EXPLORATORY RESEARCH

sesar JOINT UNDERTAKING

2nd Workshop report

Deliverable ID:	**D7.3**
Dissemination Level:	**PU**
Project Acronym:	**MAHALO**
Grant:	**892970**
Call:	**H2020-SESAR-2019-2**
Topic:	**SESAR-ER4-01-2019**
Consortium Coordinator:	**Deep Blue**
Edition date:	**27 June 2022**
Edition:	**00.02.00**
Template Edition:	**02.00.05**

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & Approval

Authors of the document

Name / Beneficiary	Position / Title	Date
Supathida Boonsong (LFV)	Project member	2022-05-11

Reviewers internal to the project

Name / Beneficiary	Position / Title	Date
Matteo Cocchioni (DBL)	Project contributor	2022-05-26
Martin Christiansson (LFV)	Project member	2022-05-26
Brian Hilburn (CHPR)	WP2 leader	2022-05-26
Carl Westin (LiU)	WP6 leader	2022-05-26
Clark Borst (TU Delft)	WP4 leader	2022-05-26

Reviewers external to the project

Name / Beneficiary	Position / Title	Date

Approved for submission to the SJU By - Representatives of all beneficiaries involved in the project

Name / Beneficiary	Position / Title	Date
Stefano Bonelli (DBL)	Project Coordinator	2022-05-31
Stefano Bonelli (DBL)	Project Coordinator	2022-06-27

Rejected By - Representatives of beneficiaries involved in the project

Name and/or Beneficiary	Position / Title	Date

Document History

Edition	Date	Status	Name / Beneficiary	Justification
00.00.01	2022-05-11	ToC	Supathida Boonsong	ToC
00.00.02	2022-05-17	Draft	Supathida Boonsong	Initial draft of the deliverable

00.00.03	2022-05-26	Internal review	Matteo Cocchioni, Martin Christiansson, Brian Hilburn, Carl Westin, Clark Borst	Deliverable improvement
00.00.04	2022-05-31	Internal review	Matteo Cocchioni	Deliverable improvement
00.01.00	2022-05-31	Final release	Stefano Bonelli	Deliverable improved for submission
00.01.01	2022-06-21	Revision	Supathida Boonsong	Reviewer's comments
00.01.02	2022-06-27	Internal review	Matteo Cocchioni	Deliverable reviewed
00.02.00	2022-06-27	Final release	Stefano Bonelli	Deliverable improved for submission

MAHALO

MODERN ATM VIA HUMAN / AUTOMATION LEARNING OPTIMISATION

This deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 892970 under European Union's Horizon 2020 research and innovation programme.

Abstract

The MAHALO (Modern ATM via Human/Automation Learning Optimisation) project has developed a Machine Learning (ML) modelling system, i.e. Personalized model [Supervised Learning (SL)] and Optimal model [Reinforcement Learning (RL)], that is coupled to an enhanced Ecological User Interface (E-UI). The project experimentally evaluated the models using human-in-the-loop (HITL) simulations performed by active Italian (provided by ANACNA) and Swedish (provided by Luftfartsverket, LFV) Air Traffic Controllers (ATCO) during winter 2021 and spring 2022 respectively. More details of the overall experimental design for the simulations are provided in *D6.1 Experiment design document* [MAHALO Project, 2021]. In mid-May, the MAHALO project presented the ML models (i.e. SL and RL) and preliminary results of HITL simulations to the Advisory Board (AB) members and relevant Air Traffic Management (ATM) stakeholders. The ATCOs who have participated in the simulations were also invited to attend the workshop.

Table of Contents

Abstract 4
1 Introduction 6
1.1 Purpose and scope of the document 6
1.2 Structure of the document 6
1.3 List of Acronyms 6
2 Workshop framework 8
2.1 Objectives 8
2.2 Agenda 9
2.3 Participants 10
2.4 Workshop materials 10
3 Results and Conclusions 13
3.1 Workshop results 13
3.2 Conclusions 15
4 References 18
Appendix A Workshop recordings 19
Appendix B Workshop minutes 20
MAHALO workshop 4
Workshop objectives 4
Organisation of the participants 4
Agenda 4
Minutes 5

List of Tables

Table 1 List of Acronyms 7
Table 2 Workshop agenda 9
Table 3 List of organisations of the participants 10
Table 4 Assessment of workshop objectives 16

List of Figures

Figure 1 Miro board used to facilitate the discussions and collect feedback during group works 12

1 Introduction

1.1 Purpose and scope of the document

This deliverable *D7.3 Workshop report* presents the organization of the 2nd workshop with the MAHALO Advisory board members and relevant ATM stakeholders. This event is part of *WP7 Dissemination*, task *T7.2 Workshop organisation* involving two workshop occasions to be conducted by the consortium within the project life cycle. Note that the 1st workshop was carried out in October 2021 as an online event. The workshop report has been submitted to SJU as a public deliverable *D7.2 Workshop report* [MAHALO Project, 2021].

Hereby, this document:

- details the 2nd workshop organization covering workshop agenda and structure, list of participants, workshop activities, links to public access documents and related materials that were presented, used and collected during the workshop;
- presents workshop results, conclusions and the workshop minutes.

1.2 Structure of the document

This document consists of three chapters and contains the following details:

- Chapter 1 provides the purpose and scope, as well as the structure of this deliverable.
- Chapter 2 describes the framework of the 2nd workshop organization including workshop agenda, objectives and organization of the participants as well as materials used during the workshop.
- Chapter 3 summarizes the results and conclusions of the conducted workshop.

