INDUSTRIAL RESEARCH

D4.1.100-PJ.10-W2-96 ASR-TRL6 TVALR

Deliverable ID:	**D4.1.100**
Dissemination Level:	**PU**
Project Acronym:	**PROSA**
Grant:	**874464**
Call:	**H2020-SESAR-2019-1**
Topic:	**Separation Management and Controller Tools**
Consortium Coordinator:	**DFS**
Edition Date:	**30 April 2023**
Edition:	**01.00.00**
Template Edition:	**02.00.05**

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & Approval

Authors of the document

Beneficiary	Date
LDO	16/09/2022
ENAIRE/CRIDA	30/09/2022
DLR	07/02/2023
SINTEF	15/12/2022

Reviewers internal to the project

Beneficiary	Date
LDO	21/02/2023
ENAIRE/CRIDA	06/02/2023
DLR	07/02/2023
SINTEF	27/02/2023
INDRA	06/02/2023
ANS CR	27/02/2023

Reviewers external to the project

Beneficiary	Date

Approved for submission to the S3JU By - Representatives of all beneficiaries involved in the project

Beneficiary	Date
DLR	23/02/2023
ENAIRE/CRIDA	28/02/2023
SINTEF	27/02/2023
INDRA	28/02/2023
ANS CR	27/02/2023
ACG	28/02/2023
INTEGRA	28/02/2023
CROCONTROL	28/02/2023

Rejected By - Representatives of beneficiaries involved in the project

Beneficiary	Date

Document History

Edition	Date	Status	Beneficiary	Justification
00.00.01	29/08/2022	Creation	LDO	Content/Structure
00.00.01	30/09/2022	First Draft	LDO, ENAIRE	Content
00.00.02	12/11/2022	Draft	DLR	Questionnaire results for exercise 002d in appendix B
00.00.03	05/12/2022	Draft	ENAIRE/CRIDA	Appendix C EXE-003
00.00.04	15/12/2022	Draft	SINTEF	Appendix D EXE-005
00.00.05	23/12/2022	Draft	LDO	Appendix A EXE-001
00.00.06	02/02/2023	Draft	DLR	Update section 4&5; Appendix B EXE-002
00.00.07	07/02/2023	Draft	DLR	Review of DLR inputs
00.00.08	06/02/2023	Draft	INDRA	Review and inputs
00.00.09	06/02/2023	Draft	ENAIRE/CRIDA	Review and inputs
00.00.10	06/02/2023	Draft	SINTEF	Update Appendix D EXE-005
00.00.11	14/02/2023	Draft	LDO	Update Appendix A EXE-001
00.00.12	21/02/2023	Draft	LDO	Implementation of all comments and contributions from the partners after review
00.00.13	23.02/2023	Final	DLR	Review performed, no changes
00.00.14	27/02/2023	Final	SINTEF	Changes in Appendix D and section 4.1 for EXE-05
00.01.00	28/02/2023	Final	LDO	Clean version ready for submission
00.01.01	29/03/2023	Draft	LDO	Implementation of SJU comments
00.01.02	30/03/2023	Draft	SINTEF	Additional table in Appendix D about EXE-05 results.
00.02.00	13/04/2023	Final	LDO	Final version ready for resubmission.

00.03.00	20/04/2023	Final	LDO	Final version ready for resubmission including SJU comments.
01.00.00	30/04/2023	Final	SJU	Approved by the SJU

PJ.10-W2 PROSA

SEPARATION MANAGEMENT AND CONTROLLER TOOL

This Technical Validation Report is part of a project that has received funding from the SESAR3 Joint Undertaking under grant agreement No 874464 under European Union's Horizon 2020 research and innovation programme.

Abstract

This Technical Validation Report presents the results for the TRL6 validation activities fulfilled for solution PJ.10-W2-96 ASR and covers the following exercises:

- EXE-PJ.10-96-ASR-TRL6-01 - A Real time simulation addressing the use of ASR to input the datalink commands in the relevant masks on the CWP reducing the ATCOs workload.EXE-PJ.10-96-ASR-TRL6-02 – A Real time human-in-the-loop simulation evaluating the determination of operational KPI values.
- EXE-PJ.10-96-ASR-TRL6-03 – A Real time simulation addressing the use of ASR to highlight callsigns and provide commands history to improve ATCO's situational awareness.
- EXE-PJ.10-96-ASR-TRL6-05 – A Real time human-in-the-loop simulation exploring how ASR in combination with traditional navigation and AG techniques in 2D and 3D visualizations of the airspace can enhance ATCO's understanding of the dynamic changes in the air space configuration.

Validation exercises were planned to address stakeholders' needs and assess the KPAs of Safety, Cost Efficiency, Fuel Efficiency and Human Performance. Deviations from the TVALP **Error! Reference source not found.** are also provided where found, along with conclusions and recommendations for future activities on the same areas.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Table of Contents

List of Tables

List of Figures

1 Executive summary

PJ.10-W2-96 ASR starts taking into account the work performed by S2020 PJ16-04 Wave 1 project, as well as the MALORCA project, executed in the context of Exploratory Research. PJ.10-W2-Sol.96 ASR's starting maturity level is TRL4 and it targets to reach TRL6 maturity at the end of Wave 2 activities.

This validation report offers an account of the TRL6 validation activities conducted for PJ.10-W2-96 ASR "Automatic speech recognition.

PJ.10-W2-96 ASR covers the following Operational Improvements and Technical Enablers:

- **POI-0055-SDM** "**Improving controller productivity by Automatic Speech Recognition (ASR) at the ER/APP CWP/HMI"** ATCOs will be supported by introducing innovative human machine interaction such as Automatic Speech Recognition that can be enhanced by the use of Machine Learning. The goal is to automatically support certain tasks of the ATCO, which are not done or done manually in today's systems / CWPs.

- **ER APP ATC 180 Controller productivity enhancements by Automatic Speech Recognition at the ER/APP CWP/HMI.** Automatic speech recognition (ASR) supported by AI/ML algorithms, which enables the recognition and translation of spoken language (e.g. ATCO commands) into the system reducing their workload and improving safety (e.g. reduce head-down times of the controller).

The current document (Technical Validation Report) gives an account of the following exercises:

- EXE-PJ.10-96-ASR-TRL6-01- A Real time simulation addressing the use of ASR to input the values of both datalink commands and RT commands in the relevant masks on the CWP reducing the ATCO's workload.
- EXE-PJ.10-96-ASR-TRL6-02 – A Real time simulation evaluating the determination of operational KPI values.
- EXE-PJ.10-96-ASR-TRL6-03 – A Real time simulation addressing the use of ASR to highlight callsigns and provide commands history to improve ATCO's situational awareness.
- EXE-PJ.10-96-ASR-TRL6-05 – A Real time simulation exploring how ASR in combination with traditional navigation and AG techniques in 2D and 3D visualizations of the airspace can enhance ATCO's understanding of the dynamic changes in the air space configuration.

Depending upon the different exercises, validation scenarios addressed different TMA/En-Route environments with medium complexity. The phase of flight of interest was the execution phase. The simulations and technical test exercises, through their validation objectives, addressed the stakeholders' needs, via an appraisal of the KPAs Safety, Technology Cost Efficiency, ATCOs' productivity, Fuel Efficiency, TMA Capacity and Human Performance. Additional activities to complement and further support the development and validation of the Operational Improvements were conducted within across the board performance assessments, most notably the safety, security, and human performance assessments.

This version of TVALR includes a detailed description of:

- EXE-PJ.10-96-ASR-TRL6-01
- EXE-PJ.10-96-ASR-TRL6-02

- EXE-PJ.10-96-ASR-TRL6-03
- EXE-PJ.10-96-ASR-TRL6-05

along with the results acquired from the validations.

The Solution has validated the operational improvements by means of the implementation of the ASR which was integrated in ATC systems. The use cases taken into account demonstrated how the developed enabler is able to reduce costs and workload ensuring the same safety level.

2 Introduction

2.1 Purpose of the document

This Technical Validation Report gives the results of the TRL6 maturity Technological Validation activities for PJ.10-W2-Sol.96 ASR. It describes how stakeholders' needs and system requirements were validated.

2.2 Intended readership

The intended audience of this document are those who are interested in how the partners involved in SESAR Solution PJ.10-W2-96 ASR validated the improvements to the management of the En-Route and TMA operations included in that solution and how those improvements may help the overall efficiency of the Air Traffic Management system. This would include, but is not limited to, other members of SESAR2020 PJ.10-W2, PJ.05-W2-97 ASR, members of PJ19 and the representatives of the S3JU. Additional readership may include public interest regarding R&D in human performance, factors, interaction with machines and computers as well as foreseeable ATM concepts focusing on novel interaction modes.

2.3 Background

PJ.10-W2-96 ASR starts taking into account the work performed by S2020 PJ16-04 Wave 1 project, as well as the MALORCA project, executed in the context of Exploratory Research. PJ.10-W2-Sol.96 ASR's starting maturity level is TRL4 and it targets to reach TRL6 maturity at the end of Wave 2 activities.

Automatic Speech Recognition matured up to TRL-4 in PJ.16-04 Wave 1 and enabled the definition of rules for transformation of a sequence of ATC words into ATC concepts (by means of the so called "ontology") and exploratory usage. PJ.16-04 concluded that speech recognition engines needed adaptation to achieve acceptable results.

PJ.10-W2-96 ASR aims to improve the speech recognition engines adaptation and use real data in an operational environment. Starting from the consolidated ontology developed in PJ.16-04, PJ.10-W2-96 ASR tries to improve the recognition engine using context data.

Furthermore the Solution aims to integrate the speech recognition engine in an operational En-Route and TMA environment exploring how the use of the recognition engine can improve and facilitate the routinely tasks of ATCOs evaluating the benefits of such a technology to reduce the ATCO's workload by highlighting the Callsigns of flights upon ATCOs and pilot utterances, prefilling command masks also for datalink messages and investigating how ASR may be used to enable faster and more predictable navigation in 3D visualizations of the airspace sectorisation when using dynamic airspace configuration (DAC).

2.4 Structure of the document

The structure of this document is based on the SESAR template for the Technical Validation Report (TVALR), and of course on the Technical Validation Plan (TVALP) and it is organized as follows:

- Chapter 1: Executive Summary: a brief summary of the key information elements contained in the TVALR document
- Chapter 2: Introduction (this chapter). Introduces the present document
- Chapter 3: Context of the Technical Validation. An explanation of the work done and to be done by each member of PJ.10-W2-96
- Chapter 4: SESAR Technological Solution PJ.10-W2-96 Validation Results. Sets the boundaries of the validation exercises and what limits are applicable to what and on which conditions
- Chapter 5: Conclusions and recommendations. A summary of conclusions at activity level for TRL6, along with recommendations for future improvements
- Chapter 6: References and Applicable Documents. A list of all references and applicable documents during TVALR editing, connected with novel human interaction modes
- Appendixes A-D: For validation exercise EXE-001 first (Appendix A), then for all remaining exercises following suit, detailed reports of each exercise are given, following the structure shown below, which indicates sections of each Appendix:
    - Summary
    - Description and scope
    - Validation objectives and success criteria
    - Validation scenarios
    - Scenarios
    - Assumptions
    - Deviations from planned activities
    - Validation results
        - Summary
        - Results per validation objective
        - Unexpected results/behaviours
        - Confidence in obtained results
    - Conclusions

2.5 Glossary of terms

Term	Definition	Source of the definition
Automatic Speech Recognition	An Automatic Speech Recognition (ASR) system gets an audio signal as input and transforms it into a sequence of words, i.e. "speech-to-text" following the recognition process. The sequence of words is transcribed into a sequence of ATC concepts ("text-to-concepts") using an ontology. For example: The word sequence "Lufthansa two alpha altitude four thousand feet on QNH one zero one four reduce one eight zero knots or less turn left heading two six zero" is transcribed into "DLH2A ALTITUDE 4000 ft, DLH2A INFORMATION QNH 1014, DLH2A REDUCE 180 OR_LESS, DLH2A	PJ.16-04

	HEADING 260 LEFT". The resulting concepts can be used for further applications such as visualization on an HMI.	
Command Recognition Rate	The number of controller commands which are correctly recognized by ASR and are not rejected before divided by number of total given commands; in other words: the percentage of given commands correctly shown on the controllers' HMI.	PJ.16-04
Command (Recognition) Error Rate	The number of controller commands which are wrongly recognized by ASR and which are not rejected divided by number of total given commands; in other words: the percentage of given commands wrongly shown on the controllers' HMI.	PJ.16-04
Command Hypotheses Predictor	Components needed for Assistant Based Speech Recognition which predicts a set of possible commands.	PJ.16-04
Command Prediction Error Rate	The number of controller commands which are not predicted by the Command Hypotheses Predictor divided by number of total given commands.	PJ.16-04
Command (Recognition) Rejection Rate	The number of recognized controller commands which are correctly or wrongly rejected (plus number of given controller commands which are not recognized at all) divided by number of total given commands.	PJ.16-04
Nuisance alert	Alerts that fail to provide useful information and can create their own human factors problems, because the ATCO receiving the alert must devote attention to deciding if the alert is valid and whether action is necessary.	
Functional Block	A logical and cohesive grouping of automated Functions in a Technical System	EATMA Guidance Material [1]

Table 1: Glossary of terms

2.6 Acronyms and Terminology

Term	Definition
ABSR	Assistant Based Speech Recognition

AI	Artificial Intelligence
ANSP	Air Navigation Service Provider
ATCO	Air Traffic Controller
APP	Approach
ASR	Automatic Speech Recognition
ASRU	Automatic Speech Recognition and Understanding
ASW	Air Situation Window
ATC	Air Traffic Control
ATM	Air Traffic Management
CHMI	Controller HMI
CR	Change Request
CV	Clearance Verification
CWP	Controller Working Position
EATMA	European Air Traffic Management Architecture
E-ATMS	European Air Traffic Management System
EN	Enabler
ER	En-Route
FB	Functional Block
FRD	Functional Requirements Document
HMI	Human Machine Interface
HP	Human Performance
HPAP	Human Performance Assessment Plan
ID	Identifier
IRS	Interface Requirements Specification
kHZ	Kilohertz
NFR	Non- Functional Requirements
NSV	NATO Systems View

POI	Performance Operational Improvement
PMP	Project Management Plan
SeAP	Security Assessment Plan
SESAR	Single European Sky ATM Research Programme
SJU	SESAR Joint Undertaking (Agency of the European Commission)
SPR-INTEROP/OSED	Safety and Performance Requirements – Interoperability Requirements / Operational Service and Environment Definition
SUT	System Under Test
TLA	Target Location Assistance
TRL	Technology Readiness Level
TS/IRS	Technical Specification/Interface Requirements Specification
TVALP	Technological Validation Plan
TVALR	Technological Validation Report
VALS	Validation Strategy
VCS	Voice Communication System
VP	Validation Plan
VR	Validation Report
V&V	Validation & Verification

Table 2: Acronyms and terminology

3 Context of the Technological Validation

3.1 SESAR Technological Solution 96 ASR: a summary

ASR in ATM domain matured in Wave 1 up to TRL-4 and enabled the definition of rules for transformation of a sequence of ATC words into ATC concepts (so called ontology). This ontology will be developed further and maybe prepared for standardization in the future.

In Wave 2 ASR will operate also on real life data and the objective is to reach TRL-6. As most input comes from the ATCO-pilot spoken dialog, ASR is the appropriate technology to reduce ATCO's workload by directly filling the command masks and radar labels using the spoken commands instead of manually inputting them into the system. This requires integration of artificial intelligence (AI) and machine learning algorithms. Highlighting of targets, user-friendly and intuitive operation is expected to increase controller productivity.

The solution operates in a medium complexity TMA and En-Route environment and ATCOs are the end users. HMI prototypes will be developed in order to present the results of the speech recognition (highlighting of targets, user-friendly and intuitive operation etc.) in the optimal way to the controllers.

ASR in the ATC environment can still be enhanced. The solution aims to reach a command recognition rate better than 85%, which was aimed in Wave1, which is quite ambitious. Therefore, comparison algorithms will be developed avoiding that erroneous recognitions are missed by the controller and used by subsequent systems. This requires integration of artificial intelligence (AI) and machine learning algorithms for the intelligent data provision to the controllers. In addition, the solution will also investigate how ASR may be used to enable faster and more predictable navigation in 3D visualizations of the air space sectorisation when using dynamic air space configuration (DAC). One of the challenges that will be addressed is how to combine using speech to communicate with pilots and other ATCOs with using speech to control the CWP.

The improvements in the HMI interaction modes and technologies aim to minimise the workload of the TMA and En-Route Controllers and may be applicable in current operations and/or in future operational concepts still under development and scope of other SESAR Solutions.

PJ.10-W2-96 ASR deals with both the current operating TMA and En-Route environment and future environments and is aligned to the current evolution happening within the SESAR 2020 Programme.

The enabler used for the validation is represented in EATMA DS22 and the architecture proposed for the validation exercises is described in the deliverable D4.1.020-PJ.10-W2-96 ASR-TRL6 Final TS/IRS section 3.

Table 3 below shows a quick breakdown of the Technical Enablers that make up the PJ.10-W2-96 ASR.

SESAR Technological Solution ID	SESAR Technological Solution Description	Master or Contributing (M or C)	Contribution to the SESAR Technological Solution short description	Enablers ref. (from EATMA)
PJ.10-W2-96 ASR				

	PJ.10-W2-96 ASR focusses on the Human Machine Interface of the CWP and is driven by operational requirements for innovative HMI needs captured from the stakeholders. PJ.16-04-02 ASR specified requirements for use of speech recognition in the En-Route and Approach, ATC environments to support controllers' productivity at the CWP	PJ.10-W2-96 ASR is the Master of this domain	PJ.10-W2-96 ASR starting from what achieved in PJ.16.04 aims to further develop the concept of Automatic Speech Recognition as part of a set of advanced technologies to be used either by themselves or as a part of a multimodal CWP to **enhance controller productivity** while maintaining safety level if not even increasing it.	**ER APP ATC 180 Controller productivity enhancements by Automatic Speech Recognition at the ER/APP CWP/HMI.**

Table 3: SESAR Technological Solution(s) under Validation

3.2 Summary of the Technological Validation Plan

3.2.1 Validation Plan Purpose

The TVALP **Error! Reference source not found.** document describes validation activities performed to take the Technical Enablers that make up Solution PJ.10-W2-96 ASR to the TRL6 complete phase of development. The objective of the TVALP – TRL6 is to set the framework for all the TRL6 activities performed by the solution members in order to validate different aspects and applications of ASR technologies that cover the research and validation needs described in the TVALP, which describes validation objectives, assumptions and exercises to be performed. Its schedule foresees Initial, Interim, and Final versions. The TVALP initial version included a detailed description of EXE-PJ.10-96-ASR-TRL6-03 , the intermediate version included a detailed description of EXE-PJ.10-96-ASR-TRL6-01. The final release of the TVALP included detailed descriptions of exercises EXE-PJ.10-96-ASR-TRL6-02 and EXE-PJ.10-96-ASR-TRL6-05. TVALP schedule was as follows:

- D4.1.030 Initial TVALP delivery, including detailed descriptions of exercises EXE-PJ.10-96-ASR-TRL6-03.
- D4.1.037 Interim TVALP delivery, including detailed descriptions of exercises EXE-PJ.10-96-ASR-TRL6-01.

- D4.1.040 Final TVALP, including also detailed descriptions of exercises EXE-PJ.10-96-ASR-TRL6-02 and EXE-PJ.10-96-ASR-TRL6-05.

3.2.2 Summary of Technological Validation Objectives and success criteria

Validation objectives were adapted taking into account that the solution under validation is technical in nature, as well as BIMs and feedback from safety and human performance assessment recommendations. Maturity gate criteria for TRL6 are held into account and include guidelines to ensure future coverage of such criteria. Individual exercises link to objectives in order to show how success criteria support the overall solution maturity. A traceability matrix Excel sheet showing links between High Level Validation Objectives and individual exercises is shown as an Appendix in the TVALP document **Error! Reference source not found.**, in its sections **Error! Reference source not found.** and **Error! Reference source not found.** .

Validation Identifier	Name	Primary Text	Success Criteria 1	Success Criteria 2	Success Criteria 3	Success Criteria 4	Success Criteria 5
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0010	ASR technical feasibility and interoperability (TFI)Analysis of new algorithm for collision avoidance	To assess the technical feasibility of the integration of the ASR system and its sub-systems into CWP and interoperability between the ASR sub-systems and the existing CWP systems and tools. To validate that the SESAR Technological Solution	The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the existing CWP systems and tools. Availability of systems and tools and their performan				

		XX new algorithm reduces the number of nuisance alerts	ce remains at 100%				
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0020	ASR performance stability (PST)	To assess the stability of the ASR system performance.	The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)	The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.			
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0030	ASR operational feasibility (OPF)	To assess the impact on the human performance of the integration of the ASR system and its sub-systems into operations in a realistic environment.	The introduction of the ASR system into the context of application is operationally viable, ATCos workload with ASR is equal or better than in baseline (without ASR support).	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognitio	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. Controllers' feedback with respect to Human Factors questionnaire is better	Changes in the design of the user interface (input devices, visual displays/output devices, alarm& alerts) support ATCos in carrying out the tasks.	The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are

				n Error Rate).	than for baseline.		identified).
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0040	Safety (SAF)	To assess the impact of the introduction of the ASR system on safety	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.	The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.	The level of ATCo's situational awareness is not reduced with the introduction of ASR system (ATCo is able to perceive and interpret task relevant information and anticipate future events/actions).	The level of ATCos' workload is maintained or decreased with the introduction of ASR system.
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0050	TMA Capacity	To assess the impact of the introduction of the ASR system on capacity.	The workload of ATCO after introduction of an ASR system is adequate to increase TMA capacity. The workload of ATCOs is less when working with ASR compared to baseline.	ASR allows ATCOs to safely manage a higher amount of aircraft, increasing the throughput in TMA.			

OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0060	Fuel reduction	To assess the impact of the introduction of the ASR system on Fuel efficiency	Aircraft will be able to improve their route Efficiency (fuel burnt) due to the higher throughput in TMA thanks to the introduction of ASR.				The Solution has not detect any impact on fuel reduction. The flight trajectory were not impacted by the use of ASR and therefore the fuel consumption didn't change.
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0070	Visualization navigation	To assess the impact of the introduction of the ASR in visualization navigation in Dynamic Airspace Configuration (DAC).	ATCos are able to perform a faster and more predictable navigation when using ASR to support DAC.				

Table 4 Validation Objectives and Success Criteria

3.2.3 Technological Validation Assumptions

Assumptions applicable to Solutions PJ.10-W2-96 ASR and which might have had an impact on the validation results are listed below. Such assumptions are applicable to all the validation exercises in the Validation Plan. Additional validation assumptions at exercise level are captured in each Appendix. Validation assumptions below are consistent with those available in the EATMA architecture. At the moment of publication, there are no validation assumptions to be found – as such – in EATMA. Deviations identified, are justified and reported in this document.

The list of validation assumptions presented in the table below contains the pre-requisites for the activities under the scope of the Validation Plan e.g. pre-Step 1, SESAR 1 SESAR Solutions required for validation.

Identifier	Title	Description	Justification	Impact on Assessment
AS-GEN-01	Actor Compliance	General compliance by all actors with existing standards and guidelines	This general compliance does not exclude occurrences of failures in the respect of the guidelines; it does not exclude possible deviations in early stages of implementation. Their likelihood as well as their consequences must be taken into account when defining the most important abnormal scenarios and performing the related Safety assessments.	N/A

AS-GEN-02	Standards	Separation standards and responsibilities unchanged.	N/A	N/A
AS-GEN-03	Flight Plan	Very high proportion (> 95 %) of commercial and military flights with Extended Flight Plan	N/A	N/A
AS-GEN-04	Training	Airborne and ATC have appropriate training and competencies.	Similar considerations as AS-GEN-01 regarding "exceptions".	N/A
AS-GEN-05		It is assumed that the SESAR Solution Catalogue related concepts are already validated and implemented.	N/A	N/A
AS-GEN-06		Validation activities uses reference scenarios to measure performance changes.	Check PJ19 guidance if needed.	N/A
AS-GEN-07		During the validation activities, it is assumed that the simulated traffic in the validation scenarios encompass the one corresponding to the FOC of the	N/A	N/A

		OI step to be validated.		
AS-GEN-08		Widely shared information among all necessary actors about key turnaround milestones, during planning and execution.		N/A

Table 5: Technological Validation Assumptions overview

3.2.4 Technological Validation Exercises List

Shown below is a short list of the validation exercises planned to achieve TRL6 maturity, and how they will contribute to cover the R&D needs at the solution level. An explanation follows as to why the set of planned validation exercises are required and sufficient to ensure that PJ.10-W2-96 ASR will progress from the TRL4 maturity level to TRL6.

			Exercises	*001*	*002*	*003*	*005*
			Leader	*LDO*	*ACG/DLR*	*ENAIRE /INDRA*	*SINTEF*
	OI step	*Tech nology*	*Simulation execution*	*May-22*	*Oct-22*	*Nov-21*	*Sep-22*
SOL 96	***POI-0055-SDM — Improving controller productivity by Automatic Speech Recognition (ASR) at the ER/APP CWP/HMI***	*ASR + AI/ML*	Assessment of efficiency improvements in datalink communications by means of the use of Automatic Speech Recognition.	X			
			Assessment of Recognition Performance and ATM Efficiency Improvements		X		

EUROPEAN PARTNERSHIP

Co-funded by the European Union

			Assessment of improvements in situational awareness thanks to the use of Automatic Speech Recognition			X	
			Assessment of how ASR may be used to enhance the ATCOs management of changes when using dynamic air space configuration (DAC).				X

In order to achieve TRL6-complete maturity, various pieces of information need to be collected, including projected levels of benefits in the related KPAs, identifying areas of possible impacts on Safety, on Human Performance, Workload Reduction and possible Security risks. In addition, the concept needs to show that its development has reached a stable state and that no major conceptual changes are foreseen in further activities.

3.2.4.1 Technological Validation Exercises

[EXE]

Identifier	EXE-PJ.10-96-ASR-TRL6-TVALP-01
Title	Assessment of efficiency improvements in datalink communications by means of the use of Automatic Speech Recognition.
Description	To validate that the SESAR Technological Solution PJ.10-96-ASR has achieved an acceptable Automatic Speech Recognition Performance so that ATCO’s workload can be reduced resulting in an increase of ATM efficiency at least in medium traffic scenarios and in particular in datalink communications.
Expected achievements	Workload Reduction, Increase Operational Efficiency
TRL	<TRL6>
T. Validation Technique	RTS
Start Date	17/05/2022

End Date	20/05/2022
T. Validation Coordinator	LDO
T. Validation Platform	Lead In Sky
T. Validation Location	Rome
Status	<in progress>
Dependencies	Other dependent exercises

[EXE Trace]

Linked Element Type	Identifier
<SESAR Solution>	SESAR Solution 96
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0010
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0020
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0030
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0040

Identifier	EXE-PJ.10-96-ASR-TRL6-TVALP-02
Title	Assessment of Recognition Performance and ATM Efficiency Improvements
Description	To validate that the that the SESAR Technological Solution PJ.10-96-ASR has achieved an acceptable Automatic Speech Recognition Performance so that ATCO's workload can reduce resulting in an increase of Operational efficiency at least in high density traffic scenarios.
Expected achievements	Workload Reduction, Increase Operational Efficiency
TRL	<TRL6>
T. Validation Technique	Speech Recognition Performance Test in real life environment Laboratory Test at DLR for Workload Reduction and ATM Efficiency Increase
Start Date	11/10/2022
End Date	14/10/2022

T. Validation Coordinator	Austro Control
T. Validation Platform	DLR
T. Validation Location	Vienna for Speech Recognition Performance Test, final validation Braunschweig
Status	<in progress>
Dependencies	Other dependent exercises

[EXE Trace]

Linked Element Type	Identifier
<SESAR Solution>	SESAR Solution 96
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0010
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0020
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0030
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0040
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0050
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0060

[EXE]

Identifier	EXE-PJ.10-96-ASR-TRL6-TVALP-03
Title	Assessment of improvements in situational awareness thanks to the use of Automatic Speech Recognition.
Description	To support ATCOs to quickly identifying new flights entering the sector or flight crews requesting actions from ATCOs by highlighting their callsigns in the ATCOs CWP, in illuminating the callsign coming from a controller utterance to provide a safety check to the controller, and in providing access for consultation to a historical annotation of the R/T exchanged with the aircraft.
Expected achievements	Workload Reduction, Increase of situational awareness
TRL	TRL6
T. Validation Technique	RTS
Start Date	08/11/2021
End Date	12/11/2021

T. Validation Coordinator	ENAIRE
T. Validation Platform	SACTA
T. Validation Location	Madrid
Status	<in progress>
Dependencies	Other dependent exercises

[EXE Trace]

Linked Element Type	Identifier
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0010
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0020
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0030
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0040

[EXE]

Identifier	EXE-PJ.10-96-ASR-TRL6-TVALP-05
Title	Assessment of how ASR may be used to enhance the ATCOs management of changes when using dynamic air space configuration (DAC).
Description	To validate that the SESAR Technological Solution PJ.10-W2-96 ASR has achieved an acceptable Automatic Speech Recognition Performance so that ATCOs increase their understanding of how sectorization changes when using DAC, and how this influences the traffic being controlled. The exercise will address the en route phase, using the Milan ACC.
Expected achievements	Increased situational awareness and safety
TRL	<TRL6>
T. Validation Technique	RTS
Start Date	26/09/2022
End Date	29/09/2022
T. Validation Coordinator	SINTEF
T. Validation Platform	SIMADES ATC

T. Validation Location	Oslo, Norway, using a mix of Italian and Norwegian ATCOs
Status	<in progress>
Dependencies	Exercise is conducted in cooperation with EXE5 in PJ09-W2-Sol44

[EXE Trace]

Linked Element Type	Identifier
	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0010
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0020
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0030
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0040
<V&V Objective>	OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0070

3.3 Deviations

3.3.1 Deviations with respect to the S3JU Project Handbook

No deviation.

3.3.2 Deviations with respect to the Technological Validation Plan

No deviation.

4 SESAR Technological Solution 96 ASR Validation Results

4.1 Summary of SESAR Technological Solution 96 ASR Validation Results

In general Assistant Based Speech Recognition, based on Automatic Speech Recognition techniques, has established itself as a reliable support tool for many Air Traffic Management environments, and the validation exercises seem to bring encouraging results.

Please refer to Appendixes A-D for detailed information on each exercise and on their detailed results.

For the outcome of the Technological Validation Objective Status at Solution level it was stated OK when all Exercises had an OK result, it was stated POK when at least one Exercise had an OK result, and NOK when all Exercises had a NOK result.

SESAR Technological Solution Technological Validation Objective ID	SESAR Technological Solution Technological Validation Objective Title	SESAR Technological Solution Success Criterion ID	SESAR Technological Solution Success Criterion	SESAR Technological Solution Validation Results	SESAR Technological Solution Technological Validation Objective Status
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0010	ASR technical feasibility and interoperability (TFI	CRT-Sol.96ASR-TRL6-TVALP-0010.001	The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the existing CWP systems and tools. Availability of systems and tools and their	EXE-01, EXE-03, EXE-02, EXE-05, OK	OK

			performance remain at 100%		
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0020	ASR performance stability (PST)	CRT-Sol.96ASR-TRL6-TVALP-0020.001	The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)	EXE-01,POK EXE-03,POK EXE-02 OK, EXE-05 POK	POK
		CRT-Sol.96ASR-TRL6-TVALP-0020.002	The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.	EXE-01 NOK, EXE-02 OK, EXE-03 NOK, EXE-05 OK	POK
OKOBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0030	ASR operational feasibility (OPF)	CRT-Sol.96ASR-TRL6-TVALP-0030.001	The introduction of the ASR system into the context of application is operationally viable, ATCos workload with ASR is equal or better than in baseline (without ASR support).	EXE-01, EXE-03, EXE-02 OK EXE-05 POK	POK
		CRT-Sol.96ASR-	The accuracy of the information provided by	EXE-01, EXE-03, EXE-02 OK EXE-05 POK	POK

		TRL6-TVALP-0030.002	the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate).		
		CRT-Sol.96ASR-TRL6-TVALP-0030.003	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. Controllers' feedback with respect to Human Factors questionnaire is better than for baseline.	EXE-01, EXE-03, EXE-02 OK EXE-05 POK	POK
		CRT-Sol.96ASR-TRL6-TVALP-0030.004	Changes in the design of the user interface (input devices, visual displays/output devices, alarm& alerts) support ATCos in carrying out the tasks.	EXE-01, EXE-03, EXE-02, EXE-05, OK	OK

		CRT-Sol.96ASR-TRL6-TVALP-0030.005	The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified).	EXE-01 POK, EXE-03 OK, EXE-02 OK EXE-05 POK	POK
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0040	Safety (SAF)	CRT-Sol.96ASR-TRL6-TVALP-0040.001	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.	EXE-01 NOK, EXE-03 NOK, EXE-02 OK, EXE-05 OK	POK
		CRT-Sol.96ASR-TRL6-TVALP-0040.002	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.	EXE-01, EXE-03, EXE-02, EXE-05, OK	OK
		CRT-Sol.96ASR-TRL6-TVALP-0040.003	The number and/or severity of errors resulting from	EXE-01, EXE-03, EXE-02, OK EXE-05 POK	POK

			the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.		
		CRT-Sol.96ASR-TRL6-TVALP-0040.004	The level of ATCo's situational awareness is not reduced with the introduction of ASR system (ATCo is able to perceive and interpret task relevant information and anticipate future events/actions).	EXE-01, EXE-03, EXE-02, EXE-05, OK	OK
		CRT-Sol.96ASR-TRL6-TVALP-0040.005	The level of ATCos' workload is maintained or decreased with the introduction of ASR system.	EXE-01, EXE-03, EXE-02, OK EXE-05 POK	POK

OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0050	TMA Capacity	CRT-Sol.96ASR-TRL6-TVALP-0050.001	The workload of ATCO after introduction of an ASR system is adequate to increase TMA capacity. The workload of ATCOs in TMA is decreased by 10% when working with ASR compared to baseline.	EXE-02 OK	OK
		CRT-Sol.96ASR-TRL6-TVALP-0050.002	ASR allows ATCOs to safely manage a higher amount of aircraft, increasing the throughput in TMA.	EXE-02 OK	OK

OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0060	Fuel reduction	CRT-Sol.96ASR-TRL6-TVALP-0060.001	Aircraft will be able to improve their route Efficiency (fuel burnt) due to the higher throughput in TMA thanks to the introduction of ASR.	EXE-02, NOK	NOK
OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0070	Visualization navigation	CRT-Sol.96ASR-TRL6-TVALP-0070.001	ATCos are able to perform a faster and more predictable navigation when using ASR to support DAC.	EXE-05, POK	POK

Table 6: Summary of Technological Validation Exercises Results

4.2 Detailed analysis of SESAR Technological Solution Validation Results per Validation objective

The following tables summarize coverage of objectives and success criteria per validation exercise. Please refer to the previous section for detailed descriptions of objectives and criteria. A detailed analysis refers to every success criterion in order to further elaborate regarding whether success was actually achieved and where success was only partial, to give details regarding quantitative and qualitative aspects, as well as the results obtained in more detail, in any case covered at length in the appropriate appendixes.

4.2.1 OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0010 Results

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
CRT-Sol.96ASR-TRL6-TVALP-0010.001	A preindustrial ASR prototype with operational systems, including an operational LeadInSky CWP that provides context information in real time to the ASR (flight plan list in this approach), receives information from the ASR and presents it to the controller in a coherent manner with the rest of CWP information. The exercise also included connection with an operational voice	The ABSR system was successfully integrated into a prototypic CWP. The voice signal was received via VoIP. This signal, radar data, flight plans, and the command prediction output served as the input for the speech-to-text component. The speech-to-text output was based on a DNN. and was used by the command extraction module to recognize ATC concepts following the	The ASR was successfully integrated with operational CWP and communication system. The ASR did not impact the performance of previous systems	The speech recognition functionality was added to the SIMADES CWP. The CWP does not support voice communication between ATCOs and pseudo pilots. In the exercise, Microsoft Teams with a press-to-talk add-on was used for ATCO-pilot communication. Voice commands were turned on either by clicking on a button in the CWP HMI or preferably using a keyboard shortcut. It was automatically turned off when a command had	The ASR prototypes were successfully integrated with operational systems. There was no negative impact on the other systems of the CWP and the performance of the previous system.

Co-funded by the European Union

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
	communication system.	defined ontology. The output was presented in radar labels of aircraft. There was no negative impact on the other systems of the CWP. The functionality remained active throughout all solution runs without malfunctions.		been processed. There were special labels on the keyboard keys for ATCO-pilot communication and for giving voice commands. The ASR did not impact the performance of rest of the CWP	

Table 7 Solution 96 ASR OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0010 Results

4.2.2 OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0020 Results

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
CRT-Sol.96ASR-TRL6-TVALP-0020.001	The performance is maintained. Nevertheless, the command type recognition rate was different among command types and ATCOs varing from 94% to 99%. The callsign recognition rate from 66,5% to 87%.	The command recognition rate was 92.5% with regards to the command types that were shown to the ATCOs in the radar labels. This is the only quantified command recognition rate (not command type recognition rate) of all Sol96 validation exercises. The	The command type recognition rate for controllers is 89%-92%. The callsign recognition rate is 72%	Recognition rates of voice commands in the ordinary runs were considered to be high for most controllers. The only female controller experienced lower recognition rates, and decreasing recognition throughout the exercise. The	The command recognition rate was as required by the TS/IRS for some exercises and a bit lower for other depending on callsign or command recognition.

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
		important command recognition error rate was less than 2.5%. When performing an offline analysis on the same data, even 93.4% command recognition rate and 1.7% command error rate are achieved. The callsign recognition rate reached 97.8% with an error rate around 0.5%. If only ATCos 3-12 are considered, the command recognition rate is even beyond 95% The word error rate of automatic transcripts (not being a metric here) was 3.2%. Hence, the required rates (TS/IRS) have been outperformed as also		performance with regards to response time for the voice commands were not satisfactory and caused the controllers to not use them in situations where they might have been very useful if response time had been faster. The response time issue is inherent in the state-of-the-art technology used in the exercise, and is currently a problem for web-based speech recognition.	

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
		subjectively confirmed by ATCOs.			
CRT-Sol.96ASR-TRL6-TVALP-0020.002	Differences among different command types were greater than 2.5%, so the criteria was not satisfied	There were 17 different command types relevant for the radar label (i.e., appeared in more than 0.05% of all commands). Just five of them had a recognition rate less than 86%, but all 17 command types had recognition rates above 71%. However, those five command types made just 1.5% of all recognized commands which shows that large amount of data is required to achieve better performance. The success criterion with differences smaller 2.5% was not satisfied.	The difference between different commands used in the exercise was within the 10%	There were no observable differences in recognition rate between the different commands. In the test runs some commands had worse recognition than others, and there were some issues in recognizing certain numbers. These issues disappeared when the machine learning had worked on the data from the training sessions.	The success criterion with differences smaller 2.5% was not satisfied.

Table 8 Solution 96 ASR OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0020 Results

4.2.3 OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0030 Results

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
CRT-Sol.96ASR-TRL6-TVALP-0030.001	Solution scenarios can be considered homogeneous, with no significant peaks in workload. most ATCO responses (90%) indicated an acceptable level of WL	The ATCos spent 25.3% of their time, i.e., 12763 of 50400 seconds, just for clicking in the baseline scenario without ABSR support. They just needed 0.8% of their time for clicking in the solution runs (400 seconds). Furthermore, the secondary task (Stroop test) indicated a significant decrease of ATCo workload when being supported by an ABSR system (α of 0.2% in t-test for both run types, i.e., medium and heavy traffic density scenario when compensating sequence effects of baseline/solution runs).	ATCOs stated that the workload did not change or was decreased using the ASR technology	During the debriefings controllers gave no indication of increased workload due to use of ASR.	ATCOs stated that the workload did not change or was decreased using the ASR technology
CRT-Sol.96ASR-TRL6-	ATCOs agreed that the accuracy of the information	The good command recognition rates have already been outlined	Controller considered that the recognition rate was	ATCOs agreed that the accuracy of the information	Controller considered that the recognition rate was not enough to

 Co-funded by the European Union

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
TVALP-0030.002	provided by the ASR system was adequate to accomplish the operations although there were differences among ATCOs	above for CRT-Sol.96ASR-TRL6-TVALP-0020.001. Furthermore, four subjectively rated questionnaire items on user acceptance, confidence, complexity, and the "risk to forget something" were rated significantly better when being supported by ASRU (0%≤α<5% for both run types). Such a high statistical significance was not given for other statements such as reliability, stress, conflict resolution and focus on single aircraft. In addition, the verbal feedback of ATCos was that ASR system is adequate for the accomplishment of operations. ATCos confirmed the acceptable level of callsign recognition rates	not enough to support operations	provided by the ASR system was not always adequate to accomplish the operations.	support operations for two exercises, but in one exercise the ATCOs feedback was that ASR system is adequate for the accomplishment of operations. For another the feedback was that ASR system was mostly adequate, but not always.

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
		with an average score of 9.4 and command recognition rates with 8.4 of ASR output in the aircraft radar labels on a scale from 1 up to 10, i.e., 10 indicates the best rating option.			
CRT-Sol.96ASR-TRL6-TVALP-0030.003	ATCOs agreed that the timeliness of information was acceptable	ATCos confirmed the timeliness of ASR output in the aircraft radar labels with an average score of 8.5 on a scale from 1 up to 10, i.e., 10 indicates the best rating option.	Controllers considered that the timeliness of the callsign recognition at the beginning of the phrase should be higher. The timeliness of the callsign at the end of the utterance and event recognition was enough although could be improved	As indicated above, the response time was considered by the ATCOs to reduce the value of some of the voice commands. The ASR functionality did not provide any feedback except for a transcript of the recognized commands at the bottom of the CWP screen (for a short time), the success criterion is partly not applicable.	ATCOs agreed that the timeliness of information was acceptable.
CRT-Sol.96ASR-TRL6-TVALP-0030.004	Most controllers (nearly 70%) were generally satisfied by the interaction	ATCos confirmed that the human machine interface, i.e., the ASR output handling	Majority of responses obtained through show that the Human-	All controllers found the commands for enhancing the understanding of the	ATCos confirmed that the human machine interface, i.e., the ASR output

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
	and no issues were reported. During the debriefings, ATCOs mentioned that the interaction with CPDLC would be further improved if the latency of the tool were lower.	provided suitable access to relevant information in all situations and was both comprehensible and acceptable with an average score of 7.8 and 7.7, respectively, on a scale from 1 up to 10, i.e., 10 indicates the best rating option.	Machine Interface was adequate and appropriate to execute the simulation activity	sectorization to be useful, while the opinions diverged regarding the commands for understanding how traffic was affected by sectorization changes and for controlling the 3D visualization. 3D navigation was also mentioned in the interviews as a type of functionality that would have been more useful if the response to the voice commands had been faster. A majority of the controllers found that speech commands enhanced the CWP, and that it should be included in a CWP supporting DAC.	handling was adequate and in one exercise even very satisfactory.
CRT-Sol.96ASR-	During the debriefings,	ATCos confirmed that their trust in	Answers indicate	Interviews and questionnaires	For most exercises the

EUROPEAN PARTNERSHIP

Co-funded by the European Union

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
TRL6-TVALP-0030.005	ATCOs mentioned that ASR struggles to recognise commands in an accurate manner, unless the speech is very slow or segmented and provided at a steady pace. Therefore the average confidence was not high.	the ASR system and their satisfaction with the ASR system were high with an average score of 8.3 (both) on a scale from 1 up to 10, i.e., 10 indicates the best rating option.	that although the system was useful and understable it was not accurate or reliable enough to be confident with the system	indicate that the level of trust in the functionality of the speech recognition functionality was high.	confidence of the ATCOs in the ASR was very high, nevertheless for one was just sufficient.

Table 9 Solution 96 ASR OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0030 Results

4.2.4 OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0040 Results

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
CRT-Sol.96ASR-TRL6-TVALP-0040.001	ATCOs agreed that the accuracy of the information provided by the ASR system was most of the time adequate to accomplish the operations although there were big differences among ATCOs	As reported for CRT-Sol.96ASR-TRL6-TVALP-0020.001, the relevant command error rated stayed below 2.5%; the callsign error rate was around 0.5%. ATCos confirmed that level and quality of information provided by the system (in	Controllers considered that the accuracy was not enough to support them	There are no indications that the results presented above regarding recognition rates affected safety	For some exercises ATCOs considered that the accuracy was not enough or just sufficient to support them, but for one exercise ATCos confirmed that level and quality of information provided by the system (in the radar labels) were

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
		the radar labels) were an acceptable level with an average score of 8.8 on a scale from 1 up to 10, i.e., 10 indicates the best rating option. ATCos also confirmed that operating methods are clearly identified and consistent in all operating condition with an average score of 8.3 on a scale from 1 up to 10, i.e., 10 indicates the best rating option.			an acceptable level.
CRT-Sol.96ASR-TRL6-TVALP-0040.002	ATCOs agreed that the timeliness of information was acceptable	ATCos confirmed the timeliness of ASR output in the aircraft radar labels with an average score of 8.5 on a scale from 1 up to 10, i.e., 10 indicates the best rating option.	Controllers considered that the timeliness of the callsign recognition at the beginning of the phrase should be higher. The timeliness of the callsign at the end of the utterance and event recognition was enough although	There are no indications that the results presented above regarding response times rates affected safety	The timeliness of the ASR output was confirmed and in some cases wished to be higher.

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
			could be improved.		
CRT-Sol.96ASR-TRL6-TVALP-0040.003	33% of ATCOs either agreed or strongly agreed that ASR did not increase potential for human error compared to current operations, whereas 17% of ATCOs disagreed, and the rest (50%) were neutral about it.	ATCOs confirmed that ASR did NOT increase the potential for human errors with an average score of 3 on a scale from 1 up to 10, i.e., 10 indicates the worst rating option. Objective analysis even shows that the number of errors in the radar label cells, i.e., missing input is much less if ATCOs are supported by ASR compared to enter everything manually with mouse (α<10-7 %).	No error resulted from the introduction of the ASR	There are no indications of errors resulting from the introduction of the ASR.	ATCOs confirmed that ASR did NOT increase the potential for human errors.
CRT-Sol.96ASR-TRL6-TVALP-0040.004	All ATCOs were in agreement that the 'Hook' function improves situational awareness and could also be used as a safety barrier for avoiding	ATCOs confirmed that their situational awareness is maintained at acceptable level with ASR with an average score of 8.9 on a	All ATCOs saw their situational awareness was increased or unaffected with the introduction of the ASR system.	The use of ASR did not affect the SA negatively. This is also supported by the interviews, where no ATCOs expressed	ATCOs stated that the situational awareness was increased or unaffected with the introduction of the ASR system.

	EXE-001	**EXE-002**	**EXE-003**	**EXE-005**	Sol.96 ASR
	confusion between inbound flights with similar callsigns.	scale from 1 up to 10, i.e., 10 indicates the best rating option.		that the SA was decreased when using ASR.	
CRT-Sol.96ASR-TRL6-TVALP-0040.005	Solution scenarios can be considered homogeneous, with no significant peaks in workload. most ATCO responses (90%) indicated an acceptable level of WL	The secondary task for objective workload measurement (Stroop test) showed a statistically significant decrease (α=0.2%) of workload when ATCos are supported by ASR. The ATCo-self-rated ISA score confirmed this with the same statistical significance (α=0.2%). In the final questionnaire, ATCos confirmed that ASR supported them in maintaining workload at acceptable level with an average score of 7.9 on a scale from 1 up to 10, i.e., 10 indicates	ATCOs stated that the workload did not change or was decreased using the ASR technology.	the WL level were slightly higher for the ATCOs using ASR than the ones not using it, but this may just as well be explained by a combination of differences in the dispatch of traffic between the sessions and the unbalance between the workload between the sectors in parts of the sessions.	ATCOs stated that the workload did not change or was decreased using the ASR technology. For one exercise a secondary task was performed for objective workload measurement (Stroop test) and showed a statistically significant decrease (α=0.2%) of workload when ATCos are supported by ASR.

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
		the best rating option.			

Table 10 Solution 96 ASR OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0040 Results

4.2.5 OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0050 Results

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
CRT-Sol.96ASR-TRL6-TVALP-0050.001	N.A.	When using ASRASR in the heavy density traffic scenario, the flow could slightly be increase compared to the baseline run. The significant decrease of workload has already been reported for CRT-Sol.96ASR-TRL6-TVALP-0040.005)	N.A.	N.A.	ASR in the heavy density traffic scenario of the TMA has been investigated in one exercise. The flow could slightly be increase compared to the baseline run. The significant decrease of workload has been reported.
CRT-Sol.96ASR-TRL6-TVALP-0050.002	N.A.	Safety and throughput results have been reported in CRT-Sol.96ASR-TRL6-TVALP-0030.002, CRT-Sol.96ASR-TRL6-TVALP-0040.003, and CRT-Sol.96ASR-	N.A.	N.A.	ATCos confirmed that ASR did NOT increase the potential for human errors.

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
		TRL6-TVALP-0050.001.			

Table 11Solution 96 ASR OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0050 Results

4.2.6 OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0060 Results

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
CRT-Sol.96ASR-TRL6-TVALP-0060.001	N.A.	Throughput results have been reported in CRT-Sol.96ASR-TRL6-TVALP-0050.001, i.e. we have roughly an increase of one movement per hour in solution runs.	N.A.	N.A.	In TMA the increase of throughput is roughly of one movement per hour in solution runs

Table 12Solution 96 ASR OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0060 Results

4.2.7 OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0070 results

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
CRT-Sol.96ASR-TRL6-TVALP-0070.001	N.A.	N.A.	N.A.	In two of the sectors more flights were controlled when using ASR than when not. In the other three sectors the opposite was the case. In all sectors the difference was 4 flights or less, and summed, the non-ASR sectors had a throughput	In some sectors more flights were controlled when using ASR than when not. In other sectors it was the opposite.

	EXE-001	EXE-002	EXE-003	EXE-005	Sol.96 ASR
				of 4 flights more than the ASR sectors.	

Table 13 Solution 96 ASR OBJ-PJ.10-W2-96 ASR-TRL6-TVALP-0070 Results

4.3 Confidence in the Validation Results

4.3.1 Limitations of Technological Validation Results

Targeting TRL6, validation exercises could be considered to be operational. Technology Readiness Level 6 is defined as "System/subsystem model or prototyping demonstration in a relevant end-to-end environment (ground or space)". However, each exercise was characterised by its own limitations due to a variety of reasons. In the sections below a brief summary is provided, always bearing in mind that more detailed descriptions are available in the Appendixes.

4.3.1.1 EXE-PJ.10-W2-96 ASR-TRL6-01

Considering the simulation conditions, the results for ASR are judged to be characterised by a high level of significance, even if the training of ATCO was quite limited for time constraints reasons and this might have affected the collection of data of initial runs of each simulation day.

The exercise involved Italian ATCOs on a platform which is not the one they work with every day, but it is the more up to date of the LeadInSky system and therefore the prototype was developed on that.

4.3.1.2 EXE-PJ.10-W2-96 ASR-TRL6-02

The human-in-the-loop simulations took place in a high-fidelity air traffic control center simulation environment. It encompassed two baseline runs (without ASR) and two solution runs (with ASR), each of them with a medium and a heavy traffic scenario (next to training runs). As coordinated and updated in the course of the three pre-trials, the situation data display (SDD) has been arranged so as to be close to the SDD of AustroControl who sent the twelve study participants. Interviews confirmed the realism of the hardware and software environment for the given TRL. The handling of non-nominal situations throughout the simulation runs was very limited.

4.3.1.3 EXE-PJ.10-W2-96 ASR-TRL6-03

Air Traffic Control Operators were asked during the post-run questionnaire the following questions regarding the simulation environment:

The answer range to these questions went from 1 ("Very far from reality") to 4 ("Very realistic")

Controllers considered **the run as adapted to reality or very realistic**. There were some comments regarding the need to perform more actions than in real live in some cases (e.g. squawk change) but considered them as useful to thoroughly test the ASR tool.

