INDUSTRIAL RESEARCH

PJ.05-W2 SESAR Sol 97.1 and Sol 97.2 TVALR

Deliverable ID:	**D3.1.050**
Dissemination Level:	**CO**
Project Acronym:	**DTT**
Grant:	**874470**
Call:	**H2020-SESAR-2019-1**
Topic:	**SESAR-WAVE2-PJ05**
Consortium Coordinator:	**DLR [AT-ONE]**
Edition Date:	**19 September 2022**
Edition:	**00.01.14**
Template Edition:	**02.00.03**

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & Approval

Authors of the document

Beneficiary	Date
Leonardo	08 March 2021
NLR (At-One)	12 July 2021
DLR (At-One)	31 March 2022
Indra Navia	03 June 2022
Crida Enaire	07 June 2022
ENAV	21 June 2022
Leonardo	28 June 2022

Reviewers internal to the project

Beneficiary	Date
On (B4)	21 September 2022
Enav	27 July 2021

Reviewers external to the project

Beneficiary	Date

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

Beneficiary	Date

Rejected By - Representatives of beneficiaries involved in the project

Beneficiary	Date

Document History

Edition	Date	Status	Beneficiary	Justification
00.00.01	10 June 2021	Draft	Leonardo	Template

00.00.02	09 June 2021	First Draft	NLR	Draft Contribution EXE-001
00.00.03	12 July 2021	Initial Version	Leonardo NLR	Initial version submitted, including NLR contribution for EXE-001
00.01.01	01 April 2022	Draft	Leonardo	Deletion of activities related to Solution 97.3; new SESAR Template,
00.01.02	07 June 2022	Draft	DLR	Draft Contribution EXE-006
00.01.03	10 June 2022	Draft	Indra	Draft Contribution EXE-004
00.01.04	13 June 2022	Draft	Crida Enaire	Draft Contribution EXE-005
00.01.05	20 June 2022	Draft	ENAV UniBo	Draft Contribution EXE-002
00.01.06	28 June 2022	Draft	Leonardo	Draft Contribution EXE-007
00.01.07	01 July 2022	Draft	Crida Enaire	Modification to Contribution for EXE-005
00.01.08	25 July 2022	Draft	Leonardo	Update to chapters 4, 5
00.01.09	05 August 2022	Draft	Leonardo	Update to Draft contribution for EXE-007
00.01.10	01 September 2022	Draft	ENAV UniBo	Update to Draft contribution for EXE-002
00.01.11	07 September 2022	Draft	DLR	Changes in 6.1.2 feasibility 6.2.1 recommendations 6.2.3 standardisation
00.01.12	15 September 2022	Draft	Leonardo/NLR	Updates to chapters 4, 5
00.01.13	15 November 2022	Version for release	Leonardo/ENAV/Crida	Minor updates and fixes
00.01.14	12 January 2023	Updated release	ENAV SL	Comments by SJU (front page and page numbering)

DTT

PJ.05-W2-WP3

The present Technical Validation Report is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 874470 under European Union's Horizon 2020 research and innovation programme.

Abstract

This document, Technical Validation Report, presents the results of the TRL4 validation activities foreseen for the following solutions:

- PJ.05-W2-97.1 Virtual/Augmented reality applications for tower
- PJ.05-W2-97.2 ASR at the TWR CWP supported by AI and Machine Learning

    - EXE-05.97.1-TRL4-TVALP-VAR-001: Real-time simulation, addressing the use of an Augmented Reality device and Attention Guidance for controllers, performed on NARSIM Tower within an environment for Amsterdam Airport Schiphol.
    - EXE-05.97.1-TRL4-TVALP-VAR-002: Real-time simulation addressing Virtual/Augmented Reality Tower Tools, Tracking Labels and Air Gesture Interaction, carried out at Bologna Airport.
    - EXE-05.97.2-TRL4-TVALP-ASR-004: Real-time simulation addressing Speech Recognition in a multiple remote tower environment.
    - EXE-05.97.1-TRL4-TVALP-VAR-005: Shadow Mode validation regarding Virtual and augmented reality as well as Tracking Label and Air Gestures executed at Vitoria airport.
    - EXE-05.97.2-TRL4-TVALP-ASR-006: Real Time simulation concerning Assistant Based Speech Recognition realized at Braunschweig, simulating three generic (multiple remote) airports adapted from existing airports.
    - EXE-05.97.2-TRL4-TVALP-ASR-007: Real Time simulation addressing Speech Recognition run at Rome, simulating Sofia airport.

Validation exercises were planned to address stakeholders' needs and assess the KPAs of Safety, Cost Efficiency, Capacity (Resilience Focus Area) and Human Performance. Deviations from the TVALP [28] are also provided where found, along with conclusions and recommendations for future TRL4-TRL6 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

The present document offers an account of the TRL4 validation activities conducted for solutions PJ05-W2-97.1 and PJ05-W2-97.1 with 'HMI Interaction modes for Airport Tower' as main topic. Solutions 97.x consist of Operational Improvements and Technical Enablers, each solution covering one OI:

- SOL PJ.05-W2-97.1: POI-0039-SDM. Virtual and augmented reality were used to provide traffic information to ATCOs in order to simplify human system interaction. V&AR were used in different applications (e.g. smart screens, head-on display) in order to help tower ATCOs conducting safe operations under all meteorological conditions, all the while maintaining high taxiway and runway throughputs:
    - AERODROME-ATC-103. Introduction of new Augmented Reality vision systems with tracking labels for a/c and mobiles for improving the controller productivity through increasing heads-up vision.
    - AERODROME-ATC-104. Use of in-air gestures for user interaction, to speed up and make simpler human-system interaction.
    - AERODROME-ATC-105 use of Attention Guidance in V/AR applications to enhance the situational awareness
- SOL PJ.05-W2-97.2: POI-0040-SDM. ATCOs were supported by means of innovative human machine interaction tools such as Automatic Speech Recognition, enhanced by usage of Machine Learning. The main aim consisted in automatically supporting some ATCO tasks, which are either not performed at all or manually in currently available systems / CWPs:
    - AERODROME-ATC-106. 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-05.97.1-TRL4-TVALP-VAR-001 - A real-time simulation addressing the use of new interaction modes and attention guidance for controllers in the aerodrome control tower at Amsterdam Airport Schiphol.
- EXE-05.97.1-TRL4-TVALP-VAR-002 - A real-time simulation addressing Virtual/Augmented Reality Tower Tools, Tracking Labels and Air Gesture Interaction at Bologna Airport.
- EXE-05.97.2-TRL4-TVALP-ASR-004 - A real-time simulation addressing Speech Recognition in a multiple remote tower environment.
- EXE-05.97.1-TRL4-TVALP-VAR-005 - A Shadow mode validation addressing Virtual and augmented reality + Tracking Label and Air Gestures at Vitoria airport.
- EXE-05.97.2-TRL4-TVALP-ASR-006 - A real-time simulation addressing Speech Recognition at Braunschweig simulating at least five generic (multiple remote) airports adapted from existing airports.
- EXE-05.97.2-TRL4-TVALP-ASR-007 - A real-time simulation addressing Speech Recognition at Rome simulating Sofia airport.

Depending upon the different exercises, validation scenarios addressed different airport categories, from small to very large ones. 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, Cost Efficiency, Capacity (airport resilience) 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-05.971-TRL4-TVALP-VAR-001
- EXE-05.971-TRL4-TVALP-VAR-002
- EXE-05.97.2-TRL4-TVALP-ASR-004
- EXE-05.971-TRL4-TVALP-VAR-005
- EXE-05.97.2-TRL4-TVALP-ASR-006
- EXE-05.97.2-TRL4-TVALP-ASR-007

along with the results acquired from the validations.

Co-funded by the European Union

2 Introduction

2.1 Purpose of the document

This Technical Validation Report gives the results of the TRL4 maturity Technological Validation activities for Solutions 97.1 and 97.2. It describes how stakeholders' needs and system requirements were intended to be validated[1].

2.2 Intended readership

The intended audience of this document are people who are interested in how the partners involved in SESAR Solutions PJ.05-W2-97.1 and PJ.05-W2-97.2 validated improvements to the management of airport operations included in such solutions and how those improvements enhanced the overall efficiency of Air Traffic Management systems. Readers include other members of SESAR2020 PJ.05, members of PJ.19 and representatives of the SESAR JU. 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.05-SOL97.1 and PJ.05-SOL97.2 built on the work performed by S2020 SOL16-04 Wave 1 project, as well as RETINA and MALORCA projects, performed in the context of Exploratory Research. Starting maturity level is TRL2 and at the end of Wave 2 activities it is bound to reach TRL4 maturity.

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.05-W2-97
- Chapter 4: SESAR Technological Solution PJ.05-W2-97 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 TRL4, along with recommendations for future TRLs
- Chapter 6: References and Applicable Documents. A list of all references and applicable documents during TVALR editing, connected with novel human interaction modes

[1] The opinions expressed herein reflect the author's view only. Under no circumstances shall the SESAR3 Joint Undertaking be responsible for any use that may be made of the information contained herein.

- Appendixes A-F: 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
Air Gesture	Gesture recognition is a type of perceptual computing user interface allowing computers to capture and interpret human gestures as commands via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand. Users can use simple gestures to control or interact with devices without physically touching them.	SOL 97.1
Attention Guidance	The Attention Guidance system guides the attention of air traffic controllers via perceptual cues towards an imminent ATC situation, either determined by attention guidance logic or an external safety net system. Prioritization of events criticality (e.g. RMCA, CMAC, CTAC alert) will select how the ATCo's attention shall be raised.	SOL 97.1
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.	PJ.16-04

Co-funded by the European Union

	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 HEADING 260 LEFT”. The resulting concepts can be used for further applications such as visualization on an HMI.	
Conventional Input devices	Expression used to identify the current, legacy devices as keyboard, mouse and trackball. It is used as the reference system.	PJ.16-04
Tracking labels (in AR environment)	A label attached to a real a/c object, displaying the most important information; the tracking label displays additional information in the case of detection of any potential conflict by the Airport Safety Net Service.	SOL 97.1
Virtual/ Augmented Reality	V/AR in ATC Tower environment supports Air Traffic Controllers by blending in real-time real world images with computer-generated data (augmented reality), so that visual information can be enhanced to improve identification and tracking of a/c (or vehicles) on the airport surface. Moreover, in low visibility conditions, the lack of visual information provided by the out-of-the-tower windows view can be made up for by the massive use of synthetic vision to show digital georeferenced data supplementing the missing real vision (virtual reality). Airport operations can benefit from such advanced technologies, capable to provide beneficial automation support under low visibility conditions. Benefits are available in good visibility conditions as well, providing the controllers with additional information content in the labels to help if physical obstacles obstruct vision or to reduce head-down time.	SOL 97.1

Table 1: Glossary of terms

2.6 Acronyms and terminology

Term	Definition
ABSR	Assistant Based Speech Recognition
ADD	Architecture Definition Document
AI	Artificial Intelligence
AR	Augmented Reality
ASR	Automatic Speech Recognition
ATC	Air Traffic Control
ATCo	Air Traffic Controller
ATM	Air Traffic Management
A/C	Aircraft
COTS	Commercial Off the Shelf
CWP	Controller Working Position
E-ATMS	European Air Traffic Management System
E-OCVM	European Operational Concept Validation Methodology
HMI	Human Machine Interface
HPAP	Human Performance Assessment Plan
IRS	Interface Requirements Specification
ISA	Instantaneous self-assessment of workload technique
ML	Machine Learning
NARSIM	NLR ATC Research Simulator
OSED	Operational Service and Environment Definition
PI	Performance Indicator
SecAP	Security Assessment Plan
SESAR	Single European Sky ATM Research Programme
S3JU	SESAR3 Joint Undertaking (Agency of the European Commission)

EUROPEAN PARTNERSHIP

Co-funded by the European Union

SPR-INTEROP/OSED	Safety and Performance Requirements – Interoperability / Operational Service and Environment Definition
SUT	System Under Test
TS	Technical Specification
TSAP	Technical Safety Assessment Plan
TVALP	Technological Validation Plan
TVALR	Technological Validation Report
TWR	Tower
VALS	Validation Strategy
VP	Validation Plan
VR	Validation Report
VR	Virtual Reality
VS	Validation Strategy

Table 2: Acronyms and terminology

3 Context of the Technological Validation

3.1 SESAR Technological Solutions 97.1 and 97.2: a summary

Solutions 97.1 and 97.2 manage operational and technical objectives of the Controller Working Position, CWP. The solutions take advantage of the work already performed during Wave 1 continuing to provide significant improvements thanks to advanced interaction methods with the airport control tower human machine interface (HMI).

Solutions 97.1 and 97.2 address the development of new HMI interaction modes and technologies in order to minimise the load and mental strain on Tower controllers (especially under high traffic density situations, low visibility conditions, etc.). Such improvements may be applicable to current operations and/or in future operational concepts still under development within the scope of other SESAR solutions.

Solutions 97.1 and 97.2 refer to HMI Interaction modes for Airport Tower addressing the development of new human machine interface (HMI) interaction modes and technologies. They are applicable to the current operating airport environment as well as to future environments, being aligned to the current evolution trend taking place within the SESAR 2020 Programme. The core of activities will be oriented towards three main areas detailed in the following subsections, each according to its solution.

3.1.1 SESAR Technological Solution 97.1

SOL 97.1 Virtual and augmented reality in various applications (e.g. smart screens, head-on display) to enable tower ATCOs safe operations supervision under any meteorological conditions while maintaining a high taxiway and runway throughput. Within this specific area other technologies such as **Tracking labels** and **air gestures** and **attention guidance** were investigated.

Table 3 below gives a quick breakdown of the Technical Enablers making up PJ.05-W2-97.1.

SESAR Technological Solution ID	SESAR Technological Solution Description	Master or Contributing (M or C)	Contribution to the SESAR Solution short description	Enabler ref. (from EATMA)
PJ.05-W2-97.1	Virtual/Augmented Reality applications for tower	M	Solution 97.1 contributes completely	AERODROME-ATC-103
	Virtual/Augmented Reality applications for tower	M	Solution 97.1 contributes completely	AERODROME-ATC-104

	Virtual/Augmented Reality applications for tower	M	Solution 97.1 contributes completely	AERODROME-ATC-105

Table 3: SESAR Technological Solution(s) under Validation PJ.05-W2-97.1

3.1.2 SESAR Technological Solution 97.2

SOL 97.2 Automatic speech recognition (ASR) supported with AI/ML techniques, enables recognition and translation of spoken language into the system reducing, ATCO workload and hence improving safety. Table 4 below shows a quick itemization of Technical Enablers making up PJ.05-W2-97.2. On top of the enabler, the present solution uses also the enabler **AERODROME-ATC-50** (Advanced Airport Tower Controller Working Position (A-CWP)) as a driver to perform validation.

SESAR Technological Solution ID	SESAR Technological Solution Description	Master or Contributing (M or C)	Contribution to the SESAR Solution short description	Enabler ref. (from EATMA)
PJ.05-W2-97.2	ASR at the TWR CWP supported by AI and Machine Learning	M	Solution 97.2 contributes completely	AERODROME-ATC-106

Table 4: SESAR Technological Solution(s) under Validation PJ.05-W2-97.2

3.2 Summary of the Technological Validation Plan

3.2.1 Validation plan purpose

The TVALP [28] document describes validation activities performed to take the Technical Enablers that make up Solutions PJ.05-W2-97.1 and 97.2 to the TRL4 complete phase of development. The objective of the TVALP – TRL4 is to set the framework for all the TRL4 activities performed by the solution members in order to validate different aspects and applications of V/AR and 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 intermediate version 1 included a detailed description of EXE-05.97.1-TRL4-TVALP-VAR-001. Subsequent releases of the TVALP included detailed descriptions of exercises. TVALP schedule was as follows:

- D3.1.032 Interim TVALP delivery 2, including detailed descriptions of exercises EXE-05.97.1-TRL4-TVALP-VAR-002 and EXE-05.97.2-TRL4-TVALP-ASR-004.
- D3.1.033 Final TVALP, including also detailed descriptions of exercises EXE-05.97.3-TRL4-TVALP-MTI-004, EXE-05.97.1-TRL4-TVALP-VAR-005, EXE-05.97.2-TRL4-TVALP-ASR-006 and EXE-05.97.2-TRL4-TVALP-ASR-007.

3.2.2 Summary of Technological Validation objectives and success criteria

Validation objectives were adapted taking into account that solutions under validation are technical in nature, as well as BIMs and feedback from safety and human performance assessment recommendations. Maturity gate criteria for TRL4 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 [28], in its sections 4.2.1 and 4.2.2 .

Validation Identifier	Name	Primary Text	Success Criterion 1	Success Criterion 2	Success Criterion 3	Success Criterion 4	Success Criterion 5
OBJ-05.971-TRL4-TVALP-FEAS.1010	V/A-R TRL4 Operational feasibility	To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS.	No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Virtual or Augmented Reality and tracking labels.	No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Air Gestures.	No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Attention Guidance.		
OBJ-05.971-TRL4-TVALP-FEAS.1020	V/A-R TRL 4 technical feasibility	To identify possible technical feasibility issues and possible show stoppers	Laboratory tests (based on a prototype) have verified the technical feasibility of the use of V/AR applications in the tower environment.	Laboratory tests have verified that the integration of the V/AR applications with other related system enablers is technically feasible.			

Validation Identifier	Name	Primary Text	Success Criterion 1	Success Criterion 2	Success Criterion 3	Success Criterion 4	Success Criterion 5
OBJ-05.971-TRL4-TVALP-H103.1010 OBJ-05.971-TRL4-TVALP-H103.1010	V/A-R Tracking labels and overlays impact on ATCO tasks V/A-R Tracking labels and overlays impact on ATCO tasks	To asses that the technical systems for V/A-R Tracking labels and overlays support the ATCOs in performing their tasks To asses that the technical systems for V/A-R Tracking labels and overlays support the ATCOs in performing their tasks	Majority of ATCOs' responses (at least 75%) is that workload is mantained at acceptable level when using V/A-R Technology	Majority of ATCOs' responses (at least 75%) is that the level and quality of information is adequate, complete and acceptable when using V/A-R Technology	Majority of ATCOs' responses (at least 75%) is that the V/A-R HMI supports ATCO in maintaining an adequate level of situation awareness	Measured time spent in head-up is increased in the solution scenario with respect to the reference scenario	HMI of V/A-R tools does not overshadow the relevant information on the OTW view
			Success Criterion 6	**Success Criterion 7**	**Success Criterion 8**	**Success Criterion 9**	**Success Criterion 10**
			V/A-R HMI does not increase the potential for human error	Majority of ATCOs' responses (at least 75%) is that the trust in the system is at an acceptable level	Majority of ATCOs' responses (at least 75%) is that the level of usability is adequate when using V/A-R HMI	Majority of ATCOs' responses (at least 75%) is that the alarms and alerts are not too intrusive and support ATCOs in the early detection of ATC critical situations	Majority of ATCOs' responses (at least 75%) is that V/A-R acceptance is adequate
OBJ-05.971-TRL4-TVALP-H103.1030	V/A-R Tracking labels and overlays on ATCO role	To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/AR Tracking labels and overlays	Majority of ATCOs' responses (at least 75%) is that they can apply operating methods in an accurate, efficient and timely manner when using V/A-R	Majority of ATCOs' responses (at least 75%) is that operating methods are clearly identified and consistent in all operating conditions when using V/A-R			

Validation Identifier	Name	Primary Text	Success Criterion 1	Success Criterion 2	Success Criterion 3	Success Criterion 4	Success Criterion 5
OBJ-05.971-TRL4-TVALP-H103.1040	V/A-R Tracking labels and overlays impact on job satisfaction	To assess job acceptance and satisfaction with the introduction of V/AR Tracking labels and overlays	Majority of ATCOs' responses (at least 75%) is that job satisfaction and acceptance is adequate when using V/A-R				
OBJ-05.971-TRL4-TVALP-H104.1010	V/A-R Air Gestures impact on ATCO tasks	To assess that the technical systems for V/AR Air Gestures support the ATCOs in performing their tasks	Majority of ATCOs' responses (at least 75%) is that workload is mantained at acceptable level when using V/A-R Air Gestures Technology	Majority of ATCOs' responses (at least 75%) is that the level and quality of information is adequate, complete and acceptable when using V/A-R Air Gestures Technology	Majority of ATCOs' responses (at least 75%) is that the V/A-R Air Gestures HMI supports ATCO in maintaining an adequate level of situation awareness	Measured time spent in head-up is increased in the solution scenario with respect to the reference scenario	V/A-R Air Gestures HMI does not increase the potential for human error
			Success Criterion 6	**Success Criterion 7**	**Success Criterion 8**		
			Majority of ATCOs' responses (at least 75%) is that the trust in the system is at an acceptable level	Majority of ATCOs' responses (at least 75%) is that the level of usability is adequate when using V/A-R Air Gestures HMI	Majority of ATCOs' responses (at least 75%) is that V/A-R Air Gestures acceptance is adequate		

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Validation Identifier	Name	Primary Text	Success Criterion 1	Success Criterion 2	Success Criterion 3	Success Criterion 4	Success Criterion 5
OBJ-05.971-TRL4-TVALP-H104.1020	V/A-R Air Gestures on ATCO role	To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/AR Air gestures	Majority of ATCOs' responses (at least 75%) is that they can apply operating methods in an accurate, efficient and timely manner when using V/A-R Air Gesture	Majority of ATCOs' responses (at least 75%) is that operating methods are clearly identified and consistent in all operating conditions when using V/A-R Air Gesture			
OBJ-05.971-TRL4-TVALP-H104.1030	V/A-R Air Gestures impact on job satisfaction	To assess job acceptance and satisfaction with the introduction of V/A-R Air Gestures	Majority of ATCOs' responses (at least 75%) is that job satisfaction and acceptance is adequate when using V/A-R Air Gesture				

Validation Identifier	Name	Primary Text	Success Criterion 1	Success Criterion 2	Success Criterion 3	Success Criterion 4	Success Criterion 5
OBJ-05.971-TRL4-TVALP-H105.1010	V/A-R Attention Guidance impact on ATCO tasks	To asses that the technical systems for V/A-R Attention Guidance support the ATCOs in performing their tasks	Majority of ATCOs' responses (at least 75%) is that workload is mantained at acceptable level when using V/A-R Attention Guidance Technology	Majority of ATCOs' responses (at least 75%) is that the level and quality of information is adequate, complete and acceptable when using V/A-R Attention Guidance Technology	Majority of ATCOs' responses (at least 75%) is that the V/A-R Attention Guidance HMI supports ATCO in maintaining an adequate level of situation awareness	Measured time spent in head-up is increased in the solution scenario with respect to the reference scenario	HMI of V/A-R Attention Guidance tools does not overshadow the relevant information on the OTW view
			Success Criterion 6	**Success Criterion 7**	**Success Criterion 8**	**Success Criterion 9**	**Success Criterion 10**
			V/A-R Attention Guidance HMI does not increase the potential for human error	Majority of ATCOs' responses (at least 75%) is that the trust in the system is at an acceptable level	Majority of ATCOs' responses (at least 75%) is that the level of usability is adequate when using Attention Guidance HMI	Majority of ATCOs' responses (at least 75%) is that the alarms and alerts are not too intrusive and support ATCOs in the early detection of ATC critical situations	Majority of ATCOs' responses (at least 75%) is that V/A-R Attention Guidance acceptance is adequate
OBJ-05.971-TRL4-TVALP-H105.1030	V/A-R Attention Guidance on ATCO role	To asses that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Attention Guidance	Majority of ATCOs' responses (at least 75%) is that they can apply operating methods in an accurate, efficient and timely manner when using V/A-R Attention Guidance	Majority of ATCOs' responses (at least 75%) is that operating methods are clearly identified and consistent in all operating conditions when using V/A-R Attention Guidance			

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Validation Identifier	Name	Primary Text	Success Criterion 1	Success Criterion 2	Success Criterion 3	Success Criterion 4	Success Criterion 5
OBJ-05.971-TRL4-TVALP-H105.1040	V/A-R Attention Guidance impact on job satisfaction	To assess job acceptance and satisfaction with the introduction of V/A-R Attention Guidance	ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance				
OBJ-05.971-TLR4-TVALP-SAFE.1010	Safety Impact	To assess the impact of Virtual/Augmented Reality applications on safety	The changes related to the implementation of Virtual/Augmented Reality applications do not increase potential for human error and therefore not reducing safety levels.	ATCO's workload with the implementation of Virtual/Augmented Reality applications is maintained at acceptable level and therefore not reducing safety levels.	ATCO's situational awareness with the implementation of Virtual/Augmented Reality applications is maintained at acceptable level and therefore not reducing safety levels.	Safety assessment activities and the results are documented and integrated in the overall solution validation results	
OBJ-05.971-TLR4-TVALP- PERF.1010	TRL4 Performance Assessment	To assess the performance benefits of equivalent visual operations for tower control through the use of applications for Virtual/Augmented Reality**.**	Laboratory tests show that the use of V/A applications improves Cost Efficiency performance by reducing the cost per flight (through e.g. reduction of workload, reduction of delay times, …)	Laboratory tests show that the use of V/A reality applications improves Resilience by increasing situational awareness in Low visibility conditions while maintaining workload within acceptable limits			

Table 5: List of 97.1 Validation objectives and Success criteria

Validation Identifier	Name	Primary Text	Success Criterion 1	Success Criterion 2	Success Criterion 3	Success Criterion 4
OBJ-05.972-TLR4-TVALP-FEAS.2010	TLR4 Operational feasibility	To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS.	No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Automatic speech recognition.	No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of AI suggestions.		
OBJ-05.972-TLR4-TVALP- FEAS.2020	TLR4 Pre-industrial feasibility	To identify possible technical feasibility issues and possible show stoppers	Laboratory tests (based on a prototype) have verified the technical feasibility of the use of ASR supported by AI/ML	Laboratory tests have verified that the integration of the SESAR technological solution with other related system enablers is technically feasible		
OBJ-05.972-TRL4-TVALP-H106.2010	ASR impact on ATCO tasks	To assess that the technical systems for ASR support the ATCOs in performing their tasks	Majority of ATCOs' responses (at least 75%) is that workload is mantained at acceptable level when using ASR Technology	Majority of ATCOs' responses (at least 75%) is that the ASR supports ATCO in maintaining an adequate level of situation awareness	ASR does not increase the potential for human error	Majority of ATCOs' responses (at least 75%) is that the level and quality of information is adequate, complete and acceptable when using ASR Technology

Validation Identifier	Name	Primary Text	Success Criterion 1	Success Criterion 2	Success Criterion 3	Success Criterion 4
			Measured callsign recognition rate, command recognition rate, error rate and rejection rate of ASR system are considered within acceptable levels by the majority of ATCOS (at least 75%)	Majority of ATCOs' responses (at least 75%) is that the level of usability is adequate when using ASR system	Majority of ATCOs' responses (at least 75%) is that ASR acceptance is adequate	Majority of ATCOs' responses (at least 75%) is that the trust in the system is at an acceptable level
OBJ-05.972-TRL4-TVALP-H106.2020	ASR impact on ATCO role	To assess the role of the ATCO is consistent with human capabilities and limitations with the introduction of ASR	Majority of ATCOs' responses (at least 75%) is that they can apply operating methods in an accurate, efficient and timely manner when using ASR	Majority of ATCOs' responses (at least 75%) is that operating methods are clearly identified and consistent in all operating conditions when using ASR		
OBJ-05.972-TRL4-TVALP-H106.2030	ASR impact on job satisfaction	To assess job acceptance and satisfaction with the introduction of ASR	Majority of ATCOs' responses (at least 75%) is that job satisfaction and acceptance is adequate when using ASR			

Validation Identifier	Name	Primary Text	Success Criterion 1	Success Criterion 2	Success Criterion 3	Success Criterion 4
OBJ-05.972-TLR4-TVALP-SAFE.2010	Safety Impact	To assess the impact of Automatic Speech Recognition applications on safety.	The changes related to the implementation of Automatic Speech Recognition do not increase potential for human error and therefore not reducing safety levels.	ATCO's workload with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	ATCO's situational awareness with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	Safety assessment activities and the results are documented and integrated in the overall solution validation results
OBJ-05.972-TLR4-TVALP-PERF.2010	TLR4 Performance Assessment	To assess the performance benefits of Automatic Speech Recognition supported by AI/ML.	Laboratory tests show that the SESAR technological solution improves Cost Efficiency performance by reducing cost per flight(PER.TRL4.3) (through e.g. reduction of workload, reduction of delay times,….)			