1.3 List of Acronyms

Term	Definition
AB	Advisory Board
AI	Artificial Intelligence
ANACNA	Associazione Nazionale Assistenti e Controllori Navigazione Aerea
ANSP	Air Navigation Service Provider
ATCO	Air Traffic Controller
ATM	Air Traffic Management
CD&R	Conflict Detection and Resolution

CEST	Central European Summer Time
CHPR	Center for Human Performance Research
DQFD	Deep Q-Learning from Demonstrations
ENAV	Ente Nazionale Assistenza al Volo
EUI	Ecological User Interface
EUROCAE	The European Organisation for Civil Aviation Equipment
LFV	Luftfartsverket
LiU	Linköping University
ML	Machine Learning
RL	Reinforcement Learning
SA	Situational Awareness
SJU	SESAR Joint Undertaking
SL	Supervised Learning
TU Delft	Delft University of Technology
WP	Work Package

Table 1 List of Acronyms

2 Workshop framework

The MAHALO consortium individually invited participants (i.e. the AB members, stakeholders from ATM) by sending out a formal invitation letter in a timely manner. The invitation provided information on the project's ambitions, expected contributions from the workshop attendees and the details of the workshop itself, i.e. date, time and venue.

On Monday the 16th May 2022, between 09.00 – 13.00 CEST, the MAHALO project organized a virtual workshop using Google Meet platform. The purposes of this event were to exhibit the MAHALO ML prototypes i.e. SL model and RL model, and also to present preliminary results of HITL simulations conducted with active Italian ATCOs (winter 2021) and Swedish ATCOs (spring 2022). The consortium aimed at gathering external experts' views on how to build a future ML system to assure that the system would suit/meet particular needs of the end users and that the MAHALO research is in alignment with future visions of ATM.

The workshop was scheduled to coincide with the completion of the HITL simulations which were part of *WP6 Simulation* and was the second workshop of the project. (Note that the 1st MAHALO workshop was carried out in late October 2021 as an online event. The workshop report has been submitted to SJU as a public deliverable *D7.2 Workshop report* [MAHALO Project, 2021].) This 2nd workshop was attended by a total of twelve external guests comprising of experts from ATM domain, e.g. Air Navigation Service Providers (ANSP), standardisation body, and the ATCOs who had participated in the HITL simulations. In addition, a total of eleven members of the MAHALO consortium also attended. However, it's worth mentioning here that although the workshop invitations had been circulated approximately six weeks prior to the workshop and several reminders were sent out as follow-ups, neither academics nor ML experts accepted/responded to the invitations. The workshop, therefore, was not attended by representatives from these fields of expertise. Despite this circumstance and the fact that it was a virtual event, the MAHALO consortium ensured maximum effectiveness and high involvement from every participant throughout the workshop.

2.1 Objectives

To ensure a successful workshop and achieve the expected goals, the project outlined the following workshop objectives:

- Presentation of the MAHALO process from start to finish;
- Presentation of a proof of concept for an AI based Conflict Detection and Resolution (CD&R) tool, able to provide solutions which may be based on real air traffic controllers' strategies (conformance) or on AI-based optimisation;
- Presentation of alternative ways in which the AI can "Speak" with the controllers, providing information that enable them to understand why a specific solution has been proposed (transparency);
- Exhibition of the MAHALO ML prototypes i.e. personalized model and optimal model;

- Sharing MAHALO definitions for conformance, transparency, personalized and optimized resolutions with ATM stakeholders;
- Sharing different approaches on ML strategies;
- Presentation of preliminary results of HITL simulations carried out in winter 2021 and spring 2022;
- Assure that the MAHALO research is in alignment with future visions of ATM.

2.2 Agenda

The workshop was successfully carried out in accordance with the structure presented in Table 2. The first session was welcoming and roundtable where the MAHALO consortium welcomed the participants. During this session, each participant had an opportunity to briefly introduce themselves by providing i.e. name, organization and main area of expertise. An introductory session was then followed where a brief overview of the workshop (i.e. agenda, objectives and structure) was provided. The workshop then proceeded with three presentation sessions where the project presented its ML prototypes (i.e. SL and RL model) and preliminary results of the HITL simulations. At the end of each presentation, the participants were encouraged to ask questions, share their perspectives and provide feedback.

Time	Activity
09:00	**Opening**
09:00 – 09:20	Welcome and Roundtable
09:20 – 09:25	Workshop overview
09:25 – 09:35	The MAHALO process: from Start to Finish
09:35 – 09:50	Session 1 – MAHALO prototype: SL (Personalized model)
09:50 – 10:20	Discussion and feedback
10:20 – 10:30	*Coffee Break*
10:30 – 10:45	Session 2 – MAHALO prototype: RL (Optimal model)
10:45 – 11:15	Discussion and feedback
11:15 – 11:40	Session 3 – Preliminary results of human-in-the-loop simulations
11:40 – 11:50	*Coffee Break*
11:50 – 12:50	Discussion and feedback
12:50 – 13:00	Next steps and Wrap-up
13:00	**Closing**

Table 2 Workshop agenda

2.3 Participants

As provided at the beginning of this chapter, a total of twelve external guests attended the workshop. The majority of them were ATCOs who had participated in the HITL simulations in Italy and Sweden. The others were experts from different organizations representing different fields of expertise and professionals including standardisation body and European body. Table 3 contains the organisations of the participants.

Organisation	Type
ANACNA	ATCO association
ENAV	ANSP
EUROCAE	Standardisation body
EUROCONTROL	European body
LFV	ANSP

Table 3 List of organisations of the participants

2.4 Workshop materials

A number of materials such as PowerPoint presentations have been produced for the use during the workshop. In addition, Miro, an online collaborative whiteboarding platform, has also been created for the purpose of collecting live and instant feedback from the participants during certain session(s).

