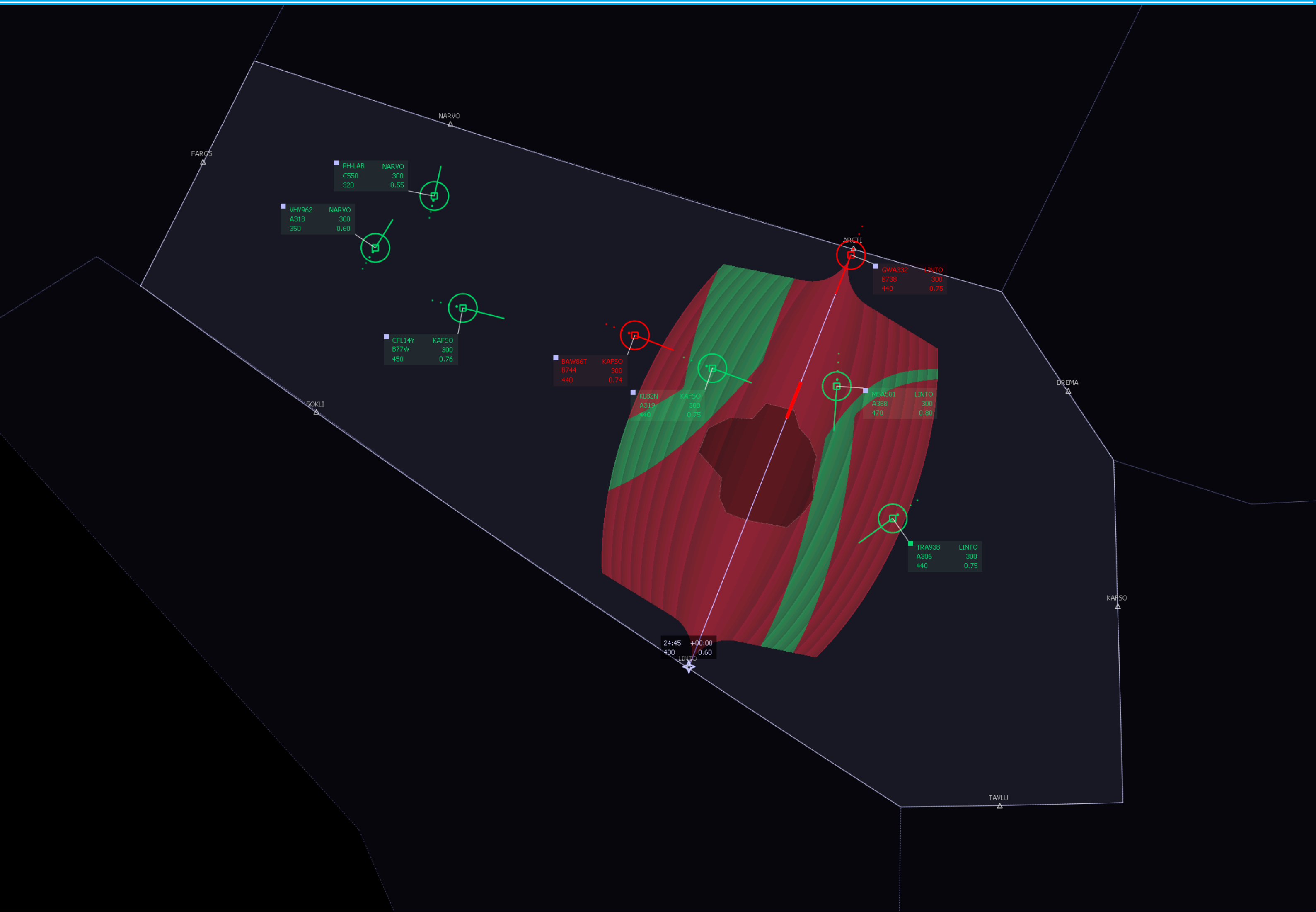
Human Factors testing results of a joint cognitive system for 4D trajectory management

Customer

SESAR Joint Undertaking

NLR-TP-2013-272 - March 2014



***NLR** - Dedicated to innovation in aerospace*

National Aerospace Laboratory NLR
Anthony Fokkerweg 2
1059 CM Amsterdam
The Netherlands
Tel +31 (0)88 511 3113
www.nlr.nl

UNCLASSIFIED

EXECUTIVE SUMMARY

Human Factors testing results of a joint cognitive system for 4D trajectory management

Report no.
NLR-TP-2013-272

Author(s)
A. Maij
A. Van Drunen
T.J.J. Bos
R.E. Klomp

Report classification
UNCLASSIFIED

Date
March 2014

Knowledge area(s)
Luchtverkeersmanagement(ATM)- en luchthavensimulatie en -validatie
Luchtverkeersmanagement(ATM)- en luchthavenoperaties
Training, Missiesimulatie en Operator Performance

Descriptor(s)
ATM
Human Factors testing
perturbation management
4D trajectories
Levels of Automation

Problem area

To keep up with the (sometimes local) increasing demand for air transport new means are necessary to maintain safe skies over Europe. The development of, and the research into these means are for a large part enabled by technological advances on the air and ground side which enable 4Dimensional operations. This vision is currently supported by the SESAR and NextGen research programmes and is foreseen to bring a paradigm shift in the work of the Air Traffic Controllers (ATCos). The first step towards 4D operations entails directing aircraft to crossroads in the sky at which they need to be at predetermined time slots. Therefore, ATCos need to ensure that aircraft adhere to strict schedules regarding the 4 dimensions when leaving their control area. This is a clear extension of their current role and is further complicated by the dynamic nature of the environment caused by changing weather, (temporary) restricted airspace(s) and increasing numbers of air movements. Automation can support

This report is based on a presentation held at the ATOS Conference, Toulouse, July 8-10, 2013.

the ATCo by taking over tasks or parts of tasks. In the past however, the implementation of automated support tools was often held back by a gap between the expectations of ATCos and the solutions to problems that were proposed. The Cognition through Shared Representations (C-SHARE) project aims to bridge this gap by using representations that keep the ATCo fully in the loop.

