VALIDATION OF THE TRAFFIC MANAGEMENT INTRUSION AND COMPLIANCE SYSTEM AS SECURITY-AWARENESS-COMPONENT AT THE CONTROLLER WORKING POSITION

M. Schaper, H. Boumann, L. Nöhren, L. Tyburzy, K. Muth, N. Carstengerdes
Deutsches Zentrum für Luft- und Raumfahrt, Lilienthalplatz 7, 38108 Braunschweig, Deutschland

Abstract

The work of an air traffic ground controller can be affected by various security events. The Traffic Management Intrusion and Compliance System uses possible individual indicators to calculate a Security Situation Indicator that provides an assessment of the current security situation to the controller at their working position. To validate the benefit of the tool, real time simulations were conducted at the Air Traffic Validation Center of DLR. This paper describes the methods, the execution and the results of human-in-the-loop experiments performed with ground and apron controllers.

1. INTRODUCTION

The Ground Air Traffic Controller is responsible for a safe, orderly and expeditious flow of the air traffic moving on the ground. Although security itself is not mentioned, it may impact the air traffic controller (ATCO) when security events happen, i.e. intentional and illegal interaction to air traffic control, as it happened e. g. in Berlin (1).
The idea to support the ATCO working on an approach position with a traffic-light colour-coded indication about the security situation at his working position was investigated in the GAMMA project (2). The SATIE project adapted this idea to a ground controller working position, re-designed the calculation of the traffic-light colour-coded indication and refined the overall concept of it. This refined version has then been termed Security Situation Indicator (SSI) (3). This paper describes the enhancements of the used controller assistance system, a surface management system (SMAN) called TraMICS (Traffic Management Intrusion and Compliance System), the design of human-in-the-loop (HITL) validation experiments and the results.

2. TRAMICS

The TraMICS is a prototypical surface management system of the DLR, enhanced with a security component. It was initially developed for the SATIE research project to assist ATCOs in supervising the security situation at their working position. It monitors the behaviour of flights and issues alerts if any non-conformances are detected. The security component, which uses amongst others the non-conformance detections as input, generates the SSI, which is shown to the ATCO and can be shared with other entities, e.g. a Security Operations Center at the airport. For the ATCO, the SSI shall be a quick reference to the current security situation: a green indicator means, that there are no specific security-related actions needed; a yellow one means, the controller should be aware and monitor, as some suspicious events were detected; a red one means that a close monitoring is recommended as there is most probably a security event.

2.1. TraMICS' SMAN Component

The SMAN-function of TraMICS was meanwhile enhanced by calculating and monitoring not only taxi-routes but also trajectories. Trajectories are routes of aircraft enhanced with temporal information, i.e. planned times at each taxi point of the route. With these trajectories a precise plan can be generated that leads to an improved and conflict free traffic flow. To generate conflict free trajectories in real time a genetic algorithm is used (4). Although TraMICS plans trajectories, the ATCO is able to change the route manually; TraMICS will come up with a fitting trajectory based on the request. It is required that the ATCO inputs his/her spoken clearances into the system. This enables the system to supervise conformance regarding various aspects of the flights:

- route deviation: If the flight deviates from the planned route, it will trigger a re-calculation of the trajectory, an alert is raised and shown to the ATCO.
- time deviation: If the flight deviates from the planned time in its trajectory and the deviation is above a threshold, the trajectory will be re-calculated.
- clearance deviation: If the flight is moving without having received the appropriate clearance, an alert is raised and shown to the ATCO.
- conflict risk: If two flights approach each other closer than a pre-configured distance, an alert is raised and shown to the ATCO.

The alerts are shown to the ATCO on the label of the flight on his traffic situation display (FIGURE 1). During verification tests preceding this validation, it became apparent that – from an operator's perspective - the TraMICS' alerts appeared too early. On several occasions, the ATCO gave a clearance verbally, the pseudo pilot implemented the clearance immediately and an alert was triggered before the ATCO had the chance to input the clearance into the system. The displayed alert will vanish as soon as the ATCO inputs the clearance; nevertheless, to prevent irritation, a suppression of route and clearance deviation alerts during the first five seconds was implemented. In other words, if a clearance or route

doi: 10.25967/570101

deviation alert is still valid after five seconds, it will be displayed and counted by TraMICS' security component (cf. section 2.2).

The traffic situation display is also used to inform and support the ATCO in implementing the trajectories planned by TraMICS. Therefore, the planned trajectories are displayed and given clearances are shown. To enable the ATCO to follow the temporal planning an advisory marker is shown. It is a coloured ring around the aircraft position indicator, which indicates what kind of action the flight needs to perform to adhere to the planning, e.g. accelerate or break.

As opposed to a previous DLR SMAN (5) the TraMICS does not advise precise taxi speeds, since pilots cannot control the speed of the aircraft this precise anyways, unless e.g. electric taxi is used. The speed is instead planned in three distinct levels (slow, normal, fast), that correspond to ten, 15 and 20 knots taxi speed approximately.

FIGURE 1. Cut-out of the traffic situation display showing 1) an alert at the label of the flight EWG657U: 2) a trajectory for flight SDM6654 and 3) an advisory marker (pink ring) at KLM82V, showing that the taxi clearance is recommended now for this flight.

2.2. TraMICS' Security Component

The TraMICS security component collects alerts from the SMAN component as well as from other sources, also TraMICS' external. In past projects a Speaker Verification Module was connected to TraMICS, that sends messages in case an unauthorized speaker is detected (3). This already verified function (6) was not in the focus of this experiments and would have required an enrolment process, so the module was not used. Instead a function to send emulated unauthorized speaker detection messages was added to the simulation environment (cp. section 3.2).