2.4.1 PowerPoint presentations

At the workshop, one PowerPoint presentation was created to be a focal document. However, there were some other presentations that were individually created by the speakers to be used during his/her respective session.

2.4.1.1 Presentation “General”

This presentation was used as a main presentation and contained general workshop contents such as objectives, agenda, list of participants and session information.

2.4.1.2 Presentation “The MAHALO Process: from Start to Finish”

This document was presented during the “The MAHALO Process: from Start to Finish” session. It provided information on the project goals and a big picture of the MAHALO process by detailing activities that have successfully been performed and accomplished by the consortium in the past 24 months period. It also provided the definitions of conformance and transparency, and partly touched upon how the HITL simulations were arranged.

2.4.1.3 Presentation “The MAHALO prototype: SL (Personalized model)”

This presentation contained high-level explanations on how the MAHALO supervised learning model was built and the pipeline for creating personal and group advisories which were simulated in the ATC

simulator “SectorX”. Additionally, the document also listed limitations and disadvantages of the SL model.

2.4.1.4 Presentation “The MAHALO prototype: RL (Optimal model)”

This document provided the audiences with information on how the MAHALO reinforcement learning model was built. It also provided high-level explanations of the RL methods which were Q-Learning and Deep Q-Learning from Demonstrations (DQFD). The main findings (drawbacks) related to the RL model were given at the end of the presentation.

2.4.1.5 Presentation “Preliminary results of human-in-the-loop simulations”

This presentation contained (1) experimental design detailing simulation protocol (i.e. training and main experiment phase), ATCO participants, validation scenarios, simulator, etc., (2) preliminary results of the HITL simulations, and (3) the findings.

Following is a link to all presentations and materials that were presented and used during the workshop. Note that the link is preferably to be opened with Google Chrome.

https://drive.google.com/drive/folders/1VkTDBf2kzgK-n1o9JP7wh1OQZYwkjbWN?usp=sharing

2.4.2 Miro whiteboard

An online interactive Miro whiteboard (Figure 1) was created to support and use during the session “The MAHALO prototype: SL (Personalized model)” with the purpose of collecting and capturing spontaneous feedback. It was used as a tool for the participants to share their perspectives on the question “**Which human CD&R factors do you consider most relevant for AI to adapt to when advising on how to solve conflicts?**” posted by the presenter. The participants could select predefined factors or add/reflect their own answers directly on the Miro board using blank post-its.

In order for a discussion to be as effective as possible, the participants were divided into 3 working groups. Additionally, one (or more) of the MAHALO consortium members was assigned to each group to facilitate and moderate the discussion. During these group-working sessions, the participants discussed and share their perspectives as well as rationale behind their choice of answers.

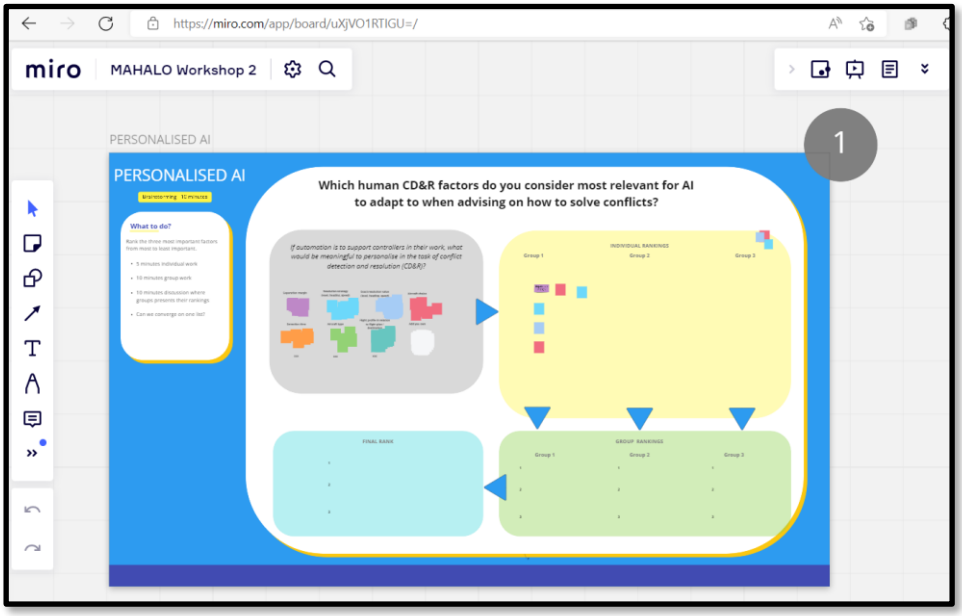


Figure 1 Miro board used to facilitate the discussions and collect feedback during group works

Co-funded by the European Union

3 Results and Conclusions

3.1 Workshop results

The high-level outcomes of the workshop could be summarized as follows:

- With regards simulation realism, ATCOs stated that the validation scenarios were quite simplified and not at all complex, when compared to real ATC environment. One of the reasons was due to the number of conflicts i.e. there was only one conflict per scenario. ATCOs also added that they could recognize traffic patterns only after some consecutive runs. In addition, conflict patterns were quite obvious to them, namely the conflicts usually occurred in the middle of the sector, which was in octagon shape for every scenario. They also noticed differences in conflict angles between runs. Another reason for scenarios being non-complex was that each conflict involved only two aircraft and not more (algorithm did not consider third aircraft involved in the conflicts). ATCOs explained that in real environment when there are more flight movements within the sector, one could expect more numbers of (potential) conflicts which sometimes may involve two or more aircraft leading to a more complex situation.