Description of work

The C-SHARE consortium has developed a system to support ATCos in their task of en-route control. This work was done within SESAR's long-term and innovative research initiatives Work Package-E. The developed system automatizes ATCo tasks while keeping the controllers fully in the loop. The system incorporates a novel representation, travel space information, which enables ATCos to (visually) re-plan air traffic within the performance, timing and separation constraints prevalent from the work domain. It uses waypoints that the ATCo can place in the control area and provides advanced system support about where these waypoints can be placed while still ensuring a timely transfer to neighboring control areas. The travel space visualization also indicates areas that may cause loss of separation. Finally, on request, the system provides information on the best possible rerouting options using an algorithm based on the 'Rapid Random Tree' (RRT) algorithm, widely used for path planning in robots. RRTs ensure preventing loss of separation but are also efficient in terms of fuel usage and environmental impact.

Automatizing operator tasks can result in several operator performance problems, such as trust, acceptance, and situation awareness problems. Therefore, it is important that the system keeps the operator in the loop. The operator needs to understand the information presented and what it implies. Also, the operator needs to use this information to predict future situations. Finally, it is crucial that the operator understands the reasoning of the automation to mitigate human-automation coordination breakdowns. To ensure that the operator needs are met, Human Factors testing in early design phases is of vital importance. The goal of this research is therefore to perform Human Factors testing of the C-SHARE system.

The goals of Human Factors testing are to identify how users experience the system and to identify which elements of the system require improvement. It is carried out with ATCos who perform typical task activities. Observers record the interactions between operators and the system and analyze the data. Preferably, Human Factors testing is part of a human centered design in which several development iterations, including testing, are done. Therefore, the C-SHARE project consists of three development phases, each of which is concluded with a test. The first phase was the conceptual development phase followed by conceptual testing. The second development phase and testing has been concluded and will be discussed in this document. The current status of the project is the start of the third and final development phase, which will be followed by final testing.

To prevent extensive testing of an immature system with scarce Subject Matter Experts, the results from the conceptual phase were transformed into prototypes which were used for a cognitive walkthrough with experts in the area of operator performance. The first development phase included creating an interactive software-based simulation of the system that facilitates the execution of scenarios. It was concluded with Human Factors testing with three ATCos from Eurocontrol (Brétigny).

Results and conclusions

The results show that the system is suitable in aiding ATCos in their future tasks. ATCos indicated that they were positive about placing waypoints, using travel space information and using RRTs to support their decision making process. They also stated that the system has good potential for future implementation after a thorough operational validation and verification procedure. Also, the Human Factors testing revealed elements of the system that should be improved in the final development phase. These elements can be divided into improvements in:

- controlling the system;
- using waypoints;
- warnings; and
- automation-generated solutions.

Applicability

SESAR's WP-E is aimed at long term and innovative research. The JCS is far from ready for implementation, but the road towards creating it has been paved in this project. Further research should be aimed at creating a user interface that supports changes in altitudes and at investigates how the JCS can be implemented in the future.

Automated systems are becoming more and more difficult to understand and therefore, to operate. The automation can help by taking over tasks or parts of tasks to aid operators, but they need to be kept in the loop to be able to troubleshoot when necessary. Adequate knowledge of joint cognitive systems allows us to improve future designs of systems that require operators to work with sophisticated automation. Also, Human Factors testing will get an even more important role in the development of joint cognitive systems in the future as it ensures that the operator can perform his tasks in cooperation with the automation.

National Aerospace Laboratory NLR
Anthony Fokkerweg 2, 1059 CM Amsterdam,
P.O. Box 90502, 1006 BM Amsterdam, The Netherlands
Telephone +31 (0)88 511 31 13, Fax +31 (0)88 511 32 10, Website: www.nlr.nl

Human Factors testing results of a joint cognitive system for 4D trajectory management

A. Maij, A. Van Drunen[1], T.J.J. Bos and R.E. Klomp[2]

[1] Thales Research and Technology
[2] Delft University of Technology

Customer
SESAR Joint Undertaking
March 2014

Human Factors testing results of a joint cognitive system for 4D trajectory management

This report is based on a presentation held at the ATOS Conference, Toulouse, July 8-10, 2013.

This publication has been refereed by the Advisory Committee AIR TRANSPORT.

Customer	SESAR Joint Undertaking
Contract number	10-220719-C
Owner	NLR + partner(s)
Division NLR	Air Transport
Distribution	Unlimited
Classification of title	Unclassified
Date	March 2014

Approved by:

Author	Reviewer	Managing department
Anneloes Maij	Henk van Dijk	Harrie Bohnen
Date 2-4-14	Date 3/4/14	Date 9/4/14

Abstract

To keep up with the (sometimes) local increasing demand for air transport, new means are necessary to maintain safe skies over Europe. This paper presents the Human Factors testing results of a joint cognitive system that is being developed for SESARs long-term and innovative research initiatives. The joint cognitive system distributes control tasks within the Air Traffic Control system over operator and automation. The system's goal is to jointly manage 4-dimensional trajectories, defined in both space and time. It provides information on areas to which aircraft can safely be rerouted to prevent perturbations without causing delays. Also, on request, the automation can advise the intended user on the best rerouting options in terms of fuel efficiency, safety, and environmental impact. Increasing levels of automation can cause surprises resulting in decreases in operator performance, such as losing situation awareness and distrusting the system. Human Factors testing can help designing an optimal cooperation between system and intended user. Therefore, this study investigated Air Traffic Controllers performing typical task activities with the new system. The results show that users are positive about the system and that it shows good potential for a future implementation. Also, the Human Factors testing revealed aspects that need more attention to improve the system. These elements can be divided into improvements in:

- controlling the system;
- using waypoints;
- warnings; and
- automation-generated solutions.

