Enhancement of Air Traffic Controller's Task Engagement for Smooth Transition from Supervisory to Manual Control

Master of Science Thesis

M.Kim

TU Delft

Enhancement of Air Traffic Controller's Task Enagegment for Smooth Transition from Supervisory to Manual Control

by

M. Kim

Student number:	4221699	
Project duration:		
Thesis committee:	Dr. ir. C. Borst,	TU Delft, daily supervisor
	Prof. dr. ir. M. Mulder,	TU Delft, supervisor
	Dr. M. A. Mitici,	TU Delft, examiner

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Cover Image: NASA ATD-2 Simulation
https://www.nasa.gov/sites/default/files/thumbnails/image/arc-2002-acd02-0050-3large_0.jpg

TU Delft

Preface

I would like to present the result of my year-long thesis research, “Enhancement of Air Traffic Controller’s Task Engagement for Smooth Transition from Supervisory to Manual Control". In this report, a scientific paper and literature survey are presented, in which I describe the performed work to graduate at Control & Simulation department of the Faculty of Aerospace Engineering at Delft University of Technology (TU Delft). It was a pleasure for me to work on the topic that involves cognitive science, as automation has become more important than ever and it is crucial to understand where humans stand now.

First of all, I would like to express my gratitude to my daily supervisor, dr. ir. Clark Borst, who has always been supportive, shared invaluable insights with me and provided great feedback on my work throughout a year of the thesis research. I really appreciated weekly meetings, as they tremendously helped me with keeping myself motivated and on track. Even during difficult times, he was always supportive and inspiring, which gave me extra motivation to push through. Also, his expertise in various fields was a great asset to the research. I can say with confidence, that I could not have had better experience of doing a MSc thesis research without him. Secondly, I would like to thank prof. dr. ir. Max Mulder, who also provided me with interesting and important insights and was always supportive during our kick-off, midterm and greenlight meetings. My gratitude also goes out to the eight participants who volunteered for my experiment to help me with gathering necessary data I needed for my research.

I could not have reached the end of my journey here without my friends. I would like to express a special gratitude to my dear friends Şan and Kostas, who halved our pain and suffering, and doubled joy and happiness throughout the bachelor’s and master’s degree program. Also, I would not been able to reach the end without my loving girlfriend Daniëlle, who has always been supportive and caring no matter the circumstances. Without her unlimited compassion, things would have turned out very differently.

Last but not least, I would like to thank my family who has always been supportive and loving, which helped me staying afloat throughout my long studies at TU Delft. Without them, I would not be where I have journeyed.

My time in the Netherlands is over with the end of my thesis research. I will return to South Korea to carry out compulsory military service, and hope to be back in Europe to influence the world with invaluable lessons I learned at TU Delft.

M. Kim
Delft University of Technology, November 2019

Contents

List of Figures

List of Tables

I

Master of Science Thesis Paper

Enhancement of Air Traffic Controller's Task Engagement for Smooth Transition from Supervisory to Manual Control

Author: Munyung Kim
Supervisors: Clark Borst, Max Mulder

***Abstract*—To meet the increasing demands of air traffic, automated systems have been introduced to help air traffic controllers cope with the increasing air traffic in the next two decades. A challenge is that the supervision and evaluation of automated conflict detection and resolution tools have to be performed by human air traffic controllers. These can suffer from vigilance and complacency problems in an extreme supervisory control environment, possibly reducing safety, together with an inability of human controllers to take over when the automation fails. In this study, a form of situation awareness feedback was used to assist controllers in maintaining their task engagement during the supervisory control, and increasing their manual control performance, in the presence of an automation failure. Results from a human-in-the-loop experiment, in which eight participants were instructed to monitor a fully automated air traffic control system and performing manual conflict resolution tasks when the automated system ceased to work, revealed a significant decrease in a workload peak briefly after the automation failure. Although the selected method of asking task-related situation awareness questions to controllers did not necessarily yield improved safety and control efficiency, the results from the experiment suggest that utilizing situation awareness feedback in line with controllers' attention is an avenue worth exploring further.**

***Keywords*—Air traffic control, Automation, Supervisory control, Task engagement, Situation awareness**

I. Introduction

As automation capabilities are being developed in aviation, most automated systems still cooperate with humans who assume a role of supervisors who are tasked with overseeing and directing the automation's performance [1]. By assuming a role of a supervisor, a human operator must stand by in case the automation cannot handle certain situations or ceases to work; most automation unfortunately suffers from brittleness, operating as intended for scenarios it is designed to handle, but requiring human intervention to cope with scenarios that the automation cannot handle [2]

Although fully automated systems offer great advantages of allowing the air traffic capacity to increase, a possible consequence is the out-of-the-loop performance (OOP) problem: the decreased capability of operators working with fully automated systems to perform manual control tasks after an automation failure [3]. Loss of situation awareness (SA) is considered as a crucial factor that contributes to the OOP problem due to over-reliance and passive monitoring of automated systems [4]. The OOP problem exposes human operators to loss of control when automated systems cease to function [5]. Research on investigating the impacts of automation categories on air traffic controller (ATCO) SA and performance revealed that with imperfect automation, automating sensory processing by improving controller SA greatly helps controllers detect upcoming conflicts earlier after an automation failure [6].

SA is a crucial component in human information processing and necessary in ATCO decision making process [7]. One key parameter for the Next Generation Air Transportation System (NextGen) is SA in order to achieve high levels of human-automation interactions. SA is scientifically acceptable concept and a proper SA training in a simulated ATC environment can be used to understand how controllers acquire and maintain their SA [8]. Single European Sky ATM Research (SESAR) envisions the future aviation environment with increasing the level of automation support, which would improve productivity of ATCOs and increase en-route capacity [9]. Eurocontrol also has acknowledged the role shift of ATCO from active controllers to monitoring supervisors. The project STRESS investigated the controller stress and workload in normal conditions and automation disruptions when human controllers must react swiftly under high stress [10]. Project MINIMA investigated possibilities of detecting the OOP problems and compensating for them through smart human-automation task distributions and live operator vigilance and attention level metrics [11]. In addition, TU Delft developed the Solution Space Diagram (SSD), a decision support tool used to assist controllers in fault diagnosis of automated device for tactical conflict detection and resolution [12].

A question arises from the complex interaction of SA and the transition of human operators from a passive supervisor to an proactive controller; how can task engagement of ATCOs be increased, to achieve a seamless transition from supervisory control to manual control when automation fails, such that safe and efficient operations can be performed? In other words, how can we ensure that ATCOs maintain high SA while they supervise a fully automated ATC system for a prolonged time without falling into complacency and boredom issues, such that a smooth takeover with minimized workload peak can occur and result in better manual control performances?

An interactive task assistant tool was developed to answer this question. The prototype was tested in a human-in-the-loop experiment to evaluate its capability to increase task engagement of operators during the supervisory control and manual control performances after the automation failed.

II. Background

A. Ironies of automation

When automation was first introduced, the aim was to minimize human aspects in manual control, planning and problem resolution by means of automated systems. Bibby and colleagues identified paradoxical characteristics of automation, which emphasizes the increased necessity and importance of human operators in highly automated systems [13]. Since the system can never be perfect and can acknowledge every possible abnormality, the presence of human controllers is necessary if an intervention is required to prevent abnormalities from negatively impacting the system safety or mitigate the consequences of the anomalies [14].

Bainbridge identified ironies of manual take-over and approaches to overcome difficulties caused in manual control skills, cognitive skills and monitoring [15]. Our research mainly focuses on the monitoring issues, as a fully automated ATC system would eventually push ATCOs to assume the role of supervisors. Based on vigilance studies, it was found that it is impossible to maintain effective visual attention for more than half an hour towards a source of information where rarely anything happens, even for a highly motivated human operator [16]. Especially for en-route ATC, where the majority of aircraft in the upper airspace is not experiencing many active altitude changes compared to lower airspace, en-route ATCOs may suffer boredom and vigilance degradation. It is nearly impossible to monitor for rarely-occurring abnormalities such as a system failure, and it is likely that an ATCO would not be able to detect such system failures unless a distinct failure alarm is provided.

On top of the vigilance issue, a more serious irony can be identified as the fact that the ATCOs are asked to monitor efficiency and validity of an automated ATC system, even though the automated ATC system is designed to perform ATCO tasks more efficiently and safely than the ATCO. The ATCO simply lacks mental capacity to follow the process of the automated system, as it utilizes more dimensions and uses more precise and specific criteria than the human ATCO [15]. Over time, the irony of monitoring can evolve into the following: even minor abnormalities in complex sociotechnical systems can worsen the consequences exponentially by human-automation interaction [14]. For ATCOs, minor errors such as misreading information of flight labels and missing the active status of aircraft in a responsible sector can result in greater accidents in the event that ATCOs are not actually aware of what caused the automation-related anomalies.

Manual control skills are also affected by the ironies of automation; the quality of physical skills regarding the refinements of timing and gain tend to degrade when they are not regularly practiced [15]. It means that an ATCO who has been an expert controller may be an inexperienced operator after monitoring the automated system. When the ATCO needs to take over, the ATCO may be forced to spend time to create feedback between the controller and system, instead of performing direct control in an open loop. This can put the operator in a dilemma, as the feedback will not precisely tell the operator whether system is malfunctioning or there is a mistake in the operator's control action. Either way, workload will increase to cope with the system malfunctioning or ineffective control actions.

B. Trust in automation

Human-machine trust, also referred as trust in automation, is defined as the progressive expectation of reliability in an automated system experiencing foreseeable changes due to interaction with it [17]. When an operator's trust in automation for a certain system is not calibrated, negative consequences can occur in the form of overtrust and distrust. Overtrust occurs when operator's trust in automation exceeds capabilities of the automated system. It can result in a loss of SA and out-of-the-loop unfamiliarity (OOTLUF) [18]. For example, an ATCO may overestimate the automated conflict detection & resolution (CD&R) function of the automated system, and lose track of current traffic status. When the automated CD&R ceases to work, the ATCO will have difficulties taking over. This is also referred as misuse of automation, which arises from a lack of monitoring by the human operator, which results in neglecting automation failures [19]. Misuse of automation is likely to occur when failure frequency is low, which can be the case in a supervisory ATC environment.

Research on factors influencing monitoring of automation [20] involved an experiment where participants were required to detect occasional automation failures by identifying malfunctions not detected by the automation [21]. It was found that participants were able to detect more than 70% of malfunctions when they performed manual control tasks, however, when the monitoring task was under automation control, the malfunction detection rate was markedly reduced [21]. Thus, it can be said that the over-reliance of automation can adversely affect supervisory performance of ATCOs supervising a fully automated ATC system, which can ultimately decrease the quality of manual control performances in presence of an automation failure.

On the other hand, distrust occurs when an operator underestimates the automation reliability and decides not to fully rely on it. Distrust prevents full benefits of the automated system to be utilized by the operator [18]. An ATCO may start questioning whether certain resolutions given to some aircraft by the automation are actually what the controller would have done, then perform manual CD&R even though the automated system would have handled the CD&R. Just like misuse, distrust is also referred as disuse of automation. Disuse is likely to improve manual control performances as operators' disuse of automation would make them more vigilant during the supervisory control, however, this is questionable for a prolonged supervisory control session with such a rare failure frequency. Considering that the ATCOs currently have shift work, meaning that they take over shifts of previous ATCOs, it is crucial to investigate the impacts of automation failure timings on trust in automation; the automated ATC system may fail more than once throughout a single shift of an ATCO, and the manual controller performance can be influenced by the automation failure timings. Wickens and Xu argued that a first automation failure leads to decreased trust of automation than subsequent failures [22]. Another research revealed that being exposed to unreliable automation yields reduction in complacency and increased performance after operators adjusted their trust in automation accordingly [23]. However, there is also evidence that so-called “first automation failure effects” are not apparent; instructions and practice trials that indicate a possibility of imperfect automation may decrease the consequences of controller performance under automation failure [24].

C. Situation awareness

While several definitions for SA exist, a definition states that SA has three levels that begin with the perception of the attributes in the environment for corresponding space and time, followed by the understanding of their meaning, and evolved to the projection of the status in the future [25]. The SA model by Endsley can be seen in Fig. 1.

Highly automated systems can directly impact SA through monitoring vigilance and complacency changes and the shift from a proactive controller to an uninvolved supervisor [27]. The changes in vigilance and complacency in monitoring are closely related to aforementioned trust in automation issues. Associated with the trust in automation, operators may decide to neglect the automation and system variables controlled by the automation, by allocating attention to other tasks, which yields low SA on the automation and system itself [28]. For the issue of assuming a passive role, there are evidences that being passive in information processing can yield worse manual control performances than actively processing information [29]. Another study revealed that subjects’ SA decreased for full automation condition compared to manual control environment for navigating an automobile [27]; their Level 2 SA was adversely affected while Level 1 SA was not affected. This means that they comprehended basic elements of the environment, but did not understand what meanings the elements had towards system goals. Endsley (1996) stated that transforming a controller into a supervisor can negatively affect SA, and taking over during automation failure can be problematic.

It is crucial to consider SA separately from the decision-making and performance stage; trained and experience operators can easily make incorrect decisions if their SA is incomplete or inaccurate, and inexperienced controllers may have acute understanding of what is happening in the environment but not know which action to take [25]. Level 1 SA corresponds to recognizing the status, attributes, and dynamics of components in the environment. ATCOs must perceive all aircraft in a responsible sector and their attributes such as aircraft ID, airspeed, heading, etc. Level 2 SA is based on integration of Level 1 SA elements, which form a comprehensive image of environment to be realized when put together [25]; Level 2 SA elements for ATC can be the impact of changing airspeed/heading of an aircraft on another aircraft, or determination of aircraft position deviations. Finally, Level 3 SA requires the ability to project near-future actions by combining the Level 1 and Level 2 SA elements. For ATC, Level 3 SA can include determining airspace capacity and availability for making a certain routing decision, and potential conflicts that may arise from such decision.

III. Task Engagement Tool design

The Task Engagement Tool (TET) is essentially an online task assistant tool in the form of a secondary task test; its aim is to increase task engagement of ATCOs by asking them task-related SA questions that the answers to the questions could be found on the plan view display (PVD) throughout their supervisory control phase, which would cycle the controllers’ SA, bridge the gap caused by the transition from supervisory control to manual control in presence of an automation failure, and allow them to perform better manual ATC. This section presents the design of the TET based on three levels of SA. System boundary was first determined to limit the scope of this research, and en-route controller task analysis was performed to connect the tasks to the three level SA questions to be constructed. Then, details of the three level SA questions used for the experiment are elaborated.

A. System boundary determination

To investigate controller performance based on aforementioned ironies of automation and trust in automation issues, upper control area (UTA) is chosen for this research. There are relatively less activities in UTA compared to lower airspace, and UTA could be the easiest airspace to automate in real life if full automation is to be achieved in ATC. Also, the less activities in UTA allowed us to assume a single flight level for the experiment.

B. En-route controller task and strategy analysis

From Integrated Task and Job Analysis of En-route ATCOs [30], five task groups are recognized as following:

1) Maintain SA: ATCOs must have ‘the picture’ to maintain SA; a mental traffic picture must be established, followed by continuous projection into future and checking of anticipation with the actual traffic status.
2) Develop and receive sector control plan: integration of flight progress information and traffic forecast into a future traffic situation plan needs to be performed to ensure a safe and efficient traffic flow.
3) Make a decision for control actions: developing and revising the sector control plan is closely related with active decision-making of control actions.
4) Solve aircraft conflicts: provision of separation and resolution of aircraft conflicts are included in this task group.
5) Provide tactical ATM: accepting an aircraft and performing handover to adjacent sectors are the main tasks in this task group.

In conjunction with the en-route controller task group categorization, some of controllers’ best practices directly related to workload and SA of ATCOs. For conflict detection, controllers typically keep look-ahead time of 5-10 minutes [31] and deploy different reaction attitude based on workload; on low workload, they tend to wait and see before taking

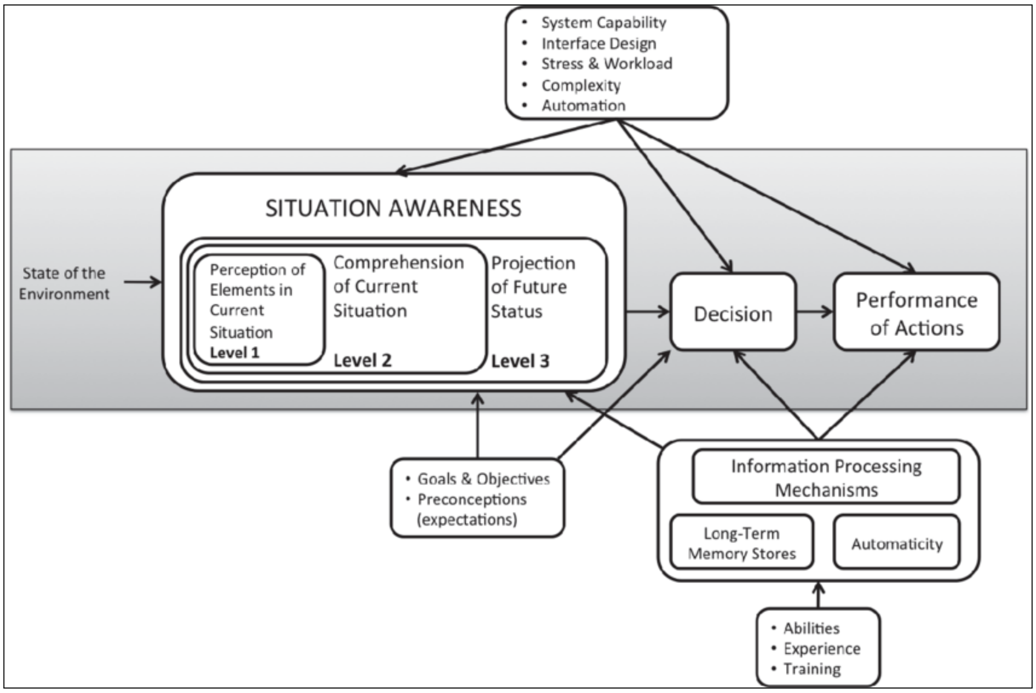


Fig. 1. Endsley's model of SA [26]

action, while they become more conservative in classifying conflicts and act immediately after detecting a conflict in high workload conditions [30], [32]. For conflict resolution, ATCOs select resolutions that requires the least monitoring and coordination, while minimizing the number of aircraft to move and additional track miles flown [33], [31].

C. Controller-automation task distribution

Specific tasks are assigned to human operators and automation for this research. As the research is concerned with the transition from supervisory control to manual control of a fully automated ATC system, a human operator is tasked with maintaining his/her SA and supervising the performance of automation for resolving conflicts and providing tactical route clearances. The automation will be responsible for the following tasks:

1) Information acquisition and integration: the automation will acquire flight information (i.e., airspeed, heading, destination, aircraft type) and project the integrated information (i.e., live position update per simulation tick) on a PVD.
2) Conflict detection: converging aircraft pairs, crossing flights, and their heading & airspeed will be checked.
3) Conflict resolution: the automation will issue heading and/or airspeed commands to aircraft, revise and organize traffic patterns and routes.
4) Tactical ATM provision: handover will be performed automatically.
5) Provision of sector clearances: automation will direct aircraft to their exit

When the fully automated system fails, the automation will only be functional in acquiring flight information, showing it on the PVD and performing automated handover. Then the human operator must take over the remaining tasks of CD&R and directing aircraft to their exit waypoint as well as own SA updating.

D. Construction of SA questions

SA information requirements for en-route ATC provides a list of controllers' major SA requirements for dynamic information broken down to the three levels of SA [25]. While the SA information requirements for en-route ATC provided an extensive library of SA elements to be considered, the performed task analysis allowed us to connect the elements together into relevant three level SA questions. The three levels are broken down as following based on the SA information requirements for en-route ATCOs [25], and combined with the en-route ATCO task analysis, as illustrated in Fig. 2.

IV. Experiment design

A. Participants

Eight participants volunteered in the experiment. All participants are students at Faculty of Aerospace Engineering, Delft University of Technology. Due to availability of volunteers who were experienced with ATC/ATM, only half of the participants had extensive knowledge or experience within the ATC domain (working on ATC/ATM related thesis researches), while the other half had a minimum degree of ATC knowledge obtained from some courses. None of them had professional ATC experience. The summary of the participant background information is shown in Table I. Given that the nature of this experiment is to expose participants to an extreme supervisory control environment before abruptly pushing them to

Level 1

Aircraft

- Aircraft ID, CID, beacon code
- Current route (position, heading, airspeed)
- Current flight plan (destination, filed plan)
- Aircraft capabilities (cruising speed, max/min speed)
- Aircraft type (medium, heavy)
- Aircraft status (on-course)

Level 2

Conformance

- Amount of deviation (airspeed, route)
- Time until aircraft reaches assigned speed, route/heading

Current Separation

- Amount of separation between aircraft along route
- Deviation between separation and prescribed limits
- Number/timing aircraft on route

Timing & deviation

- Projected time in airspace
- Projected time till clear of airspace
- Order/sequencing of aircraft
- Deviation of aircraft/flight plan

Level 3

Projected Aircraft Route (current and potential)

- Position, flight plan, destination, heading, route, airspeed
- Projected position x at time t

Projected Separation

- Amount of separation along route
- Deviation between separation and prescribed limits
- Relative projected aircraft routes
- Relative timing along route

Impact of Potential Route Changes

- Type of change required
- Time and distance till turn aircraft
- Amount of turn /new heading change required
- Aircraft ability to make change
- Projected number of changes necessary
- Cost/benefit of new clearance
- Impact of proposed change on:
  - Aircraft separation
  - Traffic flow
  - Number of potential conflicts
  - Flow requirement
  - Workload required

Fig. 2. Selected three levels of SA information requirements for en-route ATC to construct SA questions [25]

TABLE I
PARTICIPANT INFORMATION

Profile	6 M.Sc. students, 1 Ph.D. candidate, 1 Ph.D. student
Age	23-35 (average 27)

perform manual control, no specific experience was required, although knowledge and insight on how general ATC is done (i.e., controllers best practices) would benefit participants with performing manual control tasks. In order to minimize the differences between the experienced and inexperienced participants, additional verbal instructions on controllers best practices (i.e., maintain certain look ahead time, vector slower aircraft behind for a crossing conflict) were given during training runs.

B. Tasks and instructions

As illustrated in Fig. 3, the control task of the participants during the supervisory control phase was to supervise a fully automated ATC system that performed CD&R, cleared aircraft to their exit waypoints and transferred them to the adjacent sector, and answer SA questions given by the TET. One SA question was given to participants every 100 seconds after the first minute of the simulation, and remained open for 30 s. During the expiration period, participants needed to read the question, check the PVD, and indicate their response by either clicking yes or no button on a TET window (see Fig. 3). After they answered the question, they could immediately see whether their response was correct or not, by checking a numerical counter below the yes/no buttons. If a response was not given within the time window of 30 s, it counted as incorrect.

During the supervisory control task, each aircraft could be examined by a flight label which contained an aircraft ID, heading & airspeed, destination in the form of an exit waypoint name, and aircraft type. Aircraft could not be controlled manually while the automation was on. However, when automation was switched off at a chosen moment, all active automated CD&R commands were disabled and the participants had direct control of all aircraft within a given sector. They could issue heading commands by clicking aircraft and dragging its speed vector around the selected aircraft, and airspeed commands could be issued by also clicking aircraft and either scrolling up or down to increase or decrease airspeed of the selected aircraft, and were confirmed by hitting the ENTER key. For the run with an automation failure, the TET was switched off and no more questions were given to the participants as they performed manual CD&R tasks.

In order to allow participants to experience boredom and vigilance issues coming from prolonged supervisory control sessions, duration of experiment runs was not presented on the simulation display such that they could not change their attitude on coping with the vigilance issues by knowing when each experiment would end. Also, they were told that an automation failure may happen with no assurance. They knew that they may indeed have to perform manual CD&R tasks, however, they were not aware of exact timings of a rare automation failure for their experiment runs.

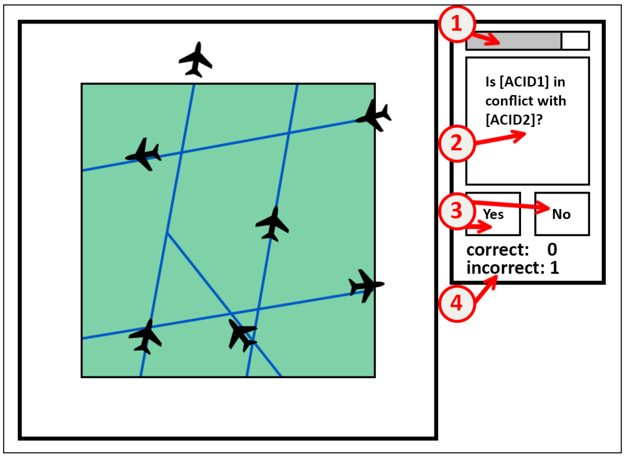


Fig. 3. The artist impression of the TET with an example question, along with a simulator screen with a sector on it. ①: = time bar, ② = TET SA question window, ③ = yes/no button, ④ = instant score counter

C. *Independent variables*

In the experiment, two independent variables were defined as following:

1) SA level of questions given by the TET, with levels 'SA1-2 (low)' and 'SA2-3 (high)' (within participants),
2) Automation failure timings, having 'early' and 'late' (within participants).