Traffic load was medium/ high and the communication load was adapted to the simulated use cases to meet the validation objectives.

The exercise simulated an **opening and closure of sectors** to gather possible requirements associated to it.

4.3.1.4 EXE-PJ.10-W2-96 ASR-TRL6-05

EXE-005 involved more than 10 ATCOs and pseudo pilots from Italy and Norway. The ATCOs have varying age and experience, although all are considered as experienced. All but one controller are male, but this may also be viewed as more representative for ATCOs than for other professions. The airspace and traffic used in the exercise are both realistic. Training on the platform to which the ASR functionality were added was limited, but as the training on the ASR system was also used for data collection, this is not considered a limitation. The basic functionality of the radar screen was limited compared to what is usually provided for the ATCOs involved. As the focus of the combined exercise was supporting ATCOs when using DAC, this is not considered a limitation neither.

4.3.2 Quality of Technological Validation Exercises Results

Quality and accuracy of exercises results are, as it usually happens, a mix of objective measurements and subjective considerations, which are then measured with qualitative assessments using various questionnaire methodologies.

All exercises in the solution can be considered as having high quality. A summary per exercise is provided below.

4.3.2.1 EXE-PJ.10-W2-96 ASR-TRL6-01

Questionnaires have been used to collect ratings from the test subjects on the different aspects of ASR as explained in section A.7: both accuracy and confidence in the collected results as well as measured indicators are judged to be of high quality to support the maturity assessment of TRL6 phase.

4.3.2.2 EXE-PJ.10-W2-96 ASR-TRL6-02

Many subjective (questionnaire) data has been collected after each simulation run. However, many of those results were confirmed by a row of objective measurements taken during the simulation runs. The pure number and topic coverage of measurements deliver a wide range of performance results on ASR usage in ATC.

4.3.2.3 EXE-PJ.10-W2-96 ASR-TRL6-03

The quality of the exercises result is considered as high.

Regarding operational significance all runs were performed on traffic scenarios and sector configurations based on the two Madrid ACC En-route sectors with adapted traffic from 2019. All controllers involved in the exercise were operational experienced professionals of Madrid ACC that volunteered for non-profit purposes.

Pseudopilots participating in the simulation have a pilot license and have previously participated in other real time simulations performed at Crida and Enaire.

The validation platform implemented is an operational SACTA.4 CWP with an operational communication system connected with a simulation engine. The simulation engines and part of the platform was used in previous simulations in SESAR wave 1 and has been updated to execute the exercise.

4.3.2.4 EXE-PJ.10-W2-96 ASR-TRL6-05

The ATCOs were interviewed after each most runs and after the exercise. The ATCOs also answered a questionnaire after the runs in which using ASR was compared with not using ASR. As explained above, the results from parts of the questionnaire are limited as there was a common questionnaire for these two sessions. Thus, the quality of the qualitative data collected is medium to high.

Various logs and observations resulted in quantitative data that are considered as having high quality.

4.3.3 Significance of Technological Validation Exercises Results

The combination of the joint outcome of the different exercises, provides the results of solution with high significance.

As in the previous sections, detailed explanations and details are given in each Appendix corresponding to the Validation Exercises.

4.3.3.1 EXE-PJ.10-W2-96 ASR-TRL6-01

The simulation exercise has been conducted on an experimental platform representing Sofia ACC environment with a high degree of fidelity providing an operational significance adequate to support the TRL6 maturity assessment, of course with limitations already mentioned in Sections A.7.1 and A.7.2.

A significant total number of runs has been conducted among 3 simulation days (12 total number of runs) as well as a significant number of test subjects (6 ATCOs) have been involved to conclude that results are significant to support the TRL6 maturity assessment, but results cannot be relied upon as having statistical significance. Considering the validation technique (real time simulation) and the executed numbers of runs, results are deemed to have a high level of significance.

4.3.3.2 EXE-PJ.10-W2-96 ASR-TRL6-02

Twelve ATCos of AustroControl (already a high number for comparable studies) participated in the final trials. As each ATCo did four runs and there were a lot of objective (ASR performance, throughput, workload, etc.) and subjective (workload, acceptance, usability, errors, etc.) measurements, we were able to perform statistical significance tests such as the t-test. Many results confirm a better performance in the solution runs compared to baseline runs with an α below 5%, i.e., a high confidence (i.e. statistical significance). For details please refer to the single results of EXE002.

4.3.3.3 EXE-PJ.10-W2-96 ASR-TRL6-03

Confidence on qualitative results is **medium as the number runs performed** by each ATCO and incidents simulated during the exercise are in line of a real time simulation.

Nevertheless, **confidence on quantitative results is low due to the number of runs being low for statistical significance.** They should be taken into account together with the comments from the debriefings and the standard deviation provided.

The statistical significance of the recognition and error rates is considered as high due to the number of utterances analysed from the real time simulation and the statistical analysis.

4.3.3.4 EXE-PJ.10-W2-96 ASR-TRL6-05

The significance of the qualitative results is considered as high. The limitations just explained of parts of the questionnaire primarily influence the quantitative data collected through the questionnaires.

The statistic significant of the quantitative data is considered as low. As explained in different parts of Section D7.2 above, it is not guaranteed that the traffic and traffic distribution between the two sessions being compared were identical in the two runs. Also, the limitations of the results for parts of the questionnaire also supports this conclusion. Furthermore, as only 5 ATCOs were involved in these two runs, differences in findings need to be large to claim statistical significance.

5 Conclusions and recommendations

5.1 Conclusions

PJ.10-W2-Sol.96 ASR is a promising step in the direction of introducing novel human machine interface methods in the TMA and En-Route environments, with the introduction of Automatic Speech Recognition and Understanding (ASR) technology. Some considerations follow regarding the outcome of the Validation Exercises which make up this SESAR Solution.

5.1.1 Conclusions on SESAR Technological Solution maturity

This Solution moves from the achievements of PJ.16-04-ASR which partly achieved TRL4. This report contains the validation results of four different exercises to demonstrate that the ASR activity in PJ.10-W2-Sol 96 ASR has achieved TRL 6 (only partly when considering just some validation exercises).

Although most of the items for achieving TRL6 are met, some items need to be reconsidered in the future. Details which items are fully achieved (OK), which are partly achieved (PLK) and which are not achieved (NOK) are provided in the Maturity Assessment Tables Appendix E.

5.1.2 Conclusions on technological feasibility

The ASR technology (incl. capture of TMA/En-route ATC instructions and input of commands into the ATC system) has shown to be feasible in an ATC TMA/En-Route environment. Results of the Validation Exercises indicate good performance and positive results of the assessment of the ASR tool made by ATCOs, depending on the specific speech-to-text engine and text-to-concept performance (basically the command recognition rate, the command recognition error rate, and the callsign recognition rate with callsign recognition error rate give a clue about performance).

	Word Error Rate	Command recognition rate	Command recognition error rate	Command type recognition rate	Command type recognition error rate	Callsign recognition rate	Callsign recognition error rate
EXE-001	N.A.	N.A.	N.A.	97.25%	2.75%	77.75%	22.25% (error + rejection rate)
EXE-002	3.1%	92.5% (offline even 93.4%)	2.4% (offline even 1.7%)	94.3% including value Not a helpful metric (number is greater	1,9% including value Not a helpful metric (number is smaller	97.8% (offline even 97.9%)	0.6% (offline even 0.5%)

				than command recognition rate as a command is only considered to be recognized correctly in this exercise if all command elements such as type, value, unit, qualifier, condition are recognized correctly, and especially the callsign.	than command recognition error rate as a command is only considered to be recognized correctly in this exercise if all command elements such as type, value, unit, qualifier, condition are recognized correctly)		
EXE-003				89% (RTS)-92% (Std)		84% (ATCo RTS)-87% (ATCo Std)) 67% (FC RTS) – 49%(FC Std)	
EXE-005	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.

Table 14 Solution 96 ASR Rates

The ASR supported by AI and Machine Learning is a functional block in EATMA linked to the functions “Command Prediction”, “Recognize voice words”, and “Apply Ontology and logical check”.
Eight Use Cases have been defined to validate the concept, all of them have been addressed by the Solution:

Use Case	EXE-01	EXE-02	EXE-03	EXE-05
Highlight of callsigns on the CWP from pilot utterances			X	

Highlight of callsigns on the CWP from controllers utterances	X	X	X	
Automatic transcription and annotation of controllers commands			X	
Pre-filling of commands of the ATCO in the CWP	X	X		
Voice commands for highlighting an upcoming sectorization change in the CWP				X
Voice commands for highlighting the flights that will be affected by an upcoming sectorization change in the CWP				X
Voice commands for navigating the 3D visualization of the air space in the CWP				X
Prefilling of Datalink commands	X			

A common ontology as PJ.16.04 inheritance has been evolved and agreed among Solution members, to define a set of commands on which the ASR components have been instructed and trained.

The prototypes were successfully integrated in the simulation /operational platforms without impact to the availability or reliability of the other functions of the CWP.

However, a list of recommendations to enhance the ASR system (in testing environment) has been made. The quantitative and qualitative feedback of ATCOs was good and motivating. All implemented ABSR prototypes fulfilled their intended functionality during the simulation runs without serious malfunctions.

5.1.3 Conclusions on performance assessments

Several subjective and objective human performance measures have been taken during and after the simulation runs to compare baseline and solution performance. The detailed results can be found in section 4 and the annexes. The mean workload and situation awareness were found to stay within acceptable limits when ATCos are supported by an automatic speech recognition and understanding system.

5.2 Recommendations

Recommendations should be focused on what it will eventually take in order to put the ASR in operation.

5.2.1 Recommendations for next phase

A set of recommendations have been figured out in order to sharpen ASR operation, supported by AI and Machine Learning, among them:

- Consider a larger amount of representative training data (especially speech data from ATC operations' rooms)

- Consider pilot utterances in order to enable reasonable callsign highlighting at ATCo side and readback error detection
- Consider further applications that use the speech recognition and understanding output such as pre-filling of CWP mask and command not sent to the pilot, advanced readback error detection, incident analysis, on-the-job training support
- Intensify the use and enhance European-wide agreed ontology for annotation of ATC utterances
- Foster standardization of ASR input and output content as well as format in order to improve system interoperability and comparability

5.2.2 Recommendations for updating ATM Master Plan Level 2

PJ.10-W2-Sol.96 ASR is currently defined as follows:

	ID	Title	Description
Solution	PJ.10-W2-Sol.96 ASR	Automatic Speech Recognition	ATCOs will be supported by introducing innovative human machine interaction such as Automatic Speech Recognition. The goal is to automatically support certain tasks of the ATCO, which are either not performed at all or performed manually in today's ER APP systems / CWPs.
OI Step	POI-0055-SDM	Improving controller productivity by Automatic Speech Recognition (ASR) at the ER/APP CWP/HMI	Innovation of human machine interaction through the use of Automatic Speech Recognition (enhanced by AI algorithms and machine learning techniques) for tower controllers. The goal is to automatically support certain tasks of the ATCO, which are not done or done manually in today's systems/CWPs.
Enablers	ER APP ATC 180	Controller productivity enhancements by Automatic Speech Recognition at the ER/APP CWP/HMI	Introduction of new automated functions for Automatic Speech Recognition at the CHMI Management ER/APP for improving the controller productivity.

The definitions have been revised several times, consolidated and processed through continuous Data Set roll out process.

5.2.3 Recommendations on regulation and standardisation initiatives

From the work performed in the Solution a need of standardization has raised. In particular a standardization of ASR input and output content as well as format would be very useful in order to improve system interoperability and comparability. This concerns speech-to-text with a number of word sequence hypotheses, text-to-concept based on the ontology for ATC utterances, and preparations in order to feed succeeding applications such as command error detection, plus formats such as JSON for content transmission, and many aspects more to enable comparability and interoperability. This can be an important topic for future research and collaboration among parties.

6 References

6.1 Applicable Documents

This TVALR complies with the requirements set out in the following documents:

Content Integration

[1] EATMA guidance material and report, 16 December 2019, Ed. 01.00.01

Content Development

[2] SESAR project handbook, 27 April 2017, Ed. 01.00.01

Performance Management

[3] SESAR Performance Framework Ed 01.00.01-2019

Validation

[4] SESAR Project Handbook, (Programme Execution Guidance), 18 December 2018, Ed. 02.00.00

[5] Introduction to SESAR Maturity Criteria Ed. 01.01.03, 05 October 18System Engineering

Safety

[6] SESAR Safety Guidance Reference Material, Ed. 04.00.01

Human Performance

[7] SESAR Human Performance Guidance Reference Material, Ed. 04.00.01

Security

[8] SESAR SecRAM Security Risk Assessment methodology for SESAR 2020, 25 September 17, Ed. 02.00.00

Programme Management

[9] SESAR 2020 PJ05-W2_D1.1_PMP Ed. 00.01, 06 January 2020

[10] SESAR 2020 PJ05-W2_874470_Annex1-DoW-PartB Ed. 1.12

6.2 Reference Documents

[11]ED-78A GUIDELINES FOR APPROVAL OF THE PROVISION AND USE OF AIR TRAFFIC SERVICES SUPPORTED BY DATA COMMUNICATIONS.[1]

[12] D4.1.040-PJ.10-W2-96 ASR-TRL6 Final TVALP - Part I_00.05.00

[13] D4.1.020-PJ.10-W2-96 ASR-TRL6 Final TS/IRS

[14]SESAR D4.1.070-PJ.10-W2-96 ASR –TRL6 EXE-03 Availability Note Ed00.01.00

[15]SESAR D4.1.060-PJ.10-W2-96 ASR-TRL6 EXE-02 Availability Note Ed00.02.00

[16]SESAR D4.1.050-PJ.10-W2-96 ASR-TRL6 EXE-01 Availability Note Ed00.01.00

[17]SESAR D4.1.090-PJ.10-W2-96 ASR-TRL6 EXE-01 Availability Note Ed00.01.00

[18]Ontology for transcription of ATC Speech Commands of SESAR 2020 solution PJ.16-04. Sept 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)

[19]EUROCAE ED-137B Interoperability Standards for VoIP ATM Components

1

Appendix A Technological Validation Exercise #01 Report

A.1 Summary of the Technological Validation Exercise #01 Plan

The present section is a report of the technological validation exercise EXE-PJ.10-96-ASR-TRL6-01 run by Leonardo, on its Rome premises, situated on Via Tiburtina. Test activities ran as detailed in the Technical Validation Plan document.

A.2 Technological Validation Exercise #01 description and scope

Leonardo validation Exercise demonstrated the benefits of introducing an automatic speech recognition system in Air Traffic Management to support a set of ATCOs routine tasks in the En-Route environment.

The ASR system (ASR4ATC) was integrated in Leonardo Lead In Sky CWP, interacting with the system to support and improve the efficiency of ATCOs' control tasks by means of prefilling a set of appropriate system masks which otherwise ATCOs would be filling in manually ("speech-to-text").

Such a module makes use of artificial intelligence techniques and heuristics for recognition of word patterns in recorded speech, as well as machine learning techniques in the implementation of the speech to text model, based on a predefined training set. Sequences of words were transcribed into sequences of ATC concepts ("text-to-concepts") according to a defined ontology.

ASR used the contents of ATCO R/T verbal communication (as well as stand-alone verbal commands) to update the information concerning individual radar tracks and/or command masks, both RT and datalink asking ATCOs to approve it once it is prefilled. To this purpose, a set of valid ATCO instructions and routine tasks were identified, along with their corresponding HMI masks/features.

Simulations were run in Rome, at Leonardo Tiburtina site, mid of May 2022 (17 - 20 May). Validation took place in the shape of a human-in-the-loop real-time simulation in En-Route environments, simulating scenarios at Sofia ACC. Leonardo Lead In Sky Controller Working Position was used. The remaining SW elements of the platform all belong to the Lead In Sky suite. Tests were run simulating traffic in and around Sofia ACC, with the support of six ATCOs all working in turn as En-Route specialists. Two pseudo-pilots were also present, utilizing a proprietary simulation tool, able to run adaptively scenarios generating radar tracks on the fly. There were two main different scenarios, simulating overfly air traffic in Sofia, with En-Route operations until the transfer of control. Sofia can be considered to be a medium complexity ACC.

An initial objective of the exercise was to compare performance for a reference scenario during normal operation and with the aid of ASR4ATC, the ASR module allowing controllers to issue commands with their voices. Simulations mainly consisted of issuance by ATCOs of ATC commands, such as climb, descend, turn right, turn left, turn left heading, turn right heading, reduce speed, increase speed and so on. ASR also supported ATCO situational awareness and monitoring thanks to its "HOOK" function allowing to identify the callsign of a certain a/c in its own sector, highlighting its track label.

The applicable use cases were three (see Technical Validation Plan for reference):

- UC.2 **Highlight of callsigns on the CWP from controller's utterances**. The highlight of the callsign coming from a controller utterance is expected to support the ATCO Situation awareness in the tactical management phase (e.g. a/c inbound to the sector or during inter sector coordination) or in case of a/c requesting actions (e.g. deviation, flight level change or request information).

- UC.4 **Prefilling of commands in the CWP.** In this use case ATCOs will be presented with the recognized (and validated) command types together with the command values in the CWP. ATCOs will then be able to accept/reject or manually correct the commands.

- UC.8 **Prefilling of Datalink commands.** After the Logon and Connecting procedure have been fulfilled, ATCOs will utter the datalink commands and the CWP will prefill the command values to be accepted/rejected or manually corrected by the ATCOs.

The exercise follows two complementary approaches:

The first approach seeks the operational feedback from controllers by means of a Real Time Simulation. ATCOs will control two En-Route sectors performing their task as usual with (solution scenario) and without (reference scenario) the automatic speech recognition system enabled. Pseudo pilots will manage flights and interact via voice with the controllers.

Debriefings, observations and ATCOs' feedback will be gathered by means of a dedicated set of questionnaires. A set of specific system data log will be also recorded to corroborate the qualitative data.

The second approach will analyse the ASR performance by means of post processing of the data recorded during the Real Time Simulation. Data logs collection and analysis is foreseen in order to evaluate the performance of the ASR module (e.g. command/callsign recognition rate, command /callsign error rate).

Intermediate steps have been scheduled to support the evolution of the concept and prototype. A dry-run with an early prototype was performed in February 2022 and the outcome has been used to improve the performance.

The exercise aims to reach a TRL-6 maturity as a pre-industrial prototype and will be integrated in an operational platform, including the operational communication system.

This exercise addresses several Validation Targets:

- Human Performance (Positive impact is expected on the human performance, in particular with potential benefit in terms of reduced workload due to the support of automatic voice recognition and increase of situational awareness.)
- Safety (positive impact on safety is expected, enhancing ATCOs' situational awareness)
- Controllers productivity (related to the workload reduction associated to the ASR support in early identification of aircraft, automatic filling of command masks and datalink command masks to perform routinely tasks)

A.2.1 Validation platform/tool & Validation technique

The validation platform consisted in two Leonardo Lead In Sky Working Positions plus another acting as a feeder, each running CentOs 7, and connected to the Lead In Sky infrastructure. One WP had been assigned to the SU sector, while the other to the SD one. Pseudo-pilots were using test track generators running, one on a Linux machine and the other on a Windows computer, injecting flight related data

EUROPEAN PARTNERSHIP

Co-funded by the European Union

into the systems, while being in a physically separated room and communicating with ATCOs over simulated R/T. The ASR4ATC module was installed and run on a different virtual machine, along with the Context Based Data server, also running on the same virtual machine. In order to simulate radio telephony, Mumble was run across WPs and Pseudo Pilot machines. Imtradex USB headsets were used for ATCOs and pseudo-pilots.

The room layout was as shown on figure below

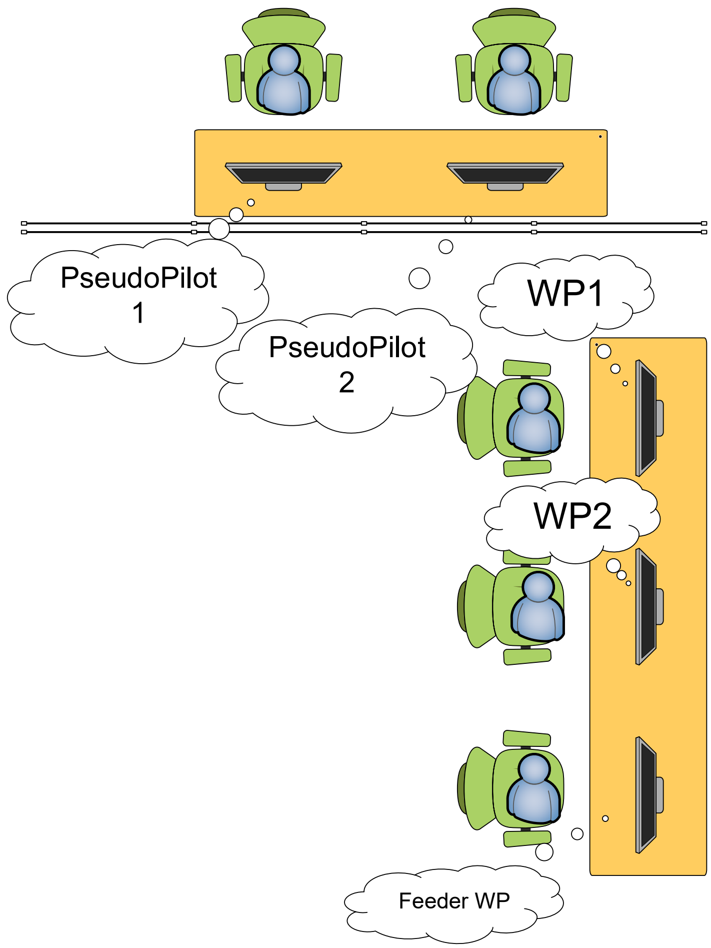

Figure A-1 Simulation room layout for EXE-01

Here in the following some pictures of the simulation room organization.

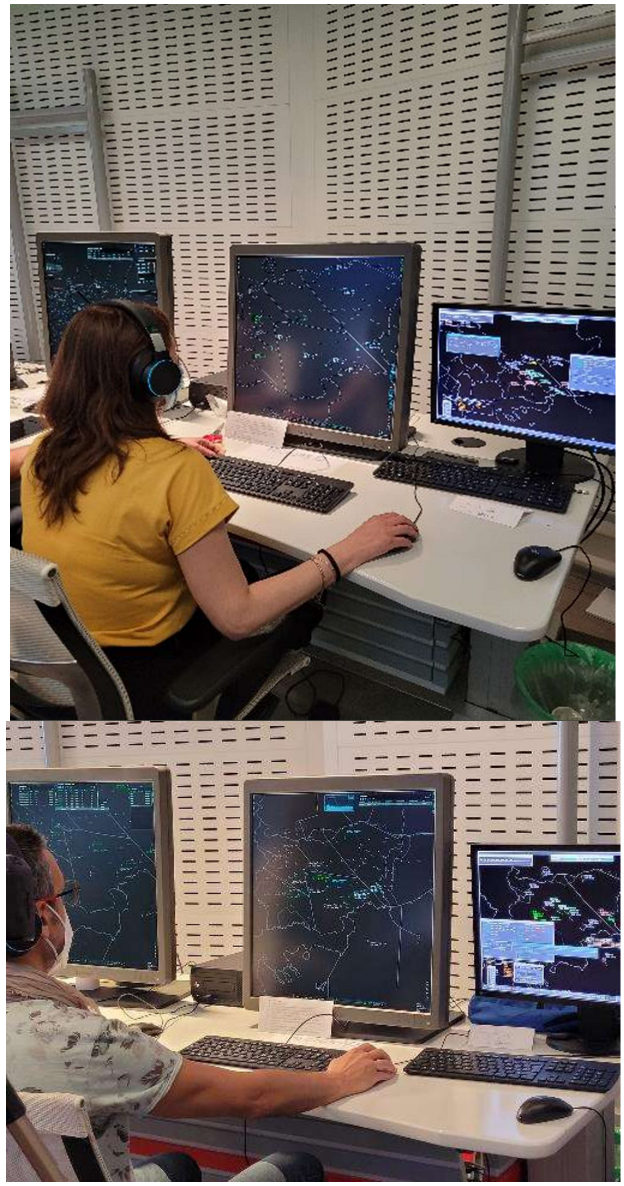

Figure A-2: ATCOs WPs for EXE-01

The simulation was run on the platform as shown in Figure A-3 below.

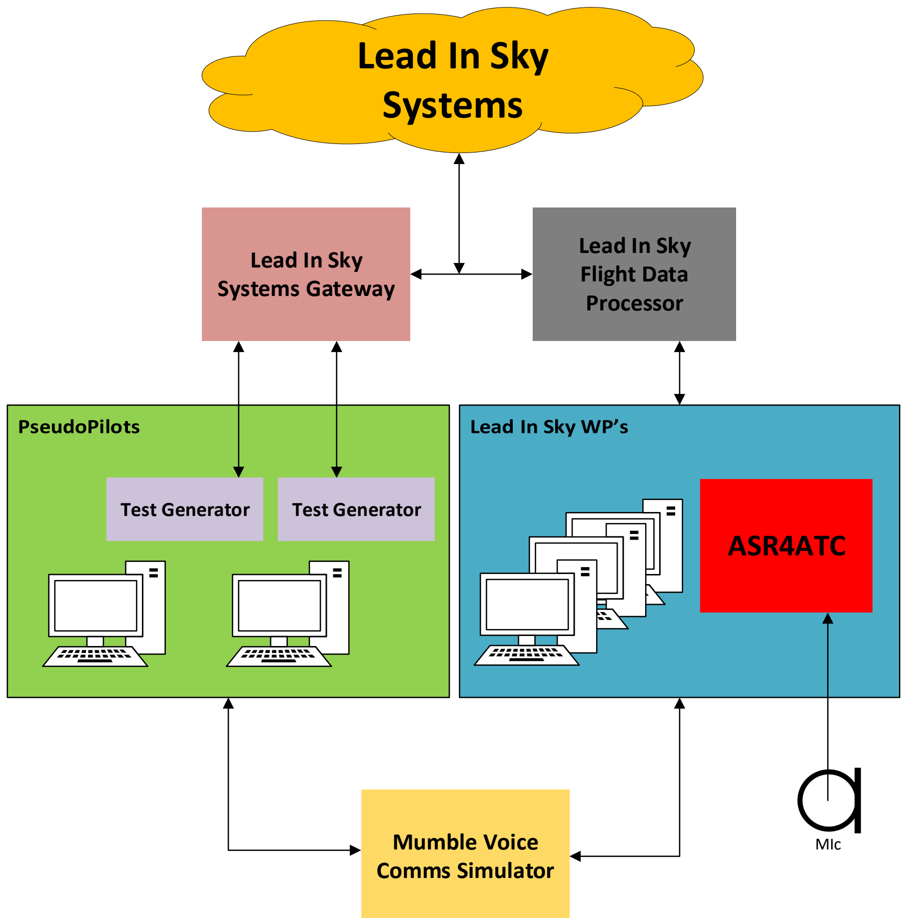

Figure A-3: Simulation platform for EXE-01

Simulations were run as Humans In The Loop real time simulations, with pseudo-pilots injecting data into the systems, and ATCOs interacting with WP, simulating operations at a medium scale, both in terms of complexity and in number of movements. The Sofia ACC qualifies as a medium complexity, in terms of all aspects pertaining the validation. Logging and metering features were available, to save all necessary information, and to process it in an automatic flow at a later time, for diagnostic and reporting purposes. All audio recordings were saved and archived for offline analysis and processing at a later date, if necessary.

1. ASR4ATC

In the picture below, a schematic diagram of how the ASR module works within the LDO platform, is given. It is a first attempt at designing an ASR module, using Kaldi as main building block. Speech corpora used were widely available ones, and no special agreement with ANSP or provider was possible given the Exercise and development timescales. No specific annotation task was performed, either. The ontology used was a subset, aptly reduced for the validation scope

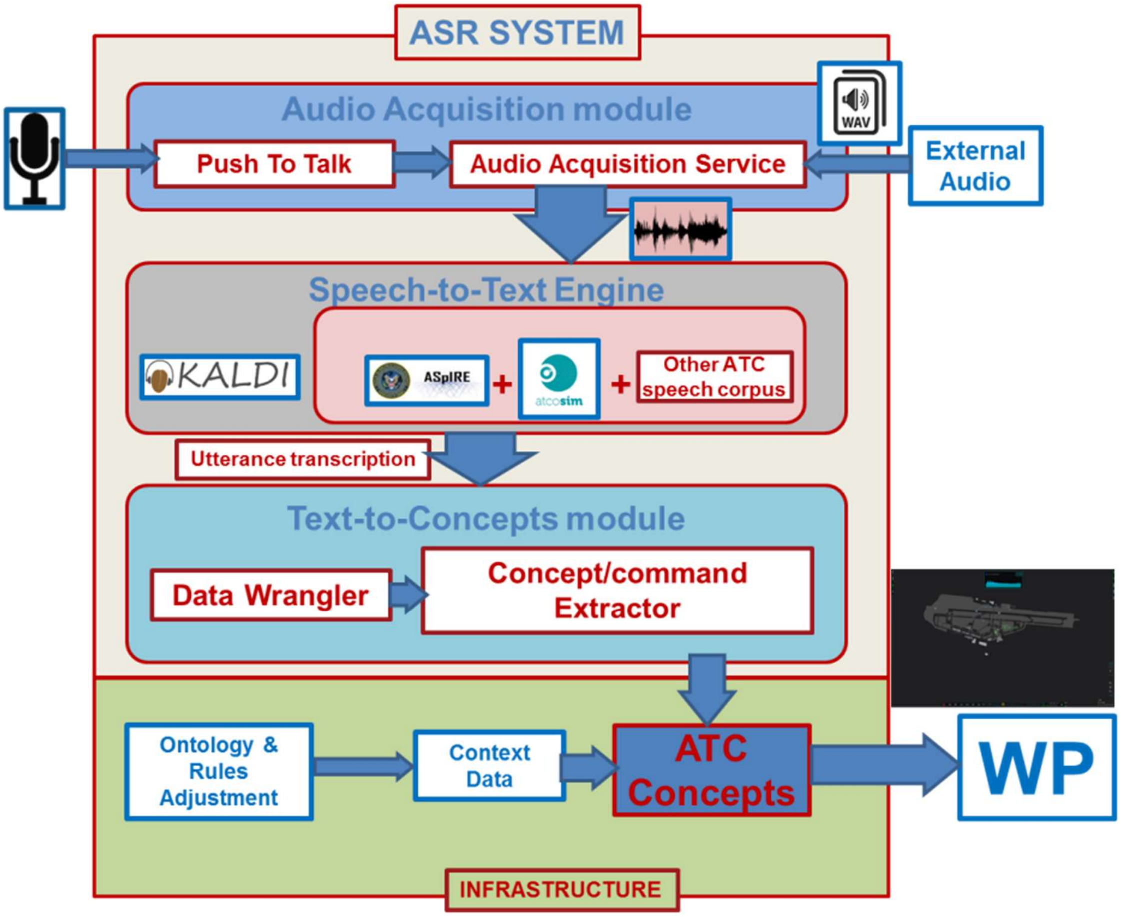

Figure A-4: A block diagram of the Leonardo ASR4ATC module, used for EXE-01

ASR4ATC is generated going through several compilation steps, and its end result is a docker, running independently of any other software, installed on a dedicated Virtual Machine also running CentOs 7, with ASR4ATC and with 8 CPU and 4GB of RAM. Preliminary tests did not indicate measurable improvement with an increase in either the number of CPUs or the amount of RAM.

2. Audio recording widget

When ASR is invoked, a widget is superimposed to the WP HMI, in order to provide ATCOs with graphical feedback, see Figure A-55 below. The widget alerts users that recording is taking place, giving also a graphical feedback on the audio content, with a small spectrum analyser showing roughly frequency content and sound levels. With a red round symbol it shows recording is taking place, and it is deactivated when the button or pedal is released.

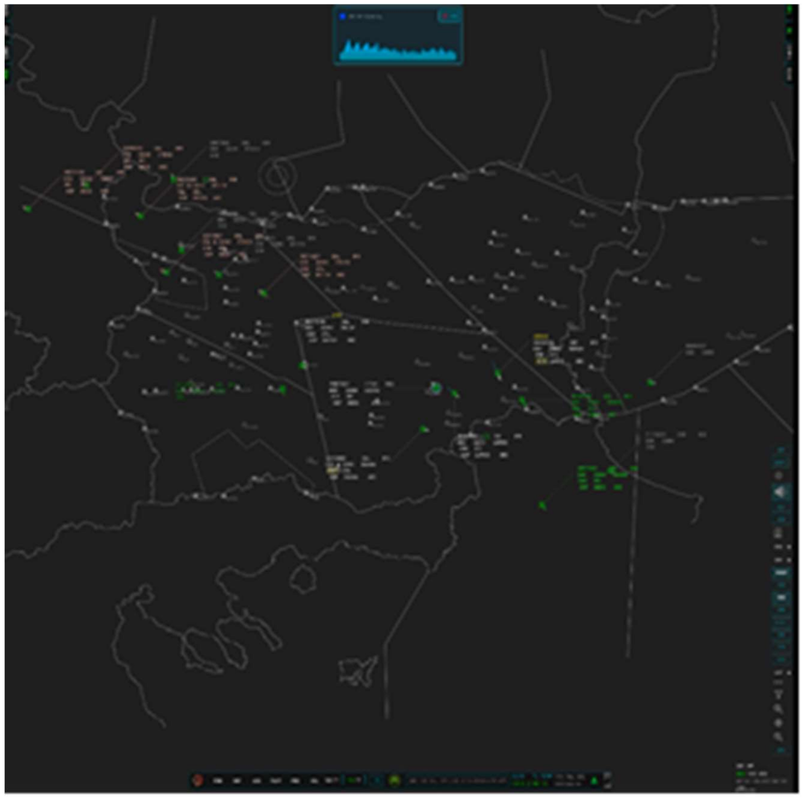

Figure A-5: A screenshot of WP showing the audio recording widget while recording

3. Proxies and Data Logging software

In order to handle data exchange with WP, with ASR4ATC and with the Lead In Sky Infrastructure, a proxy is configured, taking care of appropriate data forwarding and configuration. Even if the naming convention is not completely appropriate, proxies also look after data logging. As a result of simulation runs, results and diagnostics data were produced in abundance and subsequently required manual browsing and, not only for annotation, but also for recollection of transcriptions, instructions and measurements.

4. Context-based data generator

Context-based data are crucial in order to improve ASR performance. In the EXE-01 platform, context data were only a list of applicable call signs, which was updated every minute. The Context-based Data Generator (CDG) runs alongside ASR4ATC and does a simple job of extracting from the FDP DB a list of call signs of searchable flights in permitted states. Then the text list is forwarded to ASR4ATC, helping greatly callsign recognition.

5. GTG custom Leonardo simulator

In order to accomplish two results, designing simulation scenarios and running them, a custom tool designed for internal Leonardo use was utilized. It is a web based application, therefore it can be run on any computer running a recent browser. The tool is used to design scenarios given an underlying map structure, and to generate interactively tracks, running scenarios in various ATC environments.

GTG was also used by pseudo-pilots to run the actual scenarios and to make instant changes to them, if and when needed.

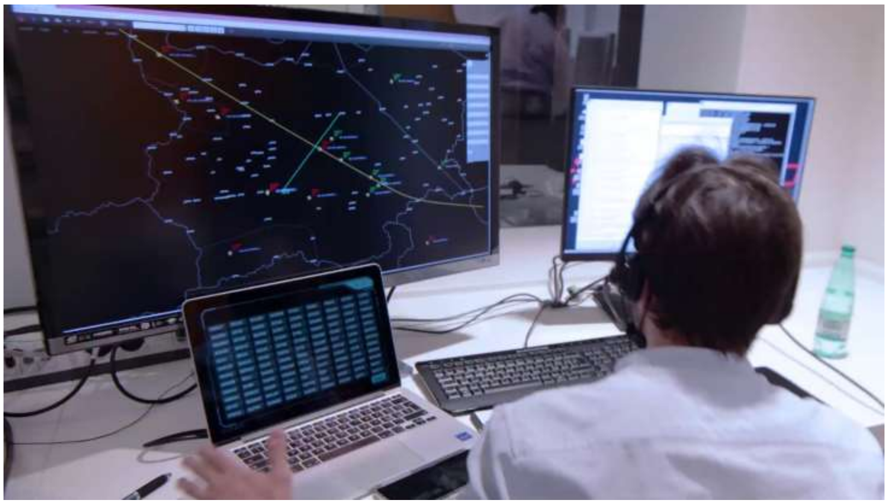

Figure A-6: Pseudo-pilot in a separate room running a GTG scenario

The designed scenarios could be reused in different environments and to run different tests, always situated in the Sofia ACC.

6. ACC Layout

The Sofia ACC, as it shown from the picture below, is a medium complexity En-Route volume. Two sectors were taken into account, the south east SD and the north west SA.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

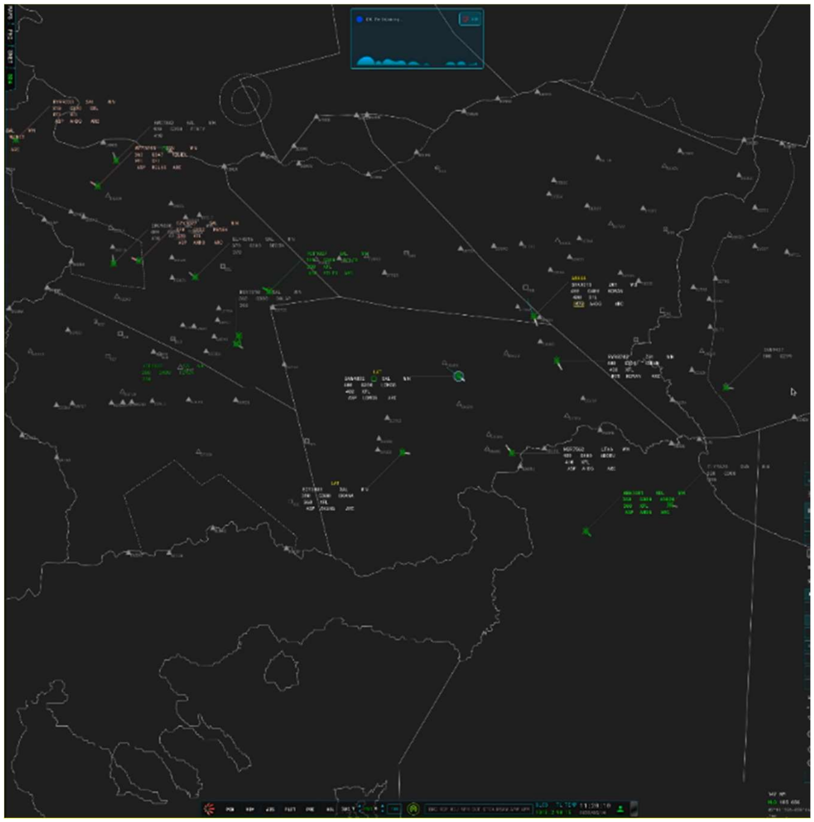

Figure A-7: Sofia ACC operational view via Leonardo WP

7. Controller Working Position

The Controller Working Position utilized was a purpose custom version of the Leonardo Lead In Sky product; maps and geographical information were all based on Bulgaria and Sofia AIP.

8. Traffic main characteristics

The simulation traffic was designed having in mind reliable En-Route operation, without the need of high traffic situations, and for an expected duration of about 40 minutes per run. A/c were all commercial aviation, no military, General Aviation or VFR flights, Traffic was orchestrated by the two pseudo-pilots who in turn would establish radio contact with ATCOs in order to simulate R/T voice communication.

A.3 Summary of Exercise 1 Technological Validation Objectives and success criteria

The first four SESAR Solution Validation Objectives with their respective Success criteria for Solution 96 as defined in section 4.3.2 of the TVALP, are covered in Exercise 01 without modified Exercise Validation Objectives and/or Exercise Success Criteria.

Please refer to the TVALP section 4.3.2 for specific validation objectives and associated Success Criteria.

SESAR Solution Validation Objective	SESAR Solution Success criteria	Coverage and comments on the coverage of SESAR Solution Validation Objective in Exercise 01	Exercise Validation Objective	Exercise Success criteria
OBJ-Sol.96ASR-TRL6-TVALP-0010	CRT-Sol.96ASR-TRL6-TVALP-0010.001	Fully covered	EX01-OBJ-Sol.96ASR-TRL6-TVALP-0010 same description as OBJ-Sol.96ASR-TRL6-TVALP-0010 To assess the technical feasibility of the integration of the ASR system and its sub-systems into CWP and interoperability between the ASR sub-systems and the existing CWP systems and tools.	EX01-CRT-Sol.96ASR-TRL6-TVALP-0010.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0010.001 The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the existing CWP systems and tools. Availability of systems and tools and their performance remain at 100%
OBJ-Sol.96ASR-TRL6-TVALP-0020	CRT-Sol.96ASR-TRL6-TVALP-0020.001	Fully covered	EX01-OBJ-Sol.96ASR-TRL6-TVALP-0020 same description as OBJ-Sol.96ASR-TRL6-TVALP-0020 To assess the stability of the ASR system performance	EX01-CRT-Sol.96ASR-TRL6-TVALP-0020.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0020.001 The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)
	CRT-Sol.96ASR-TRL6-TVALP-0020.002	Fully covered		EX01-CRT-Sol.96ASR-TRL6-TVALP-0020.001 same description as

				CRT-Sol.96ASR-TRL6-TVALP-0020.002 The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.
OBJ-Sol.96ASR-TRL6-TVALP-0030	CRT-Sol.96ASR-TRL6-TVALP-0030.001	Fully covered	EX01-OBJ-Sol.96ASR-TRL6-TVALP-0030 same description as OBJ-Sol.96ASR-TRL6-TVALP-0030 To assess the impact on the human performance of the integration of the ASR system and its sub-systems into operations in a realistic environment	EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.001 The introduction of the ASR system into the context of application is operationally viable, ATCos workload with ASR is equal or better than in baseline (without ASR support)
	CRT-Sol.96ASR-TRL6-TVALP-0030.002	Fully covered		EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.002 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.002 The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate)
	CRT-Sol.96ASR-TRL6-TVALP-0030.003	Fully covered		EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.003 same description as

				CRT-Sol.96ASR-TRL6-TVALP-0030.003 The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. Controllers' feedback with respect to Human Factors questionnaire is better than for baseline)
	CRT-Sol.96ASR-TRL6-TVALP-0030.004	Fully covered		EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.004 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.004 Changes in the design of the user interface (input devices, visual displays/output devices, alarm& alerts) support ATCos in carrying out the tasks.
	CRT-Sol.96ASR-TRL6-TVALP-0030.005	Fully covered		EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.005 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.005 The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified)
OBJ-Sol.96ASR-TRL6-TVALP-0040	CRT-Sol.96ASR-TRL6-TVALP-0040.001	Fully covered	EX01-OBJ-Sol.96ASR-TRL6-TVALP-0040 same description as OBJ-	EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.001 same description as

			Sol.96ASR-TRL6-TVALP-0040 To assess the impact of the introduction of the ASR system on safety.	CRT-Sol.96ASR-TRL6-TVALP-0040.001 The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.
	CRT-Sol.96ASR-TRL6-TVALP-0040.002	Fully covered		EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.002 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.002 The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.
	CRT-Sol.96ASR-TRL6-TVALP-0040.003	Fully covered		EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.003 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.003 The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.
	CRT-Sol.96ASR-TRL6-TVALP-0040.004	Fully covered		EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.004 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.004 The level of ATCO's situational awareness

				is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions).
	CRT-Sol.96ASR-TRL6-TVALP-0040.005 The level of ATCos' workload is maintained or decreased with the introduction of ASR system.	Fully covered		EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.005 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.005 The level of ATCOs' workload is maintained or decreased with the introduction of ASR system.
OBJ-Sol.96ASR-TRL6-TVALP-0050 To assess the impact of the introduction of the ASR system on capacity.	CRT-Sol.96ASR-TRL6-TVALP-0050.001 The workload of ATCO after introduction of an ASR system is adequate to increase TMA capacity. The workload of ATCOs is the same or less when working with ASR compared to baseline. The average flight time of the aircraft is expected to be reduced with respect to baseline without ASR support due to less time needed by the ATCO to complete task for one aircraft. ATCO should then have more time available for other aircraft and more timely task execution with regard to the flight progressing through their airspace. This should	Not covered		

	result in more optimum trajectories.			
	CRT-Sol.96ASR-TRL6-TVALP-0050.002 ASR allows ATCOs to safely manage a higher amount of aircraft, increasing the throughput in TMA	Not covered		
OBJ-Sol.96ASR-TRL6-TVALP-0060 To assess the impact of the introduction of the ASR system on Fuel efficiency	CRT-Sol.96ASR-TRL6-TVALP-0060.001 Aircraft will be able to improve their route Efficiency (fuel burnt) due to the higher throughput in TMA thanks to the introduction of ASR	Not covered		
OBJ-Sol.96ASR-TRL6-TVALP-0070 To assess the impact of the introduction of the ASR in visualization navigation	CRT-Sol.96ASR-TRL6-TVALP-0070.001 ATCos are able to perform a faster and more predictable navigation when using ASR for 3D visualization	Not covered		

A.4 Summary of Technological Validation Exercise #01 Validation scenarios

Figure A-8 Sofia ACC Operations simulated on Leonardo WP with ASR

Two executive ATCOs per day participated in four runs lasting 40 minutes each. The first run was a reference scenario without ASR, the second and the third were solution scenarios with ASR and the last was a safety scenario with ASR where the ATCOs were asked to make mistakes in calling the callsign. Each Each scenario consisted of 20 overflies. The traffic sample was different in each scenario to minimize learning effects.

The overflying traffic structure has one main axe, north west/south east axe with traffic in both directions.

Real Time Simulation SOL. 96 - EXE. 001			
Time frame	18.05.2022	19.05.2022	20.05.2022
10:00 -10:05	Briefing	Briefing	Briefing
10:05 - 10:50	REF1_96	REF1_96	REF1_96
10:50 - 11:05	Questionnaire	Questionnaire	Questionnaire
11:05 - 11:20	Break	Break	Break
11:20 -11:25	Briefing	Briefing	Briefing
11:25-12:10	SOL1_96	SOL1_96	SOL1_96
12:10-12:30	Questionnaire & Debriefing	Questionnaire & Debriefing	Questionnaire & Debriefing
12:30-12:35	Briefing	Briefing	Briefing
12:35-13:20	SOL2_96	SOL2_96	SOL3_96
13:20-13:40	Questionnaire & Debriefing	Questionnaire & Debriefing	Questionnaire & Debriefing
13:40-14:45	Lunch	Lunch	Lunch
14:45-15:30	SOL3_96	SOL3_96	SOL3_96
15:30-16:05	Questionnaire & Final Debriefing	Questionnaire & Final Debriefing	Questionnaire & Final Debriefing

Figure A-9 Real Time Simulation Schedule

A.4.1 Reference Scenario

The reference scenarios will address the current En-Route operational environment of two Bulgarian ACC En-route sectors. Figure 7 and 8 present the sectors that will be simulated.

In such operational context, the controller issues a datalink and R/T command to the flight crews by R/T communications. The flight crew is expected to confirm the clearance by read-back, in an accurate and timely manner. As soon as he/she receives the readback, the ATCO manually updates the system (using the mouse) in order to input the command issued and align the CWP data.

The sectors will be controlled by two executive controllers in single operation position (without planning controller). A pseudo pilot position (PWP) will support the scenario execution.

The controller working position will have the ASR DISABLED during the reference scenario simulation.

This setting will be used as baseline against which the solution scenario- implementing the ASR module- will be compared.

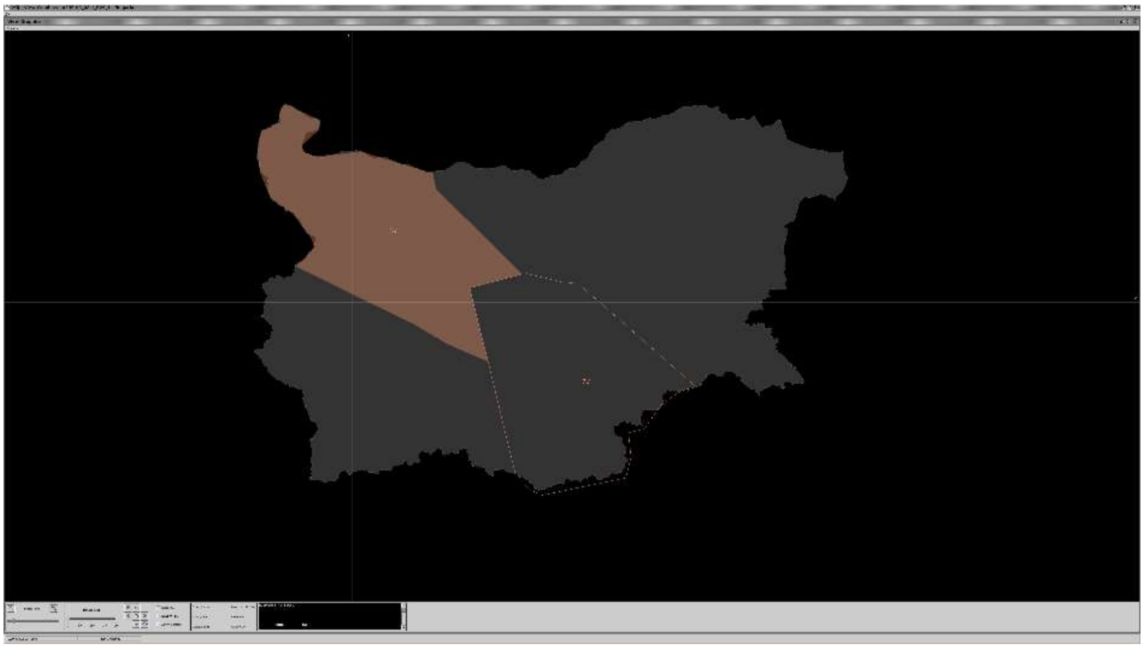

Figure A-10 EXE-001 simulated sector SA

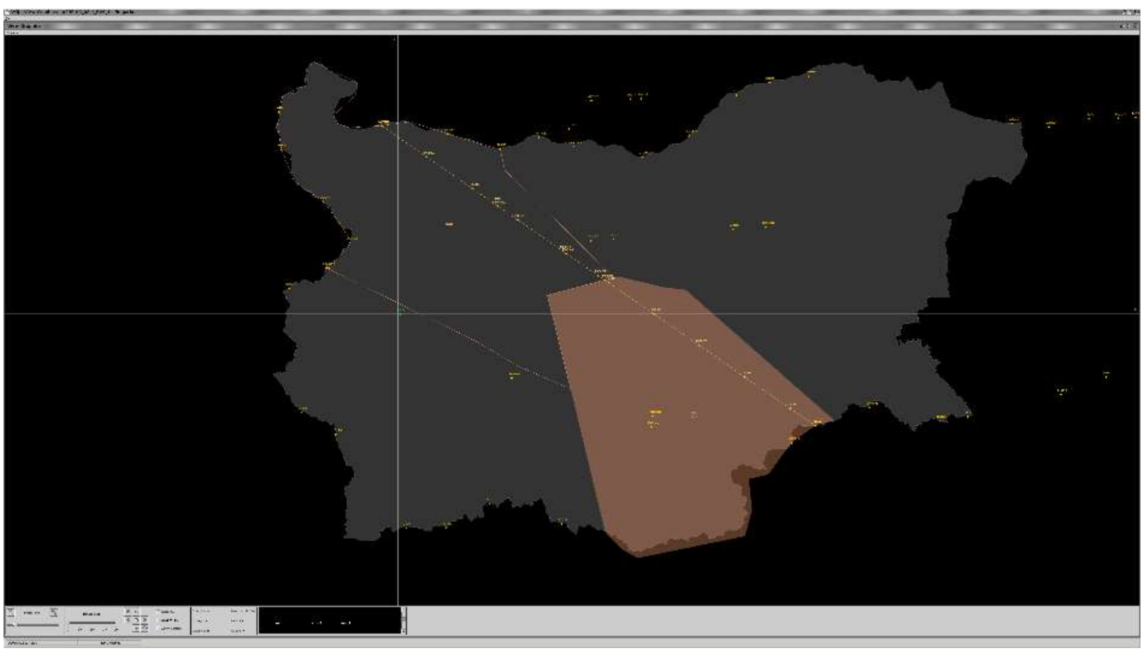

Figure A-11 EXE-001 simulated sector SD

A.4.2 Solution Scenario

The solution scenarios of the real time simulation are the two Bulgarian ACC En-Route sectors used for the reference scenario. The sectors will be controlled by two executive controllers in single configuration. A pseudo pilot will follow the instructions of the controllers and perform any necessary action to test the ASR system. These actions will also be performed in the reference scenario for comparison.