Table 6: List of 97.2 Validation objectives and Success criteria

3.2.3 Technological Validation assumptions

Assumptions applicable to solutions 97.1 and 97.2 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	Mixed Equipage	Mixed A/C equipage; mixed ground vehicles equipage.	No use of radar data (not related to a/c equipage) (EXE001 and 002) EXE-005 a/c and vehicles ADS-B out	N/A

Identifier	Title	Description	Justification	Impact on Assessment
			equipped. No equipage mix. EXE004 006 007 no Datalink use	
AS-GEN-04	User Diversity	Consideration of diversity of users: mainline, regional, business, rotorcraft, GA, RPAS and/or drones.	To be adapted by every Validation Plan.	N/A
AS-GEN-05	Flight Plan	Very high proportion (> 95 %) of commercial and military flights with Extended Flight Plan / RBT/RMT associated to.	N/A	N/A
AS-GEN-07	Training	Airborne, ATC and vehicles staffs have appropriate training and competencies.	Similar considerations as AS-GEN-01 regarding “exceptions”.	N/A
AS-GEN-10	conformance	General conformance and compatibility between airborne and ground data bases.	Similar considerations as AS-GEN-01 regarding “exceptions”.	N/A
AS-GEN-14	baseline	It is assumed that relevant concepts from the SESAR Solution Catalogue are already validated and implemented (i.e. SESAR 1 is baseline).	N/A	N/A
AS-GEN-17	traffic	During the validation activities, it is assumed that the simulated traffic in the validation scenarios encompasses those corresponding to the FOC of the OI step to be validated.	N/A	N/A
AS- GEN-18	Information sharing	Widely shared information among all necessary actors about key turnaround milestones, during planning and execution.	N/A	N/A

Identifier	Title	Description	Justification	Impact on Assessment
AS- GEN-19	Data and tools	Up-to-date and comprehensive capacity data and information from ANSPs and airports are available, as well as the appropriate tools to process them and assure coordination.	N/A	N/A

Table 7: Technological Validation assumptions overview

3.2.4 Technological Validation Exercises list

Shown below is a short list of the validation exercises planned to achieve TRL4 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 SESAR Solutions 97.x will progress from the TRL2 maturity level to TRL4. Validation exercises planned for each solution and their main features addressed are shown in Table 8~~Table 8~~.

			Exercises	*001*	*002*	*004*	*005*	*006*	*007*
			Leader	*NLR*	*ENAV*	*INDRA*	*ENAIRE*	*DLR*	*LDO*
	OI step	*Tech nology*	*Simulation execution*	*Feb-21*	*Mar-22*	*Dec-21*	*Feb-22*	*Mar-22*	*May-22*
SOL 97.1	***POI-0039-SDM Virtual/Augmented Reality, attention guidance and air gesture for tower controllers.***	*VAR + TKL*	*Virtual and Augmented Reality and tracking labels*	X	X		X		
		AG	*Air Gestures*		X		X		
		VAR+AtGu	*Attention Guidance*	X					
SOL 97.2	***POI-0040-SDM Improving controller productivity by ASR at the TWR CWP***	*ASR + AI/ML*	*Automatic Speech Recognition + AI / Machine Learning algorithms*			X		X	X

		A-CWP	*Advanced Airport CWP (A-CWP)* *[required use]*			X		X	X

Table 8: Technological Solutions validation exercises list

In order to achieve TRL4-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, Capacity (airport resilience focus area) 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 S97.1 technological validation exercises

Exercises in this list investigate POI-0039-SDM Equivalent visual operations for tower control through the use of applications for Virtual/Augmented Reality. Each table provides information regarding each exercise.

[EXE]

Identifier	EXE-05.97.1-TRL4-TVALP-VAR-001
Title	Validation of AR Interaction Modes for Schiphol Tower with a Focus on Attention Guidance
Description	This validation activity investigated AR applications for a conventional tower environment of Schiphol Airport. Alerts currently given in that environment, which are Runway Incursion Alerting and Go-around Alerting, are generated and the differences in attention getting in comparison with traditional tower control as well as the advantages of attention guidance were evaluated.
Achievements	The optimal way to guide attention was assessed by exposing controllers to different presentations of alert information, different symbology and/or audio alerts within the AR device. The AR device was used in different scenarios with different traffic situations, different types of alerting with different levels of severity at selected locations in the airport movement areas. Controller reaction times, attention distribution and decision-making effectiveness for the situation to be solved were measured and compared. Reaction times were deduced from controller input into the system, attention distribution and decision-making effectiveness was evaluated by an experienced operational expert involved in the experiment design. Generally, it should be observed that Situational Awareness increased, workload was reduced and operational effectiveness was improved.
Addressed OIs	POI-0039-SDM

TRL	<TRL4>
T. Validation Technique	<RTS>
Start Date	12 February 2021
End Date	14 April 2021
T. Validation Coordinator	Royal NLR
T. Validation Platform	NARSIM Tower (at Royal NLR, Amsterdam)
T. Validation Location	Amsterdam Airport Schiphol
Status	<complete>
Dependencies	N/A

[EXE Trace]

Linked Element Type	Identifier

[EXE]

Identifier	EXE-05.97.1-TRL4-TVALP-VAR-002
Title	Augmented Reality Multimodal Control Tower Interaction
Description	The exercise objective is to mature the results obtained in the previous RETINA validation campaign including additional features not considered at exploratory research level, such as: Adaptive HMI, working positions, multimodal interaction and safety nets visualization.
Expected achievements	• Increase in situational awareness of the controller; • Reduction of controller workload; • Increased ATCO efficiency; • Improved HMI and usability and performance of interactions; • Increased or maintained safety.
Addressed OIs	POI-0039-SDM
TRL	TRL4

T. Validation Technique	<RTS>
Start Date	18 January 2022
End Date	26 January 2022
T. Validation Coordinator	ENAV
T. Validation Platform	CAVE V-LAB
T. Validation Location	Bologna Airport
Status	<complete>
Dependencies	N/A

[EXE Trace]

Linked Element Type	Identifier

[EXE]

Identifier	EXE-05.97.1-TRL4-TVALP-VAR-005
Title	V2 Augmented Reality in the Tower Environment
Description	The objectives of this exercise are to validate that the ATCO Head up Interface so as to: • Increase the out-the-window viewing time. • Reduce the changes of controller gaze from out-the-window to computer screen. • Increase situational awareness in low visibility conditions. • Accurately track a/c motion.
Expected achievements	Main achievements expected are: • Increased Situational Awareness • Increased ATCO efficiency; • Improved HMI and usability;
Addressed OIs	POI-0039-SDM;
TRL	TRL4
T. Validation Technique	<shadow mode>
Start Date	7 February 22

End Date	18 February 22
T. Validation Coordinator	ENAIRE
T. Validation Platform	ITWP
T. Validation Location	Vitoria Airport
Status	<complete>
Dependencies	N/A

[EXE Trace]

Linked Element Type	Identifier

3.2.4.2 S97.2 technological validation exercises

Exercises listed in this section investigate POI-0040-SDM: improving controller productivity by ASR at the TWR CWP.

[EXE]

Identifier	EXE-05.97.2-TRL4-TVALP-ASR-004
Title	Improved controller productivity by using speech recognition in a multiple remote tower environment
Description	This exercise investigates the benefits of ASR technology, using HungaroControl's and Indra's system. Functions implemented using the new technology are adapted into a system simulator. This part of the validation exercise with its main focus on ASR is expected to take place during December 2021 at Asker.
Expected achievements	• Reduction of Controller Workload, • Increase of controllers' productivity and situational awareness by introducing new way of controller input
Addressed OIs	POI-0040-SDM
TRL	TRL4
T. Validation Technique	<RTS>
Start Date	22 November 2021

End Date	10 December 2021
T. Validation Coordinator	INDRA Navia
T. Validation Platform	InNOVA Remote, InNOVA ITWP
T. Validation Location	Asker
Status	<complete>
Dependencies	ASR – PJ05.35

[EXE Trace]

Linked Element Type	Identifier

[EXE]

Identifier	EXE-05.97.2-TRL4-TVALP-ASR-006
Title	Assistant Based Speech Recognition in a Multiple Remote Tower Environment
Description	EXE-006 investigates the benefits of an Assistant Based Speech Recognition (ABSR) system for a simulated Multiple Remote Tower environment mainly with respect to a reduction of controller workload. Radar data, flight plan data, and meteorological data are used to predict controller commands using machine learning algorithms. Those forecast command hypotheses support an automatic speech recognition engine to automatically recognize word sequences. Word sequences thus obtained are then in turn automatically analysed to extract meaningful ATC concepts such as commands with callsigns, type, values, units, etc. Resulting content is then used to enable enhanced support functionalities for the controller such as a/c flight strip maintenance with given clearance contents and callsign highlighting.
Expected achievements	• Reduction of Controller Workload, • Improved use of HMI enabling enhanced functions with extracted Controller Commands
Addressed OIs	POI-0040-SDM;
TRL	TRL4
T. Validation Technique	<RTS>
Start Date	14 February 22

End Date	04 March 22
T. Validation Coordinator	DLR
T. Validation Platform	DLR Remote Tower with integrated ABSR
T. Validation Location	Braunschweig
Status	<complete>
Dependencies	N/A

[EXE Trace]

Linked Element Type	Identifier

[EXE]

Identifier	EXE-05.97.2-TRL4-TVALP-ASR-007
Title	Assistant Based Speech Recognition as support to ATCOs
Description	The exercise performs integration of a speech recognition system in a next-gen Surface CWP in order to achieve the following operational goals: Facilitate the ATCOs work by prefilling an appropriate system mask (which contains clearances updated according to ATCO directives) using the content of verbal communication. The exercise simulates activity at Sofia airport.
Expected achievements	• Increase of ATCO productivity, with consequent reduction of cost per flight; • Positive impact on ATCO workload and in terms of decrease of human error.
Addressed OIs	POI-0040-SDM;
TRL	TRL4
T. Validation Technique	<RTS>
Start Date	02 May 2022
End Date	05 May 2022
T. Validation Coordinator	LDO
T. Validation Platform	LDO platform

Co-funded by the European Union

T. Validation Location	Rome
Status	<complete>
Dependencies	N/A

[EXE Trace]

Linked Element Type	Identifier

3.3 Deviations

3.3.1 Deviations with respect to the S3JU Project Handbook

No deviations.

3.3.2 Deviations with respect to the Technological Validation Plan

No deviations.

4 SESAR Technological Solution 97.1 and 97.2 Validation Results

4.1 Summary of SESAR Technological Solution 97 Validation Results

In the tables below a very synthetic view of the validation exercises' results is shown. Please refer to Appendixes A-F for detailed information on each exercise and on their detailed results.

In general V/A-R has shown to be a very promising technology whose adoption in Air Traffic Management Tower and Ground environments seems to be on the horizon. However, there are still limitations commercial devices are affected by, and this means provisions and improvements will have to be made in order to proceed to higher TRL levels and finally to operation in Control Rooms and Control Centres. In spite of several POK results reached, V/A-R has shown to be a thoroughly dependable technological enabler, whose importance and results are very promising reaching TRL4.

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.

4.1.1 Synopsis of SESAR Technological Solution 97.1 Validation Results

Topic	EXE001	EXE002	EXE005	Solution Result	Objective Status
TRL4 feasibility	OK	OK	OK	No show stoppers found in the validation exercises	OK
V/A-R Tracking labels and overlays impact on ATCO tasks	N/A	POK	POK	Reservations regarding information quality, extent and resulting situation awareness	POK
V/A-R Tracking labels and overlays impact on ATCO role	N/A	POK	POK	Improvable results in timeliness, accuracy and efficiency	POK
V/A-R Tracking labels and overlays impact on job satisfaction	N/A	OK	OK	Generally well received though better methods will be established	OK
V/A-R Air Gestures impact on ATCO tasks	N/A	POK	POK	Methods to evolve	POK

Topic	EXE001	EXE002	EXE005	Solution Result	Objective Status
V/A-R Air Gestures impact on ATCO role	N/A	POK	POK	Improvable results in timeliness, accuracy and efficiency	OK
V/A-R Air Gestures impact on job satisfaction	N/A	OK	OK	Favourable impact on job satisfaction	OK
V/A-R Attention Guidance impact on ATCO tasks	POK	N/A	N/A	Positive remarks. Symbology could be improved	POK
V/A-R Attention Guidance impact on ATCO role	POK	N/A	N/A	Usability and trust not yet acceptable. HMI improvable.	POK
V/A-R Attention Guidance impact on job satisfaction	OK	N/A	N/A	Response time, clarity, all at least neutrally affected	OK
V/A-R impact on Safety	OK	OK	OK	Reduction or stability of human error, workload and situational awareness	OK
V/A-R impact on Performance	OK	OK	OK	Even with simulations of a limited extent impact was deemed positive	OK
TRL4 Performance Assessment	POK	POK	POK	Results achieved with limitations	V/A-R OK as a technological enabler

Table 9: Synopsis of Solution 97.1 exercises results

4.1.2 Synopsis of SESAR Technological Solution 97.2 Validation Results

Topic	EXE004	EXE006	EXE007	Solution Result	Objective Status
TRL4 feasibility	OK	OK	OK	No show stoppers found in the validation exercises	OK
ASR impact on ATCO tasks	Not Addressed POK	OK	OK	Evaluation of the results showed no detrimental effects	OK

ASR impact on ATCO role	Not Applicable	OK	OK	Evaluation of the results showed no detrimental effects	OK
ASR impact on job satisfaction	Not Addressed	OK	OK	All metrics show beneficial effects	OK
ASR impact on Safety	Not Addressed /POK	OK	POK	In EXE007 no detrimental effects were measured	POK
ASR impact on Performance	Not Addressed	OK	POK	Positive measured impact in several metrics	POK
TLR4 Performance Assessment	Not Applicable	OK	OK	Overall results show TRL4 is achievable	OK (006-007)

Table 10: Synopsis of Solution 97.2 exercises results

4.1.3 SESAR Technological Solution 97.1 Validation Results

As well as to the following table, please refer to Appendixes A, B, D for detailed Technological Validation Exercises 001, 002 and 005 results and thorough explanation.

<table>
<tr><th>SESAR Technological Validation Objective ID/Description/Title</th><th>SESAR Technological Validation Success Criteria ID/Description</th><th>SESAR Technological Validation Results</th><th>SESAR Technological Validation Objective Status</th></tr>
<tr><td rowspan="3">OBJ-05.971-TRL4-VALP-FEAS.1010</td><td>CRT-05.971-TLR4-TVALP- FEAS -1011
No operational show-stoppers identified during laboratory tests (on a prototype) related to the use of Virtual or Augmented Reality and tracking labels.</td><td>001,002, 005 OK</td><td rowspan="3">OK</td></tr>
<tr><td>CRT-05.971-TLR4-TVALP- FEAS -1012
No operational show-stoppers identified during laboratory tests (on a prototype) related to the use of Air Gestures</td><td>002,005 OK</td></tr>
<tr><td>CRT-05.971-TLR4-TVALP- FEAS -1013
No operational show-stoppers identified during laboratory tests (on a prototype) related to the use of Attention Guidance.</td><td>001 OK</td></tr>
</table>

EUROPEAN PARTNERSHIP

Co-funded by the European Union

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
OBJ-05.971-TRL4-TVALP-FEAS.1020 To identify possible technical feasibility issues and possible show stoppers. TRL4 Pre-industrial feasibility	CRT-05.971-TLR4-TVALP- FEAS -1021 Laboratory tests (based on a prototype) have verified the technical feasibility of the use of V/AR applications in the tower environment	001, 002: OK	OK
	CRT-05.971-TLR4-TVALP- FEAS -1022 Laboratory tests have verified that the integration of the V/AR applications with other related system enablers is technically feasible.	001,005: OK 002: N/A	
OBJ-05.971-TRL4-TVALP-H103.1010 To assess that the technical systems for V/A-R Tracking labels and overlays support the ATCOs in performing their tasks. V/A-R Tracking labels and overlays impact on ATCO tasks	CRT-05.971-TLR4-TVALP- H103-1011 Majority of ATCOs (at least 75%) responses is that VA-R supports ATCO in maintaining workload at acceptable level	002:OK 005:POK	POK
	CRT-05.971-TLR4-TVALP- H103-1012 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R	002,005: OK	
	CRT-05.971-TLR4-TVALP- H103-1013 Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO in maintaining an adequate level of situation awareness	002:OK, 005:NOK	
	CRT-05.971-TLR4-TVALP- H103-1014 Measured time spent in head up is increased in the solution scenario with respect to the reference scenario	002,005 :OK	
	CRT-05.971-TLR4-TVALP- H103-1015 HMI of V/A-R tools does not overshadow the relevant information on the OTW view	002:POK, 005: OK	
	CRT-05.971-TLR4-TVALP- H103-1016 V/A-R HMI does not increase the potential for human error	002: POK, 005: OK	

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
	CRT-05.971-TLR4-TVALP- H1033-1017 ATCOs (at least 75%) trust in the system is at an acceptable level	002,005: OK	
	CRT-05.971-TLR4-TVALP- H103-1018 Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R HMI	002: POK, 005: OK	
	CRT-05.971-TLR4-TVALP- H103-1019 Majority of ATCOs (at least 75%) provide feedback that alarm and alerts are not too intrusive and supports ATCOs in the early detection of ATC critical situations	002: OK	
	CRT-05.971-TLR4-TVALP- H103-1020 ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R tool	002,005: OK	
	CRT-05.971-TLR4-TVALP- H103-1021 Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO team (GND and TWR) in maintaining an acceptable level of situation awareness	002: OK	
OBJ-05.971-TRL4-TVALP-H103.1030 To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Tracking labels and overlays V/A-R Tracking labels and overlays on ATCO role	CRT-05.971-TLR4-TVALP-H103-1031 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner	002,005: OK	POK
	CRT-05.971-TLR4-TVALP-H103-1032 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions	002:NOK, 005:OK	
OBJ-05.971-TRL4-TVALP-H103.1040 To assess job acceptance	CRT-05.971-TLR4-TVALP-H103-1041 ATCOs' (at least 75%) provide positive feedback on job satisfaction and acceptance	002, 005: OK	OK

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
and satisfaction with the introduction of V/A-R Tracking labels and overlays V/A-R Tracking labels and overlays impact on job satisfaction			
OBJ-05.971-TRL4-TVALP-H104.1010 To assess that the technical systems for V/A-R Air Gestures support the ATCOs in performing their tasks V/A-R Air Gestures impact on ATCO tasks	CRT-05.971-TLR4-TVALP-H104-1011 Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures supports ATCO in maintaining workload at acceptable level	002: NOK, 005: POK	POK
	CRT-05.971-TLR4-TVALP-H104-1012 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R Air Gestures	002: NOK, 005: OK	
	CRT-05.971-TLR4-TVALP- H104-1013 Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures HMI supports ATCO in maintaining an adequate level of situation awareness	002: POK, 005: NOK	
	CRT-05.971-TLR4-TVALP-H104-1014 Measured time spent in head up is increased in the solution scenario with respect to the reference scenario	002: OK, 005: OK	
	CRT-05.971-TLR4-TVALP- H104-1015 V/A-R Air Gestures HMI does not increase the potential for human error	002: NOK, 005: OK	
	CRT-05.971-TLR4-TVALP- H104-1016 ATCOs (at least 75%) trust in the system is at an acceptable level	002, 005: OK	
	CRT-05.971-TLR4-TVALP- H104-1017 Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R Air Gestures HMI	002: NOK, 005: OK	

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
	CRT-05.971-TLR4-TVALP- H104-1018 ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R Air Gestures tool	002: POK, 005: OK	
OBJ-05.971-TRL4-TVALP-H104.1020 To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Air Gestures V/A-R Air Gestures on ATCO role	CRT-05.971-TLR4-TVALP-H104-1021 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner	002:OK, 005:NOK	POK
	CRT-05.971-TLR4-TVALP-H104-1022 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions	002,005: OK	
OBJ-05.971-TRL4-TVALP-H104.1030 To assess job acceptance and satisfaction with the introduction of V/A-R Air Gestures V/A-R Air Gestures impact on job satisfaction	CRT-05.971-TLR4-TVALP-H104-1031 ATCO (at least 75%) provide positive feedback on job satisfaction and acceptance	002,005: OK	OK
OBJ-05.971-TRL4-TVALP-H105.1010 To assess that the technical systems for V/A-R Attention Guidance support the ATCOs in performing their tasks V/A-R Attention Guidance impact on ATCO tasks	CRT-05.971-TLR4-TVALP-H105-1011 Majority of ATCOs (at least 75%) provide feedback that V/A-R Attention Guidance supports ATCO in maintaining workload at acceptable level	001: POK	POK
	CRT-05.971-TLR4-TVALP-H105-1012 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R Attention Guidance	001: POK	
	CRT-05.971-TLR4-TVALP- H105-1013 Majority of ATCOs (at least 75%) provide feedback that V/A-R Attention Guidance HMI	001: POK	

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
	supports ATCO in maintaining an adequate level of situation awareness		
	CRT-05.971-TLR4-TVALP- H105-1014 Measured time spent in head up is increased in the solution scenario with respect to the reference scenario	001: POK	
	CRT-05.971-TLR4-TVALP- H105-1015 HMI of V/A-R Attention Guidance tools does not overshadow the relevant information on the OTW view	001: POK	
	CRT-05.971-TLR4-TVALP- H105-1016 V/A-R Attention Guidance HMI does not increase the potential for human error	001: OK	
	CRT-05.971-TLR4-TVALP- H105-1017 ATCOs' (at least 75%) trust in the system is at an acceptable level	001: POK	
	CRT-05.971-TLR4-TVALP- H105-1018 Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R Attention Guidance HMI	001: POK	
	CRT-05.971-TLR4-TVALP- H105-1019 Majority of ATCOs (at least 75%) provide feedback that alarms and alerts are not too intrusive and support ATCOs in the early detection of ATC critical situations	001: POK	
	CRT-05.971-TLR4-TVALP- H105-1020 ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R Attention Guidance tool	001: POK	
	CRT-05.971-TLR4-TVALP- H105-1021 Majority of ATCOs (at least 75%) provide feedback that V/A-R HMI supports ATCO team (GND and TWR) in maintaining an acceptable level of situation awareness	001: POK	

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
OBJ-05.971-TRL4-TVALP-H105.1030 To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Attention Guidance V/A-R Attention Guidance on ATCO role	CRT-05.971-TLR4-TVALP-H105-1031 Majority of ATCOs (at least 75%) provide feedback that ATCOs can apply operating methods in an accurate, efficient and timely manner	001: OK	OK
	CRT-05.971-TLR4-TVALP-H105-1032 Majority of ATCOs (at least 75%) provide feedback that operating methods are clearly identified and consistent in all operating conditions	001: OK	
OBJ-05.971-TRL4-TVALP-H105.1040 To assess job acceptance and satisfaction with the introduction of V/A-R Attention Guidance V/A-R Attention Guidance impact on job satisfaction	CRT-05.971-TLR4-TVALP-H105-1041 ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance	001: OK	OK
OBJ-05.971-TLR4-TVALP-SAFE.1010 To assess the impact of Virtual/Augmented Reality applications on safety Safety Impact	CRT-05.971-TLR4-TVALP- SAFE -1011 The changes related to the implementation of Virtual/Augmented Reality applications do not increase potential for human error and therefore not reducing safety levels.	001,002, 005: OK	OK
	CRT-05.971-TLR4-TVALP- SAFE -1012 ATCO's workload with the implementation of Virtual/Augmented Reality applications is maintained at acceptable level and therefore not reducing safety levels.	001,002, 005: OK	
	CRT-05.971-TLR4-TVALP- SAFE -1013 ATCO's situational awareness with the implementation of Virtual/Augmented Reality	001,002, 005: OK	

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
	applications is maintained at acceptable level and therefore not reducing safety levels.		
	CRT-05.971-TLR4-TVALP- SAFE -1014 Safety assessment activities and the results are documented and integrated in the overall solution validation results	001,005: OK, 002:N/A	
OBJ-05.971-TLR4-TVALP-PERF.1010 To assess the performance benefits of equivalent visual operations for tower control through the use of applications for Virtual/Augmented Reality. TRL4 Performance Assessment	CRT-05.971-TLR4-TVALP-PERF-1011 Laboratory tests show that the use of V/A applications improves Cost Efficiency performance by reducing the cost per flight (through e.g. reduction of workload, reduction of delay times, ...)	001,002, 005:OK	OK
	CRT-05.971-TLR4-TVALP-PERF-1012 Laboratory tests show that the use of V/A reality applications improves Resilience by increasing situational awareness in Low visibility conditions while maintaining workload within acceptable limits	002,005: OK	

Table 11: Solution 97.1 technological validation exercises results

4.1.4 SESAR Technological Solution 97.2 Validation Results

Please refer to Appendixes C, E, F for detailed Technological Validation Exercises 004, 006, 007 results.