This page is intentionally left blank.

Content

Abbreviations

Abbreviation	Description
4D	4 Dimensional
AC	Aircraft
ATCo	Air Traffic Controller
ATM	Air Traffic Management
CSE	Cognitive Systems Engineering
CTO	Controlled Time Over
EID	Ecological Interface Design
JCS	Joint Cognitive System
NLR	National Aerospace Laboratory
WDA	Work Domain Analysis
WP	Waypoint

1 Introduction

To keep up with the increasing demand for air transport new means are necessary to maintain safe skies over Europe. The development of, and the research into these means are for a large part enabled by technological advances on the air and ground side which enable 4D operations. In fact, a re-distribution of tasks within the socio-technical system of Air Traffic Management (ATM) may be necessary, enabled through higher levels of automation. Introduction of such automation often introduces new problems in a complex socio-technical system [1]. Such problems can be caused by a lack of understanding between automation and human agents: humans may have difficulty getting the automation to do what they want, and the way the automation works is often poorly understood by the human operator.

The C-SHARE project provides a solution to the problem of understanding the automation. The project does that by starting with the design of a common representation, which are used by both the human actor and the automation system for generating and evaluating solutions. The user interface and the automated tools use this representation as a common ‘model of the world’ in the joint effort of safely and efficiently managing the air traffic. This was coined the Joint Cognitive Representation which is part of the framework of a Joint Cognitive System (JCS) entailing the system and its human actor. The term JCS is lent from the field of Cognitive Systems Engineering.

The C-SHARE project follows an iterative development process consisting of three design cycles that each includes a testing phase. This article describes the test phase of the second cycle. The introduction outlines the fundaments of the prototype design.

2 Background

2.1 Cognitive Systems Engineering

Cognitive Systems Engineering (CSE) emerged in the early 1980s as a theoretical framework to analyse the performance of socio-technical systems within safety-critical domains, such as aviation and the nuclear power domain. Within the framework of CSE, socio-technical systems are identified to be so-called Joint Cognitive Systems (JCSs) [2]. JCS is a system that consists of two components of which at least one is a cognitive system [2]. A cognitive system is a system that can modify its behaviour based on past experience to achieve specific goals even under disruptive influences (normally the human operator is one of the cognitive components in such a system). JCSs are in control of a process or an environment, and act in complex situations, which requires the balancing of multiple goals to meet the contextual demands.

The underlying principle of the JCS as proposed in the C-SHARE project is based upon presenting a common ground -or shared representation- of the work domain of future 4D ATM. This common ground has been identified through performing a Work Domain Analysis (WDA, [3]) and defines the relevant elements, functions, constraints and opportunities within the environment on various levels of abstraction. The WDA reveals what needs to be done and how this can be achieved irrespective of whom should actually perform this task, the human operator(s), the automation, or both.

The en-route tactical revision of 4D trajectories was taken as the main control task which is to be effectively achieved by means of joint human-automation coordination. It is foreseen that when a shared representation based upon this common ground is somehow made visible in a JCS, it will act as a basis for enabling productive collaboration, team thinking and a deeper understanding between the cognition of the operator and the automated agents.

2.2 Travel Space Representation

As a starting point, a prototype of a constraint-based shared representation was designed for the task of the in-flight manipulation and revision of 4D trajectories by Air Traffic Control (ATC). The manipulation and placement (position and timing) of such waypoints is a task to be shared between the human users and the automation. Re-planning of waypoints might become necessary in case one or more inherent or intentional constraints active on the aircraft's trajectory cannot be satisfied due to any number of unforeseen events in a prior planning phase. Inherent constraints can be for example other traffic, terrain, weather and intentional constraints such as restricted airspace or procedures.

2.2.1 Safe Field of Travel

When considering the task of in-flight re-planning of a 4D trajectory by ATC, the relevant functions and constraints which govern the work domain can be derived from the WDA. The overall goal of the aircraft is to execute its subsequent trajectory segments and pass all waypoints within the timing constraints in agreement with ATC. Now consider that either the aircrew or air traffic controller intends to introduce an intermediate waypoint into a trajectory segment, for instance to solve a conflict with another aircraft. Any arbitrary (4D-)placement of that waypoint will lead to a new definition of the trajectory. The feasibility of such a trajectory can be tested against the relevant constraints which govern the work domain (e.g., adherence to locomotion, obstruction and perturbation management [4]). Then, the subset of all trajectories, which adhere to these constraints, are feasible solutions and, by definition, forms a so called 'safe field of travel'.

2.2.2 Graphical Representation

By means of one-to-one mapping, a correspondence-driven [5] translation of this safe field of travel can be made on the air traffic controller's plan view display, indicating the real-world spatial locations of feasible waypoint placements and their timing implications. For the design of a shared representation, the Ecological Interface Design (EID-) [6] approach has been adopted. EID is a design method elaborated within the framework of CSE, and argues that for effective human-machine interface design, the constraints and underlying relationships governing the work domain should be somehow made visible to the human operator. In the present project it is hypothesized that, by visualizing the task-relevant functional constraints that arise from the work-domain, and using these same constraints to guide automated actions, humans will get a deeper understanding of why automation proposes a particular action.

2.2.3 Representation Breakdown

Figure 1 shows the basic composition of the travel space representation. Aircraft AC1 is flying along a pre-agreed 4D trajectory towards a certain metering fix (point FIX) at the sector border. The Controlled Time Over (CTO) at the fix is taken as a hard constraint (i.e., it must be met). When considering constraints which follow from the aircraft performance envelope (in combination with the time constraint at the fix), an area can be bounded in which intermediate waypoint placement is feasible. The aircraft turn characteristics determine the rounded shape of the travel space close to the current aircraft position and the metering fix. Furthermore, any intermediate waypoint that does not lie directly on the current trajectory segment implies an increase in track length, and thus an increase in required ground speed. The outer edges of the travel space are therefore bounded by the maximum achievable speed within the aircraft performance envelope.