Based on the three level SA model, SA1-2 was defined as a combination of Level 1 and Level 2 SA, while SA2-3 level SA was defined as a combination of Level 2 and Level 3 SA. The rationale for splitting the SA levels in such way was to prevent intentional manipulations of experiment outcomes regarding controller SA during the supervisory control; if SA1-2 only contained Level 1 SA questions and SA2-3 contained Level 2 and 3 SA questions, the level differences of SA feedback would be too great. For instance, if SA1-2 condition only provided participants with questions regarding low level information such as current airspeed and heading, the TET would naturally be disturbing participants with such questions. At the same time, if participants were getting higher level SA questions regarding projection onto near-future status, it would be easily expected that participants have higher SA compared to the them getting Level 1 SA questions only.

The automation failure timings were varied based on a period until an automation failure occurred, as illustrated in Fig. 4. The 'early' failure timing corresponded to having a short period before experiencing an automation failure; as the supervisory control phase (indicated as AUTO in the status row in Fig. 4) lasted 7 minutes, an early failure occurred after a single block of supervisory control phase. The late failure occurred after prolonged supervisory control phases of 29 minutes, considering that uninterrupted attention span of average human adults can go up to approximately 20 minutes [34]. It can be seen that for the second set of experiment runs, the location of the early failure has shifted behind a late failure, which still resulted in 29 minutes of supervisory control for the late failure scenario, and 7 minutes of supervisory control for the early failure scenario.

Failure timing	EARLY						LATE	
SA level	low SA (1,2)		low SA (1,2)		low SA (1,2)		low SA (1,2)	
scene #	scene 1A (10 min)		scene 2B (10 min)		scene 3A (10 min)		scene 4B (10 min)	
status	AUTO	MANUAL	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL
duration [min]	7	4	7	4	7	4	7	4
	break + questionnaires							
Failure timing					LATE		EARLY	
SA level	high SA (2,3)		high SA (2,3)		high SA (2,3)		high SA (2,3)	
scene #	scene 2A (10 min)		scene 3B (10 min)		scene 1B (10 min)		scene 4A (10 min)	
status	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL	AUTO	MANUAL
duration [min]	7	4	7	4	7	4	7	4
	questionnaires							

Fig. 4. An experiment matrix for one participant, which indicates two independent variables for the experiment

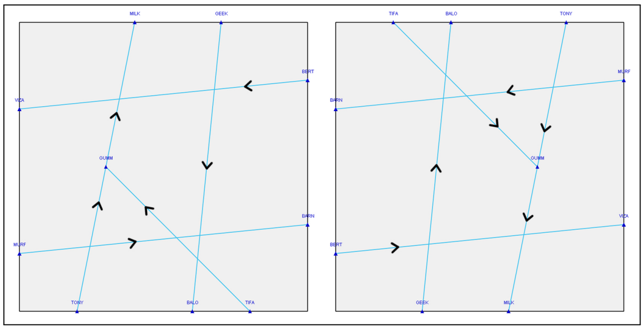


Fig. 5. 0° and 180° sectors used in the simulator

D. *Traffic scenarios and automation*

The sector used for all scenarios had two variants of a single base scenario; they shared identical traffic routes, but one was rotated over 180° in order to prevent participants from noticing similarities between individual experiment run. A combination of the 0° sector and 180°sector was given to each participant to prevent participants from noticing similarity between each scenario. Also, waypoint names and aircraft ID were modified for each scenario. The sector featured crossing points close to each corner of the sector, and had a concentrated area of crossing points near the area where two airways merge into one, as illustrated in Fig. 5. The two variants were then duplicated into Scenario 1, 2, 3, and 4; each scenario contained 30 aircraft in total, of which average 12 aircraft were present at any given moment. Each scenario had 10 conflicts within the 11 minute run, and the runs with the automation failure had 4 conflicts after the automated CD&R ceased to work until the end of the runs. Each scenario ended with the same traffic complexity as it had when the run started.

To test multiple experiment runs per participant and ensure that there would be fast enough activities to interact with during the manual control phase, the simulation ran twice as fast as the real time. This yielded a real time scenario of 1,320 s, which lasted for 660 s in the simulation. It was selected in such way that a SA question could be given every 105 seconds, which would give 6 SA questions on a run without an automation failure, and 4 SA questions on a run with an automation failure.

E. *Task Engagement Tool questions*

For the four scenario variants with the SA level manipulation, the questions were assigned as following. Note that [N] corresponds to a number, [ACID] corresponds to an aircraft ID, and [WAYPOINT] corresponds to a waypoint name. Each attributes were carefully randomized for each scenario. Examples of the SA questions are shown as following:

- Level 1: Did the sector have [N] crossing points?
- Level 1: Is [ACID] a heavy aircraft?
- Level 2: Did [ACID] receive a heading change and a direct-to command?
- Level 2: Are there [N] aircraft on the route [WAYPOINT]-[WAYPOINT]?
- Level 3: Does [ACID 1] need to take over [ACID 2] to avoid a further conflict with [ACID 3]?
- Level 3: Will it take approximately [N] seconds before a conflict with [ACID 1] and [ACID 2] becomes critical?

F. Control variables

The control variables are shown as following:

- *Degree of freedom* All aircraft were flying on the same altitude of flight level 290, which resulted in a 2D control task on the horizontal plane only. The simplification ensured that results between participants would be more comparable, as they could only change heading and airspeed of aircraft when the automation failed.
- *Aircraft type* All aircraft were either medium- or heavy-type aircraft. The medium type had indicated airspeed (IAS) envelope of 200 kts - 290 kts, and the heavy type had IAS envelope of 230 kts - 350 kts.
- *Aircraft count* All scenarios had 12 aircraft simultaneously inside the sector at any given moment.
- *Automation reliability* Automation reliability was defined as the percentage time of perfectly functioning ATC automation in 88 minutes of the experiment. 82% of automation reliability was chosen to assess the effects of different automation failure timings while ensuring that there would be sufficient time for supervisory control characteristics to be excited.
- *Task engagement tool questions* the SA level ratio of the questions was kept 1:1, meaning that the low SA level runs gave participants three level 1 SA questions and three level 2 SA questions, while the high SA level runs provided three level 2 SA questions and three level 3 SA questions. For the runs with the automation failure gave 2 questions for each level, as there were four questions in total. All questions had an expiration time of 30 s.
- *Scripted automation* the automation was scripted by an experienced person who was familiar with the simulator and underwent a five-day area control center course at National Aerospace Laboratory (NLR).

G. Dependent measures

The dependent measures were as following:

- *TET SA question scores* measured if participants' responses were correct or incorrect, and the counter for the correct and incorrect responses was shown to participants.
- *TET SA question response time* measured the time between the moments when a question was displayed on the TET window and participants gave their response by pressing either yes or no button within the 30-s expiration window.
- *The number of heading and airspeed commands* measured the number of commands given during the manual control phase for the scenarios with the automation failure.
- *Short Term Collision Avoidance (STCA) ratio* measured the ratio of the duration of activated STCA and all logpoints of the scenarios with the automation failure.
- *Average minimum separation distance* measured the average minimum separation distance during the manual control phase for the scenarios with the automation failure.
- *Average track deviation* measured the average of additional aircraft tracks due to deviation during the manual control phase for the scenarios with the automation failure.
- *Workload ratings* measured the overall subjective workload for each scenario and were measured with a slider bar with a scale of 0-100 every 2 minutes of each scenario.
- *Trust in automation ratings* measured overall trust in automation by asking participants to fill in SHAPE Automation Trust Index (SATI) [35] after each block of runs (twice in total). The trust in automation was then investigated per question.
- *Situation Awareness for SHAPE (SASHA)* [36] measured overall SA by asking participants to fill in SASHA_Q questionnaires after each block of runs (twice in total). SA was assessed per question.

H. Procedure

Few days before the experiment, participants received a pre-experiment briefing which contained the information regarding tasks to be performed. On the day of the experiment, the experiment started with a short briefing regarding an overview of ATCO tasks, followed by five training scenarios which lasted for half an hour in total. Complexity for training scenarios gradually increased from the basic level of controlling a single aircraft pair in a crossing conflict to the level of the actual experiment scenarios where they had to first supervise a sector with multiple airway crossings for five minutes, then perform manual control for another five minutes; they had chances to supervise the fully automated ATC system and experience the automation failure, which led to the manual control phase to detect and resolve conflicts themselves.

After the briefing and training, participants did four experiment runs of 11 minutes each. A short break of 5 minutes was then given, along with SASHA and SATI questionnaires. Participants did another set of four experiment runs of 11 minutes each again, followed by SASHA and SATI questionnaires. After the experiment was over, a short debriefing was given and participants filled in final questionnaires to provide feedback about the TET. The experiment took 2.5 hours per participant in total.

I. Hypotheses

It was hypothesized that being asked higher levels of SA questions during the supervisory control phase before an automation failure would result in: (1) increased overall workload but decreased workload difference around the transition from supervisory to manual control, (2) worse TET scores and higher response time, (3) higher average minimum separation distance, (4) lower STCA ratio, (5) lower average track deviation, (6) lower number of commands needed to resolve conflicts and (7) higher overall SA.

The main rationale for the hypotheses was that having higher levels of SA would be the most important when taking over control when automation fails, which would result in increased manual control performances in the form of the aforementioned performance criterion. To maintain the higher SA during the supervisory control, participants would have to answer more difficult SA2-3 questions compared to SA1-2 questions, which would raise their overall workload as they supervise the automated system. As workload would be raised already, a sharp workload peak upon the transition after the automation failure would be less.

Also, it was hypothesized that experiencing an early automation failure would result in (1) lower trust in automation, (2) increased manual control performances in terms of average minimum separation distance, STCA ratio, average track deviation and number of commands and (3) higher SA. The main rationale for the hypotheses was that an early automation failure would decrease participants' trust in automation and raise their vigilance to actively supervise the system until it failed, which would yield increased manual control performances.

V. Results

A. Workload

The workload ratings measured by the Instantaneous Self Assessment (ISA) method are shown in Fig. 6. The ISA workload ratings were measured every 2 minutes, which gave 5 workload ratings per experiment run. Raw workload ratings from 0 to 100 were normalized per participant, and z-scoring was applied to the normalized workload ratings. From two-way ANOVA, it was revealed that the TET SA manipulation (SA1-2 vs. SA2-3) had a significant effect ($F(1,7) = 6.784$, $p = 0.035$) at 8 minutes, but also showed no significant interactions of TET SA level and automation failure timing manipulations; the automation failure occurred at 420 s (7 min), thus the ISA workload ratings at 8 min reflects the participants' workload after the failure. This follows but also goes against to the hypotheses. A Friedman test revealed a significant effect of the within-participant manipulation ($\chi^2(3) = 12.570$, $p = 0.016$), where pairwise comparisons between 'SA1-2 late failure' and 'SA2-3 late failure' conditions, and between 'SA2-3 early failure' and 'SA2-3 late failure' conditions.

B. Task Engagement Tool metrics

TET metrics were defined as the number of correct responses (maximum 4 for each run with the automation failure) and the response time in seconds, which can be seen in Fig. 7 and Fig. 8. Virtually there was no difference between the different TET SA levels, and a Friedman test on the within-participant manipulations revealed that the TET SA level and automation failure timing manipulations did not have a significant effect ($\chi^2(3) = 1.453$, $p = 0.693$). The results can be explained by the observed behaviors and feedback, in which participants felt that some of SA2-3 questions were vague and ended up guessing as the chance of getting a correct response was 50%. Thus, the number of correct responses cannot be considered as a metric to be connected with the participants' SA and performance, even though participants may have correctly updated their mental pictures based on the correct/incorrect counter on the TET window.

For the average response time for the TET SA questions, it can be observed that participants spent more time answering SA2-3 questions than SA1-2 questions. Repeated measures ANOVA revealed that the TET SA level manipulation had a significant effect ($F(1,7) = 7.824$, $p = 0.027$), while the automation failure timing manipulation and the interaction of two manipulations did not have a significant effect.

C. Average minimum separation distance

As a part of the safety criterion, the average minimum separation distance was recorded for each experiment run. The average minimum separation distance for the TET SA level and automation failure timing manipulations is illustrated in Fig. 9. Repeated measures ANOVA revealed that only the automation failure timing manipulation had a significant effect on the average minimum separation distance ($F(1,7) = 12.144$, $p = 0.01$), while the TET SA level manipulation and the interaction of two manipulations were found to be insignificant.

It can be observed that early the average minimum separation distance was higher with the early failure for SA1-2, while it was higher with the late failure for SA2-3. This result ran counter to the hypotheses, as lower average minimum separation distance was expected for the SA2-3 late failure scenario.

D. STCA ratio

As a part of the safety criterion, the STCA ratio was calculated based on simulation logpoints that had STCA activated over the total logpoints for each experiment run. As shown in Fig. 10, the differences between each manipulation are minimal. A Friedman test reported that the manipulations were not significant ($\chi^2(3) = 4.219$, $p = 0.239$).

Lower STCA ratio for the late automation failure timing manipulation did follow the hypothesis despite the insignificant results. The reason why the STCA ratio for the different manipulations did not differ to a significant magnitude could be that each scenario had the same number of aircraft evenly distributed in the sector, and most participants responded to conflict pairs before STCA was triggered. As conflict pairs would first turn amber before the STCA would turn the pairs red, the visual indication immediately grabbed participants' attention for issuing a resolution.

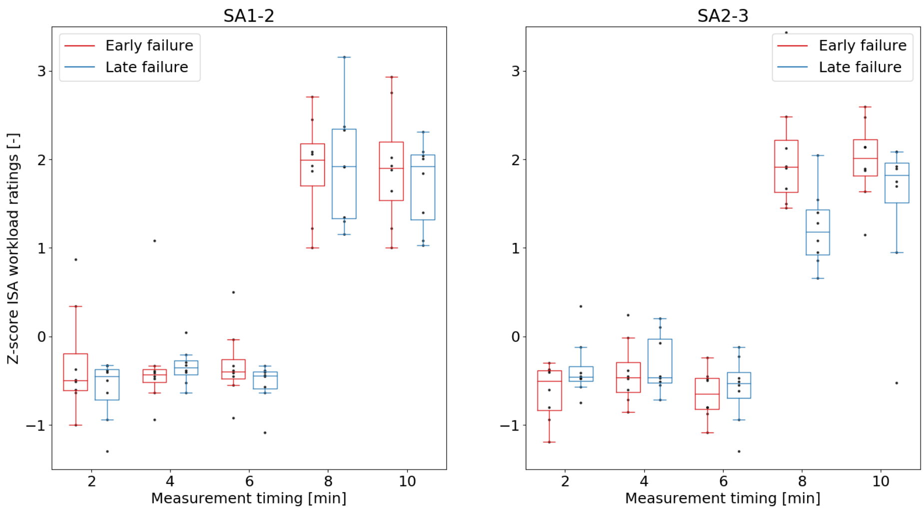


Fig. 6. Z-score of ISA workload ratings for TET SA levels of SA1-2 and SA2-3 and automation failure timings of early and late, measured every 2 minutes of each experiment run

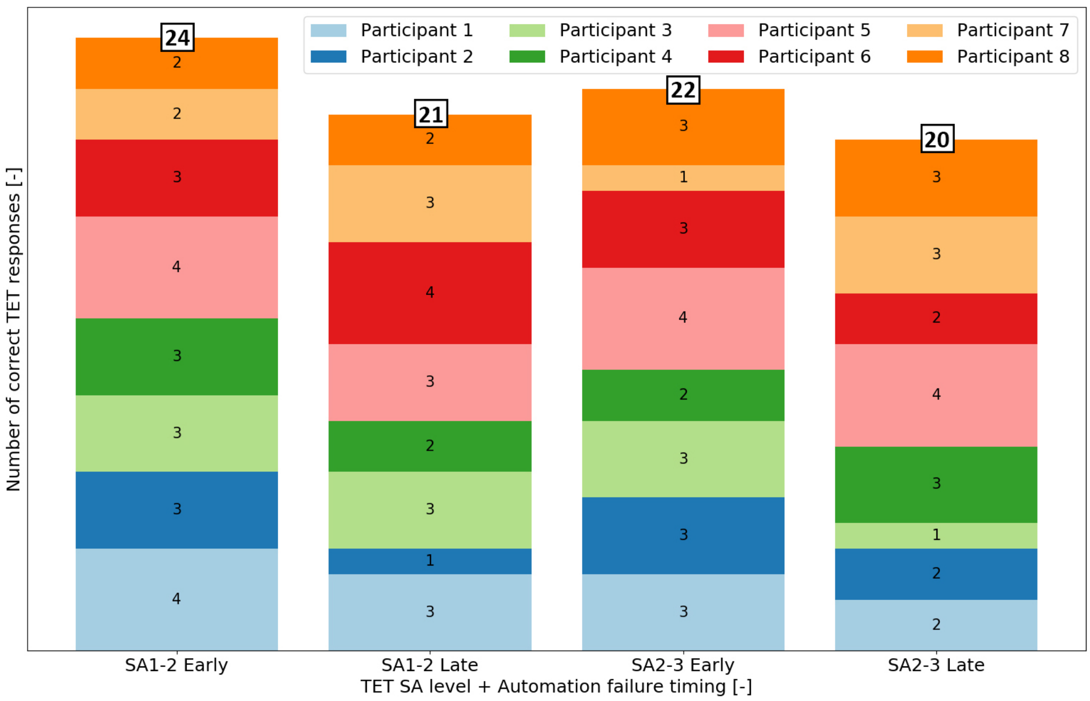


Fig. 7. The number of correct responses to the TET SA questions, for different TET SA levels (SA1-2 vs. SA2-3) and automation failure timings (early vs. late) per participant

E. Average track deviation

Control efficiency criterion include the average track deviation from original direct routes for all aircraft in the sector. Fig. 11 illustrates the average track deviation in NM for both the TET SA level and automation failure timing manipulations. It can be observed that participants deviated aircraft from original routes more for the late automation failure scenario. This may been caused by the vigilance issue for experiencing a prolonged supervisory control session before the late failure occurred; the observation on participants showed that they were less vigilant in terms of monitoring and preparing for a take-over for the late failure scenario, and had to issue inefficient heading and airspeed changes briefly after the automation failure, which led to greater track deviation distance. However, repeated measures ANOVA revealed that the failure timing manipulation was not significant to the results. The TET SA

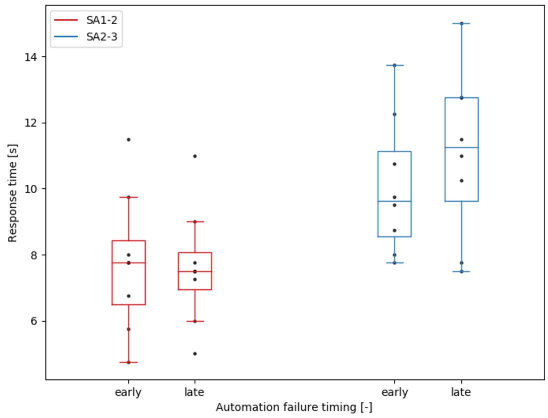


Fig. 8. Average time spent to answer TET questions for the SA levels and failure timings

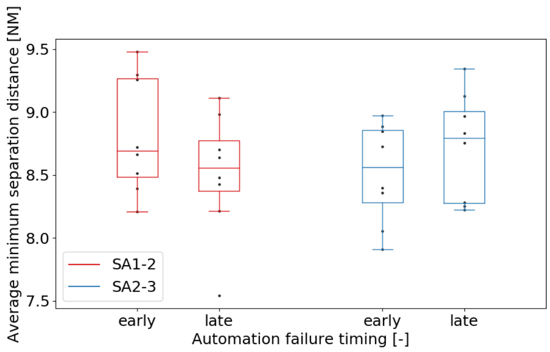


Fig. 9. Average minimum separation distance between all aircraft for TET SA and failure timing manipulations

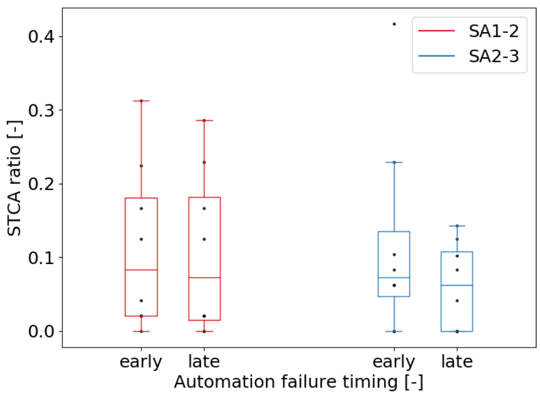


Fig. 10. STCA ratio throughout experiment runs for TET SA and failure timing manipulations

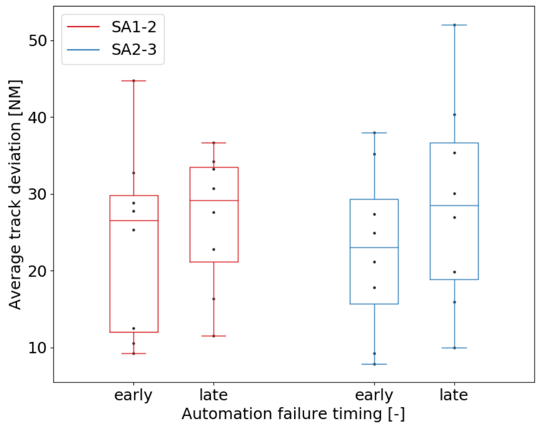


Fig. 11. Average track deviation from a direct route of each aircraft in the sector

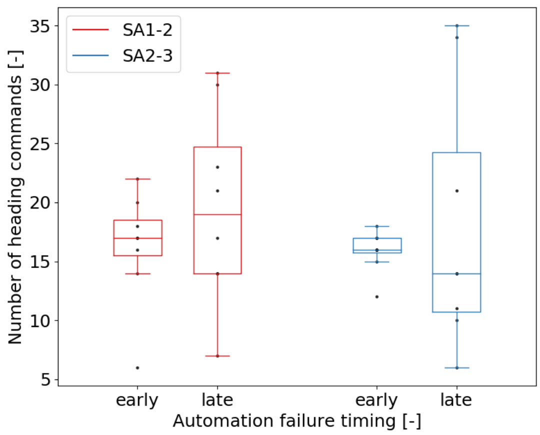


Fig. 12. The number of heading commands issued to aircraft during the manual control phase

level manipulation and the interaction of the two manipulations was also not significant.

F. Number of heading and airspeed commands

The number of heading and airspeed commands given to aircraft after the automated system failed were measured as a control efficiency criteria. The TET SA level and automation failure timing manipulations did not have a significant effect on the number of heading commands. It can be observed that the number of heading commands for the late failure scenario had a greater standard deviation than the early failure counterpart, which is illustrated in Fig. 12. The results could be from participants' boredom and complacency issues after being exposed to the prolonged supervisory control phase (28 minutes) for the late failure scenario; participants who already had experiences with the simulator or worked on ATC/ATM thesis researches had already developed strategies that decreased the number of commands given to aircraft, while inexperienced participants failed to develop an efficient strategy to perform the manual conflict resolution. Both manipulations and their interaction did not have a significant effect on the number of the speed commands given to aircraft during the manual control phase. The number of airspeed commands for the failure timing variations did not show significant difference in standard deviation, which is shown in Fig. 13.

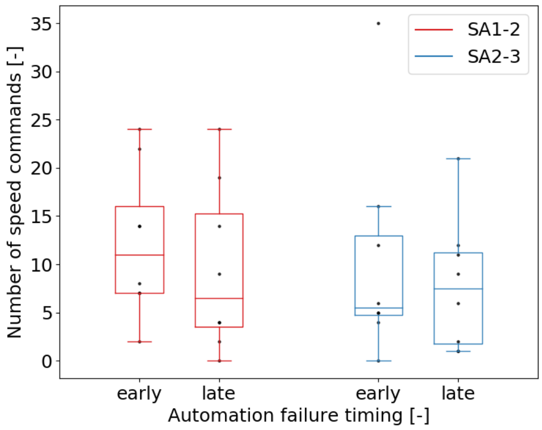


Fig. 13. The number of airspeed commands issued to aircraft during the manual control phase

G. Situation awareness

SA was assessed by comparing raw scores of SASHA questionnaires for the TET SA level manipulation, which can be seen in Fig. 14. A Friedman test was performed for each question, and revealed that the TET SA level manipulation was not significant. However, it can be observed that participants felt that they were able to foresee the traffic evolution more for the SA2-3 condition, while SA2-3 did not quite help them planning their control tasks; the reason why they felt that planning of control tasks was not affected could be that the TET was only available during the supervisory control phase, and questions they received may have become irrelevant during the manual control phase due to the fact that some of aircraft already left the sector or became clear of conflict. Participants found themselves focusing more on a single element on the PVD for SA2-3; participants spent more time answering SA2-3 questions than SA1-2 questions. Participants felt that SA2-3 questions were actually not as useful as SA1-2 questions, as they mentioned that some questions were simply too trivial (mainly SA1-2) or difficult to understand and answer (SA2-3). Experienced participants responded that questions regarding conflict pairs did help them improve their SA. For allocating too much attention in interacting with the TET, the responses varied but the average stayed closer to ‘never’. Finally, it can be observed that SA2-3 did assist participants with obtaining a better understand of the traffic situation than SA1-2. Trust in automation may have been interfering with the interest in knowing the current traffic situation, regardless of whether questions are being asked or not. Thus, for a very reliable system, it is likely that a task assistant tool such as the TET can be deemed as not helpful.