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
OBJ-05.972-TLR4-TVALP-FEAS.2010	CRT-05.972-TLR4-TVALP- FEAS -2011 No operational show-stoppers have been identified during laboratory tests (based on a	004,006, 007:OK	OK

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS. TLR4 Operational feasibility	prototype) related to the use of Automatic speech recognition.		
	CRT-05.972-TLR4-TVALP- FEAS -2012 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of AI suggestions.	004,006, 007:OK	
OBJ-05.972-TLR4-TVALP-FEAS.2020 To identify possible technical feasibility issues and possible show stoppers TLR4 Pre-industrial feasibility	CRT-05.972-TLR4-TVALP- FEAS -2021 Laboratory tests (based on a prototype) have verified the technical feasibility of the use of ASR supported by AI/ML	004,006, 007:OK	OK
	CRT-05.972-TLR4-TVALP- FEAS -2022 Laboratory tests have verified that the integration of the SESAR technological solution with other related system enablers is technically feasible	004,006, 007:OK	
OBJ-05.972-TRL4-TVALP-H106.2010 To assess that the technical systems for ASR support the ATCOs in performing their tasks ASR impact on ATCO tasks	CRT-05.972-TLR4-TVALP-H106-2011 Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining workload at acceptable level	004:N/A 006,007: OK 004:no measurement	POK
	CRT-05.972-TLR4-TVALP-H106-2012 Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining an adequate level of situation awareness	004:N/A 006,007: OK 004:no measurement	
	CRT-05.972-TLR4-TVALP- H106-2013 ASR does not increase the potential for human error	004:N/A 006,007: OK 004:no measurement	
	CRT-05.972-TLR4-TVALP- H106-2014 ATCOs (at least 75%) provide positive feedback	004:POK, 006,007: OK	

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
	on adequacy (level and quality) of information provided by ASR		
	CRT-05.972-TLR4-TVALP-H106-2015 Measured callsign recognition rate, command recognition rate, error rate and rejection rate of ASR system are at acceptable levels are considered within acceptable levels by the majority of ATCOS (at least 75%)	004:POK, 006,007: OK	
	CRT-05.972-TLR4-TVALP- H106-2016 Majority of ATCOs (at least 75%) confirm adequate usability of ASR system	004:POK 006,007: OK	
	CRT-05.972-TLR4-TVALP- H106-2017 Majority of ATCOs (at least 75%) provide positive feedback on acceptance of ASR tool	004:POK 006,007: OK	
	CRT-05.972-TLR4-TVALP- H106-2018 ATCOs (at least 75%) trust in the system is at an acceptable level	004:NOK 006,007: OK	
OBJ-05.972-TRL4-TVALP-H106.2020 To assess the role of the ATCO is consistent with human capabilities and limitations with the introduction of ASR ASR impact on ATCO role	CRT-05.972-TLR4-TVALP- H106-2021 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient, and timely manner	004:N/A 006,007: OK	POK
	CRT-05.972-TLR4-TVALP-H106-2022 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions	004:N/A 006,007: OK	
OBJ-05.972-TRL4-TVALP-H106.2030 To assess job acceptance and satisfaction with the introduction of ASR ASR impact on job satisfaction	CRT-05.972-TLR4-TVALP-H106-2031 ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance	004:N/A 006,007: OK	POK

SESAR Technological Validation Objective ID/Description/Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
OBJ-05.972-TLR4-TVALP-SAFE.2010 To assess the impact of Automatic Speech Recognition on safety. Safety Impact	CRT-05.972-TLR4-TVALP- SAFE -2011 The changes related to the implementation of Automatic Speech Recognition do not increase potential for human error and therefore not reducing safety levels.	004:N/A 006,007: OK	POK
	CRT-05.972-TLR4-TVALP- SAFE -2012 ATCO's workload with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	004:N/A 006,007: OK	
	CRT-05.972-TLR4-TVALP- SAFE -2013 ATCO's situational awareness with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	004:N/A 006,007: OK	
	CRT-05.97B-TLR4-TVALP- SAFE -2014 Safety assessment activities and the results are documented and integrated in the overall solution validation results	004: in TSAR 006,007: OK	
OBJ-05.972-TLR4-TVALP-PERF.2010 To assess the performance benefits of Automatic Speech Recognition supported by AI/ML. TLR4 Performance Assessment	CRT-05.972-TLR4-TVALP-PERF-2011 Laboratory tests show that the SESAR technological solution improves Cost Efficiency performance by reducing cost per flight (PER.TRL4.3) (through e.g. reduction of workload, reduction of delay times,….)	004:N/A 006,007: OK	POK

Table 12: Solution 97.2 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 criteria 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 Solution 97.1 Objectives and Success Criteria

4.2.1.1 OBJ.05.971-TRL4-TVALP-FEAS.1010

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-FEAS-1011	ATCOs happy about AI guidance (tracking labels) But felt a heavy head by the end of the day due to wearing the hardware for long	Concept feasible when addressing V or A-R in different visibility settings with tracking labels	Improvements needed in data stability, size of the markers in HMI and weight of the glasses	OK
CRT-05.971-TLR4-TVALP-FEAS-1012	-	Concept feasible when addressing Air Gestures to issue non-time critical clearances on the TWR GND post. Ideas given to improve air gesture functions	Advances needed in training and use of air gestures. The glasses used were 1st gen Hololens. Newer glasses include air gestures more intuitive to learn	OK
CRT-05.971-TLR4-TVALP-FEAS-1013	Concept feasible when addressing Attention Guidance with A-R device. Progress to be made in symbology and attention guidance cues timing	-	-	OK

Table 13: Solution 97.1 OBJ.05.971-TRL4-TVALP-FEAS.1010 results

4.2.1.2 OBJ-05.971-TRL4-TVALP- FEAS.1020

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-FEAS-1021	A-R symbology correlated well with objects in the simulated outside view and tracking labels followed a/c. Visibility of the symbology at times competing with external reflections	A-R overlays correctly collimated with the simulated out-of-the-tower view, including dynamic objects (a/c, gnd vehicles).	Lab tests (prototype based) have verified the technical feasibility of the use of V/A-R applications in the tower environment.	OK
CRT-05.971-TLR4-TVALP-FEAS-1022	A-R attention guidance correctly received information from the alerting system	-	Laboratory tests have verified that the integration of the V/A-R applications with other related system enablers is technically feasible	OK

Table 14: Solution 97.1 OBJ-05.971-TRL4-TVALP-FEAS.1020 results

4.2.1.3 OBJ-05.971-TRL4-TVALP-H103.1010

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-H103-1011	-	Feedback from ATCOs indicates acceptable workload level	ATCOs state that weight and shape of the A-R device influence greatly perceived workload	POK
CRT-05.971-TLR4-TVALP-H103-1012	-	90% ATCOs give positive feedback on information amount delivered by V/A-R. 50% ATCOs give positive view on information quality	Majority of ATCOs responded favourably	OK
CRT-05.971-TLR4-TVALP-H103-1013	-	90% of ATCOs respond that V/A-R HMI supports keeping enough situation awareness	50% of ATCOs responded favourably	POK
CRT-05.971-TLR4-TVALP-H103-1014	-	Measured time spent in head up is increased in the solution scenario	Measured head up time increased according to 55% of ATCOs, compared	OK

	EXE-001	EXE-002	EXE-005	Solution 97.1
		with respect to the reference scenario.	with reference scenario	
CRT-05.971-TLR4-TVALP-H103-1015	-	80% ATCOs agreed they always had adequate field of view with V/A-R to perform their task. 50% ATCOs agreed tracking label and airport overlay provided by V/A-R did not generate confusion nor disturbance. It was due to labels overlapping, covering the background and not always being clearly aligned with related A/C's	HMI of V/A-R tools does not minimize relevant information on OTW view. Average normalized responses: 63% favourable rate	POK
CRT-05.971-TLR4-TVALP-H103-1016	-	60% ATCOs agreed V/A-R system did not increase potential for human error compared to current operations.	V/A-R HMI does not increase potential for human error. Average of normalized responses: 64% favourable rate	POK
CRT-05.971-TLR4-TVALP-H103-1017	-	90% ATCOs trust the system's reliability	ATCOs trust in the system: 75% in favour. Average of normalized responses: 60% favourable rate	POK
CRT-05.971-TLR4-TVALP-H103-1018	-	80% ATCOs state: system easy to use, clear and complete interface, device is physically comfortable. 50% ATCOs would like to use the system frequently 70% ATCOs think most people can learn to use the system very quickly	ATCOs confirm adequate level of usability of V/A-R HMI. 100% of users responded favourably. Average of normalized responses: 73% favourable rate	POK

	EXE-001	EXE-002	EXE-005	Solution 97.1
		and felt confident using the system and the device did not cause adverse physical concerns like eyestrain. 40% ATCOs think there were not too many inconsistencies		
CRT-05.971-TLR4-TVALP-H103-1019	-	100% ATCOs think alerts in the V/AR prototype are effective and not intrusive. 90% ATCOs think alerts support them in the early detection of critical situations (conflicting clearances, runway incursions)	-	OK
CRT-05.971-TLR4-TVALP-H103-1020	-	80% ATCOs give positive feedback on V/A-R tool acceptance	ATCOs response is favourable (75%). Average of normalized responses: 65% favourable rate	OK
CRT-05.971-TLR4-TVALP-H103-1021	-	100% ATCOs state V/A-R HMI supports ATCO team (GND, TWR) in retaining enough situational awareness	-	OK

Table 15: Solution 97.1 OBJ-05.971-TRL4-TVALP-H103.1010 results

4.2.1.4 OBJ-05.971-TRL4-TVALP-H103.1030

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-H103-1031	-	90% ATCOs indicate they can apply operating methods in an	75% ATCOs indicate they can apply operating methods in an accurate, efficient and timely manner.	OK

	EXE-001	EXE-002	EXE-005	Solution 97.1
		accurate, efficient and timely manner	Average of normalized responses: 59% favourable rate	
CRT-05.971-TLR4-TVALP-H103-1032	-	50% ATCOs think operating methods are clearly identified and consistent in the simulated operating conditions	75% ATCOs replied favourably. Normalized responses average: 55% favourable rate	POK

Table 16: Solution 97.1 OBJ-05.971-TRL4-TVALP-H103.1030 results

4.2.1.5 OBJ-05.971-TRL4-TVALP-H103.1040

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-H103-1041	-	85% ATCOs answer positively with regard to job satisfaction and acceptance of V/A-R for tracking labels and overlays	75% ATCOs replied favourably. Normalized responses average: 65% favourable rate	OK

Table 17: Solution 97.1 OBJ-05.971-TRL4-TVALP-H103.1040 results

4.2.1.6 OBJ-05.971-TRL4-TVALP-H104.1010

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-H104-1011	-	40% ATCOs think V/A-R Air Gestures support ATCO in maintaining workload at acceptable levels. Several ATCOs had difficulties using Air Gestures, increasing their workload.	75% ATCOs responded favourably. Average of normalized replies: 66% favourable	POK
CRT-05.971-TLR4-TVALP-H104-1012	-	Qualitative feedback collected during debriefings due to this being a technical test. ATCOs said they had difficulties using Air Gestures as the system did not always identify them	75% ATCOs answered favourably. Average of normalized responses: 60% favourable rate	POK
CRT-05.971-TLR4-TVALP-H104-1013	-	60% ATCOs responses is V/A-R Air Gestures HMI supports ATCO in maintaining an adequate level of situation awareness. Due to not always being able to issue clearances with air gestures	50% ATCOs of the users responded favourably. Average of normalized responses: 69% favourable rate	POK
CRT-05.971-TLR4-TVALP-H104-1014	-	Measured head up time is increased in the Air Gesture solution scenario with respect to the reference scenario	Average of normalized responses show a 55% favourable response rate to a head up time increase	OK

CRT-05.971-TLR4-TVALP-H104-1015	-	40% ATCOs state that V/A-R Air Gestures increase potential for human error due to usability issues	Average of normalized responses shows 69% favourable response rate. Increase in situational awareness reduces likelihood of human error such as allowing two simultaneous a/c at the same low visibility block. Human error is also assessed through observations. Initially ATCOs experienced some difficulties for gestures to be recognized due to deficient training. No human errors were observed	POK
CRT-05.971-TLR4-TVALP-H104-1016	-	80% ATCOs trust the prototype for V/AR Air Gestures	75% ATCOs trust the test system. Average of normalized responses: 60% in favour	OK
CRT-05.971-TLR4-TVALP-H104-1017	-	20% ATCOs stated V/A-R Air Gestures have no impact on usability. 80% ATCOs believes that usability is negatively affected, from an ergonomic point of view.	ATCOs responded favourably regarding usability. Normalized positive responses average: 67%	POK
CRT-05.971-TLR4-TVALP-H104-1018	-	One ATCO thought air gestures should be removed and two ATCOs suggested that air gesture orders be avoided for runway authorizations/critical cases	All ATCOs replies were positive for acceptance of V/A-R Air Gestures tool. The average of normalized replies is 63% positive	POK

Table 18: Solution 97.1 OBJ-05.971-TRL4-TVALP-H104.1010 results

4.2.1.7 OBJ-05.971-TRL4-TVALP-H104.1020

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-H104-1021	-	ATCOs mentioned no specific impacts of Air Gestures on operation methods	75% positive replies normalized average positive replies:59%	OK
CRT-05.971-TLR4-TVALP-H104-1022	-	ATCOs mentioned no specific impacts of Air Gestures on operation methods	75% positive replies normalized average positive replies:55%	OK

Table 19: Solution 97.1 OBJ-05.971-TRL4-TVALP-H104.1020 results

4.2.1.8 OBJ-05.971-TRL4-TVALP-H104.1030

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-H104-1031	-	80% ATCOs give positive feedback on job satisfaction and acceptance for V/AR Air Gestures	normalized average positive replies:63% (all questions count in evaluation)	OK

Table 20: Solution 97.1 OBJ-05.971-TRL4-TVALP-H104.1030 results

4.2.1.9 OBJ-05.971-TRL4-TVALP-H105.1010

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-H105-1011	No significant differences in workload found between baseline and V/A-R in post-run ratings. Expected influence of the A-R on workload was rated 'positive' to 'very positive'. Some extra workload can be explained by the acknowledgement of alerts (clicking) and due to taxi nuisance conflicts	-	-	POK
CRT-05.971-TLR4-TVALP-H105-1012	ATCOs could provide instructions immediately as location and call signs were visible in the A-R device, no need to look down to the flight strips. ATCOs were happy with the call signs, type and location of the alert but extra information in the middle of the field of view was not appreciated.	-	-	POK

	ATCOs liked labels of all a/c being visible.			
CRT-05.971-TLR4-TVALP-H105-1013	The influence of the A-R on Situational Awareness was rated 'positive' to 'very positive'. The post-run ratings of Situational Awareness show that the level of SA is not decreasing.	-	-	POK
CRT-05.971-TLR4-TVALP-H105-1014	No head up time was measured, but controllers mentioned they appreciated to be able to stay heads-up in case of an alert.	-	-	POK
CRT-05.971-TLR4-TVALP-H105-1015	The test shows the HMI can be improved regarding some of the symbology, in particular the (re)appearance of the alert notification.	-	-	POK
CRT-05.971-TLR4-TVALP-H105-1016	No errors were noticed in the experiment. Furthermore, the system did not interfere with actions to be taken by the controller.	-	-	OK
CRT-05.971-TLR4-TVALP-H105-1017	ATCO trust in the system was rated with a trend towards being not acceptable. One controller was positive and the other more negative. Automated reappearance of alerts not fully understood and taxi conflicts were experienced as nuisance alerts.	-	-	POK
CRT-05.971-TLR4-TVALP-H105-1018	ATCO ratings of Usability (System Usability Scale) were 40 to 52.5 (on scale of 1 to 100), not yet acceptable	-	-	POK
CRT-05.971-TLR4-TVALP-H105-1019	Alert notification presentation can be improved by avoiding showing notification label in the middle of the field of view. Labels and dotted lines to indicate expected movement	-	-	POK
CRT-05.971-TLR4-TVALP-H105-1020	The HMI needs improvement before being acceptable regarding the (re)appearance of alerts.	-	-	POK
CRT-05.971-TLR4-TVALP-H105-1021	Team SA was not part measured in the experiment, because there was no team set-up. The expected effects of the A-R on team SA were rated 'neutral' to 'positive' in the post-experiment questionnaire.	-	-	POK

Table 21: Solution 97.1 OBJ-05.971-TRL4-TVALP-H105.1010 results

4.2.1.10 OBJ-05.971-TRL4-TVALP-H105.1030

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-H105-1031	ATCOs could react immediately to alerts and rated the influence of the A-R on identifying and situating the involved a/c as well as ATCO response time as 'positive' to 'very positive'.	-	-	OK
CRT-05.971-TLR4-TVALP-H105-1032	A-R provides additional information to ATCO but has no effect on the operating methods or procedures. Information is consistent with the alert content presented head-down	-	-	OK

Table 22: Solution 97.1 OBJ-05.971-TRL4-TVALP-H105.1030 results

4.2.1.11 OBJ-05.971-TRL4-TVALP-H105.1040

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-H105-1041	ATCOs rated the question whether they would like to use the system frequently as 'neutral' to 'positive'.	-	-	OK

Table 23: Solution 97.1 OBJ-05.971-TRL4-TVALP-H105.1040 results

4.2.1.12 OBJ-05.971-TLR4-TVALP-SAFE.1010

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-SAFE-1011	ATCOs rated influence of A-R on safety as 'neutral' to 'very positive'. As ops procedures do not change it is not expected to influence human error rate	ATCOs agreed V/A-R system helped them in the early detection of critical situations and indicated that Safety Nets positively impacted the potential for Human Error	An average of the normalized responses show 69% positive response. Increased situational awareness reduces the likelihood of human error such as allowing two simultaneous a/c at the same low visibility block	OK
CRT-05.971-TLR4-TVALP-SAFE-1012	A-R influence on workload was rated 'positive' to 'very positive'.	Workload was satisfactory without reduction.	V/A-R applications improve safety performance by reducing ATCO	OK

	Post-run ratings of Workload show that the Workload is maintained as acceptable		workload as shown by a 69% rate of normalized responses	
CRT-05.971-TLR4-TVALP-SAFE-1013	The level of SA is maintained (post-run ratings) or expected to be increased (post-experiment ratings).	ATCO's situational awareness with the implementation of Virtual/Augmented Reality applications is maintained at acceptable level and therefore not reducing safety levels, as positive replies rates were above limits	V/A-R applications improve safety performance by increasing situational awareness. As indicated by the average of normalized responses (80% positive rate)	OK
CRT-05.971-TLR4-TVALP-SAFE-1014	ATCOs indicate SA is built up faster and instructions can be given earlier since users do not have to look down for the alert type, location or call signs. Improved presentation and understanding of safety-relevant events is expected to have a positive impact on safety.	-	The only comment regarding Safety was related to some data dropouts of the altitude and speed indicators on the flight tags. An average of the normalized responses show a 57% favourable response rate.	OK

Table 24: Solution 97.1 OBJ-05.971-TLR4-TVALP-SAFE.1010 results

4.2.1.13 OBJ-05.971-TLR4-TVALP- PERF.1010

	EXE-001	EXE-002	EXE-005	Solution 97.1
CRT-05.971-TLR4-TVALP-PERF-1011	A-R guidance is allowing the ATCO to be more rapid with instructions concerning safety-relevant events. Validation has shown that	V/A-R prototype can contribute to an assessment of Cost Efficiency performance by having a positive impact on situation awareness, workload and	V/A applications improve Cost Efficiency performance by reducing the cost per flight (through e.g. reduction of workload, reduction of delay	OK

	ATCOs perceive a positive impact on SA and workload	efficiency of ground operations.	times). Average of the normalized responses is 100% positive	
CRT-05.971-TLR4-TVALP-PERF-1012	-	Laboratory tests showed V/A-R applications improve Resilience by increasing situational awareness in Low visibility conditions while maintaining workload within acceptable limits	V/A-R reality applications improve Resilience by increasing SA in low visibility keeping workload at acceptable levels. Average normalized replies: 100% positive. Outcome needs to be confirmed as the SA outcome was below expectation	OK

Table 25: Solution 97.1 OBJ-05.971-TLR4-TVALP-PERF.1010 results

4.2.2 Solution 97.2 Objectives and Success Criteria

4.2.2.1 OBJ-05.972-TLR4-TVALP-FEAS.2010

	EXE-004	**EXE-006**	**EXE-007**	**Solution 97.2**
CRT-05.972-TLR4-TVALP-FEAS-2011	The validation exercise did not reveal any operational showstoppers	No operational showstopper has been identified. The mean value of a questionnaire item regarding ASR show-stoppers was in the acceptable range. Experiment confirmed that the concept is operationally feasible when addressing the ASR use cases	No ASR malfunctions occurred, other than some learning curve for ATCOs in using the ASR recording widget (human errors, retries etc…)	OK
CRT-05.972-TLR4-TVALP-FEAS-2012	The validation exercise did not reveal any operational showstoppers	The mean value of a questionnaire item regarding operational AI show-stoppers was in the acceptable range. However, the AI itself was not transparent to	No hiccups during the runs and some minor issues before simulations were due to the WP and simulation platform	OK

		the ATCo while working		

Table 26: Solution 97.2 OBJ-05.972-TLR4-TVALP-FEAS.2010 results

4.2.2.2 OBJ-05.972-TLR4-TVALP-FEAS.2020

	EXE-004	EXE-006	EXE-007	Solution 97.2
CRT-05.972-TLR4-TVALP-FEAS-2021	It is technically feasible to use ASR supported by AI/ML to assist, or automate, selected Aerodrome ATC system inputs.	The mean value of a questionnaire item regarding technical ASR/AI show-stoppers was in the acceptable range	AI/ML techniques were employed to generate a working ASR module. Usage of context-based data has improved performance with no issues	OK
CRT-05.972-TLR4-TVALP-FEAS-2022	The validation exercise verified integration between an ASR module and an Aerodrome ATC system.	The ABSR system was perceived as well integrated. The ABSR system was supporting ATCos throughout all solution runs.	Integration with WP and in general LIS platform did not show issues during runs	OK

Table 27: Solution 97.2 OBJ-05.972-TLR4-TVALP-FEAS.2020 results

4.2.2.3 OBJ-05.972-TRL4-TVALP-H106.2010

	EXE-004	EXE-006	EXE-007	Solution 97.2
CRT-05.972-TLR4-TVALP-H106-2011	Not applicable as workload could not be measured during the tests	No significant differences in workload found between reference and solution scenarios. The secondary task (sorting cards) and ISA tend to show a workload reduction of ATCos when being supported by ASR, NASA-TLX and Bedford Workload Scale tend into the opposite way. The statement “I think that ASR supports me in maintaining workload at	100% ATCO feedback shows that ASR supports controllers in maintaining an acceptable level of workload	POK

		acceptable level" was rated with 7.8 on a 10 point scale (90% ATCos rated with 7 or above)		
CRT-05.972-TLR4-TVALP-H106-2012	Not applicable as situational awareness could not be measured during the tests	No significant differences in SASHA score (SA assessment) found between reference and solution scenarios. The statement "I think that ASR supports me in maintaining an adequate level of situation awareness" was rated with 7.7 on a 10 point scale (90% ATCos rated with 7 or above)	100% of ATCOs responded that ASR supports them in maintaining an adequate level of situational awareness	POK
CRT-05.972-TLR4-TVALP-H106-2013	Not applicable as human error could not be measured during the tests	As per ATCos questionnaire ratings, potential for human errors is not increased	66% of ATCOs agreed that ASR did not increase potential for human error compared to current ops	POK
CRT-05.972-TLR4-TVALP-H106-2014	66% of ATCOs indicated level of information provided by ASR is adequate.	ASR adequacy result is supported by the analysis of word error, call sign recognition error and command recognition error rates, and callsign prediction rates. System performance was made worse due to a software issue (see numbers in Appendix E for online and offline	100% of ATCOs gave positive feedback on adequacy of ASR feedback	POK

		recognition in detail)		
CRT-05.972-TLR4-TVALP-H106-2015	The subjective perception of callsign, clearance recognition and understanding other parameters were below threshold(75%).However, speed and accuracy of callsign recognition were highly regarded.	Accuracy of ASR was rated to be good (callsign highlighting 8.9/10, other values for e.g., commands around 7/10)	100% of ATCOs gave positive replies on callsign and command recognition rates. 66% of ATCOs on callsign and command rejection rates	POK
CRT-05.972-TLR4-TVALP-H106-2016	Usability in terms of ASR performance was not optimal, and HMI related questions were below criteria limits, perhaps as ASR HMI was not designed for an operational validation.	The system usability scale (SUS) score was 75 for solution (with ASR) compared to 71 for baseline (without ASR).	100% of ATCOs provided positive feedback on usability of the ASR system	POK
CRT-05.972-TLR4-TVALP-H106-2017	The design of the ASR module was overall acceptable for the ATCOs and also many improvement ideas have been gathered to further enhance the system	80% of ATCos stated with 8/10 or more points that they would accept such an ASR system in their normal CWP	100% of ATCOs provided positive feedback on acceptance of the ASR tool.	POK
CRT-05.972-TLR4-TVALP-H106-2018	The usability in terms of ASR performance was far from optimal, and the SATI results did not meet the cut-off score criteria either	80% of ATCos stated with 6/10 or more points (so above scale mean) that they had trust in the ASR system	83% of ATCOs provided positive feedback on trust in the ASR tool	POK

Table 28: Solution 97.2 OBJ-05.972-TRL4-TVALP-H106.2010 results

4.2.2.4 OBJ-05.972-TRL4-TVALP-H106.2020

	EXE-004	**EXE-006**	**EXE-007**	**Solution 97.2**
CRT-05.972-TLR4-TVALP-H106-2021	Not applicable as it was not a fully operational validation therefore this objective could not be measured during the tests	80% of ATCos stated with 8/10 or more points that they could apply operating methods in a timely manner	ATCOs mentioned that they were satisfied about latency and feedback provided which allowed them to apply operating methods in an accurate, efficient, and timely manner.	POK

			They also mentioned they experienced no change in operating methods.	
CRT-05.972-TLR4-TVALP-H106-2022	Not applicable as it was not a fully operational validation therefore this objective could not be measured during the tests	80% of ATCOs stated with 6/10 or more points (so above scale mean) that operating methods were clearly identified and consistent in all operating conditions	All ATCO responded they were able to apply unchanged operating methods in an accurate, efficient, and timely manner	POK

Table 29: Solution 97.2 OBJ-05.972-TRL4-TVALP-H106.2020 results

4.2.2.5 OBJ-05.972-TRL4-TVALP-H106.2030

	EXE-004	**EXE-006**	**EXE-007**	**Solution 97.2**
CRT-05.972-TLR4-TVALP-H106-2031	Not applicable as it was not a fully operational validation therefore this objective could not be measured during the tests	ATCOs were satisfied working with the system overall - 100% of ATCOs stated this with 6/10 or more points (above scale mean).	ATCOs provided positive feedback on job satisfaction and acceptance.	POK

Table 30: Solution 97.2 OBJ-05.972-TRL4-TVALP-H106.2030 results

4.2.2.6 OBJ-05.972-TLR4-TVALP-SAFE.2010

	EXE-004	**EXE-006**	**EXE-007**	**Solution 97.2**
CRT-05.972-TLR4-TVALP-SAFE-2011	Not applicable as human error could not be measured during the tests.	Due to the ATCos questionnaire ratings, the potential for human errors have not been increased	66% of ATCOs responded ASR does not increase the potential for human error compared to current operations	POK
CRT-05.972-TLR4-TVALP-SAFE-2012	Not applicable as workload could not be measured during the tests	No significant differences in workload were found between reference and solution scenario	ATCO's workload with the implementation of Automatic Speech Recognition is maintained at acceptable level. 83% ATCOs agreed	POK

			safety level was at least equal to today operations	
CRT-05.972-TLR4-TVALP-SAFE-2013	Not applicable as situational awareness could not be measured during the tests	No significant differences in SASHA score found between reference and solution scenarios	ATCO situational awareness is maintained at acceptable levels with ASR and therefore does reduce safety levels	POK
CRT-05.972-TLR4-TVALP-SAFE-2014	Safety assessment activities, conducted according to SESAR SRM, are documented in TSAR.	Safety assessment activities and results are documented and integrated in overall solution validation results	The safety aspect related to the ASR was investigated across the whole validation exercise. No specific safety related issue has been identified.	POK

Table 31: Solution 97.2 OBJ-05.972-TLR4-TVALP-SAFE.2010 results

4.2.2.7 OBJ-05.972-TLR4-TVALP-PERF.2010

	EXE-004	**EXE-006**	**EXE-007**	**Solution 97.2**
CRT-05.972-TLR4-TVALP-PERF-2011	The exercise was not designed as an operational use of ASR and to measure workload	A significant reduction of workload or delay times could not be shown in the laboratory trials. Due to the nature of the multiple remote tower setup (no dense traffic at neither of the three airports), an improvement in cost efficiency could not be shown. However, due to the verbal feedback of ATCos, a support of ASR at a tower CWP can be assumed	Qualitative feedback from ATCOs indicates at least no detrimental effect of ASR on performance	POK

Table 32: Solution 97.2 OBJ-05.972-TLR4-TVALP-PERF.2010 results

4.3 Confidence in the Validation Results

4.3.1 Limitations of Technological Validation Results

Targeting TRL4, validation exercises could not be considered to be operational. Technology Readiness Level 4 is defined as “technology validated in lab”, therefore with some inherent limitations. 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 Solution 97.1

4.3.1.1.1 EXE-001

Alerting and related attention getting and guidance were meant to direct the focus of the controller on a particular situation that, without intervention, was expected to lead to a conflict: no limitations or impact on the level of significance for the A-R device operating method. The main difference was a limited scope, because the exercise had a focus on certain parts of the operation in the tower, in particular the work of the tower controllers in a specific critical situation. The focus was thus put on Safety and Human Performance issues: more attention was given to Tower Runway controller or Tower Ground controller. Task sharing among the controller team in the tower was not assessed: one working position was measured at a time, while the second controller had an observer role. No performance issues concerning the complete Schiphol operation were assessed since the focus was put on specific situations to resolved with Attention Guidance.