- In terms of the simulator, SectorX was equipped with necessary menus required for the execution of ATC clearances, e.g. heading/altitude clearance and direct-to clearance. The menus were described as similar to those that existed in current ATC system. Therefore, it only took a small amount of time for the ATCOs to get familiar with the simulator and its available menus. ATCOs, however, complained that some crucial information for building up situational awareness (SA) and for planning for conflict situations such as coordination with adjacent sectors, destination airports (presented on aircraft flight plans), waypoints along the intended routes etc., were missing. This contributed to unrealistic traffic situations/scenarios compared to real operational environment.

- ATCOs do not consider themselves consistent when it comes to solving conflicts. Many factors could affect their decision-making, and thus actions. This includes aircraft flight plan, the proximity to a destination (which may favour a descent rather than a climb), airline, aircraft type (e.g. medium vs heavy wake turbulence category), and weather factors.

- Many layers of information (i.e. automation transparency) provided together with automation advisories are neither necessarily useful nor matter much for ATCOs. Instead, it is more important and essential to have a workable solution proposed to ATCOs at a reasonable time.

- Participants generally agreed that a workable solution is more important than an optimal solution. One participant noted that the definition of an optimal solution is subjective. Another participant argued that analysing if the solution is optimal or not is a secondary level of analysis. Participants stated that, when presented with an advisory, they do not actively explore solution alternatives to determine if the proposed solution is optimal or not. Rather, they check if the solution does not cause any adverse effects, such as secondary conflicts or

unnecessary large deviations. However, what a controller deems workable is highly subjective and the workshop results did not clarify what factors are considered when making that judgement. Somewhat contradictory, participants also stated that the MAHALO advisory system and the proposed solutions worked quite well.

- Personalized solution advisories seem to increase acceptance from ATCOs. The project, in future works, shall be able to confirm this provided that there is more collected data.
- The project, in future works, shall reclassify the term “acceptance” in a more fine-grained way, and not in a simple binary (accept versus reject) way. The project realizes in its analysis that there are interim options.
- In future works, a better algorithm/ML model will be required in order for the agent training to perform better and to improve efficiency.
- Although the workshop was not able to receive feedback on external ML perspectives because of the absence of ML experts, the MAHALO will, on other dissemination occasions, e.g. ATM-related conferences or workshops, reach out to relevant AI/ML projects to exchange knowledge and create synergy. Additionally, it is worth mentioning that a couple of members of the MAHALO team are sitting on Advisory Board of other AI/ML projects and representing MAHALO.
- Using a Miro board, participants discussed the question “Which human CD&R factors do you consider most relevant for AI to adapt to when advising on how to solve conflicts?” Participants were in the first step asked to reflect on the question individually and select and rank post-it notes with pre-defined factors in order of importance. They were asked to rank the three most important factors from most to least important. These factors were partly derived from the previous Workshop 1 and partly derived by the MAHALO team. The factors were:
    - Separation margin
    - Resolution strategy
    - Exact resolution value (level, heading, speed)
    - Aircraft choice
    - Detection time
    - Aircraft type
    - Flight profile in relation to flight plan

  Participants could also add their own factors. In the second step, participants were instructed to discuss the factors that they had ranked in their respective groups (there were three groups). Each group then presented their finding to all participants. Group 1 ranked *Resolution strategy* as most important (i.e. heading, level, speed, or direct-to). *Aircraft choice* was ranked second, and *Flight profile* ranked third. Group 2 could not agree on a ranking of individual factors but instead tied the ranks for *Detection time* and Resolution strategy. These were considered more important than the tied ranks for *Exact value* and *Flight profile*. Group 3 considered *Separation margin* as most important, followed by *Detection time* and *Flight profile*.

- Participants were asked what type of ML CD&R support system they would like to have with regards to conformance and transparency. For this question, the below figure was shown, but with empty cells. Participants responded that the "perfect box" would be "high conformance and low transparency". In the MAHALO team we have described this box as "Confusing automation". Participants stated that the perfect solution to a conflict should be the same every time and that this can be based on the most frequently used solution made by a group of controllers. Others agreed and pointed out that transparency is not needed if the system is consistent.

		TRANSPARENCY	
		Low	High
CONFORMANCE	Low	**Stupid automation:** *"It's doing a strange thing, and I don't understand why…"*	**Peculiar automation:** *"It's doing a strange thing, but I understand why…"*
	High	**Confusing automation:** *"It's doing the right thing, but I don't understand why…"*	**Perfect automation:** *"It's doing the right thing, and I understand why…"*

  -

3.2 Conclusions

3.2.1 Assessment of workshop objectives

Objective	Assessment	Rationale
Presentation of the MAHALO process from start to finish	Achieved	The consortium successfully presented the entire MAHALO process from the initial literature review of ML advances to field simulations, to data analysis and report.
Presentation of a proof of concept for an AI based Conflict Detection and Resolution (CD&R) tool, able to provide solutions which may be based on real air traffic controllers' strategies (conformance) or on AI-based optimisation	Achieved	Using the simulator SectorX, the project demonstrated MAHALO scenarios implemented at the HITL simulations. MAHALO also provided an insight into how the conformal and optimal solutions were achieved.
Presentation of alternative ways in which the AI can "Speak" with the controllers, providing information that enable them to understand why a specific solution has been proposed (transparency)	Achieved	SectorX coupled with the Ecological User Interface (EUI) consisting of several visual elements and providing an insight into the inner workings of the ML agent, was demonstrated to the workshop attendees.