The solution scenarios' simulation addresses the experimental condition in which ASR support is enabled. The three use cases (highlight of flights, execution of datalink commands and execution of R/T commands) will be addressed.

The ASR system is intended to support and expedite the ATCO performing routinely tasks (e.g. updating the FL on the a/c label) by automatically recognizing a set of verbal clearances/values extracted from the verbal communication and by filling the appropriate command masks.

If the command mask generated by the ASR is correct, the ATCO acknowledges it with the mouse. The command is executed and the flight plan and radar label are then updated according to the ATCO's command.

If the ATCO verifies that the ASR output is incorrect, he/she has to correct manually the recognized command/values, or reject the command/values altogether.

The ASR will also extract and log relevant information from controller's utterances for further analysis and statistical investigation. In fact during the measured runs is foreseen the ASR data log recording in order to collect relevant information to analyse the ASR performance . The performance results will allow to derive the percentage of recognition and the error rate for callsigns and command types

Communications between controller and pilot will be performed using Mumble, the communication system of LDO and flight plan data will be sent to the ASR prototype and updated by the Lead in Sky platform.

6.2.1.1 Airspace Information

Bulgaria ACC is a class C airspace categorized as medium density/complexity environment.

The main communication means between Controllers and Flight Crew is Radio communication and datalink.

The scenario has a high percentage of the traffic evolving in the overflying routes mainly through the axis south east to north west and vice versa of the Bulgarian airspace.

6.2.1.2 Validation scenario

The configuration selected for this exercise is within Bulgaria ACC airspace, with two En-route sectors one on the south east side called SD and the other on the north west side called SA.

The overflying flights allow the execution of many commands through datalink, such as climb/descend to, turn right/left, maintain/increase/decrease speed etc.

The sectors selected gather two elementary volumes, SAT, SAU, SDT, SDU.

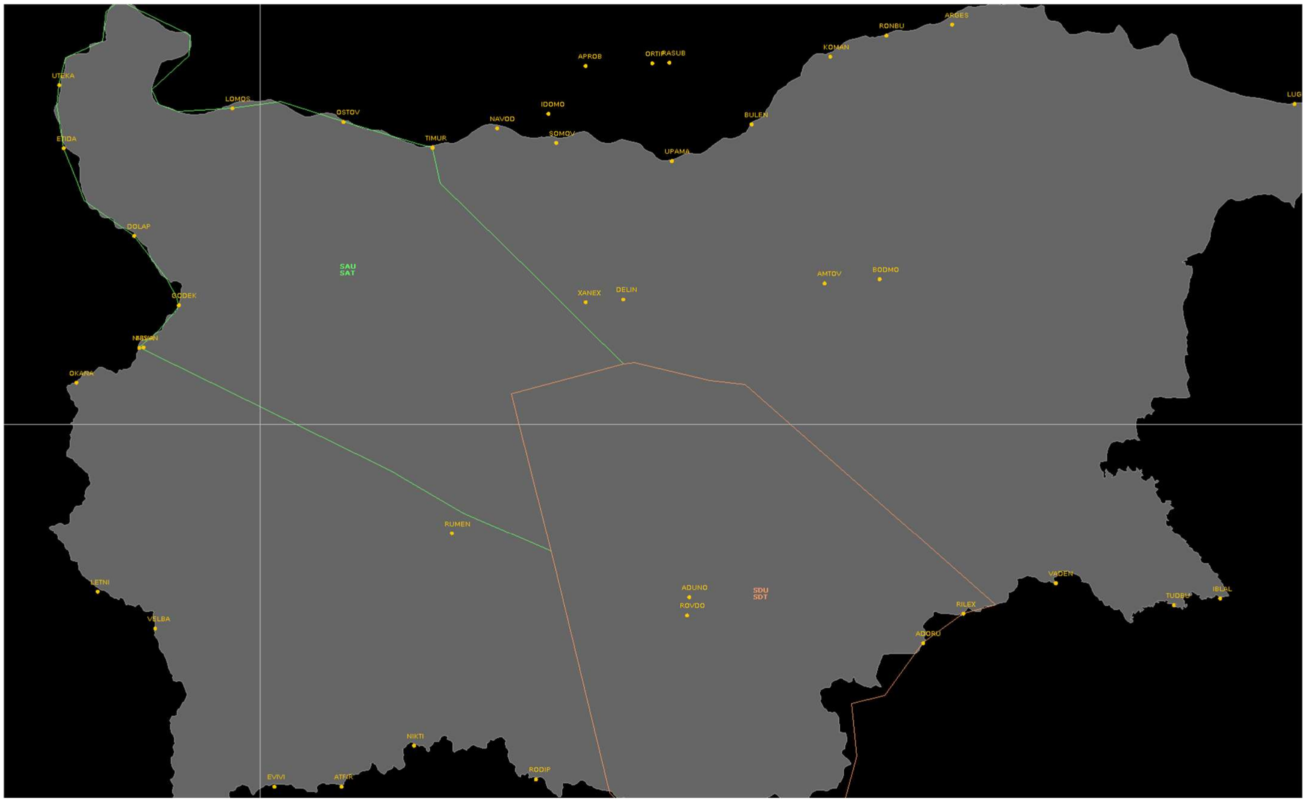

Figure A-12 Sofia ACC elementary sectors

A.5 Summary Technological Validation Exercise #01 Assumptions

Identifier	Title	Description	Justification	Impact on Assessment
AS-EX01-01	Datalink	Data link will be used during the exercise	The exercise is centred on the radio communication and the datalink between controllers and pilots	N/A
AS-EX01-02	Language	ATCOs and pilots will communicate in English	This language is operational in Bulgarian airspace	N/A

AS-EX01-03	Ambient noise	Noise generated by conversation in a control centre Room	Ambient noise will be limited to what can be generated by conversation of other ATCOs present in a control room. No simulation of aircraft generated noise is foreseen	Medium
AS-EX01-04	Weather	Normal/good	No impact analysis of abnormal weather conditions	Low
AS-EX01-5	Traffic conditions	Traffic conditions will be regular in terms of flow and amount	The test has to reflect a "normal" setting	Low

Table 15 Technological Validation Assumptions overview

A.6 Deviation from the planned activities

The following deviation from the initial exercise plan is to be reported:

- Manual correction rate - the data log related to this indicator are not available. ATCO were asked to report a qualitative estimation (in %) of the manual correction required during the validation exercise.
- In the post run questionnaire, one ATCO response is missing for each Safety related statement

One response on the SASHA questionnaire is missing, related to the statement 'I started to focus on a single problem or a specific area of the sector' on the SASHA questionnaire

A.7 Technological Validation Exercise #01 Validation Results

A.7.1 Summary of Technological Validation Exercise #01 Results

Exercise #01 Validation Objective	Exercise #01 Success Criterion	Exercise #01 Results	Exercise #01 Validation Objective Status
EX01-OBJ-Sol.96ASR-TRL6-TVALP-0010 same description as OBJ-Sol.96ASR-TRL6-TVALP-0010 To assess the technical feasibility of the integration of the ASR system and its sub-systems	EX01-CRT-Sol.96ASR-TRL6-TVALP-0010.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0010.001 The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and	A preindustrial ASR prototype with operational systems, including an operational LeadInSky CWP that provides context information in real time to the ASR (flight plan list in this approach), receives	OK

into CWP and interoperability between the ASR sub-systems and the existing CWP systems and tools.	availability of the existing CWP systems and tools. Availability of systems and tools and their performance remain at 100%	information from the ASR and presents it to the controller in a coherent manner with the rest of CWP information. The exercise also included connection with an operational voice communication system.	
EX01-OBJ-Sol.96ASR-TRL6-TVALP-0020 same description as OBJ-Sol.96ASR-TRL6-TVALP-0020 To assess the stability of the ASR system performance	EX01-CRT-Sol.96ASR-TRL6-TVALP-0020.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0020.001 The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)	The performance is maintained. Nevertheless, the command type recognition rate was different among command types and ATCOs varing from 94% to 99%. The callsign recognition rate from 66,5% to 87%.	OK
	EX01-CRT-Sol.96ASR-TRL6-TVALP-0020.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0020.002 The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.	Differences among different command types were greater than 2.5%, so the criteria was not satisfied	NOK
EX01-OBJ-Sol.96ASR-TRL6-TVALP-0030 same description as OBJ-Sol.96ASR-TRL6-TVALP-0030 To assess the impact on the human performance of the integration of the ASR system and its sub-systems into operations in a realistic environment	EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.001 The introduction of the ASR system into the context of application is operationally viable, ATCos workload with ASR is equal or better than in baseline (without ASR support)	83% ATCOs declare that ASR supports controllers in maintaining an acceptable level of workload.	OK
	EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.002 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.002 The accuracy of the information provided by the ASR system is	83% of ATCOs provided positive feedback on callsign recognition rate.	POK

	adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate)	33% of ATCOs provided positive feedback on callsign rejection rate. 17% of ATCOs provided positive feedback on command recognition rate. No positive feedback was provided on command rejection rate. 17% of ATCOs provided positive feedback on command manual correction required. 66% of ATCOs provided positive feedback on the frequency of wrong highlighted callsign. 17% of ATCOs provided positive feedback on the frequency of wrong ASR command. 44% of ATCOs rated the acceptance of the system above the acceptable minimum of 5. 17% of ATCOs agreed to the fact that job satisfaction increases when using ASR.	
	EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.003 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.003 The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. Controllers' feedback with respect to Human Factors questionnaire is better than for baseline)	17% of ATCOs agreed that the latency of the ASR feedback was acceptable. 17% of ATCOs agreed that the latency of the ASR command recognition was acceptable 66% of ATCOs were satisfied with the latency of the 'Hook' function.	POK

		50% of ATCOs found operating methods for ASR to be clear, complete and exhaustive.	
	EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.006 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.006 Changes in the design of the user interface (input devices, visual displays/output devices, alarm& alerts) support ATCos in carrying out the tasks.	70% of ATCOs were generally satisfied with the ASR interaction with CPDLC and no issues were reported. 70% of ATCOs disagreed that the level the feedback support provided by the ASR was adequate and clear and did not disturb them. 66% of ATCOs responded that the usability of the ASR system is 'Sufficient'.	POK
	EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.007 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.007 The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified)	33% of ATCOs evaluated their trust in the ASR system as 'Sufficient'.	NOK
EX01-OBJ-Sol.96ASR-TRL6-TVALP-0040 same description as OBJ-Sol.96ASR-TRL6-TVALP-0040 To assess the impact of the introduction of the ASR system on safety.	EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.001 The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.	70% of ATCOs disagreed that the level the feedback support provided by the ASR was adequate and clear and did not disturb them. 50% of ATCOs rated positively the level of safety during the ASR runs.	NOK
	EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.002 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.002 The timeliness of the information provided by the ASR system is	See results for EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.003.	POK

	adequate for the accomplishment of operations.		
	EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.003 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.003 The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.	33% of ATCOs agreed that ASR did not increase potential for human error compared to current operations.	NOK
	EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.004 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.004 The level of ATCO's situational awareness is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions).	50% of ATCOs responded that ASR supports ATCO in maintaining an adequate level of situation awareness. 100% of ATCOs agreed that the 'Hook' function improves situational awareness.	POK
	EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.005 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.005 The level of ATCOs' workload is maintained or decreased with the introduction of ASR system.	See results for EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.001.	OK

Table 16: Technological Validation Results Exercise 1

1. Results on technological feasibility

The validation activities showed the technological feasibility of introducing ASR4ATC tool as a new input mode to manage air traffic in En-Route scenario. In particular, ATCOs thought as effective usage of the Hook function (call sign highlighting), supported by context-based data (a list of currently applicable call signs gathered from the LIS FDP), which greatly improved performance and accuracy of the tool. ASR4ATC and its integration with WP and at large the LIS suite did not show blocking issues.

It is worth bearing in mind that:

a. ASR4ATC was developed in a very focused way, optimizing performance on a limited range of utterances/commands, which, accordingly, permit usage on a narrowed down number of use cases.

b. The phonetic model and its training are usually one of the main hurdles when developing such a tool. Using British English required ATCOs who speak English as a foreign language some degree of adaptation. However feedback was positive and performance was satisfactory. ATCOs pointed out some training was also necessary in order for them to adapt to the accent and pronunciation included in the model.

In the following a list of refinements and enhancements which can improve ASR4ATC technological feasibility is reported:

- Train the phonetic model to accept local English as a foreign language accents
- Allowing more than one command per utterance and widen the command choice, always based on the SESAR shared ATC Ontology
- Despite current response latency was considered acceptable by ATCOs, one significant step forward could be concept-by-concept recognition and transcription, also referred to as online transcription
- Making callsign range wider, including military, GA, more formats and airline operators

An assumption which was under inspection of participating ATCOs was the choice of a pedal in order to keep hands of controllers free: there is no generalized consensus since some Controllers found using a pedal beneficial while others did not and reverted to using the keyboard as normal.

Some commands such as "contact" were used by ATCOs and the resulting utterances have been removed from the stats, since they were not implemented in the WP instruction set or in the ASR module, as in one instance in which an incorrect radio name was associated to an ICAO tri-letter code (PEGASUS associated to PGT instead of the correct SUNTURK uttered by ATCOs).

2. Results per KPA

Method

In total, 6 professional ATCOs participated in the validation exercise.

For the purpose of the validation, each ATCO was assigned to both sectors changing at each run according to a predefined seating plan.

During the validation exercise, data were collected in the form of subjective qualitative assessment and objective quantitative measurement on the following aspects:

- ATCO Situational Awareness
- ATCO Workload
- ASR overall and ASR HOOK Function
- ATCO Acceptance & Job Satisfaction
- Trust in the system
- ASR Usability/ Ergonomics
- ASR Callsign & Command Recognition
- ASR Interaction with CPDLC
- Human Error

ASR recognition and rejection rates were measured by the validation platform and provided in the form of a data log. Remaining data were collected by means of subjective questionnaire provided at the end of runs (Post-Run Questionnaires) and at the end of the simulation participation (Post-

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Simulation/Exercise Questionnaires) and interviews during debriefing at the end of the run and at the end of the validation exercise. A training day was planned for all 6 ATCOs.

The table below includes the daily agenda for the entire duration of the simulation:

Real Time Simulation SOL. 96 - EXE. 001			
Time frame	**18.05.2022**	**19.05.2022**	**20.05.2022**
10:00 -10:05	Briefing	Briefing	Briefing
10:05 - 10:50	REF1_96	REF1_96	REF1_96
10:50 - 11:05	Questionnaire	Questionnaire	Questionnaire
11:05 - 11:20	Break	Break	Break
11:20 -11:25	Briefing	Briefing	Briefing
11:25-12:10	SOL1_96	SOL1_96	SOL1_96
12:10-12:30	Questionnaire & Debriefing	Questionnaire & Debriefing	Questionnaire & Debriefing
12:30-12:35	Briefing	Briefing	Briefing
12:35-13:20	SOL2_96	SOL2_96	SOL3_96
13:20-13:40	Questionnaire & Debriefing	Questionnaire & Debriefing	Questionnaire & Debriefing
13:40-14:45	Lunch	Lunch	Lunch
14:45-15:30	SOL3_96	SOL3_96	SOL3_96
15:30-16:05	Questionnaire & Final Debriefing	Questionnaire & Final Debriefing	Questionnaire & Final Debriefing

Table 17 RTS Agenda

As Table 17, above, shows, ATCOs participated in four different runs:

- REF scenario: baseline run with reference scenario (No ASR)
- SOL1: solution scenario, same as reference, with ASR usage
- SOL2: solution scenario, equivalent in terms of volume complexity and duration to the reference one, with ASR, different flights and trajectories
- SOL3: solution scenario, also equivalent in complexity terms to the reference, with ASR, different flights and trajectories.

The post-run questionnaire contained 6 questionnaires, including Bedford for workload, SASHA questionnaire for Situational Awareness, and CARS for user acceptance. Results analysis of this last

EUROPEAN PARTNERSHIP

Co-funded by the European Union

questionnaire were obtained by comparison of feedback regarding the solution scenario against comment regarding the reference scenario. The post-exercise questionnaire contained 35 questions, aimed at collecting the final ATCOs response about *usage* of ASR across all the validation scenarios.

Additionally, during the final debriefing, a Want/Have Matrix was used to collect data, as shown in the figure below. The purpose of the validation exercises is to proceed from TRL 2 to TRL 4 and the Want/Have Matrix was used to let ATCOs envision upcoming developments of the technologies applied to the exercise. They were questioned about what they liked about the system, what they did not like, what they would like to be added to the current concept and what they wish to avoid (even if not experimented) for the technology. The Want/Have matrix was judged as an appropriate tool fitting well with the level of maturity.

Table 18 WANT/HAVE MATRIX

Performance

The level of ASR performance was found to be acceptable in terms of callsign and command recognition rates. The "Hook" function was also reported to be effective and was very appreciated by ATCOs. Such a function's effectiveness depends heavily on the availability of context-based data, which proved to be highly beneficial, even during the initial stages of development. A limited ontology had to be adopted also to factor in the limitations introduced by the WP platform integration. The list of accepted commands was pared down to a minimum, in order to handle ground/tower movements.

Human Performance

- Situational Awareness

Although according to the SASHA questionnaire results, situational awareness was higher in the reference scenarios compared to the solution scenarios, in the post simulation questionnaire, the majority of ATCOs (83%) rated Situational awareness between 'Sufficient' and 'High' during the ASR exercise compared to the reference scenario.

Moreover, all ATCOs agreed that the ‘Hook’ function improves situational awareness with respect to the reference scenarios.

- Workload

The average level of workload reported for the solution scenarios (4.1 out of 10) was below the maximum tolerable WL level (6), however it was higher to the average workload level calculated for the reference scenario (2) which indicates that ATCOs perceived a higher level of workload during the scenarios with ASR. The scoring was motivated by the low system recognition rates and current ASR activation means which require mental effort.

- Usability

Results in the usability area indicate that ATCOs see a great benefit of using the ASR system, however at this stage of development, the poor recognition rates, the high latency, and errors of the system impact its usability. The recognised phraseology is to be enriched and more complex commands are to be integrated. There were no agreement of the ‘preferred’ ASR activation means however some valid suggestions were provided by the ATCOs to improve the overall usability of the system such as the implementation of a ‘diagnostics’ window or the integrations of various functions within the system e.g assuming or transferring traffic etc.

- Acceptance & job satisfaction

The average Acceptance resulting from the CARS questionnaire is below the acceptable minimum of 5 (4.8). Negative results are also observed for the overall acceptance investigated in the post simulation questionnaire, with nearly 70% of ATCOs ratings ranging from ‘Slightly Low’ to ‘Low’. Job satisfaction also received negative feedback from 50% of the ATCOs.

However, when asked about the callsign recognition rate, most ATCOs (83%) provided positive feedback.

The results that at this stage of development suggest that controllers are not entirely satisfied with the systems and that further improvements need to be envisaged.

- Trust

According to the SATI questionnaire, ATCOs trust in the system is rather low, from various perspectives: reliability, accuracy, usefulness, understandability and robustness. Moreover, the average confidence in working with the system is also low.

The overall trust in the system in the post simulation questionnaire was mostly reported by ATCOs between ‘sufficient’ and ‘low’, which might be due to the low recognition rates and latency of the system.

Want-Have Matrix

REMOVE (have but don’t want)	**PRESERVE (have and want)**
• The latency (4) • ASR errors	• The ‘hook’ function (5) • Integration with CPDLC

• Activation pedal • ASR manual activation (2) • The fact that the choice window does not appear close to the aircraft track • Manual command acknowledges • Partial recognition function	• The current recognised commands (2) • Separate activation means between ASR and radio frequency • Modification of ASR issued command at any time
AVOID (don't have and don't want) • Poor callsign recognition • Overlap of ASR and the frequency • Automatic command implementation	**ACHIEVE (don't have but want)** • Lower latency • Diagnostic window • Alert/alarm in case of ASR malfunction • Complex commands and instructions (4) • Readback/hear back check for both ATCOs and pilots (2) • More obvious highlight of the 'hooked' aircraft • Higher recognition rate • Integrate other different 'hook' sub-functions • ASR disable function • Integrate an 'Assume' and a 'Transfer' function (2)

Table 19 WANT/HAVE MATRIX RESULTS

Numbers shown in brackets indicate the number of ATCOs who repeated the comment.

- Preserve

Most ATCOs (5) were positive about the 'Hook' function and was said to be very useful for determining a certain a/c's position and to maintain situational awareness. Some positive comments were also received regarding command recognition which ATCOs specifically mentioned in the 'preserve' category of the matrix.

One ATCO particularly liked the idea of integrating ASR with CPDLC while another one mentioned that the separate activation means between ASR and radio frequency are very useful.

ATCOs would also like to maintain the possibility to modify an ASR issued command at any time, which would ensure that any potential errors could be rectified once identified, which ensure that any potential errors can be rectified once identified.

- Achieve

In terms of ASR improvements, most ATCOs wrote down that it would be useful for ASR to recognise more complex commands and instructions, or for example to integrate into the ontology a command for frequency change, or one for the correction of a given command.

Some ATCOs also mentioned that integrating the ‘assume’ and ‘transfer’ functions within ASR, which would assume traffic in the area of responsibility or transfer traffic to another area would be very useful.

An alert or alarm in case of ASR malfunction would also be useful according to ATCOs, together with an ASR ‘diagnostic window’, which would display useful information on request, situated always in the same place, similar to a chat window.

A more obvious highlight of the ‘hooked’ aircraft on the HMI was also mentioned, as the current indication was said to not be visible enough. It was suggested that the aircraft label is also highlighted when the ‘hook’ function is activated.

Improvements were also suggested in terms of the latency of the tool and its recognition rate.

One of the ATCOs suggested that ASR could be activated by the ‘push to talk’ button instead of using the keys or the pedal, while another controller would like to use ASR and R/T communication simultaneously. A better integration with A-SMGCS was also suggested, in which ASR would recognise and display the taxi route given to an a/c by an ATCO. Two ATCOs commented that ASR could be further integrated with other functions on the ASMGC-S, for example by displaying a runway as ‘occupied’ when recognising that a vehicle using it is in contact with the Tower or by highlighting a closed taxiway on the HMI.

An ASR ‘disable’ function was also suggested, that allows the ATCO to enable or disable ASR in case it causes issues.

Another suggestion from ATCOs was to integrate another function into the ASR system, that of recognising ATCOs and pilots’ communication simultaneously and double checking the information and possibly identifies potential mismatches a safety barrier.

Moreover, the ‘Hook’ function was said to have great potential and could be further improved to include other sub functions for example zooming out on an aircraft that calls from outside the ATCOs controlled airspace. It could also identify an aircraft through mode S, even if the FDP data is not present or simply highlight an aircraft when the ATCOs communicate with it.

- Remove

Most of the ATCOs (4) indicated that the latency of ASR is currently high, and this is concerning also in view of integrating ASR with CPDLC, because the latter system already has its own latency.

A couple of comments were also received regarding ASR manual activation by keys, which was found burdensome, and suggested that ASR could be activated by voice to avoid higher workload levels. However, some ATCOs expressed their concerns regarding automatic activation, as ASR could pick on some of ATCOs’ ‘offline’ coordination and internal discussions and erroneously provide commands based on that.

Another ATCO suggested to only remove the activation pedal which was found impractical and outdated for one of the ATCO, but to keep the manual activation of ASR via a key or a button.

The ASR error rate was also mentioned in the 'Remove section'. ATCOs were of the opinion that at this stage, the tool is not completely reliable as a result of the high error rate.

Another ATCO suggested that the fact that the choice window does not appear close to the aircraft track be confusing and counterproductive. The manual command acknowledge was also said to cause additional unnecessary workload by some ATCOS, who suggested that a 'timeout' functionality could be implemented, which would allow for an automatic ASR implementation of a given command, if during a set period, the ATCO has not corrected it. Other ATCOs were happy about acknowledging a command as they consider it need to be the controller's final responsibility.

Lastly, one ATCO mentioned that although the partial recognition of the tool is useful for ATCOs for identifying callsigns, in the case of flight level or heading instructions, it could send incorrect or undesired information to aircraft, which could impact safety.

- Avoid

In terms of functions that are not currently present with ASR and should be avoided, ATCOs mentioned the overlap of ASR and the frequency and poor callsign recognition especially that there are situations when aircrafts with similar callsigns fly in the airspace, and a wrong callsign recognition could have an impact on safety.

ATCOs suggest that for future development, it should also be avoided to overlap the ASR system and the radio frequency especially if the aircraft to which the ATCO intends to communicate is using CPDLC.

Safety

Considering the results on Human Performance, the perceived potential for Human Error did not clearly increase for controllers when using ASR. The system did not clearly impact the perceived potential for Human Error; however, safety concerns have been raised during the debriefing in relation to the system's latency and error rates at this level of development.

The overall safety level was perceived by half of the controllers the same as in today's operations.

A.7.2 Analysis of Exercise 1 Results per Technological Validation objective

1. OBJ-Sol.96 ASR-TRL6-TVALP-0010 Results

Objective description: *To assess the technical feasibility of the integration of the ASR system and its sub-systems into CWP and interoperability between the ASR sub-systems and the existing CWP systems and tools.*

Validation Objective ID	Success Criterion ID	Success Criterion Status	Validation Objective Status
EX03-OBJ-Sol.96ASR-TRL6-TVALP-0010	EX03-CRT-Sol.96ASR-TRL6-TVALP-0010.001	OK	OK

The ASR was successfully integrated with operational CWP and communication system. The ASR did not impact the performance of previous systems

EXE-01 connected a preindustrial ASR prototype with operational systems, including an operational LeadInSky CWP that provides context information in real time to the ASR (flight plan list in this approach), receives information from the ASR and presents it to the controller in a coherent manner with the rest of CWP information. The exercise also included connection with an operational voice communication system.

There was no impact on other systems or tools of the CWP.

All the controllers agreed that the ASR-system does not interfere with the availability and/or reliability (of other systems and components installed on the CWP).

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0010.001	The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the existing CWP systems and tools. Availability of systems and tools and their performance remain at 100%.	OK

2. OBJ-Sol.96ASR-TRL6-TVALP-0020 Results

Objective description: To assess the stability of the ASR system performance.

Success Criterion ID	Success Criterion	Success Criterion Status	Validation Objective Status
EX03-OBJ-Sol.96ASR-TRL6-TVALP-0020	EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001	POK	NOK
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.002	NOK	

Success Criteria:

CRT-Sol.96ASR-TRL6-TVALP-0020.001 The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.).

CRT-Sol.96ASR-TRL6-TVALP-0020.002 The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.

A dedicated data analysis has been run on the ASR data collected from the data logs and audio recordings taking into account callsign recognition, utterance recognition and event annotation.

The statistical analysis was obtained by transcribing manually the recordings, creating the callsign and event annotation standard, and then comparing it against ASR outcome.

The table below presents the total number of command issued for each one of the ATCOs involved in the simulation and the recognition rates for callsigns and for command types and the two combined.

Table 20 Callsign and Command type recognition rates

No callsign was wrongly recognized as only complete callsigns were detected. Feedback from

ATCO	COMMANDS ISSUED	CALLSIGN RECOGNITION RATE	COMMAND TYPE RECOGNITION RATE	CALLSIGN + COMMAND TYPE RECOGNITION RATE
ATCO 1	98	77%	99,5%	69,5%
ATCO 2	86	79%	94,5%	78,5%
ATCO 3	64	77%	99%	80%
ATCO 4	83	80%	95,5%	79%
ATCO 5	78	66,5%	97%	66%
ATCO 6	175	87%	98%	87%

controller indicated that they would like to have higher recognition rates. The allowed error is something to be investigated.

It catches the eye the fact that for one ATCO in particular the recognition was lower than the average for the other ATCOs. This can be due to his accent or lower attitude towards adapting to the ASR tool.

Table 46The table below presents the command types taken into account in the validation. There are differences in the recognition rate between different command types. This suggests how the ASR tool has to be trained to increase performance.

COMMAND (all 9 solution runs)	COMMANDS ISSUED	RECOGNISED COMMANDS	COMMAND RECOGNITION RATE
Climb to	122	33	27,04%

Continue present heading	44	14	78%
Descend to	48	30	61%
Hook	124	82	66%
Increase speed	21	13	62%
Maintain speed	5	4	80%
Maintain	11	8	72,7%
Proceed direct to	130	67	51,5%
Reduce speed	11	5	45%
Resume normal speed	10	9	90%
Turn left	15	9	60%
Turn left heading	8	5	63%
Turn right	17	12	71%
Turn right heading	18	11	61%

The climb to command has a not so high recognition rate due to the fact that often the qualifier, i.e. the FL was not recognized or only partially recognized.

According to the TS/IRS the recognition rates should be:

- The Command Recognition Error Rate of ASR should be less than 2.5% for ATCos.
- The Command Recognition Rate of ASR of ATCos should be higher than 85%.

The TS/IRS criterion is met for some ATCOs. Command Recognition rates increase if we don't consider the qualifier, i.e FL, Knots etc values.

The average percentages of recognition rate do not meet the criterion for callsign and command, but it does for command type recognition only, i.e. turn left, descend to and so on without value.

- Callsign Recognition Rate: 77.75%
- Command Type Recognition Rate: 97.25%
- Command Recognition Rate: 76.66%

Success Criterion ID	Success Criterion	Success Criterion Status

EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001	The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)	**POK**

The requirement regarding difference between commands is not meet as differences are greater than 2.5%.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.002	The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises	**NOK**

3. OBJ-Sol.96ASR-TRL6-TVALP-0030 Results

Objective description: To assess the impact on the human performance of the integration of the ASR system and its sub-systems into operations in a realistic environment.

Success Criteria:

EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.001

The introduction of the ASR system into the context of application is operationally viable, ATCOs workload with ASR is equal or better than in baseline (without ASR support).

In the post-run questionnaire, ATCOs were asked to evaluate the level of workload experienced during the run, on a (Bedford) scale from 1 to 10. For the solution scenarios, no significant difference in the average workload level is observed between the three scenarios (4.5 in Sol.1 as compared to 3.7 in Sol. 2 and 4.0 in Sol.3), all values being below the maximum tolerable WL level (6). Furthermore, the values of the standard deviation calculated for these scenarios (2.0 for Sol.1, 1.2 for Sol2 and 2.2 for Sol 3) are not significantly different either. As a result, the three solution scenarios can be considered homogeneous, with no significant peaks in workload.

However, the average Workload for the Reference scenario was estimated at 2.0, with a lower standard deviation compared to the other scenarios' (0.6). A slightly higher average WL is observed for the solution scenarios (4.1) but still in the acceptable limits. During the post run debriefings, ATCOs motivated the scoring with the fact that the limited number of utterances implemented in ASR in this phase, impacted the pacing of ATCO activity which requires a lot of concentration and mental effort. Another reason was the use of both the ASR and the frequency, which requires additional manual skill required to the ATCO to manage ASR (RT + ASR pedal/ keybord activation) .Moreover, the use of both ASR and frequency also required manual handling of the keyboard and the pedal which also adds to the workload.

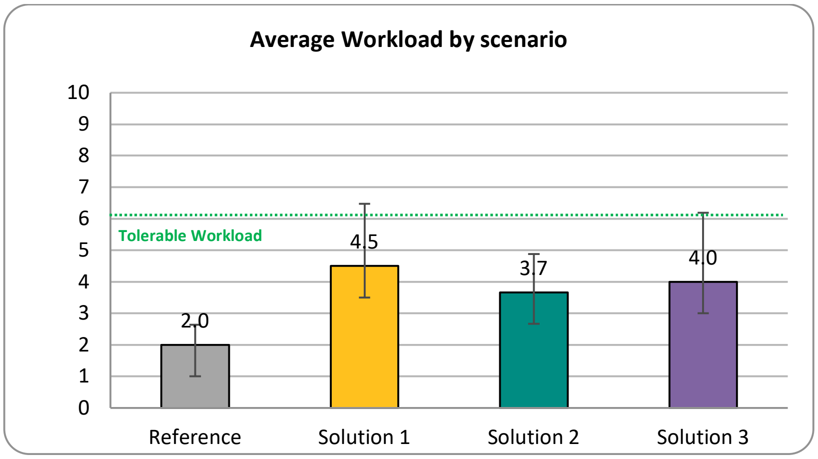

Figure A-13 - Post-run questionnaire – average workload by reference and each solution scenario

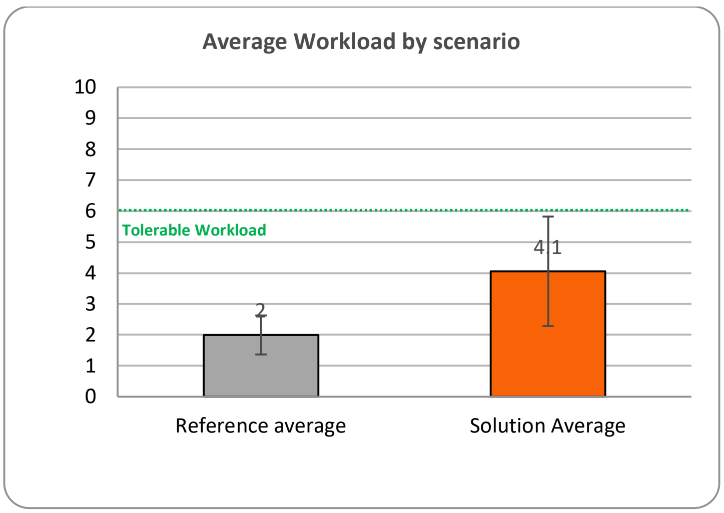

Figure A-14 – Post-run questionnaire – average workload by reference scenarios vs. solution scenarios together

For the solution scenarios, most ATCO responses (90%) indicated an acceptable level of WL, below the maximum tolerable WL level (5), while 11% of responses were below the threshold (8). Figure 8 shows the distribution of answers (on the 10-point Bedford scale), expressed also in percentage. Moreover, in the comments section of the Post Run questionnaire on workload, ATCO commented that training on the system will help with reducing mental workload significantly.

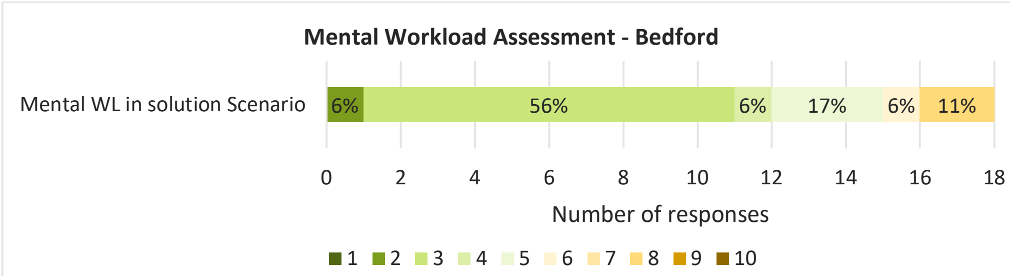

Figure A-15 - Number of responses for Mental Workload in the Solution scenario

Figure 9 below shows the average mental workload for solution scenarios according to controller's assigned sector. All mean values are below the maximum tolerable workload level (5) and there is no significant difference between the sectors (3.78 for the SA sector and 4.33 for the SD sector). This indicates that the two sectors are homogenous and don't generate different Workload levels.

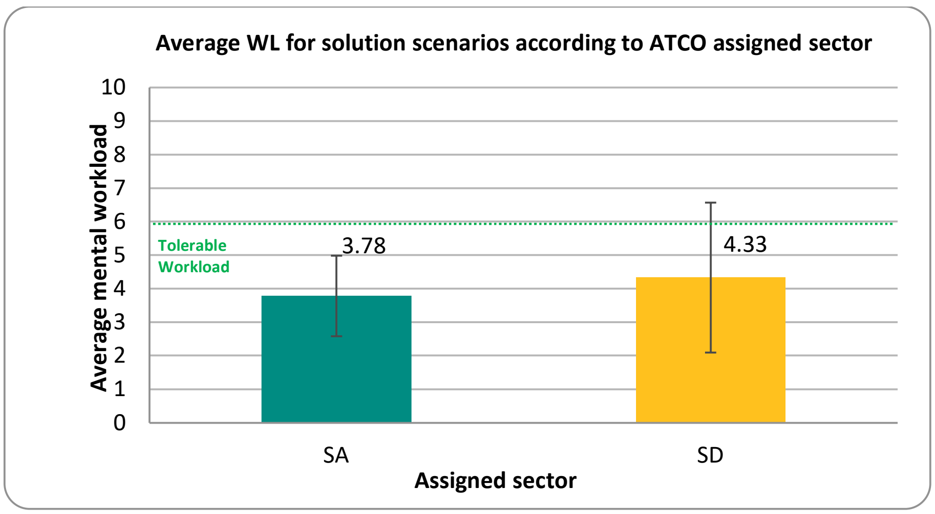

Figure A-16 - Post-run questionnaire – Average mental workload for solution vs reference scenarios according to ATCO assigned sector

During the post run questionnaires, the ATCOs were also asked to fill in a NASA TLX questionnaire, and a scale from 1 to 20 was used. The responses to the questionnaire can be seen in Figures 11 and 12 and are in line with the results collected through the Bedford questionnaire.

For the 'frustration' element, ATCOs reported relatively low values for both solution and reference scenarios, ranging on average from 4.8 (ref) to 8.3 (Sol.1). The fact that the value for Sol. 1 was slightly higher that the other scenarios might be due to this being the first ASR scenario for each simulation day and as a result, ATCOs needed some time to adjust to the ASR system.

Similarly, for the 'Effort' element, a high value is observed for Solution 1 (11) as compared to the rest of scenarios which ranged from 3.8 to 7 as well as for the 'Mental demand' element, where for Sol. 1, the average value is 8, higher than the rest of scenarios which range from 4.2 to 6.3. The same trend is observed for the 'Performance' element, where performance was lower for Sol. 1 compared to the rest of scenarios.

For the 'temporal' and 'physical demand elements' the average values are low for both reference and solution scenarios, with no significant peaks.

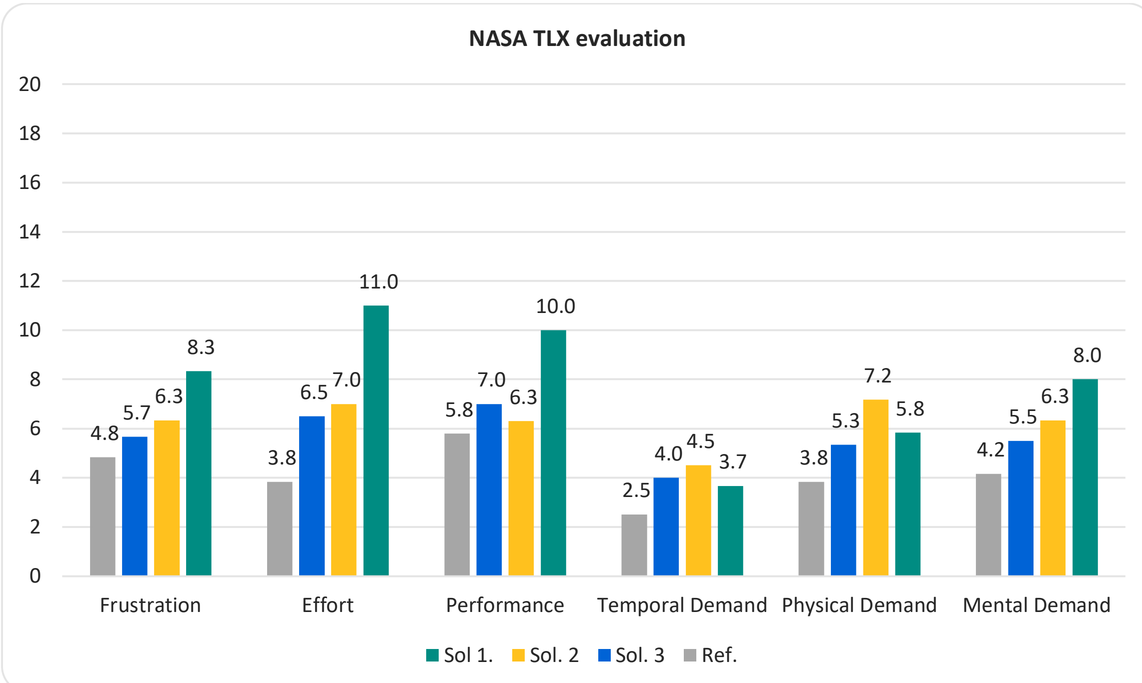

Figure A-17 - Post Run Questionnaire – Mental workload with NASA TLX

Overall, for all elements on the NASA TLX questionnaire, higher scores are observed for the solution scenarios which indicates a higher level of workload when using the ASR system, as compared to the reference. This is also observed in the diagram below.

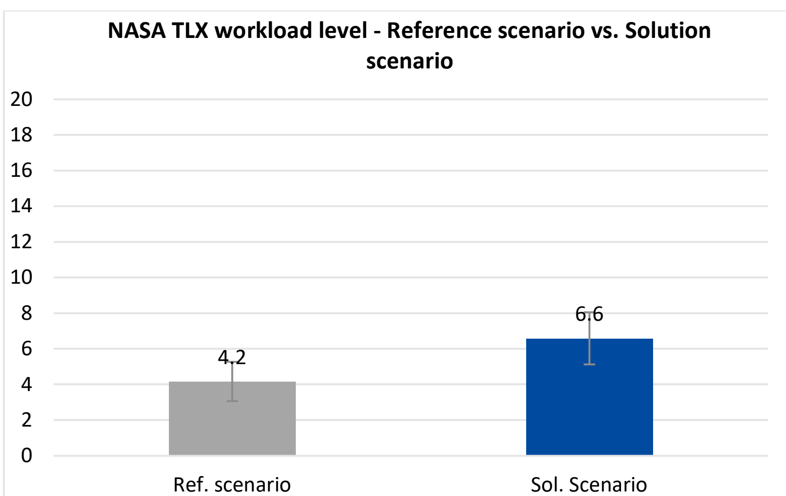

Figure A-18 - Post Run Questionnaire – Mental workload with NASA TLX – Reference vs. Solution scenarios

In the post-exercise questionnaire, 50% of the controllers reported 'acceptable' workload in the ASR scenarios compared to the reference scenarios, while the 34% of them reported very light or light workload levels. During the final debriefings, the system's latency and poor command recognition rates were said to be the main drivers for the negative Workload ratings among the controllers.

This is also represented in Figure 13 below which shows the distribution of answers on the 7-point Likert scale, also expressed in percentage.

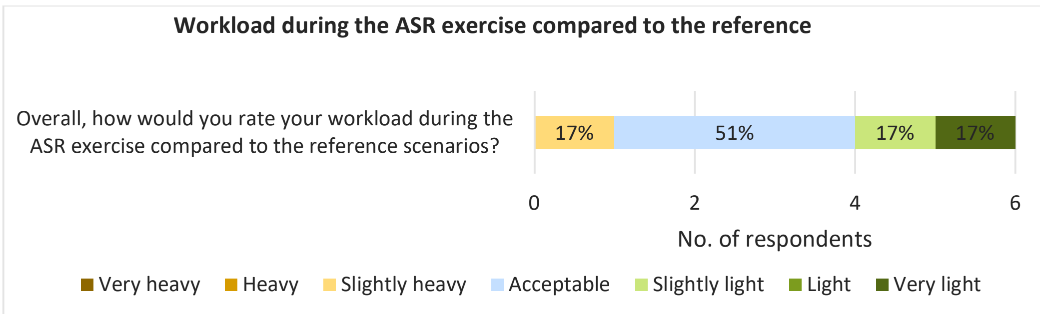

Figure A-19 - Post exercise questionnaire – Workload during the ASR exercise

In the post-exercise questionnaire, ATCOs were also asked if the ASR filling commands based on ATCO speech, improved workload with respect to the reference situation. 50% of controllers agreed, while the rest of them had divided opinions with answers ranging from 'neither agree nor disagree' to 'disagree'. This division in opinions might be due to the poor command recognition rates and high tool latency, which were also mentioned by ATCOs in the debriefings.

This is also represented in Figure 14 below which shows the distribution of answers on the 7-point Likert scale, also expressed in percentage.

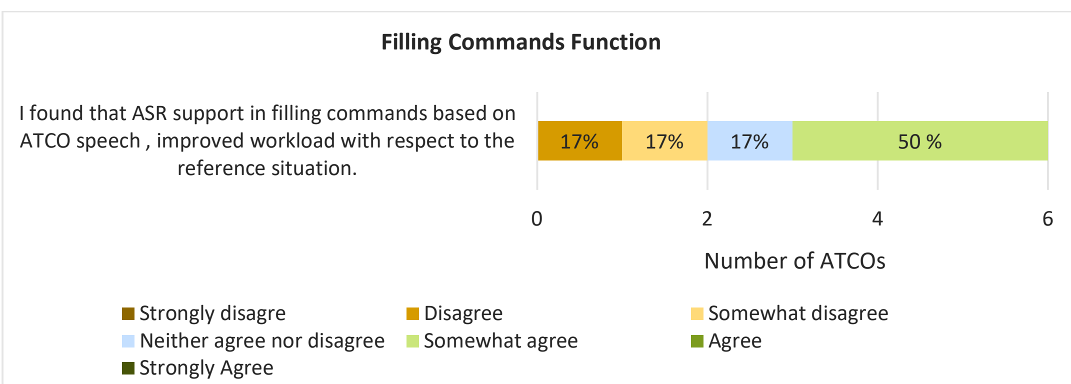

Figure 20 - Post exercise questionnaire – Workload improvement due to the 'filling commands' function

EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.002

The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate).

Please see the results for OBJ-Sol.96ASR-TRL6-TVALP-0020 above.

In the post exercise questionnaire, ATCOs were asked to rate the acceptability of callsign and command recognition and rejection as well as command manual correction, as seen in the Figure below.

Most ATCOs (83%) found the callsign recognition rate as either acceptable or 'slightly acceptable'. The manual correction required for the command was also found acceptable by most ATCOs (5) while the evaluation of the remaining ATCO was neutral.

In terms of callsigns rejection rate, opinions were divided. A third of the ATCO found it 'Acceptable' while the other two thirds were either neutral about it or found it 'Slightly unacceptable'

The command recognition rate was found acceptable by only 17% of ATCOs. Half of the ATCOs were not satisfied about the command recognition rate, rating it either 'Unacceptable' or 'Totally Unacceptable', while the rest of the ATCOs (33%) were neutral about it.

Similarly for the command rejection rate, nearly 70% of the ATCOs rating ranged from 'Totally unacceptable' to 'Slightly unacceptable' while the rest of the ATCOs were 'neutral' about it.

In terms of command manual correction required, only 17% of ATCOs rated it as 'slightly acceptable'. The rest of the ratings were divided between 'neutral' (33%) and 'totally unacceptable'.

During the debriefings, it was revealed some ATCOs felt that the command recognition is rather poor and provided a qualitative estimate for it, ranging between 25% and 50%. Others on the other hand were more positive about the command recognition rate and estimated it between 60% and 70%. This difference in experience might be caused by the fact that some ATCOs participated in the ASR training sessions and had more time to adjust to the ASR system and learned how to adapt to it for a higher recognition rate.

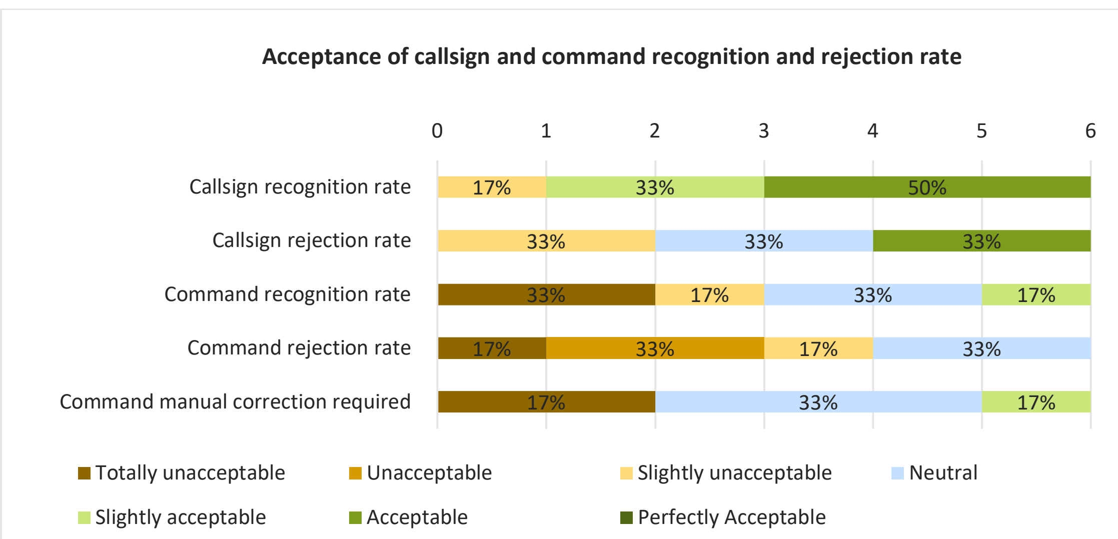

Figure A-21 - Acceptance of callsign and command recognition and rejection rate

In the post-exercise questionnaire, ATCOs were also asked about the acceptance of ASR wrong callsign and commands frequency. Most ATCOs (66%) were in agreement that the frequency of wrong highlighted callsign was acceptable.

However, only 17% of ATCOs somewhat agreed that the frequency of wrong recognised ASR commands was 'Slightly acceptable' while 17% were neutral about it and the rest of the ATCOs were in disagreement. This was motivated by ATCOs in the debriefings, who's perception was that the command recognition was rather poor, at this stage of development.

The difference in perception on the tool recognition rates could be motivated by the fact that the ATCOs that attended the ASR Operational Acceptance Testing (OAT) and Technical Acceptance Testing (TAT) have been more exposed to the tool as compared to the ones who only attended the training day and the simulation itself, and therefore had more time to adapt to the tool, which at this stage of development is not as flexible.

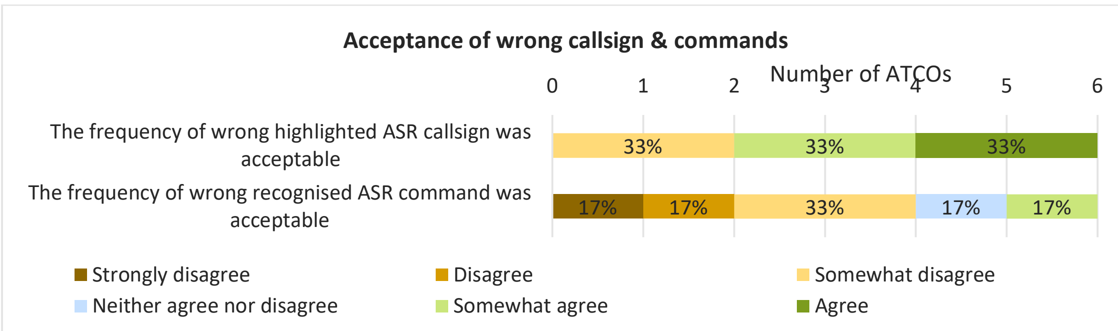

Figure 22 - Post exercise questionnaire – Acceptance of wrong callsign & commands frequency

In the CARS post-run questionnaire, ATCOs reported an average Acceptance level of 8,1 for the ASR tool. 44% of ATCOs rated the acceptance of the system above the acceptable minimum of 5. See the figure below for the distribution of answers collected for each point of the 10-point CARS scale.