An important aspect to point out is that while Schiphol was the geographical environment used, traffic was reduced to an off-peak situation (1 landing, 1 departing runway, as opposed to normal operation which is 2-1 or 1-2), so traffic and controllers (only one ground/apron/taxi controller) were reduced compared with normal operation.

4.3.1.1.2 EXE-002

Considering the simulation conditions, the results for V/A-R TL are judged to be characterised by a high level of significance, even if the training of ATCO was quite limited for time constraint reasons and it might have affected the collection of data of initial runs of each simulation day. This lack of training impact is limited considering the very intuitive tools employed in the simulations. The scope of V/A-R Air Gesture was reduced to a technical test: each run duration was 15 minutes, not enough to measure HP & SAF quantitative indicators, hence subjective feedback during debriefings was collected.

4.3.1.1.3 EXE-005

Technologically speaking, since it was a shadow mode exercise, using real ADS-B data, performed in an actual tower environment, and overlaying HMI on actual traffic, confidence was very high with a target maturity level 4.

4.3.1.2 Solution 97.2

4.3.1.2.1 EXE-004

The validation exercise was neither a pure operational validation nor a pure technical validation: it used ATCOs for collecting data, but still not using something that could be regarded as an operational ASR based input of clearances/instructions into an TWR ATC system. The validation exercise was more a hybrid between the two validation types. The validation platform was created more for validating the technical feasibility of using ASR to automate inputs in a TWR ATC system.

4.3.1.2.2 EXE-006

The validation is based on real-time simulation environment addressing speech recognition in a multiple remote tower environment – three towers named Vilnius, Kaunas and Palanga. 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. For more detail, please see section E.7.4.1.

4.3.1.2.3 EXE-007

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.

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.

4.3.2.1 Solution 97.1

4.3.2.1.1 EXE-001

The level of representativeness and quality of the simulation is considered as high. This is due to the proven record of the NARSIM Tower validation platform in carrying out simulations of Amsterdam Airport Schiphol for ATC the Netherlands (LVNL). These activities not only consisted of research projects, but very often focused on very practical applications, such as the Schiphol Winter Training and the course programme for Schiphol Tower Control given to ATC operational experts as well as experts from related disciplines. More recently, the same environment was also used for training tower controllers in the use of the acquired EFS system. For that purpose, the actual EFS system was integrated into the simulation platform to allow for all desired manipulations. In summary, the simulation platform used was considered to be of high realism and quality.

The simulated scenarios were considered moderately realistic. In particular the scenario with taxiway conflicts was seen as less realistic. The reason for this is that the researchers needed to elicit taxiway conflicts in order to expose the ATCOs to the warnings and guidance associated with these events. After all, without these conflicts, validation of the attention guidance would have been more difficult (see also the comment on nuisance alerts in Section A.6).

The differences with the real tower environment that were reported are, obviously, the lighting conditions. A simulator cannot offer the same brightness as real sunlight. Apart from the differences mentioned above, the ATCOs found that the simulator offered a realistic representation of the tower environment that they know.

The same goes for the A-R device. Usage of the A-R device in such a (very realistic) simulation environment did not lead to any limitations from a purely conceptual point of view. Even though lighting conditions in the simulation environment were very different from the conditions in a real tower environment and were even less favourable, the A-R device managed to keep track of its own position and presented the holographic objects with high precision. System engineers tested this thoroughly and found that there was excellent alignment between a/c positions on the projection screen and within the HoloLens. While tuning to the specific situation of the observer will still play a role (tuning it is a HoloLens function but controllers initially might need assistance from people familiar with the device) this fact certainly contributed to the quality of the exercise.

4.3.2.1.2 EXE-002

One issue recorded for simulation day 2 was the temporary failure of HoloLens audio system that was overcome by the use of headsets. This issue anyway is judged not affecting the collection of data or the provided results, considering also the feedback provided by the ATCOs involved in the specific day.

Questionnaires have been used to collect ratings from the test subjects on the different aspects of V/A-R TL as explained in section A.7: 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 TRL4 phase.

4.3.2.1.3 EXE-005

The quality of the results for a TLR4 exercise is considered high but the fact that no antenna coverage analysis was performed prior to the execution of the exercise should still be taken into consideration

4.3.2.2 Solution 97.2

4.3.2.2.1 EXE-004

The quality of the results in the area of recognizing commands using ASR technology was good. This was also the focus of the exercise and validation platform set-up.

The quality of the results relating to using the recognized commands to automate ATC system HMI input was not that good, as the focus was more on trying to automate as many inputs as possible. The validation platform setup was also missing some inputs that often is performed for every flight, but those were not in the predefined list of clearances and did not have any impact on strip status (i.e. “backtrack” or “joint traffic circuit” were frequently used but those instructions did not create an event to update the flight strip)

The measurements of recognizing call signs and clearances and data collection were performed during all the simulation runs.

During the runs, apart from the predefined scenario, there were also practice sessions and other experimental activities with the system, free tests, and on occasion some end-users deliberately pronounced call signs incorrectly, to see how the system responds. This definitely influenced the quality of the result. To perform a pure laboratory test with scripted call signs and clearances could show a result with better quality of exercise result.

4.3.2.2.2 EXE-006

The quality of the validation results is determined as medium 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.

4.3.2.2.3 EXE-007

Questionnaires have been used to collect ratings from the test subjects on the different aspects of ASR as explained in section F.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 TRL4 phase.

4.3.3 Significance of Technological Validation Exercises Results

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

4.3.3.1 Solution 97.1

4.3.3.1.1 EXE-001

The simulation exercise was based on the participation of two (former Schiphol) air traffic controllers, with one controller actively involved in control and guidance activities, and the other controller providing assistance. Both controllers were wearing a HoloLens, but only the controller giving ATC instructions was considered an exercise subject and was filling out questionnaires. The work between controllers was evenly divided and both controllers took part in debriefings and interviews.

In summary, this means that this approach, as a first introduction of both a technological enabler and an operational concept for Attention Guidance in a limited Schiphol environment, does obviously not give reliable results concerning an impact on the Schiphol operation as a whole or even in part. That will only be possible, if the recommended changes in the Attention Guidance symbology and logic are realized and re-evaluated and a larger operational scope including adaptation of required information to different controller roles and team working aspects has been investigated.

Nevertheless, the feedback obtained from the simulations led to new ideas regarding all aspects of the Attention Guidance, namely the cues used, the information provided and the triggers for different concept phases (e.g. when to remind the controller of an alerting situation). As such the exercise results are very valuable and can be built upon when continuing research in this area and scaling it up towards different controller roles and interaction between controllers.

As a consequence, the conclusions and recommendations that will follow are also based heavily on the debriefing and interview results, as these results offer more insight into the procedural aspects and the experience that the controllers had with both the HoloLens and the Attention Guidance concept. All other results must also be seen in the light of the limitations of the set-up and the fact that this was the first time indeed that controllers from Schiphol airport were confronted with the use of an Augmented Reality device in their working environment.

As an output of the validation exercise, the following results were obtained in the workload assessment:

Reference	Solution	WL Reduction
1.8	1.0	44,44%

4.3.3.1.2 EXE-002

The simulation exercise have been conducted on an experimental platform representing Bologna airport environment with a high degree of fidelity providing an operational significance adequate to support the TRL4 maturity assessment, of course with the limitations already mentioned in above sections 1 and 2.

A significant number of total run have been conducted among 5 simulation days (25 total number of runs) as well as a significant number of test subjects (10 ATCOs) have been involved to conclude that results are significant to support the TRL4 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.

As an output of the validation exercise, the following results were obtained in the workload assessment:

Reference	Solution	WL Reduction
9.3	8.8	5.38%

4.3.3.1.3 EXE-005

When looking at the questionnaire results, the fact that there were only 4 participants can lead to one outlier score skewing the results. However, since the results broadly follow previous results from RETINA, a medium confidence in the results of this exercise can be assumed.

As an output of the validation exercise, the following results were obtained in the workload assessment:

Reference	Solution	WL Reduction
2.8	2.0	28.57%

4.3.3.2 Solution 97.2

4.3.3.2.1 EXE-004

As this technical validation was executed in a very operational environment and closely related to an operational validation of remote tower operations, it gave an impression of an "unfinished" functionality, compared to other functionality validated in the remote tower validation. The results however demonstrate the feasibility of automating ATC system HMI input based on ASR technology. Validating ASR in this hybrid environment also highlights that the set of inputs required by the ATC system shall be explicitly defined, as there will be instructions that will not affect the automatic EFS update process but could have an impact on user acceptance.

The validation exercise was based on assumption that it is the spoken instructions and clearances by the ATCO that can be recognized and used as automated system input. There could also be ASR of the requests, readback and other utterance from the flight crew transmitted by radio.

As an output of the validation exercise, no quantitative results were obtained in the workload assessment.

4.3.3.2.2 EXE-006

Each of the ten ATCos did two runs per day: one reference and one solution run (alternating order to avoid learning effect in the data). Each ATCO was working simultaneously with three airports and corresponding displays.

As each ATCo was at DLR from 8:30 to 16:30, we also did exactly 10 validation days, i.e., we had some days in between where there was no ATCo at DLR, e.g., because PANSA cancelled their participation. Start was Feb 14, end was Mar 3. Hence, in sum ten ATCo feedbacks were collected for the whole validation. For operational significance the existing airspace and applicable procedures and corresponding letters of agreement were applied. All participants were holders of an active tower ATCo licence.

Statistical significance was rather absent in the quantitative questionnaire results of ATCos due to high standard deviations. However, the recognition and error rates of the ABSR system are based on a lot of utterances. Hence, these numbers have higher significance.

As an output of the validation exercise, the following results were obtained in the workload assessment:

Reference	Solution	WL Reduction
4.0	3.2	20.00%

4.3.3.2.3 EXE-007

The simulation exercise has been conducted on an experimental platform representing Sofia Airport environment with a high degree of fidelity providing an operational significance adequate to support the TRL4 maturity assessment, of course with limitations already mentioned in Sections F.7.1 and F.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 TRL4 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.

As an output of the validation exercise, the following results were obtained in the workload assessment:

Reference	Solution	WL Reduction
3.5	2.1	40.00%

4.3.4 Workload Assessment

The assessment of mental workload is the main basis to carry out an evaluation of benefits generated by the solution.

From an analysis of the results shown in the Appendix, which can be referred to for detailed information, the outcome is a positive feasibility, both qualitative and quantitative, for Workload reduction.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

5 Conclusions and recommendations for 97.1

5.1 Conclusions

Solution 97.1 is a promising step in the direction of introducing novel human machine interface methods in the Tower Ground environments, with the introduction of Virtual and Augmented Reality technology. Some considerations follow regarding the outcome of the Validation Exercises which make up this SESAR Solution.

5.1.1 Maturity

Maturity depends on a number of factors, and given the expected level of readiness is TRL4, this technological solution can be considered to be valid at a Laboratory level, albeit running realistic simulation scenarios.

5.1.2 Technological feasibility

The validation trials presented in this TVALR all use an VAR device based on COTS developments, consisting of Head Mounted display, used in different versions, that have been each time configured to enable specific features according to each exercise objectives.
The concept of Virtual and Augmented Reality in Tower environment has been proven technically feasible in all the validations which have addressed it. No operational showstoppers have been identified during laboratory tests related to the use of Tracking Labels, Attention Guidance or Air Gestures in VAR environment. The exercises took into consideration both virtual environment with simulated traffic and physical tower environment with real airport traffic.

The configurations allowed to have tracking labels correctly displayed, collimated to the OTW view, correctly tracked to associated flight and visible. Smoothness of traffic data is also important and must be ensured through an adequate rate of update or interpolation algorithm.

The VAR technology has introduced new functional blocks in EATMA, such as Virtual and Augmented Reality Display, Air Gestures Detector, Attention Guidance. All implementations rely on the Virtual and Augmented Reality Display. Only EXE-001 addresses the Attention Guidance.

On ergonomic perspective, the technology was deemed acceptable particularly for last generation devices, while the previous models could lead to experience some heavy head. Furthermore, as the HoloLens limits the augmented visual range, ATCO can feel overstimulated to move their head in order to see the augmented information or they can find difficult to fill in the paper strips.

The visualisation of augmented information displayed on HoloLens can improve in the real environment where the outside view has more contrast, so that HoloLens images do not appear too bright on top of the background.

5.1.3 Performance assessments

EUROPEAN PARTNERSHIP

Co-funded by the European Union

The validation exercises tried to show that controllers' workload could be reduced thanks to Virtual and Augmented Reality technology. Reduced workload was expected to result in increased controller productivity. Performance was assessed by means of a set of measurable quantity to be evaluated. On this respect, PJ.19 has defined quantitative validation targets for solution 97.1, hereafter summarized:

SOL. CODE	SAF	FEEF1	TEFF1	CAP3	CAP1	CAP2	PRD1	PUN1	CEF2	CEF3	HP
PJ.05-W2-97.1	ISI	N/I	N/I	N/I	N/I	N/I	N/I	N/I	1	N/I	YES

Actually, while no reduction between the Solution and Baseline scenarios was found in terms of **Workload** (still maintained at an acceptable level, see appendices related to EXE001,002,005), the influence of VAR technology on **Situational Awareness** was rated 'positive' to 'very positive', both during normal operations and in case of an alert, stating that the V/A-R HMI can be, in the majority of the cases, very supportive for ATCOs in increasing or at least maintaining an adequate level. Thanks to the VAR technology, the Situational Awareness could be built faster and easily maintained, being the information (traffic, weather, conflicts etc.) displayed to the controllers on head up view and avoiding the continuous need to switch from head up to head down attitude. Anyhow, it is true that (potential) degradations in SA can be due to the information representation such as overlaps or missing information; thus, the feedback is strongly related to the quality and quantity of the presented information.

As a result, **Cost Efficiency** performance could be further improved by increasing the system's positive impact on situation awareness and workload. This can be achieved e.g. improving the label design.

The ATCOs rated the influence of the A-R on **safety** as 'neutral' to 'very positive': V/AR with safety nets improves the perceived safety performance by reducing human error. On the contrary, ATCO's workload and situational awareness with the implementation of Virtual/Augmented Reality Air Gesture application was not maintained at acceptable level, therefore potentially reducing safety levels.

The majority of ATCO involved in the simulations, provide positive feedback on job satisfaction and **acceptance** for what concerns Tracking Labels and Air Gesture. Concerning the Attention Guidance, although the concept is highly acceptable, its implementation still needs to be further improved to be acceptable, specifically alert design needs improvement and nuisance warnings (taxi conflicts in this case) had to be solved.

The V/A-R Air Gestures negatively impacted workload and SA as ATCOs had difficulties using them. The system should be further developed so that it recognises air gestures better. Potentially, also training and familiarisation will improve ATCOs performance using air gestures.

5.1.4 Workload Assessment

As also shown in section 4.3.4, the assessment of mental workload is the main basis to carry out an evaluation of benefits generated by the solution and its enabler.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

From an analysis of the results shown in the Appendix, which can be referred to for detailed information, the outcome is a positive feasibility, both qualitative and quantitative, for Workload reduction.

5.2 Recommendations for Solution 97.1

Though the potential and feasibility of VAR solution has been demonstrated, some technical recommendations have been figured out to further improve the usability of the technology itself and associated performance.

5.2.1 Next phase

The presentation of the information was deemed satisfactory, with some mentions of improvements for future phases:

- HMI: position, width, brightness… of symbols should be refined in order to avoid visual interference;
- The addition of an altitude filter to allow the controller to filter out a/c that are either flyovers or outside the scope of their control
- Choice of the device: the latest generations devices are preferable due to lower weight and a wider angle of view, thus improving the experience comfort;
- It was found that controllers thought it would be enough to alert them only once for serious events, such as a runway incursion or a go-around. After acknowledgement via focussing on the area of interest, they would only need guidance from that point on (e.g. location of conflict, label information). Monitoring the actions of controllers to repeat alerts was not appreciated. For future work this means that we have to look into the question. Whether the nuisance was perceived because of the time values used, or whether a repeating alerts would make sense in other conditions, such as Alerts with several severity levels (repeat alert if a new severity level is reached and the controller does not pay attention to the area of interest) or Simple warnings of high traffic intensity in certain areas of the airport (with less intrusive symbols or aural alerts).
- No distinction between different controller roles was made (e.g. runway controller, ground controller, assistant, supervisor), while in fact both roles may require another, more customized way of presenting the necessary information.
- Other static or dynamic information on the airport surface could be presented, such as buildings, and taxiway and runway edges (in reduced visibility), stop bars and their statuses, protected areas, closed runways etc.
- Automatic Speech Recognition could be used in the future to identify certain situations in the system (e.g. a pilot calling) and signalling to the AR device to highlight particular information (e.g. aircraft label).
- Strip-less working methods could be investigated adding planning aspects to the outside view, making it superfluous to build a mental picture with flight status strips.

- Use of the technology could also lead to a new definition of controller roles and responsibilities, where the AR logic determines (or is fed with) the sequence of operations and the course of actions that need to be carried out by a particular individual in the tower. Obviously, such novel arrangements would require a high degree of automation and a clear delegation of authority, particularly in system failure situations.
- Additional features could be integrated into the AR device view, such as video streams from cameras at gate positions that cannot be seen very well from the tower or video that zooms in on certain aspects of the operation at the gate to give an indication of the statuses for boarding and de-boarding, fuelling, catering and baggage handling.
- For some areas, it might be useful to offer detailed (camera) views inside the device, e.g. for runways where thresholds are far away from the tower or where part of the runway cannot fully be seen (gap fillers).
- For attention capturing and guidance mechanisms (without an AR device), there could be advantages when used in multiple remote tower set-ups, where one or more controllers need to maintain a mental picture of the operational situation at two different airports.

5.2.2 Updating ATM Master Plan Level 2

Solution 97.1 is currently defined as follows:

	ID	Title	Description
Solution	97.1	Virtual/Augmented Reality applications for tower	This Technological Solution aims to support the Air Traffic Controllers by means of Virtual and Augmented Reality application in Tower Environment. The technology involves the use of Tracking Labels, Air Gestures and Attention Guidance. These applications are enabled by devices like Head-Mounted See-Through display, that allow • to visualise equivalent out-of-the-window view to good visibility even in LVC, • to augment the out of the window view by tracking labels, • to provide interaction with V/A-R interface by air gestures and • to guide controller's attention to critical ATC situations. The need to switch from head up to head down and vice versa is expected to decrease, with benefits on ATCO productivity and situation awareness.

OI Step	POI-0039-SDM	Virtual/Augmented Reality, attention guidance and air gesture for tower controllers	Use of V/A-R technologies to present head up visual information to tower controllers when watching aircraft in landing, taxing to/from the gate/stand and take-off, to present equivalent out of the window view in low visibility conditions, to interact with V/A-R interface by air gestures and to guide controller's attention to critical ATC situations. These are expected to improve ATCO productivity and situation awareness.
Enablers	AERODROME-ATC-103	Virtual and Augmented Reality systems for Tower ATC	Introduction of new Augmented Reality vision systems superimposed onto the out of the tower view stimulate the ATCO to work in head-up position resulting in an improvement of the controller situational awareness and productivity in any visibility conditions.
	AERODROME-ATC-104	Controller productivity enhancements by Air gestures for Tower ATC	Air Gesture interactions with V/A-R interface will reduce the need for head-down, resulting in improvement of ATCO human performance.
	AERODROME-ATC-105	Attention Guidance in V/AR applications for aerodrome tower operations	Introduction of new automated functions for attention guidance in V/AR applications for improving situational awareness of aerodrome tower controllers.

Table 33: Solution 97.1 definition

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

As an outcome of the Technical Validation Exercises, the following Functions have been created in EATMA to be linked to AERODROME-ATC-104:

- Detect Information Menu Navigation Request
- Detect Navigation through Menus
- Detect Information Selection

The link between them and AERODROME-ATC-104 has been documented through the appropriate Change Request.

5.2.3 Regulatory and standardisation initiatives

After the execution of the different PJ.05-W2-97.1 Technical Validation Exercises, a communication of the findings and results to EUROCAE Technical Advisory Committee has been carried out, in order to identify the impact on Regulatory and Standardisation initiatives. As a result, the following items have been agreed as relevant with regards to the potential impact:

- EUROCAE ED-87E "MASPS for A-SMGCS including Airport Safety Support Service Routing Service and Guidance Service" should be considered for the Virtual and Augmented Reality functionalities, e.g. for the identification and alignment of elements in the V/AR devices.
- EUROCAE ED 255 "MASPS for a combined vision systems for rotorcraft operations" may contain relevant information for the safety critical requirements to comply with.
- The usage of COTS products in safety critical operations could imply some risks. It is recommended that a specific prototype for ATM is developed, or the existing COTS products are improved in a way that comply with the safety requirements for this environment.

These findings have been fed into the proper section in the PJ.05-W2-97.1 TS/IRS Part I [27].

6 Conclusions and recommendations for 97.2

6.1 Conclusions

Solution 97.2 is a promising step in the direction of introducing novel human machine interface methods in the Tower Ground environments, with the use of Assistant Based Speech Recognition technology. Some considerations follow regarding the outcome of the Validation Exercises which make up this SESAR Solution.

6.1.1 Maturity

This Solution moves from the achievements of 16.04.02 which partly achieved TRL4. This report contains the validation results of three different exercises to demonstrate that the ASR activity in Sol 97.2 has fully/partly achieved TRL 4.

Although most of the items for achieving TRL4 are met, some items need to be reconsidered in Wave-2 starting in 2020. 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 J.

6.1.2 Technological feasibility

The ASR technology (incl. capture of Aerodrome ATC instructions and input of clearances into the ATC system) has shown to be feasible in an ATC tower 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).

	Command recognition rate	Command recognition error rate	Callsign recognition rate	Callsign recognition error rate
EXE-004	76.0%	-	81.2%	7.8%
EXE-006	91.4%	4.5%	98.4%	0.9%
EXE-007	64.6%	5.1%	89.8%	10.2% (error + rejection rate)

Table 34: Solution 97.2 ASR Rates

The ASR supported by AI and Machine Learning has introduced Automatic Speech Recognition as a new functional block in EATMA linked to the functions "Command Prediction" and "Concept Extraction".
Four Use Cases have been defined to validate the concept:

- Highlighting of recognised callsign,
- Showing full recognised command in HMI,
- Manual manipulation of an ASR output,
- (Automatic) acceptance of ASR output,

and all of them have been addressed by each validation exercise.

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

However, a list of recommendations to enhance the ASR system (in testing environment) have been made. The quantitative and qualitative feedback of ATCOs was good and motivating to go beyond TRL4 and would have been even better if the full potential of ABSR accuracy have been offered to them.

6.1.3 Performance assessments

Performance was assessed with some main measurable quantity to be evaluated, mainly Human Performance qualitative measures.

Mean Solution workload was found to be within acceptable levels.

As also shown in section 4.3.4, the assessment of mental workload is the main basis to carry out an evaluation of benefits generated by the solution and its enabler.

From an analysis of the results shown in the Appendix, which can be referred to for detailed information, the outcome is a positive feasibility, both qualitative and quantitative, for Workload reduction.

6.2 Recommendations

Recommendations should be focused on what it will eventually take in order to reach TRL6 in a possible next phase of the project. The transition should be feasible, but a certain number of steps must be taken in order to take all the validation platforms to the next phase in terms of readiness level, overcoming shortcomings which could be acceptable for TRL4 validations.

6.2.1 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 ABSR experience and functionality for aircraft cockpits
- Consider further applications that use the speech recognition and understanding output such as pre-filling of radar labels and flight strips, 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 ABSR input and output content as well as format in order to improve system interoperability and comparability

6.2.2 Updating ATM Master Plan Level 2

Solution 97.2 is currently defined as follows:

	ID	Title	Description
Solution	97.2	Automatic Speech Recognition at the TWR CWP supported by AI and Machine Learning	This technological solution aims to support the Tower Controllers by means of Automatic Speech Recognition supported by AI and Machine Learning algorithms, to improve usability and task efficiency.
OI Step	POI-0040-SDM	Automatic Speech Recognition with AI/ML at the TWR CWP	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	AERODROME-ATC-106	Automatic Speech Recognition supported by AI and ML algorithms for aerodrome tower operations	Introduction of new automated functions for Automatic Speech Recognition using AI and Machine Learning Techniques at the Aerodrome CWP/HMI for improving the controller workload.

Table 35: Solution 97.2 definition

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

6.2.3 Regulatory and standardisation initiatives

After the execution of the different PJ.05-W2-97.2 Technical Validation Exercises, a communication of the findings and results to EUROCAE Technical Advisory Committee has been carried out, in order to identify the impact on Regulatory and Standardisation initiatives. As a result, the following items have been agreed as relevant with regards to the potential impact for ASR:

- There is not an existing standard for Voice Recognition in the ATM environment. An assessment of this need should be performed in further stages of developments.

- Outside the ATM environment, there is an existing standard that is relevant for the solution: ISO/IEC 30122-2:2017Fehler! Verweisquelle konnte nicht gefunden werden., which provides the technical criterions and test methods of voice commands and its speech recognition engine. It is recommended that this standard is taken into account when developing the ASR functionality for ATM.

- Proposals for standardisation of the content and the format for input and output of assistant based speech recognition systems should be identified, i.e., 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 runway error detection, formats such as JSON for content transmission, and many aspects more to enable comparability and interoperability.

- The usage of commercial-off-the-shelf products are not feasible for the ATM environment. Therefore, dedicated products developed for this environment would match better the expectations and requirements for deploying the concept in ATM.

These findings have been fed into the proper section in the PJ.05-W2-97.2 TS/IRS Part I [27].

6.2.4 Workload Assessment

As also shown in section 4.3.4, the assessment of mental workload is the main basis to carry out an evaluation of benefits generated by the solution and its enabler.

From an analysis of the results shown in the Appendices, which can be referred to for detailed information, the outcome is a positive feasibility, both qualitative and quantitative, for Workload reduction.