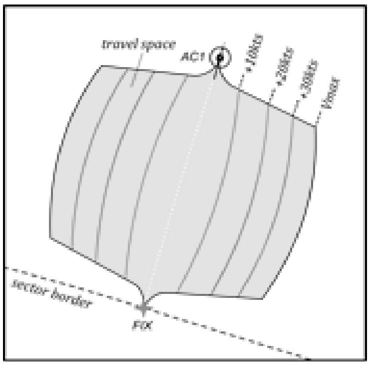


Figure 1. Travel Space Representation for a single aircraft

Figure 2 introduces other traffic in the form of a second aircraft (AC2). When taking the separation constraints for both aircraft into account, an area within the travel space for AC1 becomes restricted (i.e., an intermediate waypoint in that area will result in a 4D trajectory which is in conflict with the other aircraft at a certain point in time). This area is indicated in the figure as the restricted field of travel. The restrictive area visualizes the locations where an intermediate waypoint would not lead to the resolution of the conflict.

The above procedure is only applicable when aircraft deviate from their planned 4D trajectory. To ensure the planned time over the fix is still met, the speed of an aircraft will need to deviate from its optimal speed leading to a (small) increased fuel usage. The upside is that because the aircraft still meets its planned time at the fix, there are no emergent effects on the stability of the 4D system as a whole and therefore trajectories of other aircraft are not affected. The speed increase to ensure timely arrival at the fix will obviously lie within the speed envelope of the aircraft. For example, the difference in normal cruise speed and maximum speed of a Boeing 737 is about 12% (780 versus 876 km/h). This means that in theory the path towards the fix can be stretched by 12%. For a segment of 100NM the path may then be stretched up to 112NM at most.

Figure 2 also shows how the travel space representation can be used by the human or automation to select an appropriate position for an intermediate waypoint in a conflict situation. By placing the waypoint (WP1) inside the safe field of travel within the travel space, the constraints following from aircraft performance, separation, and timing are all met. Note that the timing of the introduced waypoint is set such that it corresponds with the constraints visualized by the representation (e.g., constant speed along both segments and fixed timing at the final waypoint).

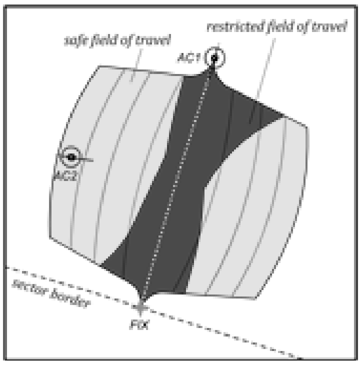


Figure 2. Travel Space Representation with traffic causing restricted field

2.2.4 Automation-generated Solutions

Taking the concept one step further, the automation is also able to suggest potential solutions. The travel space allows for maximum freedom for the controller to choose a solution. While such freedom undisputedly has benefits, in more challenging or complex scenarios, it may take time to interpret the travel space and select a proper solution. Using the same criteria as the travel space, the automation is able to suggest one or a few potential new routes for flights in conflict, all within the travel space, outside the restricted field of travel.

Using an approach loosely based on the Rapid Random Trees algorithm [6], the automation is able to automatically generate solutions to conflicts while taking into account the following criteria:

- Timely detect and solve potential conflicts such that the impact on the 4D trajectory is minimized;
- Solve by adding a single waypoint;
- Present only the most efficient solution;
- Ensure that an aircraft is able to continue its intended 4D trajectory after solving the conflict;
- Plan around forbidden areas (closed airspace, bad weather etc.), either clockwise or anti-clockwise, such that a predictable stream of traffic is created.

The suggested solutions are presented on controller demand and plotted in the Travel Space, presented by the light line in figure 3. As the proposed solutions are presented in and based upon the same rules as the Travel Space, a controller can quickly interpret the suggestion and either accept it or create his own solution. When the ATCo choses to create his own solution he can still use the auto-generated solution as a guide.

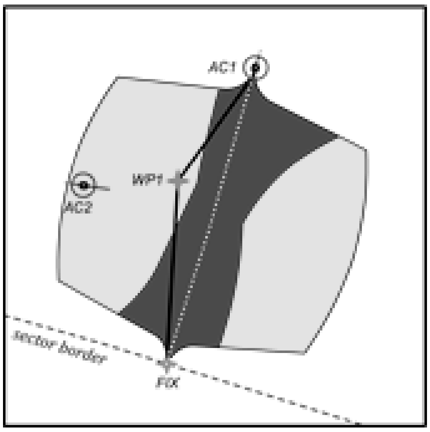


Figure 3. Travel Space Representation with intermediate waypoint ensuring deconfliction

2.3 Human Factors Research

Human Factors research is commonly perceived as the investigation of a system's effectiveness, efficiency, and the emotional response it provokes [7]. Thus, it answers questions such as: does the system deliver what it was built for; how many steps are required to perform certain actions with the system; and do potential users see the benefit of using the system? The goal of Human Factors testing is to identify which elements of the system are eligible for improvement. Advantages of Human Factors testing can be found in the areas of system performance, lower costs for training and support and increased user satisfaction [8].

If developers fail to perform Human Factors testing they risk missing important flaws in their design. Once implemented, users may search for shortcuts to deal with these flaws causing non-optimal system performance or the users may not be willing to use the system at all. Also, when flaws go unnoticed in a safety sensitive environment, such as ATC, they may contribute to serious incidents and accidents. Thus, proper Human Factors testing is an essential part of development. However, even though much literature is devoted to elements of Human Factors testing and specific evaluation set-ups, there does not seem to be one 'best practice' for Human Factors testing [9]. To still ensure fruitful and meaningful Human Factors testing without this one path to success, decisions need to be taken between best elements of Human Factors testing, such as number of participants, means for data gathering and data analysis.

