencing factors on ATC is essential for further development in this research area.

The ATC task is divided into en route control, approach control, and tower control (Mensen 2014). The majority of studies focus on the en route and approach control workplace (e.g. Durso et al. 1998; Edwards 2013; Karikawa et al. 2014; Lee et al. 2012), but only a few have focused on ATCOs performing tower control operations (e.g. Fürstenau et al. 2013; Lange

2014; Papenfuss et al. 2010). The ATCOs who are responsible for tower control use air/ground communications, radar, direct view on the airfield and binoculars/zoom cameras (Hadley et al. 1999) to regulate the speed, heading, and flight level of aircraft on the ground and in the direct proximity of the airport. Considering the expected increase in traffic, this article addresses the shortcoming in understanding the influence of workload onto the tower control operations by conducting research with operational experts in a high-fidelity simulator. This will prepare the foundation for introducing new technologies with the purpose of increasing safety and efficiency.

According to EUROCONTROL/FAA (2010), "human performance […] refers to the adequate performance of jobs, tasks and activities by operational personnel – individually and together" (p. 8). This performance depends on the person (capabilities, skills, knowledge, motivation, etc.) and the context of work (systems, organization, and environment). Consequently, human performance of ATCOs is important for the overall system safety and effectiveness of aviation operations (Hadley et al. 1999). Human performance is influenced by performance-shaping factors determining human error probability (Ambroggi and Trucco 2011). Performance-shaping factors have been mainly investigated separately in research on human performance in ATC (Cox-Fuenzalida 2007; Edwards 2013). The lack of research on the interaction of these factors was addressed at an OPTICS[1] Expert Workshop in 2014 as the most important goal in human factors research (Kirwan et al. 2014). Based on a literature review of over 400 incident reports[2] of Eurocontrol, Edwards (2013) identified the following nine

---

[1] Observation Platform for the Technical and Institutional Consolidation of Safety Research; OPTICS, www.optics-project.eu/

[2] According to EUROCONTROL Safety Regulatory Requirements (ESARRs), incidents (potentially) affecting safety must be reported to the Accident/Incident Data Reporting (ADREP) of the International Civil Aviation Organization (ICAO), www.skybrary.aero/index.php/Reportable_Incidents

performance-shaping factors as influential: Communication, teamwork, trust, fatigue, stress, vigilance, attention, mental workload, and situation awareness (SA). Especially the factors mental workload and SA are often used as performance measurement and frequently considered in relation to each other (e.g. Endsley 1995b; Vidulich 2000; Wickens 2002), but the results of their interaction differ (Edwards 2013; Lee et al. 2012). Performance as an objective measurement for all participants can be interpreted in terms of safety, efficiency and orderliness (Griffin et al. 2000).

2 WORKLOAD, SITUATION AWARENESS, AND CONTROL STRATEGIES

2.1 Workload

When engaging in a cognitive task, the active information processing requires cognitive resources which are limited for mental operations (Kahneman 1973). With increasing task difficulty, more resources are needed, so the operator’s remaining capacity diminishes. High mental workload will occur when the performance of a task needs the majority of the available resources, while low workload indicates unused capacities. A distinction must be made between task load and workload. Task load is the external stress or demand through the task or the system, e.g. complexity or time pressure. Mental workload, in contrast, is the resulting strain or impact of these stressors on the person, depending on their individual constitution, resources, abilities, etc. (International Organization for Standardization 1991). The relationship between task load and workload is not straightforward, as a change in task load can lead to different levels of perceived workload in different individuals (Kerkau 2005). However, in ATC, the number of aircraft in the control traffic region (CTR) is, in experiments, frequently used to manipulate task load, which has been shown to affect mental workload: The more aircraft’s to control, the higher the mental workload (Ahlstrom and Friedman-Berg 2006; Brookings et al. 1996; Lee et al. 2012).

Young et al. (2015) provide an overview on studies investigating the relationship between workload and performance. In total, there is strong evidence that high workload leads to poor performance and a higher number of errors. It can be assumed that “if demands begin to exceed capacity, skilled operators can either adjust their strategy to compensate or else performance necessarily degrades” (Young et al. 2015).

2.2 Situation Awareness

Endsley (1995b) defines SA as a hierarchical construct comprising the perception of elements in the environment together with their states and attributes, the understanding of these elements and their significance, and the projection of future states of the elements and the situation. Durso et al. (1998) consider SA as a process which is not necessarily present in consciousness but acting implicitly: “SA may sometimes involve simply knowing where in the environment to find a particular piece of information, rather than remembering what that piece of information is”. Chiappe et al. (2012) extend this idea by stating that individuals tend to rely on external props instead of forming a detailed mental representation: This spares the limited capacity of working memory and attention. Thus, information is not only stored internally in memory but operators also make use of external storage since “it is not necessary to have all the information stored in the mind to keep track of and perform well in complex situations” (Kraemer and Süß 2015).

SA is an often-investigated factor in high-risk-environments such as the medical operating room (Shelton et al. 2012), submarine track management (Loft et al. 2015), or nuclear power plants (Yang et al. 2012). There are fields of study that indicate age and experience have an influence on SA (e.g. Bolstad 2001; Kass et al. 2007; Patten et al. 2006), that makes it necessary to analyze the sample in accordance with those results. Especially considering the cascade model (Fry and Hale 1996; Fry and Hale 2000; Gregory et al. 2009) that tries to explain this influence and also indicates that continuous training reduces the influence of age on SA. Yet, there are critical views on the concept of SA, doubting its usefulness as a scientific construct (Dekker 2015; Dekker and Woods 2002). However, several researchers address this criticism and conclude that the outlines of SA have been thoroughly drawn over the past years of research (Endsley 2015; Parasuraman et al. 2008; Vidulich 2000). Some literature defines what SA is, by which means it can be measured, which constructs are linked to it, and which are not. It is true that the concept of SA may also be tangible for non-scientists, as interviews with ATCOs show (Edwards 2013). The authors consider SA as a necessary and measurable construct for understanding human computer interaction, especially in relation to workload.

There are three approaches to measuring SA: (1) subjective measures in terms of self-rating or assessment by experts, (2) implicit performance measures, and (3) probe techniques (Bacon and Strybel 2013; Durso et al. 1998). Because SA is linked to a mental picture only accessible by the particular person, the first approach, self-rating, is the one most often used. However, a difficulty is “that people are not always aware of what they don’t know” (Jeannot 2000). So self-rating leads to a statement of how certain the person feels about his or her SA (Endsley 1995a), rather than to a valid estimation of the real degree of SA. Still, some subjective rating tools for SA are often used, such as the 3D SART (Situational Awareness Rating Technique, Taylor 1990), or for the context of ATC, the SASHA (Dehn 2008) which was derived from SASHA_Q (Jeannot et al. 2003).

The second approach, assessing implicit performance measures focusing on observable actions and errors, may be recommendable, especially when considering their non-intrusiveness (Jeannot 2000). The difficulty is finding the adequate performance indicator, because the relation between SA and performance is not entirely clear (Salmon et al. 2009). Additionally, the resulting performance is not the product of SA alone but several other influences (Endsley 1995a).

The third approach is the probe technique, which consists of asking the subject questions about the current situation (Salmon et al. 2009). Probe measures can be divided into freeze techniques and online techniques. Freeze techniques interrupt the current task and blank the displays. As all sources of information are removed, the subject has to recall information from memory to answer questions about the situation, e.g. Situation Awareness Global Assessment Tool (Endsley 1995a). In contrast to that, online techniques pose questions about the situation without interruption, e.g. SPAM (Durso et al. 1998). Compared to freeze techniques, online techniques possess a less intrusive nature and allow for a more realistic performance in simulation settings (Salmon et al. 2009) but are, however, not applicable for operational environments. Based on the SPAM method, Kraemer and Süß (2015) developed SARA-T which analyses the log files of simulation in real-time, allowing for an actual real-time assessment of SA. However, online probes have the disadvantage of representing extra workload for the operator, who has to deal with a further task demand (Jeannot 2000). Salmon et al. (2009) emphasize the difficulty of proposing a consistent measurement method, since several concepts of SA exist which are often linked to a certain suggestion of how to measure it. In consequence, selecting the best-fitting procedure depends on the particular testing assumptions and preconditions. Given the advantages and disadvantages of each approach, the best strategy is to combine different approaches.