7 References

7.1 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

System and Service Development

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 18

System Engineering

[6] E-OCVM Version 3, February 2010

[7] SESAR 2020 Requirements and Validation Guidelines, Ed. 00.01.01 ,20 November 2017

Safety

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

Human Performance

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

Environment Assessment

Security

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

Programme Management

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

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

7.2 Reference documents

[13] ED-78A GUIDELINES FOR APPROVAL OF THE PROVISION AND USE OF AIR TRAFFIC SERVICES SUPPORTED BY DATA COMMUNICATIONS.[2]

[14] SESAR 2020 Requirements and Validation Guidelines, Ed. 00.01.01 20 November 2017

[15] RETINA D4.2 Operational Concepts Description Update, Ed. 00.00.02. 15 January 2018

[16] RETINA D6.2 Project Conclusions, Ed. 00.01.02. 30 August 2018

[17] D3_2_020-SESAR 2020 PJ_16-04 TRL4 TVALR-ASR, Ed. 02.00.00 30 September 2019

[18] D3_2_020-SESAR 2020 PJ_16-04 TRL4 TVALR-ASR -Appendix G, Ed. 02.00.00, 30 September 2019

[19] D3_2_020-SESAR 2020 PJ_16-04 TRL4 TVALR-MTI, Ed. 02.00.00 30 September 2019

[20] D3_2_020-SESAR 2020 PJ_16-04 TRL4 TVALR-MTI -Appendix E, Ed. 02.00.00 30 September 2019

[21] D3_2_020-SESAR 2020 PJ_16-04 TRL4 TVALR-AG, Ed. 02.00.00 30 September 2019

[22] D3_2_020-SESAR 2020 PJ_16-04 TRL4 TVALR-AG - Appendix C, Ed. 02.00.00 30 September 2019

[23] D2 SESAR 2020_PJ_16 FINAL PROJECT REPORT, Ed. 01.01.00, 13 December 2019

[24] D1_2 SESAR 2020_PJ_03b FINAL PROJECT REPORT, Ed 02.00.00, 15 November 2019

[25] MALORCA D5-3 Final Project Results Report Ed 2.00, 30 April 2018

[26] MALORCA D1-1 Operational Concept Document Ed. 2.00 14 September 2016

[27] PJ_05-W2 SESAR Solution 97 TS_IRS - Part I (1_19)

[28] SESAR 2020 - PJ05-W2 Sol 97 D3.1.033 - Technical Validation Plan (TVALP) Final version

[29] SESAR 2020 – PJ05-W2 Sol 97 TVALP Part II – Safety Assessment Plan (06_3)

2

[30] SESAR 2020 – PJ05-W2 Sol 97 TVALP Part III – Security Assessment Plan (1_6)

[31] SESAR 2020 – PJ05-W2 Sol 97 TVALP Part IV – Human Performance Assessment Plan (1_0)

[32] SESAR 2020 EXE-05.97.1-TRL4-TVALP-VAR-001 Availability Note (D3.1.041), V00.01.00, 21 April 2021

[33] SESAR 2020 EXE-05.97.1-TRL4-TVALP-ASR-004 Availability Note (D3.1.044), V00.00.01, 19 November 2021

[34] SESAR 2020 EXE-05.97.1-TRL4-TVALP-VAR-002 Availability Note (D3.1.042), V00.01.00, 16 February 2022

[35] SESAR 2020 EXE-05.97.1-TRL4-TVALP-ASR-006 Availability Note (D3.1.046), V01.00.00, 26 January 2022

[36] SESAR 2020 EXE-05.97.1-TRL4-TVALP-VAR-005 Availability Note (D3.1.045), V00.01.00, 04 February 2022

[37] SESAR 2020 EXE-05.97.1-TRL4-TVALP-ASR-007 Availability Note (D3.1.047), V01.00.01, 09 May 2022

[38] SESAR ATC Ontology CommandTypeValues, V01.05.01, 04 Feb 2022

Appendix A Technological Validation Exercise 001 Report

A.1 Summary of EXE-001 plan

The following represents the technological validation exercise report for EXE-05.97.1-TRL4-TVALP-VAR-001 (EXE-001 for short). There were no major deviations from the validation plan as described in [5]. Other than reported in that deliverable, though, attention distribution was not measured with the built-in eye tracker. For the system to notice that the attention was captured, it was sufficient to detect that the tower controller looked at the area of interest (as measured by the A-R device in terms of head movement). Furthermore, while controller reaction times were measured in the situation wearing the A-R device, it was not possible to do so in the reference scenario. Instead, controllers were asked about the perceived difference in reaction times, which means that the measurements were subjective.

A.2 EXE-001 description and scope

EXE-001 was carried out by Royal NLR and investigated A-R applications for a conventional tower environment of Schiphol Airport (EHAM) with a special focus on attention getting and attention guidance. To that end different types of alert that are currently given in the Schiphol tower environment were generated in an A-R device (HoloLens) and the differences with traditional tower control were evaluated.

The relevant use cases for the exercise (see also Figure A-1) were defined in the TS/IRS document [27]. They were UC-97-TRL4-TS-101 and UC-97-TRL4-TS-102, which can be described as:

Guiding ATCO attention via perceptual cues...

- ...in case of potentially critical ATC situations
- ...in case of potentially missed command actions

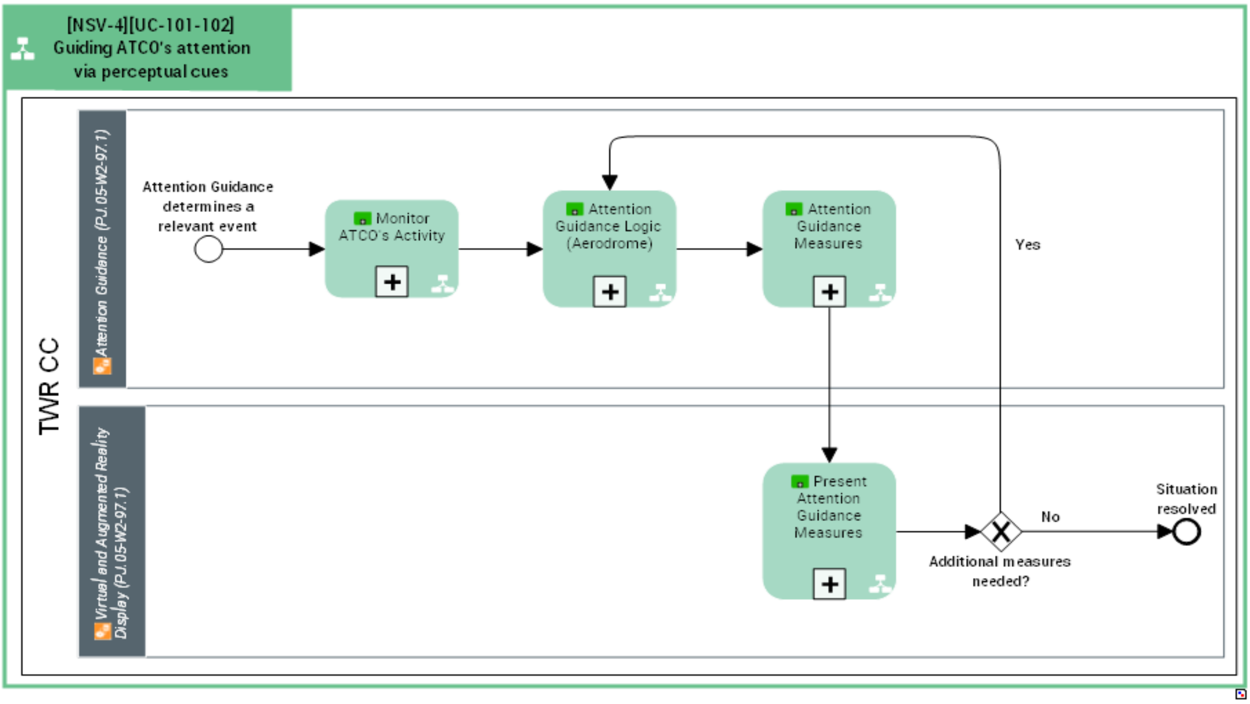


Figure A-1: Relevant use cases for EXE-001

While use case UC-97-TRL4-TS-105 for the use of Tracking Labels to show conflict detection alerts was considered as well, the main focus was on the presentation of the relevant alert information for attention guidance purposes, rather than a standardized integration of a specific alerting element in the tracking label. Tracking labels in this exercise were also reduced to show callsign information only, although the technology would certainly allow for more information elements to be displayed.

The optimal way to guide the attention of air traffic controllers was thus assessed by exposing them to different presentations of alert information, symbology and audio alerts within the A-R device. The A-R device was used in different scenarios with different traffic situations, different types of alerting with different levels of severity at selected locations in the airport movement areas.

The main focus of the exercise was to contribute to the goal for A-R device applications to reach TRL-4 maturity and show improvements in attention getting and guidance. This required the collection of information regarding the impact on Human Performance, Safety and Capacity (airport resilience) for the mentioned improvement areas.

The exercise was performed as a real-time simulation on the SESAR IBP NARSIM Tower, used in earlier SESAR validation activities for airport guidance and safety nets by consortium LVNL in co-operation with ENAV. The A-R device was a HoloLens 2^{TM} (from Microsoft) that was prepared for use within the NARSIM Tower environment with its large 360-degree projection screen (with a field-of-view of 40 degrees vertically and a diameter of 11 metres). The real-time simulation was carried out as a Human-in-the-Loop activity and with realistic ground movement scenarios of Schiphol traffic.]

A.3 Summary of EXE-001 objectives and success criteria

The following is a summary of the table of validation objectives that can be found in [28] (Section 5.1.3) and briefly explains which objectives were addressed by EXE-001. Exercise validation objectives and success criteria are generally identical to the overall objectives and success criteria for Solution 97.1

(V/A-R), but focus on the Attention Guidance prototype that was tested. The other exercises in Solution 97.1 will complete the picture and look at other relevant aspects of usage of A-R devices in the conventional tower. Should validation objectives and success criteria for EXE-001 differ, it will briefly be indicated.

SESAR Solution Validation Objective	SESAR Solution Success Criteria	Exercise Validation Objective	Exercise Success Criteria
OBJ-05.971-TRL4-TVALP-FEAS.1010 Operational Feasibility	CRT-05.971-TLR4-TVALP-FEAS-1011 No showstoppers for use of V/A-R and tracking labels.	EX1-OBJ-05.971-TRL4-TVALP-FEAS.1010 Restricted to Use Case for Attention Guidance with A-R device.	EX1-CRT-05.971-TLR4-TVALP-FEAS-1011 Idem.
	CRT-05.971-TLR4-TVALP-FEAS-1013 No showstoppers for Attention Guidance.		EX1-CRT-05.971-TLR4-TVALP-FEAS-1013 Idem.
OBJ-05.971-TRL4-TVALP-FEAS.1020 Technical Feasibility	CRT-05.971-TLR4-TVALP-FEAS-1021 Technical feasibility for use of V/A-R applications in the tower environment is verified.	EX1-OBJ-05.971-TRL4-TVALP-FEAS.1020 Restricted to Attention Guidance with A-R device. The link with the relevant safety nets will be considered.	EX1-CRT-05.971-TLR4-TVALP-FEAS-1021 Idem.
	CRT-05.971-TLR4-TVALP- FEAS -1022 Technical feasibility for integration of the V/A-R applications with other related system enablers is verified.		EX1-CRT-05.971-TLR4-TVALP- FEAS-1022 Idem.
OBJ-05.971-TRL4-TVALP-H105.1010 ATCO Task Support	CRT-05.971-TLR4-TVALP-H105.1011 Workload maintained at acceptable level.	EX1-OBJ-05.971-TRL4-TVALP-H105.1010 Restricted to Attention Guidance with A-R device. Impact on human performance is performed and documented.	EX1-CRT-05.971-TLR4-TVALP-H105.1011 Idem.
	CRT-05.971-TLR4-TVALP-H105.1012 Information provided (level and quality) is adequate.		EX1-CRT-05.971-TLR4-TVALP-H105.1012 Idem.
	CRT-05.971-TLR4-TVALP-H105.1013		EX1-CRT-05.971-TLR4-TVALP-H105.1013 Idem.

	SA maintained at adequate level.		
	CRT-05.971-TLR4-TVALP-H105.1014 Head up time is increased with respect to reference.		EX1-CRT-05.971-TLR4-TVALP-H105.1014 Only good weather conditions are assumed (and thus qualitative assessment only).
	CRT-05.971-TLR4-TVALP-H105.1015 HMI blocking of OTW view.		EX1-CRT-05.971-TLR4-TVALP-H105.1015 Idem.
	CRT-05.971-TLR4-TVALP-H105.1016 HMI does not increase potential for human error.		EX1-CRT-05.971-TLR4-TVALP-H105.1016 Idem.
	CRT-05.971-TLR4-TVALP-H105.1017 ATCO system trust at acceptable level.		EX1-CRT-05.971-TLR4-TVALP-H105.1017 Idem.
	CRT-05.971-TLR4-TVALP-H105.1018 Adequate level of HMI usability.		EX1-CRT-05.971-TLR4-TVALP-H105.1018 Idem.
	CRT-05.971-TLR4-TVALP-H105.1019 Alarms and alerts not too intrusive.		EX1-CRT-05.971-TLR4-TVALP-H105.1019 Not an assessment of alarms and alerts given, but of presentation.
	CRT-05.971-TLR4-TVALP-H105.1020 Positive feedback on acceptance.		EX1-CRT-05.971-TLR4-TVALP-H105.1020 Idem.
	CRT-05.971-TLR4-TVALP-H105.1021 ATCO team maintains acceptable level of SA.		EX1-CRT-05.971-TLR4-TVALP-H105.1021 Idem.

OBJ-05.971A-TLR4-TVALP- H105.1030 Consistency with Human Capabilities and Limitations	CRT-05.971-TLR4-TVALP-H105.1031 Operating methods applied accurately, efficiently and in timely manner.	EX1-OBJ-05.971-TRL4-TVALP-H105.1030 Restricted to Attention Guidance with A-R device. An assessment of the consistency of the roles of ATCOs with human capabilities and limitations is carried out.	EX1-CRT-05.971-TLR4-TVALP- H105.1031 Idem.
	CRT-05.971-TLR4-TVALP-H105.1032 Operating methods clearly identified and consistent in all operating conditions.		EX1-CRT-05.971-TLR4-TVALP-H105.1032 Idem.
OBJ-05.971-TRL4-TVALP- H105.1040 Job Acceptance and Satisfaction	CRT-05.971-TLR4-TVALP-H105.1041 Positive feedback on job satisfaction and acceptance.	EX1-OBJ-05.971-TRL4-TVALP-H105.1040 Restricted to Attention Guidance with A-R device.	EX1-CRT-05.971-TLR4-TVALP-H105.1041 Idem.
OBJ-05.971-TLR4-TVALP-SAFE.1010 Safety Impact	CRT-05.971-TLR4-TVALP-SAFE-1011 Improvement of safety performance by reducing human error.	EX1-OBJ-05.971-TLR4-TVALP-SAFE.1010 Restricted to Attention Guidance with A-R device. Relevant input about safety issues is gathered.	EX1-CRT-05.971-TLR4-TVALP-SAFE-1011 Changes related to implementation of A-R applications do not increase potential for human error.
	CRT-05.971-TLR4-TVALP-SAFE-1012 Improvement of safety performance by reducing ATCO workload.		EX1-CRT-05.971-TLR4-TVALP-SAFE-1012 ATCO workload is shown to be maintained at acceptable level.
	CRT-05.971-TLR4-TVALP-SAFE-1013 Improvement of safety performance by increasing SA.		EX1-CRT-05.971-TLR4-TVALP-SAFE-1013 ATCO SA is shown to be maintained at acceptable level.
	CRT-05.971-TLR4-TVALP-SAFE-1014 Safety assessment activities and results are documented and integrated in overall solution validation results.		EX1-CRT-05.971-TLR4-TVALP-SAFE-1014 Idem.

OBJ-05.971-TLR4-TVALP-PERF.1010 Performance Benefits	CRT-05.971-TLR4-TVALP-PERF-1011 Improvement of Cost Efficiency by reducing cost per flight.	EX1-OBJ-05.971-TLR4-TVALP-PERF.1010 Restricted to Attention Guidance with A-R device. To Relevant input to an assessment of performance benefits is gathered.	EX1-CRT-05.971-TLR4-TVALP-PERF-1011 Contributions to the assessment are made by investigating positive impact on SA, workload and efficiency of ground operation.
	CRT-05.971-TLR4-TVALP-PERF-1012 Improvement of Resilience by increasing SA in LVC while maintaining workload within acceptable limits.		N/A: the chosen focus is restricted to good visibility operations. Non-nominal alarms and alerts will be applicable in all visibility conditions.

Table A-1: Summary of Validation Objectives addressed in EXE-001

A.4 Summary of EXE-001 validation scenarios

A.4.1 Reference scenarios

In the reference scenario, ATCOs were working with traffic that was comparable to the traffic in the solution scenario, but they were not using the A-R device and the symbology that was developed for attention capturing and guidance. Alerts given by the A-R device were shown on the Traffic Situation Display instead, in about the same way as alerts which are currently presented to controllers in the Schiphol Tower. What was different from the current working procedures at Schiphol was the fact that paper strips were used instead of an EFS system which has been available for about two years now. This choice was made because the introduction of the EFS in the simulation would have led to a more complex set-up and former controllers involved in the simulations would have needed special training in using EFS. Since the set-up with one measured tower controller was already reduced, the introduction of an EFS system would consequently have been a disproportionate addition of realism and complexity.

The reference scenarios could therefore be run for comparison between ATCO-behaviour and performance with the solution scenario because the only difference was the presence of the A-R device. Thus, if performance was influenced in either a positive or negative way, this could directly be attributed to the solution offered for Attention Guidance and the impact of wearing an A-R device (HoloLens).

A.4.2 Solution scenarios

In the solution scenario, the tower controller was confronted with a busy traffic situation. That situation would require most of the attention of the controller. In the beginning, even though training runs had been carried out, the ATCO would still need some time to familiarise with the A-R device and the traffic situation. This meant that traffic scenarios were already running for several minutes (mostly more than 10 minutes) before a first alert situation could be elicited.

After a while, the ATCO would be fully engaged in controlling the traffic and talking to the pilots. At that point, an event that required the immediate attention of the ATCO was created. The generated events were go-around situations, runway incursions with entering or crossing of a runway without permission, or potential taxiway conflicts, which were considered less urgent alerts and were therefore mostly used to cause nuisance or distractions. The runway incursion and go-around alerts were based on the existing Schiphol alerting systems RIAS and GARDS. The taxiway conflict alerting is not currently present at Schiphol and was based on a prototype developed by NLR, tested earlier in the same environment as part of SESAR project PJ03B.

A.5 Summary of EXE-001 assumptions

The main assumption for EXE-001 concerned the weather conditions foreseen. Only good weather conditions with no impact on visibility were considered. This means that a possible impact on thresholds caused by visibility (bad weather or night time conditions) was not investigated. The focus in this exercise was thus put on traffic stream complexity and high levels of traffic. The limited scope of the simulation obviously had an impact on the efficiency results, as it was not possible to obtain them for a complete airport operation. As this was not the focus of EXE-001 though, the impact on the assessment itself was rather limited.

Identifier	Title	Description	Justification	Impact on Assessment
AS-EXE.001-01	Weather conditions	Good weather conditions throughout the simulation	No impact	Low
AS-EXE.001-02	Limited Simulation Scope	Simulation focused on the work of one tower controller in changing roles (runway or ground control)	The choice of having one controller carrying out the tower operation had a limited impact on the operation itself, but a rather large impact on the perceived realism of the events that the controller was exposed to.	Medium

Table A-2: Technological Validation assumptions overview

A.6 Deviations from planned activities

Prototype development as well as development of specific NARSIM validation scenarios for Schiphol Airport were already described in [32]. With regard to that plan, a few changes were made after the first day of experiments and discussions with ATCOs.

For prototype development, one of the features of attention capturing would be an update event in the case that the controller disregards an alert or does not focus on the area of attention. In [32] this is described as follows:

An update of the event, in accordance with Use Cases UC-97-TRL4-TS-101/102 will occur when…

a) …the tower controller acknowledges an attention cue but then disregards the area of attention (after a given time interval) and the situation persists.

b) …the tower controller acknowledges the attention cue, keeps focussing on the area of attention and the seriousness of the situation (e.g. the severity of a conflict alert) increases.

c) …the tower controller does not acknowledge the cue within a given time period.

Feedback after the first day of experiments showed that c) would only occur if nuisance alerts for potential taxiway conflicts were disregarded due to more serious events and would then lead to even more nuisance. Based on that feedback, it was decided to remove that kind of alert for taxiway conflicts and not run scenarios that would lead to such events. The primary reason for this was that it would distract controllers from the actual goal of validating Attention Guidance in the case of serious events such as go-arounds and runway incursions. Furthermore, taxiway conflicts would usually be handled by a separate ground controller at Schiphol. Since only one measured controller role was available and that controller already had control of the runway, additional nuisance caused by apron or taxiway events farther away from the runway would have been less realistic. This caused a reduced number of runs.

While several other comments on the concept were made, it was mainly this change that was eventually introduced in order to reduce confusion about the concept and reduce complexity in the number of runs to be performed. All other comments regarding concept elements were taken as feedback on the concept, but did not subsequently lead to a change in the set-up, as the specific situations noted would not occur often enough to cause further confusion.

Another deviation from the original plan concerned the measurement of head up time and reaction times. While both could be measured in the solutions runs, it was not possible to find a way to make objective measurements in the reference runs. Head up time was never considered to be measured in this exercise anyway, because one of the assumptions was the presence of good weather conditions in the simulations. This had already been indicated in Section A.5. The same, however, would account for reaction times to the events, as the reference situation did not allow to measure exactly when the controller would have noted the event. The reason for this lies in the fact that the solution more or less replaces the indication on the radar screen with the one in the outside view. Thus, while it was possible to measure a reaction time (implicit acknowledgement by focussing on the area of attention) in the solution, the reference situation would have required measurements with an eye tracker to discover attention focus on the alerts and on the area of focus subsequently. Even then, it would have been difficult to exactly determine a comparable reaction time, as the focus areas would not have been exactly the same for the solution and the reference. This would have meant that the eye tracker had to be configured in the same way as the HoloLens, thereby causing the same kind of intrusive situation. This consideration, in conjunction with the needed additional complexity and effort for setting-up eye-tracking, led to the decision to only gather subjective results.

Prototype development as well as development of specific NARSIM validation scenarios for Schiphol Airport were already described in Ref. With regard to that plan, a few changes were made after the first day of experiments and discussions with ATCOs.

For prototype development, one of the features of attention capturing would be an update event in the case that the controller disregards an alert or does not focus on the area of attention. In [32] this is described as follows:

An update of the event, in accordance with Use Cases UC-97-TRL4-TS-101/102 will occur when…

a) …the tower controller acknowledges an attention cue but then disregards the area of attention (after a given time interval) and the situation persists.

b) …the tower controller acknowledges the attention cue, keeps focussing on the area of attention and the seriousness of the situation (e.g. the severity of a conflict alert) increases.

c) …the tower controller does not acknowledge the cue within a given time period.

Feedback after the first day of experiments showed that c) would only occur if nuisance alerts for potential taxiway conflicts were disregarded due to more serious events and would then lead to even more nuisance. Based on that feedback, it was decided to remove that kind of alert for taxiway conflicts and not run scenarios that would lead to such events. The primary reason for this was that it would distract controllers from the actual goal of validating Attention Guidance in the case of serious events such as go-arounds and runway incursions. Furthermore, taxiway conflicts would usually be handled by a separate ground controller at Schiphol. Since only one measured controller role was available and that controller already had control of the runway, additional nuisance caused by apron or taxiway events farther away from the runway would have been less realistic. This caused a reduced number of runs.

While several other comments on the concept were made, it was mainly this change that was eventually introduced in order to reduce confusion about the concept and reduce complexity in the number of runs to be performed. All other comments regarding concept elements were taken as feedback on the concept, but did not subsequently lead to a change in the set-up, as the specific situations noted would not occur often enough to cause further confusion.

Another deviation from the original plan concerned the measurement of head up time and reaction times. While both could be measured in the solutions runs, it was not possible to find a way to make objective measurements in the reference runs. Head up time was never considered to be measured in this exercise anyway, because one of the assumptions was the presence of good weather conditions in the simulations. This had already been indicated in Section A.5. The same, however, would account for reaction times to the events, as the reference situation did not allow to measure exactly when the controller would have noted the event. The reason for this lies in the fact that the solution more or less replaces the indication on the radar screen with the one in the outside view. Thus, while it was possible to measure a reaction time (implicit acknowledgement by focussing on the area of attention) in the solution, the reference situation would have required measurements with an eye tracker to discover attention focus on the alerts and on the area of focus subsequently. Even then, it would have been difficult to exactly determine a comparable reaction time, as the focus areas would not have been exactly the same for the solution and the reference. This would have meant that the eye tracker had to be configured in the same way as the HoloLens, thereby causing the same kind of intrusive situation. This consideration, in conjunction with the needed additional complexity and effort for setting-up eye-tracking, led to the decision to only gather subjective results.

A.7 EXE-001 validation results

A.7.1 Summary of Technological Validation Exercise EXE-001 results

The A-R symbology for attention guidance in case of a runway incursion, a go-around or a taxi conflict was presented in a HoloLens. An experiment took place in the NARSIM Tower simulator with two air traffic controllers. Controllers found the A-R guidance very effective and efficient for the runway incursion and go-around detection. As the labels (call signs) of the involved a/c were in view, there was no need to look down to consult the flight strips. The experiment results yielded several potential improvements to the design of the guidance and the controllers experienced some negative symptoms resulting from the hardware that need to be fixed before the concept can be introduced.