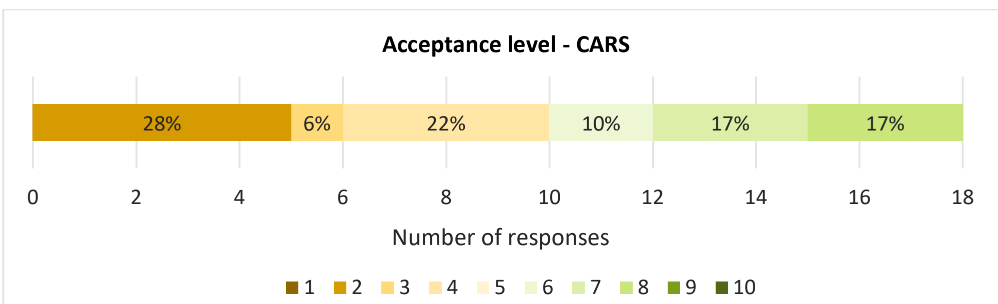

Figure 23 – Post run questionnaire – CARS

The average Acceptance level for each sector covered by ATCOs are illustrated in the figure below. All mean values are at or below the acceptable minimum of 5, which might be caused by the tool's poor recognition rates and high latency, at this stage of development.

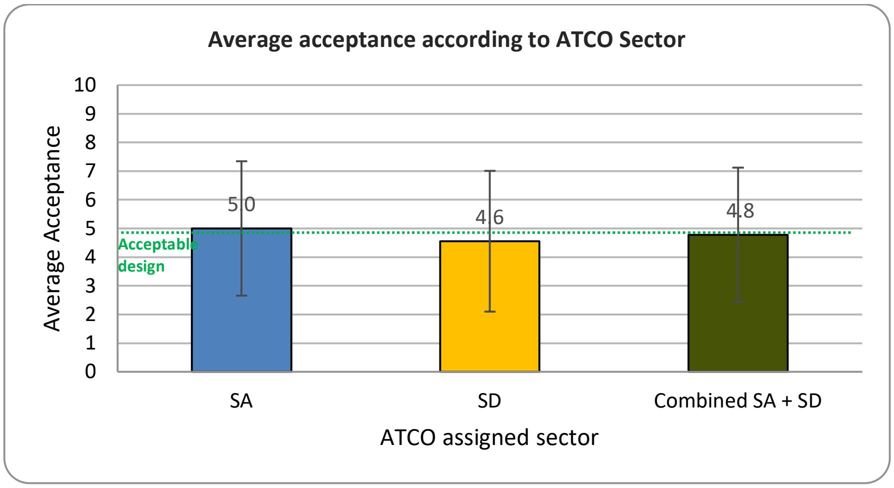

Figure 24 – Post exercise questionnaire – Average acceptance rate according to assigned sector

The job satisfaction was also investigated during the post-exercise questionnaire. Only 17% of ATCOs agreed to the fact that job satisfaction increases when using ASR, while 33% were neutral and the rest either 'Disagreed' or 'Somewhat disagreed'.

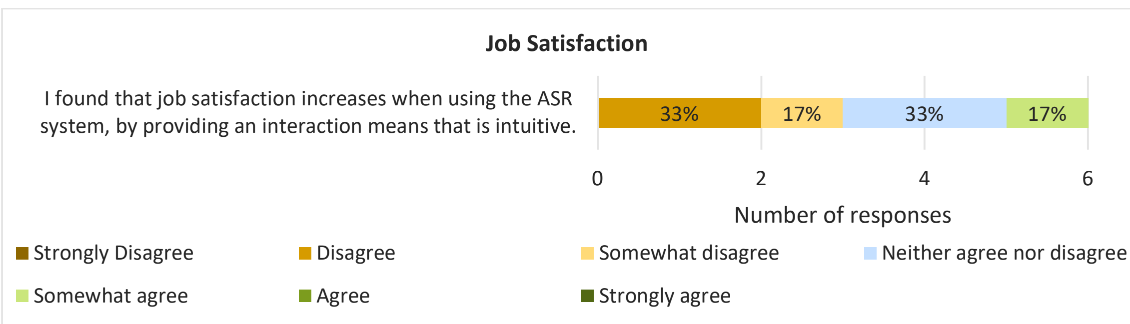

Figure 25 – Post exercise questionnaire – Job Satisfaction

During the debriefing, the ATCOs mentioned that due to the poor tool recognition, job satisfaction is impacted.

When asked to rate the overall acceptance of the system in the post exercise questionnaire, a general agreement can be observed between ATCOs. Nearly 70% of ATCOs ratings ranged from 'Slightly Low' to 'Low' while the rest of the ratings were 'neutral'.

During the debriefings, the low acceptance rates were motivated by the poor command recognition of the tool and the high latency which have an impact on controllers' execution of tasks in a timely and efficient manner.

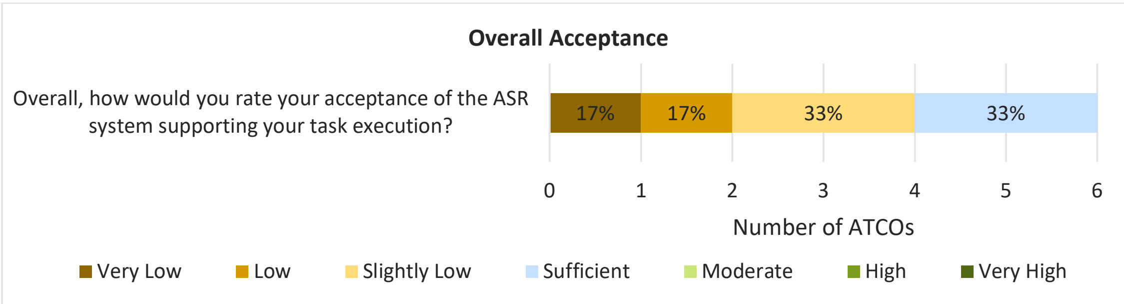

Figure 26 – Post exercise questionnaire – Overall Acceptance of ASR

EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.003

The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. Controllers' feedback with respect to Human Factors questionnaire is better than for baseline).

<u>ASR Latency</u>

The ASR system latency was also explored in the post-exercise questionnaire. As seen in the figure below, two thirds of the ATCOs were satisfied with the latency of ASR's 'Hook' function.

However, when it comes to ASR's feedback and command recognition, only 17% of ATCOs 'Somewhat agreed' that they are acceptable.

During the debriefings, ATCOs mentioned that the system has a high latency both in terms of feedback and in terms of command recognition, and that the pace of the task execution is much slower when using ASR, as compared to using radio communication and CPDLC. However, most ATCOs mentioned they are satisfied by the latency when it comes to the 'hook' function of the tool.

The figure below illustrates the distribution of answers for ASR latency related questions.

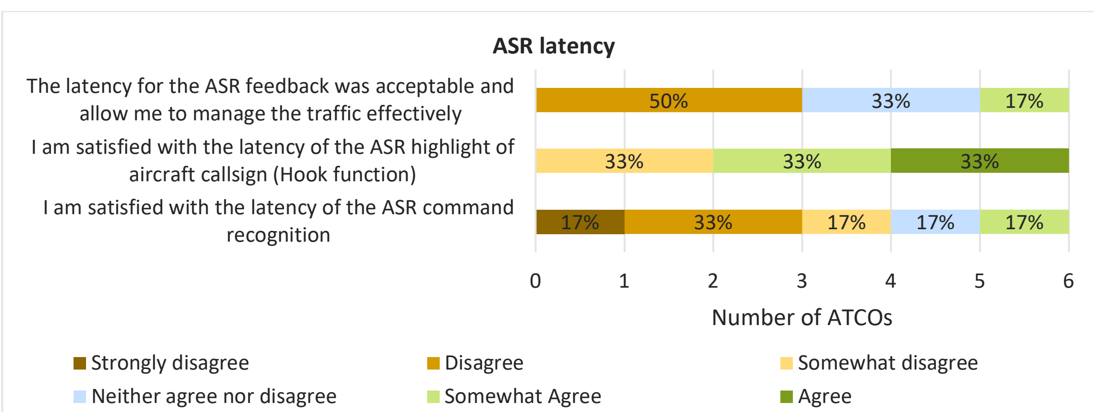

Figure 27 - Post exercise questionnaire – Latency of ASR response

Operating methods

In the post-exercise questionnaire, ATCOs agreed in 50% of the cases that they found operating methods for ASR to be clear, complete and exhaustive, under all operating conditions. The fact that half the ATCOs either disagreed or somewhat disagreed indicates that they were dissatisfied about applying operating methods when using ASR, which might be a result of tool's latency and high error rate (as shown in figures 21) which did not fully allow them to apply operating methods in an accurate, efficient, and timely manner.

Figure 22 below shows the number of answers collected for each point of the 7-point Likert.

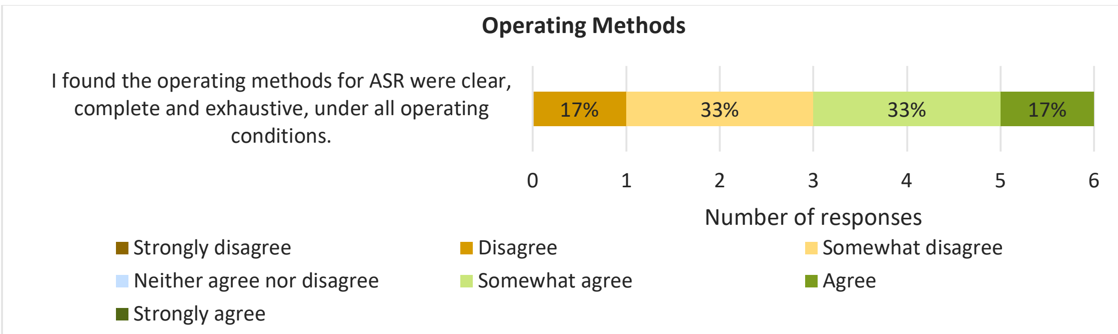

Figure 28 - Post exercise questionnaire - Operating Methods

EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.004

Changes in the design of the user interface (input devices, visual displays/output devices, alarm& alerts) support ATCOs in carrying out the tasks.

ASR & CPDLC

The ATCOs were also asked about the overall interaction between ASR and the CPDLC system. Most controllers (nearly 70%) were generally satisfied by the interaction and no issues were reported. During the debriefings, ATCOs mentioned that the interaction with CPDLC would be further improved if the latency of the tool were lower.

The figure below shows the number of answers collected for each point of the 7-point scale.

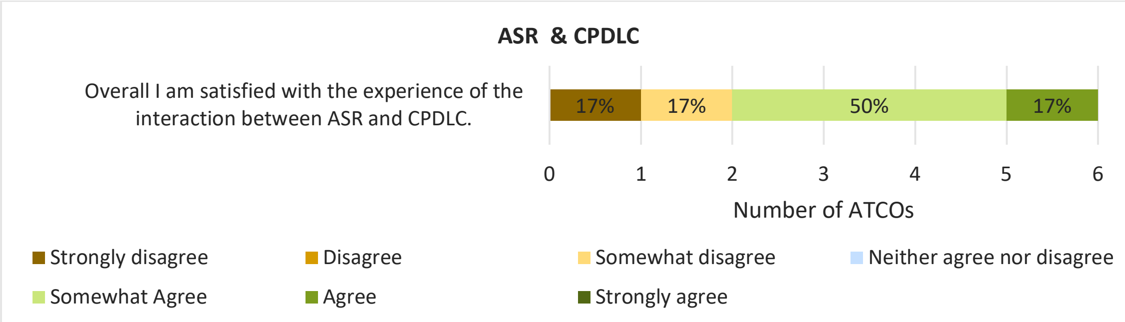

Figure 29 - Post exercise questionnaire – Interaction of ASR with CPDLC

<u>**HMI Feedback**</u>

In the post exercise questionnaire, the HMI feedback provided by ASR was also investigated. Most ATCOs disagreed (nearly 70%) that the level the feedback support provided by the ASR was adequate and clear and did not disturb them while the rest neither agreed nor disagreed.

During the debriefings, ATCOs mentioned that the colours used for the ASR dedicated windows should be more obvious as compared to the current ones and that the choice menu for manual input should be displayed close to the track of the aircraft of interest.

See the figure below for the number of answers collected for each point of the 7-point Likert scale, also expressed in percentage.

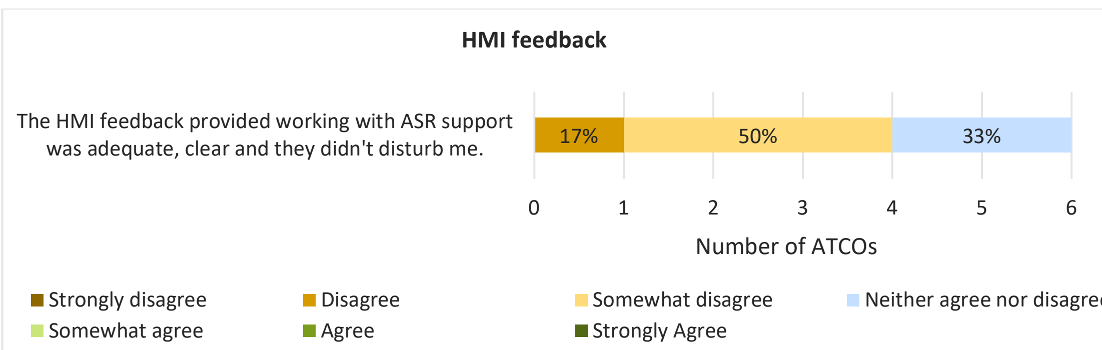

Figure 30 - Post exercise questionnaire – Adequacy of HMI feedback provided

<u>**The 'Hook' function**</u>

In terms of the 'Hook' function, all ATCOs agreed that they use it when identifying flights in complex and dense traffic, with answers ranging from 'Strongly agree' to 'Somewhat agree'.

Most ATCOs agreed to use the function to identify flights when entering their sector (nearly 80%) while the remaining 17% were neutral (neither agree nor disagree).

Half the ATCOs agreed they use it when arriving traffic calls for info or requests while the other half neither agreed nor disagreed.

Most ATCOs (nearly 70%) 'somewhat agreed' to using the function when they need to manage inbound traffic or deviating traffic while the remaining answers were divided between neither agree nor disagree and somewhat disagree.

During RT communication with other sector, 50% of ATCOs agreed that they use the 'hook' function, with ratings ranging from 'Agree' to 'Slightly agree'. The rest of the ATCOs were neutral (33%) or slightly disagreed (17%).

The difference in ATCO's responses related to the use of the 'Hook' function indicates that controllers have different preferences in this regard. The controllers might use the function according to their tasks or needs in their current role or based on their previous experience.

See the figure below for the distribution of answers, also expressed in percentage.

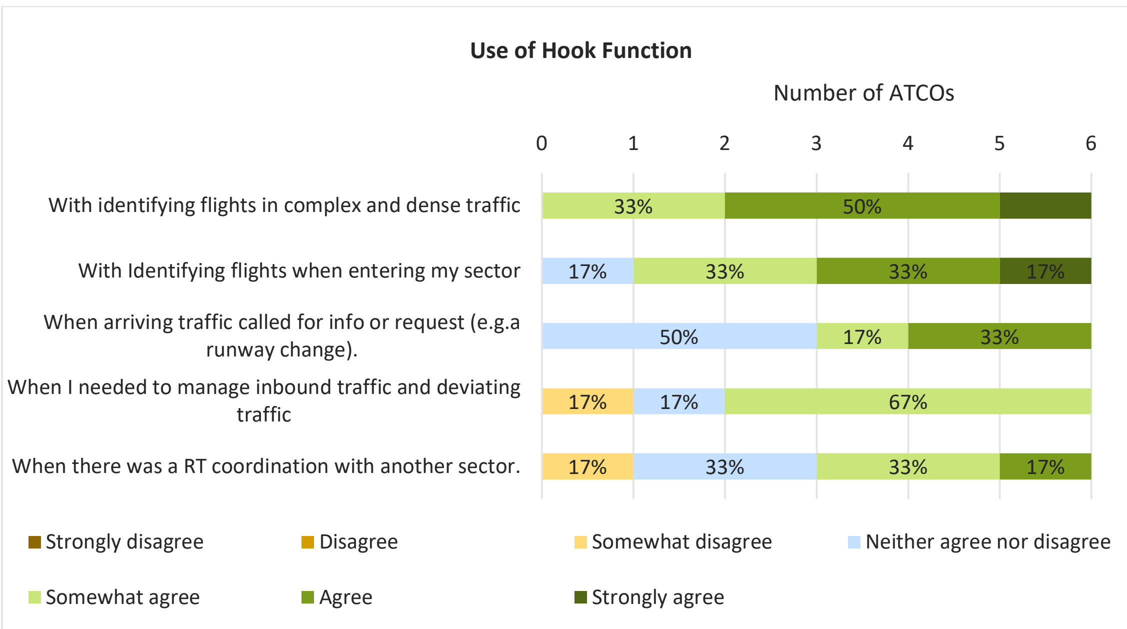

Figure 31 - Post exercise questionnaire – Use of 'Hook Function'

<u>ASR Use</u>

In the post- exercise questionnaire, ATCOs were also asked if they would like to only use ASR for call sign recognition (Hook function) or in routinary ATC tasks, without impacting CPDLC. As can be seen in the diagram below, there was no general agreement between the ATCO. 17% of the controllers prefer to use ASR only in routinary tasks, while 34% either 'disagreed' or 'strongly disagreed'. The rest of controllers (approx.50%) were neutral. This indicates that ATCOs would also like to use ASR for more complex tasks.

When asked if they prefer to use ASR for callsign recognition, the opinions were divided. Half of the controllers (51%) agreed, with answers ranging from 'strongly agree' to 'somewhat agree' while the rest either strongly disagreed (17%) or 'neither agreed nor disagreed' (33%).

The divided opinions related to the ASR use could be related to ATCOs experience or roles they cover, which might involve different controller needs.

During the debriefing, it was revealed that controllers found the 'Hook' function very useful and would be happy to use ASR for more complex tasks, granted that the system performs well in terms of latency and recognition rates of more complex commands, and it incorporates other 'nice to have' functions, listed in the 'Achieve' section of the WANT/HAVE matrix.

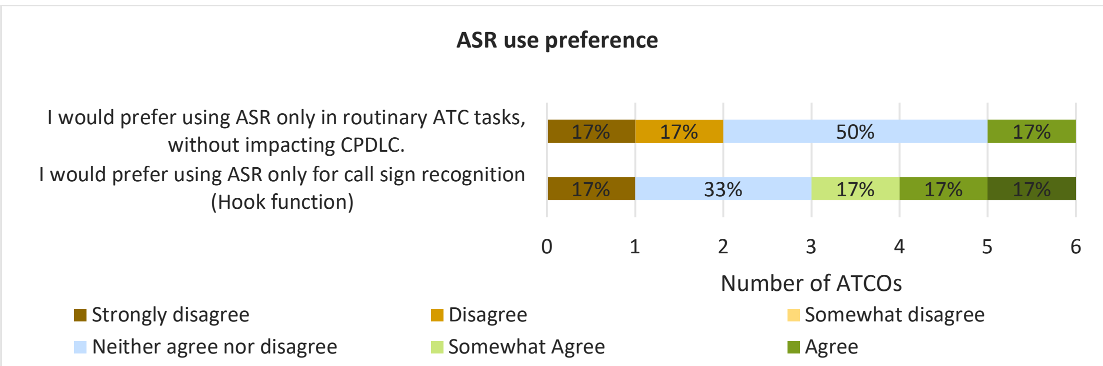

Figure 32 - Post-exercise questionnaire – ASR use preference

ASR Usability

In the Post-Exercise Questionnaire, the majority of ATCOs (5 out of six) were neutral when asked if they would like to use the system frequently while the remaining one strongly disagreed. Half of the ATCOs found the system unnecessarily complex, with responses ranging from 'Agree' to 'Strongly Agree'. Only a quarter of the ATCOs found the system easy to use and its functions were well integrated.

When asked if they would need the support of a technical person to use the system, the opinions were divided: two ATCOs agreed while the rest either disagreed or were neutral about it. The majority of ATCOs (4) thought there was too much inconsistency in the system. The system was found cumbersome to use by a quarter of the ATCOs, while the rest remained neutral about it. Only half the controllers felt confident in using the system.

However, when asked if they needed to learn a lot of things before using ASR, half the ATCOs disagreed, and the other half remained neutral. Also, when asked if most people would learn to use the system quickly, the majority of ATCOs (4) were in agreement.

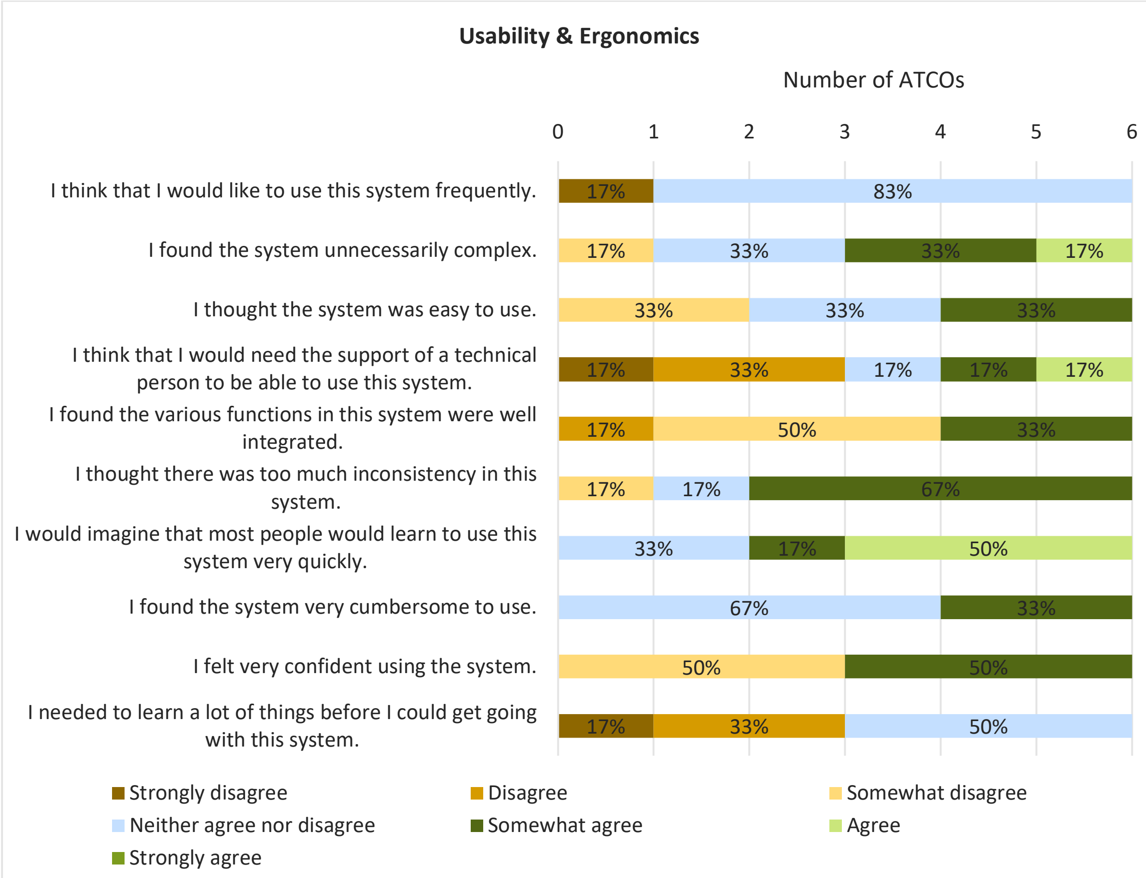

Figure 33 - Post exercise questionnaire – Usability (SUS) questionnaire distribution of responses per statement

When asked to rate the overall usability of the system, the majority of ATCOs (4) rated it as 'Sufficient' while the remaining ones rated it as either 'Low' or 'Very Low'. This indicates that at this stage of development, ATCOs are not entirely happy with the ASR's system usability. Th**Error! Reference source not found.** figure below shows the distribution of responses.

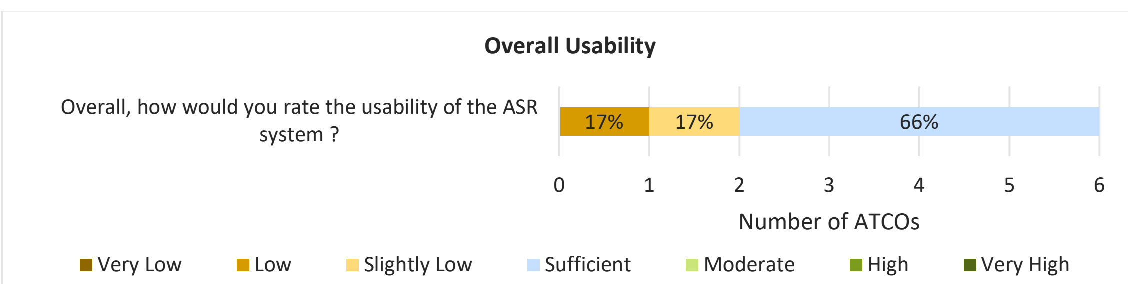

Figure 34 - Post exercise questionnaire – Overall usability rating

The general feedback on usability during the debriefings was that ATCOs see a great benefit of a system automatically filling in commands issued by voice, if the recognition rates were higher, and the latency were lower. These two aspects currently impact the complexity of the system and its use, together with the errors and inconsistencies identified. It seems that at this stage, the controllers do not trust the system enough to use it confidently. Please see the results for EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.007 in relation to trust, in the following section.

Moreover, in the debriefing sessions, ATCOs suggested that the ASR recognised phraseology is to be enriched, as it currently does not recognise some important commands e.g., 'direct to, 'turn right/left' 'contact' etc.

During the debriefing sessions ATCOs were also asked for feedback about the use of certain keys on the keyboard to activate ASR as opposed to the use of the pedal. There was no general agreement between ATCOs as some claimed they prefer to have hands free and use the pedal, but some others found the pedal a bit outdated. A few ATCOs would like to have ASR always active, without pressing any keys or pedals. Some ATCOs expressed some concerns about the latter, as as ASR could pick on some of ATCOs' 'offline' coordination and internal discussions and erroneously provide commands based on that.

ATCOs also provided some suggestions which would improve the usability of the system such as the implementation of an ASR 'diagnostics window which would display useful information such as history of commands issued, ASR issues etc. For more information on ATCOs suggestions for an improved usability, please see the 'Achieve' section in the Want/Have Matrix results.

Training

In the post-exercise questionnaire, ATCOs were also asked whether the effective use of ASR in operation requires a dedicated training (i.e., classroom, simulator, on-the job training). As the graph below shows, the majority of ATCOs (5) either agreed or strongly agreed on the need of dedicated training.

In the debriefing sessions, ATCOs mentioned that training on the ASR system is fundamental especially regarding the use of the pedal or switch between keys, if these will remain the main activation or input modes.

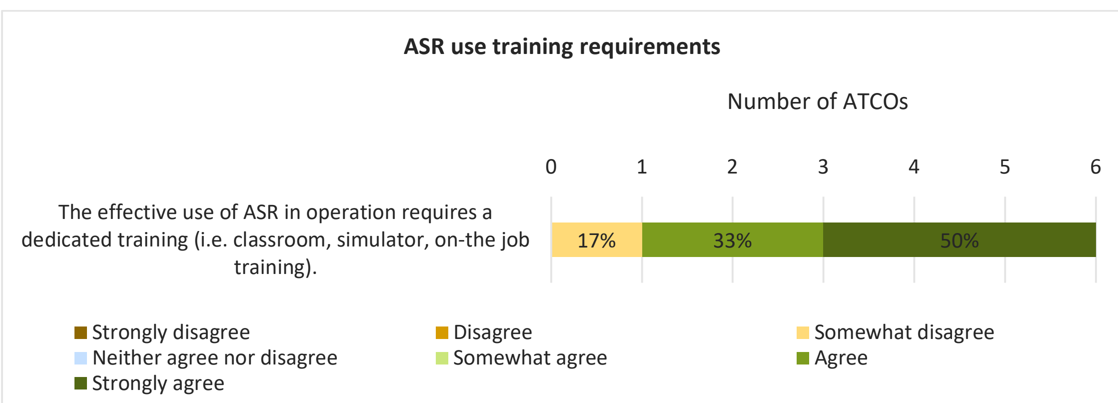

Figure 35 - Post Exercise Questionnaire - ASR use training requirements

EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.006

The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified)

In the post-run questionnaire, the standard SATI questionnaire was used for the evaluation of Trust. ATCOs were asked to select for each of the six statements, the experienced frequency on a seven-point scale (never (0), seldom (1), sometimes (2), often (3), more often (4), very often (5), always (6)) that better represents their experience during each run.

As regards to solution scenarios, no significant difference was observed between the three, in terms of average experienced frequency for any of the trust statements (average ratings ranging from 1.3 to 2.8 as seen in the figure below). This means that the three solution scenarios can be considered homogeneous.

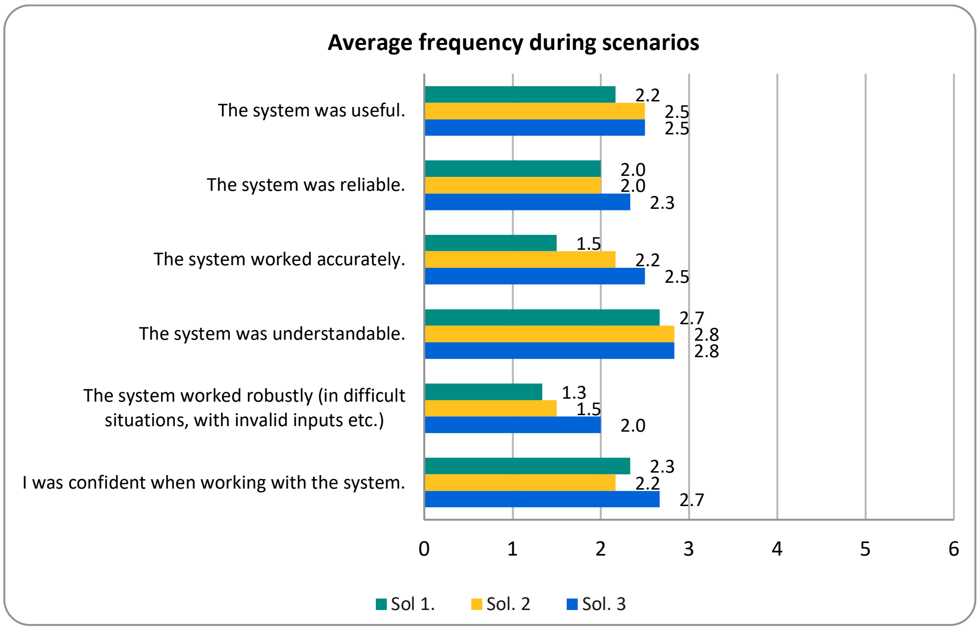

Figure 36 - Post run questionnaire: Average frequency per scenario type for each statement on the SATI standard questionnaire

The Figure below illustrates the average frequency for each one of the SATI statements, for the reference scenario responses against the solution scenarios responses altogether.

Although the data sample for the reference scenarios is smaller than the one for the solution scenarios together, after applying Welch's t-test, it was revealed that the means of the two are not statistically different.

Overall, higher average values are observed for the Reference scenario, as compared to the solution scenarios, for all statements of the SATI questionnaire.

The low averages for the solution scenarios for all statements (ranging from 1.6 to 2.4) indicate that overall, ATCOs trust in the system is rather low, from various perspectives: reliability, accuracy, usefulness, understandability and robustness. Moreover, the average confidence in working with the system is also low (2.4).

During the debriefings, ATCOs mentioned that ASR struggles to recognise commands in an accurate manner, unless the speech is very slow or segmented and provided at a steady pace. ATCOs noticed that if they hesitate or take a break in speech while providing the command, the ASR does not recognise it. For higher trust levels, ATCOs suggested the tool needs to be more consistent and precise.

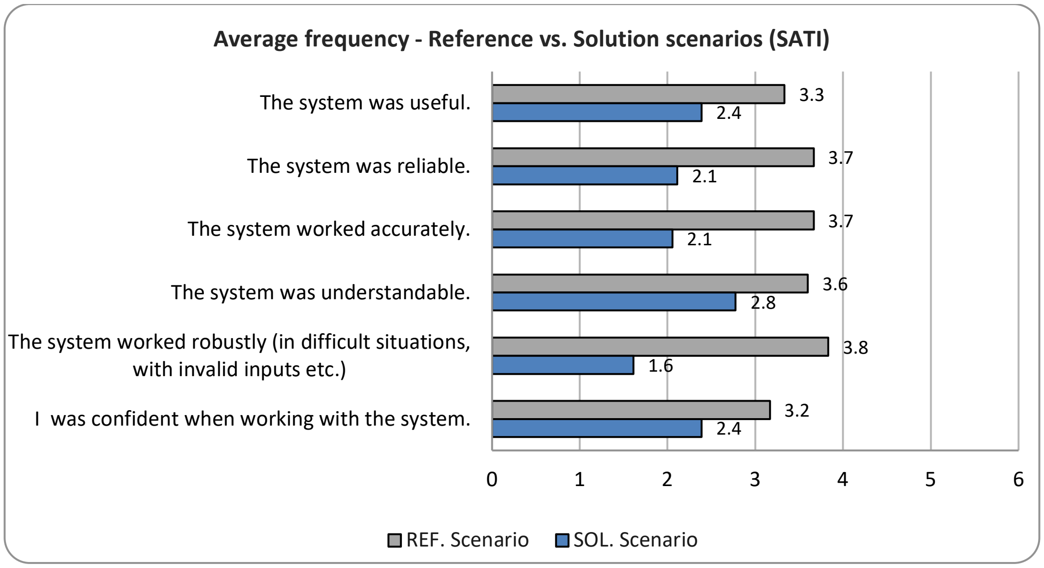

Figure 37 - Post run questionnaire – Average experienced frequency per SATI questionnaire statement for reference and solution scenarios

Also, during the post exercise questionnaire, the majority of ATCOs rated Trust poorly, with ratings ranging from 'Slightly low' to 'Low', while the rest rated it as *'Sufficient'*. This indicates that the majority of ATCOs does not trust the ASR system. The low ratings were also motivated during the debriefings by the high latency of the system and high error rates.

See figure below for the number of answers collected for each point of the 7-Point Likert Scale.

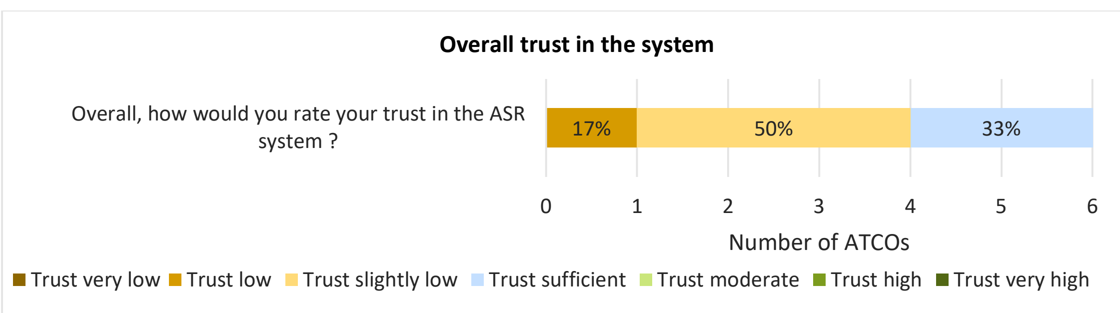

Figure 38 - Post exercise questionnaire – Overall trust in the ASR system

4. OBJ-Sol.96ASR-TRL6-TVALP-0040 Results

To assess the impact of the introduction of the ASR system on safety.

EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.001

The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.

See results on criteria EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.006 on HMI feedback.

In the post run questionnaire, controllers were asked whether the level of safety was acceptable, and no degradation or safety concern was raised during the ASR scenarios. Most responses show a positive attitude regarding the level of safety since on nearly 60% of occasions ATCOs either 'strongly agreed', 'agreed', or 'somewhat agreed' to the statement. The rest of the responses were however divided- (18%) were neutral (neither agree nor disagree), 18% somewhat disagreed while 6% strongly disagreed.

When asked if they were able to detect an ASR degradation and switch to manual mode without a decrease in safety in the previous run, the majority of responses (approx. 65%) were positive, with ratings between 'Somewhat agree' to 'Strongly agree'. Also, 12% of answers were neutral and on 12% of the occasions, ATCOs 'disagreed'.

The different perceptions on the level of safety for the two questions could be explained by some of the ATCO comments, which mentioned that during the simulation, the readback of the pseudo pilot arrived before ASR finished to process the command provided which indicates a high system latency. In a real traffic situation this would not allow the ATCOs to follow the dynamic flow of the operations and might impact safety because of an overlap in communication. As a result, at this level of development, ASR is seen more like a redundant system by ATCOs. As mitigation, ATCOs suggested an exclusive datalink communication, without the use of radio frequency.

See the figure below for the number of answers collected for each point of the 7-point Likert scale, for each of the two safety related questions.

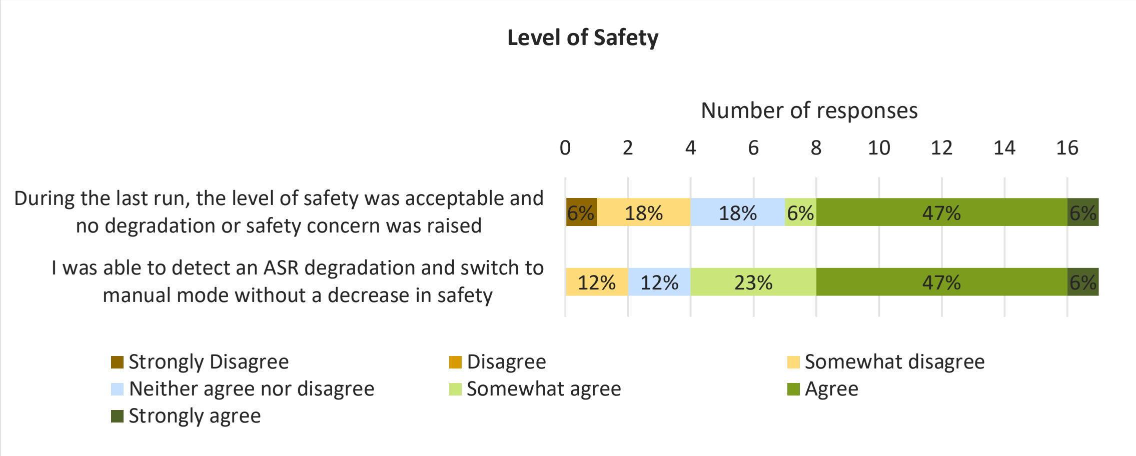

Figure 39 - Post run questionnaire – Level of Safety during Solution Scenarios

**In the post run questionnaire, one ATCO response is missing for each Safety related statement*

In the Post exercise questionnaire, ATCOs were asked if the overall level of safety was at least as the today operations during the ASR exercise execution. 50% of ATCOs agreed, while only 17% disagreed, and 33% remained neutral. This suggests that the overall level of safety was satisfactory for only half of the controllers. The lack of agreement in the other controllers might be the result of the system's latency and error rate, mentioned during the debriefings and discussed above Figure 34

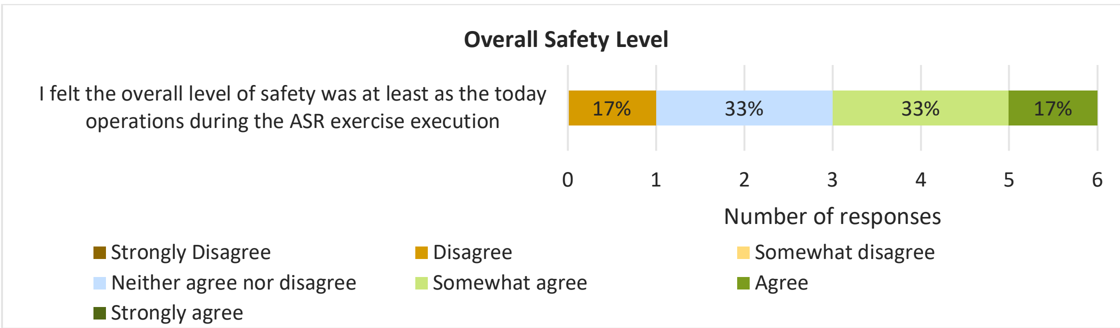

Figure 40 - Post exercise questionnaire - Overall Safety level with ASR

A few improvements were suggested by ATCOs regarding safety, for example integrating another function into the ASR system, which supports with recognising ATCOs and pilots' communication simultaneously and with identifying potential mismatches by double checking that the information transmitted via radio is correct.

The 'hook' function could also be used as a safety barrier as it could support with avoiding confusion between inbound flights with similar callsigns, which is a risk that ATCOs are often exposed to.

EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.002

The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.

See results for EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.003.

EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.003

The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.

33% of ATCOs either agreed or strongly agreed that ASR did not increase potential for human error compared to current operations, whereas 17% of ATCOs disagreed, and the rest (50%) were neutral about it. This does not indicate a clear ASR system impact on the potential for human error, although safety concerns have been raised during the debriefing in relation to the system's latency and error rates at this level of development, which according to ATCOs could impact safety. Please see the results for EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.001 above for ATCO's safety concerns related to ASR.

See the figure below for the number of answers collected for each point of the 7-point Likert scale, also expressed in percentage.

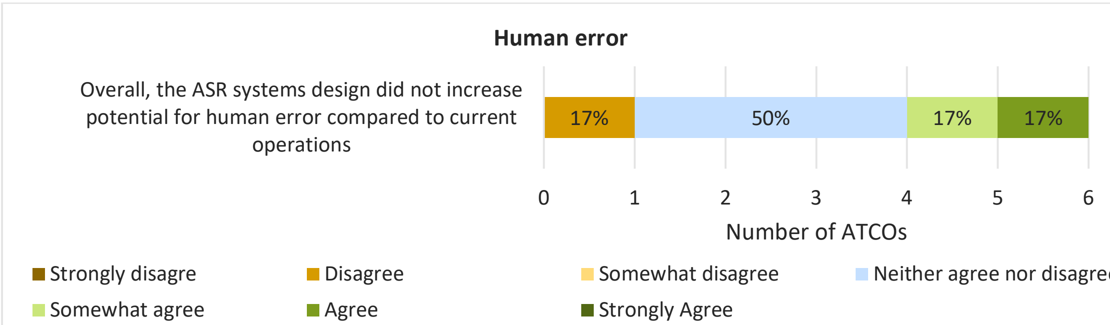

Figure 41 – Post exercise questionnaire – Distribution of responses for potential increase of Human Error

For ASR error rates, see results for EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.002

EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.004

The level of ATCO's situational awareness is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions).

After each run, the standard SASHA questionnaire was used for the evaluation of Situational Awareness. ATCOs were asked to select for each of the six statements, the frequency on a 7-point scale (from 0-never to 6-always*[2]) that better represents their experience during each run. No significant difference was observed between the three solution scenarios, in terms of average experienced frequency, for any of the situational awareness statements, as seen in the figure below. This means that the three solution scenarios can be considered homogeneous.

[2] Please note that the scale has been inverted for statements 2,3,5 & 6 as per the SASHA scoring key

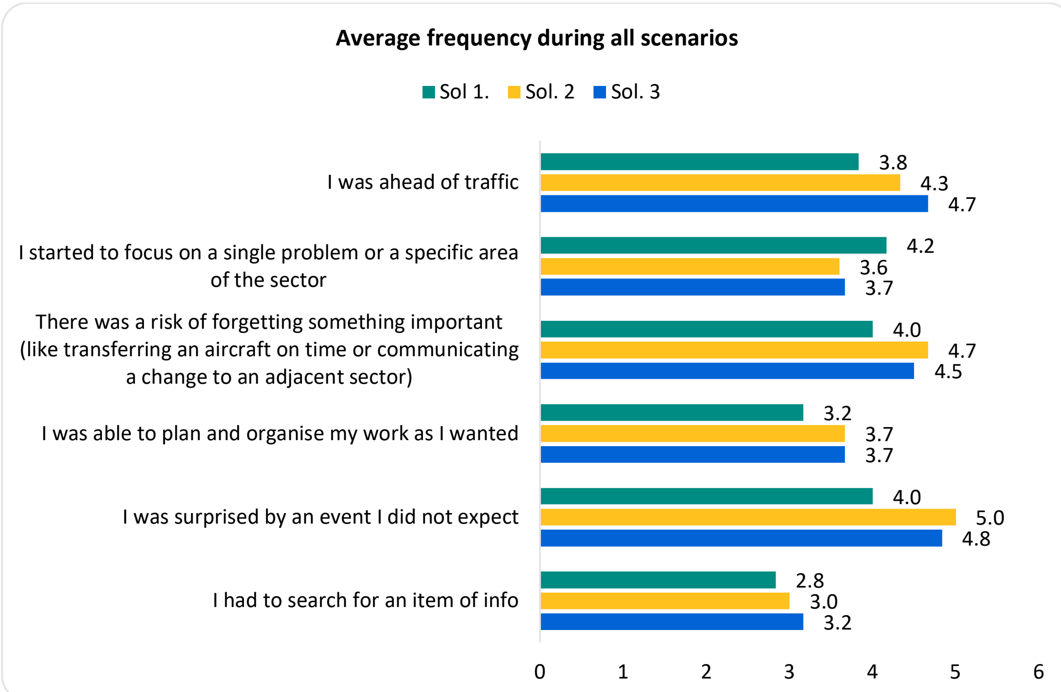

Figure 42 - Post run questionnaire – Average experienced frequency per each type of scenario for each statement on the SASHA questionnaire

When comparing the reference to the solution scenarios as seen in the figure below, it can be observed that overall, for all six statements, Situational Awareness was higher in the reference scenarios compared to the solution scenarios.

More specifically, ATCOs said that they were very often ahead of traffic during the reference scenario (5.7), however lower values are observed during solution scenarios (4.3

On average, ATCOs claimed that they almost never had to focus on a single problem or a specific area of the sector during the reference scenario (5.7), as compared to the solution scenario where the average value is (3.8), which indicates that ASR might sometimes distract ATCOs or demand their attention.

Similarly, in reference to the statement about the 'risk of forgetting something important', the average ATCOs responses were estimated at an average of 5.8 in the reference scenarios, as compared to 4.4 in the solution scenarios, which indicate that ATCOs feel that they would sometimes be exposed to that risk when using ASR.

On average, when asked if they were able to plan their work as they wanted during the reference scenarios, a value of 5.2 is observed. However, a much lower value was reported for the solution scenario (3.5), which means that when using ASR, ATCOs feel like they can't plan and organise their work as well as without the system.

When asked if they were surprised by an unexpected event, an average frequency of 5 is reported by ATCOs for the reference scenarios and a slightly lower value is observed in the solution scenarios (4.6), which indicates that this situation happens seldomly whether ASR is being used or not.

An average frequency of 3.5 was reported when ATCOs were asked if they had to search for an item of information during the Reference scenario and a slightly lower value can be observed for the solution scenario (3). This indicates that ATCOs might need to be provided with more required information, without having to look for it, whether or not they use ASR.

See the figure below for the average frequency reported for each one of the questions in the SASHA questionnaire, for both reference and solution scenarios together.

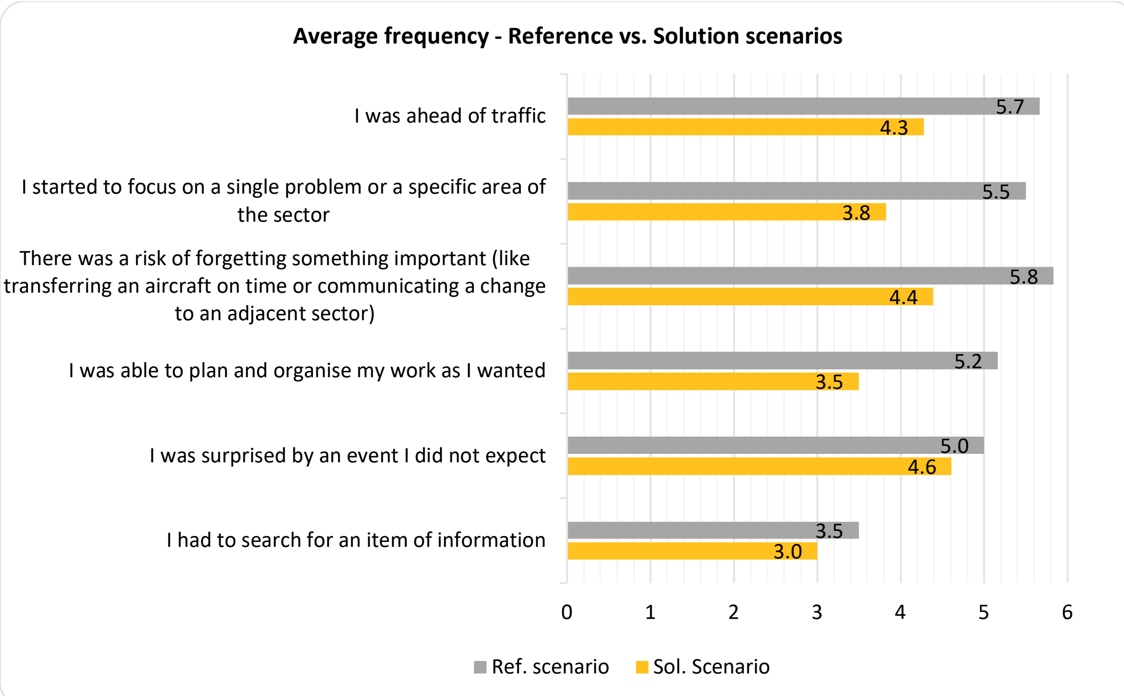

Figure 43 - Post run questionnaire- Average reported frequency in the post- run questionnaire for Situational Awareness Statements

See also the figure below for the number of answers collected for each point of the 7-Point Likert Scale, for each one of the questions on the SASHA questionnaire, also expressed in percentage*. [3]

[3] Please note that the percentage calculation is an approximation of the percentage they represent, to avoid cluttering the chart.

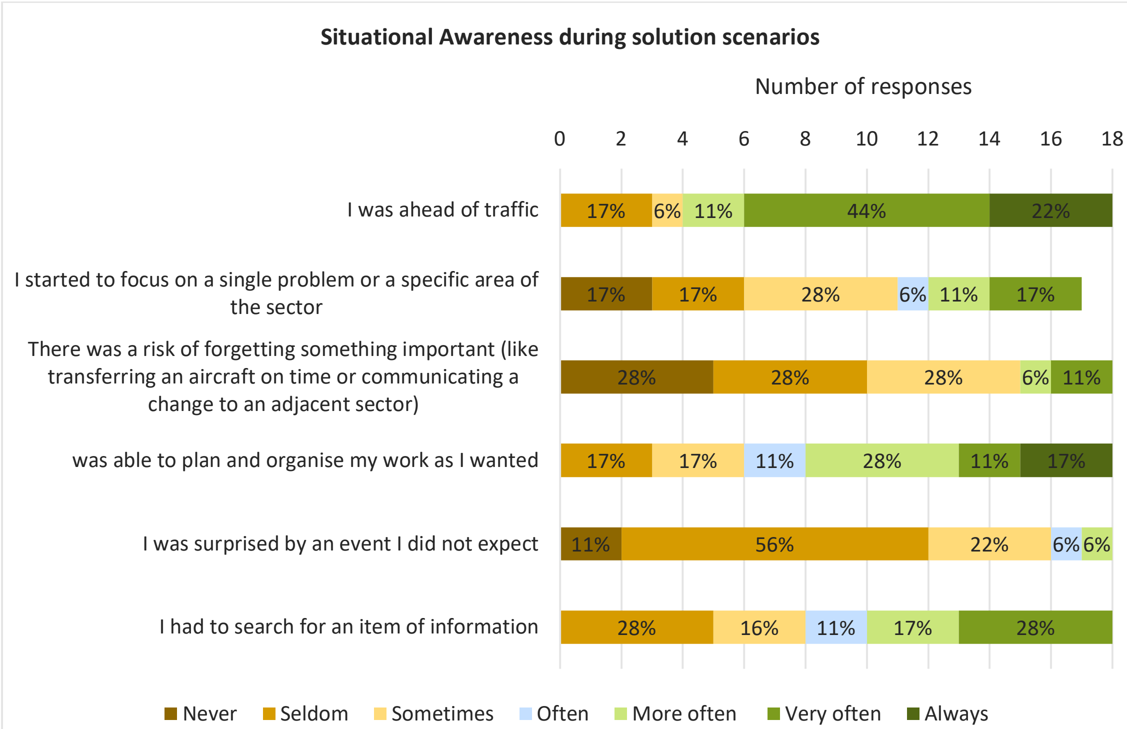

Figure 44 - Situational Awareness during solution scenarios

*One of the ATCOs omitted the statement ‘I started to focus on a single problem or a specific area of the sector’.

