E.02.08-D13-MUFASA-Final Project Report

Document information	
Project Title	**Mu**ltidimensional **F**ramework for **A**dvanced **S**ESAR **A**utomation
Project Number	E.02.08
Project Manager	Lockheed Martin UK IS&S Ltd
Deliverable Name	Final Project Report
Deliverable ID	D13 (D0.9 in contract)
Edition	01.00.00
Template Version	03.00.00
Task contributors	
Lockheed Martin UK IS&S Ltd, CHPR, TUD, IAA	

Abstract

The final report of the **Mu**ltidimensional **F**ramework for **A**dvanced **S**ESAR **A**utomation project provides a publishable summary of the results. In addition it lists all deliverables, dissemination activities, eligible costs, deviations, bills and lessons learned.

Authoring & Approval

Prepared by - *Authors of the document.*		
Name & Company	Position & Title	Date
[redacted] CHPR	[redacted]	02/09/2013
[redacted] LM UK IS&S Ltd		02/09/2013

Reviewed by - *Reviewers internal to the project.*		
Name & Company	Position & Title	Date
[redacted] CHPR	[redacted]	03/9/2013

Approved for submission to the SJU by - *Representatives of the company involved in the project.*		
Name & Company	Position & Title	Date
[redacted] LM UK IS&S Ltd	[redacted]	03/9/2013

Document History

Edition	Date	Status	Author	Justification
00.00.01	03/09/2013		[redacted]	New Document

Intellectual Property Rights (foreground)

This deliverable consists of SJU foreground.

Table of Contents

List of tables

Publishable Summary

Roughly 60 years ago, English mathematician Alan Turing famously posed the ultimate test for artificial intelligence: that its performance be indistinguishable from that of a human. If one could converse with an unseen agent, and mistake computer for human responses, then that computer could truly be said to "think." This notion has driven research into artificial intelligence for over half a century.

We are at a point in the evolution of automation that we routinely turn over to computers many of the "thinking" tasks previously performed only by humans. Our planes, trains and even automobiles rely on more, and more sophisticated, automation, than ever before. Despite various achievements, however, there remains some gap between the theory and practice of automation design. For instance, as of this date not a single computer has passed the Turing test. Nor have we realised in any meaningful way the highest levels of autonomous systems. As Sheridan noted, we still have no idea how to program computers to "take care of children, write symphonies or manage corporations…." [1]

The recently completed MUFASA (Multi-dimensional Framework for Advanced SESAR Automation) project started, in a sense, from the opposite perspective: What if we could build perfect automation, which behaved and solved complicated problems exactly as a human? Would the human accept its operation? Or might humans reject their very own solutions simply because these were offered by automation? In such a situation, anything less than perfect acceptance might be evidence of automation bias, a prejudice against using automation.

These questions are not just academic. MUFASA started from the assumption that future ATM will increasingly rely on automation that can assume control of the cognitive and strategic aspects of ATM. Successfully introducing such advanced new forms of automation might rely heavily on initial air traffic controller acceptance. We set out to explore whether controllers would accept automation that is, by definition, perfectly matched to the human strategy, since "automation" in our experimental design was (unbeknownst to the controller participants) in fact an unrecognisable replay of the controller's very own performance [2].

MUFASA laid out an initial predictive framework for automation usage and acceptance, and set out to explore how these would be impacted by the possibly interactive effects of three factors:

- Traffic complexity;
- Level of automation; and
- Strategic conformance.

The MUFASA project introduced the concept of *strategic conformance*, which we defined as follows:

> *The degree to which automation's behaviour and apparent underlying operations match those of the human.*

Prototype automation was built on the Solution Space Diagram (SSD) interface under development at the Technical University of Delft. Three simulations were conducted in total, and each was progressively more sophisticated in terms of interface design, automation capability, and experimental design. These three simulations are described a follows:

- Simulated Baseline Automation (SIMBA);
- Preliminary Update / Modified Baseline Automation (PUMBA);
- Nominal Advisory Level Automation (NALA).