Technological Validation Exercise #001 Validation Objective ID and Title	Technological Validation Exercise #001 Success Criterion ID and Title	Technological Validation Exercise #001 Results	Technological Validation Exercise #001 Validation Objective Status
EX1-OBJ-05.971-TRL4-TVALP-FEAS.1010 Operational Feasibility for Attention Guidance	EX1-CRT-05.971-TLR4-TVALP-FEAS-1011 No showstoppers for use of A-R and tracking labels.	Controllers were enthusiastic about the potential of the AI guidance, especially the tracking labels, but experienced a heavy head by the end of the day as a consequence of wearing the hardware.	OK
	EX1-CRT-05.971-TLR4-TVALP-FEAS-1013 No showstoppers for Attention Guidance.	The experiment confirmed that the concept is operationally feasible when addressing the Use Case for Attention Guidance with an A-R device. Several suggestions on how to improve elements of the chosen concept (for symbology and timing of attention guidance cues) were given.	OK
EX1-OBJ-05.971-TRL4-TVALP-FEAS.1020 Technical Feasibility for Attention Guidance	EX1-CRT-05.971-TLR4-TVALP-FEAS-1021 Technical feasibility for use of A-R applications in the tower environment is verified.	No technical showstoppers were experienced in the simulated tower environment. The exercise does not provide clarity, though, on potential showstoppers in the actual tower environment. The A-R symbology correlated accurately with the objects in the simulated outside view and tracking labels followed the a/c. Visibility of the symbology was sometimes competing	OK

		with reflections of light coming from the surroundings.	
	EX1-CRT-05.971-TLR4-TVALP- FEAS-1022 Technical feasibility for integration of the A-R applications with other related system enablers is verified.	The A-R attention guidance received information from the alerting system. No integration issues were noted.	OK
EX1-OBJ-05.971-TRL4-TVALP- H105.1010 ATCO Task Support with Attention Guidance	EX1-CRT-05.971-TLR4-TVALP-H105.1011 Workload maintained at acceptable level.	No significant differences in workload were found between baseline and A-R condition in post-run ratings. But the expected influence of the A-R on workload was rated 'positive' to 'very positive'. Some extra workload can be explained by the acknowledgement of alerts (clicking) and due to taxi conflicts that controllers experienced as nuisance.	POK
	EX1-CRT-05.971-TLR4-TVALP-H105.1012 Information provided (level and quality) is adequate.	The controllers could provide instructions immediately as the location and the call signs were visible in the A-R device. There was no need to look down to the flight strips. They were happy with the callsigns, type and location of the alert but the extra information in the middle of the field of view was not appreciated. Generally, controllers appreciated that the labels of all a/c were visible.	POK
	EX1-CRT-05.971-TLR4-TVALP-H105.1013	The influence of the A-R on Situational Awareness was rated 'positive' to 'very positive'. The post-run ratings of Situational	POK

	SA maintained at adequate level.	Awareness show that the level of SA is not decreasing.	
	EX1-CRT-05.971-TLR4-TVALP-H105.1014 Head up time is increased with respect to reference (qualitative assessment only).	No head up time was measured, but controllers mentioned they appreciated to be able to stay heads-up in case of an alert.	POK
	EX1-CRT-05.971-TLR4-TVALP-H105.1015 HMI blocking of OTW view.	The test shows the HMI can be improved regarding some of the symbology, in particular the (re)appearance of the alert notification.	POK
	EX1-CRT-05.971-TLR4-TVALP-H105.1016 HMI does not increase potential for human error.	No errors were noticed in the experiment. Furthermore, the system did not interfere with actions to be taken by the controller.	OK
	EX1-CRT-05.971-TLR4-TVALP-H105.1017 ATCO system trust at acceptable level.	The ATCO trust in the system was rated with a trend towards being not acceptable. One controller was positive and the other more negative. The automated reappearance of alerts was not fully understood and taxi conflicts were experienced as nuisance alerts.	POK
	EX1-CRT-05.971-TLR4-TVALP-H105.1018 Adequate level of HMI usability.	The ATCO ratings of Usability (System Usability Scale) were 40 to 52.5 (on scale of 1 to 100), which is not yet at an acceptable level.	POK
	EX1-CRT-05.971-TLR4-TVALP-H105.1019	The presentation of the alert notification can be improved by avoiding presentation of the notification label in the	POK

	Alarms and alerts not too intrusive (concerns presentation of alarms and alerts only).	middle of the field of view. Presentation of labels and dotted lines to indicate expected movement was good.	
	EX1-CRT-05.971-TLR4-TVALP-H105.1020 Positive feedback on acceptance.	The HMI needs improvement before being acceptable regarding the (re)appearance of alerts.	POK
	EX1-CRT-05.971-TLR4-TVALP-H105.1021 ATCO team maintains acceptable level of SA.	Team SA was not part measured in the experiment, because there was no team set-up. The expected effects of the A-R on team SA were rated 'neutral' to 'positive' in the post-experiment questionnaire.	POK
EX1-OBJ-05.971-TRL4-TVALP- H105.1030 Consistency of Attention Guidance with Human Capabilities and Limitations	EX1-CRT-05.971-TLR4-TVALP- H105.1031 Operating methods applied accurately, efficiently and in timely manner.	The ATCOs could react immediately to alerts and rated the influence of the A-R on identifying the involved a/c, on locating the a/c and the ATCO response time as 'positive' to 'very positive'.	OK
	EX1-CRT-05.971-TLR4-TVALP-H105.1032 Operating methods clearly identified and consistent in all operating conditions.	The A-R provides additional information to the controller but has no effect on the operating methods or procedures. The information is consistent with the alert information that is presented head-down.	OK
EX1-OBJ-05.971-TRL4-TVALP- H105.1040 Job Acceptance and Satisfaction with Attention Guidance	EX1-CRT-05.971-TLR4-TVALP-H105.1041 Positive feedback on job satisfaction and acceptance.	The ATCOs rated the question whether they would like to use the system frequently as 'neutral' to 'positive'.	OK
EX1-OBJ-05.971-TLR4-TVALP-SAFE.1010	EX1-CRT-05.971-TLR4-TVALP-SAFE-1011	The ATCOs rated the influence of the A-R on safety	OK

Safety Impact of Attention Guidance	Changes related to implementation of A-R applications do not increase potential for human error.	as 'neutral' to 'very positive'. As operating procedures do not change there is no expectation that it influences the human error rate.	
	EX1-CRT-05.971-TLR4-TVALP-SAFE-1012 ATCO workload is shown to be maintained at acceptable level.	The influence of the A-R on workload was rated 'positive' to 'very positive'. The post-run ratings of Workload show that the Workload is maintained at an acceptable level.	OK
	EX1-CRT-05.971-TLR4-TVALP-SAFE-1013 ATCO SA is shown to be maintained at acceptable level.	The level of SA is maintained (post-run ratings) or expected to be increased (post-experiment ratings).	OK
	EX1-CRT-05.971-TLR4-TVALP-SAFE-1014 Safety assessment activities and results are documented and integrated in overall solution validation results.	The comments of ATCOs show that SA is built up faster and instructions can be given earlier since ATCOs do not have to look down for the alert type, location or call signs. The more efficient and timely presentation and understanding of safety-relevant events is expected to have a positive impact on safety.	OK
EX1-OBJ-05.971-TLR4-TVALP-PERF.1010 Performance Benefits of Attention Guidance	EX1-CRT-05.971-TLR4-TVALP-PERF-1011 Contributions to cost efficiency assessment are made by investigating positive impact on SA, workload and efficiency of ground operation.	The A-R guidance is not changing the role of the ATCO, it is only allowing the ATCO to be timelier with instructions concerning safety-relevant events. Validation has shown that ATCOs perceive a positive impact on SA and workload. While this should improve the efficiency of the ground operation as a whole, this	OK

		experiment could not show it due to the limited set-up.	

Table A-3: Technological Validation Results EXE-001

A.7.1.1 Results on technological feasibility

The A-R symbology correlated accurately with the objects in the simulated outside view. Controllers appreciated the use of A-R for presenting the call signs of each a/c. The ATCOs found the A-R guidance very effective and efficient for both runway incursion and go-around detection. As the labels/call signs of the involved a/c are in view there is no need to look down to consult the flight strips.

Some difficulties were reported that regarded the hardware. Both ATCOs reported to have a somewhat heavy head by the end of a day of experiments. Also, one ATCO experienced some reflections of surrounding light, which seemed to be hardware related as the problem was experienced more with one of the two HoloLens devices. Also, the use of the HoloLens in combination with the personal glasses of ATCOs could increase this problem. Further, 3D audio alerts were not integrated with the radio head-sets in the prototype and, as such, were not part of the evaluation.

The visual alert presented in the middle of the visual field was sometimes obstructing the view of the A/C. The alert disappeared when the controller was facing the right direction but reappeared after a certain interval, when the separation was still below the set criteria. Attention guidance for Runway incursions and Go-Around detection was appreciated very much. For taxi-conflicts the attention guidance was less appreciated, partially because the algorithm was not as advanced (not taking into account the clearances that were given) and partially because the level of urgency is lower. Besides that, taxi conflicts are not part of the runway controller's responsibility.

A.7.1.2 Results per KPA

Feasibility

In the simulated environment the a/c labels correlated really well with the a/c visible in the outside view. No technical showstoppers were encountered in the simulated tower environment.

Human Performance

The evaluation exercise provides subjective evidence that ATCO reaction times will decrease with the A-R guidance, because the controllers do not have to look to head-down displays for information. This was not objectively measured. Controllers commented that it was efficient and convenient not having to look down for call signs of the a/c concerned.

ATCOs reported to experience a 'heavy head' by the end of a day of experiments, to a certain degree. The controllers never mentioned this to be a showstopper, expecting that A-R hardware in the future will become leaner and less intrusive, seeing the great advantages of it.

The current prototype had not reached a development stage sufficient to gain a stable level of trust from the controllers.

Safety

In general, it is expected that safety will increase because with A-R guidance controllers can give instructions more efficiently (with improvement of the notification presentation). In the experiment no negative effects on Workload or Situational Awareness were found. The set-up of the system (add-on and consistent with other system info) and the outcome of the experiment do not give reason to believe that it will have a negative effect on error rates.

A.7.2 Analysis of EXE-001 Results per Technological Validation objective

The upcoming sections of this document describe the results of EXE-001 per Technological Validation Objective specified [28]. While all of the objectives mentioned were addressed, as can be seen from **Table A-3**, not all of them could be fully exploited due to the fact that the AR device was still considered to be an unfinished prototype and the simulation set-up was reduced (not the full operational set-up of Schiphol Tower). Nevertheless, wherever possible, air traffic controller feedback and subjective results from questionnaires were used to get as close to the essence of an objective as possible. This means that, in some cases, participants were asked to imagine a situation with a larger scope and to give an estimate about the projected performance in a certain area. This limitation of the simulation should be considered before drawing far-reaching conclusions on performance values, both regarding technology impact and human performance.

A.7.2.1 EX1-OBJ-05.971-TRL4-TVALP-FEAS.1010

The experiment confirmed the concept is operationally feasible when addressing the Use Case for Attention Guidance with A-R device.

For one subject (wearing glasses), the visibility of the symbology was sometimes competing with reflections of light coming from the surroundings. Both subjects reported to experience a 'heavy head' to a certain extent by the end of a day of experiments.

The visual alert presented in the middle of the visual field was sometimes obstructing the view of the A/C. The alert disappeared when the controller was facing the right direction but reappeared after a certain interval, when the separation was still below the set criteria.

The attention guidance for Runway incursions and Go-Around detection was appreciated very much. For taxi-conflicts the attention guidance was less appreciated, partially because the algorithm was not as advanced (not taking into account the clearances that were given) and partially because the level of urgency is lower. Besides that, taxi conflicts are not part of the runway controller's responsibility and were therefore considered nuisance alerts.

A.7.2.2 EX1-OBJ-05.971-TRL4-TVALP-FEAS.1020

The experiment showed that the concept is technically feasible. A-R symbology correlated accurately with the objects in the simulated outside view. The A-R attention guidance had no interaction with other systems.

A.7.2.3 EX1-OBJ-05.971-TRL4-TVALP- H105.1010

The responses to the post-experiment questionnaire show that the influence of A-R on workload was rated 'positive' to 'very positive', both during normal operation and in case of an alert. Nevertheless, the post-run ratings of Workload show no significant differences between A-R and baseline.

The controllers mentioned that it was effective and efficient to get the alert, the location and the labels in the A-R device so that there was no need to look down to the strips before giving an instruction to solve the conflict.

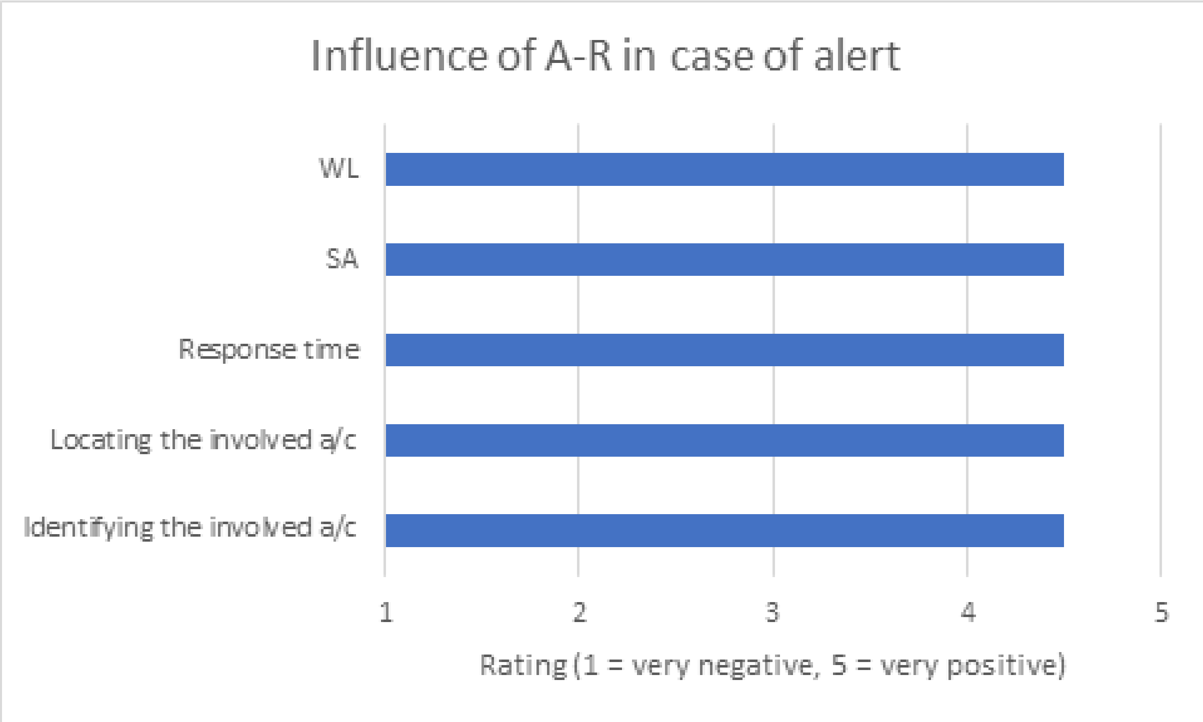


Figure A-2: Post-experiment rating on the influence of A-R guidance on Workload, Situational Awareness, Response time to alerts, Locating the involved a/c, and identifying the call signs.

The influence of A-R on Situational Awareness was rated (post-experiment) 'positive' to 'very positive', both during normal operation and in case of an alert. The post-run ratings of Situational Awareness show no significant differences between A-R and baseline, which means that a sufficient level of SA is maintained. The controllers stated that, not having to search for information about where the conflict is and which a/c (call signs) are involved, was a substantial improvement.

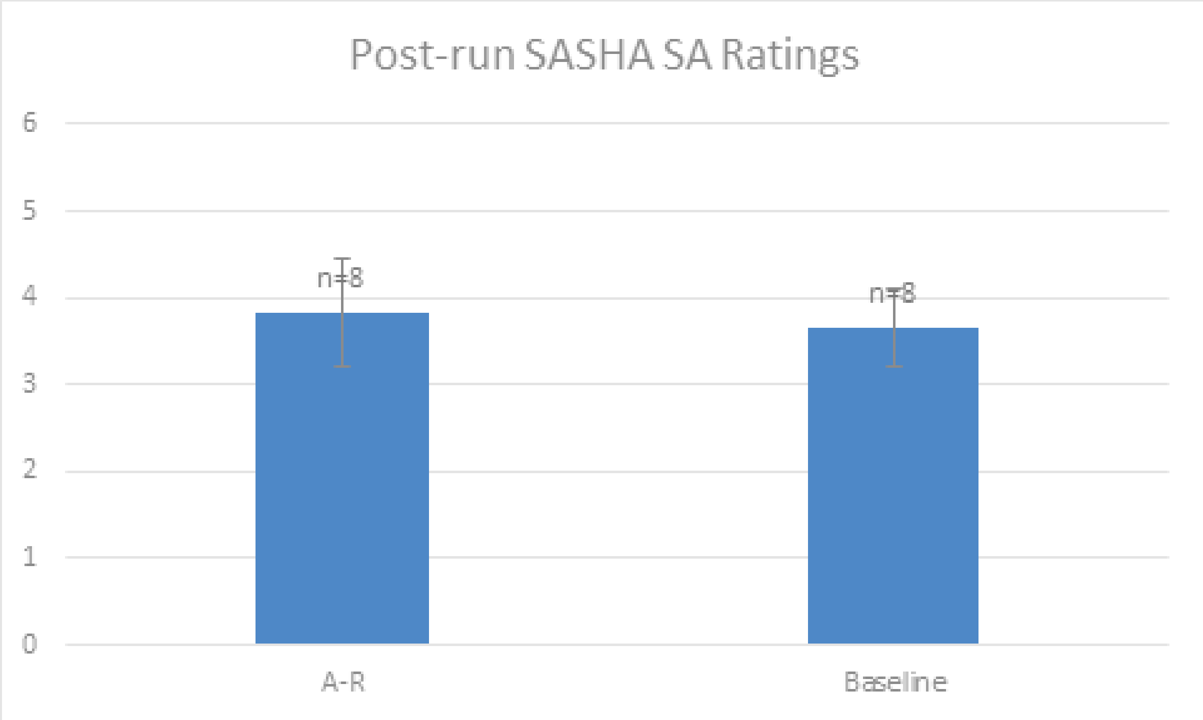


Figure A-3: The average Situational Awareness ratings of 8 comparable runs

Head up time was not measured, but controllers mentioned they appreciated to be able to stay heads-up in case of an alert, as the location and the call signs of the involved a/c appear in the A-R symbology.

The HMI can be improved, by not placing the notification in the middle of the field of view and improving the rules for reappearance of the notification. The controllers mentioned that the alert notice disappeared when facing the direction of the conflict, but reappeared after a certain interval when the separation remained below the set minima, even if the conflict was solved.

No errors were noticed in the experiment. Furthermore, the system does not interfere with actions to be taken by the controller.

The ATCO ratings of Trust (SHAPE Automation Trust Index) for the A-R were 2.3 to 3.6 which is one point less compared to the baseline. Explanations for this lower rating: the reflections of surrounding light that one of the controllers experienced, placement of the alert notice in the middle of the field of view, the reappearance of the notice was not always understood or not desired, and the taxiway conflict alerts were experienced as nuisance. The taxiway conflicts are not part of the responsibility of the runway controller and the algorithm of this type of conflict could be improved.

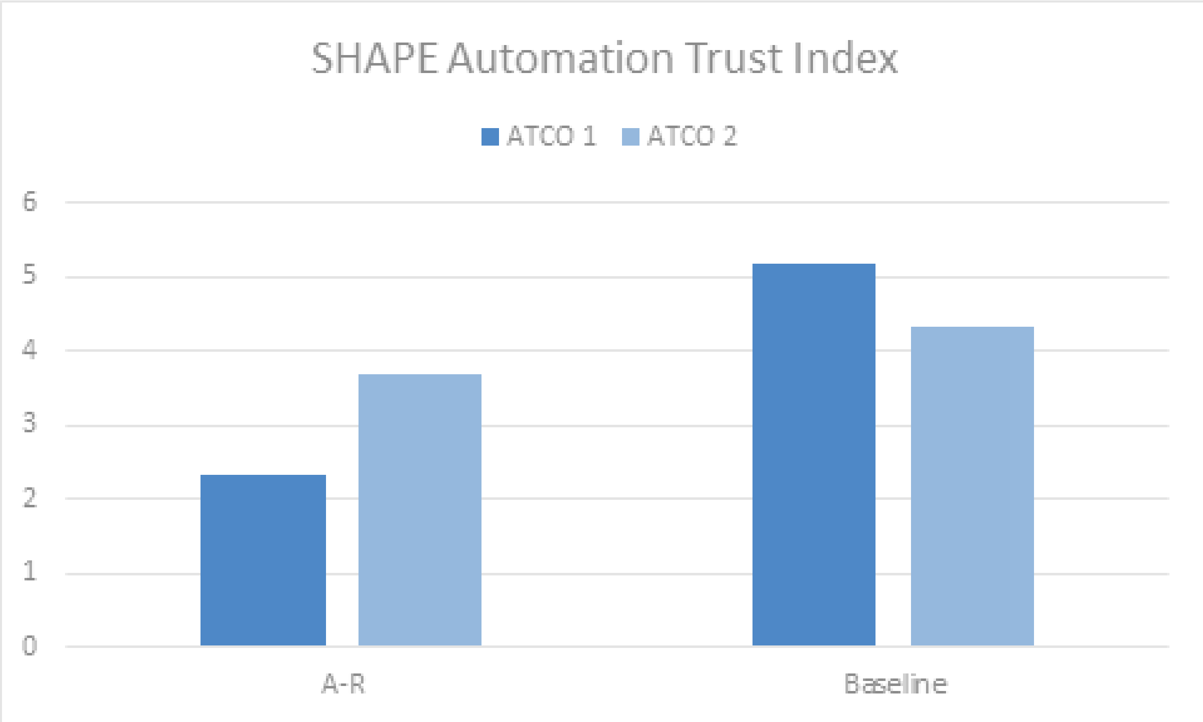


Figure A-4: ATCO ratings on the SHAPE Automation Trust Index (SATI) after Baseline and after A-R usage (values are based on a single rating)

The ATCO ratings of Usability (System Usability Scale) were 40 to 52.5 (on scale of 1 to 100), which is not yet at an acceptable level. The explanations given for success criterion 1017 (above) apply here too. The presentation of the alert notification can be improved, as specified before.

The post-experiment ratings on Controller Acceptance Rating Scale were 1 and 7 (on a scale to 1 to 10). The ATCO that rated the acceptance '1' mentioned that the HMI needed improvement before being acceptable. Improvements are specified under success criterion 1015.

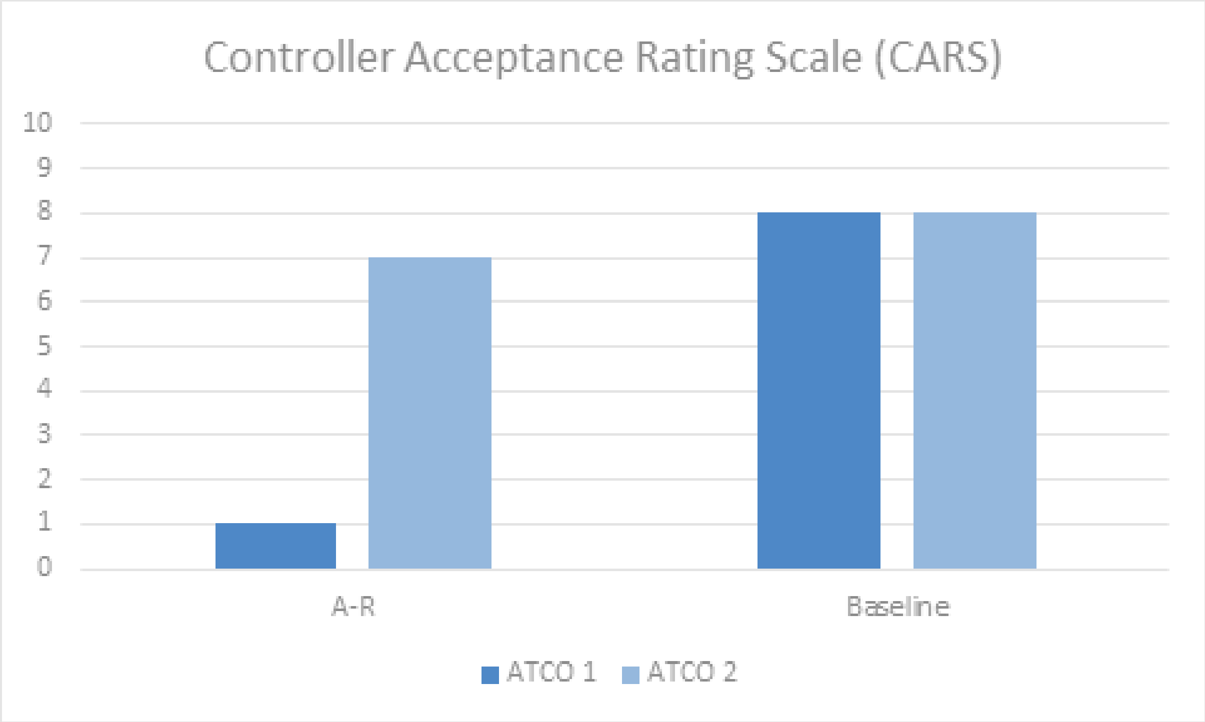


Figure A-5: Controller Acceptance Ratings after the baseline and after the A-R usage

The expected effects of A-R on team SA were rated 'neutral' to 'positive' in the post-experiment questionnaire. This was not part of the experiment itself.

A.7.2.4 EX1-OBJ-05.971-TRL4-TVALP- H105.1030

The experiment proves that ATCOs can apply operating methods for the prototype for Attention Guidance with an A-R device in an accurate, efficient and timely manner.

The ATCOs rated the influence of the A-R (in case of an alert) on identifying the involved a/c, on locating the a/c and the ATCO response time as 'positive' to 'very positive'. The controllers mentioned that it was effective and efficient to get the alert, the location and the labels in the A-R device so that there was no need to look down to the strips for giving an instruction to solve the conflict.

The A-R device provides additional information to the controller but has no effect on the operating methods or procedures. This information is consistent with the alert info presented head-down.

The Controllers rated the effect on their response time to alerts ‘positive’ to ‘very positive’.

A.7.2.5 EX1-OBJ-05.971-TRL4-TVALP- H105.1040

The experiment showed that acceptance of the prototype evaluated was not achieved yet. The ATCOs rated the question in the System Usability Scale regarding whether they would like to use the system frequently 'neutral' to 'positive'. The Controller Acceptance Rating Scale (CARS) was rated 7 by one controller and 1 by the other (scale 1 to 10). The controller who gave a negative acceptance rating, mentioned that the concept had great potential but that alert design needs improvement and nuisance warnings (taxi conflicts in this case) had to be solved.

The A-R guidance does not change the role of the ATCO. Therefore, job satisfaction is not expected to be affected and was not specifically addressed in the questionnaires.

A.7.2.6 EX1-OBJ-05.971-TLR4-TVALP-SAFE.1010

The ATCOs rated the influence of the A-R on safety as 'neutral' to 'very positive'. As operating procedures do not change there is no expectation that it influences the human error rate.

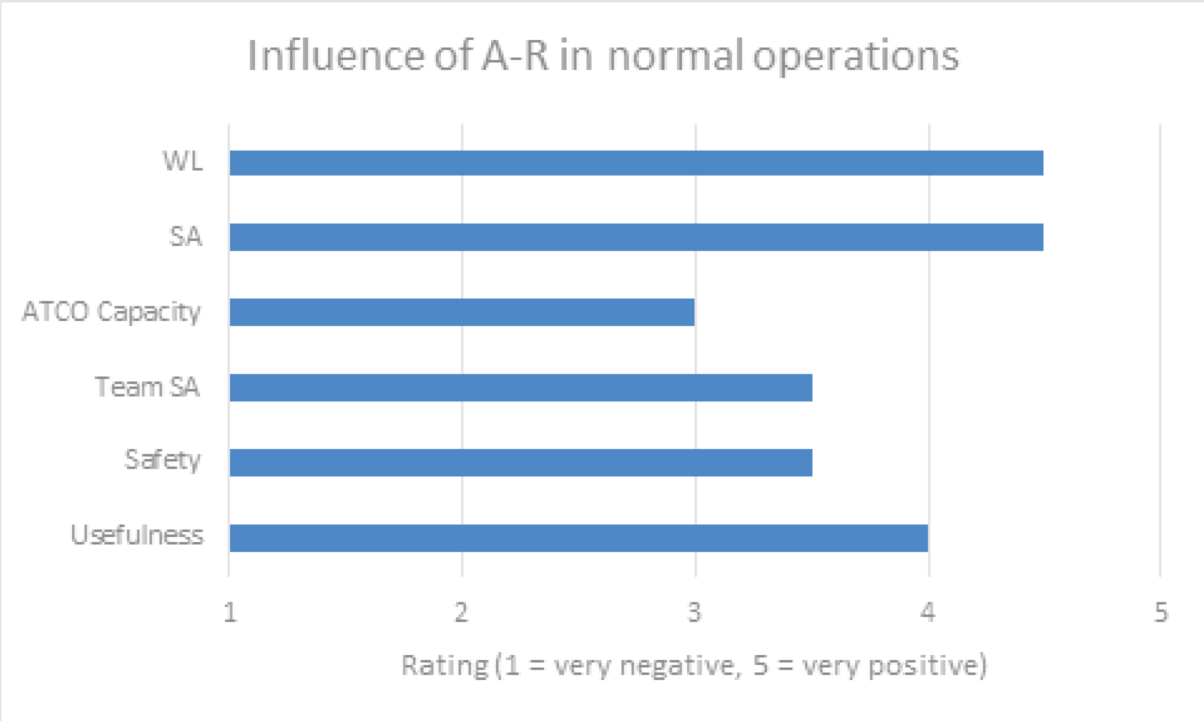


Figure A-6: Post-Experiment ratings regarding the influence of the A-R in normal operations on Workload, situational Awareness, ATCO capacity, Team Situational Awareness, Safety and Usefulness of the A-R

As stated previously, the responses to the post-experiment questionnaire show that the influence of A-R on workload was rated 'positive' to 'very positive', both during normal operation and in case of an alert. The post-run ratings of Workload show no significant difference between A-R and baseline. This, in turn, means that Workload is maintained at an acceptable level.