H. Trust in automation

Trust in automation was assessed in the same way as SA. The results of trust in the automated CD&R system questionnaires are shown in Fig. 15. A Friedman test on all questions revealed that the automation failure sequence timing manipulation did not have a significant effect on the trust. Participants found the automated CD&R very useful, and also responded that it is highly reliable; for the reliability, participants had higher trust in the automated CD&R, as expected that experiencing an early failure first would lower trust in automation. Their assessment in the accuracy followed the same trend. Understanding the automated system was equal for both failure sequences. Participants did not like monitoring the automated system overall, however, preferred monitoring it for the late-early failure sequence; this could be also from having higher trust in automation for the late failure. In terms of supervision difficulty, there was no virtual difference between the failure sequence manipulations.

The results of trust in the TET can be found in Fig. 16. A Friedman test on the responses revealed that the automation failure sequence timing manipulation did not have a significant effect on the trust. In terms of usefulness, participants found the TET more useful when they experienced an early failure first, followed by a late failure. This could be explained by an observation during the experiment that participants were more actively finding answers to the TET questions after experiencing an early failure, compared to the late failure counterpart. However, they also responded that the TET annoyed them more when they experienced an early failure compared to a late failure. It would be the case that they were maintaining vigilance and actively supervising the system, and the TET interaction annoyed them. This also explains the fourth question regarding distraction, as participants’ attitude towards supervising the system would be more passive, and the occasional interaction with the TET would come across as distraction from the rapid attention reallocation. At last, participants found it easier to answer the TET questions for the early-late failure sequence. This is also connected with the aforementioned vigilance status, that participants were more active in building their mental pictures while supervising the system. Thus, answering the SA questions was easier then.

VI. Discussion

The goal of this research was to investigate the impact of increasing the task engagement of controllers during a supervisory ATC task by asking them task-related questions in presence of an automation failure. The primary focus was on maintaining situation awareness while supervising a fully automated air traffic control system and performing manual conflict detection and resolution after the automated system ceased to work. Based on previous researches on situation awareness and trust in automation, it was assumed that keeping situation awareness high during the supervisory control would lead to a smooth transition to a manual air traffic control environment and make manual control safer and more efficient.

The results showed that keeping high situation awareness by providing controllers task-relevant questions during the supervisory control did affect the transition from the supervisory to manual control when an automation failure occurred; for example, the sharp rise of workload briefly after the automation failure was reduced when participants received higher levels of situation awareness questions during the supervisory control phase. It was expected that the rise of participants’ workload

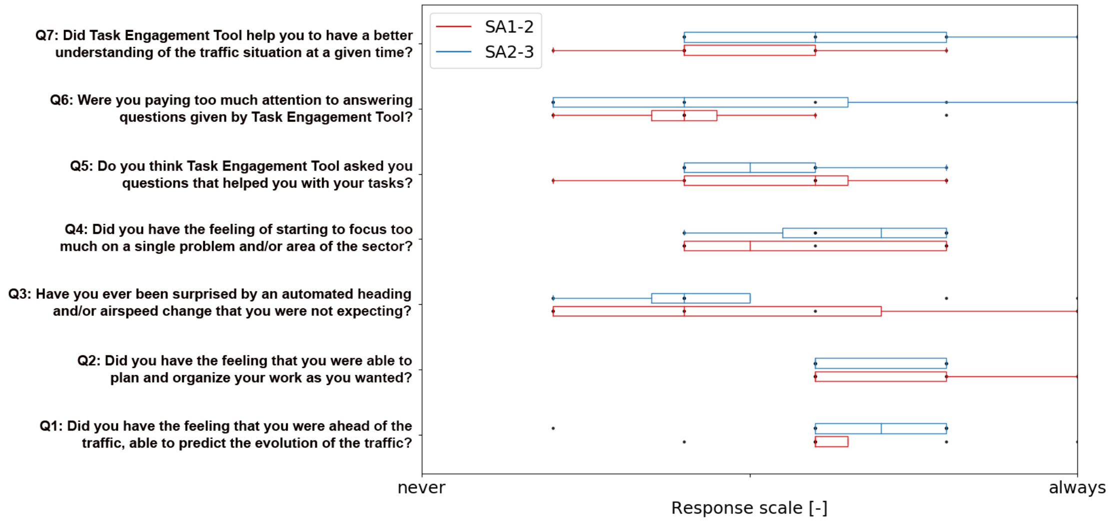


Fig. 14. Responses to post-experiment SASHA questionnaires

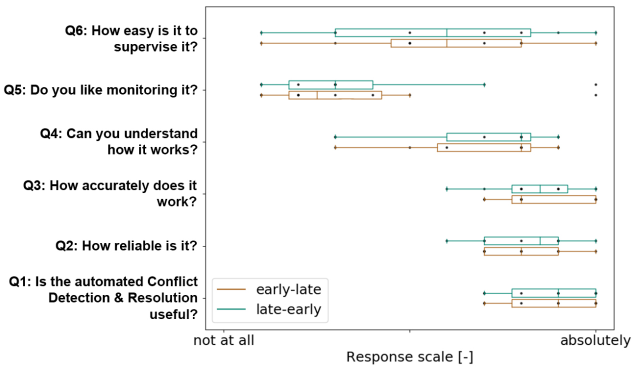


Fig. 15. Responses to post-experiment SATI questionnaires on trust in the automated CD&R system during the supervisory control phase until the automation failure occurred

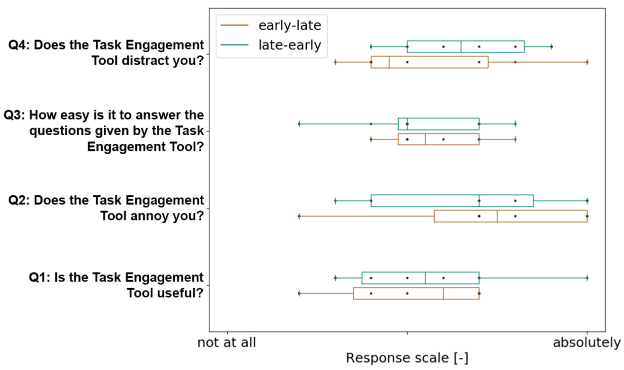


Fig. 16. Responses to post-experiment SATI questionnaires on trust in the TET during the supervisory control phase until the automation failure occurred

for the late automation failure scenarios would be greater than the early automation failure counterparts, however, it was instead observed that the participants' workload for late automation failure scenarios was actually lower. In hindsight, this was caused from history and maturation effects that can jeopardize internal validity; each experiment run shared the same principle of supervising a fully automated air traffic control environment and taking over to perform manual air traffic control tasks, on a same sector with predefined routes. The majority of participants had difficulties in the first two runs, but developed certain strategies to cope with the boredom and tried to improve their manual control performances when they had runs with the automation failure as the experiment proceeded. The experiment consisted of two blocks of four individual runs each, and experienced participants were able to anticipate the number of failure runs in the second block. Also, the length of the experiment can be questionable in hindsight; was it really long enough in time for boredom effects to start affecting participants' supervisory control? Even though there was maximum of 29 minutes of supervisory control before the automation failed, it is important to note that the mentioned period was not a single session. As each run of 11 minutes was paused before participants continued to next runs, accumulated boredom and fatigue may have been partially lost. Ensuring less runs with longer time period would be recommended for future work.

As the automated conflict resolution system was scripted, participants who had insight in general air traffic control/management initially questioned whether they could already perform manual conflict resolution themselves even when the automation was at work; they did not agree with some of the automated resolutions, and this resulted in higher vigilance during the supervisory control phase at the cost of reduced trust in automation and increased workload, which naturally led to better manual control performances. To improve vigilance, it can be said that allowing controllers to admit the imperfection of the automation can improve vigilance and manual control performance when a take-over is needed. However, exposing controllers to such inferior automation will limit capabilities of automation to increase, so that it cannot cope with increasing air traffic. A possible way is to use advanced augmentation such as the SSD; using the SSD for this experiment was not considered due to its extensive training requirements, but the SSD could have increased controller acceptance and monitoring performance, as participants could have been able to easily assess what the automation would do by checking integrated information given by the SSD. In hindsight, only participants with a certain degree of knowledge in air traffic control should have been chosen, although this was not possible at the time of conducting the experiment due to unavailability of experienced students. In combination of an extremely small sample size of 8 participants, this resulted in a big standard deviation in the majority of measured data.

Another factor that may have caused such deviation in the data could be varying scenarios. In order to minimize the history and maturation effects, four scenarios in each block had alternating scenarios in terms of the rotation angle. Thus, the two failure scenarios were always different such that one failure scenario was set at 0°, while the other failure scenario was rotated by 180°. Even though the traffic complexity was kept the same, a potential effect of the rotation cannot be completely ignored.

Regarding the Task Engagement Tool, four participants found some of the higher situation awareness questions ambiguous; for instance, a question asked, ‘will it take approximately 40 seconds before a conflict with aircraft 1 and aircraft 2 becomes critical?’ Even though the aim of the question was to give participants a basis of projecting elements onto future, participants mentioned that they felt the need of guessing instead of carefully integrating level 1 and level 2 situation awareness elements to determine the answer to the question. All participants responded that they were not paying too much attention answering questions of the Task Engagement Tool. In terms of the tool helping participants performing manual control tasks, participants felt that the tool indeed helped them with their manual control tasks when the tool brought their attention to specific aircraft pairs that would later be in conflict. On the other hand, lower situation awareness level questions were found to be too simple and not relevant to the control tasks.

For future work, the use of a decision support tool such as the SSD is recommended; with the SSD, more advanced task engagement method would be possible. For example, the researched task engagement method was to ask controllers task-related questions, and it certainly had a limitation of leaving a room for guessing answers, as they felt the necessity to answer all the questions by randomly guessing answers if they did not know the answer. If the task engagement method is directly linked with the SSD, it will become more interactive and have possibilities to increase vigilance and give controllers opportunities to hone their skills in using the SSD as well as performing manual control tasks.

VII. Conclusion

This paper presented the empirical investigation of increasing task engagement in a supervisory air traffic control environment, on achieving a smooth transition from supervisory to manual control upon an automation failure, by means of providing task-related situation awareness questions. Although a significant impact in lowering a sharp rise in workload upon transitioning from supervisory to manual control was observed, the experiment exposed that the selected method of increasing task engagement was not effective in improving safety and control efficiency. A plausible explanation is that the provided questions were not always bringing participants' attention to elements that participants considered helpful. This suggests a further research opportunity of utilizing situation awareness questions such that they are on the same line of attention related to controllers' primary tasks, and using an advanced decision support tool to further evolve task engagement method from simply answering questions to interacting more with the automation environment.

References

[1] M. R. Endsley, “From here to autonomy: Lessons learned from human-automation research,” *Human Factors*, vol. 59, no. 1, pp. 5–27, 2017.

[2] D. Woods and R. Cook, *Incidents: Markers of resilience or brittleness?* Ashgate, 2006, pp. 69-76.

[3] C. E. Bilings, “Human-centered aviation automation: Principles and guidelines,” NASA, Tech. Rep., 2 1996.

[4] M. R. Endsley, “Measurement of situation awareness in dynamic systems,” *Human Factors*, vol. 37, no. 1, pp. 65–84, 1995.

[5] M. R. Endsley and E. O. Kiris, “The out-of-the-loop performance problem and level of control in automation,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 2, pp. 381–394, 1995.

[6] A. Sethumadhavan, “Effects of automation types on air traffic controller situation awareness and performance,” *Human Factors and Ergonomics Society*, vol. 53, no. 1, pp. 1–5, 2009.

[7] T. Nguyen, C. P. Lim, N. D. Nguyen, L. Gordon-Brown, and S. Nahavandi, “A review of situation awareness assessment approaches in aviation environments,” Institute for Intelligent System Research and Innovation, Deakin University, Waurn Ponds, VIC, 3216, Australia, Tech. Rep., 2019.

[8] D. Chiappe, K. P. L. Vu, and T. Strybel, “Situation awareness in the nextgen air traffic management system,” *International Journal of Human-Computer Interaction*, vol. 28, no. 2, pp. 140–151, 2012.

[9] “European ATM master plan: executive view,” European Union, Eurocontrol, Tech. Rep., 2015.

[10] P. T. et al, “STRESS final project results report,” EU, Eurocontrol, Tech. Rep. 3, 6 2018.

[11] J. Kraemer, F. D. Crescenzio, B. Berberian, O. Ohneiser, and G. D. Flumeri, “MINIMA final project results report,” EU, Eurocontrol, Tech. Rep. 2, 5 2018.

[12] C. Borst, V. A. Bijsterbosch, M. M. van Paassen, and M. Mulder, “Ecological interface design: supporting fault diagnosis of automated device in a supervisory air traffic control task,” *Cogn Tech Work*, vol. 19, no. 4, pp. 545–560, 2017.

[13] K. Bibby, F. Margulies, J. Rijnsdorp, and R. Withers, “Man's role in control systems,” *IFAC Congress*, vol. 8, no. 1, pp. 664–683, 1975.

[14] B. Strauch, “Ironies of automation: Still unresolved after all these years,” *IEEE Transactions on Human-Machine Systems*, vol. 48, no. 5, pp. 419–433, 2018.

[15] L. Bainbridge, “Ironies of automation,” *Automatica*, vol. 19, no. 6, pp. 775–779, 1983.

[16] N. Mackworth, *Researches on the Measurement of Human Performance*. H.M.S.O London, 1950, pp 174-331.

[17] B. M. Muir, “Trust between humans and machines, and the design of decision aids,” *International Journal of Man-Machine Studies*, vol. 27, no. 5, pp. 527–539, 1987.

[18] J. D. Lee and K. A. See, “Trust in automation: Designing for appropriate reliance,” *Human Factors*, vol. 46, no. 1, pp. 50–80, 2004.

[19] A. D. Bronwyn French and A. Heathcote, “Trust in automation: A literature review,” University of Tasmania, Tech. Rep., 5 2018.

[20] R. Parasuraman, R. Molloy, and I. Singh, “Performance consequences of automation-induced “complacency”,” *International Journal of Aviation Psychology*, vol. 3, no. 1, pp. 1–23, 1993.

[21] R. Parasuraman and V. Riley, “Humans and automation: Use, misuse, disuse, abuse,” *Human Factors*, vol. 39, no. 2, pp. 230–253, 1997.

[22] C. D. Wickens and X. Xu, “Automation trust, reliability and attention hmi 02-03,” Aviation Human Factors Division, Institute of Aviation at University of Illinois, Tech. Rep., 10 2002.
[23] J. L. Merlo and C. D. Wickens, “Effect of reliability on cue effectiveness and display signaling,” *Proceedings of the 4th Annual Army Federated Laboratory Symposium*, pp. 27–31, 2000.
[24] C. D. Wickens, J. Helleberg, and X. Xu, “Pilot maneuver, choice, and workload in free flight,” *Human Factors*, vol. 44, no. 2, pp. 171–188, 2002.
[25] M. R. Endsley and M. D. Rodgers, “Situation awareness information requirements for en route air traffic control,” U.S. Department of Transportation, Federal Aviation Administration, Tech. Rep., 1 1995.
[26] J. Oxstrand, J. Ohara, and K. L. Blanc, “Development of an initial model of human- automation collaboration results from a needs analysis,” Idaho National Laboratory, Tech. Rep., 3 2013.
[27] M. R. Endsley, *Automation and human performance: Theory and applications*. Lawrence Erlbaum, 1996, pp. 163-181.
[28] R. Parasuraman, M. Mouloua, and R. Molloy, *Human performance in automated systems: Current research and trends*. Erlbaum, 1994, pp. 8-14.
[29] N. Cowan, “Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system,” *Psychological Bulletin*, vol. 104, no. 2, pp. 163–191, 1988.
[30] K. W. Kallus, D. van Damme, and A. Dittmann, “Integrated task and job analysis of air traffic controllers - phase 2: Task analysis of en-route controllers,” Eurocontrol, Tech. Rep. 1, 8 1999.
[31] B. Kirwan and M. Flynn, “Investigating air traffic controller conflict resolution strategies,” Eurocontrol, Tech. Rep. 1, 3 2014.
[32] J.-F. D’Arcy and P. S. D. Rocco, “Air traffic control specialist decision making and strategic planning - a field survey,” U.S. Department of Transportation, Federal Aviation Administration, Tech. Rep., 3 2001.
[33] T. L. Seamster, R. E. Redding, J. R. Cannon, J. M. Ryder, and J. A. Purcell, “Cognitive task analysis of expertise in air traffic control,” *The International Journal of Avition Psychology*, vol. 3, no. 4, pp. 257–283, 1993.
[34] D. D. David Cornish, *The Essential 20: Twenty Components of an Excellent Health Care Team*. RoseDog Books, 2009, pp 72-73.
[35] P. Goillau, C. Kelly, M. Boarrdman, and E. Jeannot, “Guidelines for trust in future atm systems: Measures,” Eurocontrol, Tech. Rep., 5 2003.
[36] E. Jeannot, C. Kelly, and D. Thompson, “The development of situation awareness measures in atm systems,” Eurocontrol, Tech. Rep., 6 2003.

II

Literature Review and Preliminary Research (*Graded under AE4020*)

1

Automation in Air Traffic Control and Implications

1.1. Today's Automation in Air Traffic Control

Gathering raw information such as airspeed & heading of aircraft and integrating them into live aircraft position for a safe ATC operation has become automated such that human ATCOs do not actively have to search for specific raw information and integrate it to higher level of flight information. However, ATC of today still relies heavily on human ATCOs to make decisions and supervise the automated systems. ATCOs primarily inspect a Plan View Display (PVD) to monitor a designated sector with an aid of DSTs. DSTs typically automate process of gathering and integrating information for ATCOs such that decision-making process is assisted by the automation. Several DSTs are already implemented in the current ATC, while the others are being researched and developed. The following subsections introduce modern DSTs with their detailed components.

1.1.1. Conflict Resolution Assistant

The Conflict Resolution Assistant (CORA) is a tool which provides a assistance to en-route ATCOs in the conflicted airspace and increase the safe throughput of the ATC system [4]. The CORA tools consist of the following:

- Trajectory Predictor (TP): the TP determines the potential conflict areas with corresponding aircraft by predicting routes of aircraft.

- Monitoring Aid (MONA): the MONA checks if aircraft with received instructions such as a heading or altitude change shows non-conformance or deviations from the intended routes.

- Medium Term Conflict Detection (MTCD): the MTCD detects potential conflicts within a medium time window up to 20 minutes ahead. It detects all possible conflicts between every aircraft pair examined. It is important to note that the MTCD does not provide advisory to ATCOs.

1.1.2. Highly Interactive Problem Solver

Highly Interactive Problem Solver (HIPS) developed by Eurocontrol is a tool that provides assistance to the ATCOs to enable timely and efficient resolution of problems [5]. On top of a base radar display, it is augmented with Conflict and Risk Display (CARD), which is shown in Figure 1.1; a conflict is displayed as a set of triangles, which is placed on a 2D axis system of x-axis being Time to Closest Point of Approach (TCPA) and y-axis being Distance Between Aircraft at CPA (DCPA). The conflict probe detects conflicts to be resolved and displays them on the CARD.

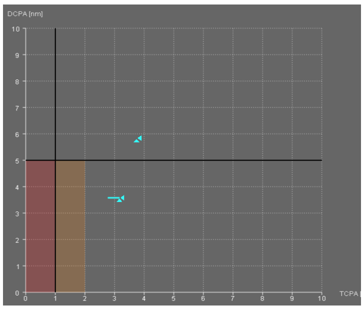


Figure 1.1: HIPS CARD

HIPS also utilizes TP, as a part of the navigator subsystem. HIPS actively attempts to ensure that safe trajectories are calculated with minimal assumptions with regards to other trajectories [5]. By combining the radar display, CARD and TP, the problem solver display was built, which is shown in Figure 1.2.

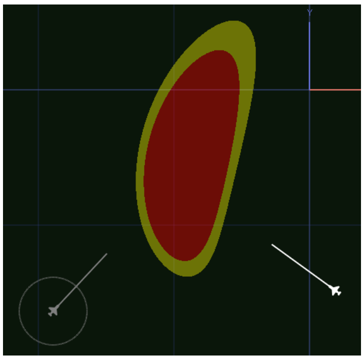


Figure 1.2: Problem solver display of HIPS

1.2. Future Air Traffic Control Trend

The visions of leading ATM/ATC researches were investigated in order to understand the future ATC trend. SESAR master plan describes the vision which aims to achieve 'high-performing aviation for Europe' by 2035, which is based on trajectory-based operations and a progressive increase of the level of automation support; the increase in automation support facilitates tactical coordination, which increases productivity of ATCOs and would allow for increased en-route capacity [6]. The overview can be seen in Figure 1.3 and the relevant elements for this thesis are marked with magenta boxes.

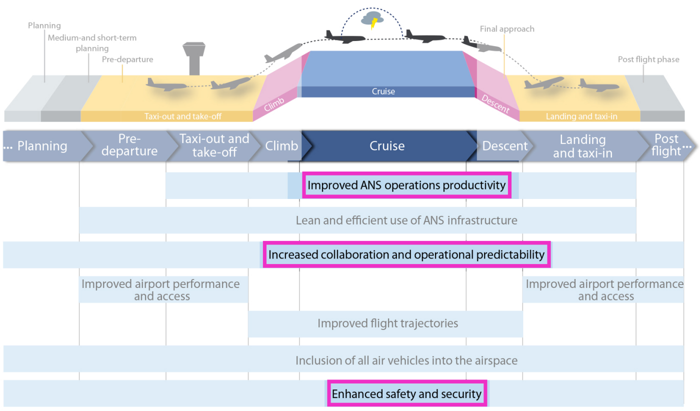


Figure 1.3: The Vision of the Latest SESAR Master Plan (2015)

Federal Aviation Administration's (FAA) NextGen has a similar vision. Improvements to automated controller tools will allow increased access to planning information and assist controllers with better aircraft sequenc-

ing; the automated tools aim to reduce traffic complexity as well as cognitive workload, and provide opportunities for en-route ATCOs to manage more aircraft simultaneously and accommodate optimized tactical ATM provision with less required intervention for spacing and sequencing [7].
The following subsections introduce state-of-the-art researches that provided inspirations for the concept to be proposed for this thesis research.

1.2.1. STRESS

The project STRESS by Eurocontrol has acknowledged the role shift of ATCOs from active controllers to monitoring supervisors. The research is focused on the controller stress and workload in normal conditions and automation disruptions when human controllers must react swiftly under high stress. STRESS proposed a multidisciplinary approach intended to apply the high time resolution neurophysiological measurement of ATCOs' mental status to the extension of operational tasks, within a simulated ATC environment reproducing complexity of future airspace scenarios and associated supporting technologies [8]. The human factors that were mainly focused in the research are stress, attention, mental workload and cognitive control behavior, which splits to skill-based, rule-based and knowledge-based. They were assessed using neurophysiological measurement tools such as electroencephalography, eye-tracker, electrocardiography, electrooculography and galvanic skin response tools. The research gap for STRESS was customizing the neurophysiological indicators to future ATC tasks, as it is believed that stress in ATC corresponds to different patterns of neural activity compared to everyday stress [8]. As a result, a customized neurophysiological measurement toolbox was constructed such that human performance can be measured for the future ATM.

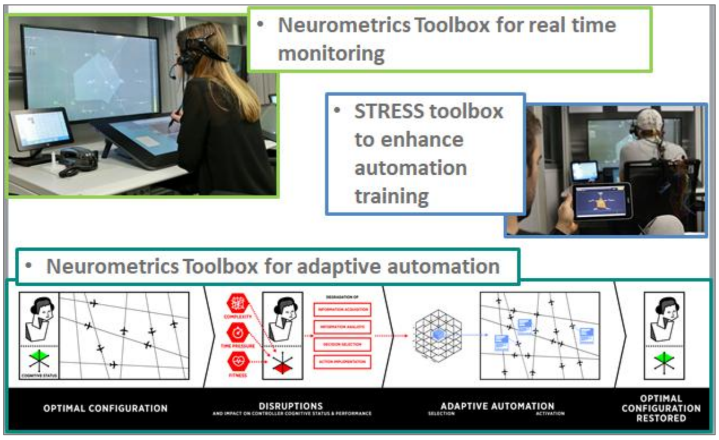


Figure 1.4: The STRESS toolbox as an enabler for future automation solutions

It is stated that the capability and reliability of selected neurophysiological indicators were proven to provide additional and useful information for assessing an operator, however, types of sensors and neurophysiological signals to consider must be carefully selected depending on the task and environment [8]. Also, it is noted that the knowledge and use of neurometrics in ATC environments is not common yet, and more efforts are required in order to implement the use of the neurometrics to support ATCOs.
An interesting and relevant recommendation from the STRESS project is that the validation of the solutions to mitigate ATCO performance drawbacks at highly automated environments should also cover the rather extreme cases of automation failure where a controller has to step in to resume control.