The security component collects and counts the appearance of following types of alerts and derives the SSI based on rules and configurable thresholds:

- non-conformant movement, i.e. clearance or route deviation,
- conflict detection,
- ADS-B spoofing, i.e. a message is received that for a specific aircraft the received ADS-B (Automatic Dependent Surveillance – Broadcast) data is spoofed,
- unauthorized speaker.

The occurrences of alerts in the pre-configured sliding time window are counted type-wise and then compared to type-specific red and yellow thresholds. If the red threshold of one type of alerts is exceeded, the SSI will turn red. If there are less alerts of all types than the red thresholds but one type-wise count exceeds its specific yellow threshold, the SSI will be yellow. If for all types of alerts the quantity is less the yellow thresholds, the SSI will be green. The SSI is re-assed periodically. For the validation trials the reassessment period was set to one minute and the sliding window to five minutes, i.e. each 60 seconds all alerts received in the last five minutes will be considered. The SSI (i.e. the colour) as well as the triggering condition is shown to the ATCO on the traffic situation display (FIGURE 2).

FIGURE 2. Cut-out of the traffic situation display showing the received flight-independent alert (upper line) and the SSI (lower line) beginning with the coloured dot.

2.3. Goals

There are two facets of TraMICS that have not been validated previously: The SSI for the ground controller working position and the new concept of how the ATCO is supported in controlling the flights according to the trajectory-based planning.

The goal of this study was to examine the usefulness and acceptance of these two facets exploratively. For this purpose, the following aspects were explored: event detection rates, mental workload, situation awareness, usability and trust.

3. VALIDATION METHODS

The validation was conducted as HITL experiments at DLR's Air Traffic Validation Center (7) in Braunschweig, Germany.

3.1. Sample

The sample comprised eight ground and apron controllers (one female, seven male) from four German airports (Berlin, Hannover, Celle and Braunschweig). Participants were aged between 24 and 45 years ($M = 35.00$, $SD = 8.62$). Their average work experience as controllers was 9.56 years ($SD = 7.47$) and ranged between two and 20 years. Participants provided written consent and received monetary compensation.

3.2. Simulation Environment

DLR's tower simulator with a 360° out-of-the-window view was used to conduct the validation experiment. It was configured with an outdated but not simplified layout of Hamburg airport, which has two crossing runways. Two controller working positions were installed: the tower controller, responsible for the runways, and the ground controller, responsible for taxiways and apron. The ground controller had also to transmit the clearances to cross the runway to the pilots when approved by the tower controller. The role of tower controller was performed by DLR personnel, whereas a validation participant worked as ground controller. The ground controller working position was equipped with a screen containing the traffic situation display on the left and electronic flight strips sorted in bays in the right side of the display (FIGURE 3). A mouse was used for interaction. The ground controllers were equipped with a hand-held radio and loudspeaker to communicate with the pseudo-pilots. The pseudo-pilots were located in a separate room. They operated the aircraft, according to the ATCOs instructions or to the scenario design, which required deviations at some times. The simulation supervisor was sitting next to the pseudo pilots and had the possibility to trigger sending emulated message for ADS-B spoofing detections and unauthorized speaker detections when required by the exercises (cf. section 3.5).

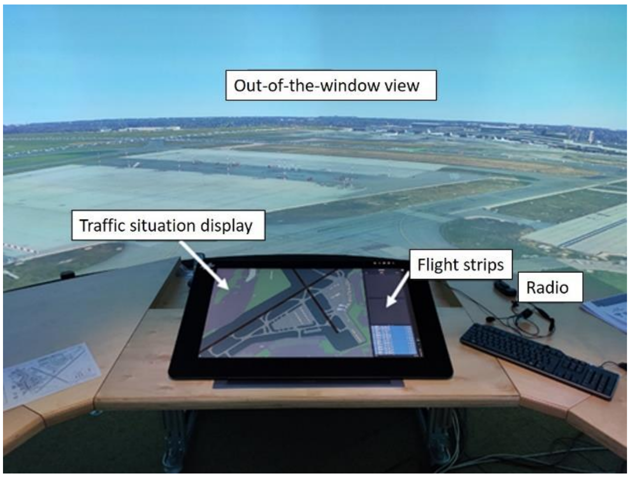

FIGURE 3. Ground Controller working position in the simulation environment.

3.3. Observation Data

During all exercises, two observers monitored the participants. They were tasked with noting down remarks made by the participants and watching for any indication that a security or safety event has been detected by a participant. Participants were instructed to tell the observers about unusual events. Time stamps for the detections were noted if possible.

3.4. Mini Scenarios

Like in (8), the mini scenarios have the advantage of already having established a specific traffic situation the ATCO has to handle, and that events can be planned for all participants nearly deterministically. This is in contrast to long scenarios, where the ATCO builds the traffic situation according to his habits; The resulting traffic situation may vary in different scales from ATCO to ATCO, which hinders comparison. The mini scenarios were used to measure the event detection time of TraMICS system and ATCO.
Each of the nine mini scenarios lasted five minutes at maximum. Two of them had no event planned, in two mini scenarios a flight each deviated from its route and in five mini scenarios a flight each moved without appropriate clearance. In one of the five the departure started taxi although a pushed flight is in front of it.