When looking at situational awareness related average frequency experienced by the controllers according to each assigned sectors in the solution scenario, it can be observed that there is no significant difference in experienced situational awareness, for any of the six statements in the SASHA questionnaires. This indicates that the sector allocation does not influence situational awareness.

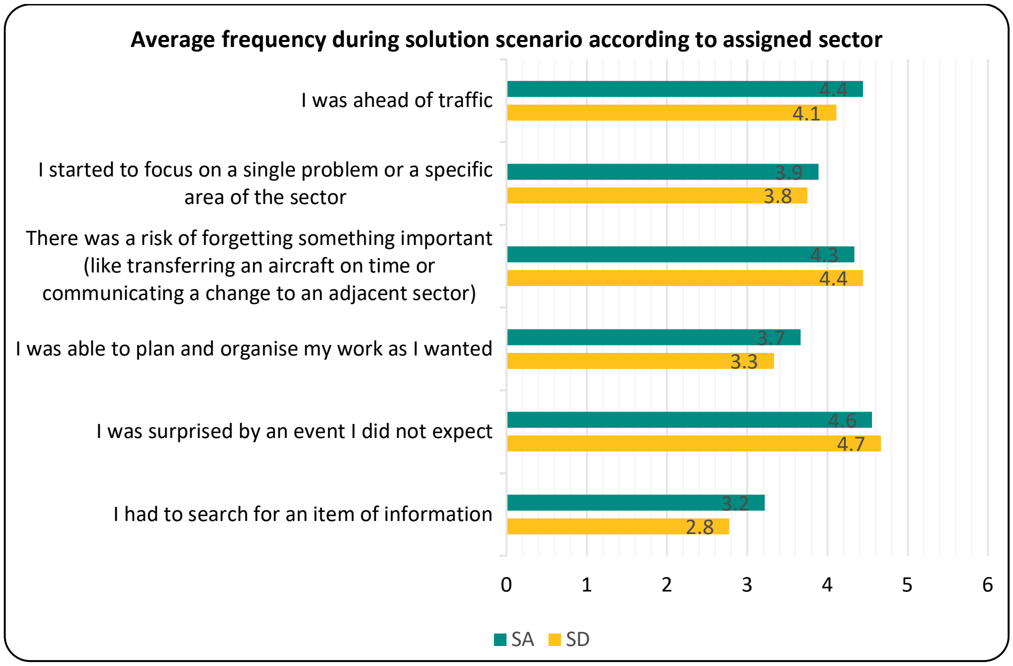

Figure 45 - Post run questionnaire – Average reported frequency in the post-run questionnaire for Situational Awareness statements according to ATCO role

In the post exercise questionnaire, ATCOs were also asked if the ASR "HOOK" function used to highlight the callsign improves situation awareness with respect to the reference situation. The answers ranged from 'agree' (33%) to 'somewhat agree' (67%). This suggests that all ATCOs were in agreement that the 'Hook' function improves situational awareness.

During the debriefings, ATCOs mentioned that the 'Hook' function is well implemented, especially in situations of congested traffic or when one is not familiar with the respective airline callsigns. The function could also be used as a safety barrier for avoiding confusion between inbound flights with similar callsigns.

See figure below for the number of answers collected for each point of the 7-Point Likert Scale, also expressed in percentage.

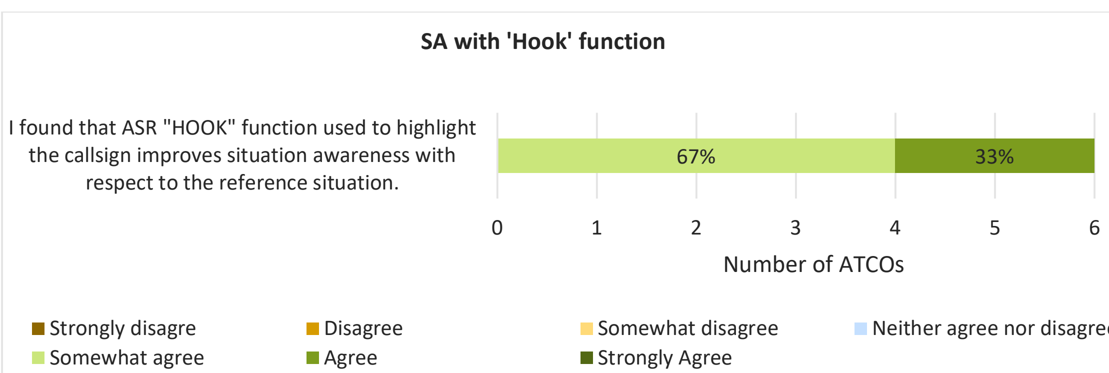

Figure 46 - Post exercise questionnaire – Situational awareness improvement with 'Hook' function

In the post exercise questionnaire, out of the six ATCOs participating in the simulation, 50% rated the overall Situational Awareness during the ASR exercise as *High*, while 17% rated it as 'degraded' and the rest as 'Sufficient'. This indicates that the Situational awareness was satisfactory for only half the ATCOs when using ASR. See figure below for the number of answers collected for each point of the 7-Point Likert Scale.

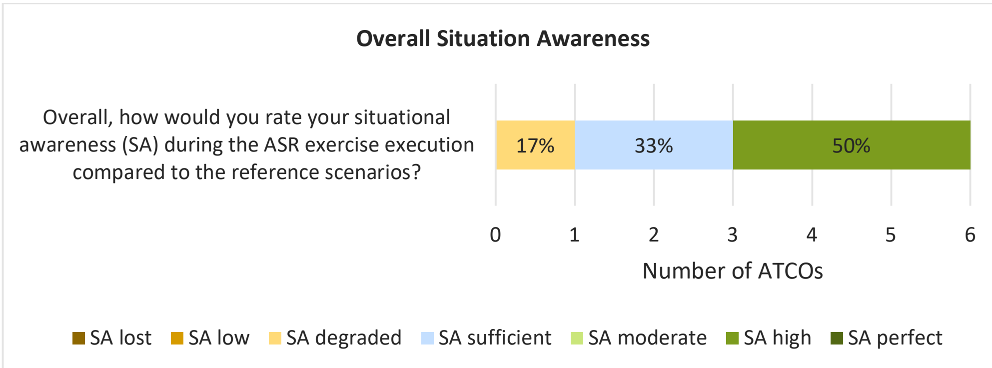

Figure 47 - Post exercise questionnaire – Overall situational awareness during the ASR exercise

While the 'hook' function was said to support with situational awareness, the reason why half the ATCO's rated Situational Awareness poorly could be related to the system's high latency and error rates, which were said to also impact workload acceptance/ trust in the system.

EX01-CRT-Sol.96ASR-TRL6-TVALP-0040.005

The level of ATCOs' workload is maintained or decreased with the introduction of ASR system.

See results for EX01-CRT-Sol.96ASR-TRL6-TVALP-0030.001.

A.7.3 Unexpected Behaviours/Results

The platform used for the validation was an operation ATM system LeadInSky provided by LDO. The system is not the one used during daily work by the ATCOs involved in the validation.

Some configurations of the HMI were very different from the one they are used to and some ATCOs didn't like any of them. This affected their attitude towards the system and their feedback of the ASR.

A.7.4 Confidence in Results of Validation Exercise 1

1. Level of significance/limitations of Technological Validation Exercise Results

Simulation EXE-001 has involved a wide range of Test subjects (6 ATCOs) with different background and expertise in a simulation environment representing Sofia En-Route operational environment with a high level of fidelity. Considering the simulation conditions, the results for ASR are judged to be characterised by a high level of significance, even if the training of ATCO was quite limited for time constraints reasons and this might have affected the collection of data of initial runs of each simulation day.

2. Quality of Technological Validation Exercises Results

Questionnaires have been used to collect ratings from the test subjects on the different aspects of ASR as explained before both the accuracy and the confidence on the collected results and measured indicators are judged at a high quality to support the maturity assessment of TRL6 phase.

3. Significance of Technological Validation Exercises Results

The simulation exercise has been conducted on an experimental platform representing Sofia En-Route environment with a high degree of fidelity providing an operational significance adequate to support the TRL6 maturity assessment, of course with the limitations already mentioned in above.

A significant number of total runs have been conducted among 3 simulation days (12 total number of runs) as well as a significant number of test subjects (6 ATCOs) have been involved to conclude that results are significant to support the TRL6 maturity assessment, but it cannot be considered that the results have statistical significance. Considering the validation technique (real time simulation) and the executed numbers of runs it is judged the results have a high level of significance.

A.8 Conclusions

ASR4ATC represents a first step in the development of an ASR computing platform in order to provide support to ATCOs for Leonardo. In spite of compromises were made and the limitations in allowed utterances, results of the Validation Exercise indicate good performance and satisfactory results of the assessment of the ASR tool made by ATCOs. ATCOs believed that the system could be useful in the near future when the technology is more developed, faster and accurate than it is today for LDO’s prototype. Looking ahead, such a tool can improve greatly and provide a more effective means to significantly reduce ATCO workload, which in turn would entail a higher throughput of flights and finally a higher capacity.

1. Conclusions on Technological feasibility

The exercise confirmed the feasibility of integrating an Automatic Speech Recognition tool in an operational Controller working position with an operational communication system to support the use cases:

- UC.2 **Highlight of callsigns on the CWP from controller’s utterances**. The highlight of the callsign coming from a controller utterance is expected to support the ATCO Situation awareness in the tactical management phase (e.g. a/c inbound to the sector or during inter sector coordination) or in case of a/c requesting actions (e.g. deviation, flight level change or request information).

- UC.4 **Prefilling of commands in the CWP.** In this use case ATCOs will be presented with the recognized (and validated) command types together with the command values in the CWP. ATCOs will then be able to accept/reject or manually correct the commands.

- UC.8 **Prefilling of Datalink commands.** After the Logon and Connecting procedure have been fulfilled, ATCOs will utter the datalink commands and the CWP will prefill the command values to be accepted/rejected or manually corrected by the ATCOs.

The integration was performed without impacting the performance of other console systems coherently with the rest of the HMI. The ASR preindustrial prototype was able to receive information online of the flights entering and leaving the sectors and was able to provide to the HMI the necessary input to be displayed to the controller.

2. Conclusions on performance assessments

- **Cost Efficiency Performance**:

Cost efficiency of ASR is evaluated in the context of a Cost Benefit Analysis.

- **Human Performance**

Impacts of the solution on the following relevant topics were addressed through questionnaires and debriefings: *ATCO Situational Awareness, ATCO Workload, ASR and ASR HOOK Function, ATCO Acceptance & Job Satisfaction, Trust in the system, ASR Usability/Ergonomics, ASR Callsign & Command Recognition, ASR Interaction with CWP and Human Error*. This was accomplished in combination with a realistic simulation in which end-users performed realistic tasks.

- Results confirmed the benefits associated to the solution in terms of human performance as well as its open issues. The outcomes indicated that at the current level of development, the latency and recognition rate of ASR4ATC may negatively impact workload and situation awareness. On the other hand, beneficial effects arising from the support offered by the 'Hook' function and the 'commands filling' function on situational awareness and workload resulted from simulations. In order to further enhance efficiency of ASR support, some improvements were identified: lower latency and higher recognition rates, as well as a need to enrich the phraseology recognised by ASR4ATC Also, changing the background colours of ASR-generated pop-up windows and a more evident highlight of 'Hooked' a/c on the HMI would also increase situational awareness.

- Low levels of trust in the system and acceptance from ATCOs indicate that the level of ASR technical performance needs further improvement to be consistent with human capabilities. However, positive feedback was provided in terms of integration of ASR4ATC with CPDLC.

- Favourable reactions were received from ATCOs regarding the automatic filling in of commands issued by voice, which suggest a great potential of the system and its functions to enhance user experience, if the latency and error rates were improved. Some degree of training would be required for ATCOs to better understand "behaviours" of ASR4ATC and also to learn how to proactively adapt their speech to the tool.

- **Safety**

- Concerns have been raised in relation to the system's latency and error rates at this level of development which could impact safety.

- No other specific safety issues were identified during the validation exercise; however, safety aspects were addressed across all runs.

A.8.2 Recommendations

1. Technological feasibility

The ASR technology has shown to be feasible in an ATC En-Route environment. However, ASR technological feasibility could benefit from some refinements and improvements as shown in the following:

- Train phonetic models to accept local English as a foreign language accents
- Allowing more than one command per utterance and widen the command choice, always based on the SESAR shared ATC Ontology
- Current response latency was considered not fully acceptable by ATCOs, one significant step forward could be concept-by-concept recognition and transcription, also referred to as online transcription
- Making callsign range wider, including military, GA, more formats and airline operators
- Incorporate the entire SESAR Ontology, in order to enlarge recognition capabilities in the view of a wider choice of sectors in which for ASR4ATC to operate.

2. Cost Efficiency Performance

Improvements to ASR4ATC could further reduce ATCO workload, improving overall capacity and throughput of ATM infrastructure of an airport, improving Cost Efficiency of the ATM platform.

3. Human Performance

With the current ASR4ATC configuration, the data collected during the simulation does not indicate an enhancement in some of the Human Performance aspects investigated; however, the following recommendations based on feedback from controllers are expected to generate a significant improvement.

Workload:

- ATCOs recommend enriching the type and number of ATC commands to be integrated in ASR4ATC, by automatically recognising and executing commands. Voice operated commands were considered as beneficial and expected to provide further support to workload and situational awareness, providing that the system latency and error rates are improved.
- Integration of other functions such as "assume" and "transfer" would rimprove workload.

Situational Awareness

- Changing the background colours of the ASR pop-up window could increase situational awareness.
- A more marked highlight of the ‘Hooked’ a/c on the HMI would also help improve SA.
- Integrating other sub-functions into the ‘Hook’ function such as highlighting an aircraft when the ATCOs communicate with it would improve its effectiveness.
- The implementation of an alert in case of ASR malfunction would also improve Situational Awareness.

Usability:

- Better knowledge of a preferable ASR activation means, since some ATCOs preferred to have their hands free and use the pedal, while some others found the pedal a bit outdated.
- Introduction of an ‘ASR diagnostics window’ which would display logs and transcripts on requests, always in the same place, similar to a chat window, in order for ATCOs to inspect ASR operation when and if necessary.
- Introduction of an ‘ASR disable’ function in case ASR malfunctions.
- Better knowledge of the implementation of the ‘partial recognition function’, since some ATCOs were concerned that it might send incorrect information to an aircraft of a wrongly recognised command.
- Avoiding to overlap the ASR system and the radio frequency if the aircraft use CPDLC.
- Further investigation of the manual command acknowledge function, since some ATCOs said the function causes additional unnecessary workload, while some were happy about having to acknowledge a command

Acceptance & Job Satisfaction

-
- Another proposal is to foresee dedicated ASR training for ATCOs, who need to be aware of the tool *behaviours* to optimize use for more effective performance; moreover, ASR would also require ATCOs to better conform to standard ATC ICAO phraseology, and to use a dedicated subset of ATC commands (as in the present validation exercise).

ASR operational target

- During the debriefing, ATCOs suggested to extend the use of ASR by the recognition of pilots’ communication, which would also act as additional safety net to mitigate the risk of potential mismatches/ misunderstandings ground-air.

4. Safety

As discussed in the recommendations section above, under ‘ASR Operational Target’, ATCOs suggested to extend the use of ASR by the recognition of pilots’ communication, which would also act as additional safety net.

No other direct safety recommendations have been identified, however, most improvements in human performance listed above also have an indirect yet essential contribution to Safety.

Appendix B Technological Validation Exercise #02 Report

B.1 Summary of the Technological Validation Exercise #02 Plan

In Exe-002 DLR together with Austro Control supported by CCL (Human Factors) and Integra (Safety) use automatic speech recognition and understanding (ASRU) technology to support Approach controllers in their tasks. **The ASRU technology is integrated in the NARSIM operational platform, including the operational communication system between controller and pilot.** The communication from pilot to the ATCO is not sent to ASRU.

Three use cases are addressed in the exercise. The description of the UCs is extracted from PJ.10-W2-96 ASR Technical Specification:

- UC.2 **Highlight of callsigns on the CWP from controller's utterances**. The illumination of the callsign coming from a controller utterance provides a safety check to the controller that will be able to detect if there is a difference between the callsign mentioned and the flight in the CWP where the command is being introduced. The ATCO will **quickly identify** the callsign to which they are speaking, immediately knowing whether ASRU will be able to update the spoken commands in the radar label or whether a manual input will be needed.
- UC.3 **Annotation of controller's commands**. A historical annotation of the provided commands will be available to the controller for consultation. This will, at the same time, increase controllers' situational awareness and provide a safety check as they will be able to check the instructions given to the flights. This UC is an intermediate step before the semiautomatic/automatic input of commands in the CWP using ASR.
- UC.4 **Prefilling of commands in the CWP.** In this use case ATCOs will be presented with the recognized command types together with the command values in the CWP. The ATCOs will then be able to accept/reject or manually correct the commands.

The exercise follows two complementary approaches:

The first approach seeks the operational feedback from controllers by means of a Real Time (Human-in-the-Loop) Simulation. ATCOs will control a sector performing their tasks as usual with and without the automatic speech recognition system enabled. Simulation Pilots will manage flights and interact via voice with the controllers. **Subjective feedback will be gathered by means of questionnaires**, **debriefings and observations. Objective data regarding system performance will be recorded** (e.g. flown trajectory length and command recognition rates). The real time exercise will take place at DLR's premises in Braunschweig.

The second approach will obtain statistically significant objective data regarding the ASR performance from offline evaluation of data from ops room and lab environment of Vienna airspace.

Intermediate steps have been scheduled to support the evolution of the concept and prototype. A dry-run with an early prototype was performed in December 2021 and March 2022. Another final check with a two run is planned for

The outcome is being used to improve the approach.

The exercise aims to reach a TRL-6 maturity as a **pre-industrial protype** will be integrated in an operational platform, including the operational communication system, and operational recordings will be used.

This exercise addresses several **Validation Targets**:

- **Human Performance** (increase of situational awareness, decrease of workload)
- **Safety** (due to double check and increase of situational awareness)
- **Controllers productivity** (related to the workload reduction associated to the early identification of contacting aircraft)

B.2 Technological Validation Exercise #02 description and scope

The whole validation trials of exercise 002d consisted of six days, in which 12 different ATCOs from Austro Control participated, one female and 11 male ones. The basic setup is shown in Figure B-48.

The first ATCOs participated on September, the 14th and 15th in the exercises, which resulted in some modifications of the scenario setups, so that sometimes only the last ten ATCOs are considered for the evaluation. 8 ATCOs performed the exercises from 11th to 14th of October and the last two ones on 3rd of November 2022. 13th of October and 3rd of November were also open days, i.e., open for interested stakeholders.

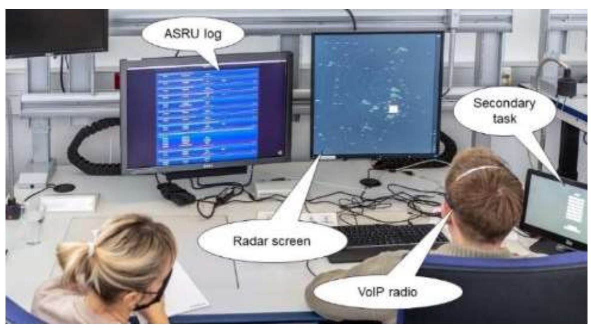

Figure B-48 – Validation setup for exercise 002 at DLR Braunschweig.

Each ATCO participated in four runs lasting 35 minutes each (except the first two ATCO runs lasted 40 minutes each). Each ATCO conducted two runs with a medium scenario and each ATCO also participated in a heavy scenario. The heavy scenario consists of 42 arrivals per hours and the medium one of roughly 30 arrivals per hour. Each medium and each heavy scenario were runs twice by each ATCO, one with automatic speech recognition and understanding (ASRU) support and one without. Each ATCO started with a medium run. Half of the ATCOs started with ASRU support and the other half started without ASRU support to minimize learning effects.

After each run the ATCO filled our several questionnaires and after the last validation the ATCOs did a final questionnaire on top. An unformal debriefing with open interviews follows after the final validation run.

B.3 Summary of Exercise 2 Technological Validation Objectives and success criteria

Please refer to section 4.1.

B.4 Summary of Technological Validation Exercise #02 Validation scenarios

Each ATCO participated in four runs lasting 35 minutes each (except the first two ATCO runs lasted 40 minutes each). Each ATCO conducted two runs with a medium scenario and each ATCO also participated in a heavy scenario. The heavy scenario consists of 42 arrivals per hours and the medium one of roughly 30 arrivals per hour. Each medium and each heavy scenario were runs twice by each ATCO, one with automatic speech recognition and understanding (ASRU) support and one without. Each ATCO started with a medium run. Half of the ATCOs started with ASRU support (the solution scenarios) and the other half started without ASRU support (the baseline scenario) to minimize learning effects.

B.5 Summary Technological Validation Exercise #02 Assumptions

Identifier	Title	Description	Justification	Impact on Assessment
AS-EXE.002-01	Sequence-Effects	Sequence effects can be quantified and eliminated	In order to compare baseline and solution runs the same scenarios are used for both. Re-doing a scenario has effects on the results. Baseline or solution should also have an effect.	If assumption is valid, which seems to be the case, sequence effects can be eliminated
AS-EXE.002-02	Limited Simulation Scope	Simulation focused on the work of one approach ATCo without neighbouring ATCos	The choice of having one ATCo carrying out the approach operation had a limited impact on the operation itself, but a rather large impact on the perceived realism of the events that the ATCo was exposed to.	Medium

Table 21: Technological Validation Assumptions overview

The technique of eliminating sequence effects is described below:

Due to learning effects, the results in the second run were mostly better than in the first run of the ATCO and from second to third they also slightly improved. This averages out, because 50% started with baseline and 50% with solution run, i.e. with ABSR support.

For reducing the variance, the averages of all 12 ATCOs for the first run, the second run, third run and fourth run were calculated. The average for the heavy runs with ABSR in the beginning and the heavy runs with ABSR as number two were calculated. These two averages were used to correct the resulting answer for each question. This is the explanation, why the numbers in the following table are no multiples of one. Keep in mind the overall averages for heavy runs with ABSR, medium runs without ABSR etc. did not change at all. Only their SIGMA changes.

To clarify the approach, we show the uncorrected answers of the 12 ATCOs to the question "How mentally demanding was the task?"

1	Medium	with ASR	3

2	Heavy	with ASR	9
1	Heavy	without ASR	4
2	Medium	with ASR	7
1	Medium	without ASR	3
2	Heavy	without ASR	9
1	Heavy	with ASR	8
2	Medium	without ASR	7
3	Medium	without ASR	7
4	Medium	with ASR	5
3	Heavy	with ASR	9
4	Heavy	without ASR	8
3	Medium	with ASR	4
4	Medium	without ASR	3
3	Heavy	without ASR	9
4	Heavy	with ASR	7
5	Medium	without ASR	5
6	Medium	with ASR	3
5	Heavy	with ASR	7
6	Heavy	without ASR	6
5	Medium	with ASR	4
6	Medium	without ASR	1
5	Heavy	without ASR	8
6	Heavy	with ASR	5
7	Medium	without ASR	3
8	Medium	with ASR	4
7	Heavy	with ASR	6
8	Heavy	without ASR	5
7	Medium	with ASR	2
8	Medium	without ASR	1
7	Heavy	without ASR	5
8	Heavy	with ASR	4
9	Medium	without ASR	7
10	Medium	with ASR	5
9	Heavy	with ASR	8
10	Heavy	without ASR	8
9	Medium	with ASR	4
10	Medium	without ASR	3
9	Heavy	without ASR	9
10	Heavy	with ASR	6

11	Medium	without ASR	6
12	Medium	with ASR	3
11	Heavy	with ASR	7
12	Heavy	without ASR	7
11	Medium	with ASR	4
12	Medium	without ASR	3
11	Heavy	without ASR	7
12	Heavy	with ASR	3

The following table shows e.g. in row 1, that the average answer value was 5.83, if the first run was the medium scenario in baseline run, whereas row 6 shows, that the average value of all six ATCOs was only 2.33, when the baseline run was the second run, i.e. the baseline run was run after the solution run.[4]

run 1	Medium	without ASR	5.83
run 2	Medium	with ASR	3.83
run 1	Heavy	with ASR	7.67
run 2	Heavy	without ASR	6.33
run 1	Medium	with ASR	4.17
run 2	Medium	without ASR	2.33
run 1	Heavy	without ASR	7.83
run 2	Heavy	with ASR	5.50

All in all, we got the following correction factors for this question, which means that we subtract 0.69 from the answer for the first medium traffic run (independent of being done as baseline or as solution) and we add 0.69 to the value of the answer, if it is the second medium run. For the heavy run we only subtract 0.17 from the first run and add 0.17 to the second.

First Medium	1	-0.69
Second Medium	3	0.69
First Heavy	2	-0.17
Second Heavy	4	0.17

We see that mean values do not change, because we add 0.69 from 6 six runs and we also subtract 0.69 from the other six medium runs.

The following table shows now the answers to this question after normalization

1	Medium	with ASR	2.31

4

2	Heavy	with ASR	8.83
1	Heavy	without ASR	3.83
2	Medium	with ASR	7.69
1	Medium	without ASR	3.69
2	Heavy	without ASR	9.17
1	Heavy	with ASR	8.17
2	Medium	without ASR	7.69

3	Medium	without ASR	6.31
4	Medium	with ASR	4.31
3	Heavy	with ASR	8.83
4	Heavy	without ASR	7.83
3	Medium	with ASR	4.69
4	Medium	without ASR	3.69
3	Heavy	without ASR	9.17
4	Heavy	with ASR	7.17

5	Medium	without ASR	4.31
6	Medium	with ASR	2.31
5	Heavy	with ASR	6.83
6	Heavy	without ASR	5.83
5	Medium	with ASR	4.69
6	Medium	without ASR	1.69
5	Heavy	without ASR	8.17
6	Heavy	with ASR	5.17

7	Medium	without ASR	2.31
8	Medium	with ASR	3.31
7	Heavy	with ASR	5.83
8	Heavy	without ASR	4.83
7	Medium	with ASR	2.69
8	Medium	without ASR	1.69
7	Heavy	without ASR	5.17
8	Heavy	with ASR	4.17

9	Medium	without ASR	6.31
10	Medium	with ASR	4.31
9	Heavy	with ASR	7.83
10	Heavy	without ASR	7.83
9	Medium	with ASR	4.69
10	Medium	without ASR	3.69
9	Heavy	without ASR	9.17
10	Heavy	with ASR	6.17

EUROPEAN PARTNERSHIP

Co-funded by the European Union

11	Medium	without ASR	5.31
12	Medium	with ASR	2.31
11	Heavy	with ASR	6.83
12	Heavy	without ASR	6.83
11	Medium	with ASR	4.69
12	Medium	without ASR	3.69
11	Heavy	without ASR	7.17
12	Heavy	with ASR	3.17

After this normalization of the answers, a paired T-Test is performed assuming a normal distribution of the answer, which is not fully true, but a good approximation.

B.6 Deviation from the planned activities

There were no deviations from the planned activities.

B.7 Technological Validation Exercise #02 Validation Results

B.7.1 Summary of Technological Validation Exercise #02 Results

All objectives were found to be validated as OK (see section 4.2) apart from the Fuel Efficiency.

1. Results on technological feasibility

The ABSR system was technically feasible and worked without malfunction throughout the 48 simulation runs.

2. Results per KPA

In the following, we show the subjective and objective results per KPA.

B.7.2 Analysis of Exercise 2 Results per Technological Validation objective

1. OBJ-10.96-TRL6-TVALP Results

- **Workload**

  Significance of NASA TLX

The used NASA TLX was performed with the following questions:

- How mentally demanding was the task?
- How physically demanding was the task?

- How hurried or rushed was the pace of the task?
- How successful were you in accomplishing what you were asked to do?
- How hard did you have to work to accomplish your level of performance?
- How insecure, discouraged, irritated, stressed, and annoyed were you?]

For each of the questions an integer value between 1 (low) and 10 (high) was possible.

We got results – after compensation of sequence effects - for only the medium runs, only the heavy runs and for all runs together.

We performed paired T-Test for all medium runs. The following table shows the results for the medium traffic scenarios:

	Mental	Physical	Stressful	Successful	Effort	Annoyed
Average	-0.20	-0.48	-0.16	-0.38	-0.75	-0.43
Sigma	1.04	0.99	1.23	0.81	1.22	1.63
With ASRU better	possible	possible	possible	possible	possible	possible
min α	**no result**	**no result**	**no result**	**no result**	**no result**	**no result**
Without ASRU better	possible	reject	possible	reject	reject	possible
min α	**25.7%**	**5.4%**	**33.1%**	**5.6%**	**2.2%**	**18.4%**
difference boundries		-0.9		-0.7	-1.2	
for α=10%		-0.1		-0.1	-0.3	

Table 22: Statistical Significance of NASA TLX for medium traffic scenarios

The average value of the differences (solution minus baseline) is minus 0.2 with a sigma of 1.04 for the 12 measurements. Minus 0.2 means that the mental demand is slightly less, when the ATCO is supported by speech recognition. Due to the high sigma, this result is not statistically significant. We only get a p-value (α) of 25.7, i.e. if we would have repeated the experiments with the 12 ATCOs 1000 times, in 257 cases we can expect that the mental workload is weighted higher in the solution runs. Therefore, the result is that we cannot assume, that baseline or solution run result in less mental workload. We should assume equal mental demand.

For the physical demand, we get a completely different result. The average difference between solution and baseline run is minus 0.48, i.e. the ATCO feel that the physical demand is reduced by using ASRU support by 0.48 with a sigma of 0.99. The hypothesis that “Without ASRU the physical demand is less” can be rejected. The statistical error rate is only 5.4%. Assuming an α (p-value) of 10%, we can expect in 90% of the cases (100% - α) a difference in the interval of [-0.9..-0.1].

The following table shows the results for the heavy traffic scenarios

	Mental	Physical	Stressful	Successful	Effort	Annoyed
Average	-0.50	-1.08	-0.33	0.08	-0.75	0.08
Sigma	1.77	1.60	1.36	2.10	1.23	2.12
With ASRU better	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	44.6%	no result	44.7%
Without ASRU better	possible	reject	possible	possible	reject	possible
min α	16.9%	1.4%	20.3%	no result	2.3%	no result
difference boundries		-1.7			-1.2	
for α=10%		-0.5			-0.3	

Table 23: Statistical Significance of NASA TLX for heavy traffic scenarios

The following table shows the results for all runs, i.e. for medium and heavy traffic scenarios

	Mental	Physical	Stressful	Successful	Effort	Annoyed
Average	-0.35	-0.78	-0.25	-0.15	-0.75	-0.17
Sigma	1.46	1.36	1.30	1.61	1.23	1.91
With ASRU better	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	no result	no result	no result
Without ASRU better	possible	reject	possible	possible	reject	possible
min α	12.4%	0.4%	18.0%	32.4%	0.2%	32.9%
difference boundries	error	-1.1	error	error	-1.1	error
for α=10%	error	-0.4	error	error	-0.4	error

Table 24: Statistical Significance of NASA TLX for medium plus heavy traffic scenarios

Summary of NASA TLX:

The NASA TLX shows that ASR

- statistical significantly reduces the physical demand
- statistical significantly reduces the effort
- has no negative effect on mental demand, stress level, success rate of task and annoyance of the ATCO.

Significance of Bedford Workload Scale

The Bedford Workload scale presented 10 levels

- 10 Task Unsustainable due to Workload
- 9 Workload Extremely High
- 8 Workload Very High
- 7 Workload High

- 6 Workload Moderate to High
- 5 Workload Moderate
- 4 Workload Low to Moderate
- 3 Workload Low
- 2 Workload Very Low
- 1 Workload Insignificant

It asks for the average and for the peak workload.

Medium Scenario	Average	Peak
Average	-0.45	-0.38
Sigma	0.98	0.86
With ASRU better	possible	possible
min α	no result	no result
Without ASRU better	reject	reject
min α	6.2%	7.2%
difference boundries	-0.8	-0.7
for α=10%	-0.1	0.0
Heavy Scenario	**Average**	**Peak**
Average	-0.33	-0.17
Sigma	1.14	1.39
With ASRU better	possible	possible
min α	no result	no result
Without ASRU better	possible	possible
min α	16.2%	34.1%
difference boundries		
for α=10%		
Medium + Heavy Scenario	**Average**	**Peak**
Average	-0.39	-0.27
Sigma	1.07	1.16
With ASRU better	possible	possible
min α	no result	no result
Without ASRU better	reject	possible
min α	3.9%	12.9%
difference boundries	-0.7	
for α=10%	-0.1	

Table 25: Statistical Significance of Bedford Workload Scale for the different scenarios

Summary of Bedford Workload Scale

The Bedford Workload scale shows that ASRU

- ASRU statistical significantly reduces the workload average and peak in the medium scenario.
- Has always an average positive effect on the peak and average workload (the average difference between solution and baseline runs are negative).
- has no negative effect on average and peak workload.

Significance of SASHA ATCO Questionnaire

This questionnaire consists of the following questions with respect to the previous run

1. I was xxx ahead of the traffic,
2. I xxx started to focus on a single problem or a specific aircraft,
3. there was xxx a risk of forgetting something important (such as inputting the spoken command values into the labels),
4. I was xxx able to plan and organise my work as wanted,
5. I was xxx surprised by an event I did not expect (such as an aircraft call),
6. I xxx had to search for an item of information.

For xxx, the following answer options exist:

- Never (1 or 7),
- Seldom (2 or 6),
- Sometimes (3, 5),
- Often (4, 4),
- More Often (5, 3),
- Very Often (6, 2),
- Always (7, 1,).

We map the answer to a value in the interval [1..7]. The corresponding number is shown in brackets after the values “Never” to “Always”. A small number should correspond to a good performance of the ATCO. In questions 1 and 4 the value “Always” corresponds to the best performance, whereas in the other four questions "Always" corresponds to bad performance. I question 1 and 4 we, therefore, used the red coloured values, i.e. here always corresponds to 1.[5]

The following diagram shows evaluation of the answers to these six questions, when having performed the paired t-test:

Medium Scenario	ahead traffic	Single problem	risk to forget	plan work	surprise	search
Average	-0.09	0.18	-0.55	-0.27	0.00	-0.09
Sigma	0.63	0.69	0.53	0.73	0.57	1.24
With ASRU better	possible	possible	possible	possible	possible	possible
min α	**no result**	**19.6%**	**no result**	**no result**	**no result**	**no result**
Without ASRU better	possible	possible	reject	possible	possible	possible
min α	**32.0%**	**no result**	**0.1%**	**11.4%**	**50.0%**	**40.5%**
difference boundries			-0.8			
for α=10%			-0.3			
Heavy Scenario	ahead traffic	Single problem	risk to forget	plan work	surprise	search
Average	-0.50	-0.17	-0.42	-0.33	0.00	-0.58
Sigma	0.62	0.69	1.03	0.90	1.70	0.95
With ASRU better	possible	possible	possible	possible	possible	possible
min α	**no result**	**no result**	**no result**	**no result**	**no result**	**no result**
Without ASRU better	reject	possible	reject	possible	possible	reject
min α	**0.5%**	**20.5%**	**8.8%**	**10.6%**	**no result**	**2.2%**
difference boundries	-0.7		-0.8			-0.9
for α=10%	-0.3		0.0			-0.2
Medium + Heavy Scenario	ahead traffic	Single problem	risk to forget	plan work	surprise	search
Average	-0.30	0.00	-0.48	-0.30	0.00	-0.35
Sigma	0.66	0.71	0.84	0.82	1.29	1.12
With ASRU better	possible	possible	possible	possible	possible	possible
min α	**no result**	**no result**	**no result**	**no result**	**no result**	**no result**
Without ASRU better	reject	possible	reject	reject	possible	reject
min α	**1.6%**	**no result**	**0.4%**	**4.1%**	**no result**	**7.2%**
difference boundries	-0.5		-0.7	-0.5		-0.7
for α=10%	-0.1		-0.3	-0.1		0.0

Table 26: Statistical Significance of SASHA ATCO for the different scenarios

The scaling has an advantage for the fast reader: Negative values in the average rows give hints that the ATCO performs better with ASRU support.

Summary of SASHA ACTO:

- ASRU statistical significantly helps the ATCO to be ahead of the traffic,

[5] The ATCO might be also confused by this change of the meaning of "Never" etc. Sometimes it means good performance and sometimes bad performance. Questions 1 and 4 should be reformulated, i.e. 1"I was xxx **behind** the traffic" and "I was xxx **not** able to plan and organise my work as wanted."

- ASRU has no positive and also no negative effect that the ATCO starts to focus on a single problem or a specific aircraft,
- ASRU statistical significantly reduces the risk of forgetting something important (such as inputting the spoken command values into the labels),
- ASRU statistical significantly supports the ATCO to plan and organise the work as wanted,
- ASRU has neither positive no positive effect that the ATCO might be surprised by an event which is not expected (such as an aircraft call),
- ASRU statistical significantly supports the ATCO to avoid searching for an item of information.

Significance of AIM-s

This questionnaire consists of the following questions with respect to the previous run. How much effort did it take to …

1. …prioritise tasks?
2. …identify potential conflicts?
3. …scan radar or any display?
4. …evaluate conflict resolution options against the traffic situation and conditions?
5. ...anticipate the future traffic situation?
6. ...recognise a mismatch of available data with the traffic picture?
7. ...issue timely commands?
8. ...evaluate the consequences of a plan?
9. ...manage flight data information?
10. ...recall necessary information?
11. ...anticipate team members' needs?
12. .... prioritise requests?
13. ...scan flight progress data?
14. ...access relevant aircraft or flight information?
15. ...gather and interpret information?

The answer alternatives were:

- Extreme (7)

- Very Much (6),
- Much (5),
- Some (4),
- Little (3),
- Very Little (2),
- None (1).

In brackets the numbers to which the answers were mapped to be able to calculated average values etc.

Medium Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Average	0.26	0.34	-0.07	0.20	0.68	0.72	0.00	0.56	-0.28	0.07	0.40	0.57	0.53	0.00	0.09
Sigma	1.29	1.49	1.38	1.36	1.17	1.34	1.29	1.35	1.08	1.23	1.14	1.10	1.24	1.15	1.12
With ASRU better	possible	possible	possible	possible	reject	reject	possible	reject	possible	possible	possible	reject	reject	possible	possible
min α	24.4%	22.0%	no result	30.6%	2.9%	4.5%	no result	8.3%	no result	42.6%	14.0%	4.3%	7.8%	no result	39.6%
Without ASRU better	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible
min α	no result	no result	42.9%	no result	no result	no result	49.7%	no result	19.4%	no result	no result	no result	no result	49.6%	no result
difference boundries					0.2	0.2		0.0				0.2	0.1		
for α=10%					1.1	1.3		1.1				1.0	1.0		
Heavy Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Average	-0.08	-0.67	-0.75	-0.50	0.00	0.21	-0.42	-0.17	-0.67	-0.25	-0.28	-0.42	-0.58	-0.50	-0.42
Sigma	0.86	0.97	1.27	1.03	0.94	1.13	1.23	1.01	1.54	1.06	1.10	0.75	0.92	0.65	0.63
With ASRU better	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	no result	no result	27.6%	no result	no result	no result	no result	no result	no result	no result	no result	no result
Without ASRU better	possible	reject	reject	reject	possible	possible	possible	possible	reject	possible	possible	reject	reject	reject	reject
min α	37.0%	1.3%	2.7%	5.3%	no result	no result	12.7%	28.7%	7.4%	21.1%	21.6%	3.4%	1.9%	0.7%	1.7%
difference boundries		-1.0	-1.2	-0.9					-1.3			-0.7	-0.9	-0.7	-0.7
for α=10%		-0.3	-0.3	-0.1					-0.1			-0.1	-0.2	-0.3	-0.2
Medium + Heavy Scenario	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Average	0.09	-0.16	-0.41	-0.15	0.34	0.46	-0.21	0.20	-0.47	-0.09	0.06	0.08	-0.03	-0.25	-0.17
Sigma	1.11	1.35	1.37	1.25	1.12	1.27	1.28	1.25	1.34	1.16	1.17	1.06	1.23	0.96	0.94
With ASRU better	possible	possible	possible	possible	reject	reject	possible	possible	possible	possible	possible	possible	possible	possible	possible
min α	34.8%	no result	no result	no result	7.2%	4.7%	no result	22.4%	no result	no result	41.0%	36.4%	no result	no result	no result
Without ASRU better	possible	possible	reject	possible	possible	possible	possible	possible	reject	possible	possible	possible	possible	possible	possible
min α	no result	27.8%	7.4%	28.2%	no result	no result	21.2%	no result	4.6%	35.0%	no result	no result	45.6%	10.4%	19.8%
difference boundries			-0.8		0.0	0.1			-0.8						
for α=10%			0.0		0.6	0.8			-0.1						

Table 27: Statistical Significance of AIM-s for the different scenarios

The answer "none" is always mapped to 1 and always means less effort for the ATCO. This scaling has again an advantage for the fast reader: Negative values in the average rows give hints that the ATCO performs better with ASRU support.

Summary of AIM-s:

- In the medium scenarios the ATCO sometimes feels uncomfortable with ASRU support, i.e. there is statistical significance, that without ASRU support the ATCO can
    a. better anticipate the future traffic situation,
    b. better recognise a mismatch of available data with the traffic picture,
    c. better evaluate the consequences of a plan,
    d. better prioritise requests
    e. and better scan flight progress data.

- In the heavy scenarios there for eight questions statistical significance, that the ATCO performs better with ATCO support (identify potential conflicts, scan radar or any display, evaluate conflict resolution options, manage flight data information, prioritise requests, scan flight progress data, access relevant aircraft or flight information, gather and interpret information.
- In the heavy scenario there is no statistical evidence against using ASRU.

Significance of SATI

The ATCOs were asked the following questions:

In the previous working period, I felt xxx that

- … the system was useful,
- ...the system was reliable
- ...the system worked accurately
- ...the system was understandable
- ...the system worked robustly (in difficult situations, with invalid inputs, etc.).
- ...I was confident when working with the system.

For xxx, the following answer options exist:

- Never (7),
- Seldom (6),
- Sometimes (5),
- Often (4),
- More Often (3),
- Very Often (2),
- Always (1).

In brackets the numbers to which the answers were mapped to be able to calculated average values etc.

Medium Scenario	useful	reliable	accurate	understandable	robust	confident
Average	-1.19	0.09	0.12	-0.72	-0.16	-1.21
Sigma	1.66	1.65	1.62	1.55	1.73	1.70
With ASRU better	possible	possible	possible	possible	possible	possible
min α	no result	43.4%	40.6%	no result	no result	no result
Without ASRU better	reject	possible	possible	reject	possible	reject
min α	1.8%	no result	no result	8.1%	39.0%	1.8%
difference boundries	-1.9			-1.4		-1.9
for α=10%	-0.5			-0.1		-0.5
Heavy Scenario	useful	reliable	accurate	understandable	robust	confident
Average	-2.07	-0.45	-0.60	-1.05	-1.45	-1.11
Sigma	1.28	1.23	1.16	1.16	0.99	2.15
With ASRU better	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	no result	no result	no result
Without ASRU better	reject	possible	reject	reject	reject	reject
min α	0.0%	12.0%	5.1%	0.3%	0.0%	5.1%
difference boundries	-2.6		-1.1	-1.5	-1.8	-2.0
for α=10%	-1.6		-0.1	-0.6	-1.1	-0.2
Medium + Heavy Scenario	useful	reliable	accurate	understandable	robust	confident
Average	-1.65	-0.19	-0.25	-0.89	-0.84	-1.16
Sigma	1.54	1.47	1.45	1.37	1.54	1.95
With ASRU better	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	no result	no result	no result
Without ASRU better	reject	possible	possible	reject	reject	reject
min α	0.0%	27.4%	21.2%	0.2%	0.8%	0.5%
difference boundries	-2.1			-1.3	-1.3	-1.7
for α=10%	-1.2			-0.5	-0.4	-0.6

Table 28: Statistical Significance of SATI for the different scenarios

The scaling has an advantage for the fast reader: Negative values in the average rows give hints that the ATCO performs better with ASRU support.

Summary of SATI:

- There is no statistical evidence against the hypotheses, that "the ATCO performs better with ATCO". The contrary is the case.
- Especially in the high traffic runs, there is statistical significance that
  a. the system was useful,
  b. the system worked accurately
  c. the system was understandable
  d. the system worked robustly (in difficult situations, with invalid inputs, etc.)
  e. I was confident when working with the system.

- The p-value for the hypothesis that the system was reliable was only 12%, whereas for the others we get at least 5%. Further analysis is needed here.

Significance of System Usability Scale (SUS)

The following statements were presented to the ATCOs:

- I think that I would like to use this system frequently,
- I found the system unnecessarily complex.
- I thought the system was easy to use.
- I think that I would need the support of a technical person to be able to use this system.
- I found the various functions in this system were well integrated.
- I thought there was too much inconsistency in this system.
- I would imagine that most people would learn to use this system very quickly.
- I found the system very cumbersome to use.
- I felt very confident using the system.
- I needed to learn a lot of things before I could get going with this system.

The ATCO could answer with 1, 2, 3, 4 and 5. For 50% of the question the "1" means that the ATCO "fully agrees" and "5" means that the ATCO "fully disagrees". For the red marked answers, however "1" means "fully disagree" and "5" means "fully agree", so that always a small number means that the ATCO likes the system with respect to this question.[6]

[6] The red answers should be reformulated, so that misunderstandings are avoided.

Medium Scenario	Use it	Complex	easy to use	technician	integration	inconsistent	use quickly	cumbersome	confidence	Much to Learn
Average	-1.13	-1.21	-1.50	0.22	-0.84	-0.38	-0.23	-1.20	-0.80	-0.35
Sigma	1.50	1.07	0.50	0.81	0.95	0.76	0.87	0.59	0.87	1.14
With ASRU better	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	19.7%	no result	no result	no result	no result	no result	no result
Without ASRU better	reject	reject	reject	possible	reject	reject	possible	reject	reject	possible
min α	1.4%	0.1%	0.0%	no result	0.6%	6.5%	21.0%	0.0%	0.5%	17.3%
difference boundries	-1.8	-1.7	-1.7	error	-1.2	-0.7	error	-1.5	-1.2	error
for α=10%	-0.5	-0.8	-1.3	error	-0.4	-0.1	error	-0.9	-0.4	error
Heavy Scenario	Use it	Complex	easy to use	technician	integration	inconsistent	use quickly	cumbersome	confidence	Much to Learn
Average	-1.00	-0.80	-1.00	0.70	-0.94	0.00	-0.30	-1.40	-0.70	-0.10
Sigma	1.08	0.98	0.87	0.90	1.07	0.80	0.90	1.44	1.09	0.90
With ASRU better	possible	possible	possible	reject	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	1.2%	no result	50.0%	no result	no result	no result	no result
Without ASRU better	reject	reject	reject	possible	reject	possible	possible	reject	reject	possible
min α	0.4%	1.0%	0.1%	no result	0.9%	no result	15.2%	0.3%	2.8%	36.4%
difference boundries	-1.5	-1.2	-1.4	0.3	-1.4	error	error	-2.0	-1.2	error
for α=10%	-0.5	-0.4	-0.6	1.1	-0.5	error	error	-0.8	-0.2	error
Medium + Heavy Scenario	Use it	Complex	easy to use	technician	integration	inconsistent	use quickly	cumbersome	confidence	Much to Learn
Average	-1.07	-1.01	-1.25	0.46	-0.89	-0.19	-0.26	-1.32	-0.75	-0.22
Sigma	1.31	1.05	0.75	0.89	1.01	0.80	0.89	1.17	0.99	1.03
With ASRU better	possible	possible	possible	reject	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	1.3%	no result	no result	no result	no result	no result	no result
Without ASRU better	reject	reject	reject	possible	reject	possible	reject	reject	reject	possible
min α	0.0%	0.0%	0.0%	no result	0.0%	14.8%	9.5%	0.0%	0.1%	16.9%
difference boundries	-1.4	-1.3	-1.5	0.2	-1.2	error	-0.5	-1.7	-1.0	error
for α=10%	-0.7	-0.7	-1.0	0.7	-0.6	error	0.0	-0.9	-0.5	error

Table 29: Statistical Significance of SUS for the different scenarios

The scaling has an advantage for the fast reader: Negative values in the average rows give hints that the ATCO performs better with ASRU support.

Summary of SUS (System Usability Scale):

- There is strong statistical evidence, that the ASRU improves system usability.
- For the two questions "I think that I would need the support of a technical person to be able to use this system" and "I needed to learn a lot of things before I could get going with this system", there is, however, statistical evidence against the ASRU support. More training might help here. The p-value for the hypothesis that the system was reliable was only 12%, whereas for the others we get at least 5%. Further analysis is needed here.

Significance of Controller Acceptance Rating Scale (CARS)

Here the ATCO got the following task:

Please read the descriptors and score your overall level of user acceptance experienced during the run. Please check the appropriate number. The ten options were:

- Improvement mandatory. Safe operation could not be maintained (1).
- Major Deficiencies. Safety not compromised, but system is barely controllable and only with extreme controller compensation (2).
- Major Deficiencies. Safety not compromised but system is marginally controllable. Considerable compensation is needed by the controller (3).

- Major Deficiencies. System is controllable. Some compensation is needed to maintain safe operations (4)
- Very Objectionable Deficiencies. Maintaining adequate performance requires extensive controller compensation (5)
- Moderately Objectionable Deficiencies. Considerable controller compensation to achieve adequate performance (6)
- Minor but Annoying Deficiencies. Desired performance requires moderate controller compensation (7)
- Mildly unpleasant Deficiencies. System is acceptable and minimal compensation is needed to meet desired performance. (8)
- Negligible Deficiencies. System is acceptable and compensation is not a factor to achieve desired performance. (9)
- Deficiencies are rare. System is acceptable and controller doesn't have to compensate to achieve desired performance. (10)

Here low values show bad performance, which is contrary to the previous questions, which is not a good choice.[7]

Medium Scenario	CARS
Average	-1.85
Sigma	2.64
With ASRU better	possible
min α	**no result**
Without ASRU better	reject
min α	**2.0%**
difference boundries	-3.0
for α=10%	-0.7
Heavy Scenario	CARS
Average	-1.12
Sigma	1.75
With ASRU better	possible
min α	**no result**
Without ASRU better	reject
min α	**3.7%**
difference boundries	-1.9
for α=10%	-0.3
Medium + Heavy Scenario	CARS
Average	-1.43
Sigma	2.26
With ASRU better	possible
min α	**no result**
Without ASRU better	reject
min α	**0.4%**
difference boundries	-2.1
for α=10%	-0.8

Table 30: Statistical Significance of CARS for the different scenarios

The scaling has an advantage for the fast reader: Negative values in the average rows give hints that the ATCO performs better with ASRU support.