1.2.2. MINIMA

The project MItigating Negative Impacts of Monitoring high levels of Automation (MINIMA) developed a Vigilance and Attention Controller (VAC) to mitigate OOTLUF phenomenon as the global air traffic grows and implementing higher levels of automation has become prominent. The aim of MINIMA was to develop tools capable of detecting the negative of OOTLUF and compensating for them through smart human-automation task distributions [9]. Utilizing neurophysiological methods, there are three key results to be reviewed. First, the MINIMA project concluded that one of the main factors causing OOTLUF phenomenon is the vigilance and attention decrements issue. Thus, they decided to use the current vigilance and attention levels of ATCOs as metrics to identify OOTLUF phenomenon, in combination with neurophysiological methods and implementation of adaptive automation. Second, the VAC was developed, which integrates a neurophysiological

based tool to assess real-time vigilance data and activates a trigger for adaptive automation solutions to maintain the controller vigilance and attention at a desired level.
The conclusion from the project points out that lack of task engagement in highly automated systems results in decreased vigilance, which leads to reduced safety. It can be noted that this thesis research will prioritize increasing task engagement of controllers. Also, the MINIMA project concluded that neurophysiological metrics turned out to be intrusive and complex to be suitable for operational use.

1.2.3. Solution Space Diagram

SSD shows locomotion constraints for the selected aircraft by displaying triangular velocity zones in the airspeed envelope, which is illustrated in Figure 1.5.

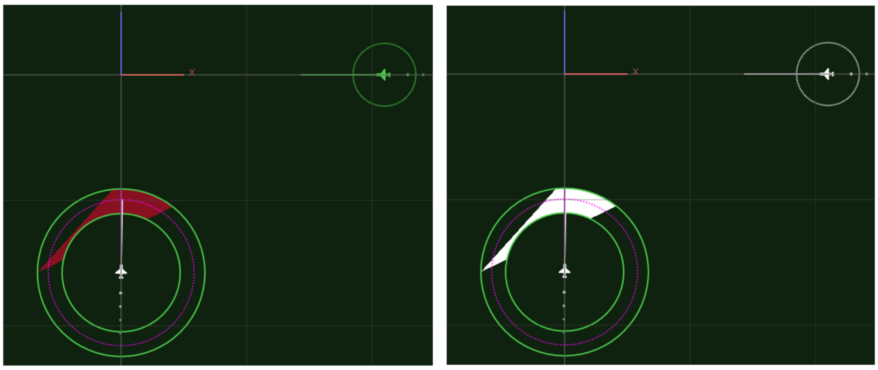


Figure 1.5: SSD demonstration

The SSD has been utilized to help controllers with diagnosing faulty automation performances for conflict detection and resolution [10]. The research utilized Ecological Interface Design (EID) in combination with the AH analysis for the fault diagnosis. It is important to point out that the interfaces should show all the relevant information to the operator, which is important for unexpected events such as sensor failures [11]. The SSD is particularly interesting for this thesis research; as the aim of the experiment to be conducted is to determine human operators' interaction with a fully automated ATC system (supervisory) and the same ATC system with a minimal implementation of automation (manual), it is important to select a platform which has capabilities to automate the previously mentioned stages of automation described in section 1.4. The SSD is capable of simulating a fully-automated ATC scene as well as limiting desired automation capabilities (e.g. turning off an automated conflict resolution function).

1.3. Ironies of Automation

With reasonable confidence, the ironies of automation will not be resolved [12]. Automated ATC introduces ironies of automation to ATCOs; two main ironies of automation is that designer errors can be a major source of operating problems, and the designer who tries to eliminate the operator skill leaves the operator to do the tasks which the designer cannot think how to automate [13]. An operator may be expected to monitor the validity of the automated system, or take over the control otherwise. There are several implications of having human operators as supervisors, which are elaborated in the following categories.

1.3.1. Manual Control Skills

According to several studies showing differences between inexperienced and experienced operators making a step change, the experienced operator tends to perform the step change with the minimum number of actions, and the process output transfers to the new level swiftly and smoothly. Compared to the experienced operator, the inexperienced operator has a process output oscillating around the target value. However, physical skills including fine gain and timing often deteriorate when they are not frequently used. This means that a formerly experienced operator who has been monitoring an automated process may now be an inexperienced one [13]. This means that experienced ATCOs may have difficulties with maintaining their high manual control skills as automation takes over majority of their work, and the supervisory control aspect increases. A possible solution to overcome this irony is to include ATCOs within a loop of automation routines, such that they get opportunities to practice manual control.

1.3.2. Cognitive Skills

An operator will only be successful in coming up with effective strategies for unusual situations if the operator has sufficient knowledge of the process. The retrieval of the knowledge depends on frequency of the usage and availability of feedback about its effectiveness. Also, it is important to note that online decision making process is done within context of the operator's knowledge on the current state of the system. For example, an operator at an automated plant may come into the control room half an hour before the operator is due to take over control in order to grasp a feeling for what the automated process is doing. It is suggested that a possible solution is to allow the operator to perform hands-on control for a short period during each shift. For the case of using TP and MTCD, the tools do not actively include ATCOs in terms of allowing them to develop a mental picture of the traffic situation in a sector, as integrated results are merely given to ATCOs. When the automated CD&R fails, ATCOs will have difficulties understanding the traffic situation as a proper mental picture would not be built by the time they have to take over and perform manual CD&R.

1.3.3. Monitoring

It is not possible for even a motivated and skilled operator to maintain effective visual attention more than half an hour when things happen rarely on an active source of information. Unlikely abnormalities cannot be monitored without an automatic alarm system usually connected to sound signals. Also, the automated control system has been introduced because it can perform tasks better than the human operator, however, the operator still has to monitor the automated system to see if it is doing the right thing effectively. A serious problem with the aspect is that a computer can make specified decisions much faster while considering more dimensions and using more accurately specified criteria than a human operator can; this means that a human operator has no way to check in real-time that the computer is following set rules correctly and performing tasks accordingly. A possible way to overcome the aforementioned vigilance issues is to increase the signal rate artificially. Over time, the irony of monitoring has evolved into the following: even minor anomalies in complex sociotechnical systems can escalate the severity of potential consequences through human operator interaction with automation [12]. For ATCOs, minor errors such as misreading information of flight labels and missing the active status of aircraft in a responsible sector can result in greater accidents in the event that ATCOs are not actually aware of what caused the automation-related anomalies.

1.4. Stages of Automation

Stages of automation framework suggested by Parasuraman, Sheridan and Wickens is chosen for this research, as it allows specific automation failures to be designed and implemented for the research. An example of the framework can be seen in Figure 1.6, which shows two systems with different levels of automation.

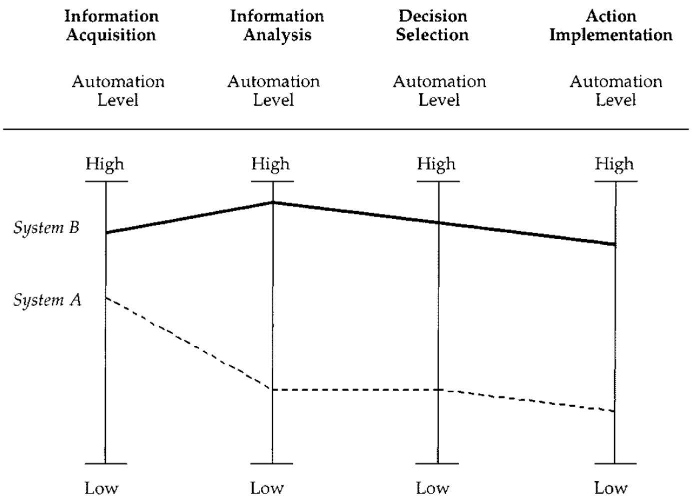


Figure 1.6: Levels of automation for different stages of automation [1]

1.4.1. Acquisition Automation

The first stage of automation corresponds the automation of detecting and registering available input data. A typical example can be radars used in ATC. Currently moderate levels of automation for the acquisition automation stage exist in the form of electronic flight strips, as ATCOs can sort aircraft by handling priority.

1.4.2. Analysis Automation

Then analysis automation utilizes inferential processes and working memory. For the lower level of this automation, incoming data can be processed using prediction algorithms. For example, cockpit predictor displays show the future trajectory of aircraft in its proximity [14]. For the ATC application, the converging runway display remove the necessity for ATCOs to predict and project the approach paths of landing aircraft onto others on a converging runway [15]. It can be seen that human operator perception and cognition are augmented with the analysis automation.

1.4.3. Decision Automation

This stage of automation involves selection of a decision among available decision alternatives by varying human decision options with automated decision-making [1]. Figure 1.7 shows the 10 levels for the levels of automation for the decision automation stage. Typically the level 5 is referred as "Management by Consent" (MbC) and the level 6 can be regarded as "Management by Exception" (MbE). For MbC, the human operator concurs a decision that the automation suggests, then the automation implements the suggested solution. The automation essentially does the same for MbE, but the human operator can only veto the suggested solution. Both levels are being widely used and implemented in the aviation field, however, varying levels of automation at this stage will not be considered for this research for the following reasons. First of all, using the varying levels of automation for a HITL experiment will cause branches of outcomes to sprout; some human operators may trust automation and decide to accept given automation advisories, while the others may be inclined to veto the automation advisories due to distrust in automation. As a result, outcomes of each participant will allow not allow meaningful and valid comparison to be performed. In addition to the issue with the HITL experiment design, there are several researches on why MbC and MbE will not work as intended. A study which investigated the impacts of conflict type, time, and display design on controllers' capability to come up with decisions about suggested automation goals and actions in a MbC context found that MbC does not necessarily ensure effective control, as the complexity and low observability of automated systems can make it nearly impossible for controllers to provide informed consent to the automation goals and actions [16]. MbE also has issues that it makes human operators get trapped into a dilemma that can't be resolved by training and experience. Early intervention leaves barely any room for justification, while late intervention yields no time to resolve the problem, even though the problem may have evolved into a greater one [17]. Thus, only the highest (level 10, full automation) and the lowest (level 1, manual) levels of automation will be considered for this research and HITL experiment to be conducted.

HIGH	10. The computer decides everything, acts autonomously, ignoring the human.
	9. informs the human only if it, the computer, decides to
	8. informs the human only if asked, or
	7. executes automatically, then necessarily informs the human, and
	6. allows the human a restricted time to veto before automatic execution, or
	5. executes that suggestion if the human approves, or
	4. suggests one alternative
	3. narrows the selection down to a few, or
	2. The computer offers a complete set of decision/action alternatives, or
LOW	1. The computer offers no assistance: human must take all decisions and actions.

Figure 1.7: Levels of Automation of Decision and Action Selection [1]

1.4.4. Action Automation

The last stage is the action automation stage, which refers to the execution of the chosen decision & action and generally eliminates human operators action [1]. An example for ATC is the automated handoff, as a single key press allows ATCOs to automatically perform handover into the adjacent airspace.

1.5. Situation Awareness

To ensure minimum separation and safe & efficient operations, sorting and projecting paths of aircraft rely on SA of ATCOs who must maintain live assessments of aircraft locations and their future locations relative to each other, along with other important aircraft parameters such as destination and airspeed [18]. A model for SA was constructed based on dynamic goal selection, attention to appropriate critical cues, expediencies regarding future states, and the relationship between SA and typical actions.

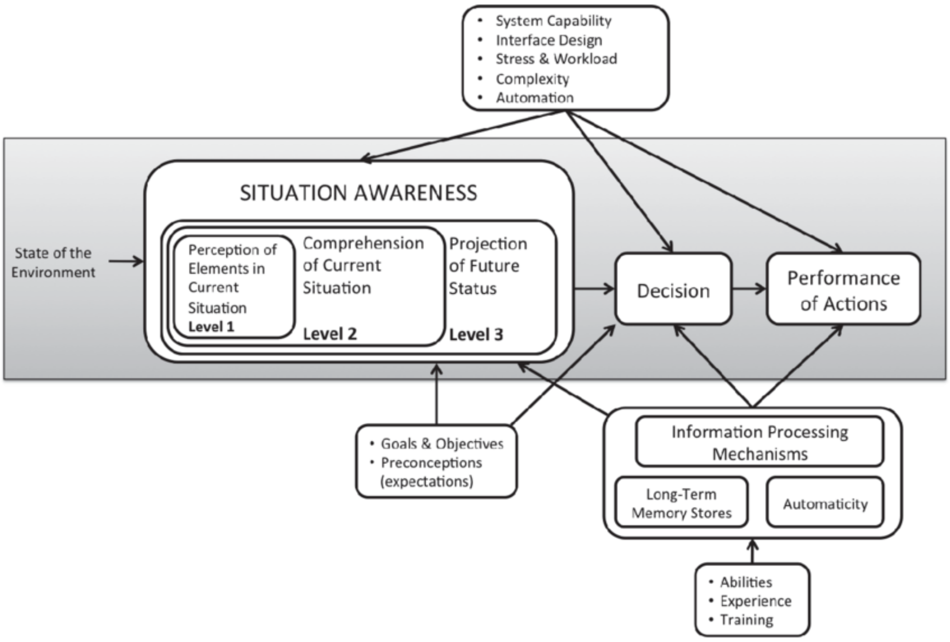


Figure 1.8: Model for SA in dynamic decision-making

The SA model can be seen in Figure 1.8. The main focus of this research on this model is the three levels of SA.

The first level is perception of elements in the environment. SA can be achieved by perceiving the status, attributes, and dynamics in the environment. An ATCO would perceive elements such as aircraft, heading and speed of each aircraft, or alarms on a radar screen.

The second level is comprehension of the current situation. It is based on a synthesis of first level elements. Based on a combination of the first level elements, a decision maker can form a comprehensive picture of the environment, which allows the decision maker to recognize and understand the significance of objects and events. An example can be that an ATCO assesses altitude, heading and airspeed of two aircraft and realizes that they are currently in conflict.

The third and highest level is projection of future status. It is obtained by acquiring knowledge of the attributes and dynamics of the components along with understanding of the current state, which is gained by the synthesis the first and second level of SA. For instance, an ATCO may acquire this level of SA upon inspecting a pair of aircraft in conflict and figuring out that giving one aircraft a certain heading change can resolve the conflict without putting both aircraft in a new conflict with other aircraft in a sector.

Endsley categorized the following SA measurement techniques that can be used for empirical measurement of situation awareness [19].

1.5.1. Measurement Techniques

Physiological Techniques

Physiological techniques can be utilized to check whether environment elements are perceived and processed by subjects, however, the techniques are not capable of determining registration of information and the degree of operators' comprehension on the elements. For example, eye-tracking methods can only provide researchers necessary measures to understand the processes subjects use for achieving SA. Even though physiological techniques can provide useful data, they are not considered as a suitable method to measure SA for this research, since high and rapid eye and cursor movements do not translate to maintaining high levels of SA. There are several state of the art researches carried out by EU and Eurocontrol, which will be elaborated in section 1.2.

Performance Measures

In general, performance measures usually excel in being objective and non-intrusive. Conducting system simulations on computers allows performance data collection extremely easy as the performance data can be automatically collected. However, there are several limitations in using performance measures to derive SA. Global measures of performance only give the end result of long and complex cognitive processes, which leads to lack of understanding on the cause of poor performance in a given scenario, as poor performance could occur from sources that are not related to SA.
A different type of performance measures is to artificially change certain information or remove parts of information, then measure the time required for the operator to cope with the event [20]. A major flaw with this measure is that results regarding SA may be highly misleading, as it is assumed that an operator will act in an expected manner even though operators tend to use alternative schemes to compensate for unexpected circumstances. Also it can directly affect attention and SA itself as the tasks to be performed can be changed.
Another type of performance measures is called imbedded task measures, which examine the operator performance on primary or specific sub-tasks. It can mean than the aforementioned global measures because detailed performance measures can provide more accurate and relevant SA inferences. However, improved SA on certain elements may lead to a decrease in SA on other elements.

Subjective Techniques

Self-rating is a simple method to ask operators to rate their SA experienced subjectively. It can be done either during or after an experiment. The downside with collecting SA ratings during an experiment is that operators' own SA estimation can be limited and inaccurate as they only have perceptions of that moment. It means that they may know when they have no idea about what is happening, but will not be able to tell if their knowledge for given tasks is incomplete or inaccurate. On the other hand, asking operators to evaluate SA in a post-experiment session may allow the ratings to be contaminated by the experiment outcome.
Observer-rating is another type of subjective rating which requires observers to score the subject's SA. Although an experienced observer may have a better overview of the current state of a simulation, the observer would have limited understanding of the operator's perception on the simulation.

Questionnaires

Details about subject's SA can be gathered by using questionnaires. It can provide a fair assessment of the subject's SA. There are three types of questionnaires: post-test, on-line, freeze technique.
Post-test questionnaires can be given to subjects after each simulation run. It provides subjects sufficient time to answer a detailed and extended list of SA questions during the simulation run. However, this method can reliably measure the SA only at the end of the simulation run, as people are not great at remembering details about past mental events [18].
On-line questionnaires can overcome the previously mentioned limitation by asking subjects to assess their SA during simulation runs. However, this method also comes with several drawbacks. First of all, additional workload will be given to the subject as on-line questions would act as an ongoing secondary task, which can affect primary task performance. Also, the questions could reroute the subject's attention to the required information on displays, which can alter the subject's true SA. Another drawback is that assessing time to answer the provided questions as a SA indicator is faulty, since each subject can use different time management strategies to manage the primary tasks while answering the questions. The third questionnaire method is freeze technique. This method can be used to overcome the limitation of on-line questionnaires by freezing the simulation at randomly chosen times and asking subjects to answer questionnaires regarding their perceptions. The simulation pauses and displays become blank, which allows subjects to answer the given questions fast. As a result, SA data can be immediately gathered, which minimizes the problem of on-line questionnaires.

2

Cognitive Work Analysis: Air Traffic Control Operations at Area Control Center

CWA was carried out in order to determine the scope of the research and define relationship of ecological and cognitive elements. The outcomes of the CWA are used to construct a HITL experiment.

2.1. Determination of System Boundaries

As a general ATC is a comprehensive term for different areas and zones, the system boundaries must be defined. As the aim of this research is to increase task engagement during a highly excited supervisory control phase, it is crucial to select an appropriate environment. Thus, upper control area (UTA) is chosen as there are relatively less activities in UTA compared to other airspace areas, and it could be the easiest to automate in real life, if full automation is to be implemented. The chosen airspace highlighted in magenta can be seen in Figure 2.1.

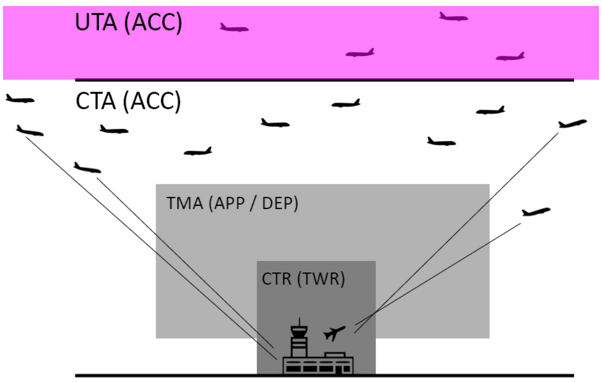


Figure 2.1: Selected airspace

2.2. Work Domain Analysis

As the system boundary was set, work domain analysis was carried out to determine the system to be investigated and its purpose. A structured and functional map of the workspace was constructed based on Rasmussen's Abstraction Hierarchy (AH) [21]. A generic form of AH is given in Figure 2.2. The functional purpose defines the designated system outputs. Next, the abstract function consists of fundamental physics which governs the selected work domain. Then, constraints of system procedures and information are determined within the generalized function. The physical function specifies intricate relationships of processes linked to interactions of different components. At the lowest level, the physical form consists of specific characteristics of objects in the system such as states, shapes and locations.

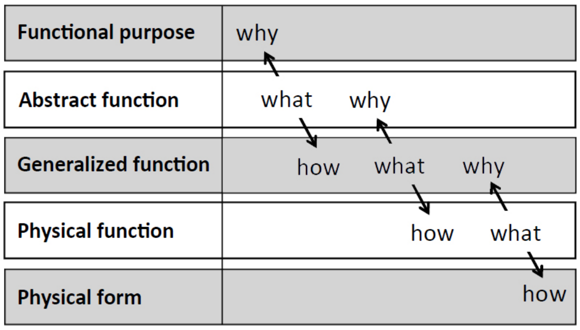


Figure 2.2: AH with means-ends relationships between levels of abstraction

En-route ATCOs at ACC primarily oversees aircraft in each designated sector to detect any potential conflicts and resolve them. Their functional purpose can be described as safety, efficiency and productivity. ATCOs ensure safety by providing separation between aircraft, and the absolute and relative locomotion is maintained by changing heading and/or airspeed. The productivity can be regarded as the level of task engagement to their tasks and the corresponding quality of their work. Efficiency can be defined as how efficient ATCOs can resolve a potential conflict using given resources such as available airspace, heading and airspace.

2.3. Control Task Analysis: En-route Controller Task Analysis

To decide which of ATCO tasks will be considered for this research, a literature survey on en-route controller tasks was performed. Eurocontrol has conducted several researches on modernizing and integrating human aspects into the future ATM system development. Task analyses for ATCOs were performed by several research groups, which can be seen in Table 2.1.

Table 2.1: ATCO task analysis

A= Redding & Seamster (1994)
B= EATCHIP (1996)
C= Cox (1994)
D= Glaser & Dahl (1995). Airport/3
E= NLR (1996). RHEA

	Core Tasks Listed by Various Task Analyses	A	B	C	D	E
1	**Maintain SA**	x	x		x	x
	Build up MP of traffic situation		x			x
	Perform routine sector 'maintenance' task			x		
2	**Develop and receive sector control plan**	x		x	x	x
	Perform actions before aircraft (a/c) arrives in sector			x		
	Handle and process flight plan information		x			x
	Manage air traffic within area of responsibility		x			
3	**Make decision for control actions**	x	x			
4	**Solve a/c conflicts**	x				
	Provide separation		x	x		
	Conduct recognition and resolution of conflicts in en-route airspace			x		
5	**Provide tactical ATM**		x			x
	Route a/c through sector airspace/manage overflight/re-route a/c	x		x	x	x
	Conduct a/c movements			x	x	
	Initiate/point out/transfer control	x		x	x	
	Receive pointout/accept a/c	x		x	x	x
	Receive handoff/carry out handover from previous controller	x		x	x	
	Initiate handoff/carry out handover to next controller	x		x	x	
	Manage arrivals	x		x		
	Manage departures	x		x		
	Issue advisory/provide pilots and colleagues with all relevant info	x	x	x	x	
	Issue safety alert	x				
	Provide assistance in abnormal situations		x			
	Manage airborne emergency			x		
	Ensure correct co-ordination		x		x	
	Conduct Radiotelephony (R/T) communication		x	x	x	
	Manage pilot-initiated communication			x		
	Perform actions after a/c has left sector			x		
	Conduct pre-shift briefing			x		
	Handle, manage Flight Progress Strips (FPSs)	x		x		
	Check technical equipment at working position		x			x
6	**Complementary tasks**		x		x	
	Train		x		x	
	Update working knowledge		x			
	Supervise control room/team		x		x	
	Co-ordinate with customers/users		x		x	
	Manage sector/position resources				x	
	Assess situational conditions				x	x

From Integrated Task and Job Analysis of En-route ATCOs [22], five task groups are recognized as following:

1. Maintain SA: ATCOs must have 'the picture' to maintain SA; a mental traffic picture must be established, followed by continuous projection into future and checking of anticipation with the actual traffic status.
2. Develop and receive sector control plan: integration of flight progress information and traffic forecast into a future traffic situation plan needs to be performed to ensure a safe and efficient traffic flow.
3. Make a decision for control actions: developing and revising the sector control plan is closely related with active decision-making of control actions.
4. Solve aircraft conflicts: provision of separation and resolution of aircraft conflicts are included in this task group.
5. Provide tactical ATM: accepting an aircraft and performing handover to adjacent sectors are the main tasks in this task group.

2.4. Controller Best Practices

Controller best practices are investigated to understand typical strategies used by ATCOs such that relevant task engagement tool questions and sub-tasks on SA can be constructed, which would ultimately assist ATCOs with making 'good' decisions and performing safe & efficient CD&R.
Best practices for conflict detection revolve around identifying a pair of aircraft in conflict and determining the best moment to act on the conflicted aircraft pairs [23]. A previous research on using adaptive automation based on ATCO's decision-making identified several key strategies. Table 2.2 shows the controller best practices for CD&R.