3.5. Long Scenarios

The long scenarios last about 45 minutes and start with empty moving area and runways. Three different long traffic scenarios were used: one for training (“training scenario”) and two for exercises (Scenario A and Scenario B). Both scenarios A and B had roughly the same amount of traffic and traffic composition, but different callsigns, timings and parking positions. The training scenario was based on Scenario A with less traffic and different times.
In contrast to the mini scenarios, where the security events were pre-planned and scheduled, they had to be inserted into these scenarios on the fly. The observer and the simulation supervisor in the pseudo-pilot room used a chat room to covertly communicate and decide how to interfere in the traffic situations. The following interferences could be used:

- deviation from the cleared route,
- commencing pushback or taxi without clearance,
- continuing taxi without clearance (e.g. crossing a runway),
- ADS-B proofing detection,
- unauthorized speaker.

The last both, ADS-B proofing detection and unauthorized speaker detection were only visible to the ATCO when the visibility of TraMICS’ alert was switched on; nevertheless, e.g. taxi without clearance could be provoked by an unauthorized speaker and some pseudo pilots also reported, that they had received the clearance already.

3.6. Questionnaires

Mental workload was assessed every five minutes during the long scenarios on a 5-point scale from 1 (under-utilised) to 5 (excessive) using the Instantaneous Self-Assessment (ISA) (9). A rating of 3 represents the optimal level of workload while values above and below represent under- or overload.
A set of EUROCONTROL’s “Solution for Human-Automation Partnerships in European ATM” (SHAPE) questionnaires (10) was administered: Situation Awareness for SHAPE (SASHA), Assessing the Impact of Automation on Mental Workload in the short version (AIM-s) and SHAPE Automation Trust Index (SATI).

- SASHA assesses SA and comprises six items that are rated on a 7-point scale from 0 (never) to 6 (always). An overall score is obtained by inverting the item scores of items 2, 3, 5 and 6 and calculating the mean of all six items scores. A higher score is desired as it indicates higher situation awareness.
- AIM-s assesses mental workload on 16 items. Ratings are given on a 7-point scale from 0 (none) to 6 (extreme). The overall score is obtained by calculating the mean of the 16 items. Mid-scale ratings are desirable while extreme values represent under- or overload.
- SATI measures the level of trust in a system on a 7-point scale from 0 (never) to 6 (always) and consists of six items. Each item represents a dimension of trust (utility, reliability, accuracy, understanding, robustness,

confidence) and can be interpreted meaningfully on its own. An overall score is obtained by calculating the mean of the six items.

Furthermore, the System Usability Scale (SUS) (11) was administered to obtain usability ratings. The SUS is made up of ten items that are rated on a 5-point scale ranging from 1 (strongly disagree) to 5 (strongly agree). To get the overall SUS score, item contributions are calculated. For items 2, 4, 6, 8 and 10, the item contribution equals five minus the scale position. For items 1, 3, 5, 7 and 9, the item contribution equals the scale position minus one. The overall SUS score is calculated by multiplying the sum of the item contributions by 2.5. A score of zero equals the worst usability possible and a score of 100 equals the best usability possible.
Additionally to the standard questionnaires a customized one was prepared. The participants received tailored questions that were rated on a 5-point Likert scale ranging from 1 (not accurate at all) to 5 (completely accurate). The statements referred to the topics of planning of routes and trajectories and security, including statements about the SSI. For some statements, participants were asked to elaborate on their ratings in a free text.

3.7. Experimental Design

A within-subjects design with the factor presence of alerting (with alerting vs. without alerting) was used to examine the dependent variables workload, situation awareness, usability and trust using an explorative approach.
After giving informed consent, the participants were briefed about the airport layout and TraMICS' routing functionality. Thereafter, they received a training to get familiar with the airport and the controller working position.
For the experiments a step-wise approach of three consecutive exercises was chosen. TraMICS is able to suppress sending controller support features to the controller working position, which was used to

A) show only planned route information;
B) show trajectories and an advisory marker to support time-based control;
C) show B) plus additional alerts (TraMICS' self-detected, e.g. non-conformances, as well as others, e.g. ADS-B spoofing) and the SSI.

The first exercise used option A). It consisted of nine consecutive mini scenarios. The order of the mini scenarios was randomized for each participant. The mini scenarios were used to measure the event detection time of TraMICS system and the participant.
After the mini scenarios, participants completed two exercises using the long scenarios. In one exercise TraMICS displayed the trajectories and an advisory marker to support time control only (option B). In the other exercise TraMICS showed alerts and the SSI in addition to that (option C). To avoid order effects, the order in which option B and option C (without alerting vs. with alerting) were presented was counterbalanced among participants. The type of scenario (Scenario A or Scenario B) was counterbalanced as well. Each participant experienced Scenario A and Scenario B once. This resulted in two variants of the agenda, see TAB 1.
Participants received a briefing about the functions they would be using and an additional training session depending on the agenda variant, but including time-based control in any case. For Variant 1, this means they were briefed and trained regarding the time-based control before Exercise #2 and received an additional briefing about security and the SSI correlation before Exercise #3. For Variant 2, participants were briefed and trained regarding time-based control as well as security and the SSI correlation all at once before Exercise #2. No further briefing took place before Exercise #3 in this case. In both Exercises #2 and #3, ISA ratings were obtained every five minutes.
After each exercise, participants received the SASHA, AIM-s, SATI and SUS. Only after having conducted all exercises, a final questionnaire comprising tailored questions was administered. A debriefing was conducted after each set of questionnaires.