The discussion about how many participants should be involved in Human Factors testing has not yet been concluded. Nielsen [10] claims that three to five subjects are sufficient to identify 60% of usability problems. Other researchers, however, emphasize that more participants are required [11], [12]. Nielsen also emphasizes that it is more important to work iteratively [13], such as developing a prototype, testing this prototype, and going back to the drawing board to develop the next prototype, etc. than to have one big Human Factors test only once in the (or worse, after the) development of a system.

Methods for gathering information on Human Factors range from using custom-made questionnaires to observing a representative sample of the target group executing scenarios comparable to their intended use of the system. Most Human Factors testing can be categorized into approaches in the lines of being heuristic evaluations, cognitive walkthroughs or user evaluations [9]. Heuristic evaluations consist of representatives of the target population of the system or Human Factors experts checking if the system complies with a number of usability heuristics, such as the quality of system status information, consistency and error prevention. During a cognitive walkthrough Human Factors specialists or intended users 'walk through' a system by executing tasks using (a mock-up of) a user interface, meanwhile using a 'think-aloud' protocol. This provides insight in how the user interface is understood by the participant. User evaluations use a sample of representatives of the target group of a system to perform system typical tasks to find flaws in the design.

All three approaches have their advantages and disadvantages and are therefore often combined to create the best possible cost-benefit ratio of Human Factors testing.

2.3.1 Human Factors Test in the First C-SHARE Development Cycle

As stated previously the C-SHARE project consists of three development cycles followed by a testing phase. The first cycle resulted in a prototype that was tested early in the design process [4]. Its objective was to evaluate the intuitiveness of the visualisations and to validate the underlying principles. This, in turn, led to improvements of the concept and its representations. For this first evaluation researchers in aviation were the invited participants, consisting of ATM experts and Human Factors experts. In 1.5 hour-sessions, a participant 'walked through' the system with two C-SHARE team members, one of which operated the prototype. An interactive computer-based implementation of the travel space representation and system-generated routes formed the basis of the evaluation. Three types of conflicts between flights under varying geometry (head-on, crossing, and take-over situation) were presented. Flights were visualized by an aircraft symbol with a speed vector, protected zone, history dots, and label. The demo was paused showing a conflict situation (as a consequence of a perturbation) allowing a stepwise walk-through. The indication of the conflict (by means of colour changes to track symbol and label) was evaluated and the sequential steps in visualizing the conflict and potential solutions to the conflict. The travel space diagram was presented to the participants. Sequentially, system generated solutions were presented which were presented by an early implementation of the system-generated conflict resolution algorithm. Additionally, other concepts were presented amongst which an indicator of costs of rerouting and a control means to reroute several flights at once. Most participants welcomed the novel concept of the travel space diagram as giving insight in the solutions for conflicts. Also the system-generated solutions were judged to meet an additional need for a quick solution in case of high-workload situations. An important improvement identified by the participants was that the solutions generated should follow the principles of the travel space diagram, which was not the case in that early implementation. The opinions on the other concepts presented varied and it was decided not to include these in a next version. In addition, the conceptual evaluation identified several user interface aspects that led to an improved next version of the concept.

2.3.2 The Second C-SHARE Development Cycle

The second development phase resulted in a user interface that presented simulated air traffic and that could be controlled by the participants. This version of the JCS was used for the second evaluation. The goals of this evaluation were to identify flaws in the system, to prioritize them and to decide which would be implemented in the JCS during the third development phase. The following parts of this document describe how this Human Factors testing was done, how the resulting information was analysed and how it was used as input for the third development phase. It also describes how the third testing sessions will be organized.

3 Methodology

3.1 Participants

One former and two active ATCos from EUROCONTROL's Research Centre in Brétigny were invited to participate in the evaluation. The selection was largely based on experience with ATC which ranged from 15 years to 25 years. All three ATCos were familiar with the SESAR program and its goals and they had excellent knowledge and extensive experience in various aspects in ATC. All participants also have experience in participating in Human Factors testing.

3.2 Materials

The first questionnaire that was used was an introductory questionnaire for gathering demographic information and information about experience and knowledge of ATC. The second was based on the Shape ATM Trust Index which measures trust in a system that aids human operators. This questionnaire was presented together with the situation awareness assessment to measure if the participant maintained awareness of what was happening in his control area. All questionnaires were presented on paper.

The experiment leader handed out consent forms and instructions on paper as well. Participants received step-by-step test cases for the first five scenarios which were intended for training and for explaining the concept, its visualisations and how to interact with the user interface. The participants were placed in the position of tactical controller/ trajectory manager in a 4D trajectory based concept and were guided through the first five scenarios by the instructions and when necessary, aided by the experiment leader. The first five scenarios included the placing and moving of waypoints in the control area; confronting controllers with travel space information and with auto-generated solutions.

The user interface was presented on a full-colour monitor and was operated with keyboard and mouse (figure 4). Operators could use the mouse to zoom in and out and move their focus point (by moving the airspace). An aircraft was selected or deselected by clicking on its label or the aircraft symbol. A movable label connected to aircraft indicating ground speed, type, and carrier was shown attached to each aircraft. Rerouting was done by adding or manipulating waypoints after selecting an aircraft. Waypoints could be added, moved and deleted with the mouse and the keyboard.

The travel space was shown in a combination of green and red shading on the background to indicate rerouting options for the active aircraft. The different bands of the travel space showed the increase in airspeed required to prevent a delay. A red zone indicated that moving a waypoint to that location would cause a conflict with another aircraft, whereas a green area indicated that a waypoint could be safely placed there without causing a conflict.