Since the 1990s, many studies on SA in the context of ATC have been conducted, focusing on different aspects of the concept of SA or its measurement (Jeannot 2000). Endsley and Rodgers (1996) investigated the attention distribution of ATCOs and concluded that tradeoffs must be made in deciding which elements are taken into account and in which detail. They traced the observed difficulties in identifying loss of separation between aircraft back to difficulties in passive monitoring. In his analysis of FAA reports on occupational error, Redding (1992) identified lacking maintenance of SA as the most probable cause for most errors. In more recent studies, SA is frequently investigated in the context of task and system design in light of the future increase in air traffic (Durso and Sethumadhavan 2008). In connection to tower control operations, findings by Teutsch and Postma-Kurlanc (2014) indicate that sight and communication (Corradini and Cacciari 2002) are an important factor for SA.

2.3 Control Strategies

The ATC workplace allows for different strategies to compensate for increased workload (Sperandio 1971; Sperandio 1978), e.g. by decreasing the amount of time spent on one single aircraft. According to Karikawa et al. (2014), there are several possible control strategies for a certain traffic situation in en route ATC that differ in safety, efficiency, or fuel economy and thus lead to different performance. They suggest that tolerance for the variability of situations could be a main factor when selecting a strategy. Edwards (2013) also stressed the significance of management strategies in ATC: “Controlling aircraft is a creative process, and controllers differ on styles of control strategy”. In her study, she suggested certain compensation strategies used by controllers which also serve as performance indicators. Facing the extremes of workload or situation awareness, different compensation strategies are possible. Papenfuss and Friedrich (2016) used eye tracking to identify the different control strategies depending on the dwell duration time participants spent on a specific source of information. They showed that different sets of workplace configuration lead to a change in the information-gathering process and subsequently in control strategy.

2.4 Relation of Workload, Situation Awareness, Control Strategies, and Performance

Processing always appears within a certain situation, which renders knowledge of the situation or mental models – and by this also SA – crucial for perception and action (Hendy 1995). Concerning the relationship of workload and SA, Hendy (1995) states that good SA can reduce the processing capacity of future decisions and thus time pressure. At the same time, though, good SA requires capacity for its formation in the time span until a decision is made. According to Lee et al. (2012), workload and SA are differently linked to performance and “these two constructs are more closely related at higher task load conditions than at lower ones.“ (p. 349). Therefore, before analyzing the influence of workload on SA and performance, different behavior in different demanding situations has to be taken into account.

The current state of research on workload, SA, control strategy, and performance indicates that none of the constructs can be measured and analyzed on its own (Wickens et al. 1997). Figure 1 presents the connections between the different constructs, beginning with the dynamic environment and its influence on workload and the selected control strategy. Workload and control strategy influence each other directly, whereas Situation Awareness can be affected if the strategy demands change. Research also suggests that the interaction is stronger when the dynamic environment generates high workload (Flin et al. 2003). If workload reaches a certain level, the control strategy is changed to keep workload on a manageable level.

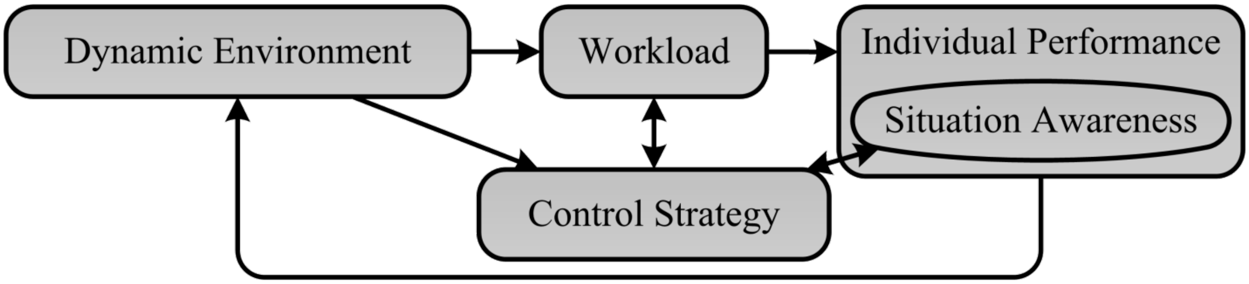

Figure 1 Connection between workload, SA, control strategy, and performance (Wickens et al. 1997)

Looking at the challenges for future tower control operations, the relation between workload, SA, control strategy, and performance still needs to be defined in more detail. Within this article, mental workload resulting from external task load will be used to influence the SA and the control strategy within the ATC tower environment. Knowledge of their influence as performance-shaping factors is necessary to evaluate new design concepts in early stages of their development. This article supports evaluation of how ATCOs interact with their environment by focusing on two aspects: whether probe techniques for SA testing are applicable for tower control operation and if it is possible to measure the influences of increased task load on the control strategy.

3 METHOD

3.1 Sample

Participants (N = 9) were recruited from the tower Braunschweig-Wolfsburg and via an internal German Aerospace Center mailing list. They were all male, with an average age of 44.22 years (SD = 11.29), ranging from 28 to 62 years. The participants' work experience ranged from 3 to 33 years with a mean of 18.56 years (SD = 10.36). One participant had been retired for 4 years; another had been in partial retirement for 6 months. Each participant obtained 100 € and compensation for travel expenses. The ethical approval was granted by the Braunschweig University of Technology.

3.2 Design

For this study, the working position of an ATCO was simulated in a high-fidelity setting using a within-subject-design 2x2 with repeated measures for one factor to investigate the dependent variables SA, eye movement, and performance under a) varying task load conditions (low vs. high; repeated measures) and b) different visual conditions (high (A) vs. low (B) visibility). The conditions were implemented in two scenarios (Table 1), each consisting of the same sequences of altering low and high task load phases (Figure 2) and differing in visibility. Both scenarios began with low task load to provide a simple start into the simulation. Each scenario was designed to take 40 minutes. In the phases of low task load, only one moving aircraft was present, while the demand increased in the phases of high task load to a maximum of 8 aircraft. The same task load phases (low vs. high) were designed to contain approximately the same amount and complexity of air traffic, without repeating exactly the same traffic flow to avoid training effects. Phases of high workload consisted of 5 moving aircraft and 3 additional requests for take-off. Traffic patterns used in this study were incoming scheduled flights, departing unscheduled flights, crossing unscheduled flights, and unscheduled flights following the traffic circuit. The scenarios were previously tested for feasibility by a former ATCO.