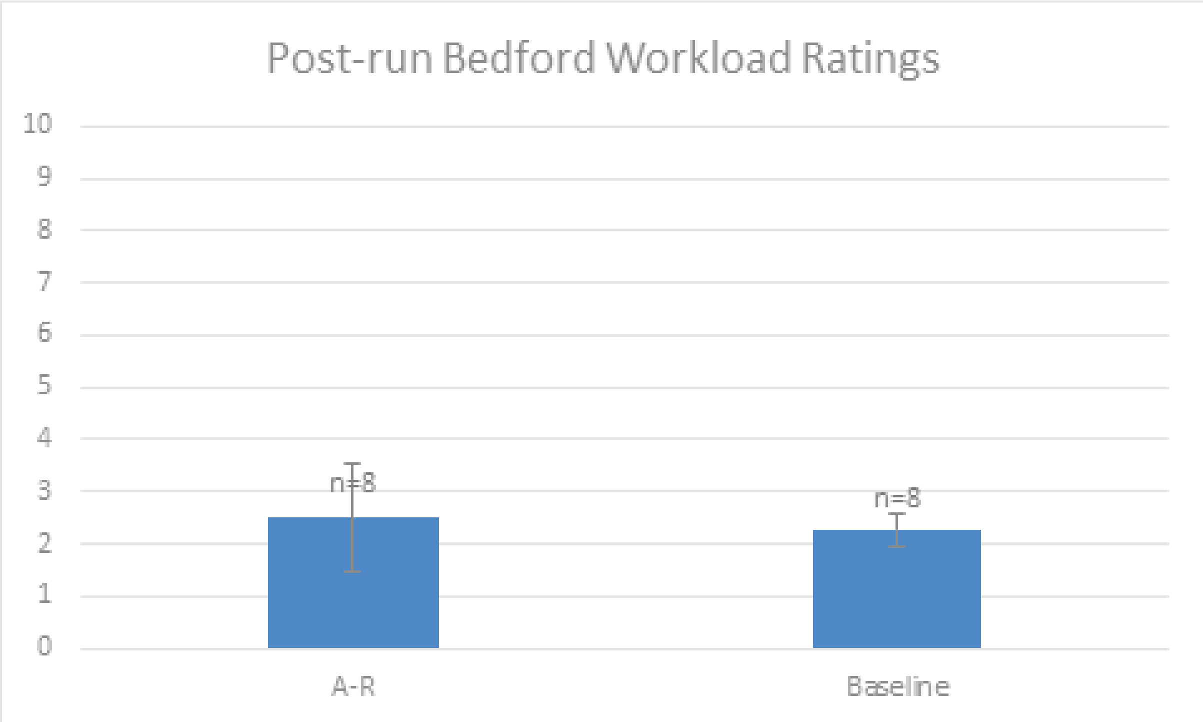


Figure A-7: Post-run ratings of Workload (for 8 comparable runs).

As stated before, the level of SA is maintained (post-run ratings) or expected to be increased (post-experiment ratings).

The A-R guidance is not changing the role of ATCOs, it is only providing information allowing ATCOs to be timelier with instructions.

A.7.2.7 EX1-OBJ-05.971-TLR4-TVALP-PERF.1010

The comments of ATCOs show that because they do not have to look down for the call signs, SA is built up faster and instructions can be given earlier.

A.7.3 Unexpected behaviours/results

There was no unexpected behaviour and there were no unexpected results regarding the simulation platform NARSIM. ATCOs made some comments on the use of the HoloLens, though.

One controller found it very difficult to adjust the A-R device in such a way that the holographic objects could be seen clearly and with the right colours. This controller also needed some more time to get adjusted to use paper flight strips at the same time.

As this unexpected behaviour could be mitigated in the actual simulation runs and directly concerns the technology of the solution, all comments made were considered to be comments on the solution itself, rather than the simulation set-up.

A.7.4 Confidence in results of EXE-001

A.7.4.1 Level of significance/limitations of Technological Validation Exercise Results

Generally, operations and responsibilities of the Aerodrome Tower Controllers did not change with the use of an A-R device. The alerting and related attention getting and guidance were thus only meant to direct the focus of the controller on a particular situation that, without intervention, is expected to lead to a conflict. This means that there were no limitations or an impact on the level of significance for the A-R device operating method.

What was different from the current working environment at Schiphol was the limited scope of the scenarios when compared to the full operation in the Schiphol centre tower. It should be noted though, that this limited scope was only chosen because the exercise had a focus on certain parts of the operation in the tower, in particular the work of the tower controllers in a specific critical situation. The focus was thus put on Safety and Human Performance issues, while an assessment of efficiency improvements for the complete operation in both Schiphol towers was too complex to be considered. This means that, depending on the safety net that was used as an attention guidance trigger, more attention was given to either the work of the Tower Runway controller or the work of the Tower Ground controller. While both controller roles were considered, task sharing among the controller team in the tower was not directly assessed. Only one working position was measured at a time, while the second controller had an observer role.

It was assumed that the focus on a specific situation that will be resolved with Attention Guidance would only be a particular aspect of the complete operation and would thus not be representative of the complete operation at Schiphol. Accordingly, no performance issues concerning the complete Schiphol operation were assessed.

A.7.4.2 Quality of EXE-001 results

The level of representativeness and quality of the simulation is considered as high. This is due to the proven record of the NARSIM Tower validation platform in carrying out simulations of Amsterdam Airport Schiphol for ATC the Netherlands (LVNL). These activities not only consisted of research projects, but very often focused on very practical applications, such as the Schiphol Winter Training and the course programme for Schiphol Tower Control given to ATC operational experts as well as experts from related disciplines. More recently, the same environment was also used for training tower controllers in the use of the acquired EFS system. For that purpose, the actual EFS system was integrated into the simulation platform to allow for all desired manipulations. In summary, the simulation platform used was considered to be of high realism and quality.

The simulated scenarios were considered moderately realistic. In particular the scenario with taxiway conflicts was seen as less realistic. The reason for this is that the researchers needed to elicit taxiway conflicts in order to expose the ATCOs to the warnings and guidance associated with these events. After all, without these conflicts, validation of the attention guidance would have been more difficult (see also the comment on nuisance alerts in Section A.6).

The differences with the real tower environment that were reported are, obviously, the lighting conditions. A simulator cannot offer the same brightness as real sunlight. Apart from the differences mentioned above, the ATCOs found that the simulator offered a realistic representation of the tower environment that they know.

The same goes for the A-R device. Usage of the A-R device in such a (very realistic) simulation environment did not lead to any limitations from a purely conceptual point of view. Even though lighting conditions in the simulation environment were very different from the conditions in a real tower environment and were even less favourable, the A-R device managed to keep track of its own position and presented the holographic objects with high precision. System engineers tested this thoroughly (as described in Ref. [7]) and found that there was excellent alignment between a/c positions on the projection screen and within the HoloLens. While tuning to the specific situation of the observer will still play a role (tuning it is a HoloLens function but controllers initially might need assistance from people familiar with the device) this fact certainly contributed to the quality of the exercise.

A.7.4.3 Significance of EXE-001 results

The simulation exercise was based on the participation of two (former Schiphol) air traffic controllers, with one controller actively involved in control and guidance activities, and the other controller providing assistance. Both controllers were wearing a HoloLens, but only the controller giving ATC instructions was considered an exercise subject and was filling out questionnaires. The work between controllers was evenly divided and both controllers took part in debriefings and interviews.

In summary, this means that this approach, as a first introduction of both a technological enabler and an operational concept for Attention Guidance in a limited Schiphol environment, does obviously not give reliable results concerning an impact on the Schiphol operation as a whole or even in part. That will only be possible, if the recommended changes in the Attention Guidance symbology and logic are realized and re-evaluated and a larger operational scope including adaptation of required information to different controller roles and team working aspects has been investigated.

Nevertheless, the feedback obtained from the simulations led to new ideas regarding all aspects of the Attention Guidance, namely the cues used, the information provided and the triggers for different concept phases (e.g. when to remind the controller of an alerting situation). As such the exercise results are very valuable and can be built upon when continuing research in this area and scaling it up towards different controller roles and interaction between controllers.

As a consequence, the conclusions and recommendations that will follow are also based heavily on the debriefing and interview results, as these results offer more insight into the procedural aspects and the experience that the controllers had with both the HoloLens and the Attention Guidance concept. All other results must also be seen in the light of the limitations of the set-up and the fact that this was the first time indeed that controllers from Schiphol airport were confronted with the use of an Augmented Reality device in their working environment.

A.8 Conclusions

This chapter describes the conclusions that could be derived from the results listed above and gives an outlook towards further research into the area of Attention Guidance with and Augmented Reality device in the aerodrome control tower.

A.8.1 Conclusions on technological feasibility

As was already shown during the preparation activities for this exercise [32], the use of the HoloLens inside a highly realistic simulation environment for the tower operation of Schiphol was technically feasible, especially because excellent alignment between the NARSIM outside view and holographic elements projected inside the HoloLens could be achieved. Furthermore, for the given lighting conditions, which are considered less favourable than in a real tower environment, this was also a very important result, as it may be assumed that orientation of the HoloLens might even be better in the real environment and that the outside view has more contrast, so that HoloLens images do not appear too brightly on top of the background. Obviously, it will still be necessary to confirm this result in a real tower environment, most importantly, to find out whether the radar sources available are sufficiently accurate to allow for the same kind of alignment. In that regard, it must be considered that a simulation always provides accurate and reliable a/c positions (because it also generates these positions).

When looking at the particular aspect of the Attention Guidance logic, its cues and the information provided, the XML via TCP solution worked as intended. It allowed to communicate a/c positions (from simulation), as well as different alerting types and situations to the HoloLens that presented the relevant cues and information when triggered accordingly by the Attention Guidance logic. Due to practicability, the monitoring of controller viewpoint (or rather the expected focus area of controller attention) was performed inside the HoloLens. This means that specific triggers to again guide the controller to the area of interest when the focus was seemingly lost, were also generated inside the HoloLens. In future research, that monitoring could be achieved differently or the monitoring information would have to be sent to the Attention Guidance logic first to re-evaluate new steps to take to capture and guide controller attention. For the purpose of this exercise, the additional communication loop between HoloLens and NARSIM logic for Attention Guidance was considered an unnecessary overhead.

Finally, the results obtained also addressed the A-R device itself, particularly regarding the tuning and use of the visor, the limited field of view of the visor that projects the holographic images, and the combined use of R/T headphones. Only a well-tuned HoloLens leads to good alignment and clear colours of the holographic images. This action needs to be trained and controllers need to be instructed well on how to use the device to that end. A difference in contrast between superimposed elements and outside view was also noticeable in the simulation environment. This might not be the case, though, in the intended (real) control room environment. The limited field of view will lead to more head movements than usual and may be tiring, as was confirmed by the controllers. Controllers mentioned that, in general, the HoloLens was surprisingly wearable, but controllers also felt they had a heavy head (due to too much muscle strain) at the end of a simulation day. Finally, the use of R/T inside the HoloLens was not tested. While speakers are already integrated in the HoloLens (they were used to give aural alerts to the controller), these speakers were not used to communicate. The same goes for the microphones. While they can be used to capture user feedback, this was not done in the exercise, and it has also not been tested whether the microphones could be used for R/T communication as well.

In summary, the HoloLens 2 that was used as A-R device in this exercise and as enabler of an Attention Guidance concept for safety-critical events on runways and taxiways, was considered as a favourable addition to the controller working environment, although there was still room for improvement. While technical performance improvements (mostly related to user comfort and general adjustments) will mainly depend on vendor development, the HoloLens used was considered a technically useful device for implementing prototypes for Attention Capturing and Guidance with aural and visual cues. This observation relates to its ability to position itself correctly within the simulation environment and to exchange the required data with the simulated ATC system. It is expected that a comparable observation will be made in a real tower cabin.

A.8.2 Conclusions on performance assessments

A.8.2.1 General

In general, the ATCOs attributed a huge potential to both technology and Attention Guidance logic. And, as with any brand-new system, they spotted room for improvement as well.

Attention Guidance via the HoloLens was primarily designed to allow ATCOs to better respond to alerts. In terms of an efficient and timely response to the alert, ATCOs considered the concept to have a positive contribution. Many of the classical human factors themes was rated as not influenced, or not improved, though. The reason for that is as follows:

Situational Awareness during the events was rated positive. What is Interesting is the fact that this result was accomplished despite the current limitations of the hardware (the HoloLens). Controllers had to get used to it, accept the coating on the glasses which makes it harder to read on paper when looking down, and had to accept the weight of the system on their heads. Despite this hassle, the ATCOs were neutral-positive about their SA, and also about their workload and the likelihood of making human errors. One of the reasons for this outcome is that the ATCOs were very well able to recognise the potential of the system. This is also reflected in their trust of, or confidence in, the system. One of them was very positive about it and the other one mildly positive. This was due to the fact that the first controller, when filling in the questionnaires, primarily focused on the potential of the system while the second one merely looked at the system as it was presented to him. ATCOs even started, by themselves, to think about and discuss the added value of the system.

A.8.2.2 Symbology

The symbology, and that included parts of the concept, was not acceptable yet in the way it was presented. The re-appearance of alerts, in some cases, was considered unnecessary and distracting. Further, the initial alert was presented in a too intrusive way in the middle of the A-R display. In the design phase the designers made a number of assumptions regarding the use of the system, and they were not able to entirely verify these assumptions with Schiphol controllers. It turned out that ATCOs do not need as much repetition and attention warnings as the designers assumed for those, quite serious, events. Therefore, the symbology and concept should, in a next step, be adjusted to this end user requirement.

A.8.2.3 Side-effects

An unanticipated result that was found was that the intuitive way of presenting the callsigns of all a/c at and around the airport, offers, in general and not just during particular events, an increased SA. After all, there is less head-d own time needed to build and maintain SA. Further, the mental workload is reduced, also because there is less head -down checking to do. These types of findings are a good reason to further develop the concept and to also use the HoloLens for other application areas than Attention Guidance (i.e. not only for the three main types of events that were studied in this experiment).

A.8.2.4 Safety

No impact on safety is expected. In particular, this is the case when some of the hardware limitations are solved. The symbology will bring events faster to the attention of the ATCOs. The call signs do increase SA in general. Those two effects are considered very positive for the safety of the operation. This is even more true, if the symbology becomes less distracting, for example by not repeating alerts too often, and such possible drawbacks in the concept are eliminated.

A.8.2.5 Main Conclusion

Attention Guidance via the HoloLens as an A-R device has a high potential, and definitely deserves more attention in order to further validate it operationally. In the described experiment a rather generic set -up was validated. For example, no distinction between runway or ground controllers was made, while in fact both roles may require another, more customised way of presenting the necessary information. Furthermore, team working aspects and coordination activities between different controllers or controller roles were not considered yet. However, as it has been proven that the Information presentation technique of the HoloLens is an interesting addition to the tower controller working environment, it will become a relevant issue to also study operational areas where such a technique may also be helpful or even more helpful.

A.8.3 Recommendations

The symbology that was offered was a first version. Prototype developers had to make a number of design choices and assumptions regarding the operation and end user requirements that could not be verified prior to experiment execution. Therefore, the experiment led to two issues that should be considered when designing an updated version of the symbology:

A.8.3.1 # 1

Do not try to re-capture controller attention. In the original design, techniques were applied to see if the ATCO indeed immediately responded as expected by the designers. If that was not the case, the system would repeat the alert and try to capture and guide the attention of the controller again. For the situations chosen in the scenarios this seemed to be a distracting and superfluous step of the concept. This was for a good reason. ATCOs stated that, once they were alerted of a serious event (such as a runway incursion and go-around), they do not need to be alerted for the same event again. What they do need, and that was indeed offered in the current designs, is solid guidance when they ask for it (rather than a system that monitors them).

A.8.3.2 # 2

Make symbols smaller (a bit less Attention Getting/Capturing). This statement follows from remarks that the symbology chosen was a bit too big and too bright. It did attract attention but did a bit too much given the circumstances. The guidance is helpful, but this felt as too much inference. Smaller and less distracting symbols would probably be the solution.

A large number of the issues that need to be solved in order to allow operational application of the technology comes from hardware that is being used and not from the information presentation and the use of sensors. However, the system is a COTS product, and the vendor (Microsoft) is aware of these issues. The vendor is continually working on them, and the expectation is that, in the coming years, a number of these issues will be solved, or at least their impact will be reduced. The main items that NLR recommends to be worked on are:

- Make the contraption less heavy. In fact, the user is wearing an entire Windows 10 laptop on his/her head. Even though it is balanced, it is still bulky and when bending the neck, the demands on the muscles in and around the neck are very different from the normal situation. The system must become lighter, either by using lighter material, making everything smaller, or taking certain components, in particular the batteries, out.
- Even though improved compared to previous versions of the HoloLens, there is still a relatively small display. Two different kinds of drawbacks of this situation are that the ATCOs need to position the system very carefully and that, once positioned correctly, there still is the need to move the head more into the direction of a situation that needs visual attention compared to the situation without the device. Less items can be seen in the periphery.
- The HoloLens display is coated. One of the disadvantages of a coating is that, for the ATCO, it seems as if s/he is looking through sunglasses. In the real tower that is acceptable in many cases, but certainly not at night, during bad weather conditions or when studying information (on paper) on the controller working position. For those situations another coating should be found that blocks less light.

A.8.3.3 Operations

- Do not solely focus on alerts related to high impact events. Also consider other means to enhance SA, reduce workload, and share knowledge with the intent to increase ATCO efficiency and capacity. The events that were chosen for the study are important, but due to that, they are also situations that ATCOs will monitor closely. Possibly, the next step would be to focus on increasing awareness of ATCOs about general operational issues at the airport and on approach. By studying the kind of information that ATCOs look up every now and then and offering that information in a more intuitive way (on a silver platter, so to speak) ATCOs will build their SA with less effort. This topic deserves more attention in next iterations regarding Attention Guidance in the tower control room.
- Provide clear guidelines about when and how to use the system. When more functions will be added to the system and, in later stages of development, the system will be in use by ATCOs, the exploratory phase of using the system is over. By then it must be clear in which situations the system should be used, what information is redundant with other sources of information that exist within the tower, and when using the system is voluntary. While such work cannot be done in this development stage, this is already a good time to become aware of the need for operational guidelines.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Appendix B Technological Validation Exercise 002 Report

B.1 Summary of EXE-002 Plan

This chapter summarises the technological validation exercise for EXE-05.97.1-TRL4-TVALP-VAR-002 carried out by ENAV, University of Bologna, and DBL as part of Solution 97.1 "Virtual/Augmented Reality Applications for Tower", in the following referred to as EXE-002.

B.2 EXE-002 description and scope

- EXE-002 was carried out by ENAV, University of Bologna and DBL and investigated the use of AR in a conventional control tower environment at Bologna airport with specific focus on adaptive HMI, multimodal interaction, and safety nets.
- Adaptive HMI and working positions: the exercise is conducted for two different working positions (i.e. Tower RWY and Tower GND Controllers). This implies tracking of the two points of view in order to customize the two different views. The system provides the user with specific information based on his/her working position, current flight status and visibility conditions.
- Multimodal Interaction: the users interacted with the system by a combination of at least two of the following interaction modes: gaze, gesture, voice. Tracking labels technology combined with V/AR and air gesture interaction enables stripless interaction with V/AR tracking labels to issue not-time-critical clearances.
- Safety net: the V/AR overlays was used to display safety warnings such as runway incursions and conflicting clearances (i.e. T_O vs LANDING, T_O vs RWY occupied).
- The exercise objective was to mature the results obtained in the previous RETINA validation campaign including additional features that were not considered at exploratory research level.
- The exercise has been conducted as a real time human in the loop simulation exploiting University of Bologna's validation platform that was initially developed for SESAR Exploratory Research RETINA project. The new version of the platform includes two Microsoft HoloLens devices integrated within the platform. No site acceptance test was foreseen, but both technical acceptance test and operational acceptance test have been executed.
- Results were collected in the following KPA: Human Performance, Cost Efficiency, Performance and Safety and they will contribute to reach Solution 97.1's target maturity TRL4.

B.2.1 Validation Platform / Tool & Validation Technique

The platform used for EXE-002 is a research validation platform of the University of Bologna. It is located at the Virtual Reality and Simulation Laboratory of the University of Bologna in Forlì, Italy and it was used in the previous validation of the Exploratory Research Project RETINA.

The exercise platform architecture is depicted in Figure B-8~~Figure B-8~~. It consists in five modules that feed three posts, namely an ATCOs GND post, an ATCOs RWY post, and a pseudo-pilot post.

The core system is the **4D model** of the reference scenario which communicates through data exchange protocols with the following four subsystems:

- Out of the Tower View Generator (OOT): it provides the ATCOs (RWY and GND) with a consistent and photorealistic view of the out of the tower scene.
- Augmented Reality Overlay Application (AR App): it derives the relevant Augmented Reality Overlays and deploys them on the Head Mounted See Through Display Microsoft HoloLens 2.

- Head Down Equipment (HDE): it consists in a simplified interface that replicates the actual head down equipment in the control tower.
- Pseudo-pilot application (PP App): it allows the pseudo-pilot to monitor and update the state of the 4D model according to the commands provided by the ATCO.

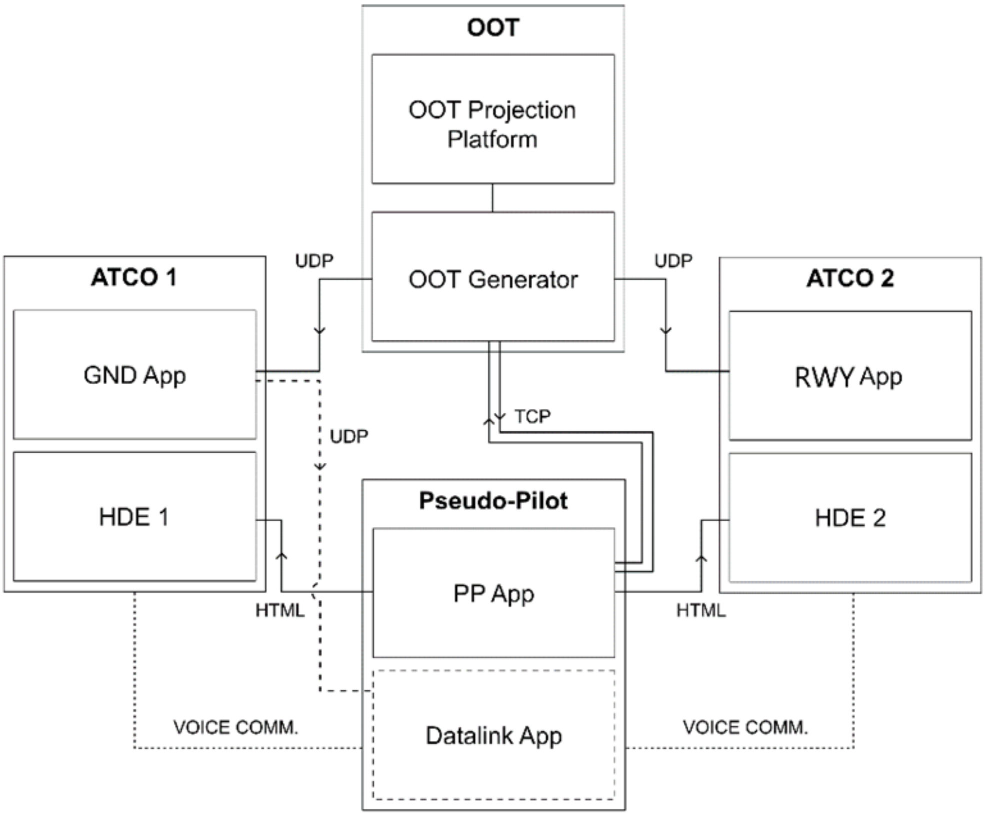


Figure B-8: UNIBO's validation platform architecture

B.2.1.1 4D Model

The four-dimensional digital model (3D+time) is the core system of the exercise platform. This model integrates all data sources and is able to manage events and respond to user inputs. The model was developed using Blender software and it includes most of the airport static features and ground signs, a library of a/c and ground vehicles, a "point and click" interface for managing a/c and ground vehicles, assigning taxi routes ad clearing take-offs and landings.

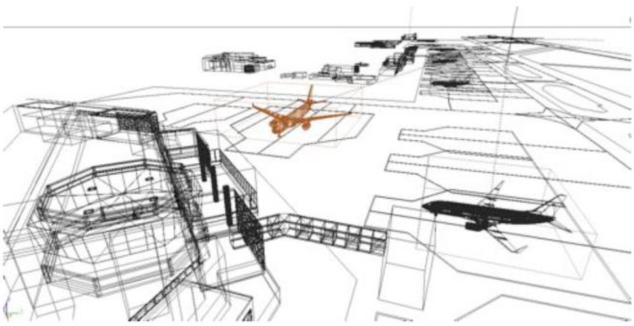


Figure B-9: 4D model of the reference scenario Bologna Airport (LIPE)

B.2.1.2 Out-of-the-Tower View Generator (OOT)

The Out-of-the-Tower View Generator derives a rendered view of the reference scenario from the 4D model and displays it on the RVE (Reconfigurable Virtual Environment). It provides the ATCOs with a semi-immersive, consistent and photorealistic view of the out of the tower scene. The Reconfigurable Virtual Environment is a CAVE-like virtual environment designed to recreate a sense of immersion by means of three, rear-projected, flat screens. The screens can be arranged in three different configurations, closed, semi-closed and wide open. Head tracking can be activated by means of a Microsoft Kinect sensor. In the OOT a custom rendering pipeline generates images based on the viewer's position providing the user with a good immersivity.

For the specific purpose of this validation, both the head tracking function and the stereoscopic option have been disabled. The RVE has been arranged in a semi-closed configuration with the side screens rotated around 10°.

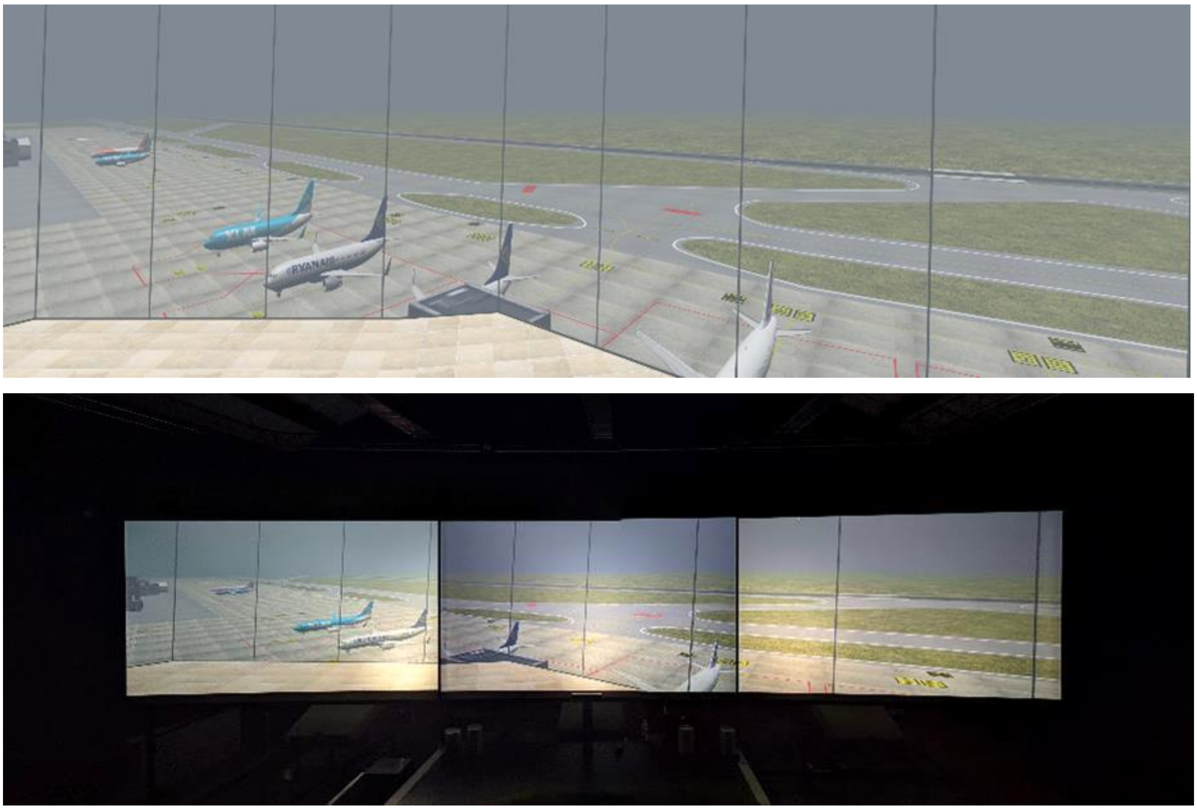


Figure B-10: Out of Tower view generator: screenshot (top) and Reconf. VE with tracking system (bottom)

B.2.1.3 Augmented Reality Overlay Application (AR App)

The Augmented Reality Overlay Application derives the relevant Augmented Reality Overlays from the 4D model and deploys them on Microsoft HoloLens 2. The ATCOs can contemporarily see both the out of the tower view and the Augmented Reality Overlays. The Augmented Reality Overlays are customized based on the specific ATCO post (either RWY or GND) and the visibility condition as depicted in Table B-4.