Table 2.2: Strategy analysis: controller best practices

#	Conflict detection
1	Keep look-ahead time of 5-10 minutes [24]
2	Be more conservative in classifying conflicts in high workload conditions [22]
3	Wait and see before taking action in low workload conditions [25]
4	Act immediately after detecting a conflict in high workload conditions [25], [22]
	Conflict resolution
5	Solutions from a mental conflict resolution library, built from training and experience [22]
6	Use conservative & safe solutions and reduce efficiency criterion in high workload conditions [26], [25], [22]
7	Use standard and routine solutions in high workload conditions [22]
8	Use the first solution that comes to mind in high workload conditions [25]
9	Select resolution that requires the least monitoring [27]
10	Select resolution that requires the least coordination [24]
11	Minimize the number of aircraft to move [24]
12	Minimize additional track miles flown [24]
13	Give initial change early on and fine-tune later [24]
14	Penalize aircraft with additional requests [24]
15	Turn slower aircraft behind for crossing conflicts [24]
16	Give faster aircraft a direct-to command in front of slower aircraft for same track conflicts [24]
17	First solve conflict pairwise and later check consequences on other traffic [24]

2.5. Social Organization

To shape an experiment for this research, tasks were split and assigned to automation and human controllers. Based on the task and strategy analysis, the following social organization between a human operator and ATC automation. When the full automation is at work (supervisory control), a human operator is tasked with maintaining SA, developing and receiving sector control plan, and supervising the automation's work for detecting and resolving conflicts. The automation is responsible for the following tasks:

1. Information acquisition and integration: acquire flight information and project integrated information on a PVD

2. Conflict detection: check converging aircraft pairs, crossing flights, heading and airspeed

3. Conflict resolution: change heading and/or airspeed, revise and organize traffic patterns and routes

4. Tactical ATM provision: perform handover

When the fully automated system fails, the automation is only left with a single task of only acquiring flight information and showing it on a PVD. Then a human operator has to take over the majority of the automated tasks as following:

1. Maintain SA

2. Conflict detection: check converging aircraft pairs, crossing flights, heading and airspeed

3. Conflict resolution: change heading and/or airspeed, revise and organize traffic patterns and routes

4. Tactical ATM provision: perform handover

2.6. Worker Competencies Analysis

CWA on terminal radar approach control (TRACON) ATCOs was carried out and worker competencies analysis was tabulated into a skill-, rule-, knowledge-based (SRK) inventory for rerouting control task [3], which can be seen in Table 2.3. The given SRK inventory was chosen as analyzing the SRK behaviors of en-route ATCOs for their conflict resolution & tactical ATM provision tasks are comparable to the rerouting task of TRACON ATCOs. Cells highlighted in blue correspond to SRK elements which are used in both supervisory and manual control, while cells highlighted in orange are only triggered when the automated system fails and the human operator has to perform tasks manually. Cells in white are not to be considered for this research, as they describe elements related to communications between ATCOs and pilots.

Table 2.3: SRK inventory for rerouting task in TRACON simulator domain [3]

Information Processing Step	Resultant Knowledge State	Skill-Based Behavior	Rule-Based Behavior	Knowledge-Based Behavior
1. Scan for aircraft presence in area of responsibility	2. Whether multiple aircraft are within area of responsibility	Monitoring of time-based spatial representation of aircraft in area of responsibility	Perceive explicit indication multiple aircraft are currently within area of responsibility	Reason, based on proposed flight plans, that multiple aircraft may be present in area of responsibility within similar time frames
3. Determine future flight vector for each aircraft	4. Whether multiple aircraft within area of responsibility have intersecting flight paths	Perceive headings of related aircraft as convergent, divergent, or parallel	Use heuristics to determine whether flight paths are intersecting	Reason, based on geospatial knowledge of to/from points for each flight, that aircraft are on convergent or divergent paths
5. Predict future, time-based location states for aircraft on convergent paths	6. Whether converging aircraft will arrive at point of convergence within a similar time frame	Perceive time-to-collision of each aircraft with the convergence point, based on spatial representations of heading and speed	Use heuristics to estimate whether aircraft will arrive at convergence point within a similar time frame	Calculate, using airspeed, heading, and location of each aircraft, the time at which each aircraft will arrive at the convergence point
7. Determine the criticality of a pending convergence	8. Whether future distances between converging aircraft will constitute a ``loss of separation' event	Perceive whether the zones of safe travel surrounding each aircraft will overlap at or near their closest point	Use heuristics to determine proximity as being greater or less than the minimum required envelope of separation	Calculate distance between each aircraft at their closest future states and compare with the minimum value of separation required for safe travel
9. Choose to modify aircraft flight paths to address future problem	10. Which aircraft flight paths must be modified to eliminate potential `loss of separation' event	Directly perceive that one or more aircraft must be redirected	Apply doctrine: e.g. if loss of separation will occur, MUST reroute one or more aircraft	Reason from knowledge of proposed flight paths, current locations, and expected future behavior that aircraft must be rerouted
11. Select specific strategy for accomplishing rerouting of aircraft	12. Desired aircraft rerouting strategy	Respond automatically to perception of loss of separation by directly manipulating a representation of aircraft flight paths	Classify loss of separation within a set of generalized scenarios and select appropriate stereotypical control rule	Develop new, optimized flight paths based on weighted criteria including urgency, flight priority, passenger convenience, efficiency, etc
13. Convey flight modifications to aircraft for execution	14. Aircraft's awareness of new flight paths	Direct, simultaneous interaction with communication equipment through control interface through input of rerouting information	Apply stereotypical control rules to select method/sequencing for conveying proposed flight path	Reason using knowledge of aircraft systems, priorities, urgency, etc., the best means and order for contacting each aircraft to convey proposed flight paths

3

Concept Proposal: Task Engagement Tool

The core idea of the task engagement tool is to reroute the operator's attention to potentially neglected or overlooked elements of the ATC tasks, by assessing the SA level via on-line questionnaires and simple tasks to perform. The concept of the task engagement tool is shown in Figure 3.1. On the left side, the PVD with a SSD augmentation can be seen. On the right side, the task engagement is present; for instance, a level 3 question, 'will a direct-to command to waypoint WENS for PA5424 result in a conflict?', is shown to a human operator at a given moment. The relevant flight label and waypoint is each highlighted and the operator can provide an answer by clicking either yes or no button. If the provided answer is correct, then the small rectangular indicator above will give a green light, otherwise a red light. Each question will have an expiration period, such that an operator has a time limit to assess the PVD and provide an answer to a given SA question. It is important to note that the SA questions have "intelligence", meaning that they will not be randomly given to operators; as a scripted automation will be used for the experiment, it is possible to carefully issue a particular SA question at a specific moment and scenario.

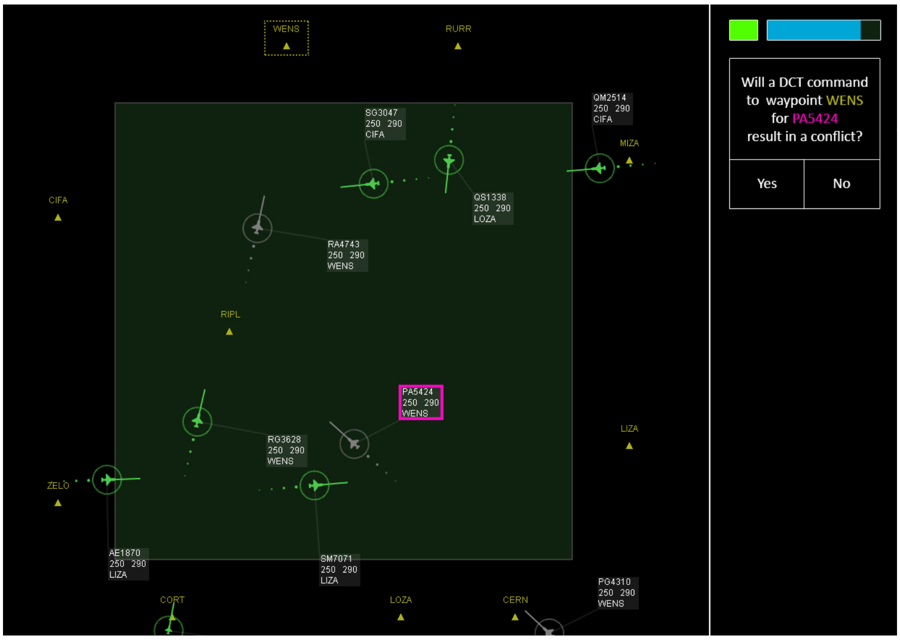


Figure 3.1: Task engagement tool concept with a base SSD platform

Situation awareness information requirements of the En Route Air Traffic Control Specialist (ACTS) were established, which includes perception (level 1), comprehension (level 2), and projection (level 3) of elements from the Endsley's SA model [28]. The detailed en-route ATC SA requirements can be seen in Table 3.1. The 3 levels follow the SA model introduced in section 1.5.

Table 3.1: Selected en-route ATC SA requirements

Level 1

Aircraft

- Aircraft ID, CID, beacon code
- Current route (position, heading, aircraft turn rate, altitude, climb/descent rate, groundspeed)
- Current flight plan (destination, filed plan)
- Aircraft capabilities (cruising speed, max/min speed)
- Aircraft type
- Aircraft status

Level 2

Conformance

- Amount of deviation (altitude, airspeed, route)
- Time until aircraft reaches assigned speed, route/heading

Current Separation

- Amount of separation between aircraft along route
- Deviation between separation and prescribed limits
- Number/timing aircraft on route

Timing

- Projected time in airspace
- Projected time till clear of airspace
- Order/sequencing of aircraft

Deviation

- Deviation aircraft/flight plan
- Projected time till clear of airspace
- Order/sequencing of aircraft

Significance

- Impact of requests/clearances on:
  - Aircraft separation/safety
  - Sector workload

Confidence level/ Accuracy of information

- Aircraft ID, position, altitude, airspeed, heading

Level 3

Projected Aircraft Route (current)

- Position, flight plan, destination, heading, route, altitude, airspeed

Projected Aircraft Route (Potential)

- Projected position x at time t

Projected Separation

- Amount of separation along route
- Deviation between separation and prescribed limits
- Relative projected aircraft routes
- Relative timing along route

Impact of Potential Route Changes

- Type of change required
- Time and distance till turn aircraft
- Amount of turn /new heading change required
- Aircraft ability to make change
- Projected no. of changes necessary
- Cost/benefit of new clearance
- Impact of proposed change on:
  - Aircraft separation
  - Traffic flow
  - Number of potential conflicts
  - Flow requirement
  - Workload required

The task engagement tool will primarily issue task-related questions to participants based on the 3 levels of SA. Also, questions regarding advisories for the highest level of SA are based on controllers' best practices. Figure 3.2 shows examples of utilizing 3 levels of SA and the en-route ATC SA requirements to generate SA questions. It can be seen that level 1 can refer to an aircraft ID, current flight plan (airspeed, heading), and beacon code. For the level 2, elements such as current amount of separation between aircraft and the number of aircraft along route are considered. Finally, the level 3 includes projection of future, which can be projected aircraft route based on position, heading and airspeed, projected separation, and impact of potential route changes.

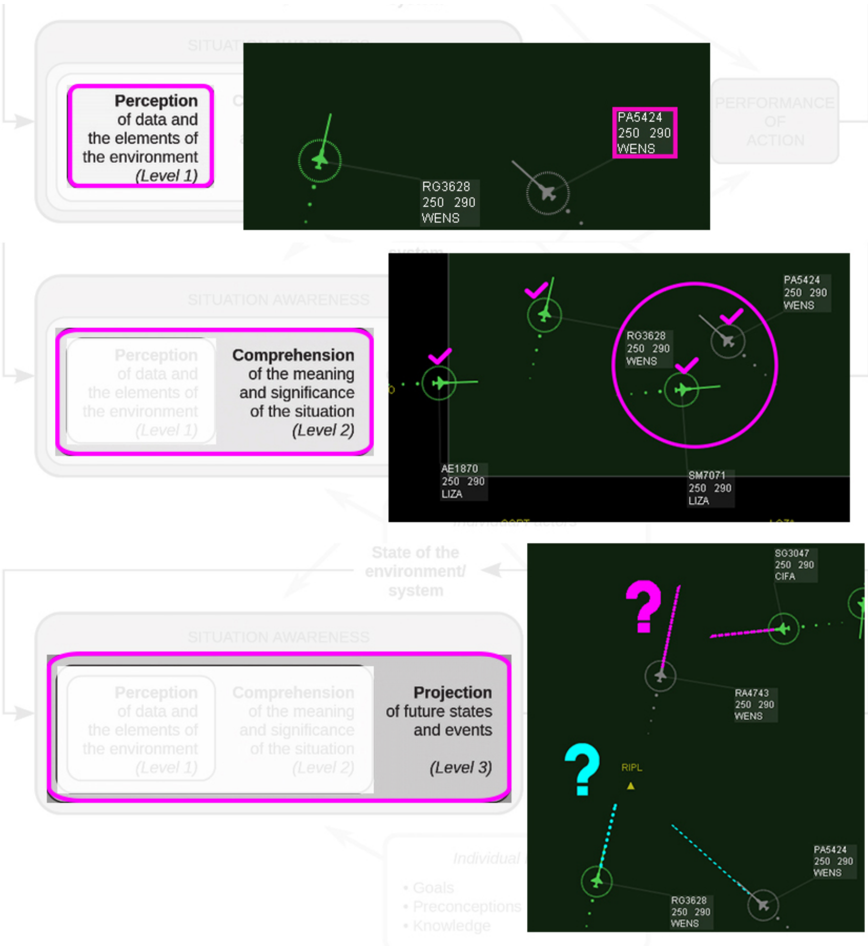


Figure 3.2: Examples of using 3 levels of SA in SSD

Examples of the level 1 SA questions are as following:

- The speed of KLM910 is 220 kts
- The heading of KLM910 is 150
- Current aircraft count is 15
- KLM910 is heading to waypoint CELO

then the level 2 SA questions are shown below.

- 3 flights are exiting via waypoint VOZA, HALO and FELO
- There are 2 flights deviating from direct routes
- Are there any flight that exceeds speed limit towards the exit waypoint HALO?
- Current traffic is concentrated at the center of the sector

Finally, the level 3 SA questions are provided as following:

- Heading change from 150 to 120 is a good solution for KLM910 & LH435 conflict
- Will a direct-to command to exit waypoint CELO for KLM910 result in conflict?
- If KLM910 did not receive an automated resolution, would LH435 be in conflict with KLM910?

It is important to validate the generated questions by checking the en-route ATCO SA requirements, ATCO task analysis, and controllers' best practices. The validation will be carried out along with the implementation of the task engagement tool on the SSD-augmented simulator. Based on the number of correctly answered SA questions, SA of each operator can be determined in a heuristic way.

4

Research Methodologies: Experiment

4.1. Hypotheses

It is hypothesized that compared to the conventional ATC simulation setup with a SSD augmentation, implementing an additional task engagement tool will increase supervisory and manual control performances of controllers at the cost of increased workload, because the task engagement tool shall increase situation awareness of controllers during the supervisory control phase by making human operators get involved in actively updating their SA mental picture. Also, it is hypothesized that human operators that experience an automation failure early in the experiment will exhibit better supervisory control before they experience another automation failure, compared to other human operators who experience an automation failure towards the end of the experiment; it is found that trust in robots/automation was affected more by early failure than later reductions in reliability [29].

4.2. Experiment Setup

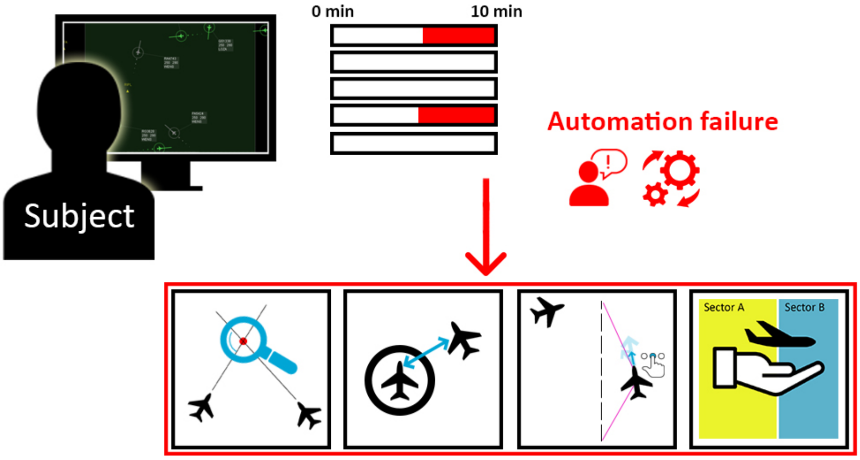


Figure 4.1: Overview of the experiment setup

The overview of the experiment to be conducted can be seen in Figure 4.1. The experiment will be 5 individual runs of 10 minutes each; for each run, participants assume the ATCO position as a supervisor. Their main task is to supervise the scripted 'automation' and take over the control if there is an automation failure. The scripted automation will do CD&R automatically, meaning that it will detect conflicts and give heading and/or speed changes to aircraft pairs in conflict. When the automation fails, the automated CD&R ceases to work, and participants have to perform the tasks manually; the tasks to be performed are conflict detection, provision of separation, conflict resolution, and provision of handovers. They can inspect aircraft pairs with SSD during the supervisory phase. Details on the experiment set-up are given in the following subsections.

4.2.1. Independent Variables

Availability of the task engagement tool

The availability of the task engagement tool during the supervisory control phase will be vary for two groups of participants for in-between participants design manner. Group A will have an access to the task engagement tool during the supervisory phase, while Group B will not have an access to the tool.

Period between automation failures

There will be two different types of experiment runs. Within each group (Group A and Group B), half of each group will experience an automation failure in the first run and be given sufficient time before another automation failure in the last run. For the others, automation failures will only occur in the last two runs. This is to assess whether trust in automation affects manual control performances of participants with the task engagement tool; it is to see if the implemented task engagement tool is effective in increasing manual control performances of participants during their increased distrust in automation.

4.2.2. Control Variables

Availability of Solution Space Diagram

Both groups will have access to SSD during the supervisory control phase. Participants can use SSD to inspect aircraft pairs during the supervisory control phase. However, SSD will be disabled during the manual control phase for both groups.

Task engagement tool questions

Participants of Group A with access to the task engagement tool will receive SA questions during the supervisory control phase such that the distribution of 3 different levels of SA is equal among the participants. The order of receiving the 3 different levels of SA questions will be pseudo-random; participants will not get the highest level of SA questions (SA level 3) as a first question, as the task engagement tool aims to help them build a mental picture by building their SA and cycling it throughout the experiment run. In other words, participants may not get exactly same questions at a given time, but in the end everyone of Group A will have received the same amount of SA questions (i.e. 6 level 1 SA questions, 4 level 2 SA questions, 5 level 3 SA questions). Also, questions will expire after 20 seconds and be counted as incorrect if left unanswered.

Duration of experiment

Considering that uninterrupted attention span of average students is roughly 10 minutes [30], it is important to note that the supervisory control phase has to be sufficiently long until an automation failure occurs. Also, accumulated fatigue over time can adversely impact experiment outcomes, as the aim of sufficiently long duration is to excite the supervisory control phase, not to load participants with high fatigue. Thus, each run will last 10 minutes and all participants will perform 5 individual runs consecutively.

Automation reliability

Automation reliability is chosen to be 60%. According to a research on automation-induced complacency, detection rate of automation failure did not differ for the low-reliability of 57.25% and high-reliability of 87.5% in a constant-reliability condition [31]. The deciding factor for selecting 60% is the excitement of supervisory control characteristics with sufficient amount of manual control actions, as there must be enough time of excited supervisory control phase before automation failure occurs.

Traffic complexity

Traffic complexity is a collective term for many individual factors, which can be seen in Figure 4.2. Note that the presented 12 traffic complexity variables were formulated according to factors relating to weather, traffic, routes, sector and other complexity measures based on the comprehensive list of factors of ATC complexity factors [2]. Also, complexity variables that are highlighted in blue do not have any meaning here, as those were chosen for research of Rahman.

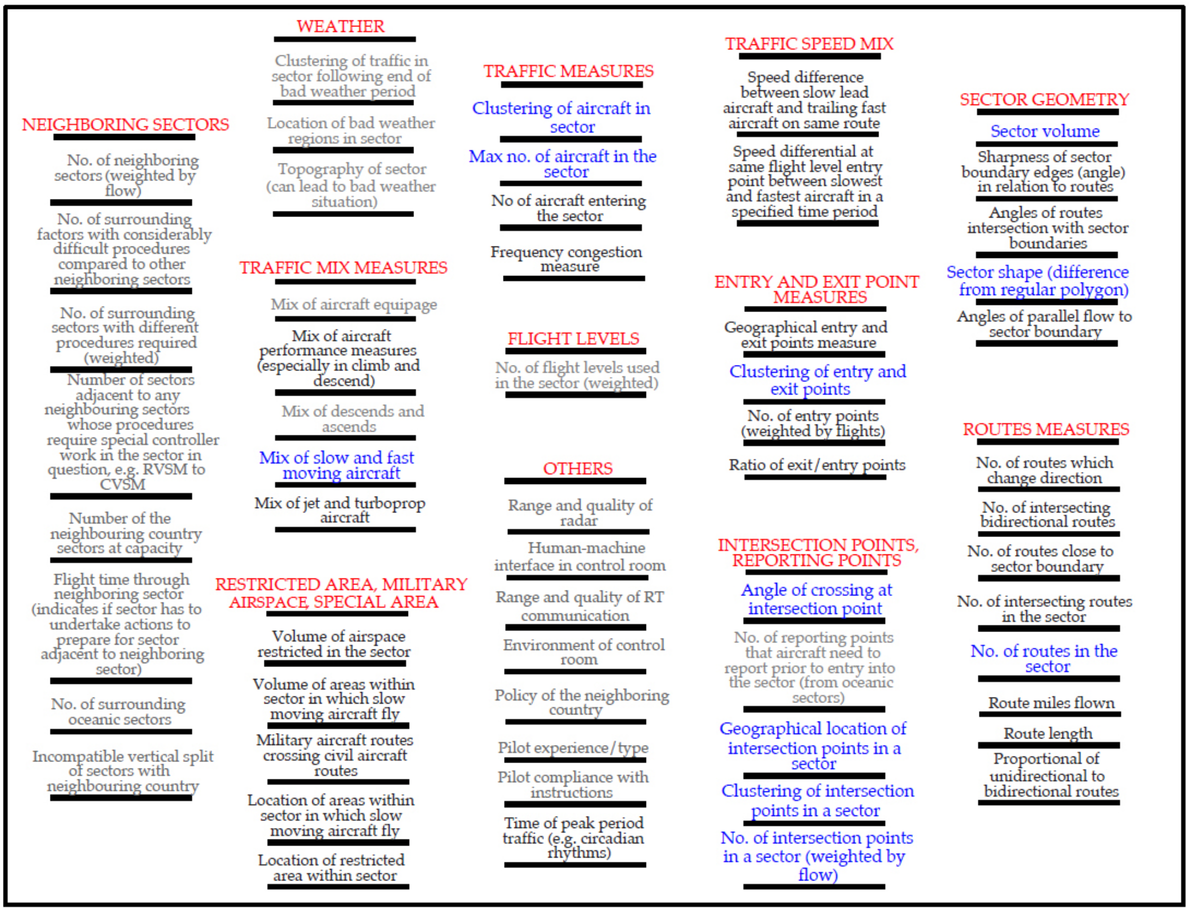


Figure 4.2: Traffic complexity variables [2]

The selected traffic complexity factors for this research can be seen in Figure 4.3.

TRAFFIC MIX MEASURES

- Mix of aircraft performance measures
- Mix of slow- and fast-moving aircraft
- Mix of aircraft types

TRAFFIC MEASURES

- Clustering of aircraft in sector
- Number of aircraft entering the sector

FLIGHT LEVELS

- Single flight level used in the sector

TRAFFIC SPEED MIX

- Speed difference between slow lead aircraft and trailing fast aircraft on same route

ENTRY AND EXIT POINT MEASURES

- Clustering of entry and exit points
- Number of entry points

INTERSECTION POINTS, REPORTING POINTS

- Angle of crossing at intersection point
- Clustering of intersection points in a sector
- Number of intersection points in a sector

SECTOR GEOMETRY

- Sector volume
- Sector shape

ROUTE MEASURES

- Number of routes in the sector
- Number of intersecting routes in the sector

OTHERS

- Range and quality of radar
- Policy of the neighbouring country
- Time of peak period traffic

Figure 4.3: Selected traffic complexity variables [2]

Starting with traffic mix measures, mix of aircraft performance measures mainly indicate different airspeed based on types of aircraft, which is divided into light, medium, and heavy. Slow- and fast-moving aircraft will be present such that takeovers can be simulated, which includes the later mentioned traffic speed mix. Then for traffic measures, there will be no major clustering of aircraft in a particular area of a sector, meaning that aircraft will be evenly distributed within a sector. The number of aircraft entering the sector will be constant, although the exact number has not been chosen yet. A single flight level will be considered. The sector-related factors (intersection points, reporting points, sector geometry, route measures) are selected based a previous research on predicting ATC workload based on 3D SSD [32]. The particular routes are chosen because crossing flights and takeovers can be simulated without excessive difficulties that may be caused by having a complex route structure. The sector shape is chosen to be a square, as it allows generations and comparison of different experiment scenarios to be simple and straightforward; as there will be 5 individual runs per subject, the sector can be rotated by 90 degrees for each run, with the 5th run being the identical sector shape as the first one with an altered traffic scenario. The sector with the traffic routes are shown in Figure 4.4. The dimensions of the sector is 40 NM by 40 NM.