1	Welcome and briefing on airport layout and controller working position, including TraMICS' routing (but no alerting and no time-based support)	
2	Training on airport layout and controller working position, including TraMICS' routing	
3	Exercise #1: Mini scenarios (no alerting, no time-based planning indicator, i.e. option A)	
4	Standard questionnaires, debriefing	
5	Briefing time-based control	
	Variant 1	Variant 2
6	Training: time-based control (option B)	Briefing: Security and correlation
7	Exercise #2 (V1): time-based control (option B)	Training: time-based control and support through alerts (option C)
8	Standard questionnaires and debriefing	Exercise #2 (V2): time-based control and support through alerts (option C)
9	Briefing: Security and correlation	Standard questionnaires and debriefing
10	Exercise #3 (V2): time-based control and support through alerts (option C)	Exercise #3 (V2): time-based control only (option B)
11	Standard questionnaires and tailored questions, debriefing	

TAB 1. Exercise procedure.

4. RESULTS

All statistical analyses were conducted using IBM SPSS Statistics Version 26 (12). A significance level of $\alpha = .05$ was used for all statistical tests. No data sets had to be excluded from the analysis.

4.1. Detection Times in the Mini Scenarios

During Exercise #1, the mini scenarios, the observers noted the detection times of the scenario-planned events. Those were compared with the times, TraMICS would have send the alert information to the controller working position. Seven of the nine mini scenarios contained one event to be detected by the ATCO. This means in sum 56 event occurrences are planned to happen for eight participants together. Overall, in six cases an ATCO managed to change the traffic situation in such a way that the planned event was prevented. For the remaining 50 event occurrences, the ATCO was quicker than the system in seven cases (14%). If the original TraMICS alert detection time would be have been used, which is five seconds earlier (cp. section 2.1), the ATCO would have been quicker only in one case (2%). Two occurrences of events (4%) were not detected by an ATCO, but by the system.

4.2. Check for Confounding Variables

As a first step, differences between Scenario A vs. Scenario B were assessed to ensure the comparability of both scenario types. For this reason, Wilcoxon tests were computed on SASHA, AIM-s, SUS and SATI mean scores. No significant effects were found, all $p > .05$, see TAB 2.

Score	Scenario A		Scenario B			
	Md	*N*	*Md*	*N*	*z*	*p*
SASHA	4.17	8	4.17	8	0.00	> .999
AIM-s	1.75	8	1.78	8	-0.70	.531
SUS	70.00	8	66.25	8	0.00	> .999
SATI	3.83	8	3.58	8	-0.63	.578

TAB 2. Wilcoxon tests examining the effect of the type of scenario (A vs. B) on SASHA, AIM-s, SUS and SATI mean scores. Exact *p*-values were used.

To check for training effects, differences between the mini-scenarios vs. Exercise #2 vs. Exercise #3 were assessed. Friedman tests were computed on SASHA, AIM-s, SUS and SATI mean scores for this purpose. No significant effects were revealed, all $p > .05$, see TAB 3.

4.3. Mental Workload

First of all, it was examined whether ISA ratings were affected by the presence of alerting and by the assessment time within the long scenarios.
For the ISA ratings, a two (alerting: with vs. without) x seven (assessment time) repeated-measures ANOVA was computed. Since the scenario durations and thus the number of ISA ratings differed between participants, missing ISA ratings were substituted with the mean of the variable. Only the first seven assessment times were included in the analysis because there were too few datapoints to do the same for assessment 8 (after 40 minutes) and assessment 9 (after 45 minutes).
The sphericity assumption was not violated, all $p > .05$, and no corrections were performed. The ANOVA revealed no significant main effect of the presence of the SSI on ISA ratings, $F(1, 7) = 2.00$, $p = .200$. No significant interaction between presence of the SSI and assessment time was found either, $F(6, 42) = 0.36$, $p = .900$. However, a significant main effect of assessment time was found, $F(6, 42) = 14.55$, $p < .001$, $\eta^2_p = .68$.
Bonferroni-corrected post-hoc pairwise comparisons showed that assessment 1 (after five minutes) was significantly lower than assessments 2, 3, 4 and 7, respectively, see TAB 4.
FIGURE 4 shows the mean ISA ratings. Overall, most ISA ratings indicated a mid-level to slightly lower than mid-level mental workload. Only assessment 1 of each exercise fell below a rating of two. On a descriptive level, ISA ratings in the exercises with alerting were slightly higher than in the condition without the SSI.

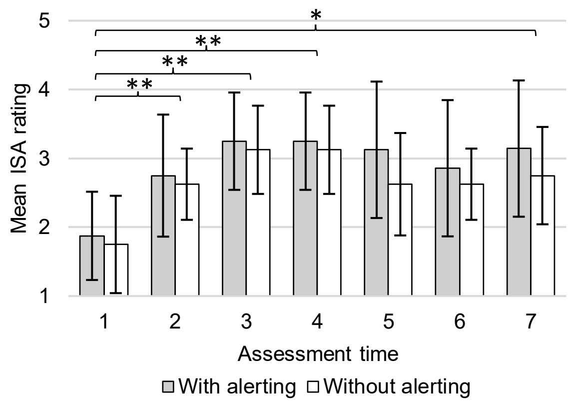

FIGURE 4. Mean ISA ratings in dependence of the presence of the SSI (with alerting vs. without alerting) and assessment time (1 – 7). Error bars represent standard deviations. Significant comparisons are marked p < .05* and p < .01**.