Exhibition of the MAHALO ML prototypes i.e. personalized model and optimal model	Achieved	The consortium successfully presented the processes and methods of building the ML models.
Sharing MAHALO definitions for conformance, transparency, personalized and optimized resolutions with ATM stakeholders	Achieved	MAHALO definitions of the mentioned terms were provided and, to some extent, discussed during the workshop. Feedback were collected for future research.
Sharing different approaches on ML strategies	Achieved	The consortium shared the MAHALO ML approaches and provided the rationales.
Presentation of preliminary results of HITL simulations carried out in winter 2021 and spring 2022	Achieved	The project successfully presented an initial analysis (i.e. preliminary results) of the data collected during the field simulations. The results were discussed with great interest from the attendees.
Assure that the MAHALO research is in alignment with future visions of ATM	Achieved	MAHALO performed a state-of-the-art review of ML advances, and recent theoretical and empirical research into the areas of human performance and ML.

Table 4 Assessment of workshop objectives

It is important to mention that within the workshop, the term “optimal” referred to what the controller think is optimal, not that the system has advised on a solution that has been optimized according to an algorithm of some sorts. Note that the statement made during the workshop does not say that a workable ("good enough") solution is necessarily different from an optimal solution. What the controller does not do is to check if the solution is the most optimal one, which would require comparing several different solutions. In other words, the controller does not care if the solution is optimal or not. The question is how to determine if a solution is workable or not to a controller. That is a subject of future research.

A finding in MAHALO is that the extent to which a solution matches the individual's preference for solving conflicts (as defined by our method for deriving personal models) appears to impact their acceptance and agreement of that advisory: the advisories that were closer to matching their preferences received higher acceptance and agreement ratings. This was the finding of the fine-grained analysis. A key parameter here was separation margin. In the simulations, very few advisories were rejected. This indicates that all advisories were considered workable. During the workshop, participants also stated that even if they nudge the heading a bit, they still considered it as accepting the advisory. However, we also saw that controllers intervened more with the advisory when the advisory differed from their preferred way of solving the conflict. What is workable to one controller may not be workable to another controllers. As such we fall back on the subjective differences between controllers, which leads us back to the strength of personalized advisories.

Overall, as a result of the interactive workshop and very interesting discussions, the MAHALO consortium was able to gather valuable feedback and perspectives from the ATM experts (especially air traffic controllers) who made much of a contribution and supports the MAHALO Solution, together with quantitative results for the time being not yet analysed, but now described in D6.2.

The presentations of the MAHALO machine learning models and the preliminary results of HITL simulations were also well received and greatly appreciated by all attendees. In addition, the workshop itself has produced essential inputs and contributed to the project's dissemination activity. In conclusion, the workshop objectives were successfully accomplished.

4 References

[1] MAHALO Project (2021). *Experimental design document*. MAHALO D6.1

[2] MAHALO Project (2021). 1st *Workshop report*. MAHALO D7.2

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Appendix A Workshop recordings

The recordings were divided into three separate files and can be viewed at a link provided below. Note that the link is preferably to be opened with Google Chrome.

https://drive.google.com/drive/folders/1r1YuBYlSCvWcNu9X-677UV023jVR3sX_?usp=sharing

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Appendix B Workshop minutes

EUROPEAN PARTNERSHIP

Co-funded by the European Union

EXPLORATORY RESEARCH

sesar JOINT UNDERTAKING

MAHALO

2nd Workshop Minutes

Project Acronym:	**MAHALO**
Grant:	**892970**
Call:	**H2020-SESAR-2019-2**
Topic:	**SESAR-ER4-01-2019**
Consortium Coordinator:	**Deep Blue**
Edition date:	**20 May 2022**
Edition:	**00.01.00**
Template Edition:	**02.00.05**

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Disclaimer

This project has received funding from the European Union's Horizon 2020 SESAR Research and Innovation action under Grant agreement No 892970.

The statements made herein do not necessarily have the consent or agreement of the MAHALO consortium. These represent the opinions and findings of the author(s). The European Union (EU) is not responsible for any use that may be made of the information they contain.

MAHALO

MODERN ATM VIA HUMAN / AUTOMATION LEARNING OPTIMISATION

Workshop Minutes

16th May 2022, 09.00 – 13.00 CEST

This workshop minutes is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 892970 under European Union's Horizon 2020 research and innovation programme.

MAHALO workshop

Workshop objectives

The purpose of the workshop was to demonstrate the whole MAHALO process and exhibit the works accomplished, especially the MAHALO machine learning (ML) prototypes, i.e. Supervised Learning (SL) model and Reinforcement Learning (RL) model, as well as to present preliminary results of the human-in-the-loop (HITL) simulations carried out with active Italian and Swedish air traffic controllers (ATCO) in winter 2021 and spring 2022 respectively.

Organisation of the participants

Organisation	Type
Associazione Nazionale Assistenti e Controllori Navigazione Aerea (ANACNA)	ATCO association
Ente Nazionale Assistenza al Volo (ENAV)	Air Navigation Service Provider (ANSP)
The European Organisation for Civil Aviation Equipment (EUROCAE)	Standardisation body
EUROCONTROL	European body
Luftfartsverket (LFV)	Air Navigation Service Provider

Agenda

Time	Activity
09:00	**Opening**
09:00 – 09:20	Welcome and Roundtable
09:20 – 09:25	Workshop overview
09:25 – 09:35	The MAHALO process: from Start to Finish
09:35 – 09:50	Session 1 – MAHALO prototype: SL (Personalized model)
09:50 – 10:20	Discussion and feedback
10:20 – 10:30	*Coffee Break*
10:30 – 10:45	Session 2 – MAHALO prototype: RL (Optimal model)
10:45 – 11:15	Discussion and feedback
11:15 – 11:40	Session 3 – Preliminary results of human-in-the-loop simulations
11:40 – 11:50	*Coffee Break*
11:50 – 12:50	Discussion and feedback
12:50 – 13:00	Next steps and Wrap-up
13:00	**Closing**

Minutes

1) Welcome by Project Coordinator

Presenter: Deep Blue (DBL)

DBL warmly greeted and welcomed the workshop attendees. A roundtable of the MAHALO consortium and the participants was then followed. Every workshop attendee had an opportunity to briefly introduce themselves by providing name, organization and area of expertise.