Table 1 Scenario overview

		Scenario A	Scenario B
a)	Task load conditions	Equal for each scenario, following Figure 2	
b)	Visibility conditions	high	low

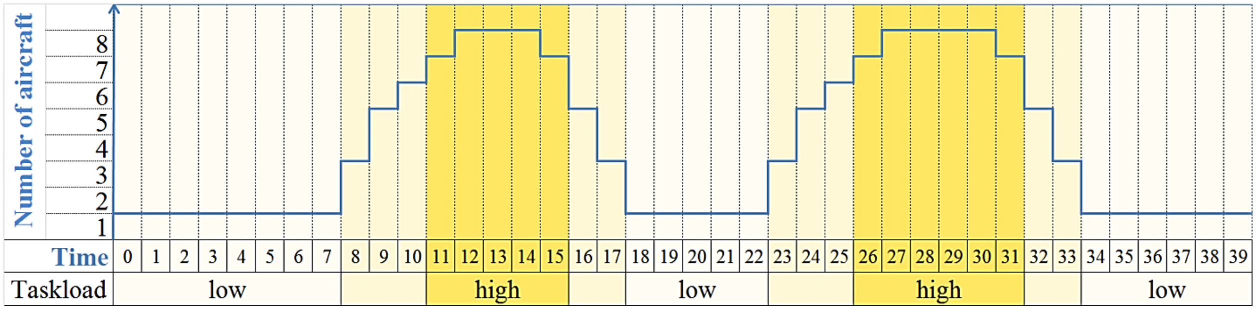

Figure 2 Overview of the scenarios in relation to task load

3.2.1 Apparatus

Figure 3 shows the experimental set-up installed in the Tower Lab of the Institute of Flight Guidance at the German Aerospace Center. The participants were seated in the center of the workspace, wearing headset and eye-tracking glasses. Five 40-inch LC displays arranged in a semi-cycle formed a compressed 180° outside view of the tower (Video Panorama). The displays in front of the participant presented Radar, Radio, digitally depicted flight strips (Flight-Strips), and Weather and Time (WaT) information. Data entry into the flight strip system was possible via mouse and keyboard. Radio communication took place with two pseudo-pilots managing the scenario from the room next door. The SARA-T probe technique was presented on a separate display.

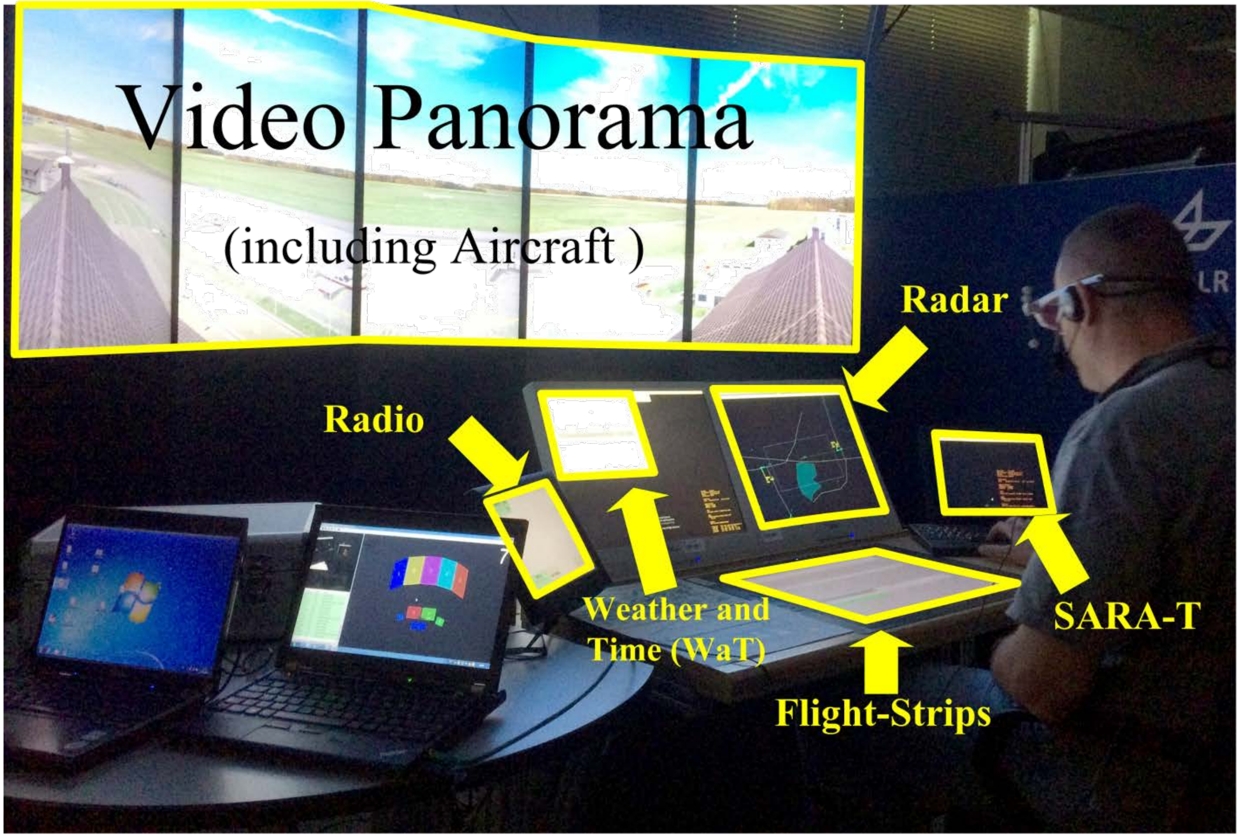

Figure 3 Tower Lab: experimental setting and AoIs

3.2.2 Performance measurement

Table 2 presents the selected performance indicators that are directly recorded by the simulation to establish a high degree of reliability and objectivity. The measurements in Table 2 were collected for each run.

Table 2 Performance measures used in the study

Performance Category	Measure	Description
Safety	Number of separation losses	A prescribed safety distance between aircraft (or aircraft and objects) must always be maintained to avoid conflict situations.
Efficiency	Number of take-offs	The scenario requires several clearances for take-off through the ATCO. The number of conducted take-offs therefore reflects the capability of the ATCO to efficiently regulate outgoing traffic.
	Aircraft taxi time	The taxi time for parking aircraft to the runway and from landing aircraft to the gate varies depending on the control style of the particular ATCO.
	Duration of radio communication	The ATCO has to communicate with the pilots via radio, giving precise instructions in a time-efficient manner to minimize the usage of the radio frequency.

3.2.3 Situation Awareness Measurement

For measuring SA, two methods were used: self-rating questionnaires after each run and probe technique during runs. In the first method, SA was evaluated through the SASHA (Dehn 2008) consisting of six questions (Table 3) about the previous workload period(s). Answers are given on a seven-point scale ranging from “never” to “always” (item values 0 to 6). The SASHA score for SA is the total sum of the single item values, inverting the scores of items 2, 3, 5, and 6, and dividing it by six. Therefore, the SASHA score ranges between 0 (low SA) and 6 (high SA).

Table 3 SASHA items with information about the scale inverting (Dehn 2008).

Item ID	Items	Scale inverted
	In the previous working period(s) …	
1	... I was ahead of the traffic.	
2	... I started to focus on a single problem or a specific area of the sector.	*
3	... there was a risk of forgetting something important (like transferring an a/c on time or communicating a change to an adjacent sector).	*

4	... I was able to plan and organise my work as I wanted.	
5	... I was surprised by an event I did not expect (like an a/c call).	*
6	... I had to search for an item of information.	*

For the second approach, the probe technique SARA-T was chosen. Derived from the Tower Traffic Management goal hierarchy by Strater et al. (2010) and by SA requirements described in Jeannot (2000), items for SA were selected. In cooperation with subject matter experts, the items were refined for tower operations. In a three-minute interval, an item is randomly selected and announced as “Incoming questionnaire” on the SARA-T display. The questions covered perception (e.g. “Please estimate the number of in-air flights.”) as well as comprehension (e.g. “Please state the call sign of the last aircraft that landed on the runway.”). While the item was being presented, the log file of the simulation was analyzed and the correct answer was made available in real-time. The correct answers and 3 distractors were presented as multiple-choice test. If the item was answered correctly, the response time was used as measurement for SA (Chiappe et al. 2015).