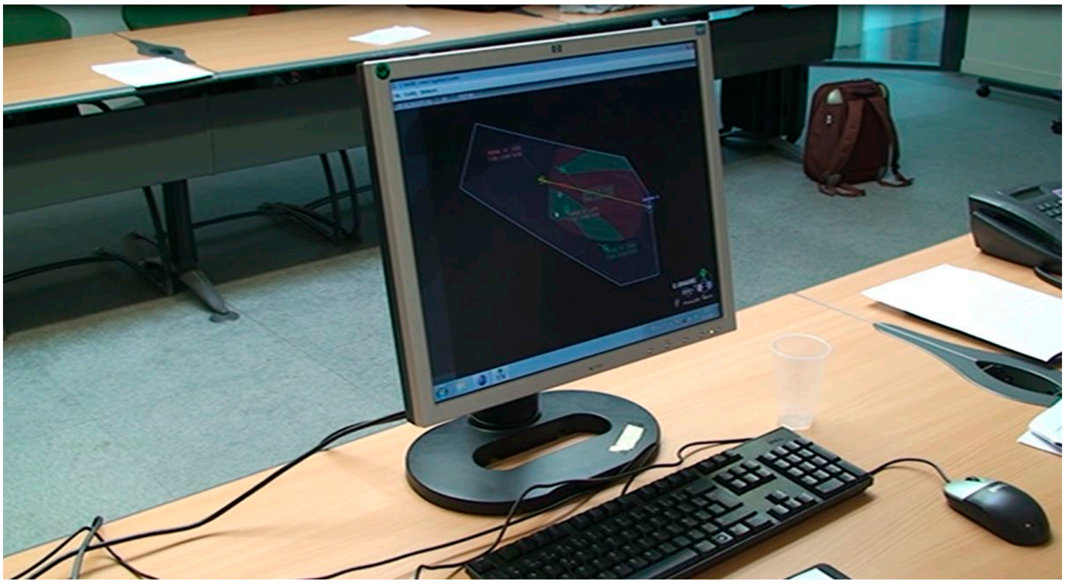


Figure 4. Experimental set-up

The time that a particular aircraft would leave the airspace could be manipulated using the mouse wheel. This was presented by a disk around the waypoint and a numerical indication of being too early or delayed in seconds. When moving the mouse wheel the delay time/upfront time was increased or decreased. If an aircraft was delayed the disk turned red. In figure 5 the implementation of the clock as a visualisation of delay is shown. The two left waypoints showed a blue section and a number indicating the number of seconds the aircraft was ahead of schedule. The third and fourth waypoints were in purple/red and indicate the number of seconds of delay. A waypoint turned red if it would be crossed with a delay of 30 seconds or more.

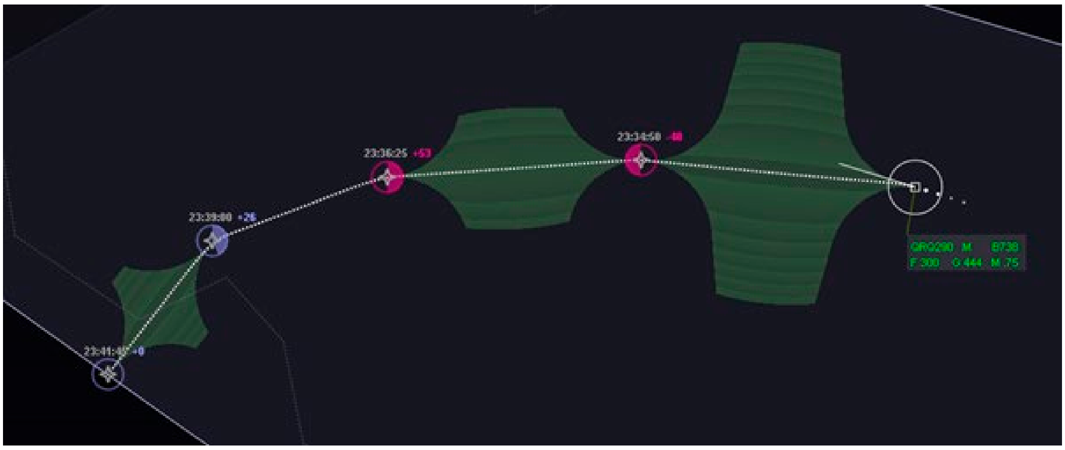


Figure 5. Waypoints with different delay times

Auto-generated solutions generated by means of RRTs were displayed as blue lines. Each suggestion was shown as a line on the screen and also in a list in the upper right corner. The suggestion in the list included information about which aircraft should be rerouted, the possible delay it would cause etc. The user could browse through different suggestions in the list and select the preferred option.

If perturbations occurred, the perturbed aircraft turned red and a list of perturbations appeared on the upper right corner of the interface. Perturbations in this list first appeared red on a dark background which inverted when the expected time of the perturbation remained less than two minutes. The information in this list also included the names and locations of the aircraft.

3.3 Procedure

The sessions were held with one participant at a time and three researchers; one operator to run the scenarios, one experiment leader to train and interview the participants and one observer to note spontaneous utterances and interview answers. Each session lasted approximately two hours. Participants first received a short introductory presentation, which included background on SESAR, perturbation management, the C-SHARE project and the outline of the experiments. Next, participants received an informed consent form and the initial questionnaire. After filling these out, each participant received instructions for executing JCS scenarios.

Participants walked through the first five scenarios for training. The last three scenarios were free-play, participants were allowed to use the user interface how they wanted without being guided by a script. However, scenario 6 emphasized manually placing waypoints, guided by the constraints and based on the travel space visualization to investigate the interaction without auto-generated visualizations. Scenario 7 emphasized the use of the auto-generated solutions to study the interaction without placing waypoints and scenario 8 gave controllers the possibility to use travel space information, using auto-generated solutions or the placement of waypoints in whichever way they preferred.

After each scenario participants were interviewed about their experiences using a structured interview format. This was followed by a short trust questionnaire based on SATI [14] and a situation awareness assessment.

The interface aspects concerned the concept of C-SHARE, the Travel Space Representation, and solving conflicts. Interview questions included intuitivity and usability, and trust.