2) Workshop overview

Presenter: LFV

LFV went through workshop objectives and structure as well as agenda.

3) The MAHALO process: from Start to Finish

Presenter: Center for Human Performance Research (CHPR)

CHPR provided information about the project goals and a big picture of the whole MAHALO process by detailing activities that have successfully been performed and accomplished by the consortium in the past 24 months period. CHPR also provided high-level definitions of conformance and transparency, and partly touched upon how the HITL simulations were organized.

4) The MAHALO prototype: Supervised Learning (Personalized model)

Presenter: Linköping University (LiU)

LiU presented high-level explanations on how the SL model was built and explained the pipeline for creating personal and group advisories, which were simulated in the Air Traffic Control simulator "SectorX". Additionally, LiU also listed limitations and disadvantages of the SL model.

Questions / Comments:

EUROCONTROL: Asked if MAHALO merged the collected data (from both Italy and Sweden) like we had said we would do at the 1st workshop. Note that the 1st workshop was carried out in October 2021. It'd be interesting to understand how group model was performing because the working environments and methods of the two populations were different.

LiU: Yes, we did merge the data. We also noticed a lot of variations and differences between the two populations.

ANACNA: Scenarios were simple in the simulations. The algorithm sometimes didn't properly consider third aircraft.

LiU: There was already a lot of information on how controllers solved conflicts.

LiU: Asked if controllers are internally consistent in their strategies?

ENAV: I'm not even consistent within myself.

LiU: I acknowledged that as well.

ANACNA: I agree with ENAV. We're not consistent between each other. It changes from one to another. I myself remembered three conflicts that I already saw conflicts and had solutions. I agree with the fact that the free route environment is in valuable fashion.

LiU: Asked if controllers, as a group, are consistent in their strategies?

EUROCONTROL: There can be differences between the controller groups based on their working environments, e.g. there's a different geometry in Terminal Manoeuvring Area (TMA), so there're some limitations. I think SL has disadvantages in Machine Learning.

LFV: That's the case for TMA. I have worked in Maastricht and Abu Dhabi. In Maastricht, they worked with levels. They had a lot more ascending traffic. Same with Abu Dhabi. But I work differently in Sweden than I did in Abu Dhabi. In the MAHALO simulations, I found it easier to use heading solutions, to just turn the aircraft because I was uncertain of the simulation airspace scale. And it was easier to turn than to try to infer climb performance. In real life, it's quite often that there are 3 – 5 aircraft involving in a conflict you need to take into account.

LFV: Different environments, e.g. free route airspace, could be a limitation.

Miro session

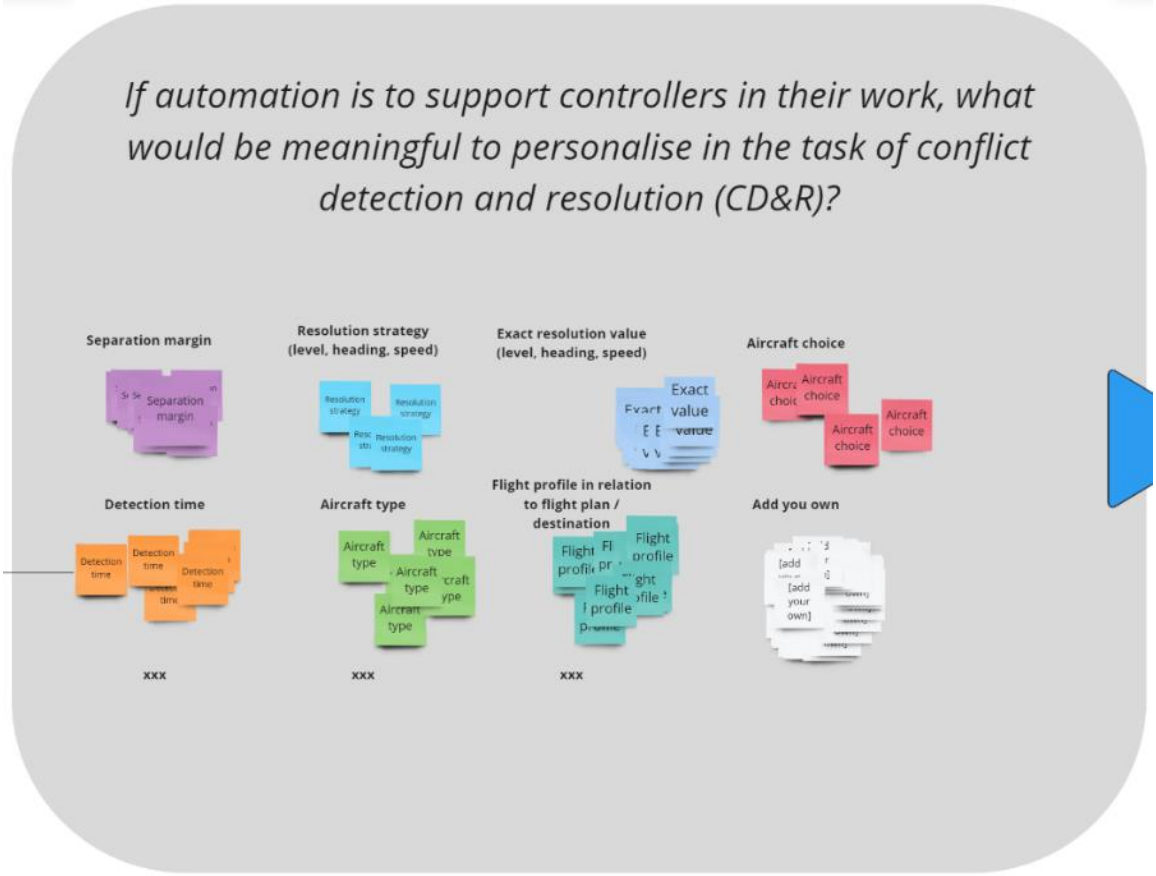
After the completion of the Questions / Comments session, the participants were divided into three working groups and invited to join Miro session/room according to their respective group. One (or more) of the MAHALO consortium members was assigned to each group to facilitate and moderate the discussions.