Score	Mini scenarios		Exercise #2		Exercise #3				
	Md	*N*	*Md*	*N*	*Md*	*N*	X^2	*df*	*p* (exact)
SASHA	3.83	8	4.25	8	4.08	8	1.46	2	.532
AIM-s	2.00	8	1.84	8	1.66	8	5.23	2	.065
SUS	65.00	8	71.25	8	65.00	8	5.87	2	.061
SATI	4.08	8	3.67	8	3.92	8	0.07	2	.991

TAB 3. Friedman tests examining the effects of exercise number (mini-scenarios vs. Exercise #2 vs. Exercise #3) on SASHA, AIM-s, SUS and SATI mean scores. Exact *p*-values were used.

Comparison		Mean difference (SE)	*p*
Assessment 1 versus	2	-0.88 (0.13)	.004**
	3	-1.38 (0.21)	.006**
	4	-1.38 (0.18)	.003**
	5	-1.06 (0.24)	.064
	6	-0.93 (0.24)	.135
	7	-1.13 (0.20)	.017*
Assessment 2 versus	3	-0.50 (0.13)	.152
	4	-0.50 (0.13)	.152
	5	-0.19 (0.19)	>.999
	6	-0.05 (0.22)	>.999
	7	-0.26 (0.21)	>.999
Assessment 3 versus	4	0.00 (0.09)	>.999
	5	0.31 (0.13)	>.999
	6	0.45 (0.17)	.752
	7	0.24 (0.17)	>.999
Assessment 4 versus	5	0.31 (0.13)	>.999
	6	0.45 (0.20)	>.999
	7	0.24 (0.14)	>.999
Assessment 5 versus	6	0.13 (0.16)	>.999
	7	-0.07(0.11)	>.999
Assessment 6 versus	7	-0.21 (0.16)	>.999

TAB 4. Bonferroni-corrected pairwise comparisons between assessment times. $^{**}p < .01$, $^{*}p < .05$.

For the mean AIM-s score, a conducted Wilcoxon test did not reveal a significant difference between exercises with alerting ($Md = 1.78$) versus without alerting ($Md = 1.63$), $z = -0.95$, $p = .343$. FIGURE 5 shows the mean AIM-s scores dependent on the presence of the alerting. In both conditions, participants rated their mental workload slightly lower than the middle of the scale.

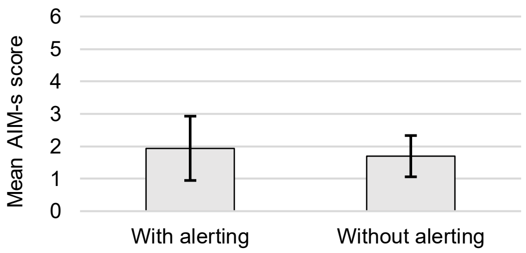

FIGURE 5. Mean AIM-s scores dependent on the presence of alerting (with alerting vs. without alerting). Error bars represent standard deviations.

4.4. Situation Awareness

TAB 5 shows the number of event occurrences by type that was implemented in the exercises with and without alerting as well as the number of events occurrences that were noticed by ATCOs, according to the assessments by the observers. Detections for speaker verification alerts and ADS-B spoofing alerts could not be obtained reliably and are therefore not included. Only one security event occurrence, a route deviation, went unnoticed by an ATCO in an exercise without alerting.
For the mean SASHA score, a conducted Wilcoxon test did not reveal a significant difference between the condition with alerting ($Md = 4.08$) and the condition without alerting ($Md = 4.25$), $z = -0.95$, $p = .343$. The mean SASHA scores are shown in FIGURE 6. In both conditions, mean SASHA scores around 4 were obtained.

	With alerting		Without alerting	
Event type	No. of occurrences	Detected	No. of occurrences	Detected
Route deviation	12	12	15	14
No clearance	16	16	19	19
Speaker verification alert	13	N. A.	N. A.	N. A.
Spoofing alert	8	N. A.	N. A.	N. A.

TAB 5. Number of event occurrences implemented in the scenarios and number of detected event occurrences by the ATCO, summarized over participants.

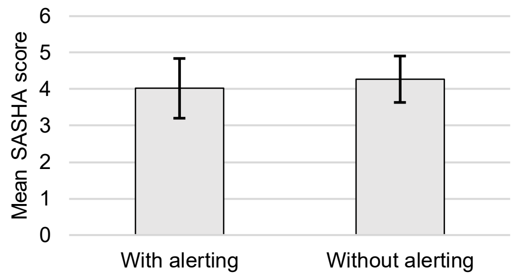

FIGURE 6. Mean SASHA scores dependent on the presence of alerting (with alerting vs. without alerting). Error bars represent standard deviations.

4.5. Acceptance

Acceptance was evaluated by measuring usability and trust using the SUS and SATI.
Participants rated the system with a mean SUS score of $M = 65.00$ ($SD = 8.66$) in the condition with alerting and $M = 68.44$ ($SD = 6.94$) without alerting. A conducted Wilcoxon test did not reveal a significant difference between the condition with alerting ($Md = 63.75$) and the condition without alerting ($Md = 70.00$), $z = -0.78$, p = .436.
The mean SATI scores were $M = 3.90$ ($SD = 0.69$) in the condition with alerting and $M = 3.69$ ($SD = 0.61$) in the condition without alerting. A conducted Wilcoxon test revealed no significant effect of the presence of alerting (with alerting vs. without alerting) on the mean SATI score, $z = -.70$, $p = .483$ ($Md = 3.83$ with alerting and $Md = 3.58$ without alerting).
For the tailored questions, means and standard deviations of the agreement ratings were calculated. The wordings can be seen in FIGURE 7, FIGURE 8 and FIGURE 9. FIGURE 7 illustrates the mean agreement to statements regarding the TraMICS routes and trajectories. Regarding statement 3, seven participants explained how their selection of routes would have differed from the TraMICS routes. Five participants stated that they would have chosen different taxiways than the TraMICS as a standard. One participant criticized the use of unrealistically short timings, e.g. between starting the engine and taxi request, and one participant would have preferred the routes to be shorter and more conflict-free.
Regarding statement 4 about the advisory marker, four participants added suggestions for the improvement. The use of unrealistic timings was criticized by two participants.