Developmental trials were used to refine experimental, interface and automation design, and culminated in NALA data collection trials with sixteen en route air traffic controllers at IAA Shannon during January – February 2013 [2].

These NALA trials were carried out in two phases. The first (or "Prequel") study was used to extract controller resolution strategies under manual (non-automated) conditions. The second (or "Conformance") study involved presenting controllers unrecognisable replays of either their own solution (i.e., a *conformal solution*) or a colleague's different solution (i.e., *non-conformal*) to a series of pending aircraft separation conflicts.

Consistent with the project's research hypotheses, conformal resolution advisories were more accepted, led to higher controller agreements, and also reduced response time to accept advisories. Most notable was the following: Of 256 conformal solutions (i.e. replays of a given controllers' very own previous performance), 23.8% were rejected by controllers.

How is it that controllers would disagree with themselves nearly one quarter of the time? This effect may have been due to automation bias (in this case, a prejudice against using automation), or to individual variability over time. In any event, it underscores the need to consider such effects when designing advanced strategic decision aiding automation.

Results also included main and interaction effects of complexity and strategic conformance in terms of acceptance, agreement and response time. Complex traffic conditions were indeed associated with higher difficulty ratings, but also with decreased response time, and increased acceptance and agreement [3].

On the basis of these results, the project proposed a refined model of controller automation acceptance, built on an expanded distinction between *agreement* and *acceptance* of automation [3]. According to this model, the decision of whether to use automation on a given trial is driven by the (possibly interactive) considerations of context (e.g., what is my current workload?), as well as strategic conformance (does automation match my way of working?) and tendency (am I inclined to use automation?). Acceptance provides feedback that informs both a conformance check, and automation usage tendency. In effect, the decision of whether to use automation is one of utility. Assessment of utility is based on a consideration of the immediate demand of the task, the perceived match between human and machine strategy, and a general tendency to use the automation.

In the words of economist and sociologist Thorstein Veblen ***"…t**he outcome of any serious research can only be to make two questions grow where only one grew before…."* The MUFASA team certainly feels that it has opened two doors for each one it has closed. The team intends, going forward, to address some of (new and remaining) issues that arose from the project, including:

- Why did controllers disagree with their very own resolutions, roughly 25% of the time? Again, is this evidence of an inherent bias against using automation, or perhaps just individual inconsistency over time in strategies controllers bring to the task?

- Are operators / controllers more biased against machine- than human agents? If, for instance, controllers believed that suggested resolutions came from a trusted colleague as opposed to automation, would they be less likely to reject such resolutions?

- To the extent that conformal rejections were due to an inherent bias, is this in fact a bias against machine advice or against any external advice whatsoever? Would controllers reject a similar percentage of conformal advice if they believed it came from a colleague, as opposed to a machine?
- How are automation acceptance and agreement impacted by automation failures?
- Are novices more or less inclined to accept automation than are experts?

1 Introduction

1.1 Purpose of the document

The purpose of this document is to:

- Summarise the technical results and conclusions of the project (Publishable Summary);
- Provide a complete overview of all deliverables;
- Provide a complete overview of all dissemination activities (past and in progress). Where appropriate, provide feedback from presentations. Describe exploitation plans.
- Provide a complete overview of the billing status, eligible costs, planned and actual effort (incl. an explanation of the discrepancies).
- Analyse the lessons learnt at project level.

1.2 Intended readership

The intended audience for this report consists primarily of representatives of the policymaking and R&D communities, who might have a stake in future ATM design. Those involved directly with aeronautical research, HMI design and human factors might also welcome the topics covered in this report. The report is written from an engineering and experimental perspective. Those from this and related domains will probably be most conversant with the terminology, but every attempt has been made to keep the document readable and approachable, so as to invite a wider audience.