[7] One ATCO selected here the wrong values. We deleted him completely from the evaluation.

Summary of all different questionnaires

Summary of all Questionnaire Groups

Medium Scenario	NASA-TLX	Bedford	SASHA ATCO	AIM-s	SATI	SUS	CARS
Average	-0.40	-0.41	-0.14	0.25	-0.51	-0.76	-1.85
Sigma	0.71	0.84	0.36	0.99	1.45	0.62	2.64
With ASRU better	possible	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	19.3%	no result	no result	no result
Without ASRU better	reject	reject	possible	possible	possible	reject	reject
min α	3.2%	5.2%	11.3%	no result	14.0%	0.1%	2.0%
difference boundries	-0.7	-0.7	error	error	error	-1.0	-3.0
for α=10%	-0.1	-0.1	error	error	error	-0.5	-0.7
Heavy Scenario	NASA-TLX	Bedford	SASHA ATCO	AIM-s	SATI	SUS	CARS
Average	-0.42	-0.25	-0.33	-0.38	-1.12	-0.55	-1.12
Sigma	0.98	1.11	0.65	0.60	1.00	0.59	1.75
With ASRU better	possible	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	no result	no result	no result	no result
Without ASRU better	reject	possible	reject	reject	reject	reject	reject
min α	7.7%	22.3%	4.6%	1.9%	0.1%	0.4%	3.7%
difference boundries	-0.8	error	-0.6	-0.6	-1.5	-0.8	-1.9
for α=10%	0.0	error	-0.1	-0.2	-0.7	-0.3	-0.3
Medium + Heavy Scenario	NASA-TLX	Bedford	SASHA ATCO	AIM-s	SATI	SUS	CARS
Average	-0.41	-0.33	-0.24	-0.06	-0.83	-0.65	-1.43
Sigma	0.86	0.99	0.54	0.88	1.27	0.61	2.26
With ASRU better	possible	possible	possible	possible	possible	possible	possible
min α	no result	no result	no result	no result	no result	no result	no result
Without ASRU better	reject	reject	reject	possible	reject	reject	reject
min α	1.2%	5.4%	2.0%	36.3%	0.2%	0.0%	0.4%
difference boundries	-0.6	-0.6	-0.4	error	-1.2	-0.8	-2.1
for α=10%	-0.2	-0.1	-0.1	error	-0.5	-0.5	-0.8

Table 31: Statistical Significance for the different questionnaires for the different scenarios

Significance of Safety Relevant Questions

In this part we evaluate only the questions which are more relevant for the safety. They are shown again in the following. All values are scaled to 10. “1” is safe and “10” is unsafe. This scaling has an advantage for the fast reader: Negative values in the average rows give hints that the ATCO performs better with ASRU support.

Medium Scenario	Annoyed	Peak	Sing Problem	Risk to forget	4	8	Reliable	Confident	Complex	CARS	Sum
Average	-0.43	-0.38	0.26	-0.78	0.29	0.80	0.12	-1.74	-2.42	-1.85	-0.49
Sigma	1.63	0.86	0.99	0.76	1.94	1.93	2.35	2.43	2.14	2.64	0.85
With ASRU better	possible	possible	possible	possible	possible	reject	possible	possible	possible	possible	possible
min α	no result	no result	**19.6%**	no result	**30.6%**	**8.3%**	**43.4%**	no result	no result	no result	no result
Without ASRU better	possible	reject	possible	reject	possible	possible	possible	reject	reject	reject	reject
min α	**18.4%**	**7.2%**	no result	**0.1%**	no result	no result	no result	**1.8%**	**0.1%**	**2.0%**	**3.0%**
difference boundries	error	-0.7	error	-1.1	error	0.1	error	-2.8	-3.3	-3.0	-0.8
for α=10%	error	0.0	error	-0.5	error	1.5	error	-0.7	-1.5	-0.7	-0.2
Heavy Scenario											
Average	0.17	-0.18	-0.13	-0.46	-0.74	-0.18	-0.65	-1.48	-1.62	-0.32	-0.63
Sigma	2.00	1.41	0.91	1.39	1.49	1.39	1.75	3.11	1.95	2.75	0.94
With ASRU better	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible
min α	**38.6%**	no result	no result	no result	no result	no result	no result	no result	no result	no result	no result
Without ASRU better	possible	possible	possible	possible	reject	possible	possible	reject	reject	possible	reject
min α	no result	**32.9%**	**31.4%**	**13.1%**	**5.0%**	**33.1%**	**11.7%**	**6.5%**	**0.9%**	**35.8%**	**1.5%**
difference boundries	error	error	error	error	-1.3	error	error	-2.7	-2.4	error	-1.0
for α=10%	error	error	error	error	-0.2	error	error	-0.2	-0.8	error	-0.3
Medium + Heavy Scenario											
Average	-0.13	-0.28	0.06	-0.61	-0.23	0.31	-0.28	-1.60	-2.02	-1.09	-0.56
Sigma	1.85	1.17	0.97	1.15	1.80	1.75	2.10	2.81	2.09	2.80	0.90
With ASRU better	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible	possible
min α	no result	no result	**39.1%**	no result	no result	**19.6%**	no result	no result	no result	no result	no result
Without ASRU better	possible	possible	possible	reject	possible	possible	possible	reject	reject	reject	reject
min α	**36.5%**	**12.4%**	no result	**0.7%**	**27.2%**	no result	**27.2%**	**0.6%**	**0.0%**	**4.5%**	**0.2%**
difference boundries	error	error	error	-0.9	error	error	error	-2.4	-2.6	-1.9	-0.8
for α=10%	error	error	error	-0.3	error	error	error	-0.8	-1.4	-0.3	-0.3

Table 32: Statistical Significance for the safety relevant questions for the different scenarios

Only the question “In the previous run, how much effort did it take to evaluate the consequences of a plan?” was evaluated significantly worse in the medium runs. Further analysis is also needed here. All in all, however, we did not find statistical evidence that using ASRU support is not as least as safe as working without ASRU support. We scale all questions by the same weight, we even get very, very high statistical significance (p-value of 0.2%) that ASRU support increases ATM safety.

Final Validation Questionnaire

The following questionnaire was asked after all sessions.

1. My situational awareness is maintained at acceptable level with Automatic Speech Recognition and Understanding (ASRU)
2. I see many safety related issues to be solved regarding Automatic Speech Recognition and Understanding (ASRU) implementation.
3. think that ASRU supports me in maintaining workload at acceptable level
4. I think that ASRU supports me in maintaining an adequate level of situation awareness
5. I think that ASR did increase the potential for human errors
6. The callsign recognition rates (and error rates) by ASRU were at an acceptable level
7. The command recognition rates (and error rates) by ASRU were at an acceptable level.

8. Overall, the level and quality of information provided by the system (in the radar labels) were an acceptable level
9. I think that the ASRU system is adequately usable
10. I would accept such an ASRU system in my future approach CWP.
11. My trust in the ASRU system is at an acceptable level
12. I can apply operating methods in an accurate, efficient, and timely manner with ASRU.
13. I think that operating methods are clearly identified and consistent in all operating conditions.
14. Overall, I was satisfied performing my task with ASRU.
15. The ASRU tool interface (HMI) provides suitable access to relevant information in all situations.
16. The ASR tool interface (HMI) does not display any non-essential information (clutter).
17. The ASRU tool display is both comprehensible and acceptable.
18. The timeliness of the ASRU tool display is within acceptable limits.
19. Procedures and operating methods are acceptable when using the ASRU tool.
20. There are no changes needed to current working methods/procedures to fully support the use of ASRU tool.
21. The ASRU tool would be operationally acceptable under either nominal or non-nominal conditions (e.g., bad weather conditions).
22. I think ASRU support can lead to reduced flight delays.
23. I think ASRU support can lead to reduce/balanced ATCo workload.
24. With Automatic Speech Recognition and Understanding (ASRU) highlighting of aircraft callsigns in the radar labels technically worked well.
25. With Automatic Speech Recognition and Understanding (ASRU) highlighting of aircraft callsigns in the radar labels supported me in quickly recognizing, which aircraft callsign has been recognized.

The ATCOs could provide answers between 1 and 10. 1 means “fully disagree” and 10 means “fully agree”. If the ATCO supports the ideas of an ASRU, in most of the cases the ATCO has to click on 10. However, for the red marked questions, (s)he has to click on 1. Here, we got a big sigma.

	1	2	3	4	5	6	7	8	9	10	11	12
	SA level	Safety Issue:	Reduce WL	Better SA	Human Errors	csg recogn	cmd recogn	Info Access	Usable	Accept ASRU	Trust	Efficient W O
Average	8.9	6.1	7.9	8.0	8.0	9.4	8.4	8.8	7.9	7.8	8.3	8.6
Sigma	1.2	2.2	1.9	1.9	1.2	0.8	1.3	1.2	1.2	1.7	1.0	0.8

Co-funded by the European Union

Table 33: Sigma and Average to Post-Validation questionnaires, part 1

	13	14	15	16	17	18	19	20	21	22	23	24	25
	Ops Concept	Satisfied	Interface	No Clutter	Compreh	timely	Op Method	No change	Bad Weathe	Less Delay	Balance WL	Csgn Highl	Csgn fast
Average	8.3	8.3	7.8	7.9	7.7	8.5	8.5	7.3	7.6	4.1	7.5	9.3	7.9
Sigma	1.2	1.4	1.3	1.7	1.3	1.1	0.9	2.5	1.8	1.9	1.6	0.8	1.7

Table 34: Sigma and Average to Post-Validation questionnaires, part 1

Text-To-Concept Accuracy

The following tables and figures present the online results of speech-to-text quality and text-to-concept accuracy per run and on command type level, respectively.

ATCo-ID	Density	Run	WER	RecogRate	ErrRate	RejRate	AllWordsCnt
1	Medium	Solution	5.83%	82.30%	5.34%	13.10%	2616
1	Heavy	Solution	7.75%	77.50%	6.02%	17.30%	3129
1	Heavy	Baseline	9.25%	69.50%	7.41%	23.90%	3034
1	Medium	Baseline	9.32%	79.10%	7.24%	14.70%	2533
2	Heavy	Solution	6.25%	84.10%	6.91%	11.30%	2824
2	Medium	Solution	5.95%	87.30%	7.10%	6.21%	2392
2	Heavy	Baseline	3.63%	92.70%	2.07%	5.44%	2805
2	Medium	Baseline	6.10%	88.80%	5.76%	7.12%	2190
3	Heavy	Baseline	2.77%	91.40%	4.02%	4.83%	2506
3	Heavy	Solution	2.54%	93.70%	1.83%	4.97%	2514
3	Medium	Baseline	1.52%	97.20%	0.94%	2.19%	2296
3	Medium	Solution	1.84%	97.80%	1.55%	0.93%	2212
4	Heavy	Baseline	2.64%	92.30%	2.67%	6.67%	2753
4	Heavy	Solution	2.57%	95.80%	2.79%	3.34%	2538
4	Medium	Baseline	2.60%	89.70%	4.50%	13.20%	2286
4	Medium	Solution	2.41%	94.50%	1.45%	5.82%	1960
5	Heavy	Baseline	5.51%	96.30%	2.56%	1.42%	2362
5	Heavy	Solution	5.16%	97.20%	1.69%	1.40%	2423
5	Medium	Baseline	1.29%	98.20%	1.06%	0.71%	1926
5	Medium	Solution	1.43%	97.60%	1.36%	1.36%	1959
6	Heavy	Baseline	2.72%	90.70%	2.03%	7.54%	2362
6	Heavy	Solution	3.80%	90.30%	0.83%	9.39%	2460
6	Medium	Baseline	4.13%	85%	1.39%	14.60%	1949
6	Medium	Solution	4.53%	84.90%	2.70%	13%	1254
7	Heavy	Baseline	3.46%	93.60%	0.50%	5.94%	2730
7	Heavy	Solution	2.28%	93%	2.11%	4.93%	2822
7	Medium	Baseline	4.59%	86.70%	3.06%	10.30%	2577
7	Medium	Solution	2.80%	94.30%	1.15%	4.58%	2396
8	Heavy	Baseline	1.52%	97.30%	0.80%	2.13%	2454

8	Heavy	Solution	1.25%	94.80%	1.91%	4.63%	2396
8	Medium	Baseline	1.22%	98.30%	1.01%	1.35%	1982
8	Medium	Solution	1.49%	95.30%	1.56%	3.75%	2144
9	Heavy	Baseline	3.43%	93.60%	3.05%	3.31%	2674
9	Heavy	Solution	2.88%	93.30%	1.92%	5.53%	2845
9	Medium	Baseline	2.54%	95%	1.86%	3.10%	2359
9	Medium	Solution	2.44%	94.60%	1.60%	4.15%	2120
10	Heavy	Baseline	1.18%	97.90%	0.93%	1.40%	3028
10	Heavy	Solution	0.70%	99.30%	0.24%	0.48%	2989
10	Medium	Baseline	0.46%	99.40%	0.30%	0.30%	2398
10	Medium	Solution	0.56%	99.70%	0%	0.29%	2507
11	Heavy	Baseline	0.75%	99.40%	0.55%	0.28%	2399
11	Heavy	Solution	0.31%	98.20%	1.02%	1.28%	2605
11	Medium	Baseline	0.27%	97.50%	0.95%	1.90%	2258
11	Medium	Solution	1.44%	96.30%	0.62%	3.72%	2279
12	Heavy	Baseline	1.91%	94.80%	1.04%	4.18%	2754
12	Heavy	Solution	2.59%	94.10%	1.33%	4.52%	2725
12	Medium	Baseline	1.56%	97.50%	0%	2.51%	2675
12	Medium	Solution	1.78%	97.80%	1.40%	1.12%	2681

Table 35: Text-To-Concept Accuracy for Radar Labels regarding the European Ontology for ATC utterance annotation.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

	LabelCell
ATCO AFFIRM	none
ATCO ALTITUDE	Alti
ATCO CALL_YOU_BACK	none
ATCO CLEARED ILS	Cleared
ATCO CLEARED VISUAL	Cleared
ATCO CLIMB	Alti
ATCO CONTACT	Cleared
ATCO CONTACT_FREQUENCY	Cleared
ATCO CONTINUE PRESENT_HEADING	Heading
ATCO CORRECTION	none
ATCO DESCEND	Alti
ATCO DIRECT_TO	Waypoint
ATCO DISREGARD	none
ATCO EXPECT ILS	none
ATCO EXPECT RNAV	none
ATCO EXPECT RUNWAY	none
ATCO FAREWELL	none
ATCO GREETING	none
ATCO HEADING	Heading
ATCO HOLDING	none
ATCO INCREASE	Speed
ATCO INFORMATION ACTIVE_RWY	none
ATCO INFORMATION ATIS	none
ATCO INFORMATION MISCELLANEOUS	none
ATCO INFORMATION QNH	none
ATCO INFORMATION TRAFFIC	none
ATCO INIT_RESPONSE	none
ATCO INTERCEPT_LOCALIZER	Cleared
ATCO MAINTAIN HEADING	Heading
ATCO MAINTAIN PRESENT_SPEED	Speed
ATCO MAINTAIN SPEED	Speed
ATCO NAVIGATION_OWN	WP-clear
ATCO NEGATIVE	none
ATCO NO_CONCEPT	none
ATCO NO_SPEED_RESTRICTIONS	Speed-Cl
ATCO RATE_OF_DESCENT	Rate
ATCO RATE_OF_DESCENT EXPEDITE	none
ATCO RATE_OF_DESCENT OWN	Rate-Cl
ATCO REDUCE	Speed
ATCO REDUCE_MIN_APPROACH_SPEED	Speed-Cl
ATCO REPORT_MISCELLANEOUS	none
ATCO REPORT_NOW ESTABLISHED	none
ATCO REPORT_NOW ESTABLISHED_ILS	none
ATCO REPORT_NOW HEADING	none
ATCO REPORT_NOW SPEED	none
ATCO RESUME_NORMAL_SPEED	Speed-Cl
ATCO SAY_AGAIN	none
ATCO SPEED	Speed
ATCO STATION	none
ATCO STOP_DESCEND	Alti
ATCO TRANSITION	Trans
ATCO TURN	none
ATCO TURN_BY	Heading
ATCO UNRECOGNIZED_CONCEPT	none
ATCO VERTICAL_RATE	Rate
ATCO VERTICAL_RATE OWN	Rate-Cl
ATCO VFR_CLEARANCE	none

Table 36: Assignment of command types to radar label cell “command type groups”.

	LabelCell	Rec-Rate	Err-Rate
ATCO ALTITUDE	Alti	73.0%	9.5%
ATCO CLEARED ILS	Cleared	91.4%	0.7%
ATCO CONTACT	Cleared	97.3%	0.0%
ATCO CONTACT_FREQUENCY	Cleared	88.4%	4.1%
ATCO CONTINUE PRESENT_HEADING	Heading	93.2%	0.0%
ATCO DESCEND	Alti	96.4%	0.8%
ATCO DIRECT_TO	Waypoint	87.6%	5.0%
ATCO HEADING	Heading	94.5%	0.8%
ATCO MAINTAIN SPEED	Speed	93.7%	0.0%
ATCO NO_SPEED_RESTRICTIONS	Speed-Cl	77.9%	8.8%
ATCO RATE_OF_DESCENT	Rate	75.2%	1.0%
ATCO RATE_OF_DESCENT OWN	Rate-Cl	71.4%	0.0%
ATCO REDUCE	Speed	93.1%	0.1%
ATCO RESUME_NORMAL_SPEED	Speed-Cl	86.7%	0.0%
ATCO SPEED	Speed	89.0%	1.8%
ATCO TRANSITION	Trans	86.6%	0.0%
ATCO VERTICAL_RATE	Rate	76.9%	11.5%

Table 37: The 17 relevant command types (shown in radar labels) with their recognition and error rate (red if below 85%).

	Rec-Rate	Err-Rate
Alti	95.9%	1.0%
Speed	90.9%	0.9%
Speed-Cl	81.2%	3.9%
Heading	94.2%	1.0%
Waypoint	87.6%	5.0%
Trans	86.6%	0.0%
Rate	75.6%	3.1%
Rate-Cl	72.1%	0.0%
Cleared	92.1%	1.6%
none	82.8%	2.7%
Rad Label	92.3%	1.5%

Table 38: Command Recognition Rates and Command Error Rates per Command Type Group (as grouped in Radar Label).

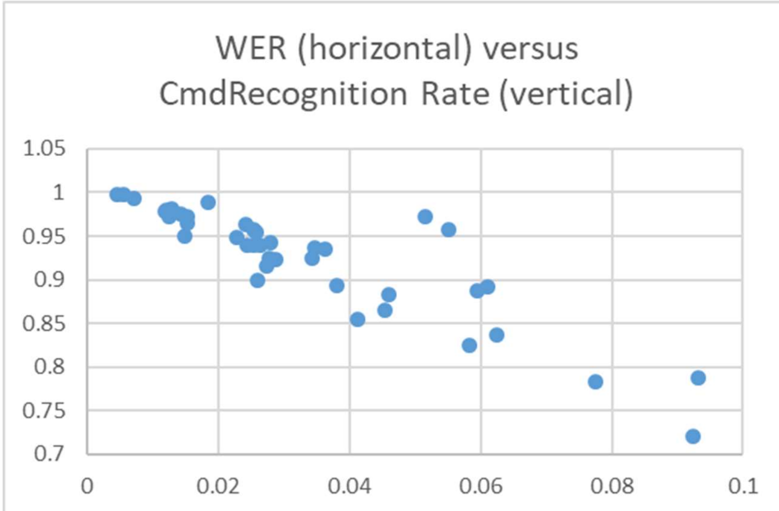

Figure 49 – Relation of word error rate and command recognition rate.

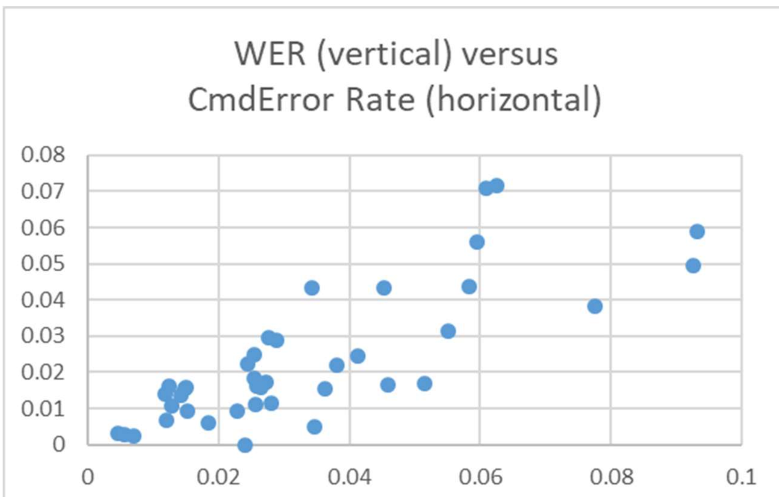

Figure 50 – Relation of word error rate and command recognition error rate.

B.7.3 Unexpected Behaviours/Results

There was no unexpected behaviour of the validation platform or any of the ATCos. However, the word error rates of the ASR engine were slightly higher in the online-mode (as experienced by ATCos) than in the later offline analysis of recorded audio files (worse speech-to-text of course also led to worse text-to-concepts).

B.7.4 Confidence in Results of Validation Exercise 002

1. Level of significance/limitations of Technological Validation Exercise Results

When considering the results of the TRL6 validation exercise 002 several factors should be considered:

- The validation is based on real-time simulation environment addressing speech recognition and understanding in an ATC center environment for Vienna approach. All displays are prototypic DLR development. Functionalities to great extent replicate the operational functions they also differ from the ones ATCo's are used to.

- Realism of the pseudo-pilot workload and task-load is not comparable to the workload of pilots in the real operational environment. The task-load of pilots in real life busy TMA environment is distributed among the flight deck crew in a more operationally-focused manner.

- In the real-time simulation, a single pseudo-pilot is responsible for keeping up with ATC instructions to numerous a/c on a single frequency very often provided in a very short time-frame with its associated limitations: reduced realism and increased opportunity of errors and omissions in particularly busy scenario.

- Only nominal situations were addressed (no bird strike or emergency landing).

- TMA ATCos participated in the TRL6 validation exercise. Therefore, it can be assumed that the results will be valid for all TMA ATCos.

- The participants with pseudo-pilot role were ATC experts from DLR - which in turn had to learn the new role of work.

2. Quality of Technological Validation Exercises Results

The quality of the validation results is determined as high due to the following:

- Experienced ATCos with appropriate ratings participated in the validation exercise.

- Unexperienced ATC experts participated in the role of pseudo-pilots, which learned and used the pseudo-pilot HMI without some difficulties. Their operational knowledge and the phraseology contributed to the quality of the results.

- The ATCos which participated in the exercise were not involved in the project in terms of participation of previous work-sessions. The participating ATCos and system engineers contributed to the developmental process in account of the validated OIs in line with real-life operational needs.

3. Significance of Technological Validation Exercises Results

The ASR recognition and error rates of the ABSR system as well as the subjective and objective measurements base on a lot of utterances/answers. Hence, these numbers promise to have statistical significance even given just 12 study participants.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

B.8 Conclusions

This section explains conclusions derived from the results detailed above and gives an outlook on future research and development work for assistant based speech recognition supported by artificial intelligence/machine learning in an en-route / TMA environment.

B.8.1 Conclusions on Technological feasibility

The ABSR technology has shown to be feasible in an ATC center environment. However, a list of recommendations on how to enhance aspects of the ASR system (and the general prototypic CWP environment) have been made. Very promising recognition rates for callsigns of 98% and for commands of 93% with error rates for callsigns of below 1% and for commands below 2.5% are possible to achieve. The quantitative and qualitative feedback of ATCos was good and motivating to go beyond TRL6.

B.8.2 Conclusions on performance assessments

In general, ATCos were able to perform their ATC tasks (even given the CWP prototypic systems) when working with ABSR support. The positive results for system usability, job satisfaction, safety questions, situation awareness, and workload measurements show the potential of ABSR in a center environment – even if some other measurements do not show any significant differences between baseline and solution. To summarize, EXE-002 has shown great potential of using the output of an ABSR system.

B.8.3 Recommendations

The amount of training data must be further improved given representative samples, i.e., the portion of female voices in ATC communication data is much less than of male. Furthermore, a big amount of data must be recorded from operations rooms (not from labs), because ATCos speak different in simulations. The success criterion “The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises” should be re-evaluated as 2.5% might be too small, i.e., it might also be acceptable if command type 1 is recognized with 97% accuracy and command type 2 with 88% accuracy.

The European-wide agreed ontology for annotation of ATC utterances as used and enhanced in this exercise should be further exploited. The continuous mutual enhancements of the ontology in the ASR projects HAAWAII (as successor of MALORCA), STARFiSH, Sol96, and Sol97 tremendously build a base for interoperability of systems. Following ASR activities (beyond TRL6) should, therefore, reuse the achieved (good) results and methods of such ABSR projects (e.g., in SESAR-3) instead of coming up with another very basic solution.

Appendix C Technological Validation Exercise #03 Report

C.1 Summary of the Technological Validation Exercise #03 Plan

This appendix contains the **Technological Validation Report for exercise** EXE-PJ.10-96-ASR-TRL6-03 performed by ENAIRE, Indra and Crida as part of Solution S96 Automatic Speech Recognition in En-route environment.

C.2 Technological Validation Exercise #03 description and scope

In EXE-03 ENAIRE together with Crida and Indra use automatic speech recognition, ASR, technology to support En-route controllers in their tasks. **The ASR technology is integrated in a simulation platform that contains a SACTA operational controller working position, CWP, and the operational communication system between controller and flight crew that is being deployed at ENAIRE, COMETA.**

In this exercise three use cases are addressed. The description of the UCs is extracted from PJ.10-W2-96 ASR Technical Specification.

UC.1 **Highlight of callsigns on the CWP from pilot utterances.** This use of the ASR technology will increase controller's situational awareness and reduce their workload by **quickly identifying new flights entering the sector or flight crews requesting actions from ATCOs** (e.g. trajectory change, flight level change or information)

UC.2 **Highlight of callsigns on the CWP from controller's utterances**. The illumination of the callsign coming from a controller utterance provides a safety check to the controller that will be able to detect if there is a difference between the callsign mentioned and the flight in the CWP where the command is being introduced.

UC.3 **Annotation of controller's commands**. A historical annotation of the provided commands will be available to the controller for consultation. This will, at the same time, increase controllers' situational awareness and provide a **safety check** as they will be able to review the instructions given to the flights. This UC is an **intermediate step before the semiautomatic/automatic input of commands in the CWP using ASR.**

As indicated in the TVALP [14], exercise 003 follows two complementary approaches:

The first approach sought the operational feedback from controllers by means of a Real Time Simulation. ATCOs controlled a sector performing their task as usual with and without the automatic speech recognition system enabled. Pseudopilots managed flights and interacted via voice with the controllers. **Subjective** feedback was gathered by means of **questionnaires, debriefings and observations.** Objective data regarding system performance was recorded. The real time exercise took place on November 10th and 11th 2021 at Crida's premises in Madrid.

The second approach obtained statistically significant objective data regarding the ASR performance. Operational recordings from real communications between ATCOs and flight crew coming from different Spanish sectors were processed through the ASR system to obtain accuracy on callsign identification and event annotation. The data obtained from the analysis has been compared with a golden standard manually corrected. The statistical study was performed between January and February 2022.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

The exercise can be considered as TRL-6 maturity as a **pre-industrial protype** has been integrated in an operational platform, including the operational communication system, and operational recordings have been used.

This exercise addresses several **Validation Targets**:

- **Human Performance** (increase of situational awareness)
- **Safety** (due to double check and increase of situational awareness)
- **Controllers' productivity** (related to the workload reduction associated to the early identification of contacting aircraft)

C.2.1 Real Time Simulation description and scope

The exercise simulated two sectors of Madrid FIR which has medium complexity. The sectors are Zamora-Toledo Integrated, LECMZTI, and Castejon-Zaragoza Integrated, LECMCZI, see Figure 51.

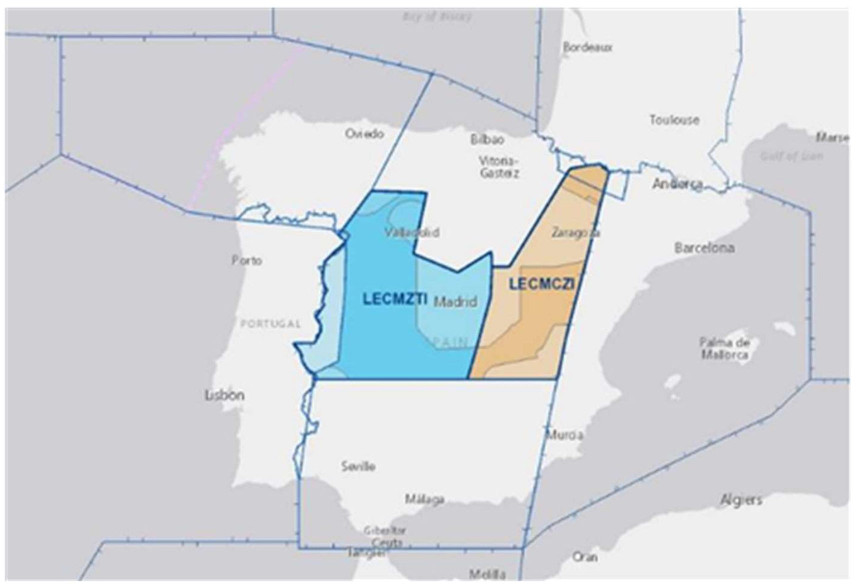

Figure 51 EXE-002 simulated sectors: LECMZTI and LECZI

This configuration is an operational configuration that is used at night. This configuration facilitates the evaluation of the validation objectives:

- The sectors have several entry points where flight crew performs their first call (related to the first UC, Highlight of callsigns on the CWP from pilot utterances).
- The sectors are quite wide and integrate nine control volumes. This implies that there are very different traffic flows that require different type of control events (related to the third UC, Annotation of controller's commands) and facilitates the creation of situations where the traffic is focused in one area or disperse along the whole sector (related to the first and second UCs, Highlight of callsigns on the CWP from pilot and controller's utterances).
- There are several airports within the control volume, the main one being Madrid- Barajas, LEMD. This airport was used in North configuration and generated traffic flows to/from both sectors.

Controllers were operational controllers from Madrid FIR, thus they were familiar with the scenario and the control rules. The control operation rules used were the operational ones with one simplification, the lower level to hand over traffic to TMAs and airport in all the volumes was the same, FL210.

Two types of exercise were used. Both had from medium to high traffic loads that supported the test of technical and operational requirements. The traffic for the exercise was created by adapting real traffic from 14th July 2019. The traffic adaptation included the modification of callsigns, flight levels and entry times. The traffic sample covered 67 different airlines plus 10 general aviation registration numbers.

The exercises were performed in an integrated controller position: one controller performs the executive and planning roles. One pseudopilot was assigned to each position. The pseudopilots have an active pilot license and have participated in previous validation activities.

Prior to the validation exercise, controllers and pseudopilots were provided with the following material:

- Participant information and agreement form.
- EXE-03 - ASR Validation Manual: that presented the exercise objectives, operational scenarios, description of prototype functionalities, and schedule of exercise activities.
- Questionnaires to be filled in (only to controllers).
- Company, and radio names of the companies in the simulation (only to pseudopilots).
- Event script per scenario type (different for controllers and pseudopilots). The event scripts for controller included different examples that they could use to test the ASR. The event scripts for pseudopilots included a list of events to be requested from controllers.

Table 39 presents the scheduled activities during the simulation execution.

	November 10th	**November 11th**
08:30 - 9:00	Welcome Presentation of ASR concept and SESAR	Presentation of scenario 2: use cases and objectives
9:00 - 10:00	Presentation of prototype functionalities Presentation of scenario 1: use cases and objectives	Scenario 2: reference run Questionnaire filling in Brief debriefing
10:00 - 10:10	Break	
10:10 - 11:20	Scenario 1: reference run Questionnaire filling in Brief debriefing	Scenario 2: first solution run Questionnaire fill- in Brief debriefing
11:20 - 11:30	Break	

11:30 - 12:40	Scenario 1: first solution run Questionnaire filling in Brief debriefing	Scenario 2: second solution run. Questionnaire fill- in Brief debriefing
12:40 - 12:50	Beak	
12:50 - 14:00	Scenario 1: second solution run Questionnaire filling in Brief debriefing	Final debriefing
14:00 - 15:00	Day summary	Final debriefing

Table 39 EXE-003 schedule

Due to some problems on the first day (see section C.6 Deviation from the planned activities), it was decided to repeat one of the scenario 1 solution runs on the second day, reducing the final debriefing time.

A very brief training was provided during the presentations the first day. Due to the familiarisation of controllers and pseudopilots with the operational scenario and consoles, and the documentation provided in advance, they considered no further training was needed**.**

C.2.2 . Statistical approach description and scope

The planned RTS in Exe-03 had some limitations i.e. the number of scheduled runs was low, the number of controllers and (pseudo)pilots was limited to two of each, and finally, the locutions that would be analysed were from a simulation environment which could impact the natural language of the speakers. To overcome these limitations a statistical approach was planned. The statistical approach includes the analysis of operational recordings from different type of sector and several actors, both controllers and flight crew.

The statistical test was performed between January and February 2022. Operational recordings fed the Voice prototype to obtain information on the callsign and event identification. The outcomes from the prototype were compared to a gold standard manually created, and rates regarding callsign and command identification have been obtained. The recordings belong to different sectors in Madrid ACC (LECM).

Table 40 presents the number of communications selected. Of the 449 communications:

- 246 were from aircraft to controllers. These are radio communications (medium to lower quality in signal to noise ratio, SNR)
- 197 are from controller to aircraft. The communication is from the jack panel (high SNR quality)
- 6 are from controller to controller. The communication is from the telephone (high SNR quality)

Canal	Radio (flight crew)	Telephone (ATCo to ACTCo)	Jack panel (ATCo)	Total
Communications	246	6	197	449

Table 40 EXE-03 Communications analysed- statistical approach

From these communications, only the ones with relevant information containing a callsign and an event were taken into account for the statistical analysis. The number of locutions finally analysed were 301.

The locutions correspond to **29 different airline companies from 18 different countries**. All the controllers were native Spanish speakers. We made the hypothesis that the flight crew native tongue was the same from the airline country (although this is not always true in reality).

C.3 Summary of Exercise 3 Technological Validation Objectives and success criteria

SESAR Solution Validation Objective	SESAR Solution Success criteria	Coverage and comments on the coverage of SESAR Solution Validation Objective in Exercise 03	Exercise Validation Objective	Exercise Success criteria
OBJ-Sol.96ASR-TRL6-TVALP-0010	CRT-Sol.96ASR-TRL6-TVALP-0010.001	Fully covered	EX03-OBJ-Sol.96ASR-TRL6-TVALP-0010 same description as OBJ-Sol.96ASR-TRL6-TVALP-0010 To assess the technical feasibility of the integration of the ASR system and its sub-systems into CWP and interoperability between the ASR sub-systems and the existing CWP systems and tools.	EX03-CRT-Sol.96ASR-TRL6-TVALP-0010.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0010.001 The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the existing CWP systems and tools. Availability of systems and tools and their

				performance remain at 100%
OBJ-Sol.96ASR-TRL6-TVALP-0020	CRT-Sol.96ASR-TRL6-TVALP-0020.001	Fully covered	EX03-OBJ-Sol.96ASR-TRL6-TVALP-0020 same description as OBJ-Sol.96ASR-TRL6-TVALP-0020 To assess the stability of the ASR system performance	EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0020.001 The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)
	CRT-Sol.96ASR-TRL6-TVALP-0020.002	Fully covered		EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0020.002 The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.
OBJ-Sol.96ASR-TRL6-TVALP-0030	CRT-Sol.96ASR-TRL6-TVALP-0030.001	Fully covered	EX03-OBJ-Sol.96ASR-TRL6-TVALP-0030 same description as OBJ-Sol.96ASR-TRL6-TVALP-0030 To assess the impact on the human performance of the integration of the ASR system and its sub-systems into operations in a realistic environment	EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.001 The introduction of the ASR system into the context of application is operationally viable, ATCos workload with ASR is equal or better than in baseline (without ASR support)

	CRT-Sol.96ASR-TRL6-TVALP-0030.002	Fully covered		EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.002 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.002 The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate)
	CRT-Sol.96ASR-TRL6-TVALP-0030.003	Fully covered		EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.003 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.003 The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. Controllers' feedback with respect to Human Factors questionnaire is better than for baseline)
	CRT-Sol.96ASR-TRL6-TVALP-0030.004 The number and/or severity of human errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact. Furthermore, more than 50% of command	Not covered		

	recognition errors and command recognition rejections are detected by the controllers and manually corrected.			
	CRT-Sol.96ASR-TRL6-TVALP-0030.005 Task allocation between human and machine, resulting from the introduction of the ASR system support, is rated as good as in baseline (with respect to feedback to Human Factors Questionnaire).	Not covered		
	CRT-Sol.96ASR-TRL6-TVALP-0030.006	Fully covered		EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.006 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.006 Changes in the design of the user interface (input devices, visual displays/output devices, alarm& alerts) support ATCos in carrying out the tasks.
	CRT-Sol.96ASR-TRL6-TVALP-0030.007	Fully covered		EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.007 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.007 The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified)

OBJ-Sol.96ASR-TRL6-TVALP-0040	CRT-Sol.96ASR-TRL6-TVALP-0040.001	Fully covered	EX03-OBJ-Sol.96ASR-TRL6-TVALP-0040 same description as OBJ-Sol.96ASR-TRL6-TVALP-0040 To assess the impact of the introduction of the ASR system on safety.	EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.001 The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.
	CRT-Sol.96ASR-TRL6-TVALP-0040.002	Fully covered		EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.002 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.002 The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.
	CRT-Sol.96ASR-TRL6-TVALP-0040.003	Fully covered		EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.003 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.003 The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.
	CRT-Sol.96ASR-TRL6-TVALP-0040.004	Fully covered		EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.004 same description as

				CRT-Sol.96ASR-TRL6-TVALP-0040.004 The level of ATCO's situational awareness is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions).
	CRT-Sol.96ASR-TRL6-TVALP-0040.005	Fully covered		EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.005 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.005 The level of ATCos' workload is maintained or decreased with the introduction of ASR system.
	CRT-Sol.96ASR-TRL6-TVALP-0040.006 The recovery means that errors resulting from the introduction of the ASR system are identified to minimise operational impact.	Not covered		
OBJ-Sol.96ASR-TRL6-TVALP-0050 To assess the impact of the introduction of the ASR system on capacity.	CRT-Sol.96ASR-TRL6-TVALP-0050.001 The workload of ATCO after introduction of an ASR system is adequate to increase TMA capacity. The workload of ATCOs is the same or less when working with ASR compared to baseline. The average flight time of the aircraft is expected to be	Not covered		

	reduced with respect to baseline without ASR support due to less time needed by the ATCO to complete task for one aircraft. ATCO should then have more time available for other aircraft and more timely task execution with regard to the flight progressing through their airspace. This should result in more optimum trajectories.			
	CRT-Sol.96ASR-TRL6-TVALP-0050.002 ASR allows ATCOs to safely manage a higher amount of aircraft, increasing the throughput in TMA	Not covered		
OBJ-Sol.96ASR-TRL6-TVALP-0060 To assess the impact of the introduction of the ASR system on Fuel efficiency	CRT-Sol.96ASR-TRL6-TVALP-0060.001 Aircraft will be able to improve their route Efficiency (fuel burnt) due to the higher throughput in TMA thanks to the introduction of ASR	Not covered		
	CRT-Sol.96ASR-TRL6-TVALP-0060.002 Aircraft will be able to improve their route Efficiency (flight time) due to the higher throughput in TMA thanks to the introduction of ASR	Not covered		
OBJ-Sol.96ASR-TRL6-TVALP-0070 To assess the impact of the introduction of the	CRT-Sol.96ASR-TRL6-TVALP-0070.001 ATCos are able to perform a faster and more predictable navigation	Not covered		

ASR in visualization navigation	when using ASR for 3D visualization			

C.4 Summary of Technological Validation Exercise #03 Validation scenarios

Next sections provide a description of the scenarios of the RTS, section C.4.1 and the statistical analysis, section C.4.4.

C.4.1 RTS Validation scenarios

The real time exercise executed two different validation scenarios:

SCENARIO 1 Description

Scenario 1 had a constant flow of aircraft that entered the controlled sectors from adjacent dependencies and neighbourhood sectors. There were traffic peaks to concatenate calls and facilitate situations where the controller was focused on one part of the sector. The traffic flight followed Instrumental flight rules, IFR. There were flights from commercial airlines and general aviation.

Next use cases were included in the scripts to promote the thorough test of the ASR functionality:

- Regarding the callsign:
  - Spell, use radio name, use company name;
  - Say numbers in units (TAP452A), tens (VLG2010), hundreds (EXY200VB), thousands (VLG3000);
  - Use “triple” (NIA6111)
  - Address a flight from an unknown (not in the ASR list) company (e.g. ACES233 – Colombian local airline);
  - Flights with very similar callsigns (TAP1135/TAP1035, RYR77MB/RHK07B;
  - Call a flight that does not enter your sector (flight plan is not recognized).

- Regarding the commands:
  - Give multiple commands in one phrase;
  - Use different commands with the same operational meaning (e.g. fly direct to, proceed direct to);
  - Use commands with mandatory and optional parts (e.g. or greater);
  - Give a command without the callsign (e.g. Controller: AEA328 do you accept flight level 380? –Pilot: Negative maximum FL360 AEA328 – Copied, maintain flight level 360;
  - Give two instructions using break break;

- Other:
  - Use Spanish and English languages;
  - ATCos using Non-Standard phraseology (e.g. “proceed by the /right/left to XX”; “proceda por la izquierda/derecha a XX”);

- Pilot answers with a wrong callsign;
- Pilots ask for: meteo information, runway information, level change, direct, alternative.

SCENARIO 2 Description

This scenario started in one sector configuration (one controller in charge of both sectors, Config R21 is used at nights with low traffic). The traffic steadily increases and the exercise leader, acting as supervisor, decides to split the sectors around minute 10. The traffic continues increasing until it starts to decrease. Near the end of the exercise, the traffic is low again and both sectors are grouped again. The exercise ends with one controller managing both sectors.

Scenario 2 allows to analyse the requirements related to sectors splitting and grouping, the rapid successive communications from different pilots, and the overlap of communication from different aircraft.

Traffic from adjacent dependencies and neighbourhood sectors contact constantly allowing the controller to see the highlight and annotation. Traffic followed Instrument Flight Rules (IFR), Visual Flight Rules (VFR) and Operational Air Traffic (OAT).

This scenario tested the same use cases as the previous ones plus the overlap of pilots calling at the same time, and successive communications.

C.4.2 RTS Reference Scenario(s)

The reference scenario are two Madrid ACC En-route sectors with adapted traffic from 2019 (pre-pandemic traffic level). The sectors were controlled by two controllers in single person operations . Figure 52 presents the sectors that were simulated. Pseudopilots followed the instructions of the controller and executed a script defined to test the ASR (e.g. request of flight level change not previously expected by the controller). These actions were also performed in the reference scenario for comparation.

The controller working position had the **ASR DISABLED** during the reference scenario simulation.

AIRSPACE AND TRAFFIC INFORMATION

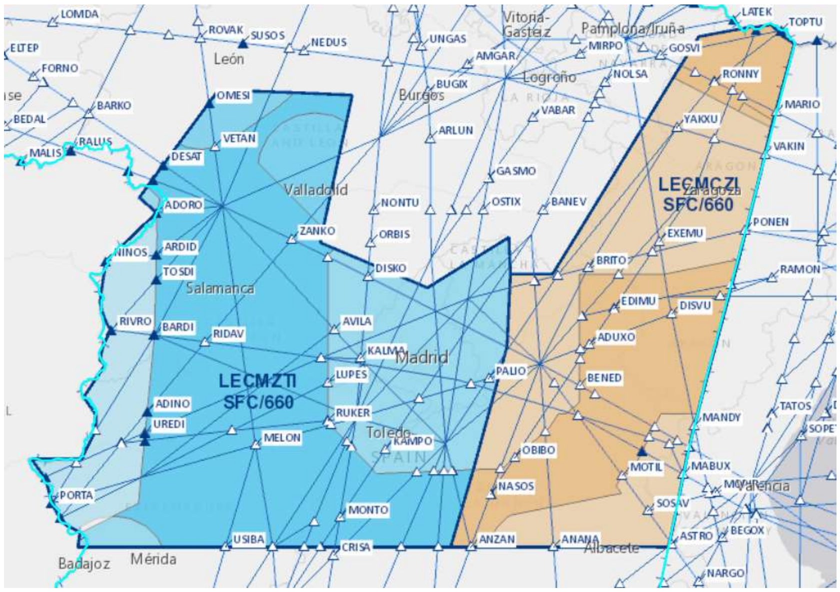

Figure 52: EXE-03 simulated sectors

Madrid ACC, Figure 52 , is a class C airspace categorized as medium density/complexity environment. Main communication means between controllers and flight crew is radio communication. Data link was not available during this simulation.

The scenario has a high percentage of the traffic evolving in the arriving/departing routes from/to Adolfo Suarez Madrid-Barajas airport to the boundary points with Seville, Lisbon and Barcelona ACCs as well with other sectors within Madrid ACC. Several SID/STARS connect the ACC routes with LEMD airport.

The overflying traffic structure has two main axes, North/South axis with traffic from North-South and vice versa, in the main traffic from/to central Europe to /from Canary Islands, and the other axis provides traffic from/to America to/from South Europe and the middle east.

The configuration selected for this exercise is within Madrid Ruta 2 airspace, Figure 52 and Table 41 , the configuration is 2A which has two En-route sectors one on the east side called CZI and the other on the west side called ZTI. The main airport of traffic destination and departure within these two groups of sectors is Madrid (LEMD).

SECTOR	VERTICAL LIMITS
CZI	SFC/FL660 except in overlapping areas
ZTI	SFC/660 except in overlapping areas

Table 41: Validation sector limits EXE-03

Control service is provided to all aircraft from FL210 to FL 660. Information service is provided from surface (SFC) to FL210 outside the TMA/airport and other control areas like airways and CTAs. TMA and airport areas are outside of the simulated sectors.

C.4.3 RTS Solution Scenario(s)

The solution scenarios of the real time simulation are the same as in the reference scenario.

The controller working position had the **ASR ENABLED** during the solution scenario simulation. It was possible to enable only controller speech recognition, pilot speech recognition or both; being the ATCOs the ones to decide. Also, the different functionalities, callsign recognition and command history window, were activated or deactivated as required to assess the three use cases separately.

Communications between controller and pilot were performed using COMETA, the communication system that is being deployed in Spain by Enaire. COMETA uses VoIP and the version used for the exercise was the latest available.

DESCRIPTION OF MAIN PROTOTYPE FUNCTIONALITIES

When a radio communication is performed the ASR is triggered. The ASR system identifies the callsign in the communication and highlights the corresponding radar track in the CWP. The ASR also extracts relevant information from controller's utterances and proceeds to annotate it in a window that the controller is able to consult.

Context information i.e. information regarding flight plans and their update was sent to the ASR prototype by the simulation platform.

As presented in Figure 53 the callsign recognition was performed by displaying a white circle around the radar track. The circle flashed for 5 seconds and then it disappeared. The functionality allowed highlighting several aircraft at the same time.

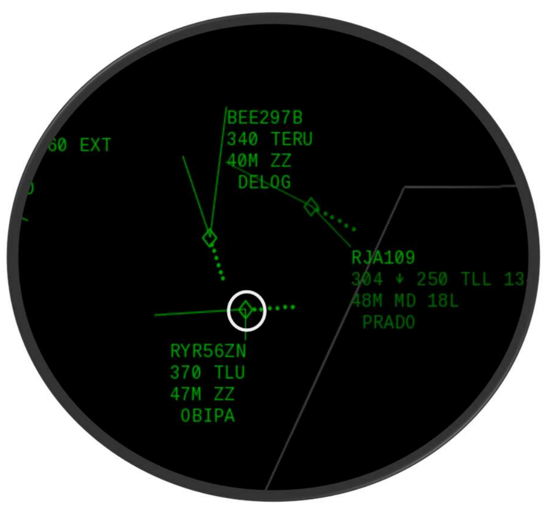

Figure 53: Callsign illumination

The annotation window, see Figure 54, contained information regarding **the commands provided by the controller**. It had the callsign of the addressed aircraft, the issuing time, the event annotation following the standard agreed within SESAR partners, and an action column. The action column presents the event information but in a coherent approach with the SACTA platform. As example, the command:

TAP1135 climb to flight level three seven zero

Is annotated following the agreed Ontology standard developed within SESAR [17] as TAP1135 CLIMB 370 FL

Is annotated following SACTA standard as CFL 370 (Cleared Flight Level 370) which is how the command is presented in other SACTA menus.

Figure 54: Annotation window

The HMI in the annotation window was coherent with SACTA HMI. In Figure 55 the flights that are in frequency with the controllers have light green colour, the flights that have just left the controller frequency have dark green colour, and no correlated flights are white. Following the feedback obtained from the dry run performed in November 2020, when the ASR detected a command addressing a flight that was not in the flight plan list/did not recognize the callsign in the phrase, the information was displayed in white.

Figure 55: Annotation window with coherent colours

The annotation functionality also allows to select one flight and display all the events exchanged with the flight. Figure 56 presents the information exchanged with the flight Aeroflot 2603 (AFL2603)

Figure 56: Flight annotation window

The ASR prototype used in the exercise is a pre-industrial prototype able to interact with an operational CWP. The pre-industrial prototype was developed by Indra and Enaire based on CRIDA's prototype used in previous phases. It included several requirements that allowed to robustly interact with an operational CWP, (see Availability Note [16]). The pre-industrial prototype covered a limited number of commands:

- Headings in grades
- Directs (to waypoint or fix)
- Speed instructions (change and maintain in knots or Mach number)
- Flight level change
- Flight transfer

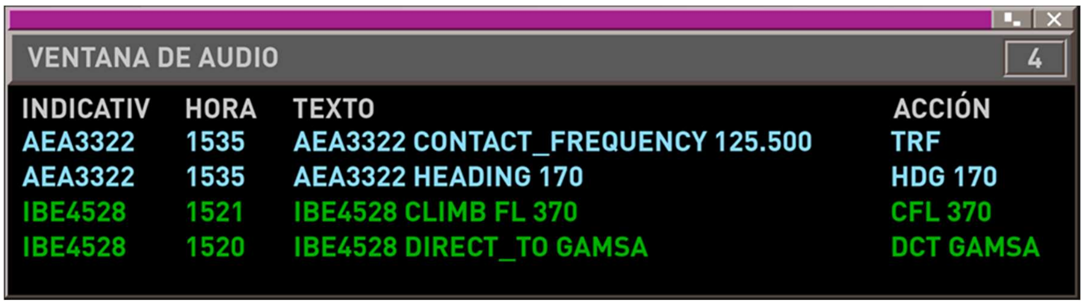

Figure 57: command annotation

The prototype was able to recognise multiple commands in one utterance and presented each command in one line. The prototype was also able to distinguish break break commands.

Conditional commands were not covered by the prototype (e.g. RYR78R climb flight level 370 to be levelled by GAMSA would be annotated: RYR78R CLIMB 370 FL)

C.4.4 Statistical approach Scenarios

The operational audios for the statistical approach belong to the sectors Lower Castejon (CJL), Upper Castejon (CJU) and Santiago (SAN), all of them from Madrid ACC (LECM). Figure 58 presents the location of the sectors within Spain airspace.