3.2.4 Eye tracking

An eye-tracking system was used to identify the source of information uses to interact with the ATC workplace. Eye data was collected with the SMI Eye Tracking Glasses 2 Wireless (SensoMotoric Instruments, Germany). This head-mounted system collects eye direction, gaze, and head position with a sampling rate of 60 Hz binocular, a gaze-tracking accuracy of 0.5° over all distances, and a gaze-tracking range of 80° horizontal and 60° vertical. A one-point calibration was carried out to make sure that the eyes were correctly tracked by the iView System. The Integration Guideline for Dynamic Areas of Interest (IGDAI) was applied to ensure the capturing of eye data and objects (dynamic and static) with a synchronized time stamp and in the same coordination system (Friedrich et al. 2016). Due to the problem of the clear distinction between fixation, saccades, and smooth pursuits in dynamic task environments (Papenmeier and Huff 2010), only the raw eye data dwell time on a defined area of interest (AoI) was used. The statistical software EyeTrackingAnalyser by DLR (Friedrich et al. 2016) was used to analyze the eye-tracking data. For every run, the areas of interest were used to classify the eye movement. Due to the application of IGDAI, dynamic areas of interest for the moving objects (aircraft) presented on the video panorama were defined and automatically classified. The sources of information were used to identify possible control strategies that interact with SA and task load.

3.3 Procedure

The participants were welcomed and the experimenter explained the procedure of the experiment. The participants were advised to perform ATC in their usual way. It was emphasized that the participation was voluntary and could be abandoned at any point. The participants signed a consent form in which the anonymity of data collection and analysis was guaranteed. The demographic information was collected via a preliminary questionnaire. The participants were then trained on the electronic flight strip system by creating flight strips, changing their status and deleting them. The eye-tracking glasses were put on and calibrated. The radio headset was adjusted and checked for proper function.

The participants were randomly assigned to begin with scenario A or B. 40 minutes into the run, the simulation was stopped. SARA-T was performed in parallel throughout the run. Immediately afterwards, the SASHA was presented on an additional laptop. Before the second run, the participants took a 5-10-minute break. After another 40 minutes, the second run was stopped and SASHA was answered once more. At the end, the final questionnaire was answered concerning the feasibility of the SA probe question technique, followed by a debriefing.

4 RESULTS

The data from the first low task load phase (Figure 1) was removed from the analysis, because the participants used the first phase as training to get acquainted with the task environment and to equalize the total duration of high and low task load phases. Also, two phases of low task load from participant 6 were removed, due to a mistake in simulation procedure resulting in a high task load during these phases.

4.1 Performance

The performance metrics (Table 2) were classified in relation to safety and efficiency. The safety-critical performance separation loss happened twice (number of separation losses) throughout the study and was caused by the pseudo-pilots. Therefore, the safety performance was the same in all conditions. The number of take-offs was analyzed for scenarios A (M = 8.0, SD = 1.32) and B (M = 8.67, SD = 0.71). Table 4 (aircraft taxi time) and Table 5 (duration of radio communication) show the descriptive efficiency performance data separated by scenario and task load. Due to technical problems, radio communication times are not available for participant 7, scenario B and participant 9, scenario A. The descriptive performance data indicate that neither visibility nor task load had an influence on safety or efficiency.

Table 4 Average taxi time in seconds, separated by scenario and task load phase.

Scenario	Task load	Average taxi time (M)	SD	MIN	MAX
A	high	269.81	95.65	104.44	559.37
A	low	260.25	89.3	130.00	559.37
B	high	266.82	69.95	153.38	435.57
B	low	238.7	58.31	132.00	362.00

Table 5 Average radio duration in seconds for the participants separated by scenario and task load phase.

Scenario	Task load	Average radio communication (M)	SD	MIN	MAX
A	high	5.04	1.18	3.41	7.28
A	low	4.15	0.76	3.06	6.25
B	high	4.45	0.96	3.28	6.69
B	low	3.96	0.79	2.59	5.50

4.2 Effects on Situation Awareness

As measurements for SA, SARA-T captures the response time for an answer to a probe as well as the correctness of it. The SARA-T response times and correctness of answers were aggregated for participants, separated by scenarios and task load phases. The response times are only considered valid if the probe answer was correct. In connection to previous studies (e.g. Kass et al. 2007; Patten et al. 2006) that indicate a connection between SA, age, and experience, the first analysis presented the correlations between these factors. Figure 4 presents the results regarding the SARA-T response times and age (left) and experience (right). The same analysis was repeated using the SASHA scores instead of the SARA-T response times, but using the Kendall’s rank test none of these correlations were significant ($.244 < p < .923$).

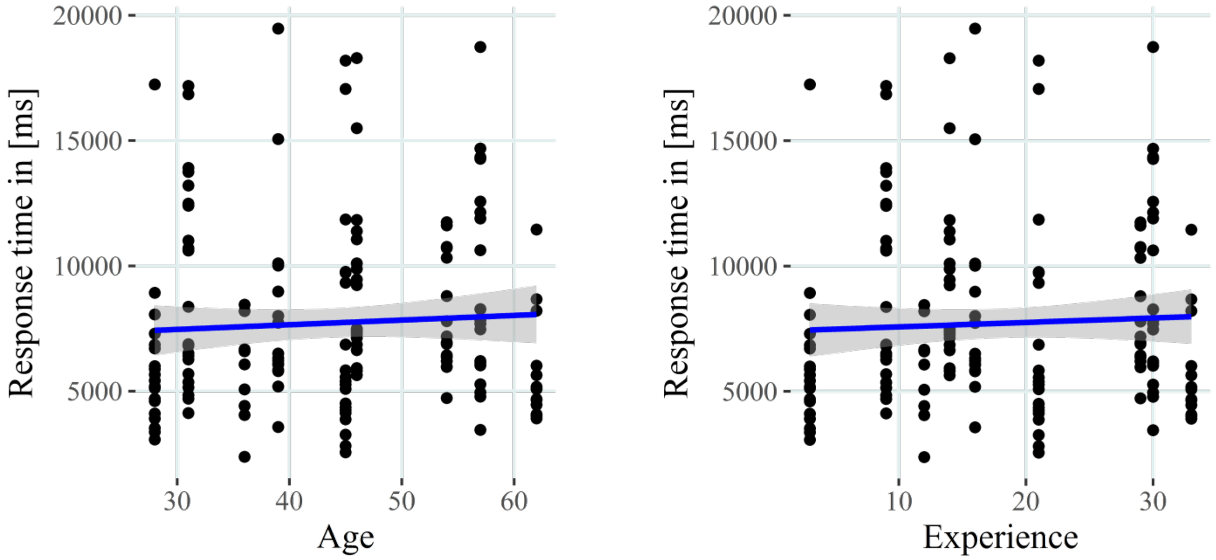

Figure 4 Correlation between SARA-T reaction time and age (left) and experience (right) for all task load phases.

Figure 5 presents the descriptive results separated by the scenario with high (A) and low (B) visibility. Although the total average of correct answers increases (Figure 5, left) from scenario A (M = 0.58, SD = 0.40) to scenario B (M = 0.80, SD = 0.25), none of the pairwise-conducted Wilcoxon Signed-Ranks test scores were statistically significantly different (.089 < p < .674). Wilcoxon Signed-Ranks test was also conducted for the response times (Figure 5, right), which also led to no statistically significantly different scores (.098 < p < .726).