1.3 Inputs from other projects

The MUFASA project built on a broad body of theoretical and empirical research into ATM automation and HMI design. The vast majority of this research is publicly accessible from academia and research organisations. The test platform for human-in-the-loop simulations was built around the Technical University of Delft's prototype Solution Space Diagram (SSD) interface, which has been under development for some years. Finally, the MUFASA project is inspired by, although reuses no specific aspect of, EUROCONTROL's Controller Resolution Assistance Project (CORA), which examined the potential benefits of en route strategic decision aiding.

1.4 Glossary of terms

Term	Definition
ANSP	Air Navigation Service Provider
CD&R	Conflict Detection & Resolution
CORA	Controller Resolution Assistance
IAA	Irish Aviation Authority
MUFASA	MUltidimensional Framework for Advanced SESAR Automation
NALA	Nominal Advisory Level Automation

Term	Definition
PUMBA	Preliminary Update / Modified BAseline
SESAR	Single European Sky ATM Research Programme
SIMBA	Simulated Baseline Automation
SSD	Solution Space Diagram

2 Technical Project Deliverables

Management deliverables such as progress reports, gate report or this final report need not be included.

Number	Title	Short Description	Approval status
D1	Report on Initial 3D Framework	Literature review of theoretical and empirical state-of-the-art with respect to human performance aspects of ATM decision aiding / automation. Specified an initial an initial theoretical model to guide subsequent human-in-the-loop simulations.	Approved
D2.1a	Experimental Design, 2D Simulation	Specified prototype HMI and intended experimental design for simulations. Outlined methods for creating test scenarios and manipulating experimental factors, as well as the dependent measures and preliminary analysis protocols.	Approved
D2.2a	Report on 2D Simulation	Outlined background, setup and initial results from the first (SIMBA) simulation, which included both Baseline and LOA studies.	Approved
D2.2b	Report on 2D Simulation: SIMBA II, LOA Study	Reviewed background, setup and initial results from the SIMBA LOA study. Notice that primary aim of the study was to pilot test HMI, test scenarios, automation concepts, and data collection / logging / analysis procedures.	Approved
D3.1	Experimental design, 3D validation	Presented final experimental design (including methods, factors and analysis protocols) for the NALA simulations.	Approved
D3.2	NALA 3D Validation	Summarised initial findings of the NALA simulations, which were conducted in Jan-Feb 2013 with sixteen controllers at IAA Shannon.	Approved
D4	Revised Framework Report	Presented more detailed analysis of the NALA data, and places the results in the context of a revised theoretical model of automation acceptance and usage. Highlighting remaining areas for future research.	Submitted

D5	Dissemination Report	This document summarises the background and results of the project, listing the project's technical deliverables and summarises the formal dissemination activities. It also describes anticipated follow-on activities.	Approved

Table 1 - List of Project Deliverables

3 Dissemination Activities

3.1 Presentations/publications at ATM conferences/journals

3.1.1 SESAR Innovation Days (Toulouse, Nov 2011)

<u>Abstract</u>

The Multidimensional Framework for Advanced SESAR Automation (MUFASA) project is exploring issues concerning the acceptance and usage of advanced decision aiding automation. Through a series of planned human-in-the-loop simulations, it aims to examine the interactive effects of automation level, air traffic complexity and strategic conformance (i.e. the fit between human and machine strategies) on automation usage. This paper outlines the theoretical background, experimental design and methodological approach underlying this effort.

<u>Citation</u>

Westin, C., Hilburn, B. & Borst, C. (2011). Mismatches between Automation and Human Strategies: An Investigation into Future Air Traffic Management (ATM) Decision Aiding. In Proceedings of the 2011 SESAR Innovation Days, Toulouse, 27-29 Nov, 2011.