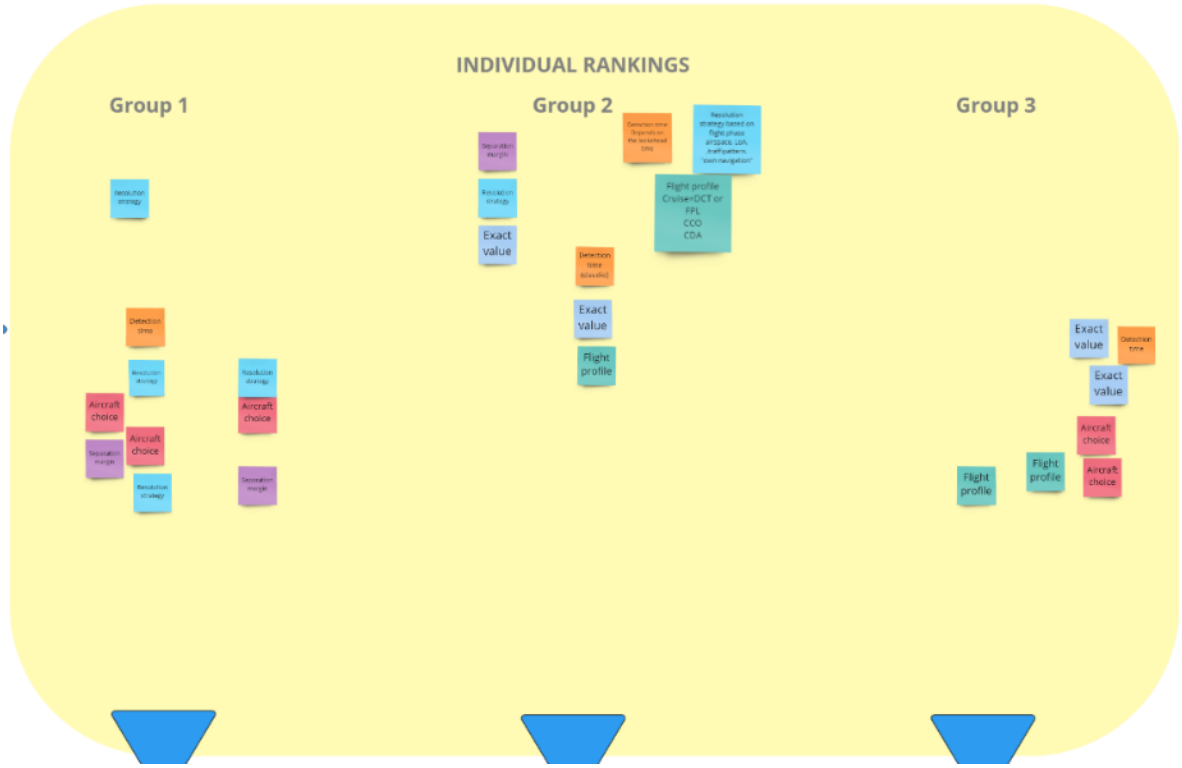
Using the online interactive tool "Miro", LiU posted the following question:

"Which human Conflict Detection & Resolution factors do you consider most relevant for AI to adapt to when advising on how to solve conflicts?"

The participants were instructed to: (1) individually consider their own preferences by selecting predefined factors or add their own answers directly on the Miro board using blank post-its, (2) share their perspectives and rationale behind their choice of answers with group members, (3) discuss within the group and come up with a group ranking of the CD&R factors.

Following are screenshots of the Miro board generated during the group work.





Co-funded by the European Union

5) The MAHALO prototype: Reinforcement Learning (Optimal model)

Presenter: Delft University of Technology (TU Delft)

TU Delft gave the audiences information on how the RL model was built. TU Delft also provided high-level explanations of the RL methods which were Q-Learning and Deep Q-Learning from Demonstrations (DQFD). The main findings (drawbacks) related to the RL model were also highlighted.

Questions / Comments:

TU Delft: Asked what determines optimality of a solution?

ENAV: The optimal is in the mind of the ATCOs. It'd be very difficult to determine what the optimal solution is for a conflict. For one conflict, there might be a lot of solutions that human can determine. This might be difficult for computer. ATCOs have rules to apply. But I think providing a solution 15 minutes before the conflicts is good. The system should provide that.

TU Delft: Asked if it's possible in some ways to build a system that offers an optimal?

ENAV: The system is good enough that shows conflicts that are timed in between Medium-Term Conflict Detection (MTCD) and Short Term Conflict Alert (STCA).

TU Delft: Asked about the reward function design.

ANACNA: For optimality, I don't care too much because the ATCOs have that. I think optimality is the second level of analysis. As long as the system provides a solution and it's valid, that is good enough. As long as the system doesn't create conflicts. Maybe optimal solution is optimal with respect to the time variable. As long as the system provides the solution in good and correct time, that'll likely be the optimal one. By the way, I don't quite understand the reward function. Please explain a bit more.