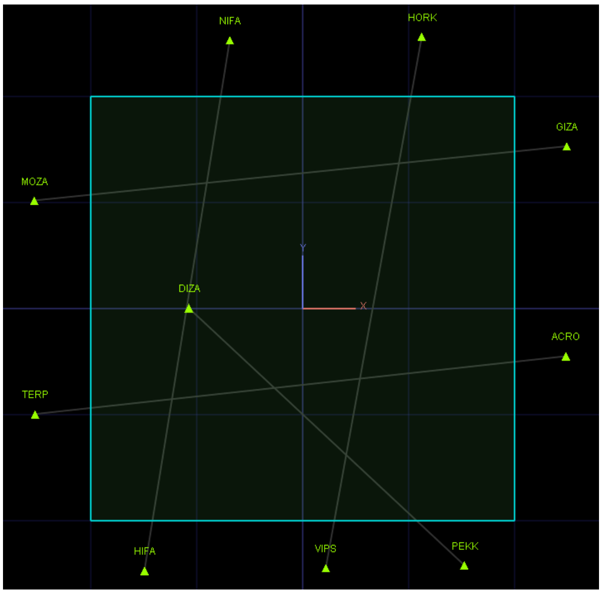


Figure 4.4: Sector and traffic routes

Traffic density will be constant, as it would simulate an ATCO taking over a shift of another ATCO in a real-life ATC environment. Each run will start with a scene with multiple aircraft already in the sector.

Automation failure types

Automation failure types are defined based on the stages of automation; the only the first stage of automation will be at the highest level, meaning that the automation acquiring information (e.g. airspeed, heading, aircraft position) and showing it to participants will still be valid. Then the rest of stages of automation stays at the lowest level, which means that participants would have to manually 1. analyze information, 2. select decision, and 3. implement action. The SSD will be disabled as information analysis automation goes to the lowest level. Automated CD&R system will also be switched off. When an automation failure occurs, participants will be notified with a visual and auditory alarm. The automation failure moment will be randomized such that it does not occur at an exact same moment of an experiment run.

4.2.3. Dependent Variables

Safety & control efficiency

Safety for this experiment is defined by a standard lateral separation on parallel or non-intersecting tracks or routes having a minimum permissible separation greater than 5 NM [33]. Thus, separation distance between aircraft during the manual control phase after automation failures will be measured as safety. Also the number of loss of separation events will be monitored and translated to safety.
Control efficiency relates to the performance metrics during the manual control phase. They are magnitude of airspeed, heading and path deviation, as a big change in airspeed is not desirable from a cost perspective

such as fuel. Average time spent to solve conflicts and the number of performed handovers will be measured as a part of control efficiency as well. Higher control safety and control efficiency metrics would correlate to higher manual control performances.

Workload

The standard Instantaneous Self Assessment (ISA) will be used to measure mental workload of participants for the real-time experiment. As the base ATC simulator has an implemented ISA workload assessment, it will be used. Participants can click on a scale bar which appears occasionally, and indicate their real-time workload. Also, they will be given with workload questions briefly in between each experiment run.

Task engagement level

Engagement level will be measured for both groups during the supervisory control phase. It will be mainly based on mouse events; for example, a participant may notice a potential conflict and click on a corresponding aircraft pair to inspect the SSD. The particular action of doing so can be regarded as an active effort to update his/her SA and mental picture. The following list of mouse events will be considered to determine engagement level of participants during the supervisory control phase:

- The number of valid SSD inspections on (potential) conflict pairs & corresponding flight labels
- The number of valid clicks on aircraft which receive any form of automated resolutions such as a speed or heading change

For the group with the task engagement tool, submitted responses and time spent to answer the questions will be recorded and analyzed.

4.2.4. Participants

Based on experiment matrices which can be seen in Figure 4.5, there will be in total of 20 participants. Red cells in the matrices indicate experiment runs with an automation failure, while white cells represent experiment runs with no failure. Participants will mostly be fellow Control & Simulation students at the aerospace engineering faculty. It is important to ensure that they are given with basic ATC knowledge and skills through a set of training runs before commencing the actual experiment.

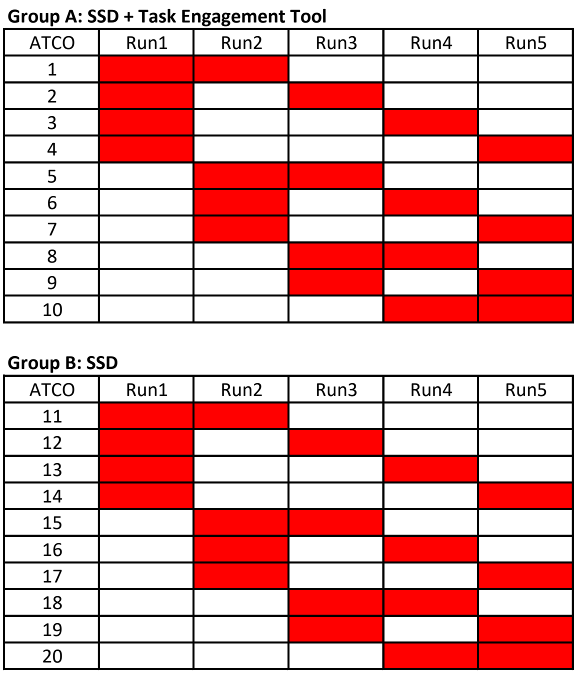


Figure 4.5: Experiment matrix

4.2.5. Experiment procedures

The experiment will begin with a briefing; participants will be given with an overview of how the experiment will proceed. Then there will be a brief training on a set of performing simple test CD&R scenarios and using the SSD in general. After the training, the actual testing will be performed for an hour in total. Then it is followed by a debriefing, to ensure disclosure of experiment details that may affect participants who have not yet completed it.

5 Conclusion and Research Planning

5.1. Preliminary Conclusion

This preliminary thesis research aimed to establish a basis for determining how to get ATCOs more engaged in their tasks, such that they can perform manual CD&R tasks safely and efficiently when automated ATC systems fail to function.

The literature study on the classical ironies of automation revealed how automation can leave ATCOs venerable in their manual control skills, cognitive skills and monitoring. Then stages of automation have been identified as an automation taxonomy to analyze, since the 4 separate stages of automation would allow a precise manipulation and analysis on disabling certain features of an automated ATC system. It was concluded that it would be the best to find a way to enhance SA in order to achieve the research aim.

The CWA was carried out to determine the scope of the research and a HITL experiment to be conducted. The outcomes of extensive en-route ATCO task analysis, strategies, and controllers' best practices were taken to a concept generation. With an idea of cycling human operators' SA by actively asking them 3 level SA questions, a concept so called task engagement tool was designed. Thus the task engagement tool would increase SA of the operators while they supervise a fully automated ATC simulation with a SSD augmentation, then the increase in SA would lead to improved manual ATC performances in terms of safety and control efficiency.

Finally, an experiment design was proposed and discussed. The main goal of the experiment is to assess whether providing real-time SA questions to ATCOs would lower the cognitive gap and increase their overall performances in presence of the transition from supervisory control to manual control when the automated system ceases to work.

5.2. Research Planning

First of all, a library of SA questions needs to be completed and validated based on en-route ATCO SA requirements. It can be done in parallel to implementing the task engagement tool to the ATC simulator platform. After the implementation of the task engagement tool, a group of participants for the actual experiment will be sought and a beta test will be performed on a student and supervisors to see if the experiment is ready to be performed. When it is deemed ready, the experiment will be performed and a thorough analysis on the outcomes will be performed to check if the suggested hypotheses are met. And finally, the analyzed data and conclusion of the research will be presented in a thesis report.

A

Experiment matrices and scenarios

The experiment matrices are shown in Figure A.1

	Failure timing	EARLY						LATE	
	SA level	SA1-2		SA1-2		SA1-2		SA1-2	
	scene #	scene 1A (10 min)		scene 2B (10 min)		scene 3A (10 min)		scene 4B (10 min)	
	status	AUTO	MANUAL	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL
	duration [min]	7	4	7	4	7	4	7	4
	break + questionnaires								
ATCo #1,5	Failure timing					LATE		EARLY	
	SA level	SA2-3		SA2-3		SA2-3		SA2-3	
	scene #	scene 2A (10 min)		scene 3B (10 min)		scene 1B (10 min)		scene 4A (10 min)	
	status	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL	AUTO	MANUAL
	duration [min]	7	4	7	4	7	4	7	4
	questionnaires								

	Failure timing					LATE		EARLY	
	SA level	SA2-3		SA2-3		SA2-3		SA2-3	
	scene #	scene 2B (10 min)		scene 3A (10 min)		scene 1A (10 min)		scene 4B (10 min)	
	status	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL	AUTO	MANUAL
	duration [min]	7	4	7	4	7	4	7	4
	break + questionnaires								
ATCo #2,6	Failure timing	EARLY						LATE	
	SA level	SA1-2		SA1-2		SA1-2		SA1-2	
	scene #	scene 1B (10 min)		scene 2A (10 min)		scene 3B (10 min)		scene 4A (10 min)	
	status	AUTO	MANUAL	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL
	duration [min]	7	4	7	4	7	4	7	4
	questionnaires								

	Failure timing	EARLY						LATE	
	SA level	SA2-3		SA2-3		SA2-3		SA2-3	
	scene #	scene 1A (10 min)		scene 2B (10 min)		scene 3A (10 min)		scene 4B (10 min)	
	status	AUTO	MANUAL	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL
	duration [min]	7	4	7	4	7	4	7	4
	break + questionnaires								
ATCo #3,7	Failure timing					LATE		EARLY	
	SA level	SA1-2		SA1-2		SA1-2		SA1-2	
	scene #	scene 2A (10 min)		scene 3B (10 min)		scene 1B (10 min)		scene 4A (10 min)	
	status	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL	AUTO	MANUAL
	duration [min]	7	4	7	4	7	4	7	4
	questionnaires								

	Failure timing					LATE		EARLY	
	SA level	SA1-2		SA1-2		SA1-2		SA1-2	
	scene #	scene 2B (10 min)		scene 3A (10 min)		scene 1A (10 min)		scene 4B (10 min)	
	status	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL	AUTO	MANUAL
	duration [min]	7	4	7	4	7	4	7	4
	break + questionnaires								
ATCo #4,8	Failure timing	EARLY						LATE	
	SA level	SA2-3		SA2-3		SA2-3		SA2-3	
	scene #	scene 1B (10 min)		scene 2A (10 min)		scene 3B (10 min)		scene 4A (10 min)	
	status	AUTO	MANUAL	AUTO	AUTO	AUTO	AUTO	AUTO	MANUAL
	duration [min]	7	4	7	4	7	4	7	4
	questionnaires								

Figure A.1: Experiment matrices

The four scenarios (with A and B variants) are shown in Figure A.2, Figure A.3, Figure A.4 and Figure A.5. The scene at the beginning and the end of each run is shown accordingly.

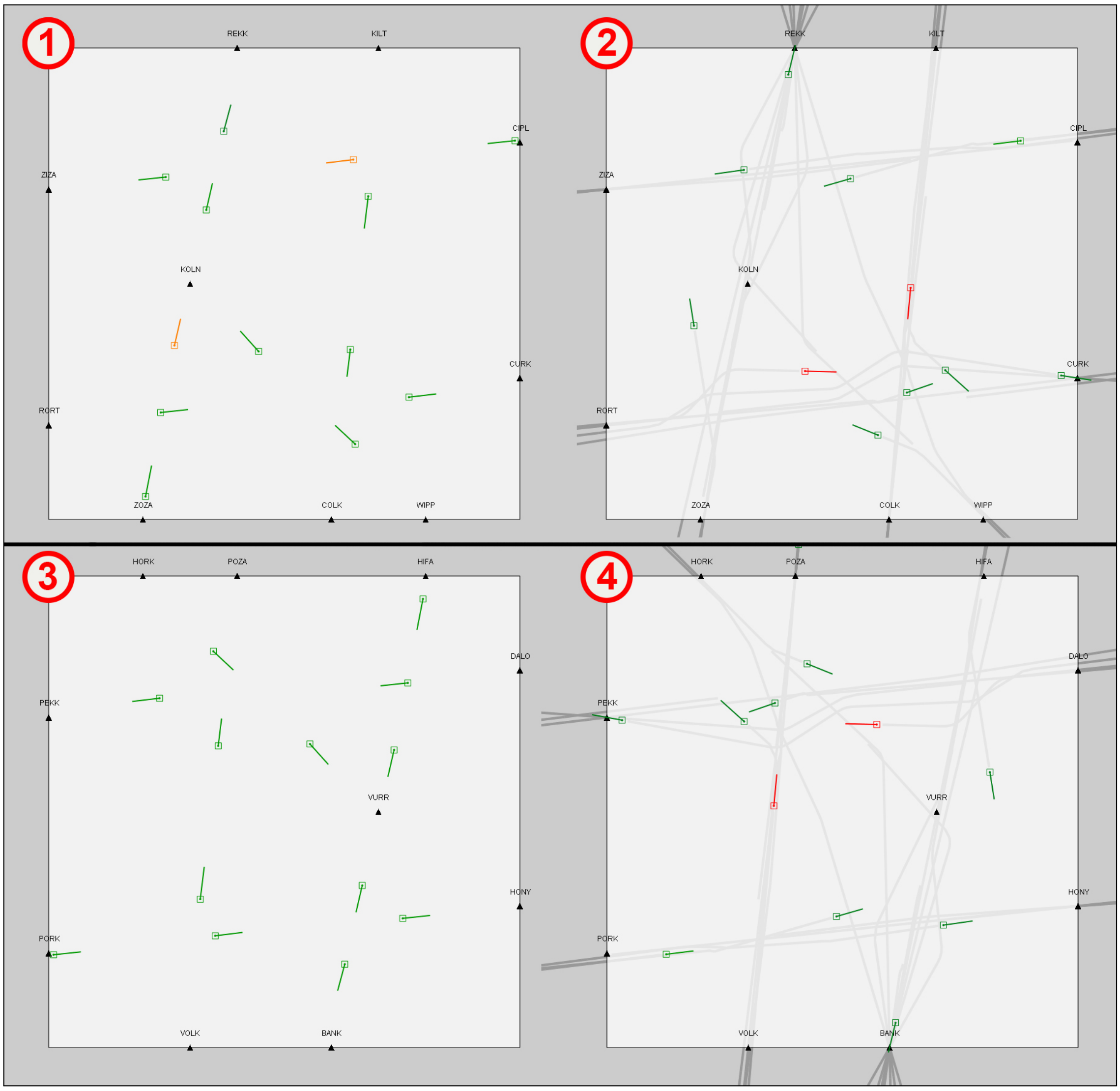


Figure A.2: Scene 1A and 1B. ①: = 1A at t = 0 s, ②: = 1A at t = 1,320 s, ③: = 1B at t = 0 s, ④: = 1B at t = 1,320 s

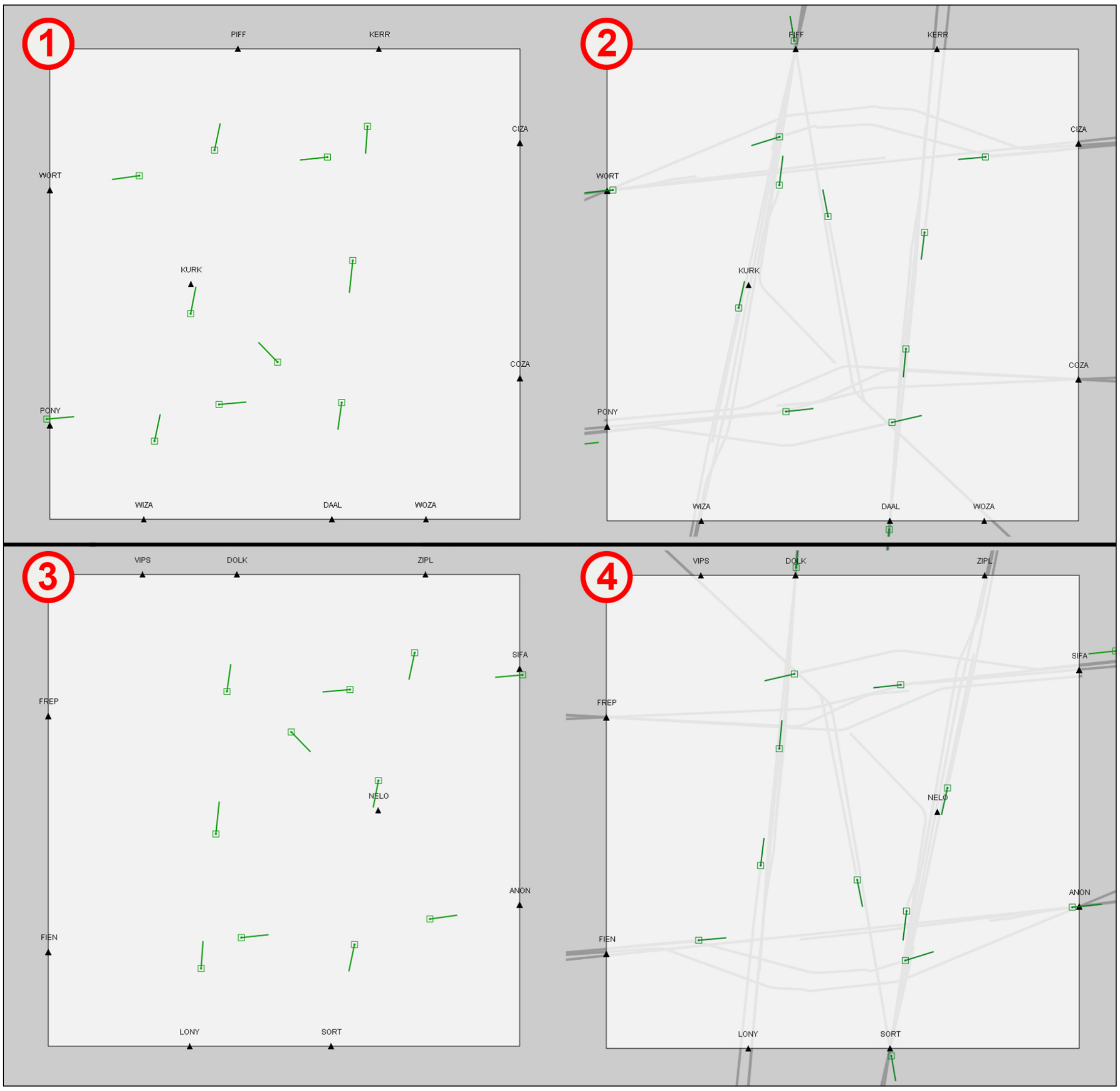


Figure A.3: Scene 2A and 2B. ①: = 2A at t = 0 s, ②: = 2A at t = 1,320 s, ③: = 2B at t = 0 s, ④: = 2B at t = 1,320 s

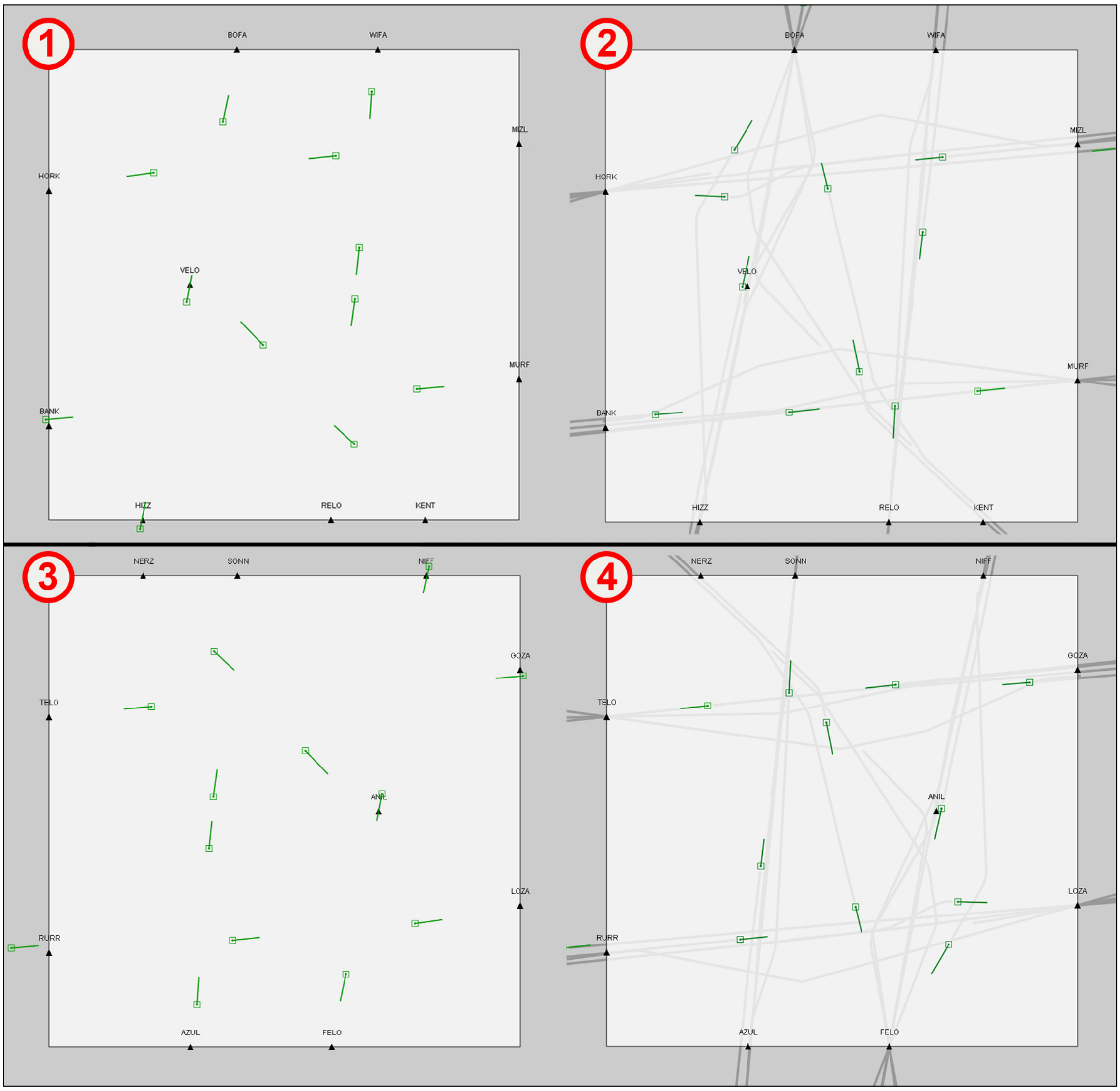


Figure A.4: Scene 3A and 3B. ①: = 3A at t = 0 s, ②: = 3A at t = 1,320 s, ③: = 3B at t = 0 s, ④: = 3B at t = 1,320 s

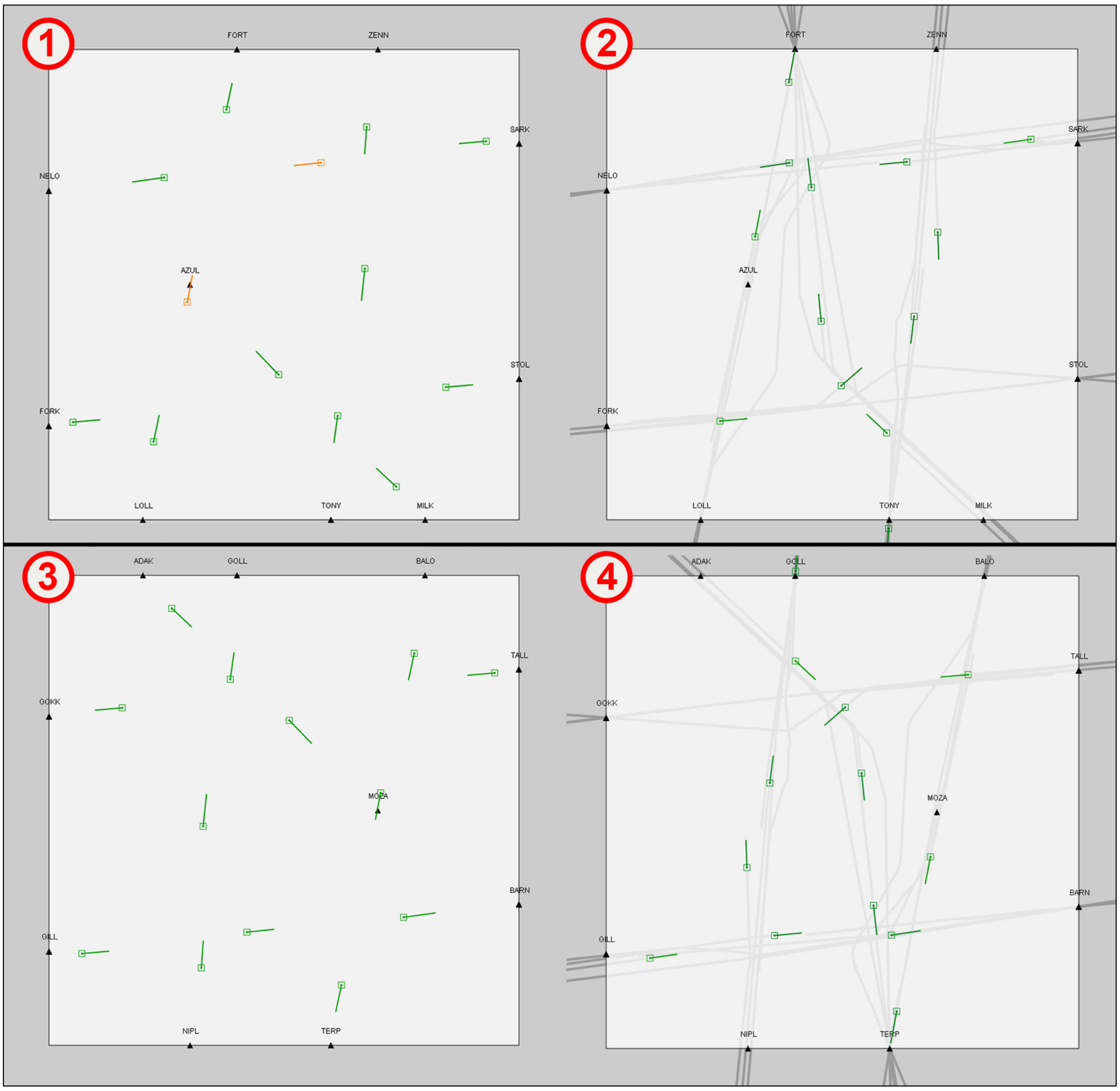


Figure A.5: Scene 4A and 4B. ①: = 4A at t = 0 s, ②: = 4A at t = 1,320 s, ③: = 4B at t = 0 s, ④: = 4B at t = 1,320 s

B

Simulator block diagram

Figure B.1 shows an overview of modifications given to the simulator for this thesis research. AUTOFAIL and FAILTIME variable nodes are added to `PlaylistConfig.java`, and `PlaylistXMLParser.java` reads AUTOFAIL and FAILTIME variables in playlist `xml` files. Scenario files of each playlist are loaded by `loadNextScenario()` in `SSDAPI.java`, then the aforementioned failure variables are turned into global variables in `GLSS.java`. `SimTimer.java` accesses the variables, and determines whether there is an automation failure for a specific scenario. If so, several `GLSS` variables are updated; they disable automated resolutions and the TET, enable manual control and play auditory alarm, which is done by `soundTET()` in `SSDAPI.java`.