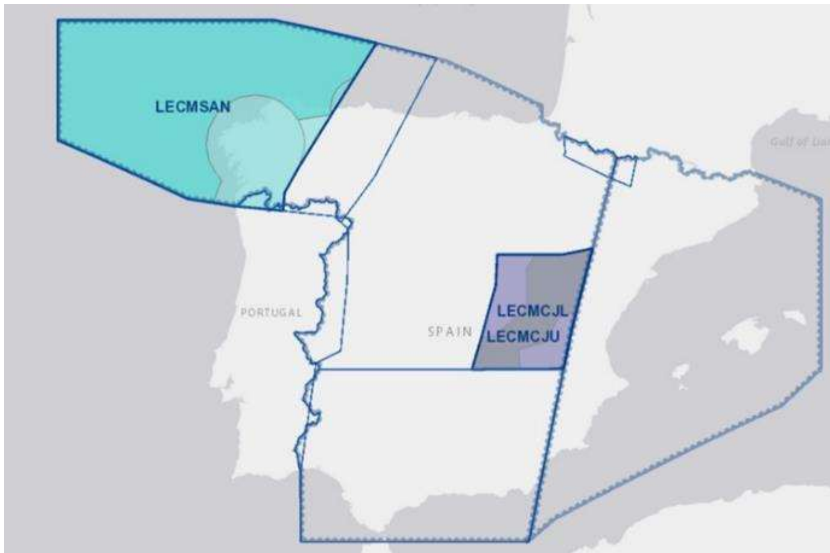

Figure 58: Statistical approach scenarios

These sectors were selected due to their complementary characteristics that provides a wide sample of technical (i.e. signal to noise ratio, native speakers origin) and operational (i.e. type of commands) characteristics:

- CJL is a sector with good radio coverage whose main traffic flows are to and from Madrid Barajas Airport, the major Spanish airport. It limits with Madrid TMA and the surface. Control service is provided to all aircraft from FL210 to FL325. Information service is provided from SFC to FL210 outside the TMA/airport areas and airways.
- CJU is a sector with good radio coverage and quality whose main traffic flows are to and from Madrid Barajas Airport, and over flights to the south of Spain. Control service is provided to all aircraft from FL325 to FL660.
- SAN sector includes a large proportion of oceanic airspace that has lower radio coverage compared to CJL and CJU. Its main flows are overflights to/from the America, and to/ from United Kingdom. The sector has Free Route airspace. Control service is provided to all aircraft from FL210 to FL660. Information service is provided from SFC to FL210 outside the TMA/airport and control areas.

The reference or gold standard was created through manually transcribing and performing the annotation following the SESAR standard.

C.5 Summary Technological Validation Exercise #03 Assumptions

The following assumptions apply to this exercise:

Identifier	Title	Description	Justification	Impact on Assessment
AS-EX03-01	Datalink	No data link will be used during the exercise	The exercise is centred on the radio communication between controllers and pilots	Low
AS-EX03-02	Language	ATCOs and pilots will communicate in English and/ or Spanish	Both languages are operational in Spanish airspace	Medium. Having to identify two languages duplicates the words that need to be identified

Table 42: Technological Validation Assumptions overview

C.6 Deviation from the planned activities

Deviations from the planned activities that do not impact objectives or success criteria:

- Deviation EXE03-1. The first day, the CWP 2 had a wrong connection with the simulator. The simulator sent the flight plans of CWP 1 to both positions. This resulted in a low callsign recognition rate. The feedback from the controllers was impacted by this low rate. A third run was performed on the second day with the same solution scenario as the first day to obtain real results for its posterior analysis. The questionnaires from CWP2, day 1 run 1 and run 2 have been removed from the analysis as they do not represent the real aptitude of this technology.

Deviations from the planned activities that impact objectives or success criteria:

- Deviation - EXE03-2. The TVALP [ref] indicates that EXE03 will not cover CRT-Sol.96ASR-TRL6-TVALP-0040.005 The level of ATCos' workload is maintained or decreased with the introduction of ASR system. of the objective OBJ-Sol.96ASR-TRL6-TVALP-0040, but workload was finally assessed in the RTS exercise.

C.7 Technological Validation Exercise #03 Validation Results

C.7.1 Summary of Technological Validation Exercise #03 Results

Technological Validation Exercise #03 Validation Objective ID	Technological Validation Exercise #03 Validation Objective Title	Technological Validation Exercise #03 Success Criterion ID	Technological Validation Exercise #03 Success Criterion	Technological Validation Exercise #03 Results	Technological Validation Exercise #03 Validation Objective Status
EX03-OBJ-Sol.96ASR-TRL6-TVALP-0010	To assess the technical feasibility of the integration of the ASR system and its sub-systems into CWP and interoperability between the ASR sub-systems and the existing CWP systems and tools.	EX03-CRT-Sol.96ASR-TRL6-TVALP-0010.001	The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the existing CWP systems and tools. Availability of systems and tools and their performance remain at 100%	The ASR was successfully integrated with operational CWP and communication system. The ASR did not impact the performance of previous systems	**OK**
EX03-OBJ-Sol.96ASR-TRL6-TVALP-0020	To assess the stability of the ASR system performance	EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001	The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)	The command type recognition rate for controllers is 89%-92%. The command text recognition rate is 72%	**POK**
		EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.002	The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.	The difference between different commands used in the exercise was within the 10%	**NOK**
EX03-OBJ-Sol.96ASR-TRL6-TVALP-0030	To assess the impact on the human performance of the integration of the ASR system and its sub-systems into operations in a realistic environment	EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.001	The introduction of the ASR system into the context of application is operationally viable, ATCOs workload with ASR is equal or better than in baseline (without ASR support).	ATCOs stated that the workload did not change or was decreased using the ASR technology.	**OK**

		EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.002	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate).	Controller considered that the recognition rate was not enough to support operations	**NOK**
		EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.003	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. Controllers' feedback with respect to Human Factors questionnaire is better than for baseline)	Controllers considered that the timeliness of the callsign recognition at the beginning of the phrase should be higher. The timeliness of the callsign at the end of the utterance and event recognition was enough although could be improved	**POK**
		EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.004	Changes in the design of the user interface (input devices, visual displays/output devices, alarm& alerts) support ATCOs in carrying out the tasks.	Majority of responses obtained through show that the Human-Machine Interface was adequate and appropriate to execute the simulation activity.	**OK**
		EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.005	The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified)	Answers indicate that although the system was useful and understandable it was not accurate or reliable enough to be confident with the system	**NOK**

EX03-OBJ-Sol.96ASR-TRL6-TVALP-0040	To assess the impact of the introduction of the ASR system on safety.	EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.001	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.	Controllers considered that the accuracy was not enough to support them	**NOK**
		EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.002	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.	Controllers considered that the timeliness of the callsign recognition at the beginning of the phrase should be higher. The timeliness of the callsign at the end of the utterance and event recognition was enough although could be improved.	**POK**
		EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.003	The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.	No error resulted from the introduction of the ASR	**OK**
		EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.004	The level of ATCO's situational awareness is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions).	All ATCOs saw their situational awareness as increased or unaffected with the introduction of the ASR system.	**OK**
		EXE03-CRT-Sol.96ASR-TRL6-TVALP-0040.005	The level of ATCos' workload is maintained or decreased with the introduction of ASR system.	ATCOs stated that the workload did not change or was decreased	**OK**

				using the ASR technology.	

Table 43: Technological Validation Results Exercise 1

1. Results on technological feasibility

The technical implementation performed by EXE-03 demonstrated that it is technically feasible to connect a preindustrial ASR prototype with operational systems, including an operational CWP that provides context information in real time to the ASR (flight plan list in this approach) and receives information from the ASR and presents it to the controller in a coherent approach with the rest of CWP information. The exercise also demonstrated the feasibility of integration with an operational voice communication system.

New requirements:

Controllers indicated the need to be able to distinguish between UC.1 and UC.2 if both approaches are implemented at the same time in the CWP. They indicated the need to be able to enable/disable each use case independently. They also suggested the use of different colours to discriminate between controller and pilot callsign highlight.

REQ-10.96-ASR-TS.003-01[8] If controller and pilot utterance highlight are implemented in the same CWP, controllers are able to distinguish between them.

REQ-10.96-ASR-TS.003-02 If controller and pilot utterance highlight are implemented in the same CWP, controllers should be able to switch them on/off independently.

Next recommendations are provided to the requirements:

Requirement REQ-10.96-ASR-TS.0080 mentions ED-137B [21] which was the one used by the exercise, nevertheless a new version of the standard, version C, has appeared. Recommendation is to update the requirement description to include possible updates.

REQ-10.96-ASR-TS-Perf.0030 Pilot Command Recognition Error Rate is out of scope of Solution 96. It is recommended to delete it or change it for Pilot *callsign* Recognition Error Rate

Next requirements where discussed:

Requirement REQ-10.96-ASR-TS.0010 (ASR is be able to process different traffic flows within the Area of Interest of the Control Unit) launched a large discussion between the technical and operational staff. While for Use cases 2 and 3 was clear the list of traffic and area of interest, controllers indicated that for Use case 1 it would be especially useful to have a broader definition of traffic of interest. They specifically referred to calls from traffic not expected in their sector that called them because they were lost or have introduced a wrong frequency. Technical staff indicated that if the flight plan of the aircraft was not assigned to the CWP, they could not highlight the flight in the console as the console did not know where that flight was. Further investigation may be needed regarding how to support controllers in these situations.

[8] The number used just refers to EXE-03, it is expected that the TS will use update them to the appropriate numbering.

EXE 03 covered requirements from Sol.96 ASR TS/IRS Table 44 Technical requirements validated y EXE-03 are listed in the Table 44. The table presents the requirements that have been validated, the UCs that have used them and any comments detected/received during the technical validation preparation or execution.

Requirement Id	Requirement	Comments
REQ-10.96-ASR-TS.0010	Cover all traffic flows with En-route and TMA	Validated UC.1 UC.2 UC.3 in En-route See comment above
REQ-10.96-ASR-TS.0020	Recognition of Commands	Validated UC.3
REQ-10.96-ASR-TS.0030	Multiple Commands	Validated UC.3
REQ-10.96-ASR-TS.0040	English and local Languages	Validated UC.1 UC.2 UC.3
REQ-10.96-ASR-TS.0050	Multiple Callsigns	Validated UC.2 UC.3
REQ-10.96-ASR-TS.0060	Different Speaker in one utterance	Validated UC.1 UC.3 In UC2 the requirement was not validated because the requirement was not applicable: The voice analysed was from the controller microphone thus there was only one speaker always.
REQ-10.96-ASR-TS.0080	ASR interoperability with VCS	Validated UC.1 UC.2 UC.3 See recommendation above
REQ-10.96-ASR-TS-HMI.0010	Callsign highlighting	Validated UC.1 UC.2
REQ-10.96-ASR-TS-HMI.0020	HMI for Command Values	Partially Validated in UC3. Instead of in the radar label they were presented in a separated window
REQ-10.96-ASR-TS-HMI.0050	ASR failure indication	Validated UC.1 UC.2 UC.3
REQ-10.96-ASR-TS-HMI.0060	ASR switch off/on	Validated UC.1 UC.2 For UC.3 was not necessary as the implementation was on a window that the controller could display or not.
REQ-10.96-01-TS-ReTi.0010	Provide callsign information immediately	Partially Validated UC.1 UC.2 The prototype sent the callsign 3 seconds after the initiation of the communication. Controlled indicated that the time should be much lower to really support them (1.5 seconds was mentioned).
REQ-10.96-ASR-TS-ReTi.0020	Reaction Time	Validated UC.3
REQ-10.96-ASR-TS-ReTi.0030	System reaction time	Validated UC.1 UC.2 UC.3

TSR-PJ10-W2-96 ASR-0010	ASR command recognition rate	Partially validated for UC3 Some recognition rates were not enough according to controllers
TSR-PJ10-W2-96 ASR-0020	Voice communications	Validated UC.1 UC.2 UC.3 UC1: Different recognition rates have been obtained in oceanic and en route sectors due to communication quality. **Pilot readback is out of scope of S96.**
TSR-PJ10-W2-96 ASR-0030	Disable ASR system from the HMI	Validated UC.1 UC.2 For UC.3 was not necessary as the implementation was on a window that the controller could display or not. (REQ-10.96-ASR-TS-HMI.0060)
TSR-PJ10-W2-96 ASR-0040	HMI display of complex clearances	Confirmed for UC3 The prototype was able to recognize several commands in one utterance but **not conditional clearances. Controller indicated the need to annotate the conditional clearances for UC.3**
TSR-PJ10-W2-96 ASR-0050	Recognition rates of different command	Not Validated. The recognition rates differed in more than 2.5% (usually up to 10%)
TSR-PJ10-W2-96 ASR-0070	ASR-system displayed on the CWP HMI	Validated UC.1 UC.2 UC.3
TSR-PJ10-W2-96 ASR-0080	The impact of ASR on other systems	Validated UC.1 UC.2 UC.3 (REQ-10.96-ASR-TS-ReTi.0030)

Table 44 Technical requirements validated y EXE-03

The callsign illumination was tested with existing functionalities of the CWP that can flash/highlight flight in the controller working position. **Controllers commented that the ASR implementation performed was clear and did not introduce confusion**. Recommendations regarding the need to take special care of this issue should be introduced in the Technical Specification.

2. Results per KPA

Objectives are stated per KPA. Please refer to the next section for the feedback regarding:

- Human performance (3 OBJ-Sol.96ASR-TRL6-TVALP-0030 Results)

- Safety (4 OBJ-Sol.96ASR-TRL6-TVALP-0040 Results)

Security related to the ASR functionality was discussed during the final debriefing. No security concern was identified in relation with the use cases implemented.

C.7.2 Analysis of Exercise 3 Results per Technological Validation objective

This section provides, per Technological Validation objective, a consolidated analysis of the Technological Validation exercise 3 results.

1. OBJ-Sol.96ASR-TRL6-TVALP-0010 Results

Objective description: *To assess the technical feasibility of the integration of the ASR system and its sub-systems into CWP and interoperability between the ASR sub-systems and the existing CWP systems and tools.*

Validation Objective ID	Success Criterion ID	Success Criterion Status	Validation Objective Status
EX03-OBJ-Sol.96ASR-TRL6-TVALP-0010	EX03-CRT-Sol.96ASR-TRL6-TVALP-0010.001	OK	OK

The ASR was successfully integrated with operational CWP and communication system. The ASR did not impact the performance of previous systems

a. Outcome Analysis

EXE-03 connected a preindustrial ASR prototype with operational systems, including an operational SACTA 4 CWP that provides context information in real time to the ASR (flight plan list in this approach), receives information from the ASR and presents it to the controller in a coherent approach with the rest of CWP information. The exercise also included connection with an operational voice communication system.

There was no impact on other systems or tools of the CWP.

All the controllers agreed that the ASR-system does not interfere with the availability and/or reliability (of other systems and components installed at the CWP).

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0010.001	The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the existing CWP systems and tools. Availability of systems and tools and their performance remain at 100%.	OK

2. OBJ-Sol.96ASR-TRL6-TVALP-0020 Results

Objective description: To assess the stability of the ASR system performance.

Success Criterion ID	Success Criterion	Success Criterion Status	Validation Objective Status
EX03-OBJ-Sol.96ASR-TRL6-TVALP-0020	EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001	**POK**	**NOK**
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.002	**NOK**	

Success Criteria:

CRT-Sol.96ASR-TRL6-TVALP-0020.001 The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.).

CRT-Sol.96ASR-TRL6-TVALP-0020.002 The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.

a. Outcome Analysis

Two statistical analyses have been performed:

The first one uses quantitative data collected from the RTS screen and audio recordings. The second one uses operational recordings from different Spanish sectors and a posterior statistical analysis of callsign recognition and event annotation was performed afterwards.

The statistical analysis was obtained by transcribing manually the recordings, crating the callsign and event annotation standard, and then comparing it against ASR outcome.

Table 45 presents the total number of callsigns present in the audios and the number of callsigns that were correctly detected by the ASR. The percentage of correctly detected callsigns is higher for ATCOs than for flight crew in both cases as the algorithm is optimized for the ATCo locutions.

Regarding the comparison between simulation and operational recordings, the percentage for ATCOs are similar but the percentage for flight crew is better in the simulation. This was already expected as the quality of the recording (ratio signal to noise) is better in the simulation and the accent (mother tongue) of the pseudopilots is unique (Spanish) while the one from the operational recordings is very diverse with 29 companies from 18 different countries.

	ATCO			Flight crew		
Analysis type	Nº of callsigns	Nº callsigns detected	Percentage	Nº of callsigns	Nº callsigns detected	Percentage

RTS recordings	859	721	**84%**	457	687	**67%**
Operational recordings	143	127	**87%**	158	77	**49%**

Table 45 EXE-003 detected callsigns

No callsign was wrongly recognized as only complete callsigns were detected. Feedback from controller indicated that they would like to have higher recognition rates even if some callsigns were incorrectly detected and highlighted. The allowed error is something to be investigated.

Table 46 presents the number of events present / detected and the callsign + event correctly detected for each analysis. We are only presenting the events that falls within the five categories for which the prototype was optimised. There were several other events that controllers used during the simulation such as squawk change, STAR assignation, information (traffic, QNH,...).

Analysis type	Events (type)	Events detected (type)	Percentage	Callsign + Event (type)	Callsign + Event detected (type)	Percentage
RTS recordings	695	619	**89%**	695	523	**75%**
Operational recordings	182	167	**92%**	182	146	**80%**

Table 46 EXE-003 detected event type

There is a 5-3% difference between the RTS and the operational recognition percentages. During the presentation of the exercise, controllers were encouraged to test the recognition system. They thus issued longer and more complex authorisations that what are usually given in operational environments. This could explain the difference between both percentages.

If not only the event type but also the information is taken into account, the numbers are lower:

Analysis type	Total Events	Event_Text detected	Percentage	Callsign + Event_Text detected	Percentag e
RTS recordings	695	498	**72%**	416	**60%**

Table 47 EXE-003 detected events

According to the TS/ASR the recognition rates should be:

- The Command Recognition Error Rate of ASR should be less than 2.5% for ATCos.
- The Command Recognition Error Rate of ASR SHOULD be less than 5% for pilots
- The Command Recognition Rate of ASR of ATCos should be higher than 85%.
- The Command Recognition Rate of ASR for pilots SHOULD be higher than 75%.

The command error rate has not been measured. **The Word Error Recognition rate which is commonly used in ASR world is 10%, which could be considered high, but it should be taken into account that the ASR used recognises words in two different languages.**

The command type recognition rate for controllers is 89%-92% thus the TS/URS criterion is meet for Event Type. If the Event text is also considered, the percentage is 72% which does not meet the criterion.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001	The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)	**POK**

Table 48 presents the number and percentage of commands per type in the RTS. As previously indicated, these were not the only commands that were issued during the simulation, but are the ones optimised in the prototype. All the event type recognition rates (except speed) are above 84% but when taking into account the information numbers and the callsigns the recognition descend greatly. When analysed, one of the main problems were complex authorisations that the prototype did not recognise: i.e. Descend fly level 220 to be levelled by TOPTU was noted as Descend TOPTU

	Total	Event_Type		Event_Text		Callsign +Event_Type		Callsign +Event_Text	
Headings	**19**	**16**	**84%**	**15**	**79%**	**14**	**74%**	**13**	**68%**
Direct	**125**	**113**	**90%**	**66**	**53%**	**98**	**78%**	**56**	**45%**
Speed	**30**	**19**	**63%**	**9**	**30%**	**15**	**50%**	**6**	**20%**
level change	**192**	**164**	**85%**	**115**	**60%**	**140**	**73%**	**98**	**51%**
Transfer	**329**	**307**	**93%**	**293**	**89%**	**256**	**78%**	**243**	**74%**

Table 48 EXE-003 RTS detected events per type

The requirement regarding difference between commands is not meet as differences are greater than 2.5%.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.002	The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises	**NOK**

3. OBJ-Sol.96ASR-TRL6-TVALP-0030 Results

Objective description: To assess the impact on the human performance of the integration of the ASR system and its sub-systems into operations in a realistic environment.

Success Criterion ID	Success Criterion	Success Criterion Status	Validation Objective Status
EX03-OBJ-Sol.96ASR-TRL6-TVALP-0030	EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.001	OK	POK
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.002	NOK	
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.003	POK	
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.006	OK	
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.007	NOK	

g

a. Outcome Analysis

Impact on the human performance was measured using post-run, post simulation questionnaires and debriefings. Next section presents the results.

1. Workload

The workload was measured using different approaches in the questionnaires. Figure 59 presents the answers to the Nasa TLX.

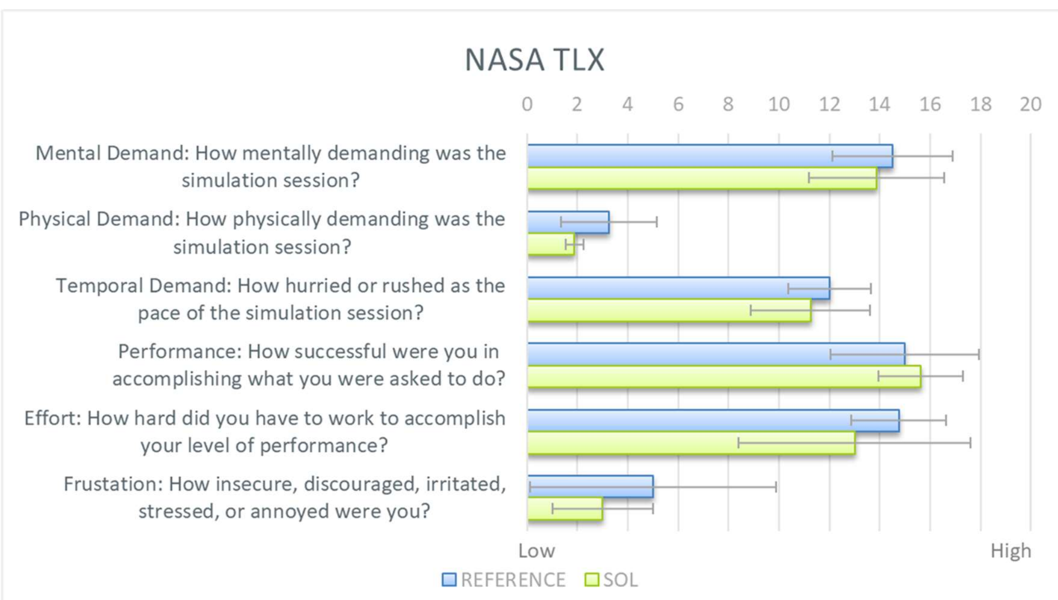

Figure 59 Questionnaire – Workload – NASA TLX

The scoring of the Nasa TLX is 9.1 for the reference questionnaire and 7.9 for the solution questionnaire. **The workload is reduced in the solution scenario**, but the outcomes need further confirmation from other exercises as the sample and difference between both scenarios are not high

Figure 60 presents other questions related to workload. The scale ranges from 0/disagree/ extremely poor to 10/agree /extremely high

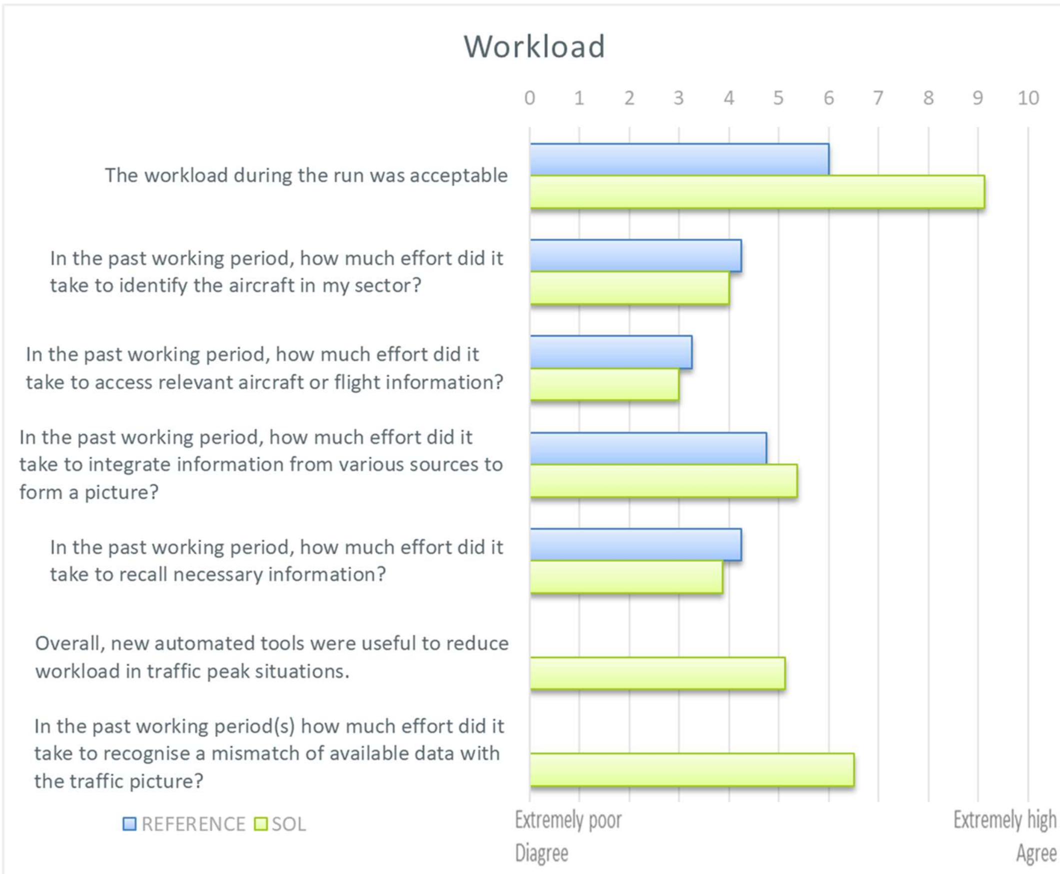

Figure 60 Questionnaire – Workload

Workload slightly improves when compared to the reference scenario. When questioned regarding the support of the new tools the answers indicates that ASR slightly supported them although mismatch of information (no recognition of the ASR) introduced some workload.

Figure 61 presents the general feeling regarding workload. The answer is ranked from 1- significantly decreased to 5- significantly increased.

Figure 61 Questionnaire – Workload II

The feedback indicated that controllers consider that ASR slightly decreased the workload which is in line with the previous answers.

During the debriefings controllers indicated that the ASR reliability, recognition rates should be improved to effectively support a workload reduction. The ASR did not increase the workload.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.001	The introduction of the ASR system into the context of application is operationally viable, ATCOs workload with ASR is equal or better than in baseline (without ASR support).	OK

2. Information completeness and timelines

As presented in section 2 OBJ-Sol.96ASR-TRL6-TVALP-0020 Results the recognition rates were of 84% and 67% from callsigns recognition from controllers and flight crew respectively. For event type detection the percentage is of 89% and 75% for the detection of both callsign and event type. The percentage lowers if not only the event type is detected but also the exact text of the event. In these cases, the outcome is 72% for the Event Text and 60% for callsign and Event Text

Subjective feedback was collected from the questionaries as presented in Figure 62. The scale ranges from 0/never to 6/always. The questions refer to the UC 1 and 2, callsign highlight, (identified as RDF mark in the questionnaire), and to UC 3, event annotation, as CAT/L window. See C.4.3 RTS Solution Scenario(s)

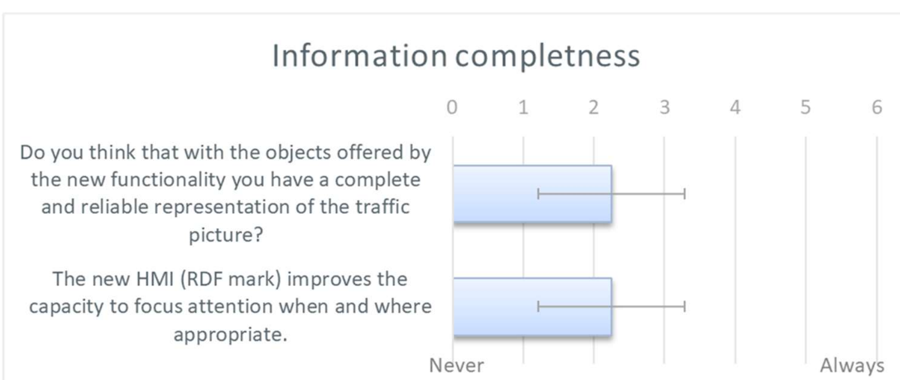

Figure 62 Questionnaire – Information completeness

The feedback indicates that **the controllers perceived the information as complete only sometimes**. In the debriefings they related the answer to the recognition rates obtained during the simulation.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.002	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate).	NOK

The information was provided with the timeliness indicated in Table 49. There is a significant processing time difference from the beginning of the phrase or the end of the phrase. This difference is related to the fact that at the end of the phrase the communication is closed, and there is a signal indicating the end of the phrase. The end signal triggers the recognition and analysis algorithms whereas at the beginning of the phrase the prototype expects further information, and the recognition and analysis algorithms keep processing.

	Mean Time (s)
Initiation of the phrase	3.02
End of the phrase	0.93

Table 49 Timeliness

This processing difference makes that callsigns that are stated at the beginning of an utterance take 3 seconds to be highlighted on the console. Callsigns at the end of the utterance and events appear on the console one second after the end of the phrase. The control event time is usually associated to the end of the phrase as controllers usually release the push-to-talk button at the end of the command.

The prototype meets the requirement REQ-10.96-01-TS-ReTi.0020 *For 99.9% of the ATCo utterances except callsign itself, the system shall be able to give the output in less than two seconds after the ATCo has released the push-to-talk button* in relation with control events and callsigns at the end of the phrase.

The prototype does not meet REQ-10.96-01-TS-ReTi.0010 *ASR may send not later than 1.0 second recognized callsign to the cooperating ATC system when the controller has pressed the push-to-talk-button and said the callsign (aircraft identifier)* for the callsigns at the beginning of the phrase.

Subjective feedback on the time was provided in the questionnaire through the following questions. The scale ranges from I totally agree/1 to I totally disagree/5.

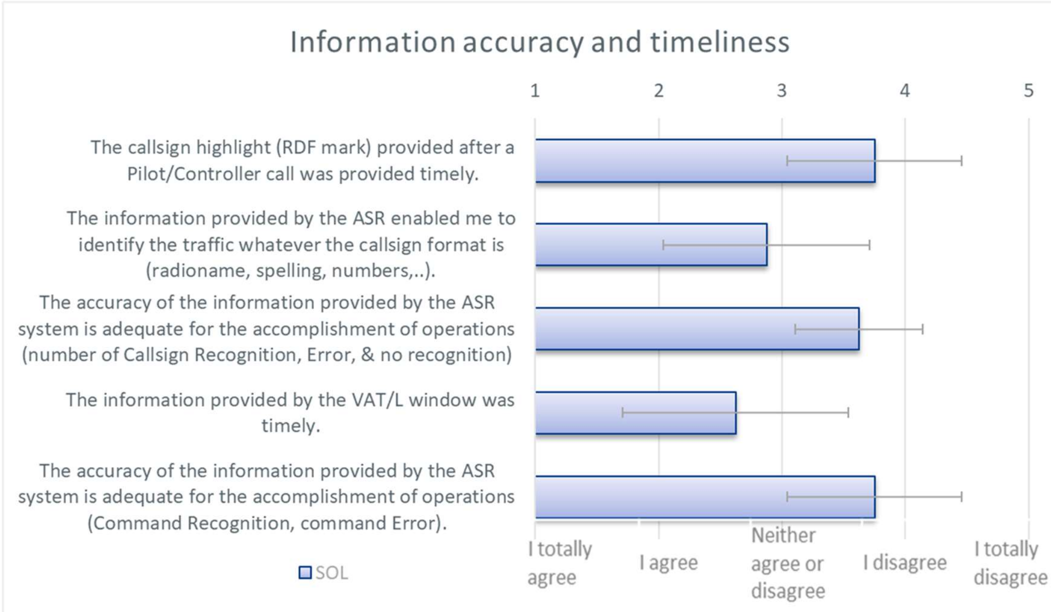

Figure 63 Questionnaire – Information accuracy and timeliness

Controllers indicated that **the callsign highlight from flight crew utterances was not provided fast enough** to support them whereas **the event information (VAT/L window), was provided timely although it could be improved**. The necessity to improve the callsign and event recognition rates was confirmed.

These questions were not answered on the reference runs as they indicated that they would be measuring their current system in the reference and only the ASR in the solution questionnaire thus their feedback should be taken into account as **absolute figures**.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.003	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. ~~Controllers' feedback with respect to Human Factors questionnaire is better than for baseline.~~	POK

3. HMI support

Controllers were questioned regarding the adaptation of the ASR to the console from a user interface design point of view. The questions were divided per use case, callsign highlight and event annotation. The question and the answers gathered are presented in Figure 64 and Figure 65. The scales range from never/0 to always/6.

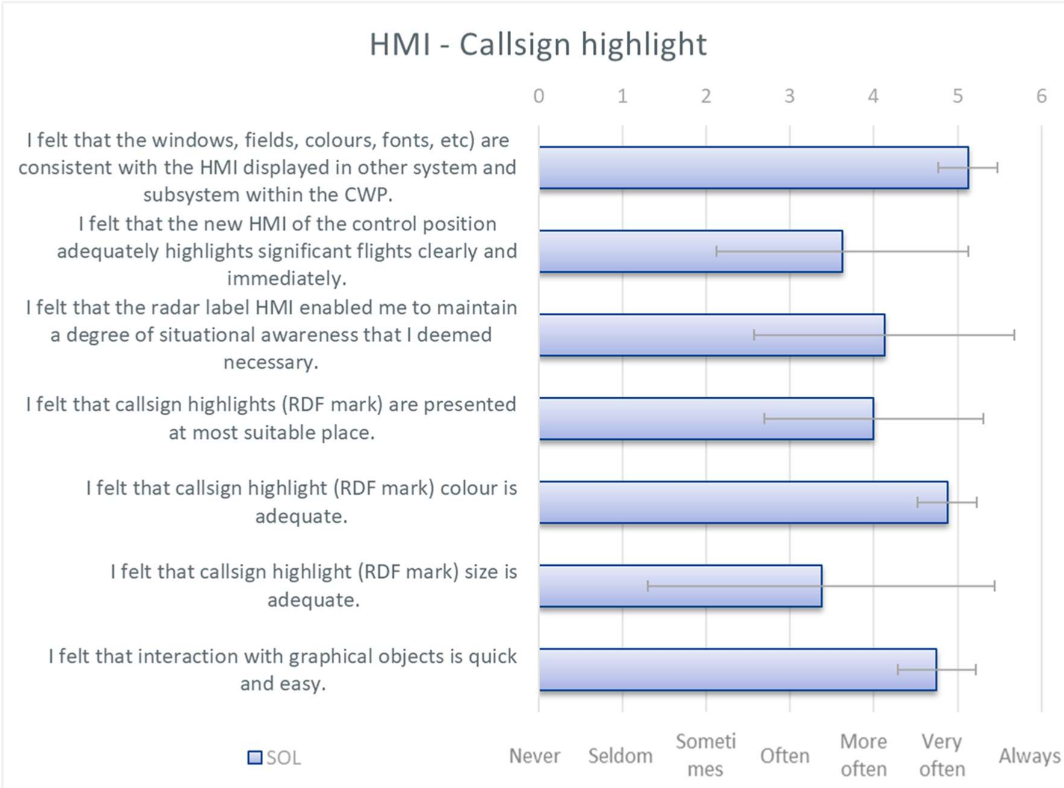

Figure 64 Questionnaire – HMI callsign highlight

Regarding the callsign highlight, most of the answers indicate that graphical object, RDF mark, was clear and consistent with the HMI of the CWP, the interaction with graphical objects was easy, and the colour was adequate**; but there is no consensus about the size, and the timeliness of the highlight. Half of the answers considered the size small, but the other half considered it correct.** Regarding **the timeliness it depended if the callsign was pronounced at the beginning (not good) or the end of the utterance (adequate).**

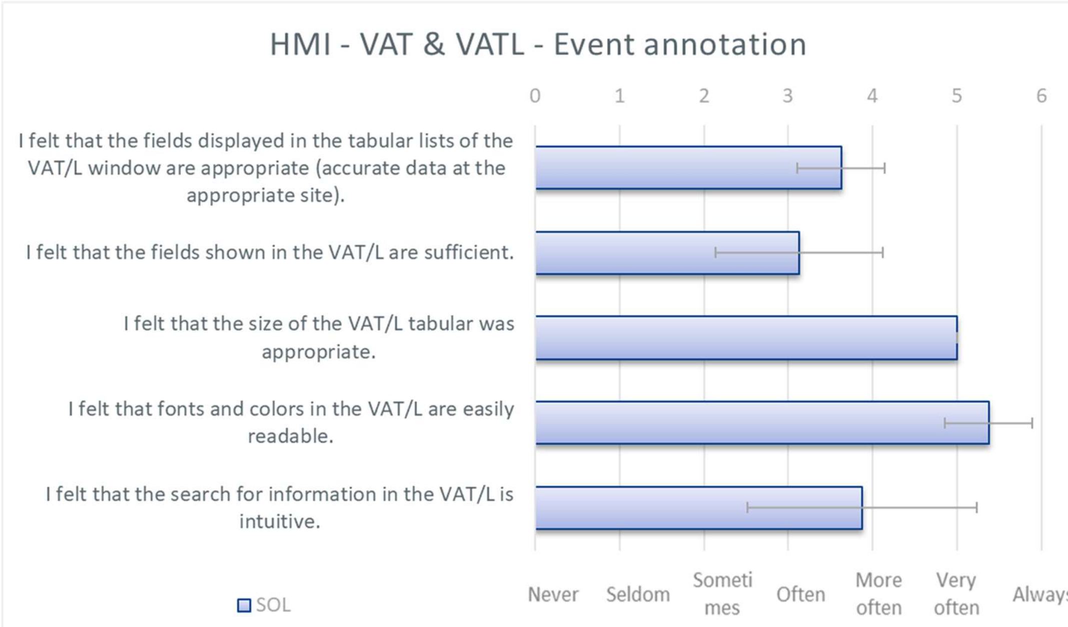

Figure 65 Questionnaire – HMI VAT/L – Event annotation

Answers regarding the event annotation windows indicate that size, colours and fonts are appropriate and well integrated with the console, although **they missed some information** (this comment is not related to the HMI but rather to the event ontology, see next section).

During the debriefings controllers indicated that the integration with SACTA HMI (existing HMI in the CWP) regarding colour consistency, tabular/window structure was very adequate and supported them to integrate the ASR in their tasks. They also expressed **the need to be able to distinguish between UC1 (callsign highlight from pilot utterances) and UC2 (callsign highlight from controller utterances**). They proposed to have different colours. Another request was the possibility to be able to switch on/off each use case independently. This request was fulfilled by the System Engineers disabling UC2, but ATCOs were not able to do that themselves. This feedback has been incorporated in section C.7.11 Results on technological feasibility as new requirements.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.006	Changes in the design of the user interface (input devices, visual displays/output devices, alarm & alerts) support ATCOs in carrying out the tasks	OK

4. Trust

Trust in the system was measured using the SATI questionnaire. The scale ranges from I totally never/0 to always/6.

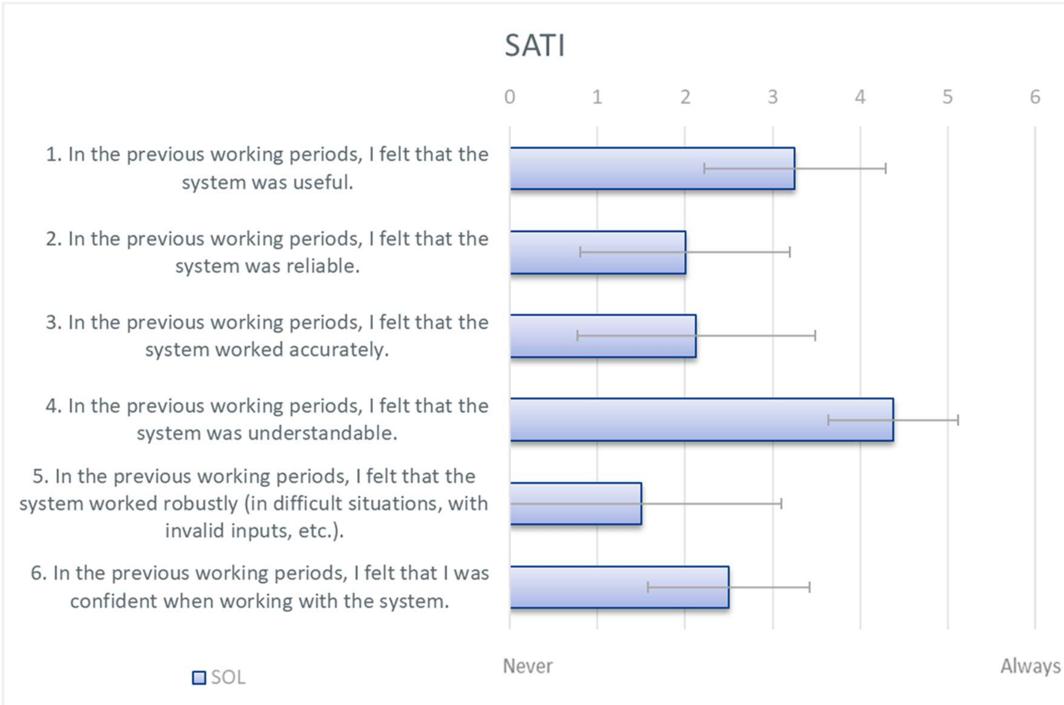

Figure 66 Questionnaire – SATI

The overall scoring of the SATI questionnaire is 2.6. Answers indicate that although **the system was useful and understandable it was not accurate or reliable enough to be confident with the system**.

During the debriefings controller indicated **the callsign recognition rates as the main factor for not trusting the system.**

They also commented on the event annotation ontology. The ontology used was the one agreed among SESAR [ref] and has been developed with objective of entering information in the CWP, thus its main focus is to be concise and include just the elements necessary to be inserted in the CWP. UC3 provides a historical annotation of commands issued to the aircraft for consultation. Controllers indicated that **the ontology was not completely adapted to their necessity and needed further information than what was presented**. It should be noted that the industrial prototype used did not cover all the events in the ontology. As example:

Air Transat four five two start descend now flight level two one zero to be levelled by AVILA

Was transcribed as:

TSC452 DESCEND 210 FL

TSC452 DIRECT TO AVILA

Which is not correct. They preferred:

TSC452 DESCEND 210 FL NOW LEVELLED_BY AVILA

In the example, the first command is correct if the information is used to update the console information, but the second example is the correct one if used to identify expected behaviour of the aircraft.

The user acceptance was measured through the CARS (Controller Acceptance Rating Scale) questionnaire. All the controllers scored that system needed improvement. The overall scoring indicates that *the system has moderately objectionable with deficiencies that require considerable controller compensation to achieve adequate performance*. The scoring improved the second simulation day.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0030.007	The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified).	NOK

4. OBJ-Sol.96ASR-TRL6-TVALP-0040 Results

Objective description: To assess the impact of the introduction of the ASR system on **safety.**

Success Criterion ID	Success Criterion	Success Criterion Status	Validation Objective Status
EX03-OBJ-Sol.96ASR-TRL6-TVALP-0040	EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.001	NOK	POK
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.002	POK	
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.003	OK	
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.004	OK	
	EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.005	OK	

Success Criteria:

CRT-Sol.96ASR-TRL6-TVALP-0040.001 The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.

CRT-Sol.96ASR-TRL6-TVALP-0040.002 The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.

CRT-Sol.96ASR-TRL6-TVALP-0040.003 The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.

CRT-Sol.96ASR-TRL6-TVALP-0040.004 The level of ATCO's situational awareness is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions).

CRT-Sol.96ASR-TRL6-TVALP-0040.005 The level of ATCos' workload is maintained or decreased with the introduction of ASR system.

a. Outcome Analysis

1. Accuracy and timeliness

The accuracy and timeliness of the information provided has already been analysed in previous objectives. This section provides a summary and incorporates some measurements.

Regarding the UC1 and UC2, callsign recognition:

- Callsign recognition rate was of 84% for callsigns from controller utterances and 67% for callsigns from flight crew utterances, which was considered as not enough by the controllers.
- The callsign error recognition rate was zero percent. **The prototype had an algorithm that unless the callsign was completely recognized, no callsign was highlighted**.
- During operational recordings the percentage are similar for callsigns from controller utterances, 87%, but are worse for callsigns from flight crew utterances, 49%.

During the debriefings controller indicated that they preferred the possibility of having a possible false positive if the recognition rates improved significantly. **There was no agreement on the acceptable percentage of false positive.** They suggested **the possibility of having a window with the transcription of the identified callsign coming from pilot utterances**. **This characteristic was implemented for controller utterances as part of the annotation window.**

Regarding the UC3, event annotation:

- Event recognition rate per Event Type was of 89% and 72% per Event Text which was considered as enough but improvable.
- Event detection rate was not measured. Word error rate, WER, was of 10%

In the debriefings controllers indicated that the recognition rates of the events were adequate although they pointed out some cases where the prototype did not react as expected. This happened in utterances with several commands. They indicated that **the main problem was that when a callsign was not recognized, the information was assigned to a "flight outside the plan" and therefore they were not able to follow the complete history of commands provided to a flight which would be the most useful feature of the use case.**

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.001	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.	NOK

The timeliness of information was analysed in a previous objective, see Table 49. As summary, controllers indicated that the callsign highlight when it was at the beginning of the phrase was not provided fast enough to support them whereas the event information (VAT/L window), was provided timely although it could be improved.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.002	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations	POK

No error resulted from the introduction of the ASR.

Most of the ATCOs did not have any safety concerns regarding the simulation and believe the system did not interfere nor compromise the task execution and availability of other systems and components installed at the CWP.

However, a remarkable percentage did not know if the ASR tool could be perceived as a safety tool or if the system was not conflicting with the existing visual indications. Only half of the answers show that the tool helped to develop the work more safely and efficiently than it is today.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.003	The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.	OK

2. Situational awareness

Figure 67 presents the outcome of the SASHA questionnaire. The answers range from Never (0) to Always (6).

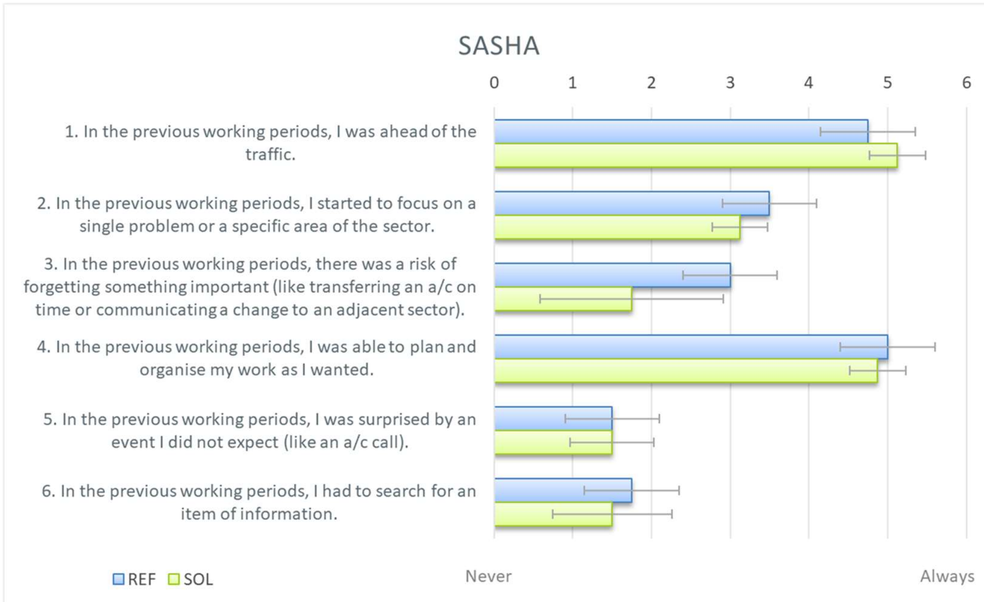

Figure 67 Questionnaire – Situational awareness -SASHA

The overall scoring of the SASHA questionnaire is 4.0 in the reference questionnaire and 4.4 in the solution questionnaire. The situational awareness slightly improved with the use of ASR.

Figure 68 presents the general feeling regarding situational awareness. The answer is ranked from 1- significantly decreased to 5- significantly increased

Figure 68 Questionnaire – Situational awareness

The feedback indicated that controllers consider that ASR slightly increased situational awareness.

During the debriefings controllers stated that situational awareness was improved but they considered that the ASR recognition rates was not high enough to allow them to completely confide and exploit the tool. They consider that higher callsign recognition rates would further improve their situational awareness.

Controllers appreciated specially UC1. They considered that with higher **robustness (meaning accuracy and timeliness**) it would support to develop their tasks more efficiently and increase situational awareness. They considered UC2 could be especially helpful when a controller needs to understand

EUROPEAN PARTNERSHIP
Co-funded by the European Union

the sector situation but is not located directly in front of the screen. In these occasions, following the performance of the controller on the radio can be difficult and having a callsign highlighting the flights from controllers' utterance will support them. At Enaire this situation happens:

- during a shift change. The entering controller may sit near the departing controller during a period of time to be able to grasp the situation before actually controlling the flights.
- When new controllers have onsite training. **The new controller may be near the experienced controller following the issued commands, or a supervisor may be near the new controller.**

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.004	The level of ATCO's situational awareness is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions)	OK

The feedback on **workload** taking into account Nasa TLX, and other related questions indicates that workload slightly improved due to the introduction of ASR. For a complete analysis please refer to CRT-Sol.96ASR-TRL6-TVALP-0030.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0040.005	The level of ATCos' workload is maintained or decreased with the introduction of ASR system.	OK

3. Other Safety considerations

Controllers indicated that the historical annotation window (UC.3) enabled the possibility to perform some checks a posteriori **but is time consuming and sometimes the information seek is not available due to the recognition rates.**

C.7.3 Unexpected Behaviours/Results

Controllers were encouraged to thoroughly test the ASR. Being a prototype with limited commands implemented, the controller feedback was impacted by the limitations, although previously informed.

C.7.4 Confidence in Results of Validation Exercise 1

1. Level of significance/limitations of Technological Validation Exercise Results

Air Traffic Control Operators were asked during the post-run questionnaire the following questions regarding the simulation environment:

The answer range to these questions went from 1 ("Very far from reality") to 4 ("Very realistic")

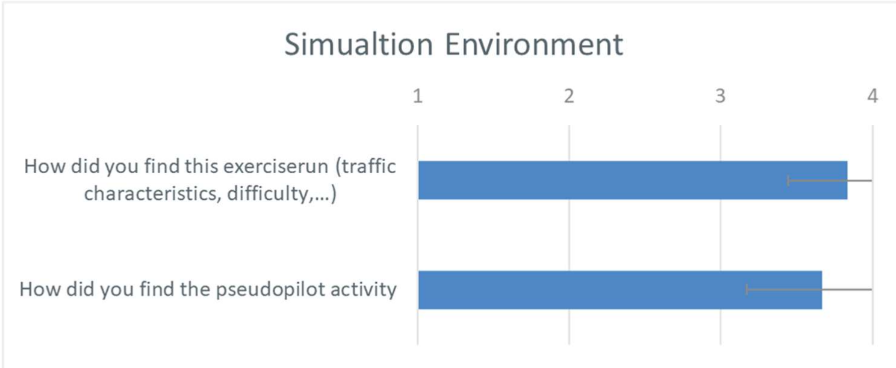

Figure 69 Questionnaire – Simulation environment

Controllers considered **the run as adapted to reality or very realistic**. There were some comments regarding the need to perform more actions than in real live in some cases (e.g. squawk change) but considered them as useful to thoroughly test the ASR tool.

Traffic load was medium/ high and the communication load was adapted to the simulated use cases to meet the validation objectives.

The exercise simulated an **opening and closure of sectors** to gather possible requirements associated to it.

2. Quality of Technological Validation Exercises Results

The quality of the exercises result is considered as high.

Regarding operational significance all runs were performed on traffic scenarios and sector configurations based on the two Madrid ACC En-route sectors with adapted traffic from 2019. All controllers involved in the exercise were operational experienced professionals of Madrid ACC that volunteered for non-profit purposes.

Pseudopilots participating in the simulation have a pilot license and have previously participated in other real time simulations performed at Crida and Eanire.