Changes in the visual design of the advisory marker were suggested twice (blinking to draw attention and using more different colours). One participant suggested additional temporal information in the flight progress strips.
Furthermore, tailored statements centred around the topic of security were presented to participants. The mean agreement ratings are shown in FIGURE 8.
FIGURE 9 shows participants' mean agreement with statements regarding the SSI. In response to statement 18, two participants proposed improvements to the labels. Both participants criticized the fact that alerts are sometimes caused by a lack of input into the system by the controller. It was proposed that only actually dangerous situations should trigger an alert or the controller should be able to delete such alerts.
As a last question, participants were asked if they would like to use a function allowing the controller to evaluate the system's assessment of the security situation by making manual inputs. It was explained that this could enable the system to learn and improve. The mean agreement to this question was $M = 4.00$ ($SD = 1.07$). One participant commented that the described function could increase work pressure.

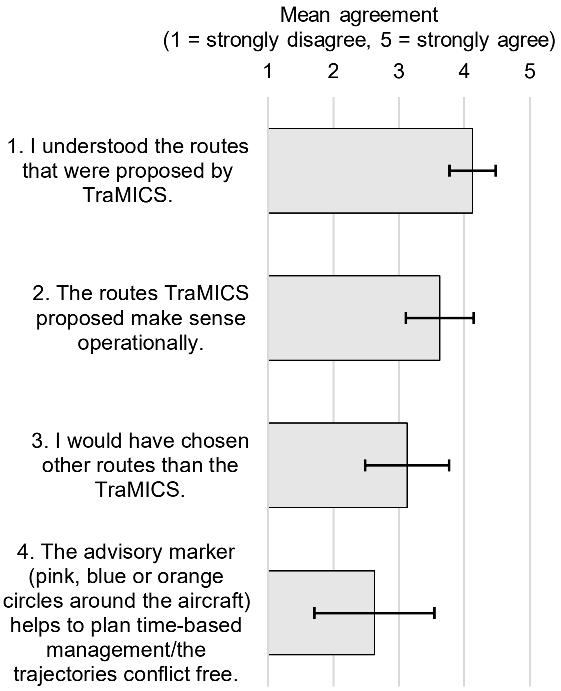

FIGURE 7. Mean agreement to custom questions regarding the routes and trajectories. Error bars represent standard deviations.

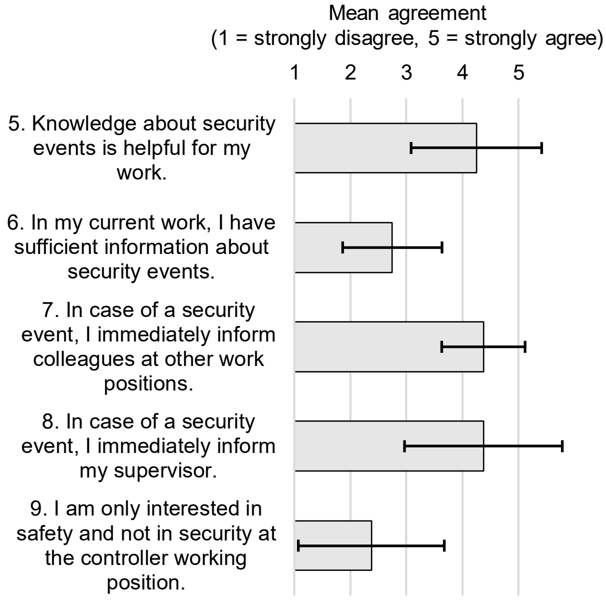

FIGURE 8. Mean agreement to custom questions regarding security. Error bars represent standard deviations.

5. DISCUSSION

The goal of the present study was to evaluate the usefulness and acceptance of two components of the TraMICS at a ground controller working position exploratively: The SSI, which gives an estimation of the current security situation at the controller working position, and TraMICS' SMAN function that calculates trajectories and shows support features to the ATCO.
One limitation to this study is the small sample size of eight controllers that reduces the explanatory power. This also means that possibly, effects might not have been detected due to reduced statistical power. Another limitation is extent of variety concerning the security event occurrences that were inserted during the long scenarios. Because the security events were inserted adaptively to match the current traffic situation, the long scenarios' comparability is impaired both between and within participants.
The comparison of TraMICS' and ATCOs' detection times of event occurrences during the mini scenarios showed that the system was mostly faster than the ATCOs. Two cases were even detected by the system only. So, the display of events has the potential to support the ATCO in detecting more events. The kind of visualization and interaction with the controller working position (e.g. clearance input by mouse vs speech recognition) can be improved.
Regarding differences between Scenario A and Scenario B, no significant effects on SASHA, AIM-s, SATI and SUS mean scores were found, suggesting both long scenarios were comparable in this regard. No effects of exercise number (mini scenarios vs. Exercise #2 vs. Exercise #3) on SASHA, AIM-s, SATI and SUS mean scores were found as well, i.e. no training effects became apparent. However, in contrast to this, several participants reported that they started to feel more familiar with the system in the course of the exercises.