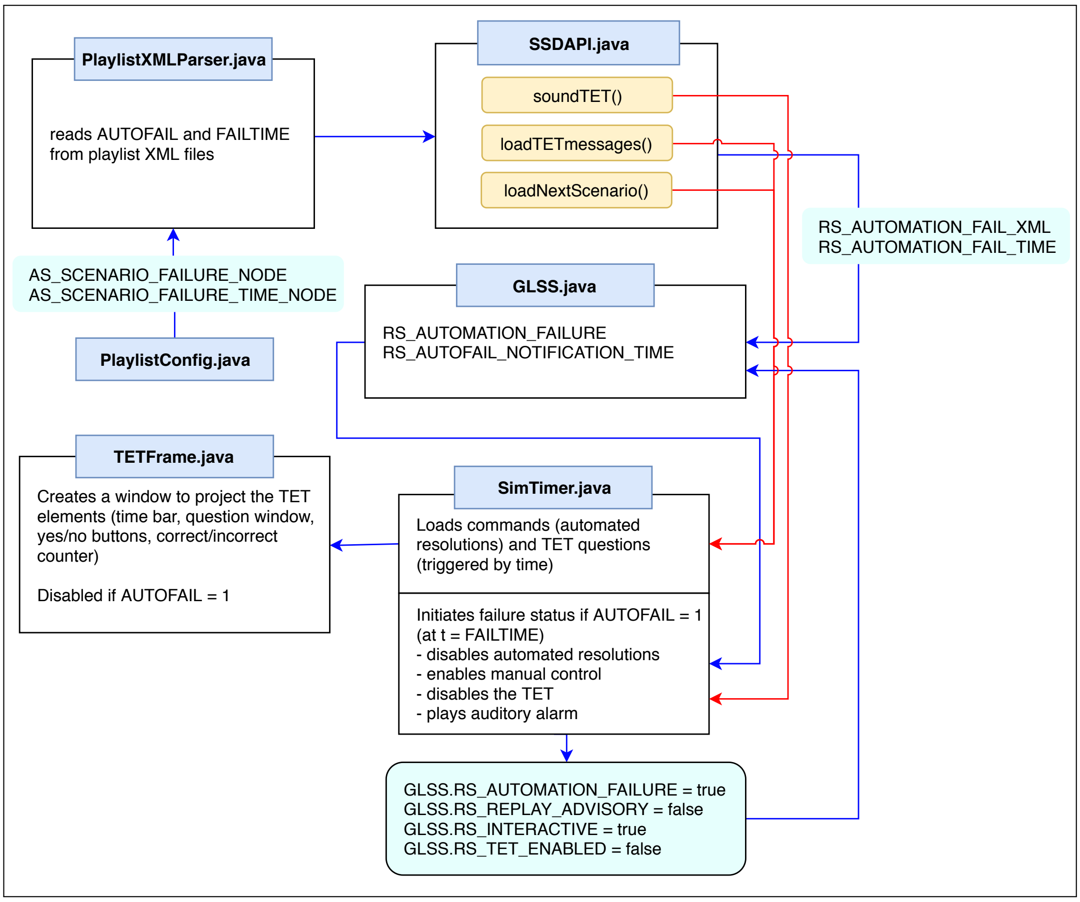


Figure B.1: Overview of work on the simulator

C
TET SA questions

The following lists show all questions used for the experiment, per run:
Scenario 1: SA1-2 with a failure

- Did the sector have [N] crossing points?
- Did [ACID] receive a heading change and a direct-to (DCT) command?
- Is [ACID] a heavy aircraft?
- Are there [N] aircraft headed to the exit waypoint [WAYPOINT] in the sector?

Scenario 2: SA1-2 with no failure

- Is airspeed of [ACID] [N]?
- Does the route [WAYPOINT]-[WAYPOINT] have the most number of aircraft in the sector at this given moment?
- Are [N] aircraft about to enter the current sector at this given moment?
- Is [ACID 1] in conflict with [ACID 2]?
- Is airspeed of [ACID] [N]?
- Did [ACID] receive a heading change?

Scenario 3: SA1-2 with no failure

- Is [ACID 1] flying slower than [ACID 2]?
- Did [ACID] receive an airspeed and heading change?
- Is [ACID] entering the sector at this given moment?
- Did [ACID] receive an airspeed change?
- Is [WAYPOINT] an exit waypoint for [ACID]?
- Is [ACID 1] in conflict with [ACID 2]?

Scenario 4: SA1-2 with a failure

- Are [N] aircraft leaving the sector at this given moment?
- Did [ACID] only receive a heading change?
- Is [ACID 1] flying faster than [ACID 2]?

- Does the bottom half of the sector have more traffic than the top half at this given moment?

Scenario 1: SA2-3 with a failure

- Will [ACID] enter the sector in next 70 seconds?
- Does [ACID 1] need to take over [ACID 2] to avoid a further conflict with [ACID 3]?
- Are there [N] aircraft on the route [WAYPOINT]-[WAYPOINT]?
- Will [ACID] experience any conflict if a heading or airspeed change is not given to it?

Scenario 2: SA2-3 with no failure

- Will [ACID] enter the sector in next [N] seconds?
- If [ACID 1] does not receive an airspeed change, will [ACID 1] be in conflict with [ACID 2]?
- Is [ACID] on course?
- Will a DCT command for[ACID] to the exit waypoint [WAYPOINT] result in a conflict?
- Is [ACID 1] in conflict with [ACID 2]?
- For [ACID 1], will a heading change to [N] result in a conflict with [ACID 2]?

Scenario 3: SA2-3 with no failure

- Did [ACID] receive a heading change?
- Will it take approximately [N] seconds before a conflict with [ACID 1] and [ACID 2] becomes critical?
- Are there currently [N] aircraft deviating from their direct routes between entry and exit waypoints?
- Will [ACID 1] experience conflicts with [ACID 2] and [ACID 3] before [ACID 1] leaves the sector?
- Are there [N] aircraft on the route [WAYPOINT]-[WAYPOINT]?
- Will a DCT command for [ACID] to the exit waypoint [WAYPOINT] result in a conflict?

Scenario 4: SA2-3 with a failure

- Will [ACID 1] in conflict with [ACID 2]?
- For [ACID 1], will an airspeed change from [N1] to [N2] result in a conflict with [ACID 2]?
- Did [ACID] receive an airspeed change?
- Will [ACID 1] require a combination of a heading and airspeed change to be clear of the incoming conflict with [ACID 2]?

The full list of TET SA questions per SA level can be found as following:
Level 1 SA

- Is [ACID] heading to [WAYPOINT]?
- Is [ACID] currently in the sector?
- Is [WAYPOINT] the exit waypoint for [ACID]?
- Is [ACID] a [TYPE] aircraft?
- Does the sector have [N] crossing points?
- Does the sector have [N] airways?
- Are there [N] aircraft in the sector?

- Is [ACID] entering/leaving the sector?
- Is [ACID 1] flying faster/slower than [ACID 2]?

Level 2 SA

- Did [ACID] receive a heading change?
- Did [ACID] receive an airspeed change?
- Did [ACID] receive a DCT command?
- Will [ACID] enter/leave the sector in next [N] seconds?
- Is [ACID 1] in conflict with [ACID 2]?
- Is the current traffic concentrated at the top/bottom half of the sector?
- Are there [N] aircraft on the route [WAYPOINT]-[WAYPOINT]?
- Are there currently [N] aircraft deviating from their direct routes between their entry and exit waypoints?
- Is [ACID] on course?
- Are there [N] aircraft headed to the exit waypoint [WAYPOINT] in the sector?

Level 3 SA

- For [ACID 1], will a heading change from [N1] to [N2] result in a conflict with [ACID 2]?
- For [ACID 1], will a speed change from [N1] to [N2] result in a conflict with [ACID 2]?
- If [ACID 1] does not receive an automated resolution, will [ACID 1] be in conflict with [ACID 2]?
- Will it take approximately [N] seconds before a conflict with [ACID 1] and [ACID 2] becomes critical (red)?
- Will [ACID 1] experience conflicts with [ACID 2] and [ACID 3] before [ACID 1] leaves the sector?
- Does [ACID 1] need to take over [ACID 2] to avoid further conflicts with other aircraft in the sector?
- Does [ACID] require a combination of a heading and airspeed change to be clear of the ongoing conflict?
- In [N] seconds, will there be a conflict on [WAYPOINT - WAYPOINT]?

D

Pre-experiment briefing

Before the experiment, the pre-experiment briefing was sent to all participants. The briefing document is shown in Figure D.1 and Figure D.2.

Pre-experiment briefing

Supervision of a Fully Automated Air Traffic Control System

This experiment aims to investigate the effects of having human air traffic controllers to supervise a fully automated Air Traffic Control (ATC) system. This pre-experiment briefing will show you the overview of the experiment setup, tasks to be performed, and the experiment procedure.

Apparatus

The experiment will take place in the practical computer area at SIMONA, located on the first floor of the SIMONA building. The lab uses a single LCD screen on which a simulated radar display of the simulation is shown. An example of the display can be seen in Figure 1.

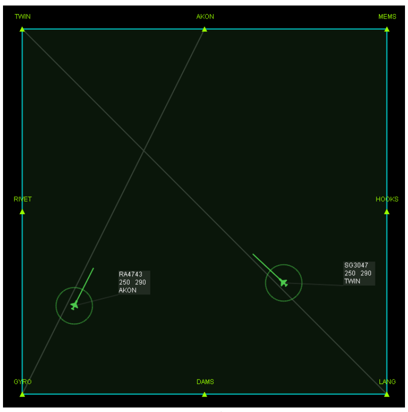


Figure 1. Example of a radar display

Control Task

For the experiment, you will assume the position of an air traffic controller who supervises a fully automated ATC system. While you are supervising the system, you do not have a direct control of aircraft.

Your main task is to supervise all aircraft in your responsible sector. The automated ATC system may cease to work, which means that you will become responsible for the following tasks:

Figure D.1: First page of pre-experiment briefing

1. Detect any potential conflict between aircraft and resolve the conflict by issuing heading and/or airspeed change to aircraft.
2. Maintain minimum separation of 5 nautical miles between aircraft at all times. Each circular zone around aircraft indicates a separation bubble of 2.5 nautical mile radius.

The tasks can be performed by "click & drag"; heading changes can be given to aircraft by first clicking an aircraft, dragging its speed vector to the desired heading around the aircraft, and pressing the ENTER key to execute the heading change. Airspeed changes can be performed in a similar manner by clicking an aircraft and using a mouse wheel to either increase or decrease airspeed within limits. Pressing the ENTER key is also required to execute the airspeed changes. Trainings on how to use the click & drag method will be given before the experiment.

Task Engagement Tool

Task Engagement Tool (TET) asks you yes/no questions related to ATC tasks to be performed. For example, TET may ask you "Is the airspeed of KLM9210 230?", and you will have 15 seconds to answer the question. TET will highlight corresponding aircraft labels if necessary. After pressing either the yes or no button, you can immediately see whether your response is correct or not so you can check your response on the go. Trainings on how to use TET will be given before the experiment.

Experiment Procedure

The experiment will take maximum 150 minutes to be completed. It will follow the following procedure:

1. Briefing about the experiment and tasks to be performed (10 minutes)
2. Training runs (30 minutes)
3. Measurement runs (110 minutes)
4. Debriefing of the experiment

After each set of measurement runs, you will be asked to fill in the following forms. You will be asked to fill in the form twice in total.

- SASHA questionnaire form: you will indicate your situation awareness during each set of runs.
- NASA Task Load Index (NASA-TLX) form: you will indicate workload experienced during each set of runs.
- SATI questionnaire form: you will indicate your trust in automation during each set of runs.

Please keep in mind that you need to fill in the forms carefully.

Thank you for your participation

Figure D.2: Second page of pre-experiment briefing

E

Questionnaires and Additional Results

The original SATI questionnaire [34] is shown in the following 6 pages:

SHAPE Automation Trust Index (SATI v0.3)

SATI Part 1 (*please complete before the start of the day's simulation runs*)

Please tell us who you are, and your forthcoming role in the simulation. Thank you.

About you:

Name:	
Nationality:	
Sex (M/F):	

About the simulation:

Date and time:	
Name of simulation project:	
Computer-assistance or automation tools available:	1. 2. 3. 4. 5.
Your simulated sector:	
Your role (planner / executive controller)	

SATI Part 1 (*continued*)

PLEASE COMPLETE AT THE START OF EACH DAY

1. What do you think of the simulation so far? (Please mark the scale with an 'X').

Bad	OK	Good
	|—|—|—|—|—|—|—|—|—|—|	

2. Are you prepared to trust the simulated system? Please give your reasons.

No	Yes

3. How much confidence do you have in the simulated system? (Please mark the scale with an 'X').

None	OK	Full
	|—|—|—|—|—|—|—|—|—|—|	
0%	50%	100%

4. Please give your reasons

SATI Part 2 (*please complete after the end of the simulation runs*)

Please write your name and your last role in the simulation. Thank you.

About you:

Name:	

About the simulation:

Date and time:	
Name of simulation project:	
Computer-assistance or automation tools available:	1. 2. 3. 4. 5.
Your last simulated sector:	
Your last role (planner / executive controller)	

SATI Part 2 (*continued*)

PLEASE COMPLETE AT THE END OF THE DAY'S RUNS

Based on today's runs

1. What did you think of the simulation? (Please mark the scale with an 'X').

Bad	**OK**	**Good**
|—|—|—|—|—|—|—|—|—|—|		

2. Were you prepared to trust the simulated system?

No	**Yes**

3. How much confidence did you have in the simulated system? (Please mark the scale with an 'X').

None	**OK**	**Full**
|—|—|—|—|—|—|—|—|—|—|		
0%	**50%**	**100%**

4. Please give your reasons. If your trust or level of confidence in the system has changed since the start of the day, please explain why.

SATI Part 2 (*continued*)

PLEASE COMPLETE A SEPARATE SHEET FOR EACH AVAILABLE AUTOMATION TOOL.

5. Please judge each automation tool against the following factors (mark each scale with an 'X').

Name of automation tool: ______________________________

1. Is the automation tool useful?

Not useful	- 5 ... 0 ... + 5	*Useful*

2. How reliable is it?

Not reliable	- 5 ... 0 ... + 5	*Reliable*

3. How accurately does it work?

Not accurate	- 5 ... 0 ... + 5	*Accurate*

4. Can you understand how it works?

Not understand	- 5 ... 0 ... + 5	*Understand*

5. Do you like using it?

Dislike	- 5 ... 0 ... + 5	*Like*

6. How easy is it to use?

Difficult	- 5 ... 0 ... + 5	*Easy*

6. Please rank these factors in order of relative importance. Number them from 1 (*least important)* to 6 (*most important*). Please use each number once only.

Name of automation tool:__________________________

Usefulness	*ranking:*
Reliability	*ranking:*
Accuracy	*ranking:*
Understanding	*ranking:*
Liking	*ranking:*
Ease of use	*ranking:*

SATI Part 2 (*continued*)

LOOKING BACK OVER THE DAY'S SIMULATION RUNS:

7. Please rate your amount of confidence in each of these five dimensions.
Please mark each scale with an 'X'.

1. Confidence in automation tools

0 50 100 %

2. Confidence in simulation

0 50 100 %

3. Self-confidence

0 50 100 %

4. Confidence in controller colleagues

0 50 100 %

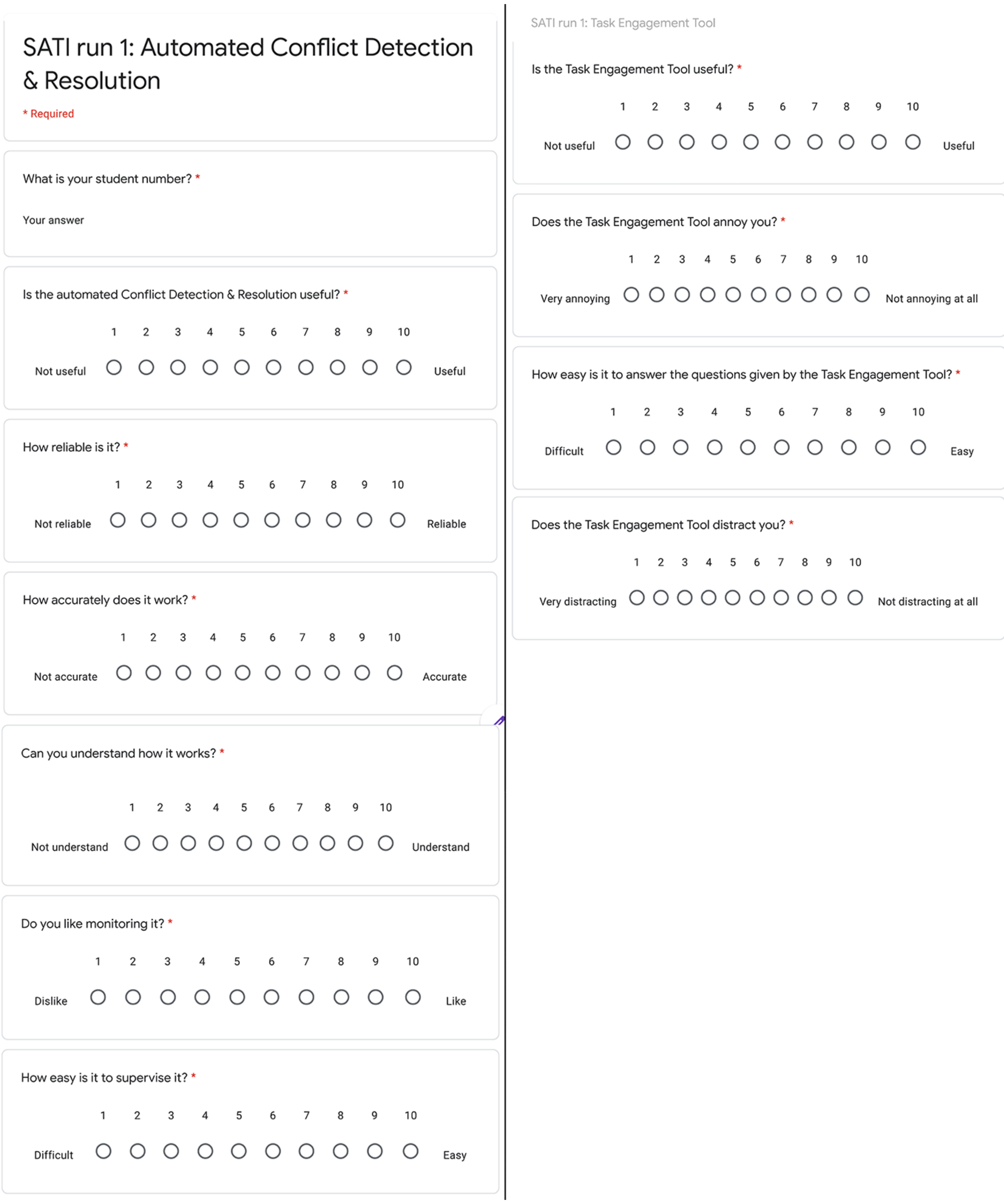
5. Confidence in pilots

0 50 100 %

8. Would you work live traffic with the tools? In your opinion, what changes would the automation need so that your trust and confidence would be increased?
If there are any other factors which influence your trust in an ATC system, or if you have any general comments, please write them here.

Thank you for completing this questionnaire.

Adapted SATI questionnaire used for the experiment are shown as following. Note that the Google Form was used for the adapted SATI questionnaires. Questions regarding the confidence and validity of the simulation were taken out, as they are meant for experienced ATCOs.



SATI run 1: Automated Conflict Detection & Resolution

* Required

What is your student number? *

Your answer

Is the automated Conflict Detection & Resolution useful? *

Not useful 1 2 3 4 5 6 7 8 9 10 Useful

How reliable is it? *

Not reliable 1 2 3 4 5 6 7 8 9 10 Reliable

How accurately does it work? *

Not accurate 1 2 3 4 5 6 7 8 9 10 Accurate

Can you understand how it works? *

Not understand 1 2 3 4 5 6 7 8 9 10 Understand

Do you like monitoring it? *

Dislike 1 2 3 4 5 6 7 8 9 10 Like

How easy is it to supervise it? *

Difficult 1 2 3 4 5 6 7 8 9 10 Easy

SATI run 1: Task Engagement Tool

Is the Task Engagement Tool useful? *

Not useful 1 2 3 4 5 6 7 8 9 10 Useful

Does the Task Engagement Tool annoy you? *

Very annoying 1 2 3 4 5 6 7 8 9 10 Not annoying at all

How easy is it to answer the questions given by the Task Engagement Tool? *

Difficult 1 2 3 4 5 6 7 8 9 10 Easy

Does the Task Engagement Tool distract you? *

Very distracting 1 2 3 4 5 6 7 8 9 10 Not distracting at all

Figure E.1: Adapted SATI questionnaire

The original SASHA questionnaire [35] is shown in the following 2 pages:

APPENDIX A: SASHA QUESTIONNAIRE

Introduction

Computer-assistance tools and other forms of automation support are being increasingly introduced into today's Air Traffic Management (ATM) systems, and are expected to be fundamental components of systems in the future. The success of such automated tool support will depend in part on the degree to which Human Factors are taken into account in the design and implementation of these tools.

As part of the overall European ATM Programme (EATMP) the Human Factors and Manpower Unit[12] within EUROCONTROL has recently initiated a new programme of work to address the human factors issues of automation in ATM systems. The programme is called 'SHAPE' (for 'Solutions for Human-Automation Partnerships in European ATM'). The present aim of SHAPE is to develop a number of measurement techniques that can be applied during real-time simulations to assess and measure the effectiveness of the automation.

This questionnaire is concerned with measuring your 'situation awareness'. It consists of ten questions. Please answer each question by ticking the box as appropriate. Add any other comments in the space provided.

Thank you for your assistance and cooperation.

SASHA Questionnaire
Q1: - Did you have the feeling that you were ahead of the traffic, able to predict the evolution of the traffic? Never ☐ ☐ ☐ ☐ ☐ Always Comments:…………………………………………………………………………………… …………………………………………………………………………………………… ……………………………………………………………………………………………
Q2: - Did you have the feeling that you were able to plan and organise your work as you wanted? Never ☐ ☐ ☐ ☐ ☐ Always Comments:…………………………………………………………………………………… …………………………………………………………………………………………… ……………………………………………………………………………………………
Q3: - Have you been surprised by an a/c call that you were not expecting? Never ☐ ☐ ☐ ☐ ☐ Often Comments:…………………………………………………………………………………… …………………………………………………………………………………………… ……………………………………………………………………………………………

[12] today known as 'Human Factors Management Business Division (DAS/HUM)'.

Q4: - Did you have the feeling of starting to focus too much on a single problem and/or area of the sector?

Never ☐ ☐ ☐ ☐ ☐ Often

Comments:…………………………………………………………………………………………………
………
………

Q5: - Did you forget to transfer any aircraft?

Never ☐ ☐ ☐ ☐ ☐ Often

Comments:…………………………………………………………………………………………………
………
………

Q6: - Did you have any difficulty finding an item of (static) information?

Never ☐ ☐ ☐ ☐ ☐ Always

Comments:…………………………………………………………………………………………………
………
………

Q7: - Do you think the <*name of tool*> provided you with useful information?

Never ☐ ☐ ☐ ☐ ☐ Always

Comments:…………………………………………………………………………………………………
………
………

Q8: - Were you paying too much attention to the functioning of the <*name of tool*>?

Never ☐ ☐ ☐ ☐ ☐ Always

Comments:…………………………………………………………………………………………………
………
………

Q9: - Did the <*name of tool*> help you to have a better understanding of the situation?