The validation platform implemented is an operational SACTA.4 CWP with an operational communication system connected with a simulation engine. The simulation engines and part of the platform was used in previous simulations in SESAR wave 1 and has been updated to execute the exercise.

3. Significance of Technological Validation Exercises Results

Confidence on qualitative results is **medium as the number runs performed** by each ATCO and incidents simulated during the exercise are in line of a real time simulation.

Nevertheless, **confidence on quantitative results is low due to the number of runs being low for statistical significance.** They should be taken into account together with the comments from the debriefings and the standard deviation provided.

The statistical significance of the recognition and error rates is considered as high due to the number of utterances analysed from the real time simulation and the statistical analysis.

C.8 Conclusions

Controllers believed that the system could be useful in the near future when the technology is more developed, **faster, and accurate than it is today**. However, they considered that the system as it is helped them to some extent with their tasks. Regarding the usefulness of the use cases tested they feedback was:

- UC.1 **Highlight of callsigns on the CWP from pilot utterances.** This use case **w**as considered the most useful use cases and they looked forward having it implemented on the CWP **when the accuracy and timeliness is improved.**
- UC.2 **Highlight of callsigns on the CWP from controller's utterances**. This use case **w**as considered as useful **in concrete situations such as shift change or controllers in training**. They considered it as a nice to have if the previous use case was implemented in the console.
- UC.3 **Annotation of controller's commands**. Controllers considered this use case as **the least useful and would rather prefer a system that automatically inputs the information in the console than a historical window.** The usefulness of this historical window was burdened by the accuracy of the callsign identification and ontology used.

1. Conclusions on Technological feasibility

The exercise confirmed the feasibility of integrating an Automatic Speech Recognition tool in an operational Controller working position with an operational communication system to support the use cases:

- UC.1 **Highlight of callsigns on the CWP from pilot utterances.**
- UC.2 **Highlight of callsigns on the CWP from controller's utterances**.
- UC.3 **Annotation of controller's commands**.

The integration was performed without impacting the performance of other console systems coherently with the rest of the HMI. The ASR preindustrial prototype was able to receive information online of the flights entering and leaving the sectors and was able to provide to the HMI the necessary input to be displayed to the controller.

A list of validated technical requirements has been provided together with comments to some of them. **New requirements** related to the integration of UC.1 and UC.2 have been identified:

- REQ-10.96-ASR-TS.03-01 If controller and pilot utterance highlight are implemented in the same CWP, **controllers shall be able to distinguish between them**.
- REQ-10.96-ASR-TS.03-02 If controller and pilot utterance highlight are implemented in the same CWP, **controllers should be able to switch them on/off independently**

2. Conclusions on performance assessments

Next conclusion can be extracted from the exercise:

In human performance, **the workload was reduced** compared to the reference scenario. Results show that the user interface of the new tools was adequate in size, colours and display, **but the technology itself was not trustable** enough at the moment for its insufficient accuracy and a delay longer than desired one for callsign highlighting.

Safety was maintained within acceptable limits and the system did not interfere with the reliability or availability of other systems. **Situational awareness was slightly improved** compared to the reference solution. However, **the prototype was not considered mature enough to be a trusted and accepted safety tool, due to the recognition rates.**

Feedback from controllers suggest that improvements in accuracy and timeliness will also provide higher reduction in workload and higher situational awareness.

C.8.2 Recommendations

Recommendations are to continue the improvement of recognition rates and timeliness to be able to implement UC1 and UC2 in the CWP. This improvement should include:

- Investigation based on machine learning on the pronunciation of company radio names by the company pilots and the local controllers.
- Investigation on the trade-off between the possibility to have false positive callsign identification and improved recognition rate.
- Improvement on the timeliness. A maximum of 1.5 second since utterance of the callsign was mentioned.

Stop the implementation of UC.3 until the accuracy of the transcription and detection is higher and concentrate in the following use cases:

- **UC.4 Pre-filling of commands in the CWP**
- **UC.5 Prefilling of Datalink commands**

The Technical Specification should be updated:

- To include the identified requirements.
- To include a recommendation to test UC1 and UC2 with existing functionalities of the CWP that can flash/highlight flight in the controller working position.

Appendix D Technological Validation Exercise #05 Report

D.1 Summary of the Technological Validation Exercise #05 Plan

This appendix contains the **Technological Validation Report for exercise** EXE-PJ.10-96-ASR-TRL6-05 performed by SINTEF as part of PJ.10-W2-Sol.96 ASR Automatic Speech Recognition in En-route environment. The exercise was conducted in cooperation with EXE-PJ.09-W2-44-V3-VALP-005 performed as part of PJ09 solution 44, involving ENAV, IDS, EUROCONTROL and SINTEF.

D.2 Technological Validation Exercise #05 description and scope

In Exe-005, SINTEF supported by ATCOs from ENAV and AVINOR used ASR technology to support En-route controllers in their tasks. The ASR technology is integrated in the CWP part of the SIMADES simulator platform.

The ASR part provides functionality supporting enhanced understanding of the changes to the airspace the sectorization changes, as well as how these changes affect the traffic being controlled. The ATCOs are offered a set of voice commands supporting DAC related navigation and information presentation in the CWP.

Three use cases were addressed in the exercise. The description of the UCs is extracted from PJ.10-W2-Sol.96 ASR Technical Specification:

UC.5 **Voice commands for highlighting an upcoming sectorization change in the CWP.** These commands will let the ATCOs quickly view how the sectorization in the upcoming airspace configuration will be. This includes both the sector being controlled by a given ATCO, and the neighboring sectors.

UC6 **Voice commands for highlighting the flights that will be affected by an upcoming sectorization change in the CWP**. These commands will let the ATCOs quickly view which flights that need special attention when the sectorization changes, particularly the flights that will change sector as a consequence of the sectorization changes.

UC7 **Voice commands for navigating the 3D visualization of the airspace in the CWP.** The SIMADES CWP provides a 3D view of the airspace. These commands will ease the visual navigation of the 3D view, among other by providing quick navigation to predefined and user defined viewpoints.

The exercise follows two complementary approaches:

The first approach collected the operational feedback from controllers by means of a Real Time Simulation. ATCOs controlled En-Route sectors performing their task as usual with (solution scenario) and without (reference scenario) the automatic speech recognition system enabled. Pseudo pilots managed flights and interacted via voice with the controllers.

Subjective feedback was gathered by means of questionnaires, debriefings and observations. A set of specific system data log were also recorded to corroborate the qualitative data. The real time exercise was conducted at SINTEF's premises in Oslo.

The second approach analysed the ASR performance by means of post processing of the data recorded during the Real Time Simulation. Data logs were collected and analysed in order to evaluate the performance of the ASR module (e.g. command recognition rate).

This exercise addressed several **Validation Targets**:

- **Human Performance** (increase of situational awareness)
- **Safety** (positive impact on safety is expected, enhancing ATCOs' situational awareness)
- **Controllers productivity** (related to the workload reduction associated to faster navigation in the CWP as well as early identification of aircraft)

Prior to the validation exercise, controllers and pseudo pilots were provided with the following material:

- Training material for the CWP
- List of voice commands with explanations
- Participant information and agreement form
- Questionnaires to be filled in (only to controllers)

Table 39 presents the scheduled activities during the simulation execution.

Date	Activity
26/9-22	Training
27/9-22	PJ09 sessions PJ10 training session with frozen traffic
28/9-22	PJ09 sessions
29/9-22	PJ10 training session with live traffic Open day Combined PJ09/PJ10 session (Ordinary session 1)
30/9-22	Combined PJ09/PJ10 session (Ordinary session 2) PJ09 session Final debriefing

Table 50 EXE-005 schedule

The speech control prototype is a front-end to the SINTEF's SIMADES CWP which is part of the SIMADES ATC platform also including the SIMADES air traffic simulator. In the context of the combined exercise EXE-PJ.10-96-ASR-TRL6-05 / EXE-PJ.09-W2-44-V3-VALP-005, these two parts of the SIMADES ATC platform works in combination with a flow manger tool (ATFM) provided by IDS, Eurocontrol's INNOVE simulator, a real-time DAC algorithm and a workload calculator, the latter two provided by SINTEF. A simplified illustration of the architecture is presented in Figure 70.

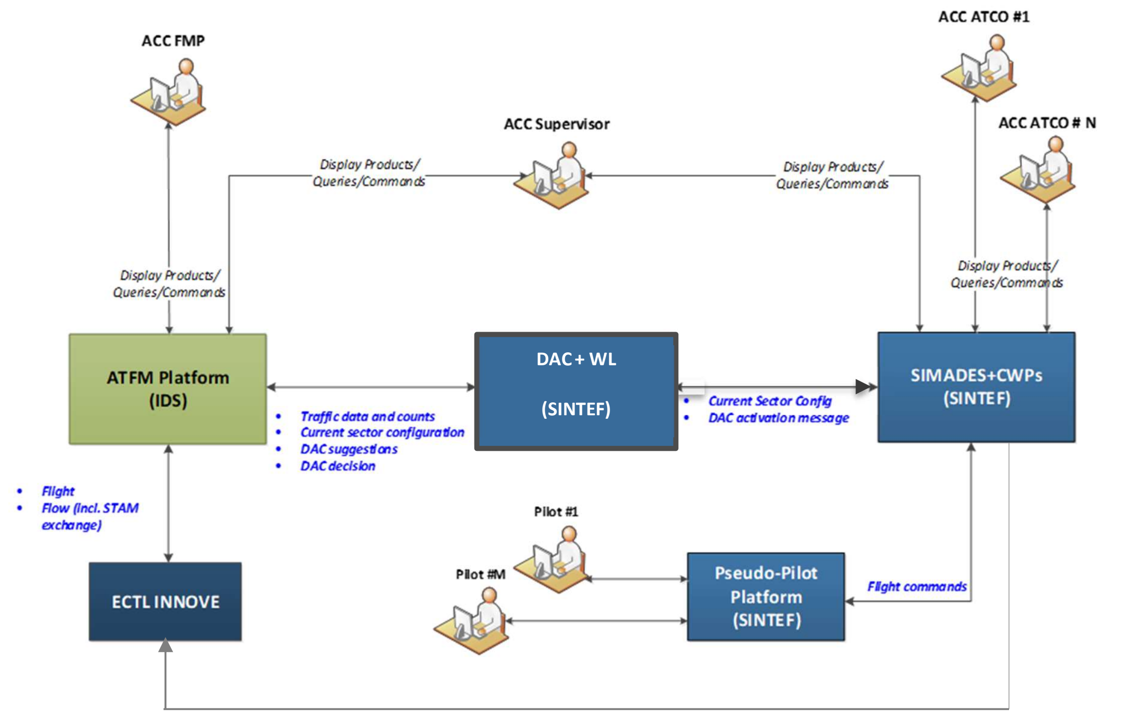

Figure 70– System architecture of the combined exercise EXE-PJ.10-96-ASR-TRL6-05 / EXE-PJ.09-W2-44-V3-VALP-005

The SIMADES CWP (see Figure 712) is web-based and provides a traditional, minimal functionality radar screen for viewing the traffic, conduct clearance, handover, etc. In addition, there are specific tools supporting the ATCO's situational awareness when DAC is being used, both for understanding the changes in the airspace and how these changes affect the air traffic.

Figure 71 – The SIMADES CWP

These tools include a sector selector, to view the current and next configuration of any sector in the airspace, a side view of the airspace and a 3D view of the airspace. The latter may be viewed from different angles and filtered to view any combination of the sectors in the current and next airspace configuration.

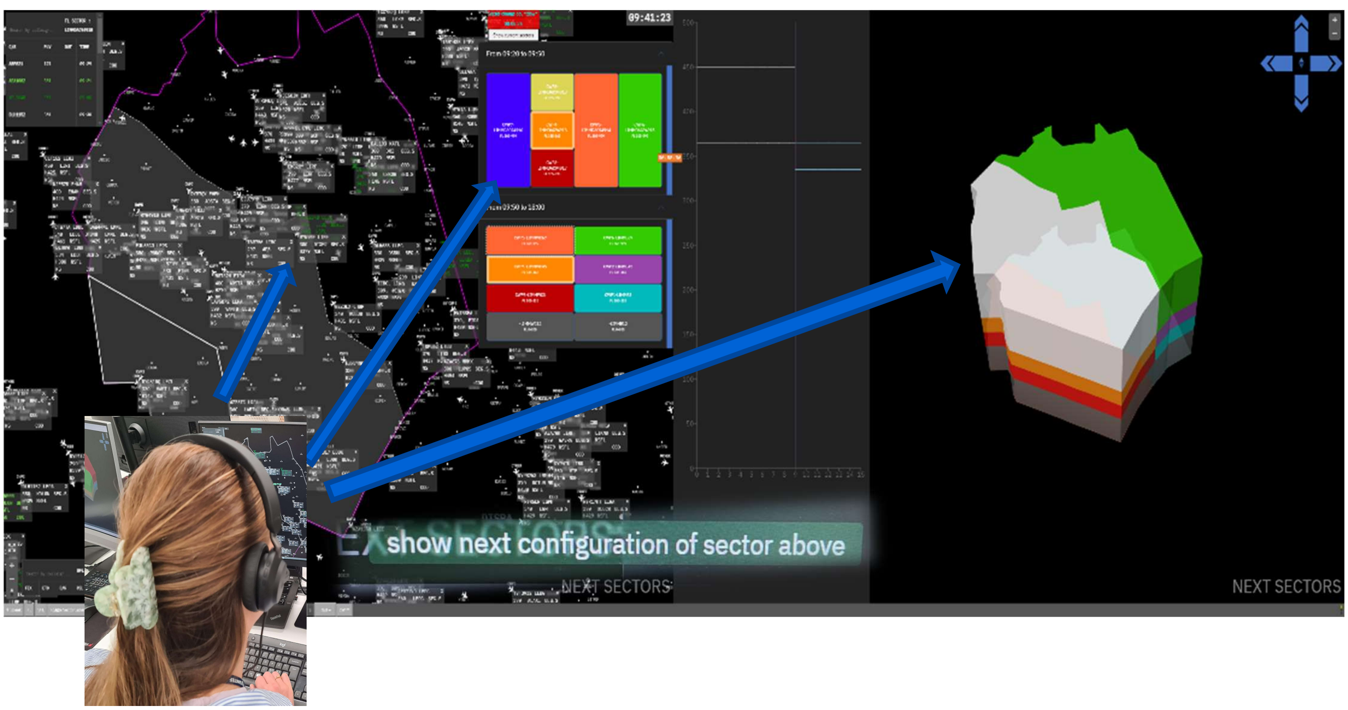

Figure 72 – The speech control front-end to the SIMADES CWP

The speech control front-end to the SIMADES CWP (see

) lets the ATCO use speech to control certain functionality in the CWP. The speech control is activated either by pressing a button on the CWP screen or by using a press-to-talk button on the keyboard (the latter is the expected preferred method).

The available voice commands include commands for

- visualizing sectors configurations
- controlling the 3D view
- highlighting flights affected by an upcoming sectorization change
- turning speed vectors and trajectories on and off, as well as setting the length of a speed vectors
- setting altitude filter
- toggle fixes
- toggle flight labels in own and other sectors
- manipulating flight lists

The speech front-end is implemented using state-of-the-art software for speech control, as shown in Figure 73.

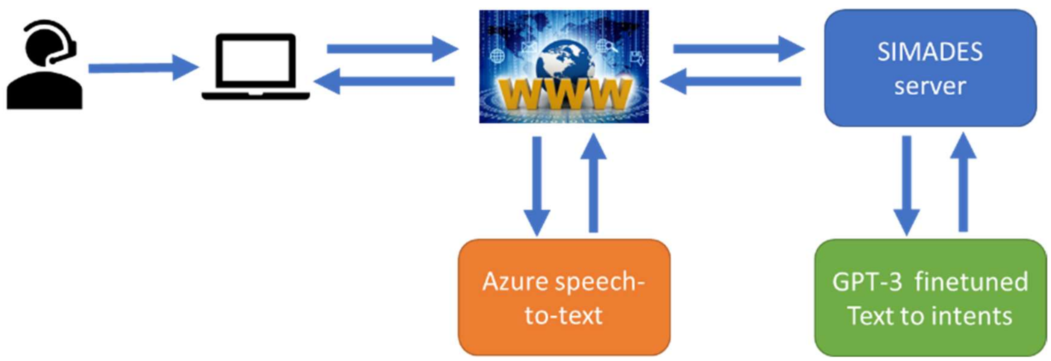

Figure 73 – Architecture for the speech control front-end to SIMADES CWP

Figure 74 – Set-up of SIMADES CWPs for five ATCO positions and a master controller

D.3 Summary of Exercise #05 Technological Validation Objectives and success criteria

SESAR Solution Validation Objective	SESAR Solution Success criteria	Coverage and comments on the coverage of SESAR Solution Validation Objective in Exercise 03	Exercise Validation Objective	Exercise Success criteria
OBJ-Sol.96ASR-TRL6-TVALP-0010	CRT-Sol.96ASR-TRL6-TVALP-0010.001	Fully covered	EX05-OBJ-Sol.96ASR-TRL6-TVALP-0010 same description as OBJ-Sol.96ASR-TRL6-TVALP-0010 To assess the technical feasibility of the integration of the ASR system and its sub-systems into CWP and interoperability between the ASR sub-systems and the	EX05-CRT-Sol.96ASR-TRL6-TVALP-0010.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0010.001 The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the

			existing CWP systems and tools.	existing CWP systems and tools. Availability of systems and tools and their performance remain at 100%
OBJ-Sol.96ASR-TRL6-TVALP-0020	CRT-Sol.96ASR-TRL6-TVALP-0020.001	Fully covered	EX05-OBJ-Sol.96ASR-TRL6-TVALP-0020 same description as OBJ-Sol.96ASR-TRL6-TVALP-0020 To assess the stability of the ASR system performance	EX05-CRT-Sol.96ASR-TRL6-TVALP-0020.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0020.001 The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)
	CRT-Sol.96ASR-TRL6-TVALP-0020.002	Fully covered		EX05-CRT-Sol.96ASR-TRL6-TVALP-0020.002 same description as CRT-Sol.96ASR-TRL6-TVALP-0020.002 The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.
OBJ-Sol.96ASR-TRL6-TVALP-0030	CRT-Sol.96ASR-TRL6-TVALP-0030.001	Fully covered	EX05-OBJ-Sol.96ASR-TRL6-TVALP-0030 same description as OBJ-Sol.96ASR-TRL6-TVALP-0030 To assess the impact on the human performance of the integration of the ASR system and its sub-systems into operations	EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.001 The introduction of the ASR system into the context of application is operationally viable, ATCos workload with ASR is equal or better

			in a realistic environment	than in baseline (without ASR support)
	CRT-Sol.96ASR-TRL6-TVALP-0030.002	Fully covered		EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.002 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.002 The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate)
	CRT-Sol.96ASR-TRL6-TVALP-0030.003	Fully covered		EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.003 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.003 The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. Controllers' feedback with respect to Human Factors questionnaire is better than for baseline)
	CRT-Sol.96ASR-TRL6-TVALP-0030.004 The number and/or severity of human errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.	Not covered		

	Furthermore more than 50% of command recognition errors and command recognition rejections are detected by the controllers and manually corrected.			
	CRT-Sol.96ASR-TRL6-TVALP-0030.005 Task allocation between human and machine, resulting from the introduction of the ASR system support, is rated as good as in baseline (with respect to feedback to Human Factors Questionnaire).	Not covered		
	CRT-Sol.96ASR-TRL6-TVALP-0030.006	Fully covered		EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.006 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.006 Changes in the design of the user interface (input devices, visual displays/output devices, alarm& alerts) support ATCos in carrying out the tasks.
	CRT-Sol.96ASR-TRL6-TVALP-0030.007	Fully covered		EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.007 same description as CRT-Sol.96ASR-TRL6-TVALP-0030.007 The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary

				mitigations are identified)
OBJ-Sol.96ASR-TRL6-TVALP-0040	CRT-Sol.96ASR-TRL6-TVALP-0040.001	Fully covered	EX05-OBJ-Sol.96ASR-TRL6-TVALP-0040 same description as OBJ-Sol.96ASR-TRL6-TVALP-0040 To assess the impact of the introduction of the ASR system on safety.	EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.001 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.001 The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.
	CRT-Sol.96ASR-TRL6-TVALP-0040.002	Fully covered		EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.002 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.002 The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.
	CRT-Sol.96ASR-TRL6-TVALP-0040.003	Fully covered		EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.003 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.003 The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.

	CRT-Sol.96ASR-TRL6-TVALP-0040.004	Fully covered		EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.004 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.004 The level of ATCO's situational awareness is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions).
	CRT-Sol.96ASR-TRL6-TVALP-0040.005	Fully covered		EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.005 same description as CRT-Sol.96ASR-TRL6-TVALP-0040.005 The level of ATCOs' workload is maintained or decreased with the introduction of ASR system.
	CRT-Sol.96ASR-TRL6-TVALP-0040.006 The recovery means that errors resulting from the introduction of the ASR system are identified to minimise operational impact.	Not covered		
OBJ-Sol.96ASR-TRL6-TVALP-0050 To assess the impact of the introduction of the ASR system on capacity.	CRT-Sol.96ASR-TRL6-TVALP-0050.001 The workload of ATCO after introduction of an ASR system is adequate to increase TMA capacity. The workload of ATCOs is the same or less when working with ASR	Not covered		

	compared to baseline. The average flight time of the aircraft is expected to be reduced with respect to baseline without ASR support due to less time needed by the ATCO to complete task for one aircraft. ATCO should then have more time available for other aircraft and more timely task execution with regard to the flight progressing through their airspace. This should result in more optimum trajectories.			
	CRT-Sol.96ASR-TRL6-TVALP-0050.002 ASR allows ATCOs to safely manage a higher amount of aircraft, increasing the throughput in TMA	Not covered		
OBJ-Sol.96ASR-TRL6-TVALP-0060 To assess the impact of the introduction of the ASR system on Fuel efficiency	CRT-Sol.96ASR-TRL6-TVALP-0060.001 Aircraft will be able to improve their route Efficiency (fuel burnt) due to the higher throughput in TMA thanks to the introduction of ASR	Not covered		
	CRT-Sol.96ASR-TRL6-TVALP-0060.002 Aircraft will be able to improve their route Efficiency (flight time) due to the higher throughput in TMA thanks to the introduction of ASR	Not covered		
OBJ-Sol.96ASR-TRL6-TVALP-0070	CRT-Sol.96ASR-TRL6-TVALP-0070.001	Fully covered	EX05-OBJ-Sol.96ASR-TRL6-TVALP-0070 same description as OBJ-	EX05-CRT-Sol.96ASR-TRL6-TVALP-0070.001 same description as

			Sol.96ASR-TRL6-TVALP-0070 To assess the impact of the introduction of the ASR in visualization navigation	CRT-Sol.96ASR-TRL6-TVALP-0070.001 ATCOs are able to perform a faster and more predictable navigation when using ASR to support DAC.

D.4 Summary of Technological Validation Exercise #05 Validation scenarios

EXE-005 simulated the Milan ACC en-route airspace. In the ordinary sessions the controllers started with a predefined, commonly used configuration (sectorization) of the Milan ACC. This configuration was chosen by an FMP based on the traffic volume. After 20 minutes the sectorization changed to a new configuration suggested by a DAC algorithm. After another 30 minutes, the sectorization changed back to another predefined configuration.

ATCOs were operational controllers from Milan ACC as well as Norwegian controllers from Avinor. The controllers from ENAV have good knowledge of the airspace, the traffic patterns and the predefined configurations, but the sectorization coming from the DAC algorithm was not known for them. The Avinor controllers had limited knowledge of the Milan ACC.

Before the exercise, the ATCOs had received a list of the voice commands and their effects. Before the runs with the CWP, the commands were presented by in a classroom session lead by the EXE leader.

The ATCOs participated in three different types of runs.

1. Training session with frozen traffic. The purpose with this session was twofold. First, the ATCOs should get accustomed to the voice control command. Second, the voice commands during the training were used to train the underlying ASR machine learning functionality. Even though the traffic was frozen, the planes were available and could be interacted with using mouse/keyboard and/or voice commands. The reason for keeping the traffic frozen was to ensure that the ATCOs had full focus on the voice commands. In this session, both ATCOs and pseudo pilots used the voice commands. As part of the debriefing, the controllers were given an opportunity to suggest additional voice commands. Some of these were implemented for the next runs.

2. Training session with live traffic. This was like an ordinary session but conducted like a usability test. In this session, all ATCOs and pseudo pilots were encouraged to use voice commands as much as possible. The training of machine learning and data collection was similar to the training session with frozen traffic. As part of the debriefing, the controllers were given an opportunity to suggest additional voice commands. Some of these were implemented for the next runs.

3. Ordinary sessions. In these sessions, the same scenario was run twice, having the same traffic, same sectorization, and same manning of the different sectors in both runs. In each run half of the controllers were encouraged to used voice commands, while the other half was asked

not to used it. Which ATCOs that were supposed to use voice control was switched between the two sessions. Pseudo pilots were free to use voice commands as they wished. The training of machine learning and data collection used in the training was also applied in these sessions. In addition, data was collected on workload, performance, and conflicts. Furthermore, a questionnaire was filled in after two sessions. Data and questionnaires were only collected for ATCOs.

The sessions were performed on integrated controller positions: one controller performed the executive and planning roles. One pseudo pilot was assigned to each position. The pseudo pilots were active pseudo pilots or ATCOs. The pseudo pilots used a version of the CWP with additional functionality for manipulating the flights in the simulator. The pseudo pilots also had the tools supporting DAC, as well as the speech control. During EXE-005 five ATCO and five pseudo pilots positions were manned. In addition, a master controller dispatched the traffic to the different sectors before the sessions started. The master controller also managed the traffic in the sectors not manned by ATCOs taking part in the session. The master controller used a variant of the CWP for pseudo pilots having some additional functionality.

The exercise used two different traffic samples. Different traffic samples were used in the training session with live traffic and the ordinary sessions. The traffic sample used for the ordinary sessions in PJ10 was also used in various PJ.09 sessions, but in these sessions the manning of sectors were different.

The same controllers and pseudo pilots participated in the PJ.09 and PJ.10 parts of the exercise. The training sessions (1 and 2 above) were sessions only for PJ.10. The first ordinary sessions were common sessions for both projects[9]. The PJ09 part had a number of sessions from which no data were collected for PJ10. Still, these sessions were valuable for PJ.10 as it made the controllers familiar with the non-speech part of the CWP. It is important to stress that the speech control is only available for parts of the CWP functionality. This means that even in the speech control sessions, the ATCOs had to use mouse/keyboard interaction for much of their work, like accepting flights, giving new speed and heading, and handing the flight over to the next sector. Furthermore, speech control was used occasionally also in these sessions, but no data were collected except that the experience from this use were collected during interviews.

D.5 Summary Technological Validation Exercise #05 Assumptions

The following assumptions apply to this exercise:

Identifier	Title	Description	Justification	Impact on Assessment
AS-EX05-01	Datalink	No data link will be used during the exercise	The exercise is centred on the radio communication between controllers and pilots	Low

[9] I.e., only the data collected for the first of the ordinary sessions were used in the PJ09 part. The questionnaires were common for the two ordinary sessions.

AS-EX05-02	Language	ATCOs and pilots will communicate in English	This language is operational in Italian (and Norwegian) airspace	N/A
AS-EX05-03	Ambient noise	Noise generated by conversation in a control Room	Ambient noise will be limited to what can be generated by conversation of other ATCOs present in a control room. No simulation of aircraft generated noise is foreseen	Medium
AS-EX05-04	Weather	Normal/good	No impact analysis of abnormal weather conditions	Low
AS-EX01-5	Traffic conditions	Traffic conditions will be regular in terms of flow and amount	The test reflects a "normal" setting	Low

Table 51: Technological Validation Assumptions overview

D.6 Deviation from the planned activities

Deviations from the planned activities that do not impact objectives or success criteria:

- Deviation EXE005-1. Initially, there were more ordinary sessions planned. The additional sessions were cancelled due to time constraints, technical issues the first two days of the common exercise, and various unforeseen technical and practical limitations.
- Deviation - EXE005-2. During the second ordinary session the simulator crashed after approximately 40 minutes (of a planned duration of 60 minutes). To be able to compare the two ordinary sessions, only data from the first 40 minutes are used in both these sessions. The main loss was workload data to be collected during the second sectorization change.

Deviations from the planned activities that impact objectives or success criteria:

- Deviation - EXE005-3. Initially, there was a plan to support UC.2 in the CWP. This turned out to be technically challenging to implement. As this use case is addressed in all the other exercises in PJ.10-W2-Sol.96 ASR, it was decided to focus on the use cases not addressed in other exercises in PJ.10-W2-Sol.96 ASR.

D.7 Technological Validation Exercise #05 Validation Results

D.7.1 Summary of Technological Validation Exercise #05 Results

Technological Validation Exercise #05 Validation Objective ID	Technological Validation Exercise #05 Validation Objective Title	Technological Validation Exercise #05 Success Criterion ID	Technological Validation Exercise #05 Success Criterion	Technological Validation Exercise #05 Results	Technological Validation Exercise #05 Validation Objective Status
OBJ-Sol.96ASR-TRL6-TVALP-0010	To assess the technical feasibility of the integration of the ASR system and its sub-systems into CWP and interoperability between the ASR sub-systems and the existing CWP systems and tools.	EX05-CRT-Sol.96ASR-TRL6-TVALP-0010.001	The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the existing CWP systems and tools. Availability of systems and tools and their performance remain at 100%	The ASR was successfully implemented in the CWP. The ASR did not impact the performance of rest of the CWP.	OK
OBJ-Sol.96ASR-TRL6-TVALP-0020	To assess the stability of the ASR system performance	EX05-CRT-Sol.96ASR-TRL6-TVALP-0020.001	The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)	Recognition rates of voice commands in the ordinary runs were considered to be high for most controllers. Response time for the voice commands were not satisfactory. .	POK
		EX05-CRT-Sol.96ASR-TRL6-TVALP-0020.002	The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.	There were no observable differences in recognition rate between the different commands.	OK
OBJ-Sol.96ASR-TRL6-TVALP-0030	To assess the impact on the human performance of the integration of the ASR system and its	EX05-CRT-Sol.96ASR-TRL6-	The introduction of the ASR system into the context of application is operationally viable, ATCos workload with ASR is equal or better	WL was in general low, but slightly higher when using ASR in a majority of the	POK

	sub-systems into operations in a realistic environment	TVALP-0030.001	than in baseline (without ASR support)	sessions. This may well have other causes than ASR.	
		EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.002	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate)	Recognition rates were good, but the response time was too slow.	POK
		EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.003	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. ~~Controllers' feedback with respect to Human Factors questionnaire is better than for baseline.~~	Response time was considered by the ATCOs to reduce the value of some of the voice commands. The success criterion is partly not applicable.	POK
		EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.006	Changes in the design of the user interface (input devices, visual displays/output devices, alarm& alerts) support ATCos in carrying out the tasks.	The voice commands were much used by most ATCOs. Controllers considered the commands to be appropriate, the ASR functionality worked satisfactorily, and the commands for enhancing the understanding of the sectorization were useful. The controllers found that speech commands enhanced the CWP, and that it should be	OK

				included in a CWP supporting DAC.	
		EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.007	The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified)	The level of trust in the functionality of the speech recognition functionality was high. Technical issues with the ARS caused some frustration for the ATCOs and reduced the trust in the running system.	POK
OBJ-Sol.96ASR-TRL6-TVALP-0040	To assess the impact of the introduction of the ASR system on safety.	EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.001	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.	Issues regarding recognition rates did not affect safety.	OK
		EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.002	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.	Issues regarding response times did not affect safety.	OK
		EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.003	The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.	No indications of errors resulting from the introduction of the ASR. There were slightly more conflicts for ATCOs using ASR than the ones not using it. This may well have other causes than ASR.	POK
		EX05-CRT-Sol.96ASR-TRL6-	The level of ATCO's situational awareness is not reduced with the	Situational awareness was in general high. The	OK

		TVALP-0040.004	introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions).	use of ASR did not affect the SA negative	
		EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.005	The level of ATCOs' workload is maintained or decreased with the introduction of ASR system.	WL was in general low, but slightly higher when using ASR in a majority of the sessions. This may well have other causes than ASR.	POK
OBJ-Sol.96ASR-TRL6-TVALP-0070	To assess the impact of the introduction of the ASR in visualization navigation	EX05-CRT-Sol.96ASR-TRL6-TVALP-0070.001	ATCos are able to perform a faster and more predictable navigation when using ASR to support DAC.	ATCOs' performance was not affected by the use of ASR.	POK

Table 52: Technological Validation Results Exercise 5

1. Results per KPA

Objectives are stated per KPA. Please refer to the next section for the feedback regarding:

- Human performance (3 OBJ-Sol.96ASR-TRL6-TVALP-0030 Results)
- Safety (4 OBJ-Sol.96ASR-TRL6-TVALP-0040 Results)

Security related to the ASR functionality was discussed during the final debriefing. No security concern was identified in relation with the use cases implemented.

D.7.2 Analysis of Exercise 5 Results per Technological Validation objective

This section provides, per Technological Validation objective, a consolidated analysis of the Technological Validation exercise 5 results.

1. OBJ-Sol.96ASR-TRL6-TVALP-0010 Results

Objective description: *To assess the technical feasibility of the integration of the ASR system and its sub-systems into CWP and interoperability between the ASR sub-systems and the existing CWP systems and tools.*

Validation Objective ID	Success Criterion ID	Success Criterion Status	Validation Objective Status
EX05-OBJ-Sol.96ASR-TRL6-TVALP-0010	EX05-CRT-Sol.96ASR-TRL6-TVALP-0010.001	OK	OK

EXE-005 added speech recognition functionality to the SIMADES CWP. The CWP does not support voice communication between ATCOs and pseudo pilots. In the exercise, Microsoft Teams with a press-to-talk add-on was used for ATCO-pilot communication. Voice commands were turned on either by clicking on a button in the CWP HMI or preferably using a keyboard shortcut. It was automatically turned off when a command had been processed. There were special labels on the keyboard keys for ATCO-pilot communication and for giving voice commands.

The ASR was successfully implemented in the CWP. The ASR did not impact the performance of rest of the CWP

a. Outcome Analysis

There was no impact on other systems or tools of the CWP.

The ASR-system did not interfere with the availability and/or reliability of the other parts of the CWP.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0010.001	The ASR system and its subsystems and functions are able to integrate with the CWP systems and subsystems without negatively affecting the performance and availability of the existing CWP systems and tools. Availability of systems and tools and their performance remain at 100%.	OK

2. OBJ-Sol.96ASR-TRL6-TVALP-0020 Results

Objective description: To assess the stability of the ASR system performance.

Success Criterion ID	Success Criterion	Success Criterion Status	Validation Objective Status
EX05-OBJ-Sol.96ASR-TRL6-TVALP-0020	EX05-CRT-Sol.96ASR-TRL6-TVALP-0020.001	POK	POK
	EX05-CRT-Sol.96ASR-TRL6-TVALP-0020.002	OK	

Success Criteria:

CRT-Sol.96ASR-TRL6-TVALP-0020.001 The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.).

CRT-Sol.96ASR-TRL6-TVALP-0020.002 The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises.

a. Outcome Analysis

Recognition rates of voice commands in the ordinary runs were considered to be high for most controllers. The only female controller experienced lower recognition rates, and decreasing recognition throughout the exercise. The performance with regards to response time for the voice commands were not satisfactory and caused the controllers to not use them in situations where they might have been very useful if response time had been faster. The response time issue is inherent in the state-of-the-art technology used in the exercise, and is currently a problem for web-based speech recognition.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001	The required ASR performance is maintained as required in TS/IRS (Command Recognition Rate, command Recognition Error Rate, etc.)	POK

There were no observable differences in recognition rate between the different commands. In the test runs some commands had worse recognition than others, and there were some issues in recognizing certain numbers. These issues disappeared when the machine learning had worked on the data from the training sessions.

Success Criterion ID	Success Criterion	Success Criterion Status
EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.002	The required level of ASR performance does not show differences greater than 2.5% among the different command types tested in the exercises	OK

3. OBJ-Sol.96ASR-TRL6-TVALP-0030 Results

Objective description: To assess the impact on the human performance of the integration of the ASR system and its sub-systems into operations in a realistic environment.

Success Criterion ID	Success Criterion	Success Criterion Status	Validation Objective Status
EX05-OBJ-Sol.96ASR-TRL6-TVALP-0030	EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.001	POK	POK
	EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.002	POK	
	EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.003	POK	

	EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.006	OK	
	EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.007	POK	

Success Criteria:

CRT-Sol.96ASR-TRL6-TVALP-0030.001 The introduction of the ASR system into the context of application is operationally viable, **ATCOs workload with ASR is equal or better** than in baseline (without ASR support).

CRT-Sol.96ASR-TRL6-TVALP-0030.002 **The accuracy of the information provided** by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate).

CRT-Sol.96ASR-TRL6-TVALP-0030.003 **The timeliness of the information provided** by the ASR system is adequate for the accomplishment of operations. Controllers' feedback with respect to Human Factors questionnaire is better than for baseline).

CRT-Sol.96ASR-TRL6-TVALP-0030.006 **Changes in the design of the user interface** (input devices, visual displays/output devices, alarm& alerts) support ATCOs in carrying out the tasks.

CRT-Sol.96ASR-TRL6-TVALP-0030.007 **The level of trust in the ASR system** and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified).

a. Outcome Analysis

1. Workload

Subjective experience of workload was assessed both during and at the end of the run. During the sessions, the controllers were asked orally to report current workload after 20 and 35 minutes. The first time was around the first sectorization change, the second was in the middle of the run with a new and unknown sectorization. WL was also collected after 50 minutes (time for second sectorization change) for the first ordinary session, but as the simulation was abandoned due to technical issues after appr. 40 minutes in the second ordinary session, these data are not used in the analyses.

At the end of run each the controllers were asked to fill-in post-run questionnaires. They were asked to report the workload they experienced during the sessions on the scale from 1 to 10 where 1 corresponds to 'workload insignificant' and 10 corresponds to 'task unsustainable'. Due to a misunderstanding, each ATCOs filled in just one questionnaire for the two ordinary sessions. Thus, only the workload data obtained during the runs may be used to compare the workload with and without using ASR. It should though be noted that the WL levels from the questionnaires are consistent with the WL data obtained during the runs.

54 shows the collected workload data for the two sessions.

	Orinary session 1		Ordinary session 2	
	Average WL trad UI	Average WL ASR	Average WL trad UI	Average WL ASR
20 min.	1.666666667	1.75	1	1.666666667
35 min.	1.333333333	1	1	1.666666667

Table 53: Workload data (1=low, 2=medium, 3=high)

In the first session, the WL is slightly higher for ATCOs using ASR after 20 minutes, and slightly lower after 35 minutes. In the second session, WL is higher for ATCOs using ASR than for the ones not using it. All average data represent low workload.

The underlying data shows that the workload is unevenly distributed, i.e., it is rather low for four of the sectors, and quite high for the fifth one. This is mainly due to a weakness in the DAC solution for used for these sessions. Furthermore, the workload data are also influenced by the dispatch work done by the master controller before and during the sessions. The WL data for all sectors indicate that the traffic distribution between the sectors were different in the two sessions, giving the ATCO in the sector with most traffic more traffic in the second than in the first session. As the ATCO in this sector used ASR in the second run, this is probably more important for the WL data than the use of ASR. During the debriefings controllers gave no indication of increased workload due to use of ASR.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.001	The introduction of the ASR system into the context of application is operationally viable, ATCOs workload with ASR is equal or better than in baseline (without ASR support).	POK

2. Information completeness and timelines

As presented in section D.7.2C.7.22 OBJ-Sol.96ASR-TRL6-TVALP-0020 Results the recognition rates were good, but the response time was too slow. The ASR functionality did not provide any feedback except for a transcript of the recognized commands at the bottom of the CWP screen (for a short time), the success criterion is partly not applicable. Thus, the criterion is rated equal to EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.002	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations with respect to requirements in TS/IRS (Command Recognition Rate, command Recognition Error Rate).	POK

As indicated above, the response time was considered by the ATCOs to reduce the value of some of the voice commands. The ASR functionality did not provide any feedback except for a transcript of the recognized commands at the bottom of the CWP screen (for a short time), the success criterion is partly not applicable. Thus, the criterion is rated equal to EX03-CRT-Sol.96ASR-TRL6-TVALP-0020.001.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.003	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations. ~~Controllers' feedback with respect to Human Factors questionnaire is better than for baseline.~~	POK

3. HMI support

The general questions in the questionnaires on the use of voice commands show that it was much used by most ATCOs (average 8.6 on a scale from 1-10, a majority of the ATCOs scoring 10). All controllers answered that the commands were appropriate, while 3 of 5 ATCOs answered that the ASR functionality worked satisfactorily. The answers to the questionnaires also show different scores for different command types regarding usefulness. All controllers found the commands for enhancing the understanding of the sectorization to be useful, while the opinions diverged regarding the commands for understanding how traffic was affected by sectorization changes and for controlling the 3D visualization. 3D navigation was also mentioned in the interviews as a type of functionality that would have been more useful if the response to the voice commands had been faster. A majority of the controllers found that speech commands enhanced the CWP, and that it should be included in a CWP supporting DAC.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.006	Changes in the design of the user interface (input devices, visual displays/output devices, alarm & alerts) support ATCOs in carrying out the tasks	OK

4. Trust

Interviews and questionnaires indicate that the level of trust in the functionality of the speech recognition functionality was high. There were though some technical issues with the software used for ASR causing it to occasionally freeze. This did not affect the rest of the CWP, but when this happened, the ATCOs had to refresh the browser page for the CWP to reactivate the speech commands. This caused some frustration for the ATCOs and reduced the trust in the running system.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0030.007	The level of trust in the ASR system and its sub-systems and functions is appropriate (potential issues related to trust and preliminary mitigations are identified).	POK

4. OBJ-Sol.96ASR-TRL6-TVALP-0040 Results

Objective description: To assess the impact of the introduction of the ASR system on **safety.**

Success Criterion ID	Success Criterion	Success Criterion Status	Validation Objective Status
EX05-OBJ-Sol.96ASR-TRL6-TVALP-0040	EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.001	OK	OK
	EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.002	OK	

EUROPEAN PARTNERSHIP

Co-funded by the European Union

	EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.003	POK	
	EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.004	OK	
	EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.005	POK	

Success Criteria:

CRT-Sol.96ASR-TRL6-TVALP-0040.001 The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.

CRT-Sol.96ASR-TRL6-TVALP-0040.002 The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations.

CRT-Sol.96ASR-TRL6-TVALP-0040.003 The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.

CRT-Sol.96ASR-TRL6-TVALP-0040.004 The level of ATCO's situational awareness is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions).

CRT-Sol.96ASR-TRL6-TVALP-0040.005 The level of ATCOs' workload is maintained or decreased with the introduction of ASR system.

a. Outcome Analysis

1. Accuracy and timeliness

There are no indications that the results presented above regarding recognition rates affected safety.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.001	The accuracy of the information provided by the ASR system is adequate for the accomplishment of operations. Command Recognition Error Rate stays in the acceptable limits.	OK

There are no indications that the results presented above regarding response times affected safety.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.002	The timeliness of the information provided by the ASR system is adequate for the accomplishment of operations	OK

There are no indications of errors resulting from the introduction of the ASR. Still, we investigated possible conflicts during the two ordinary sessions. There were 17 conflicts in the first, and 18 in the second run. The data shows that there on average were slightly more conflicts for ATCOs using ASR than the ones not using it. As with workload findings, these variations may just as well be caused by imbalance in the dispatch of the traffic between the two runs. This is supported by the fact that for three of the sectors there are fewer conflicts when using ASR than without, while there are more conflicts with ASR in the two others. The reason why the averages are higher for ASR overall is that there in general were more conflicts in the two sectors with more conflicts when using ASR.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.003	The number and/or severity of errors resulting from the introduction of the ASR system is within tolerable limits, taking into account error type and operational impact.	POK

2. Situational awareness

Data on situational awareness was collected only in the questionnaires, and as they due to a misunderstanding was filled in for the two sessions together, they cannot be used to determine changes in situational awareness. On the other hand, the data from the questionnaire shows an average SA of 8.4 on a scale from 1 to 10 (all ATCOs scored 8 or 9). This indicates that the use of ASR did not affect the SA negatively. This is also supported by the interviews, where no ATCOs expressed that the SA was decreased when using ASR.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.004	The level of ATCO's situational awareness is not reduced with the introduction of ASR system (ATCO is able to perceive and interpret task relevant information and anticipate future events/actions)	OK

3. Workload

As presented above, the WL level were slightly higher for the ATCOs using ASR than the ones not using it, but this may just as well be explained by a combination of differences in the dispatch of traffic between the sessions and the unbalance between the workload between the sectors in parts of the sessions.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0040.005	The level of ATCOs' workload is maintained or decreased with the introduction of ASR system.	POK

EUROPEAN PARTNERSHIP

Co-funded by the European Union

5. OBJ-Sol.96ASR-TRL6-TVALP-0070 Results

Objective description: To assess the impact of the introduction of the ASR in visualization navigation.

Validation Objective ID	Success Criterion ID	Success Criterion Status	Validation Objective Status
EX05-OBJ-Sol.96ASR-TRL6-TVALP-0070	EX05-CRT-Sol.96ASR-TRL6-TVALP-0070.001	POK	POK

Success criterion: CRT-Sol.96ASR-TRL6-TVALP-0070.001 ATCOs are able to perform a faster and more predictable navigation when using ASR to support DAC.

a. Outcome Analysis

To measure this success criterion the number of flights being controlled by each ATCO were analysed. 154 flights were controlled in the first session, 152 flights in the second. In two of the sectors more flights were controlled when using ASR than when not. In the other three sectors the opposite was the case. In all sectors the difference was 4 flights or less, and summed, the non-ASR sectors had a throughput of 4 flights more than the ASR sectors. A difference of appr. 1% is ignorable. Furthermore, the interviews did not reveal any indications that the use of ASR influenced the performance, neither negatively nor positively.

Success Criterion ID	Success Criterion	Success Criterion Status
EX05-CRT-Sol.96ASR-TRL6-TVALP-0070.01	ATCOs are able to perform a faster and more predictable navigation when using ASR to support DAC.	POK

D.7.3 Unexpected Behaviours/Results

The radar screen part of platform used in the combined exercise mimicked a simplified version of the CWP being used by the Italian controllers involved in the exercise. These simplifications were consciously chosen to allow using as much of the development efforts as possible on the functionality supporting DAC. These simplifications made the Italian controllers a bit reluctant to the prototype, as they were missing functionality they are used to in their day-to-day work. For the Norwegian controllers, the radar screen was significantly different from the one they are used to. In addition, they were not familiar with neither the airspace, the traffic patterns, and in some cases also the call signs. This caused quite a lot of frustration among the Norwegian controllers in the start of the combined exercise.

Most of the issues just described did not influence the ASR training sessions significantly as they focused on speech control for the functionality supporting DAC. Furthermore, the runs in which using ASR was compared with not using ASR was conducted at a point in time where these issues had been resolved.

D.7.4 Confidence in Results of Validation Exercise 5

1. Level of significance/limitations of Technological Validation Exercise Results

EXE-005 involved more than 10 ATCOs and pseudo pilots from Italy and Norway. The ATCOs have varying age and experience, although all are considered as experienced. All but one controller are male, but this may also be viewed as more representative for ATCOs than for other professions. The airspace and traffic used in the exercise are both realistic. Training on the platform to which the ASR functionality were added was limited, but as the training on the ASR system was also used for data collection, this is not considered a limitation. The basic functionality of the radar screen was limited compared to what is usually provided for the ATCOs involved. As the focus of the combined exercise was supporting ATCOs when using DAC, this is not considered a limitation neither.

Based on these considerations the representativeness of the technological validation results is considered as high.

2. Quality of Technological Validation Exercises Results

The ATCOs were interviewed after each most runs and after the exercise. The ATCOs also answered a questionnaire after the runs in which using ASR was compared with not using ASR. As explained above, the results from parts of the questionnaire are limited as there was a common questionnaire for these two sessions. Thus, the quality of the qualitative data collected is medium to high.

Various logs and observations resulted in quantitative data that are considered as having high quality.

3. Significance of Technological Validation Exercises Results

The significance of the qualitative results is considered as high. The limitations just explained of parts of the questionnaire primarily influence the quantitative data collected through the questionnaires.

The statistic significant of the quantitative data is considered as low. As explained in different parts of Section D7.2 above, it is not guaranteed that the traffic and traffic distribution between the two sessions being compared were identical in the two runs. Also, the limitations of the results for parts of the questionnaire also supports this conclusion. Furthermore, as only 5 ATCOs were involved in these two runs, differences in findings need to be large to claim statistical significance.

D.8 Conclusions

Controllers assessed the voice commands to be appropriate for their tasks, and the recognition rate was considered as high, and increasing during to exercise due to training of underlying machine learning mechanisms from the ATCOs' commands. The ASR system was a bit unstable, and the response time was too slow for the ATCOs to issue commands in parallel to working in other parts of the CWP. With the current response time, the ATCOs assessed the voice commands to be more useful for planner controllers than for executive controllers. With better response time the voice commands would be very useful for executive controllers as well. The usefulness of the use cases tested was:

- UC.5 **Voice commands for highlighting an upcoming sectorization change in the CWP.** The commands supporting this UC were considered as most useful in the current implementation.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

- UC6 **Voice commands for highlighting the flights that will be affected by an upcoming sectorization change in the CWP**. The commands supporting this UC were also considered very useful, but as the underlying functionality triggered by the commands had limitations, enhancements are needed.

UC7 **Voice commands for navigating the 3D visualization of the airspace in the CWP.** The commands supporting this UC were the ones suffering most from the response time issues. With a better response time, these commands would be useful to adjust the 3D view while working in the Radar view using mouse/keyboard.

D.8.1 Conclusions on Technological feasibility

The exercise confirmed the feasibility of integrating voice commands in a CWP supporting DAC supporting these use cases:

an Automatic Speech Recognition tool in an operational Controller working position with an operational communication system to support the use cases:

- UC.5 **Voice commands for highlighting an upcoming sectorization change in the CWP.**
- UC6 **Voice commands for highlighting the flights that will be affected by an upcoming sectorization change in the CWP**.
- UC7 **Voice commands for navigating the 3D visualization of the airspace in the CWP.**

The integration was performed without impacting the performance of other parts of CWP. The voice commands were implemented using web-based technologies. The ASR preindustrial prototype was able to process voice commands to the web based CWP and provide response to the commands using the appropriate parts of the GUI HMI.

D.8.2 Conclusions on performance assessments

The use of voice commands did not affect the performance (situational awareness, workload, throughput, and safety) significantly neither positively nor negatively. With faster response times, ATCOs assessed that voice commands could improve ATCOs' productivity.

D.8.3 Recommendations

The main recommendation from the validation is to improve response time from the ASR system. As this part of the prototype was outside the control of our implementation, and state-of-the-art technology was used, this shows that voice commands for a web-based CWP is not yet mature enough to support executive controllers. Until the underlying technology improves sufficiently, voice commands may still be used, but mainly to support tasks that are not time critical. The solution was considered to be useful for planner controllers, thus, to it is recommended to do further development and tests for planners.

Recommendations for the supported UC are:

- UC.5 **Voice commands for highlighting an upcoming sectorization change in the CWP.** Commands that are useful for planner controllers should be emphasized. Executive controllers may use these voice commands when starting a shift, and in low WL situations.

- UC6 **Voice commands for highlighting the flights that will be affected by an upcoming sectorization change in the CWP**. The underlying functionality needs to be improved. With this in place, simple commands like the ones already available have a large potential.

UC7 **Voice commands for navigating the 3D visualization of the airspace in the CWP.** With the current technology available for Web-based voice commands, the functionality supporting this UC should mainly be used by planner controllers. When the technology provides acceptable response times, the functionality may also be used by executive controllers.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

EUROPEAN PARTNERSHIP

Co-funded by
the European Union