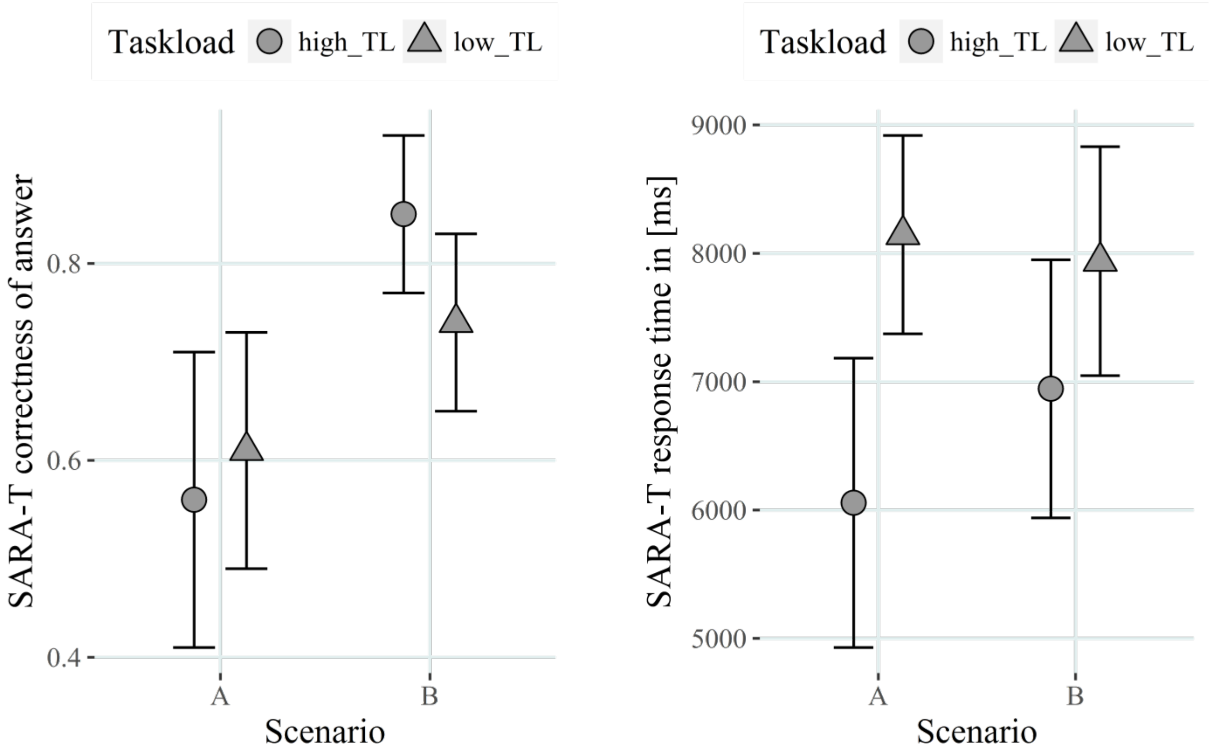

Figure 5 Answers within the scenarios (high (A) vs. low (B) visibility) and per task load phase (left, with an N=9 for each mean value) and the response times within the scenarios (high (A) vs. low (B) visibility) and per task load phase (right, with an N=9 for each mean value). The error bars in both figures represent the standard error

A Wilcoxon Signed-Ranks Test indicated that SASHA scores did not differ significantly for each scenario (Z=6; p < .087). However, a tendency can be observed: six subjects rated their SA via SASHA higher under low visibility conditions in scenario B (M = 4.37, SD = 0.57) than under high visibility in scenario A (M = 3.95, SD = 0.67). This is also supported by the SARA-T average correctness of answers.

In order to verify if there is a connection between objective probe data of particular task load phases and the retrospective self-assessment, the valid SARA-T response times were averaged per task load phase and in total and correlated separately with the SASHA scores. If the reaction times increase, the SASAH score should indicate less situation awareness. Figure 6 presents the correlation for SARA-T response times and SASHA scores in high task load (above left), low task load (above right), and in total (below). Based on the results of the Kendall’s tau test, only the response times within the high task load phases show a significant correlation with the SASHA score (r_s = -0.400, p = .08).

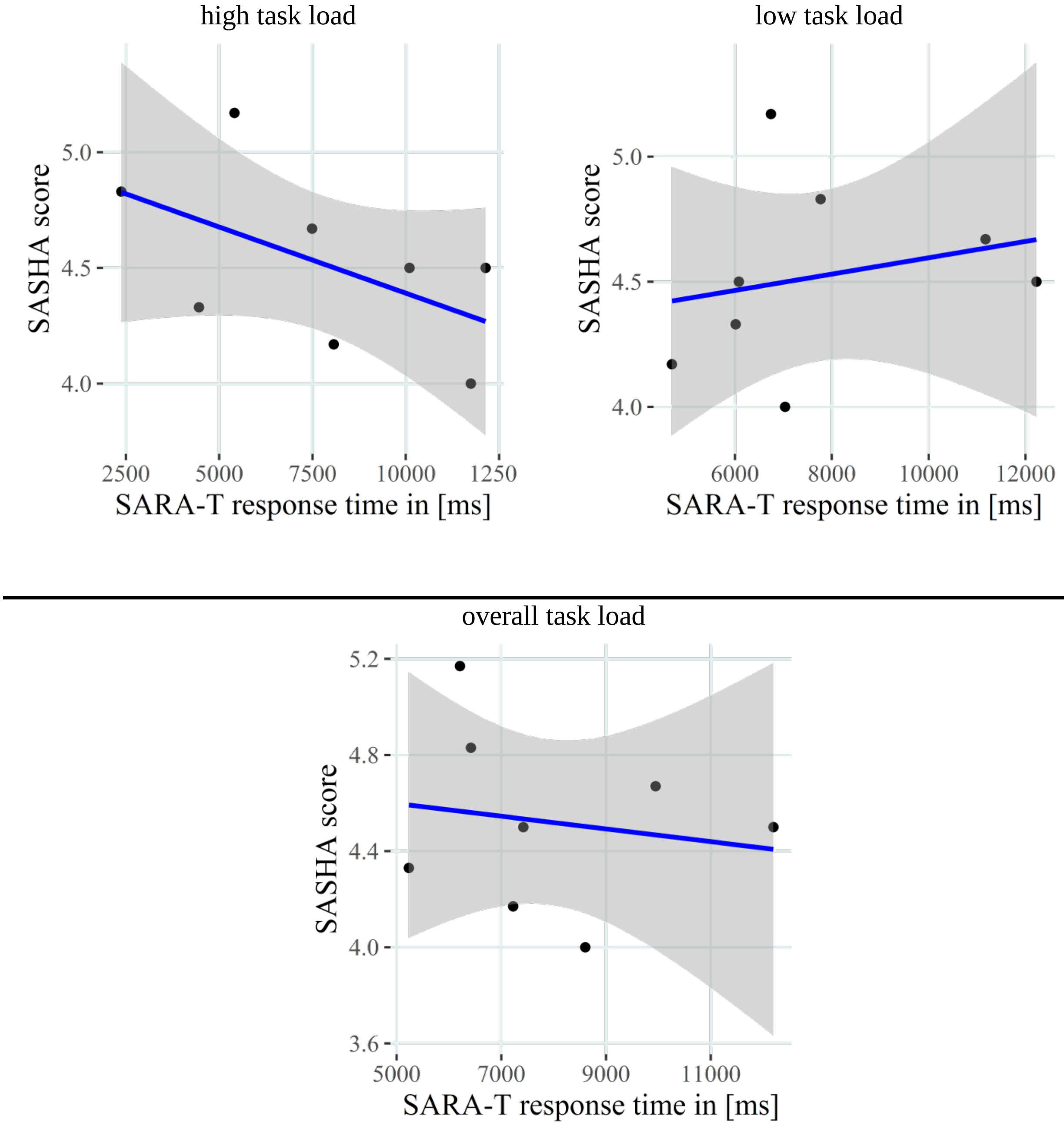

Figure 6 Correlation between SASHA and SARA-T duration for high (top left) task load, low (top right) task load, and the overall duration for scenario B

Within the debriefing, participants stated that the SARA-T questions were moderately not disruptive (M = 2.33, SD = 0.50; 1: not disruptive to 4: disruptive). The majority of the participants deemed the SARA-T questions moderately not appropriate for assessing their situation awareness (M = 2.56, SD = 0.53; 1: appropriate to 4: not appropriate). They indicated that they felt additionally challenged by the questions (M = 2.22, SD = 0.67) and that the questions were more distracting in phases with more than four aircraft (M = 2.44, SD = 1.01; 1: distractive to 4: not distractive) than in phases with less than four aircraft (M = 3.33, SD = 0.87).