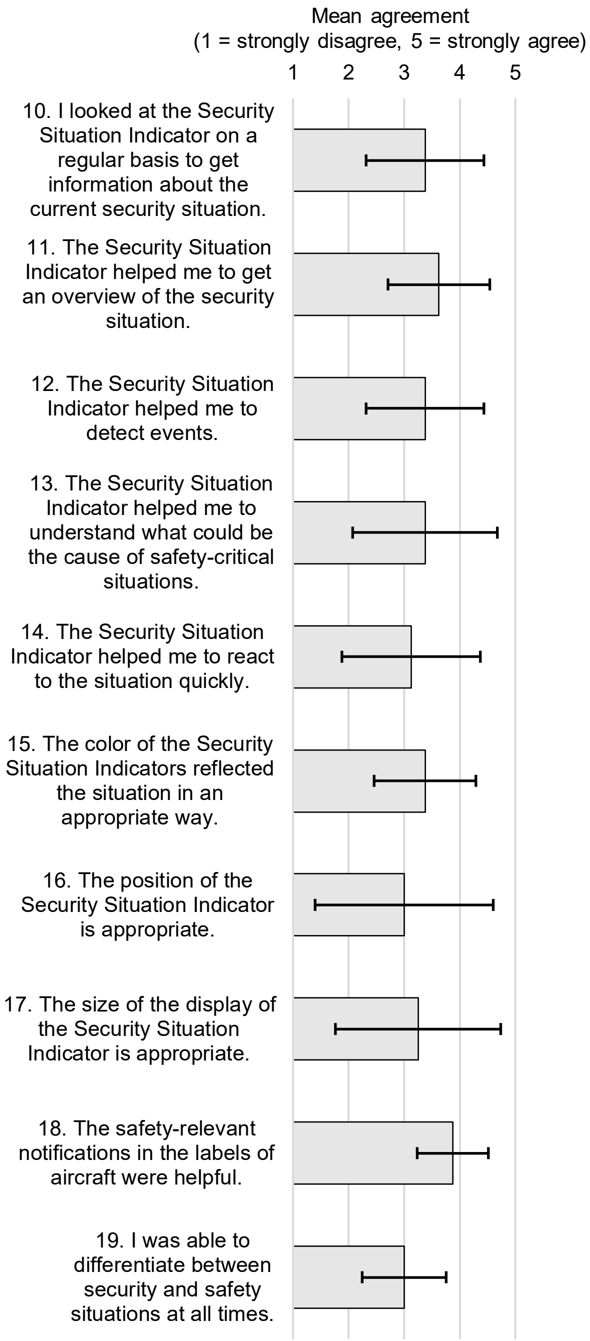

FIGURE 9. Mean agreement to statements regarding the SSI. Error bars represent standard deviations.

Next, the influence of alerting on mental workload and situation awareness was assessed. Regarding mental workload, different outcomes would have been imaginable. On the one hand, the SSI could have resulted in a reduction of mental workload because the ATCO is supported in detecting events. On the other hand, an increase in mental workload associated with the SSI could also be possible, especially when the SSI is yellow or red, as this draws the ATCO's attention to security issues in addition to safety.
Concerning mental workload measured by ISA during the exercises, no significant main effect of the presence of the alerting on ISA ratings was found, suggesting that the alerting did not influence mental workload. However, there was a significant main effect of assessment time: the first ISA assessment was significantly lower than assessments 2, 3, 4 and 7, meaning that mental workload was lower after the first five minutes than after ten, 15, 20 and 35 minutes, respectively. This effect could be attributed to the fact that the first few minutes of the long scenarios had few and easy traffic and no security events. In accordance with the ISA results, no significant effect of the presence of alerting on AIM-s mean scores was found either.
A mid-level of mental workload is generally desirable. Overall, the mean AIM-s scores suggested a slightly lower than mid-level of workload, while ISA ratings suggested a mid-level to slightly lower than mid-level of workload.
Concerning situation awareness, it could have been expected that the alerting increases the ATCOs' situation awareness as it provides information about the security situation. The number of event occurrences (excluding speaker verification alerts and ADS-B spoofing alerts) detected by participants was assessed in both exercises with and without the alerting. A higher detection rate might be indicative of better situation awareness. Only one event occurrence was missed by an ATCO in an exercise without alerting, and this ATCO confirmed that they did not detect this event in the debriefing. All other event occurrences were detected. However, these numbers have to be interpreted cautiously because they are based on observations. Even though participants were asked to tell the observers when they noticed unusual events, this instruction was not always followed. Due to experiment design the participants were not able to detect ADS-B spoofing and unauthorized speakers without the alerting, so this is not counted for the analysis. Furthermore, participants rarely reacted noticeably to alerts displayed by the SSI. Nevertheless, one participant stated in the debriefing, that the unauthorized speaker alert caused them to be more careful and stricter in his commands.
Regarding subjective situation awareness, no significant effect of the presence of alerting on SASHA was found, suggesting that alerting had no influence on situation awareness. This finding is in contradiction to the expectation that alerting might increase situation awareness. One limitation to subjective situation awareness ratings and a possible explanation for these findings could be the fact that individuals usually have no knowledge about information they missed; and thus, subjective situation awareness might not be in line with actual situation awareness.
In the case of situation awareness, a high score is generally desirable. Overall, the mean SASHA ratings suggested a slightly higher than mid-level subjective situation awareness.
Concerning usability, the mean SUS scores for the conditions with and without alerting corresponded to adjective ratings between "OK" and "good" according to Bangor, Kortum & Miller (13). Trust was not found to differ significantly between conditions with and without alerting as well.
The answers to the tailored questions regarding TraMICS' routes and trajectories (cf. FIGURE 7) suggest that the time-based support can be improved. Regarding the advisory marker, participants suggested using more realistic timings, modifications to the advisory marker's visual design or additional visual aids. Statement 3 "I would have chosen other routes then TraMICS" was rated neutral on average. This can be explained: TraMICS does not