TU Delft: We have different view on this, but agree with reasonable time for providing solutions.

TU Delft: Explained that the reward system is based on a few factors such as the smallest number of actions, Closet Point of Approach (CPA) intrusion and distance to destination. As an example, we take into account the flight direction e.g. waypoint (WP). The longer it takes to get to the WP, the

lesser reward. There're also safety factors to consider. We put the buffer so that the aircraft stays outside the grey area.

EUROCONTROL: If you are trying to make the advisory system behave like the human to make ATCOs accept the solutions, then that might not be optimal e.g. human solutions could lead to more fuel burning or extra track miles to be flown. You might be limiting yourself in the end. But I agree that as long as the solution is good enough, ATCOs might accept it.

TU Delft: SL will never be better than humans but RL can.

CHPR: If you try to find solutions to match ATCOs, you will not get better behaviours. Humans will not improve.

ANACNA: Asked If we introduce some additional parameters or increase the number of factors and complexity of the agent, e.g. able to implement altitude solution (which was not doable in the experiments), would it change the amount of training?

TU Delft: Yes, every time you add another dimension, it adds 10 times the data requirements, and there was therefore a limit on what we could do. If you expose the system to more amount of dimensions, it will expose very quickly. We need to come up with better algorithm. Better training and improve efficiency.

TU Delft: Asked with the limited dimensionality of RL applications, is the determined optimal solution realistic enough?

ENAV: The system worked quite well.

ANACNA: Agree. If the project can make another solution, it'll be good.

TU Delft: We could use the altitude change as a solution, it's still doable. Like adding one extra dimension. But it'll take a lot of time to train the agent.

TU Delft: For the current agent, it takes a few days to train. To add another dimension, it might take a week or less with high performance computer.

6) Preliminary results of human-in-the-loop simulations

Presenter: LiU

LiU explained experimental design detailing simulation protocol (i.e. training and main experiment), participants, validation scenarios, simulator, etc., and then presented preliminary results of HITL simulations, and the findings.

Questions / Comments:

LiU: Asked the participants about their viewpoints on the classification of the interaction Accept, Nudge, adjust, change and reject?

ENAV: It depends on the amounts of the heading you change, e.g. if you need to give 40 degrees turn to aircraft, it's better to change altitude instead.

LiU: Asked if this action is considered accepting the solution?

ENAV: If it's just a slight change of the heading, it's still considered accepting because that's what ATCOs do in real life. If I'm not sure that the solution will get enough separation, I'd modify it (e.g. give greater heading).

LFV: Commented on the reject option in the simulations i.e. if I accepted the advisory but later was not allowed to do the follow up or adjust. But if I chose the reject option, I could do the follow-up. In some scenarios, that solution timing was sometimes off that the advisory came quite early (i.e. conflicts occurred very early into the scenarios), that's when I tended to accept the advisories because I had no time to assess the situation. I thought that was really scary. I was pacified. I mean it's scary when you only sit and let the machine does the job for you. I would rather have a conflict alert, not a conflict resolution system.

TU Delft: Think of the system like this is the supporting tool that warns you of a conflict that you've missed or not detected it before.

LFV: The system should just support, assist and alert you with conflicts. But then you get too adapted to that and it gives you a perfect solution every time, you would stop doing your jobs and trust the system. To me that is a bit scary.

LiU: Explained the timing of the advisory to LFV's comment, that the timing has derived from the data collected during the pre-test/training phase.

LFV: I was quite competitive so I tried to solve the conflicts before the advisories popped up.

ENAV: I had the same approach because I wanted to prove that human was better than the machine. But when a scenario started with a conflict situation that was going on, I did not have situational awareness at all. But I realized that I must search for conflicts before the advisories came up.

LiU: We need to look at acceptance in a more fined-grained way and not binary (accept or reject only). We see that the transparency can lead to like or dislike the solutions.

CHPR: In MAHALO, we had assumed that more transparency would lead to more acceptance, but we were kind of wrong.

ENAV: It's good to have some information, not a lot. But it's good to have a solution presented.

ANACNA: It's good to have a software that offers a lot of options (just like a wardrobe) but you can only select one option at a time. ATCOs should be able to use information or make decision on e.g. how, which one to use and select the option that they like. But basically, when a company decides to buy a software with a lot of tools, people will use it but will select different tools inside it depending on traffic situation and environment.

LiU: Assuming that we would have only one system, which one of this matrixes would be best?

		TRANSPARENCY	
		Low	High
CONFORMANCE	Low	**Stupid automation:** *"It's doing a strange thing, and I don't understand why…"*	**Peculiar automation:** *"It's doing a strange thing, but I understand why…"*
	High	**Confusing automation:** *"It's doing the right thing, but I don't understand why…"*	**Perfect automation:** *"It's doing the right thing, and I understand why…"*

LFV: What is conformance?

ENAV: What is transparency?

LiU: Conformance is when the system is conformal to your personal strategies. For example, when Artificial Intelligence (AI) acts like the human. Transparency is when the system explains why it does things, like AI that is understandable to the human. It can relate to how it's built or how it weighs different parameters.

TU Delft: It's similar to Netflix suggestions that are based on what you've watched. There're different levels of transparency.

ENAV: My perfect box is "High conformance – Low transparency". The perfect solution is the same every time, so I care about the solution, not transparency. It's important to have a good solution and that if a group of controllers makes the same solution then it is a good solution.

ANACNA: Agree. If you're working with a system that you rely on with the tools you want, you're more familiar / adapted to it. Then you don't need high transparency. You already work well with the system. So if I have a system I can manage as I want and it helps me work better, I will work better with just the system. I don't need transparency because this will come with the job/me anyway.