Never ☐ ☐ ☐ ☐ ☐ Always

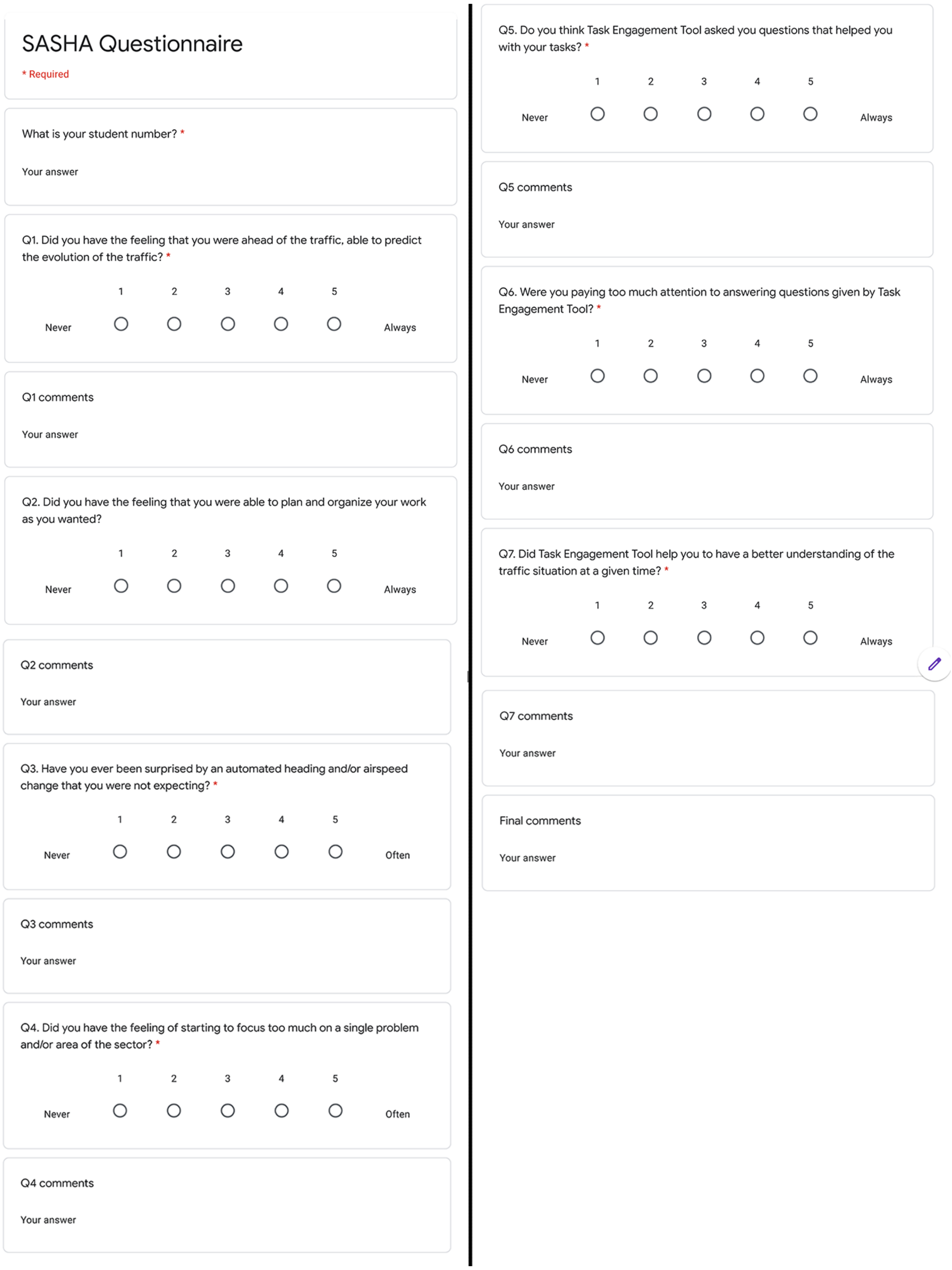
Comments:…………………………………………………………………………………………………
………
………

Q10: - Finally, how would you rate your overall situation awareness during this exercise?

Poor ☐ Quite poor ☐ Okay ☐ Quite good ☐ Very good ☐

Comments:…………………………………………………………………………………………………
………
………

The adapted SASHA questionnaire can be found as following:



SASHA Questionnaire

* Required

What is your student number? *

Your answer

Q1. Did you have the feeling that you were ahead of the traffic, able to predict the evolution of the traffic? *

Never 1 2 3 4 5 Always

Q1 comments

Your answer

Q2. Did you have the feeling that you were able to plan and organize your work as you wanted?

Never 1 2 3 4 5 Always

Q2 comments

Your answer

Q3. Have you ever been surprised by an automated heading and/or airspeed change that you were not expecting? *

Never 1 2 3 4 5 Often

Q3 comments

Your answer

Q4. Did you have the feeling of starting to focus too much on a single problem and/or area of the sector? *

Never 1 2 3 4 5 Often

Q4 comments

Your answer

Q5. Do you think Task Engagement Tool asked you questions that helped you with your tasks? *

Never 1 2 3 4 5 Always

Q5 comments

Your answer

Q6. Were you paying too much attention to answering questions given by Task Engagement Tool? *

Never 1 2 3 4 5 Always

Q6 comments

Your answer

Q7. Did Task Engagement Tool help you to have a better understanding of the traffic situation at a given time? *

Never 1 2 3 4 5 Always

Q7 comments

Your answer

Final comments

Your answer

Figure E.2: Adapted SASHA questionnaire

Frequency of each response scale (for SA1-2 and SA2-3) and comments for each question are given as following:

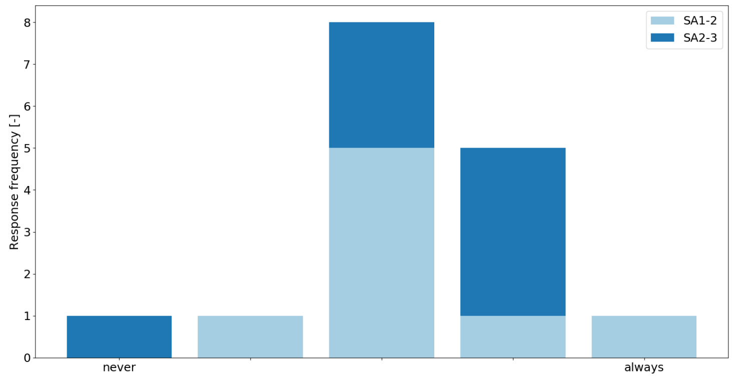


Figure E.3: Question 1: Did you have the feeling that you were ahead of the traffic, able to predict the evolution of the traffic?

Q1 comments

- Estimating the timing/speed was especially difficult.
- I relied on automation for predicting. Once it turned off I was able to predict traffic but needed time to adapt.
- I found it hard to estimate their speed.

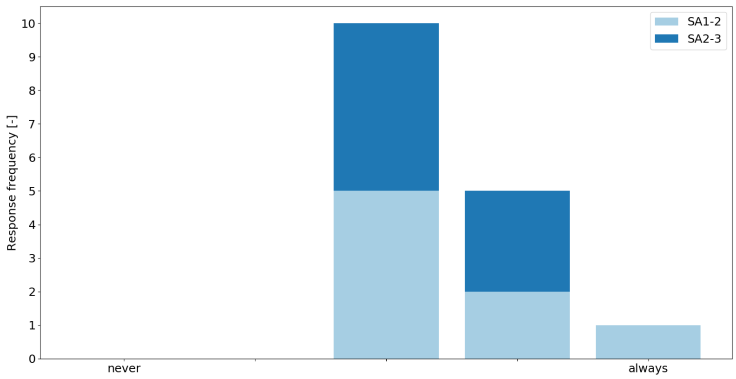


Figure E.4: Question 1: Did you have the feeling that you were able to plan and organize your work as you wanted?

Q2 comments

- When I had made a nice plan for two conflicting aircraft, a third aircraft would get involved in the conflict so I had to change my plan. Also I sometimes clicked outside the circle, so I had to re-apply my change. I wasted some time on that, in which the situation had already changed before I could make the change.
- Usually I was but under high traffic situations if a conflict is not resolved with an initial command it would need extra attention. Sometimes it might be because I am unable to predict the traffic and other times because the label interferes with given commands.

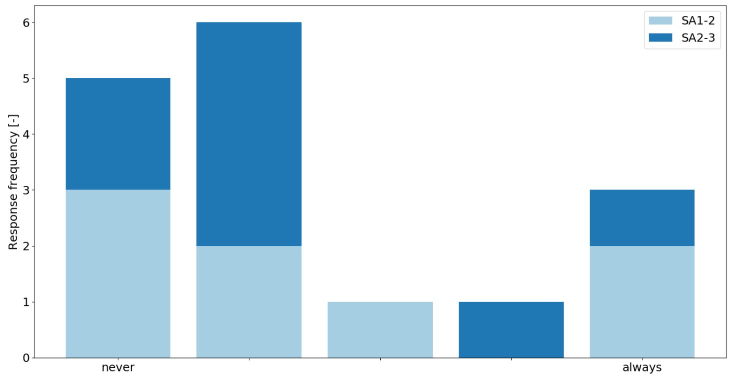


Figure E.5: Question 3: Have you ever been surprised by an automated heading and/or airspeed change that you were not expecting?

Q3 comments

- Sometimes the computer was fixing a conflict that I didn't yet identify/anticipate.
- I didn't pay too much attention to what automation is actually commanding. I trusted the automation to perform it's task as intended.
- Automation did more or less what I would expect.
- I was also sometimes expecting the automation to direct an aircraft back to the waypoint but then it didn't and the aircraft went of track more than I would have allowed.
- In manual mode the changes sometimes did not reflect what i wanted to (pressed outside of circle for example).

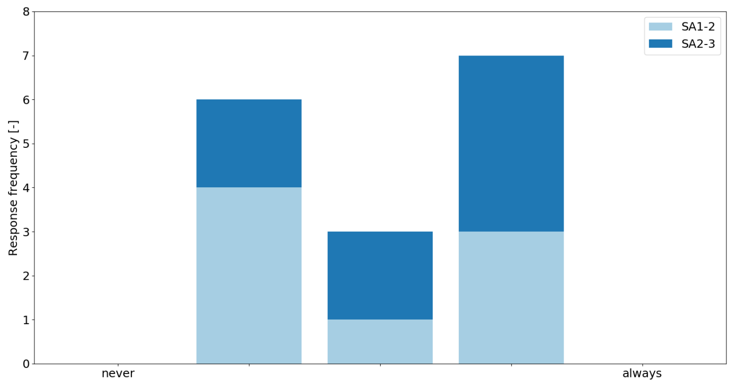


Figure E.6: Question 4: Did you have the feeling of starting to focus too much on a single problem and/or area of the sector?

Q4 comments

- With manual control, yes.
- Yes, when there's a conflict I'm totally focused on that problem and not paying any attention to the other areas. I was surprised by a second conflict for example.
- When a conflict is not resolved with an initial command I had some attention spotlight but understood that it needs to be corrected quickly so attention can be spent on other areas.

- while the automation was off i was focused on a single conflict and did not focus on possible future conflicts.

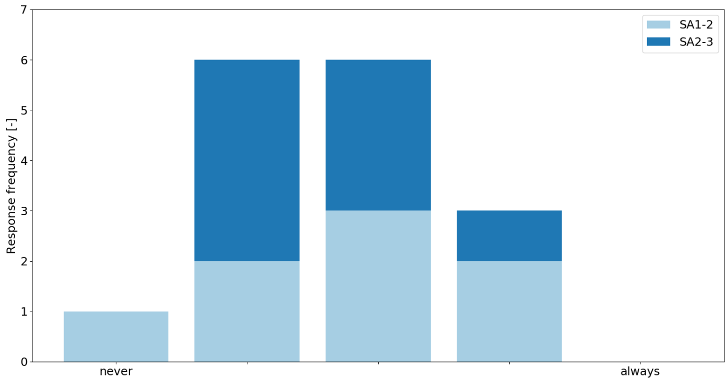


Figure E.7: Question 5: Do you think Task Engagement Tool asked you questions that helped you with your tasks?

Q5 comments

- I had that feeling one time during a scenario.
- Only the "will these two aircraft conflict" and "is a speed and heading change required to clear the conflict" where helpful. Questions about time till entry and number of aircraft on a route were not helpful.
- I do not believe that the questions that the task engagement tool were all of equal value or of equal difficulty. Sometimes I didn't understand what the tool was asking and sometimes it would be a very easy question. Sometimes the tool would also ask to count some number of objects which would take some time and sometimes I felt the questions were not clear enough to answer properly.
- I do however feel that the tool was able to interrupt my dazing off so I believe there may lie some value in it.
- The questions asked in the TET were not very task-relevant and focused on trivial matters that did not aid in understanding the traffic better.

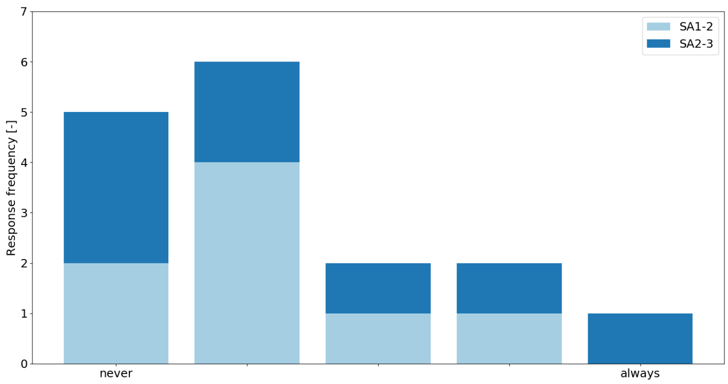


Figure E.8: Question 6: Were you paying too much attention to answering questions given by Task Engagement Tool?

Q6 comments

- Not to the effect where I had the feeling that I was missing information in the evolution of the traffic.
- The questions sometimes did take more time than I wanted.
- As mentioned, sometimes the tool would ask questions that might take some time to answer. These questions might have been too distracting but it is difficult for me to say. Each different question would have different requirements for attention.
- When the questions were asked, there wasn't an intense situation so I could pay attention to the questions.

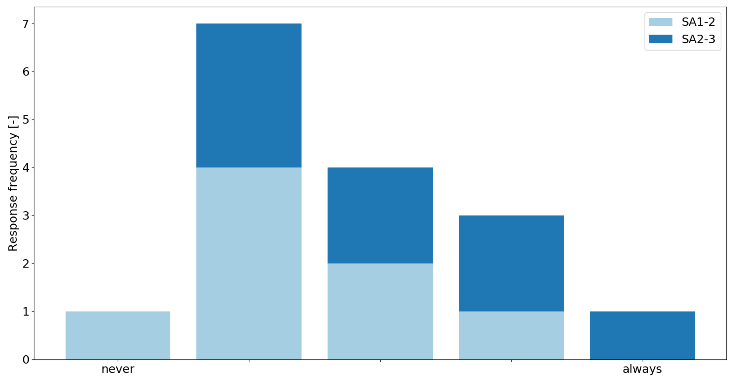


Figure E.9: Question 7: Did Task Engagement Tool help you to have a better understanding of the traffic situation at a given time?

Q7 comments

- It forces you to focus on a specific conflict or aircraft pair in the sector, leading to you actively reasoning about that pair, that can lead to a better insight (i.e., check the more detailed information such as speeds, headings, crossing geometry, etc.)
- Very rarely. Again I don't think that each question that the tool asks has the same value. I think that a separate study would need to see how each type of question engages the user. I also believe that the trust in automation may have been interfering with the interest in knowing the current traffic situation, regardless of whether questions are being asked or not.
- Very rarely.
- It did make me try to look more at the automation changes, but I didn't really care what the automation was exactly doing to be honest, as long as it did a safe job.

Final comments on overall SA

- When the automation was in control, I did not have the feeling that I was on top of the traffic, the TET did help on focusing on specific areas.
- Could do better. I've got the feeling that I am improving during the experiment, as I don't have any considerable experience using this kind of interface.
- Quite poor situation awareness when automation is active, quite good when it fails. But the time between when it is active and right after it fails I believe that it is still quite poor which can be dangerous.
- The automation was more in line with what I expected than in the other half of the experiment.

Then the final questionnaire on the TET is shown as following:
Q1: Did you have any difficulty finding information linked to questions given by the Task Engagement Tool?

- Yes, very hard to judge when something will happen (time) and what should happen (combination of speed and heading command).
- Depended on the question, some weird trivial (thus easy) such as 'is this the speed of aircraft X', some were more ambiguous such as 'are there X amount of aircraft almost about to leave the sector.'
- I think once or twice I had to zoom out to view the aircraft that the question was about. It was sometimes also impossible to find whether the automation gave heading 'and' speed changes, because the aircraft had already performed the changes when the question was asked. So if you didn't look at it in the first place and remembered the changes till the question was asked, you were unable to answer this question. I failed those quite a lot I think :-(
- yes, sometimes I felt questions were ambiguous.
- Sometimes when more than two aircraft were mentioned, it took me longer to identify each one.
- Yes, the questions which asked if an aircraft would enter the sector within a certain time: a time was not given - had to estimate. Same with if an aircraft needs to maneuver to avoid a conflict - had to guess.
- Sometimes, for example when something was asked about the heading change that already happened.
- No, the questions were clear.

Q2: Were you paying too much attention answering questions of the Task Engagement Tool?

- The time that I spent on the TET was more or less workload dependent. When the traffic was predictable, I would try to double check questions. When not, I would just make sure to have an answer before the required time.
- No, the questions only came in when there was a workable traffic situation.
- Not too much, I felt that in the time it took me to answer the question, the automation would take care of traffic. I wasn't too bothered by the amount of time it took to answer.
- No, I prioritized collision avoidance.
- In the first round yes. In the second round I made an effort to maintain situational awareness of the entire sector even when answering the questions.
- No, I thought it involved me more.
- No.

Q3: Did you feel that the Task Engagement Tool helped you performing your manual control tasks?

- Sometimes. If the question was about how many aircraft on a flow, I'd have to interact with the HMI to find the answer and I gained a lot more situational awareness than just "guessing" about what if scenarios.
- In some cases es, in other cases completely not. When the questions were task-relevant, referring to aircraft that would have actual conflicts, or to busy areas in the sectors, the TAT could push me to reason a bit more about the traffic. In other cases no, when the question was not related to any potential focus area in the sector.
- Slightly. It did keep me somewhat alert as to which aircraft were about to get into conflict. But most of the questions were about the automation I think and not so interesting for the manual control part.
- Not at all. I felt that automated control and supervisory control were completely separate from each other.
- It helped keep me in the loop when automation was active, so I was more aware when automation failed and I had to take over.

- Not really, although sometimes I would be focusing too much on a specific area and the engagement tool would remind to look at the rest of the sector.
- Yes.
- In the second run the questions were more useful.

Q4: If there is something that you would like to see changed about the task engagement tool, what would it be?

- More questions about gaining situational awareness instead of“guessing" outcomes of what if style questions.
- As mentioned previously, the TET could be more task-specific, for instance, posing questions about aircraft that the CD&R algorithm is just about to control. Or asking more specific questions about potential issues and multi-aircraft interactions (i.e., ’will this become a conflict’, or ’would it be better to steer aircraft X behind aircraft Y’).
- Maybe indicate when an aircraft from the question is outside the current zoom level. I think it would also help if the route for example would light up when the question is about a certain route. On the other hand that could also make it too easy to answer the question.
- Research into what each question asks the user to think about and how long it may take them to answer it.
- Less questions about more than 2 aircraft.
- Wouldn’t change anything.
- Sometimes questions were asked when a aircraft was just entering or leaving the sector. Whether this aircraft was included in the question or not was not sure.
- No idea.

Bibliography

[1] Raja Parasuraman, Thomas B. Sheridan, Christopher D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics*, (3):286–297, 2000.

[2] Siti Mariam binti Abdul Rahman. *Solution Space-based Approach to Assess Sector Complexity in Air Traffic Control*. PhD thesis, Delft University of Technology, 2 2014.

[3] Ann M. Bisantz, Catherine M. Burns. *Applications of Cognitive Work Analysis*. CRC Press, 2008. pp 39-41.

[4] Peter Brooker. Air traffic control automation: for humans or people? *Human Factors and Aerospace Safety*, (5):23–42, 2005.

[5] Eurocontrol Experimental Centre. Phare highly interactive problem solver. Technical Report 1, Eurocontrol, 11 1994.

[6] European ATM master plan: executive view. Technical report, European Union, Eurocontrol, 2015.

[7] The future of the nas. Technical report, U.S. Department of Transportation, Federal Aviation Administration, 2016.

[8] Paola Tomasello et al. STRESS final project results report. Technical Report 3, EU, Eurocontrol, 6 2018.

[9] Jan Kraemer, Francesca De Crescenzio, Bruno Berberian, Oliver Ohneiser, and Giannluca Di Flumeri. MINIMA final project results report. Technical Report 2, EU, Eurocontrol, 5 2018.

[10] Clark Borst, Vincent A. Bijsterbosch, M. M. van Paassen, and Max Mulder. Ecological interface design: supporting fault diagnosis of automated device in a supervisory air traffic control task. *Cogn Tech Work*, 19(4):545–560, 2017.

[11] Reising DVC, Sanderson PM. Minimal instrumentation may compromise failure diagnosis with an ecological interface. *Human Factors and Ergonomics Society*, (2):316–333, 2004.

[12] Barry Strauch. Ironies of automation: Still unresolved after all these years. *IEEE Transactions on Human-Machine Systems*, 48(5):419–433, 2018.

[13] Lisanne Bainbridge. Ironies of automation. *Automatica*, 19(6):775–779, 1983.

[14] Sandra G. Hart, Thomas E. Wempe. Cockpit display of traffic information: Airline pilots' opinions about content, symbology, and format. Technical report, NASA, 8 1979.

[15] A. Mundra. A new automation aid to air traffic controllers for improving airport capacity. Technical report, The Mitre Corporation, 1979.

[16] Wesley A. Olson, Nadine B. Sarter. Management by consent in human-machine systems: When and why it breaks down. *Human Factors*, (43):255–266, 2001.

[17] S. W. A. Dekker, D. D. Woods. To intervene or not to intervene: The dilemma of management by exception. *Cognition, Technology & Work*, (1):86–96, 1999.

[18] Mica R. Endsley. Toward a theory of situation awareness in dynamic systems. *Human Factors*, (37):32–64, 1995.

[19] Mica R. Endsley. Measurement of situation awareness in dynamic systems. *Human Factors*, 37(1):65–84, 1995.

[20] N.B. Sarter, D.D. Woods. Situation awareness: A critical but ill-defined phenomenon. *International Journal of Aviation Psychology*, (1):45–57, 1991.

[21] Jens Rasmussen. Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on System, Man, and Cybernetics*, (3):257–266, 1983.

[22] K. Wolfgang Kallus, Dominique van Damme, and Andrea Dittmann. Integrated task and job analysis of air traffic controllers - phase 2: Task analysis of en-route controllers. Technical Report 1, Eurocontrol, 8 1999.

[23] M. IJtsma. Adaptive automation based on air traffic controller's decision-making. Master's thesis, Delft University of Technology, 6 2016.

[24] Barry Kirwan and Mary Flynn. Investigating air traffic controller conflict resolution strategies. Technical Report 1, Eurocontrol, 3 2014.

[25] Jean-Francois D'Arcy and Pamela S. Della Rocco. Air traffic control specialist decision making and strategic planning - a field survey. Technical report, U.S. Department of Transportation, Federal Aviation Administration, 3 2001.

[26] P. Amaldi, M. Leroux. Selecting relevant information in a complex environment: The case of air traffic control. *5th European Conference on Cognitive Science Approaches in Process Control*, pages 89–98, 1995.

[27] T. L. Seamster, R. E. Redding, J. R. Cannon, J. M. Ryder, and J. A. Purcell. Cognitive task analysis of expertise in air traffic control. *The International Journal of Avition Psychology*, 3(4):257–283, 1993.

[28] Mica R. Endsley and Mark D. Rodgers. Situation awareness information requirements for en route air traffic control. Technical report, U.S. Department of Transportation, Federal Aviation Administration, 1 1995.

[29] M. Steinfeld H. Yanco M. Desai, P. Medvedev. Impact of robot failures and feedback on real-time trust. *IEEE International Conference on Human-Robot Interaction*, pages 251–258, 2015.

[30] Dianne Dukette David Cornish. *The Essential 20: Twenty Components of an Excellent Health Care Team*. RoseDog Books, 2009. pp 72-73.

[31] Indramani L. Singh Raja Parasuraman, Robbert Molloy. Performance consequences of automation-induced complacency. *The International Journal of Aviation Psychology*, (3):1–23, 1983.

[32] V.L.J. Somers. 3d solution space-based prediction of air traffic control workload. Master's thesis, Delft University of Technology, 5 2017.

[33] Doc 4444: Procedures for air navigation services. Technical report, International Civil Aviation Organization, 11 2016.

[34] P. Goillau, C. Kelly, M. Boarrdman, and E. Jeannot. Guidelines for trust in future atm systems: Measures. Technical report, Eurocontrol, 5 2003.

[35] E. Jeannot, C. Kelly, and D. Thompson. The development of situation awareness measures in atm systems. Technical report, Eurocontrol, 6 2003.