3.4 Data Analysis

All spontaneous utterances and answers to the structured interviews were gathered by combining the notes from the sessions with listening back to the audio data. This information was supplemented with the answers to the questionnaires and assessments to sort them by topic and facilitate the analysis of results. The sorting was done individually by different researchers and results were compared. The researchers carried out the grouping in the same way and came to the same conclusions about which answers and utterances belonged to which category. A summary of the participants' opinions was made per topic for each scenario which resulted in a list including implications, suggestions, and other feedback on the interface reflecting the opinions of all users. For each opinion on a certain aspect of the interface it was noted if all, two or only one participant gave that particular feedback. A summary of this feedback is given in the next section.

4 Results

4.1 Feedback on the Concept

As all three participants were experienced ATCos and experienced researchers, they spontaneously made comments about the tested user interface in comparison to existing automation. In general, their opinions were more often positive than critical. Moreover, the critical comments for example focused on the fact that the concept as presented was too advanced to become reality with the current positioning technologies in aircrafts. General feedback given by the participants included:

- The system is easy to understand and gives good solutions for perturbations;
- It would be nice to be able to use such a system: the job of an ATCo will become different but not less interesting;
- The end responsibility stays with the ATCo; he is in control. This is something that is very important for systems like this;
- The C-SHARE approach lacks the presentation and solutions in altitude. It does not allow solving a conflict by climbing or descending aircraft; this is in practice the most common form of solving a perturbation. This ‘simplification’ of the reality for research and development purposes allowed for the design of new visualisations without the complexity of this extra dimension. Later on this extra dimension should be incorporated in the C-SHARE system;
- The level of complexity of the test scenarios is ok. However some of the assignments are not realistic.

The list above is a summary of feedback, both positive and critical, reflecting the general opinion of all participants on the concept and rationale, which were identified as most important for this research project.

4.2 Travel Space

In the training scenarios the concept of the Travel Space Representation was briefly explained. Nevertheless, the users immediately understood, also without explanation, the meaning of the green (possible rerouting zone) and red parts (impossible area). The lines in the presentation which corresponded to the required change in speed of the rerouting to keep the same time at the sector exit were not noted by most participants and the one that did see them did not understand the meaning. After explanation the participants responded that the lines were probably not necessary. Also without lines it is clear that the farther a new route or waypoint is placed from the current one, the higher the speed increase required.

4.2.1 Solving a Conflict

The C-SHARE user interface allowed for solving a conflict (resulting from perturbations) in two ways:

- by changing a trajectory through placing a waypoint supported by the Travel Space; or
- by requesting auto-generated solutions that could be accepted or denied.

4.2.2 Placing a New Waypoint

The participants mentioned that they liked the interaction method of placing alternative waypoints to reroute flights. The participants understood the 'disk' that was included in a waypoint that indicated the delay a certain rerouting activity would give. Nevertheless, they considered this information redundant, as the delays were indicated in seconds next to the waypoint as well.

4.2.3 Auto-generated Solutions

The auto-generated solutions were explained in one of the training scenarios. Nevertheless, at presentation of the participants understood that the blue lines indicated proposed conflict solutions as given by the system.

It was observed that the participants usually checked out different solutions before choosing one. Normally the automation would provide three possible solutions for a perturbation which appeared on the screen underneath each other upon request. However, when a participant again requested a solution the first one appeared in addition to the first three solutions. This was correctly identified as a software bug.

4.3 Other

Other more general feedback that was captured is listed next.

- One point of critique concerned the flight labels that were used in this version of the C-SHARE system. The flight labels included too much information according to all users. Also all ATCos mentioned that at several moments the labels were occluding the view on the radar picture;
- The interaction in this version includes commands that can only be given by keyboard; these commands should be changed to interactions with a pointing device;
- When perturbations occurred, they were also listed in the right corner of the screen; first in red on black and inverted as the perturbed aircrafts come closer to each other. When explained, the participants understood the meaning of the list. However, before we pointed it out to them the participants did not see the list themselves. Two of them mentioned that the list should be bigger because it did not draw enough attention.

4.3.1 Trust

As described in chapter 2, during the experiment, after each scenario the trust of participants in the automation was measured by a trust questionnaire (the SATI). In addition, after each scenario the participants were explicitly asked whether they trusted the visualisations and proposed solutions generated by the automation. Following from these answers and from analysis of overall and individual ratings on the SATI statements, the participants trusted the system. The overall ratings to the trust questionnaire resulted in a mean score for all participants on all questions of 6.7 out of 7 (where the score on negatively stated questions was reversed). The individual ratings on all SATI statements (concerning the accuracy, reliability of and confidence in the system) were all positive.

Some summarized example quotes of (dis)trust that were given during the different scenarios are mentioned below:

- “As an ATCo I have to trust the system I am using and I know they are always thoroughly tested, therefore I would trust the suggestions on perturbation management from this system as well”;
- “The solutions that the system proposes (blue lines) seem logical; therefore I trust this system”.

All users indicated repeatedly throughout the experiment that they had the feeling that the system was giving them solid, good solutions for perturbation management. The users also indicated that they trusted the solutions that the automation proposed completely but still sometimes checked what the system would have proposed after they accept the solution. When asked why they did this they responded in the line of: “Of course I trust the system, however in my daily work I also always monitor a solution for a perturbation after it is carried out.”.

5 Discussion

5.1 Summary of Results

The results show that the system has potential for aiding ATCos in their task of monitoring and de-conflicting 4D trajectories 25 years from now. Users indicated that they were positive about placing waypoints, using travel space information and using auto-proposed solutions to support their decision making process. They also stated that the system has good potential for future implementation after a thorough operational validation and verification procedure. The different development cycles that included Human Factors testing revealed elements of the system that need improvement in the final development phase.

The participants were positive about the concept, particularly because of the fact that the ATCo keeps the central role in the C-SHARE concept. The novel way of presentation through the travel space has an important limitation as in the tested version it lacked the dimension altitude. Nevertheless, when a conflict needs to be solved in the lateral dimensions it provides good insight in altering a route.