3.1.2 SESAR Innovation Days (Braunschweig, Nov 2012)

<u>Abstract:</u>

The Multidimensional Framework for Advanced SESAR Automation (MUFASA) project is exploring issues concerning the acceptance and usage of advanced decision aiding automation. Through a series of human-in-the-loop simulations, it ultimately aims to examine the interactive effects of automation level, air traffic complexity, and strategic conformance (i.e. the fit between human and machine strategies) on automation usage and acceptance. This paper, however, only presents the design and results of an exploratory experiment that aimed to investigate conflict detection and resolution (CD&R) automation usage among professional air traffic controllers and novices (students) utilising a novel decision-support tool: the Solution Space Diagram (SSD). This preliminary study featured a manual air traffic control task with a fixed level of automation under high and low traffic complexity. The results indicate that students, in contrast to the controllers, reacted more immediately and promptly to conflict warnings. With no separation losses, students outperformed controllers in keeping aircraft safely separated. Controllers, on the other hand, had multiple separation losses. Observations and debriefing of controllers revealed a general scepticism towards the SSD display, its accuracy and usefulness. This allows us to speculate that controllers, compared to students, had less trust in the SSD as a CD&R tool, and rather used their own judgment in conflict management.

Citation

Borst, C., Westin, C. & Hilburn, B. (2012). An Investigation into the Use of Novel Conflict Detection and Resolution Automation in Air Traffic Management. In Proceedings of the 2012 SESAR Innovation Days, Braunschweig, 26-28 Nov, 2011.

3.1.3 International Symposium for Aviation Psychology (Ohio, USA, May 2013)

Abstract

Future air traffic management will have to rely on more, and more sophisticated, automation to accommodate predicted air traffic. However, studies across various domains have shown that user acceptance of automation decreases when the authority of decision-making automation increases. As a result, low user acceptance could lead to disuse of an automated tool and threaten potential safety and performance benefits. Through a series of human-in-the-loop simulations, the work described in this paper examined the interacting effects of air traffic complexity and strategic conformance, i.e., the fit between human and machine strategies, on automation acceptance in a conflict detection and resolution task. An experiment with 16 professional air traffic controllers showed that strategic conformance is a potentially important construct. That is, conformal resolution advisories were more accepted, led to higher controller agreement, and also reduced response time to proposed advisories.

Citation

Westin, C., Borst, C. & Hilburn, B. (2013). Mismatches between automation and human strategies: An investigation into future air traffic management (ATM) decision aiding. In Proceedings of the 2013 International Symposium for Aviation Psychology (ISAP), Ohio State University, 6-9 May 2013.

3.2 Presentations/publications at other conferences/journals

3.2.1 ATACCS (Barcelona, May 2011)

Abstract

The *Multidimensional Framework for Advanced SESAR Automation* (MUFASA) project recently kicked off as a SESAR WP-E innovative research effort. Collaborators include CHPR BV (NL), the Irish Aviation Authority IAA (EI), Lockheed Martin UK IS&S Ltd (GB), and the Technical University of Delft (NL). Over the coming 27 months, MUFASA aims to develop a framework for designing future levels of automation, based primarily on human-in-the-loop simulation. Building on the team's research into innovative space- and time-based CD&R displays, it aims to experimentally manipulate automation level, traffic complexity and, most notably, "heuristic conformity" or the fit between human and automation solutions to medium term planning and separation conflicts. This will help refine a framework addressing potential interactions and tradeoffs between these dimensions in terms of both human performance (workload, situation awareness) and system performance. Notably, it also permits a way to quantify and empirically define "automation bias." Results aim to extend current automation design principles, but also inform the design of advanced ATM automation for the mid-and far terms. Some of the notable characteristics of the MUFASA project are the following:

- It systematically evaluates algorithmic and heuristic approaches to CD&R automation.
- It is perhaps one of the first times that research has tried to empirically define and quantify user trust and acceptance in automation: as willingness to accept automated advice that fits with one's own preferred way of working (because that advice is in fact an unrecognisable replay of one's own solution).
- It brings together a consortium that offers renowned research capabilities and facilities, ANSP operational expertise and partnership with industry.
- It combines ongoing and complementary work into both optimised "machine centred" and heuristic "human centred" approaches to ATM display and automation technology.
- The project schedule calls for academic laboratory simulations to scale up for replication with the consortium's ANSP partner.
- The project would extend current state-of-the-art with respect to automation design principles.