necessarily use the same optimization criteria the controllers use. According to the participants, a kind of roundabout traffic is usual, which TraMICS does currently not consider. Concerning the tailored questions about the topic of security (cf. FIGURE 8), participants' answers suggest that they generally agreed to the relevance of knowledge about security events for their work. However, the answers to statement 6 show that there seems to be a need for more information about security events at controllers' working positions. In addition, participants stated overall agreement to informing colleagues and supervisors in case of a security event (statements 7 and 8). Overall, the tailored questions regarding the SSI were rated neutral to positive (FIGURE 9) but there seems to be potential for improvement. The position and size of the SSI were identified as areas for improvement specifically, as can be seen from the rather high standard deviations for statements 16 and 17. Independent of the SSI, the notifications in the labels seemed to be especially helpful since statement 18 received the highest average agreement with a small standard deviation. However, one point of criticism was the fact that alerts were sometimes caused by a lack of input into the system by the controller instead of an actually dangerous event.

Participants were generally in favour of the idea to use manual inputs to evaluate the systems' assessment of the security situation and use this to make the system learn, but a participant raised the concern that a function like this could increase the pressure of work. During the debriefings, some other benefits of the system were mentioned, e.g. the fact that inputting the clearances into the system could support the controllers in remembering their given clearances. As mentioned before, in the final debriefing, some participants reported that they felt better trained at that point. For further trials, longer and more experiments for each person could be feasible.

6. CONCLUSIONS

The results have to be seen in the light of the current situation: directly after the COVID pandemic restrictions and before the European energy crisis. Flight demand was recovering, but due to a lack of personnel (in air traffic control centres, security staff, ground handlers) a large number of flights had to be regulated or could not be served in time on ground. Therefore, the participants might not have had the same need for trajectory optimisation as they had before the COVID-pandemic.

The participants confirmed that knowledge about security events is helpful for their work in general and specifically, that the Security Situations Indicator helped to get an overview of the security situation. The comment, that alerts were sometimes caused by no or late controller's clearance input into the system, could be mitigated by using speech recognition. Good suggestions to improve specific visualizations and interactions have been received. These ideas have to be further developed and validated.

7. AKNOWLEDGEMENT

The project on which the presented research is based is funded by the German Federal Ministry of Education and Research (BMBF) as part of the call „Zivile Sicherheit – Kritische Strukturen und Prozesse in Produktion und Logistik“ with the funding number 13N15104. The responsibility for the content of this publication lies with the authors.

8. REFERENCES

1. **AP News.** AP. [Online] 29 01 2021. [Cited: 11 02 2022.] https://apnews.com/article/arrests-berlin-f56833f73c7ecfa34a5ed5e6461669bc.
2. *Validating an ATM Security Prototype - First Results.* **Tim H. Stelkens-Kobsch, Michael Finke, Matthias Kleinert, Meilin Schaper.** 2016. Proceedings of the 35. DASC conference.
3. *The Traffic Management Intrusion and Compliance System as Security Situation Assessment System at an Air Traffic Controller's Working Position.* **Meilin Schaper, Olga Gluchshenko, Kathleen Muth, Lukas Tyburzy, Milan Rusko, Marián Trnka.** s.l. : 31st European Safety and Reliability Conference (ESREL), 2021.
4. *Real-Time Calculation and Adaption of Conflict-Free Aircraft Ground Trajectories.* **Nöhren L, Schaper M, Tyburzy L, Muth K.** Stockholm : s.n., 2022. ICAS 2022.
5. *Taxi Routing for Aircraft: Creation and Controlling.* **Gerdes I, Temme A.** s.l. : Second SESAR Innovation Days, 2012.
6. *Speaker Authorization for Air Traffic Control Security.* **Marian Trnka, Sakhia Darjaa, Milan Rusko, Meilin Schaper, and Tim H. Stelkens-Kobsch.** s.l. : SPECOM 2021, Springer-Verlag, Berlin, Heidelberg, vol. 12997, pp. 716–725, 2021.
7. **DLR.** The Air Traffic Validation Center. [Online] [Cited: 01 09 2022.] https://www.dlr.de/fl/en/desktopdefault.aspx/tabid-1140/.
8. *A Practical Example for Validation of ATM Security Prototypes.* **Finke, Michael und Stelkens-Kobsch, Tim H.** s.l. : CEAS Aeronautical Journal. Springer. ISSN 1869-5590, 2018.
9. **Tattersall, Andrew J. and Foord, Penelope S.** An Experimental Evaluation of Instantaneous Self-Assessment as a Measure of Workload. *Ergonomics.* 2007, Vol. 39, 5, pp. 740-748.
10. **Dehn, Doris M.** Assessing the Impact of Automation on the Air Traffic Controller: The SHAPE Questionnaires. *Air Traffic Control Quarterly.* 2008, Vol. 16, 2, pp. 127-146.
11. **Brooke, John.** SUS - A Quick and Dirty Usability Scale. [ed.] Patrick W. Jordan, et al. *Usability Evaluation In Industry.* London : Taylor & Francis Ltd., 1996, pp. 189-194.
12. **IBM Corp.** IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY : IBM Corp., 2019.
13. *Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale.* **Bangor, Aaron, Kortum, Philip and Miller, James.** 2009. Vol. 4, pp. 114-123.