4.3 Effects on Control Strategy

The identification of the control strategy was based on the change in the information-gathering process that is measured by eye tracking. Eye movement is valid if the eye tracker detects the eyes and the gaze vector hits an area of interest (AoI, Figure 3). The amount of valid eye movement (valid eye point divided by total number of captured eye point) per participant was calculated and compared. The average of eye-data validity reached M = 72.6% (SD = 15.6), with only two participants (4 and 7) below 50%. Both participants wore glasses underneath the eye-tracking device that reduced the amount of captured eye data, but their general distribution of eye tracking data showed no sign of outliers and their data was therefore considered within this analysis. The average dwell time for all participants was spent on the AoI Flight-Strips (40%), Radar (20%), Video Panorama (17%), Aircraft (11%), WaT (5%), SARA-T (6%), and Radio (1%). For this analysis, the eye data was separated by scenario. A first inspection of data (Table 6) revealed that results from both scenarios do not differ significantly; only the results for scenario B (low visibility; highest average of correct SARA-T answers) are presented in detail. The eye data was categorized according to task load phases to measure their influence on the control strategy. Figure 7 shows the dwell time per AoI within high and low task load. The values represent the percentage of dwell time per AoI, and the error bars show the standard error. Summarizing the values for all AoIs results in 100% and therefore represents the complete gathering of information within the task load phases. Each participant looked at all AoIs at least once per run and task load phase, except for Radio (only 2 participants looked at it).

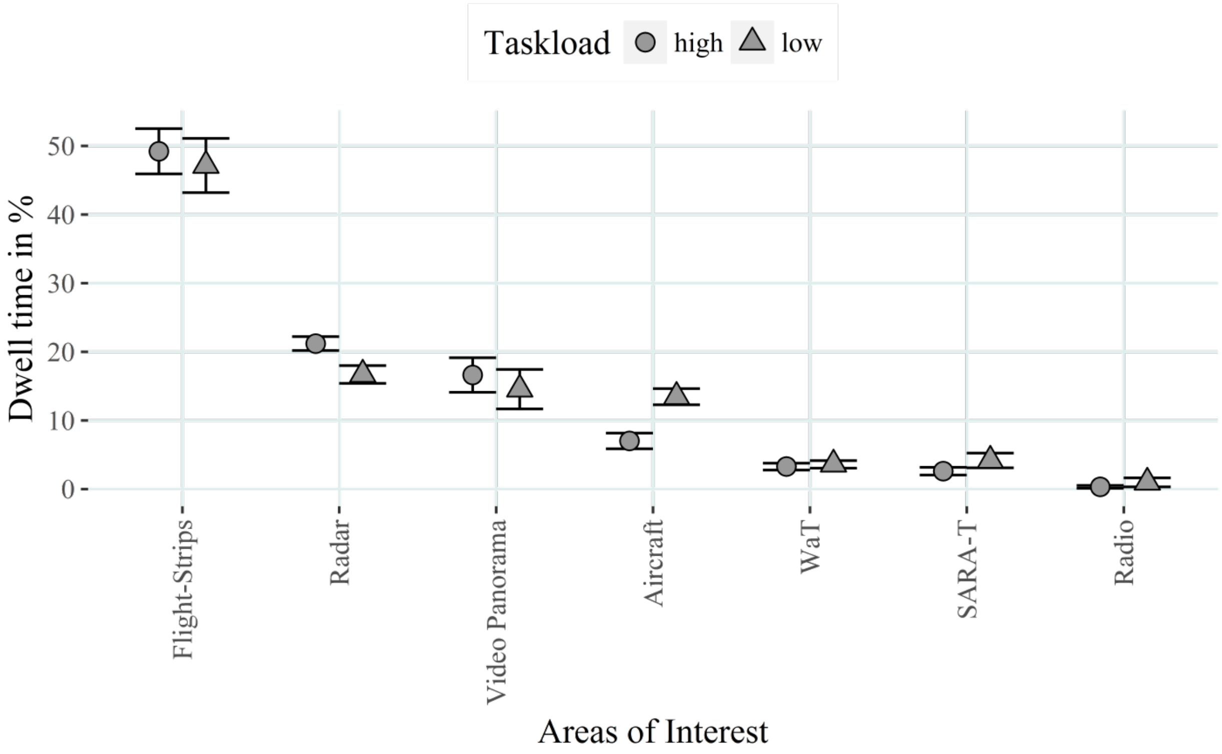

Figure 7 Dwell times for the AoIs split by task load phase for scenario B. The error bars represent the standard error

Table 6 shows the results of the Wilcoxon Signed-Ranks Tests performed for each AoI to identify significant differences between the task load phases, separated by scenario. These differences represent a change in control strategy and were identified for the AoIs Radar and Aircraft. If the task load is high, significantly more visual attention has to be directed to the Radar and less to the aircraft. This is reversed if the task load is low. No other AoI showed a significant difference depending on task load.

Table 6 F-Test results separated for each AoI and task load testing for percentage of dwell times.

AoI	Wilcoxon Signed-Ranks Tests (Scenario A)	p<0.05	Wilcoxon Signed-Ranks Tests (Scenario B)	p<0.05
Flight-Strips	V=34; p=.203		V=29; p=.499	
Radar	V=40; p=.039	*	V=40; p=.039	*
Video Panorama	V=1.88; p=.212		V=35; p=.164	
Aircraft	V=0; p=.004	*	V=0; p=.004	*
WaT	V=11; p=.203		V=18; p=.652	
SARA-T	V=7; p=.07		V=16; p=.496	
Radio	V=1; p=1		V=1; p=1	

A general criticism concerning the eye-tracking glasses was observed within the debriefing. Participants stated that wearing the eye-tracking glasses was an interference factor (M = 2.78, SD = 0.97; 1: not disturbing, to 4: disturbing).

5 DISCUSSION

Even so, in general, the sample size of this study reduces the explanatory power; the sample of nine is average for a study with ATCOs that work in the control tower (Kuk et al. 1999; Metzger and Parasuraman 2005; Teutsch and Postma-Kurlanc 2014). Interesting indications were found for the application of probe techniques for SA and the measurement of the influence of increased task load on the control strategy. The focus of the experiment was set on the application of probe techniques for SA and the detection of change in control strategy in different task load phases. Therefore, the authors used a high-fidelity simulation as experimental set-up and a sample of nine ATCOs.

The manipulation of high task load for the tower ATC is more complex than for the en route or approach control, because the ATCOs can delay take-offs if they are too busy. On the contrary, with a maximum of 6 minutes in duration, the phases of low task load were probably too short to create intense monotony or boredom. However, as the performance data showed no difference between scenario and task load phases, it is indicated that the individual performance stayed at the same level and that no overload was reached that might have impaired the results.

Concerning the influence of age and experience on SA (Kass et al. 2007; Patten et al. 2006), no correlation was found between either. This indicates that, as predicted by the cascade model (Fry and Hale 2000; Gregory et al. 2009), constant repetition in connection to a task could decrease the effects of age and experience. It is worth mentioning that ATC is a training-intensive (3 to 4 years) profession that should guarantee the same safety and performance from each ATCO.