5.2 Suggestions for Future Development Cycles

Although the current research evaluated the C-SHARE functionality by means of seven scenarios, a more thorough test with more participants, scenarios and perturbations will provide more insight in the long-term user experience. This will also give better insight in how ATCos perceive the system when they are put under higher pressure. This feeling of pressure seemed to be absent in the current evaluation, partly because the participants were invited to give feedback and talking out loud during the scenarios. A next evaluation will therefore include more training scenarios and longer and more difficult test scenarios with more traffic and perturbations. Those complex scenarios will be part of the final experiments of this project (which runs until September 2013).

A potential limiting factor is the availability of trained ATCos for validating a system by comparing it to a baseline system in terms of performance and experience. More ATCos are needed than are commonly available for user tests. A solution for this will be sought in supplementing active ATCos with former ATCos (as we did partly in the current experiment), ATCo students and domain experts (researchers in this domain).

The following suggestions will be implemented in further developments:

- The request for system-proposed solutions and selection of a solution, in the tested version, required keyboard operation. In an improved version this will only require interaction with a pointing device following the overall interaction philosophy;
- Some information will be moved out of the label and the label will be smaller. The extra information will be available on request;
- The implementation of the option of descending an ascending aircraft in software is beyond the timeframe of the C-SHARE project. Nevertheless, design efforts will be put into possible visualisation of and interaction with this extra dimension to optimally support the ATCos work in a way that is in line with the current version of the C-SHARE system;
- The list showing perturbations gives indications for the preferred order to solve them. This list should be more salient. If perturbations become time crucial; they should be much more alarming than in the current implementation;
- Adding and moving a waypoint will be possible through a pointing device in the coming implementation of the automation. The participants indicated that this type of interaction would fit the work domain better.

C-SHARE started as a first exploration of perturbation management with 4D trajectory planning with a focus on a synergy between the control by the ATCo and suggestions from the automation. To prevent that the development and testing of automation concepts in several iterations would become too complex for the runtime of the project the possibilities for perturbation management were limited to horizontal solutions. Building upon C-SHARE's results, a succeeding project should include both horizontal and vertical perturbation management. As the basic implications and drawbacks of automation of perturbation management are already analysed and developed in the current project, a next project can focus on testing the concepts on a bigger scale including more realistic perturbations and both horizontal and vertical solutions management.

Acknowledgements

The authors acknowledge EUROCONTROL and the SESAR Joint Undertaking. The work was co-financed by EUROCONTROL on behalf of the SESAR Joint Undertaking in the context of SESAR Work Package E (project C-SHARE: Joint ATM Cognition through Shared Representations). This work reflects only the authors' views and EUROCONTROL is not liable for any use that may be made of the information contained herein. In addition, the controllers and other experts participating in the Human Factors test are acknowledged for their feedback and inspiration.

References

[1] L. Bainbridge, "Ironies of automation," *Automatica*, pp. 775-779, 1983.

[2] E. Hollnagel and D. D. Woods, Joint cognitive systems; Foundations of cognitive systems engineering, Boca Raton, Florida, USA: Taylor & Francis, 2005.

[3] E. Hollnagel and D. Woods, "Cognitive systems engineering: New wine in new bottles," International Journal of Human Computer Studies, vol. 51, no. 2, pp. 339-356, 1999.

[4] R. Klomp, R. M. v. Paassen, C. Borst, M. Mulder, T. Bos, P. v. Leeuwen and M. Mooij, "Joint automation cognition through a shared representation of 4D-trajectory management," in SESAR Innovation Days, Braunschweig, 2012.

[5] K. Vicente, "Coherence- and correspondence-driven work domains: Implications for systems design," Behaviour & Information Technology, vol. 9, no. 6, pp. 493-502, 1990.

[6] J. Kuffner and S. LaValle, "RRT connect: An efficient approach to single-query path planning," in IEEE International Conference on Robotics and Automation, San Francisco, California USA, 2000.

[7] C. J. Bastien, "Usability testing: a review of some methodological and technical aspects of the method," International Journal of Medical Informatics, vol. 79, pp. e18-e23, 2010.

[8] G. M. Donahue, S. Weinschenk and J. Nowicki, "Usability is Good Business," 1999. [Online]. Available: http://half-tide.net/UsabilityCost-BenefitPaper.pdf. [Accessed 04 06 2013].

[9] A. Woolrych, K. Hornbaek, E. Frokjaer and G. Cockton, "Ingredients and meals rather than recipes: A proposal for research that does not treat usability evaluation methods as indivisible wholes," International Journal of Human Computer Interaction, vol. 27, no. 10, pp. 940-970, 2011.

[10] J. Nielsen, "Heuristic Evaluation," in Usability Inspection Methods, New York, NY, John Wiley & Sons, 1990.

[11] W. Hwang and G. Salvendy, "Number of people required for usability evaluation: the 10±2 rule," Communications of the ACM, vol. 53, no. 5, pp. 130-133, 2010.

[12] E. L. C. Law and E. T. Hvannberg, "Analysis of combinatorial user effect in international usability tests," in CHI, Vienna, Austria, 2004.

[13] Nielsen Norman Group, "Jakob Nielsen's Alert Box," NN/g, 19 03 2000. [Online]. Available: http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/. [Accessed 05 06 2013].

[14] Eurocontrol, "Guidelines for Trust in Future ATM Systems: Measures," Eurocontrol, Brussels, Belgium, 2003.

WHAT IS NLR?

The NLR is a Dutch organisation that identifies, develops and applies high-tech knowledge in the aerospace sector. The NLR's activities are socially relevant, market-orientated, and conducted not-for-profit. In this, the NLR serves to bolster the government's innovative capabilities, while also promoting the innovative and competitive capacities of its partner companies.

The NLR, renowned for its leading expertise, professional approach and independent consultancy, is staffed by client-orientated personnel who are not only highly skilled and educated, but also continuously strive to develop and improve their competencies. The NLR moreover possesses an impressive array of high quality research facilities.

NLR *– Dedicated to innovation in aerospace*