Citation

Hilburn, B., Westin, C., Kearney, P., Mulder, M. & Day, C. (2011). Would I Accept a Computer that Thinks Like Me? An Overview of the MUFASA Research Project. Presented at the ATACSS 2011 Conference, Barcelona, 26-27 May.

3.3 Demonstrations

Various informal demonstrations were made of the developmental SSD interface and simulated automated capability. These were made primarily to the aviation industry and academic research communities.

3.4 Exploitation plans

in the near term, the MUFASA team intends to further investigate aspects of the project's results. In particular, we intend to address two main research questions:

1. Is there an automation bias? Participants rejected their very own previous solutions, when they believed these were the product of automation, roughly 25% of the time. It is not clear whether this effect is due to an automation bias, or simply individual variability over time. That is, it could be that controllers are simply inconsistent over time in the strategies they use to resolve conflicts. The team intends to address this issue by a simple extension of the current design. It is hoped that a follow-on round of data collection can be arranged to explore the additional experimental combination of manual versus “automated “subsequent sessions.

2. Is the bias against automation or against advice? It could be that controllers were biased, but not against automation per se rather against receiving any sort of external advice. The question then becomes whether participants would reject a similar percentage of conformal advice if this were to come from a colleague, as opposed to a machine. We could address this question using the same basic experimental design, and modifying instructions so that participants believed their replays represented the advice of a colleague.

More generally, we intend to publish and present the project's past and ongoing results, and hope to reach an industrial audience outside of aviation and ATM. Candidates include representatives of the robotics, process control, command and control, and industrial engineering communities.

4 Total Eligible Costs

This section is based on the Project Costs Breakdown Forms of the eligible costs incurred by project participants.

Date	Deliverables on Bill	Contribution for Effort	Contribution for Other Costs (specify)	Status
29/11/11	D0.1, D1, D2.1a, D0.2	€ 73,029.74	*	P
19/04/12	D2.1b, D2.2a, D0.3	€ 109,005.87	*	P
21/09/12	D2.2b, D0.5, D0.6	€ 58,213.59	*	P
08/03/13	D3.1(includes D3.1a & D3.1b), D3.2, D0.7	€ 203,142.65	*	P
16/10/13	D4, D5, D0.8, D0.9	€75,433.50	*	Raised
GRAND TOTAL		€518,825.35		

*Costs of travel included in daily rate

Table 2 Overview of Billing

Company	Planned man-days	Actual man-days	Total Cost	Total Contribution	Reason for Deviation
LM	27	27	€19,740	€9,870	
CHPR	440	440	€363,999	€272,991	
TUD	475	475	€222,001	€222,001	
IAA	47	39.75	€27,927	€13,963	Unforeseen unavailability of staff.

GRAND TOTAL	989	981.75	€633,667	€518,825	Final invoice adjusted accordingly.

Note: No price deviation; Actual effort greater than planned, due to unforeseen challenges in simulation and experimental design, and expanded data collection.

Table 3 Overview of Effort and Costs per project participant

5 Project Lessons Learnt

What worked well?
Streamlined administrative procedures allowed us to focus most of our resources on technical work.
Network mechanism facilitated beneficial interaction between projects
Technical mentors helped refine our thinking, and improved quality control.
What should be improved?
It is hoped that administrative procedures could be streamlined even further in the future

Table 4 - Project Lessons Learnt

6 References

1. Sheridan, T.B. (2002). Humans and automation: System design and research issues. New York: Wiley.
2. MUFASA (2013a). E.02.08 – MUFASA – D3.2 - NALA 3D Validation.
3. MUFASA (2013b). E.02.08 – MUFASA – D4 – Revised Framework Report.