According to Hendy (1995), high task load at acceptable levels and good SA are linked because both involve the same internal processes. This is supported by the correlation between SARA-T and SASHA at high task load phases: when asked for a self-assessment of their SA during the previous task, participants in this study tended to refer to phases of high task load rather than to low task load, similar to Gontar and Hoermann (2015). It also indicates that retrospective self-assessment of SA is linked to the situations where a good SA is needed the most to successfully interact with the task environment. Taking into account that both performance and SA measures show no significant differences for high and low task load phases, SA itself may be an aspect of performance rather than a factor leading to it. For example, SA is considered as a cognitive skill in the NOTECHS rating scale for the assessment of pilots' Crew Resource Management skills (Flin et al. 2003). The question therefore remains as to whether SA is a factor influencing performance or if SA is the outcome of a task, thus performance itself. Considering the great number of theories, articles, and discussions on SA (e.g. Edwards 2013; Flin et al. 2003; Hendy 1995), this question cannot be answered unequivocally. After all, it may be a matter of definition.

Since SARA-T was classified as not disruptive during the debriefing, the extra workload for the operator, as proposed by Jeannot (2000), was kept to an absolute minimum. However, participants of this study considered the probe questions rather not appropriate for assessing their SA, as the debriefing showed. On the one hand, this could indicate that the tower ATCOs had a different, more holistic understanding of SA which cannot be broken down to single probe questions on concrete events during simulation. On the other hand, this calls for an improvement and validation of the probe questions used in the context of tower ATC. The SASHA tendency to contradict the assumption that reduced visibility impairs SA was supported by the analysis of SARA-T. Therefore, the assumption is supported by SASHA and SARA-T, even though both tests results are not significant. We therefore believe the results of the SARA-T are reliable and valid, which is more important than the appropriateness viewed by the participants.

In contrast to the findings of Karikawa et al. (2014), the performances between the task load phases did not differ. Furthermore, through the application of the probe technique, it was shown that the SARA-T values did also not vary significantly between the task load phases. By following the model of the interconnection (Figure 1), the control strategies should change over time. As for the en route (Karikawa et al. 2014) and the approach ATC (Sperandio 1971), the results showed the existence of different control strategies for the tower ATC environment. The analysis implies the existence of one control strategy for the high task load phases that concentrates on the radar information, and another one for low task load phases that focuses more on aircraft. Within the high task load phases, the information for managing the workplace is combined on the radar to bring all the aircraft in relative position to each other. Therefore, compared to the Video Panorama, the Radar offers the advantage of providing task-relevant information (callsign and position) at one glance. By decreasing the amount of tracking for one single aircraft, the high task load control strategies follow the same pattern as for the approach control workplace (Sperandio 1978). The low task load phases normally had one aircraft that allowed for a use of the Video Panorama to monitor aircraft more specifically. The authors believe that the change in information gathering is connected to the density of information at the radar display. As above, this also supports the model of the interconnection and even extends it by using the percentage of dwell time to show expected behavior.

6 CONCLUSION

In spite of the small sample size, some important indications were found, confirming the connection of the three factors task load, SA, and control strategy. The work presented in this article provided valuable insight into the human performance in tower ATC. The experimental design manipulated SA via visibility and task load via number of aircraft. The results showed that probe technique allows the online determination of SA within different task load phases. SARA-T is essential if the task load varies within a scenario, because the post-run self-assessment is not precise enough. Discriminating SA offline seems to be affected by the task load the participants have experienced during the runs. SARA-T is a helpful support for the analysis of SA, especially in a dynamic task environment such as the ATC environment. It provides a detailed view of the development of SA throughout the runs and helps in understanding its different states, depending on the situations.

Considering that performance and SA remained constant between the runs and the task load phases, the relevant factor had to be the switch between strategies. The analysis of the control strategy showed the influence of increased task load on the process of information gathering. The participants switched from monitoring the aircraft directly to monitoring the radar display when the task load increased.

Future research needs to focus further on interactions between different performance-shaping factors. For the development of a detailed model including these interconnections, overload situations should be used to identify ceiling effects. In this context, it would have been recommendable to create higher levels of task load by adding more aircraft to the scenarios. This is needed to analyze the connections between factors more closely. In addition, this could lead to the identification of a control strategy through which the amount of traffic is actively reduced by postponing or denying clearances.

REFERENCES

Ahlstrom U, Friedman-Berg FJ (2006) Using eye movement activity as a correlate of cognitive workload International Journal of Industrial Ergonomics 36:623-636

Ambroggi Md, Trucco P (2011) Modelling and assessment of dependent performance shaping factors through Analytic Network Process Reliability Engineering & System Safety 96:849-860

Bacon LP, Strybel TZ (2013) Assessment of the validity and intrusiveness of online-probe questions for situation awareness in a simulated air-traffic-management task with student air-traffic controllers Safety science 56:89-95

Baumann M, Krems JF (2007) Situation awareness and driving: A cognitive model. In: Modelling driver behaviour in automotive environments. Springer, pp 253-265

Bolstad CA Situation awareness: does it change with age? In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2001. vol 4. SAGE Publications Sage CA: Los Angeles, CA, pp 272-276

Brookings JB, Wilson GF, Swain CR (1996) Psychophysiological responses to changes in workload during simulated air traffic control Biological Psychology 42:361-377 doi:http://dx.doi.org/10.1016/0301-0511(95)05167-8

Chiappe D, Strybel TZ, Vu K-PL (2015) A Situated Approach to the Understanding of Dynamic Situations Journal of Cognitive Engineering and Decision Making 9:33-43 doi:10.1177/1555343414559053

Chiappe DL, Strybel TZ, Vu K-PL (2012) Mechanisms for the acquisition of situation awareness in situated agents Theoretical Issues in Ergonomics Science 13:625-647

Corradini P, Cacciari C (2002) The effect of workload and workshift on air traffic control: a taxonomy of communicative problems Cognition, technology & work 4:229-239

Cox-Fuenzalida L-E (2007) Effect of workload history on task performance. Hum Factors 49:277-291

Dehn DM (2008) Assessing the Impact of Automation on the Air Traffic Controller: The SHAPE Questionnaires Air Traffic Control Quarterly 16:127-146 doi:10.2514/atcq.16.2.127

Dekker SW (2015) The danger of losing situation awareness Cognition, Technology & Work 17:159-161

Dekker SW, Woods DD (2002) MABA-MABA or abracadabra? Progress on human–automation co-ordination Cognition, Technology & Work 4:240-244

Durso FT, Hackworth CA, Truitt TR, Crutchfield J, Nikolic D, Manning CA (1998) Situation awareness as a predictor of performance for en route air traffic controllers Air Traffic Control Quarterly 6:1-20

Durso FT, Sethumadhavan A (2008) Situation awareness: Understanding dynamic environments Hum Factors 50:442-448 doi:10.1518/001872008x288448

Edwards T (2013) Human performance in air traffic control., Dissertation, University of Nottingham

Endsley MR (1995a) Measurement of Situation Awareness in Dynamic Systems Human Factors: The Journal of the Human Factors and Ergonomics Society 37:65-84

Endsley MR (1995b) Towards a Theory of Situation Awareness in Dynamic Systems. Hum Factors 37(1):32-64

Endsley MR (2000) Theoretical underpinnings of situation awareness: A critical review. In: Endsley MR, Garland D (eds) Situation awareness analysis and measurement. Lawrence Erlbaum Associates, Mahwah, NJ, p 400

Endsley MR (2015) Situation awareness: operationally necessary and scientifically grounded Cognition, Technology & Work 17:163-167

Endsley MR, Rodgers MD Attention distribution and situation awareness in air traffic control. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 1996. vol 2. SAGE Publications Sage CA: Los Angeles, CA, pp 82-85

EUROCONTROL/FAA (2010) Human Performance in Air Traffic Management Safety: A White Paper. Action Plan 15 Safety

Flin R, Martin L, Goeters K-M, Hormann H, Amalberti R, Valot C, Nijhuis H (2003) Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills Human Factors and Aerospace Safety 3:97-120

Friedrich M, Rußwinkel N, Möhlenbrink C (2016) A guideline for integrating dynamic areas of interests in existing set-up for capturing eye movement: Looking at moving aircraft Behavior Research Methods 49 doi:10.3758/s13428-016-0745-x

Fry AF, Hale S (1996) Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade Psychological science 7:237-241

Fry AF, Hale S (2000) Relationships among processing speed, working memory, and fluid intelligence in children Biological psychology 54:1-34

Fürstenau N (2016) Virtual and Remote Control Tower: Research, Design, Development and Validation:337

Fürstenau N, Friedrich M, Mittendorf M, Schmidt M, Rudolph M (2013) Discriminability of flight maneuvers and risk of false decisions derived from dual choice decision errors in a videopanorama-based remote tower work position. vol 8020. Springer Berlin Heidelberg,

Gontar P, Hoermann HJ (2015) Interrater Reliability at the Top End: Measures of Pilots' Nontechnical Performance International Journal of Aviation Psychology 25:171-190 doi:10.1080/10508414.2015.1162636

Gregory T, Nettelbeck T, Howard S, Wilson C (2009) A test of the cascade model in the elderly Personality and Individual Differences 46:71-73 doi:https://doi.org/10.1016/j.paid.2008.08.017

Griffin M, Neal A, Neale M (2000) The contribution of task performance and contextual performance to effectiveness: Investigating the role of situational constraints Applied Psychology 49:517-533 doi:10.1111/1464-0597.00029

Hadley GA, Guttman JA, Stringer PG (1999) Air traffic control specialist performance measurement database

Hendy KC (1995) Situation awareness and workload: birds of a feather? DTIC Document,

International Organization for Standardization (1991) Ergonomic principles related to mental work-load vol ISO 10075:1991.

Jeannot E (2000) Situation awareness, synthesis of literature research EEC note

Jeannot E, Kelly C, Thompson D The Development of Situation Awareness Measures in ATM Systems. In: (HRS/HSP-005-REP-, 01) (eds), Brussels, Belgium, 2003. EUROCONTROL,

Kahneman D (1973) Attention and effort. Citeseer,

Karikawa D, Aoyama H, Takahashi M, Furuta K, Ishibashi A, Kitamura M (2014) Analysis of the performance characteristics of controllers’ strategies in en route air traffic control tasks Cognition, technology & work 16:389-403

Kass SJ, Cole KS, Stanny CJ (2007) Effects of distraction and experience on situation awareness and simulated driving Transportation Research Part F: Traffic Psychology and Behaviour 10:321-329 doi:10.1016/j.trf.2006.12.002

Kerkau F (2005) Biosignale der Pupille zur Steuerung intelligenter User-Interfaces: Untersuchung von Pupillenbewegungen zur Realisierung einer biopsychologischen Computerschnittstelle für die Mensch-Computer-Interaktion. Freie Universität Berlin

Kirwan B, Bettignies-Thiebaux B, Scholte J OPTICS: EXPERT WORKSHOP: HUMAN FACTORS. In: 1st Expert Workshop: Human Factors, Brussel, 2014.

Kontogiannis T, Malakis S (2013) Strategies in controlling, coordinating and adapting performance in air traffic control: modelling ‘loss of control’events Cognition, technology & work 15:153-169

Kraemer J, Süß H-M (2015) Real Time Validation of Online Situation Awareness Questionnaires in Simulated Approach Air Traffic Control. Procedia Manufacturing:3152-3159

Kuk G, Arnold M, Ritter FE (1999) Effects of light and heavy workload on air traffic tactical operations: a hazard rate model Ergonomics 42:1133-1148

Lange M (2014) Information Search for Decision-Making in Tower Air Traffic Control: A Field Study., Tu Chemnitz

Lee YH, Jeon J-D, Choi Y-C (2012) Air Traffic Controllers' Situation Awareness and Workload under Dynamic Air Traffic Situations Transportation Journal 51:338-352 doi:10.5325/transportationj.51.3.0338

Loft S, Bowden V, Braithwaite J, Morrell DB, Huf S, Durso FT (2015) Situation awareness measures for simulated submarine track management Hum Factors 57:298-310

Mensen H (2014) Moderne Flugsicherung: Organisation, Verfahren, Technik. Springer-Verlag,

Metzger U, Parasuraman R (2005) Automation in Future Air Traffic Management: Effects of Decision Aid Reliability on Controller Performance and Mental Workload Hum Factors 47:35-49 doi:doi:10.1518/0018720053653802

Papenfuss A, Friedrich M (2016) Head Up Only – A design concept to enable multiple remote tower operations 35th Digital Avionics Systems Conference

Papenfuss A, Friedrich M, Möhlenbrink C, Rudolph M, Schier S, Schmidt M, Fürstenau N (2010) Assessing Operational Validity of Remote Tower Control in High-fidelity Tower Simulation IFAC Proceedings Volumes 43:117-122 doi:http://dx.doi.org/10.3182/20100831-4-FR-2021.00022

Papenmeier F, Huff M (2010) DynAOI: A tool for matching eye-movement data with dynamic areas of interest in animations and movies Behavior research methods 41:179-187 doi:10.3758/BRM.42.1.179

Parasuraman R, Sheridan TB, Wickens CD (2008) Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs Journal of Cognitive Engineering and Decision Making 2:140-160

Patten CJ, Kircher A, Ostlund J, Nilsson L, Svenson O (2006) Driver experience and cognitive workload in different traffic environments Accid Anal Prev 38:887-894 doi:10.1016/j.aap.2006.02.014

Redding RE (1992) Analysis of Operational Errors and Workload in Air-Traffic-Control Proceedings of the Human Factors Society, 36th Annual Meeting, Vols 1 and 2:1321-1325

Salmon PM, Stanton NA, Walker GH, Jenkins D, Ladva D, Rafferty L, Young M (2009) Measuring Situation Awareness in complex systems: Comparison of measures study International Journal of Industrial Ergonomics 39:490-500

Shelton CL, Kinston R, Molyneux AJ, Ambrose LJ (2012) Real-time situation awareness assessment in critical illness management: adapting the situation present assessment method to clinical simulation BMJ quality & safety:bmjqs-2012-000932

Sperandio J-C (1971) Variation of Operator's Strategies and Regulating Effects on Workload Ergonomics 14:571-577 doi:10.1080/00140137108931277

Sperandio J-C (1978) The Regulation of Working Methods as a Function of Work-load among Air Traffic Controllers Ergonomics 21:195-202 doi:10.1080/00140137808931713

Strater L, Tinsley M, Costello A, Colombo D, Endsley M (2010) Situation awareness requirements for Optimized Profile Descent and integrated arrivals and departures NASA Ames Research Center

Taylor RM (1990) Situational Awareness Rating Technique(SART): The development of a tool for aircrew systems design. Paper presented at the Situational Awareness: Limitations and Enhancement in the Aviation Environment, Neuilly-Sur-Seine, France,

Teutsch J, Postma-Kurlanc A Enhanced Virtual Block Control for Milan Malpensa Airport in Low Visibility. In: Integrated Communications, Navigation and Surveillance Conference (ICNS) 2014, 2014. IEEE, pp E1-1 - E1-13. doi:10.1109/ICNSurv.2014.6819984

Vidulich MA The relationship between mental workload and situation awareness. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2000. vol 21. SAGE Publications, pp 3-460-463-463

Wickens CD (2002) Situation awareness and workload in aviation Current directions in psychological science 11:128-133

Wickens CD, Mavor AS, McGee JP (1997) Flight to the future: Human factors in air traffic control. National Academies Press,

Yang C-W, Yang L-C, Cheng T-C, Jou Y-T, Chiou S-W (2012) Assessing mental workload and situation awareness in the evaluation of computerized procedures in the main control room Nuclear Engineering and Design 250:713-719 doi:10.1016/j.nucengdes.2012.05.038

Young MS, Brookhuis KA, Wickens CD, Hancock PA (2015) State of science: mental workload in ergonomics Ergonomics 58:1-17