	GND App	RWY App

Labels	At stand	√	
	In push	√	
	In Taxi	√	√
	Take-off	√	√
	Approach	√	√
	Landing	√	√
Air Gesture interaction (Buttons)	Clearance	√	
	Push	√	
	Start-up	√	
Dynamic overlays		√	√
Safety nets			√
Dynamic HUD		√	√

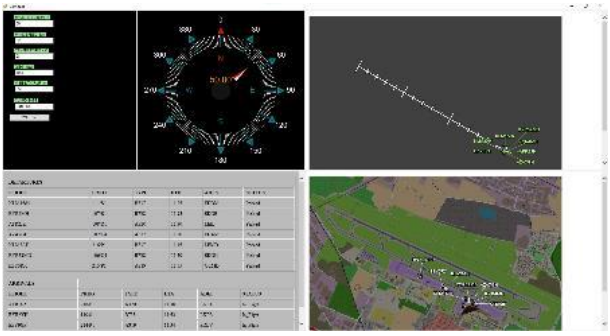
Table B-4: GND and RWY Application features

Figure B-11: ATCOs can see both OTW Augmented Reality Overlays through HoloLens at the same time. The personal view of the user is depicted in the blue square

B.2.1.4 Head Down Equipment (HDE)

It consists in a simplified interface that presents a set of data similar to that given to the ATCOs via the actual head down equipment in the control tower. It derives data from the 4D model and presents it to the ATCO on a 24 inches screen.

In each screen it is possible to find four windows with different information (Figure B-12~~Figure B-12~~).



Weather Information	Approach Radar
FDP	Ground Radar

Figure B-12: HDE interface

The weather information shown varies over time according to a predetermined time schedule and corresponding to the variation of weather conditions in the simulation. The "driving" parameter is visibility (VISIBILITY/RVR), to which the variation of the other parameters (wind, temperature, pressure, cloud base) is connected. The FDP is updated in real time with data from the simulation of the current scenario.

B.2.1.5 Pseudo-pilot application (PP App)

The pseudo-pilot application allows its operator to monitor and update the state of the 4D model. The 4D model and the pseudo pilot application communicate to keep the state consistent between them.

In order to have an overview of the airport and to represent the current state of the 4D model, the following data layers are available in the pseudo-pilot application:

- Background imagery (i.e. aerial picture of the airport and its surroundings)
- Schematic layout of the airport
- Visual representations of the navigation mesh (taxiways and waypoints)
- a/c's and ground vehicles (matching the position in the 3D model)

The pseudo-pilot operator can toggle the visibility of the layers above.

To execute the validation scenario the pseudo-pilot application provides the following actions.

- Load the initial state of a validation scenario.
- An a/c or a vehicle can be selected and a taxi route is assigned to it.
- The pseudo pilot operator can instruct the a/c/vehicle to follow the pre-planned route. The a/c/vehicle can also be ordered to stop or resume a route again (i.e. stop taxi or continue taxi).
- The pseudo pilot operator can clear an a/c for take-off.
- The pseudo pilot operator can clear an a/c to land.

The operator executes the actions above when prompted to do so by the ATCO. Each command will be sent to the 4D model which will update its state accordingly.

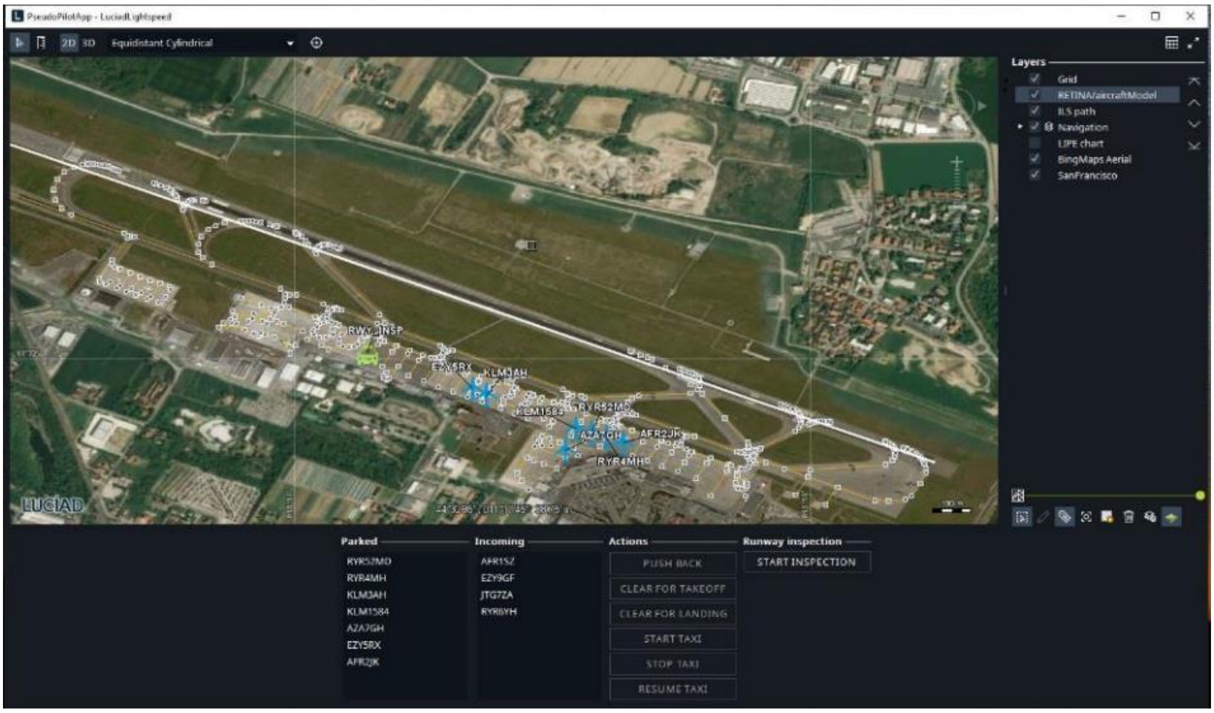


Figure B-13: Pseudo pilot application interface

Moreover, during the execution of the scenarios encompassing air gestures and safety nets, the pseudo pilot application includes an additional window with datalink messages to communicate with the ATCOs and trigger warnings and alarms (Figure B-14~~Figure B-14~~).

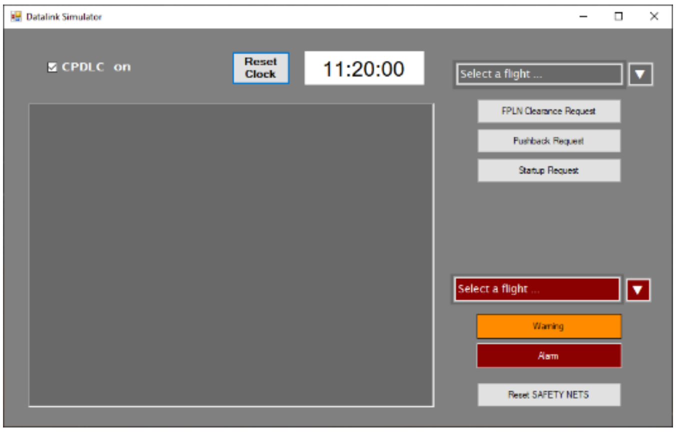


Figure B-14: Datalink interface

B.2.1.6 Architectural view: Mapping Validation Infrastructure and SUTs onto EATMA

V&V Platform Name	UNIBO's Reconfigurable Virtual Environment Simulation Platform

A.1.1	It is a new developed V&V platform?	No. The platform itself has been built in 2001 as a multipurpose research virtual environment. Since 2016 it has been customized as a Control Tower Simulator to perform research on newly conceived HMI for Airport Control Tower.
A.2	It was the first time to be used for a SESAR validation exercise?	No. UNIBO's platform was used to validate the RETINA concept at exploratory research level.
A.3	It was used the first time in a SESAR validation exercise and it needed new features to be implemented?	No. UNIBO's platform has been already used for RETINA project.
B	Which operational scenarios / improvements/etc. (general) can be validated on the new V&V Platform?	The platform is able to simulate different visibility conditions at Bologna airport (LIPE) and the consequent restrictions, including transition from good to bad visibility conditions.
B.1	Which are the ATM Domain Systems supported by the Validation Platform?	The platform provides a simplified CWP of Bologna airport. Relevant Domain System is Tower.
B.2	Which functional blocks of the Validation Platform will be provided and/or needed to support the operational concepts validation?	The validation platform simulated the air traffic controller environment for one runway controller and one ground movement controller at Bologna airport. In addition, a specially adapted COTS AR device (i.e. HoloLens) has been used by both controllers. Functional blocks associated are: • Virtual and Augmented Reality Display (PJ.05-W2-97.1) Air Gestures Detector (PJ.05-W2-97.1)
C	Which validation needs were supported by the new platform (not covered by the existing platforms)?	An adapted AR device (HoloLens) will be used in a conventional aerodrome tower environment to provide head up information to air traffic controllers depending on their role and on the visibility conditions. In addition GND ATCO will be able to interact with the overlays by means of air gestures to manage not-time-critical tasks (i.e. start-up and push-back).
D	Which validation methods can be used on the new V&V Platform?	Human-in-the-loop Real-time Simulations

Table B-5: Validation Exercise Platform/Tool mapping onto EATMA

B.2.1.7 Validation Exercise technique

The Validation technique to perform EXE-002 is a Real-time Simulation (RTS) with Humans-in-the-loop (HITL). In this case, the HITL concerned the Bologna Tower Ground and Tower Runway Controllers and pseudo-pilots. UNIBO's validation platform provided data recording for post-run analysis. Additionally, briefing, debriefing, questionnaire and interviewing techniques were used.

B.3 Summary of EXE-002 objectives and success criteria

SESAR Solution Validation Objective	SESAR Solution Success Criteria	Coverage and comments on the coverage of SESAR Solution Validation Objective in EXE-002	Exercise Validation Objective	Exercise Success Criteria
OBJ-05.971-TRL4-TVALP-FEAS.1010 To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS.	CRT-05.971-TLR4-TVALP-FEAS-1011 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Virtual or Augmented Reality and tracking labels.	Fully	EX2-OBJ-05.971-TRL4-TVALP-FEAS.1010 To confirm the concept is operationally feasible when addressing the Use Case for Virtual or Augmented Reality, tracking labels, and Air Gestures	EX2-CRT-05.971-TLR4-TVALP-FEAS-1011 Identical to CRT-05.971-TLR4-TVALP-FEAS-1011
	CRT-05.971-TLR4-TVALP-FEAS-1012 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Air Gestures.	Fully		EX2-CRT-05.971-TLR4-TVALP-FEAS-1012 Identical to CRT-05.971-TLR4-TVALP-FEAS-1012
	CRT-05.971-TLR4-TVALP-FEAS-1013 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Attention Guidance.	N/A		N/A

OBJ-05.971-TRL4-TVALP-FEAS.1020 To identify possible technical feasibility issues and possible show stoppers.	CRT-05.971-TLR4-TVALP-FEAS-1021 Laboratory tests (based on a prototype) have verified the technical feasibility of the use of V/A-R applications in the tower environment.	Fully	EX2-OBJ-05.971-TRL4-TVALP-FEAS.1020 Identical to OBJ-05.971-TRL4-TVALP-FEAS.1020.	EX2-CRT-05.971-TLR4-TVALP-FEAS-1021 Identical to CRT-05.971-TLR4-TVALP-FEAS-1021.
	CRT-05.971-TLR4-TVALP-FEAS -1022 Laboratory tests have verified that the integration of the V/A-R applications with other related system enablers is technically feasible.			EX2-CRT-05.971-TLR4-TVALP- FEAS-1022 Identical to CRT-05.971-TLR4-TVALP-FEAS-1022.
OBJ-05.971-TRL4-TVALP-H103.1010 To assess that the technical systems for V/A-R Tracking Labels and overlays support the ATCOs in performing their tasks.	CRT-05.971-TLR4-TVALP-H103.1011 Majority of ATCOs (at least 75%) responses is that V/A-R supports ATCO in maintaining workload at acceptable level.	Fully	EX2-OBJ-05.971-TRL4-TVALP-H103.1010 Identical to OBJ-05.971-TRL4-TVALP-H103.1010	EX2-CRT-05.971-TLR4-TVALP-H103.1011 Feedback from controllers (at least 75%) shows that the prototype for V/AR supports controllers in maintain an acceptable level of workload.
	CRT-05.971-TLR4-TVALP-H103.1012 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R			EX2-CRT-05.971-TLR4-TVALP-H103.1012 Identical to CRT-05.971-TLR4-TVALP-H103.1012
	CRT-05.971-TLR4-TVALP-H103.1013 Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO in maintaining an adequate level of situation awareness.			EX2-CRT-05.971-TLR4-TVALP-H103.1013 Identical to CRT-05.971-TLR4-TVALP-H103.1013
	CRT-05.971-TLR4-TVALP-H103.1014			EX2-CRT-05.971-TLR4-TVALP-H103.1014

	Measured time spent in head up is increased in the solution scenario with respect to the reference scenario.			Identical to CRT-05.971-TLR4-TVALP-H103.1014
	CRT-05.971-TLR4-TVALP-H103.1015 HMI of V/A-R tools does not overshadow the relevant information on the OTW view.			EX2-CRT-05.971-TLR4-TVALP-H103.1015 Identical to CRT-05.971-TLR4-TVALP-H103.1015
	CRT-05.971-TLR4-TVALP-H103.1016 V/A-R HMI does not increase the potential for human error.			EX2-CRT-05.971-TLR4-TVALP-H103.1016 Identical to CRT-05.971-TLR4-TVALP-H103.1016
	CRT-05.971-TLR4-TVALP-H103.1017 ATCOs' (at least 75%) trust in the system is at an acceptable level.			EX2-CRT-05.971-TLR4-TVALP-H103.1017 Identical to CRT-05.971-TLR4-TVALP-H103.1017
	CRT-05.971-TLR4-TVALP-H103.1018 Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R HMI.			EX2-CRT-05.971-TLR4-TVALP-H103.1018 Identical to. CRT-05.971-TLR4-TVALP-H103.1018
	CRT-05.971-TLR4-TVALP-H103.1019 Majority of ATCOs (at least 75%) responses is that alarms and alerts are not too intrusive and support ATCOs in the early detection of ATC critical situations.			EX2-CRT-05.971-TLR4-TVALP-H103.1019 Majority of ATCOs (at least 75%) responses show that alarms and alerts in the prototype for V/AR are not too intrusive and support ATCOs in the early detection of ATC critical situations with respect to conflicting clearances and runway incursions.

	CRT-05.971-TLR4-TVALP-H103.1020 ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R tool.			EX2-CRT-05.971-TLR4-TVALP-H103.1020 Identical to CRT-05.971-TLR4-TVALP-H103.1020
	CRT-05.971-TLR4-TVALP-H103.1021 Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO team (GND and TWR) in maintaining an acceptable level of situation awareness.			EX2-CRT-05.971-TLR4-TVALP-H103.1021 Identical to CRT-05.971-TLR4-TVALP-H103.1021
OBJ-05.971A-TLR4-TVALP- H103.1030 To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Tracking labels and overlays	CRT-05.971-TLR4-TVALP-H103.1031 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner.	Fully	EX2-OBJ-05.971-TRL4-TVALP-H103.1030 Identical to OBJ-05.971A-TLR4-TVALP-H103.1030	EX2-CRT-05.971-TLR4-TVALP- H103.1031 Identical to CRT-05.971-TLR4-TVALP-H103.1031
	CRT-05.971-TLR4-TVALP-H103.1032 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions.			EX2-CRT-05.971-TLR4-TVALP-H103.1032 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in the investigated operating conditions.

OBJ-05.971-TRL4-TVALP- H103.1040 To assess job acceptance and satisfaction with the introduction of V/A-R tracking labels and overlays	CRT-05.971-TLR4-TVALP-H103.1041 ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance.	Fully	EX2-OBJ-05.971-TRL4-TVALP-H103.1040 Identical to OBJ-05.971-TRL4-TVALP-H103.1040	EX2-CRT-05.971-TLR4-TVALP-H103.1041 Validation activities show that ATCOs (at least 75%) give positive feedback on job satisfaction and acceptance regarding the prototype for V/AR tracking labels and overlays.
OBJ-05.971-TRL4-TVALP-H104.1010 To assess that the technical systems for V/A-R Air Gestures support the ATCOs in performing their tasks.	CRT-05.971-TLR4-TVALP-H104.1011 Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures support ATCO in maintaining workload at acceptable level.	Partial (Being a technical test only, these objectives will be addressed though qualitative feedback collected during the debriefing)	EX2-OBJ-05.971-TRL4-TVALP-H104.1010 To assess that the technical systems for V/A-R Air Gestures support the ATCOs in performing the tasks under investigation.	EX2-CRT-05.971-TLR4-TVALP-H104.1011 Identical to CRT-05.971-TLR4-TVALP-H104.1011
	CRT-05.971-TLR4-TVALP-H104.1012 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R Air Gestures.			EX2-CRT-05.971-TLR4-TVALP-H104.1012 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of the interaction means supported by V/A-R Air Gestures.
	CRT-05.971-TLR4-TVALP-H104.1013 Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures HMI supports ATCO in maintaining an adequate level of situation awareness.			EX2-CRT-05.971-TLR4-TVALP-H104.1013 Identical to CRT-05.971-TLR4-TVALP-H104.1013
	CRT-05.971-TLR4-TVALP-H104.1014 Measured time spent in head up is increased in the solution scenario with respect to the reference scenario.			EX2-CRT-05.971-TLR4-TVALP-H104.1014 Identical to CRT-05.971-TLR4-TVALP-H104.1014

	CRT-05.971-TLR4-TVALP-H104.1015 V/A-R Air Gestures HMI does not increase the potential for human error.			EX2-CRT-05.971-TLR4-TVALP-H104.1015 Identical to CRT-05.971-TLR4-TVALP-H104.1015
	CRT-05.971-TLR4-TVALP-H104.1016 ATCOs' (at least 75%) trust in the system is at an acceptable level.			EX2-CRT-05.971-TLR4-TVALP-H104.1016 Laboratory tests show that ATCOs' (at least 75%) trust in the prototype for V/AR Air Gestures is at an acceptable level.
	CRT-05.971-TLR4-TVALP-H104.1017 Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R Air Gestures HMI.			EX2-CRT-05.971-TLR4-TVALP-H104.1017 Identical to. CRT-05.971-TLR4-TVALP-H104.1017
	CRT-05.971-TLR4-TVALP-H104.1018 ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R Air Gestures tool.			EX2-CRT-05.971-TLR4-TVALP-H104.1018 Identical to CRT-05.971-TLR4-TVALP-H104.1018
OBJ-05.971A-TLR4-TVALP- H104.1020 To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Air Gestures	CRT-05.971-TLR4-TVALP-H104.1021 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner.	Partial (Being a technical test only, these objectives will be addressed though qualitative feedback collected during the debriefing)	EX2-OBJ-05.971-TRL4-TVALP-H104.1020 Identical to OBJ-05.971A-TLR4-TVALP-H104.1020	EX2-CRT-05.971-TLR4-TVALP- H104.1021 Identical to CRT-05.971-TLR4-TVALP-H104.1021
	CRT-05.971-TLR4-TVALP-H104.1022 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and			EX2-CRT-05.971-TLR4-TVALP-H104.1022 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in the

	consistent in all operating conditions.			investigated operating conditions.
OBJ-05.971-TRL4-TVALP- H104.1030 To assess job acceptance and satisfaction with the introduction of V/A-R Air Gestures	CRT-05.971-TLR4-TVALP-H104.1031 ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance.	Partial (Being a technical test only, these objectives will be addressed though qualitative feedback collected during the debriefing)	EX2-OBJ-05.971-TRL4-TVALP-H104.1030 Identical to OBJ-05.971-TRL4-TVALP-H104.1030	EX2-CRT-05.971-TLR4-TVALP-H104.1031 Laboratory tests show that ATCOs (at least 75%) give positive feedback on job satisfaction and acceptance regarding the prototype for V/AR Air Gestures.
OBJ-05.971-TLR4-TVALP-SAFE.1010 To assess the impact of Virtual/Augmented Reality applications on safety.	CRT-05.971-TLR4-TVALP-SAFE-1011 Laboratory tests show that the Virtual/Augmented Reality applications improve the safety performance by reducing human error.	Fully	EX2-OBJ-05.971-TLR4-TVALP-SAFE.1010 To give relevant input about safety issues when using V/AR with safety nets	EX2-CRT-05.971-TLR4-TVALP-SAFE-1011 Laboratory tests show that the prototype for V/AR with safety nets improves the safety performance by reducing human error.
	CRT-05.971-TLR4-TVALP-SAFE-1012 Laboratory tests show that the Virtual/Augmented Reality applications improve the safety performance by reducing ATCO workload.			EX2-CRT-05.971-TLR4-TVALP-SAFE-1012 Identical to CRT-05.971-TLR4-TVALP-SAFE-1012
	CRT-05.971-TLR4-TVALP-SAFE-1013 Laboratory tests show that the use of Virtual/Augmented Reality applications improves the safety performance by increasing situational awareness.			EX2-CRT-05.971-TLR4-TVALP-SAFE-1013 Identical to CRT-05.971-TLR4-TVALP-SAFE-1013

	CRT-05.971-TLR4-TVALP-SAFE-1014 Safety assessment activities and the results are documented and integrated in the overall solution validation results.			EX2-CRT-05.971-TLR4-TVALP-SAFE-1014 Identical to CRT-05.971-TLR4-TVALP-SAFE-1014.
OBJ-05.971-TLR4-TVALP-PERF.1010 To assess the performance benefits of equivalent visual operations for tower control through the use of applications for Virtual/Augmented Reality.	CRT-05.971-TLR4-TVALP-PERF-1011 Laboratory tests show that the use of V/A applications improves Cost Efficiency performance by reducing the cost per flight (through e.g. reduction of workload, reduction of delay times).	Fully	EX2-OBJ-05.971-TLR4-TVALP-PERF.1010	EX2-CRT-05.971-TLR4-TVALP-PERF-1011 Laboratory tests show that the prototype for V/AR contributes to an assessment of Cost Efficiency performance by having a positive impact on situation awareness, workload and efficiency of ground operations.
	CRT-05.971-TLR4-TVALP-PERF-1012 Laboratory tests show that the use of V/A reality applications improves Resilience by increasing situational awareness in low visibility conditions while maintaining workload within acceptable limits.			EX2-CRT-05.971-TLR4-TVALP-PERF-1012 Identical to CRT-05.971-TLR4-TVALP-PERF-1012

B.4 Summary of EXE-002 validation scenarios

Guglielmo Marconi International Airport in Bologna (LIPE) was chosen as a reference scenario for EXE-002.

B.4.1 Aerodrome

B.4.1.1 Layout

Bologna is a single Runway (12 and 30) airport with a main taxiway T and several taxiway and a/c stand taxilanes. The runway has a 12/30 orientation with an asphalt strip of 2803x45 m. Table B-6 reports the declared distances for both runways.

13 DISTANZE DICHIARATE	DECLARED DISTANCES			
Designazione RWY RWY designator	**TORA (M)**	**TODA (M)**	**ASDA (M)**	**LDA (M)**
1	2	3	4	5
12	2803	2923	2803	2493
INT TAKE-OFF B	2400	2520	2400	-
INT TAKE-OFF C	2100	2220	2100	-
INT TAKE-OFF D	1900	2020	1900	-
30	2803	2863	2803	2442
INT TAKE-OFF J	2630	2690	2630	-
INT TAKE-OFF H	2395	2455	2395	-

Table B-6: Declared distances for both runways

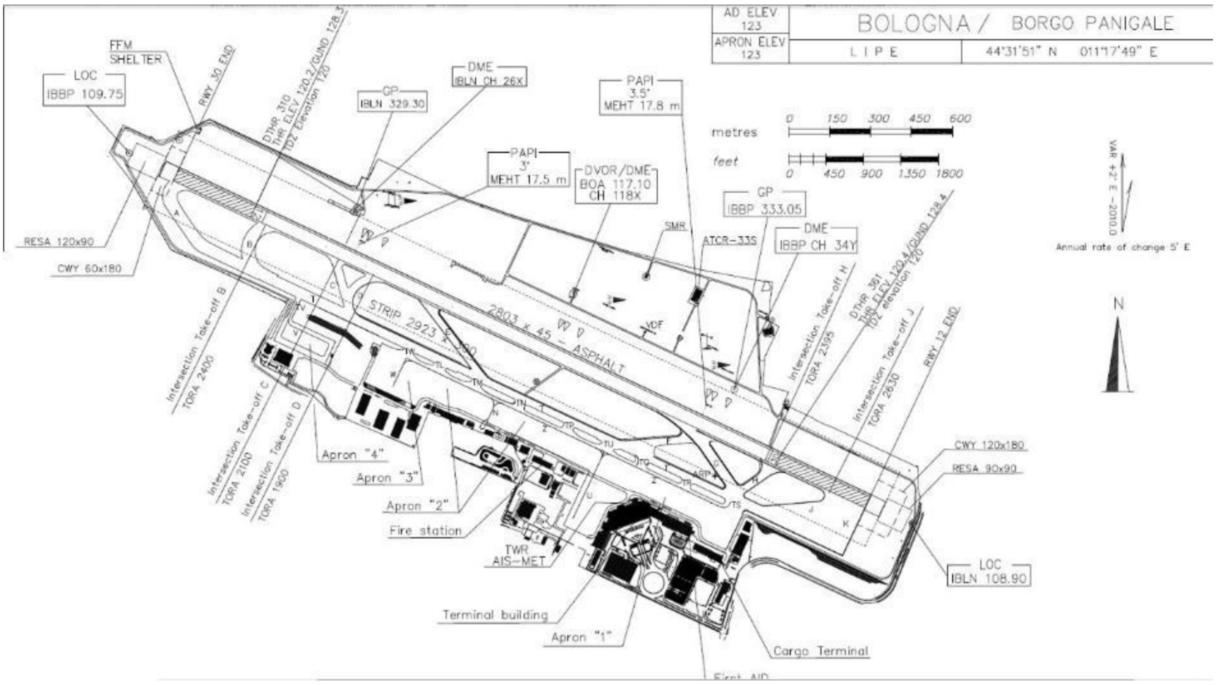


Figure B-15: Bologna airport layout

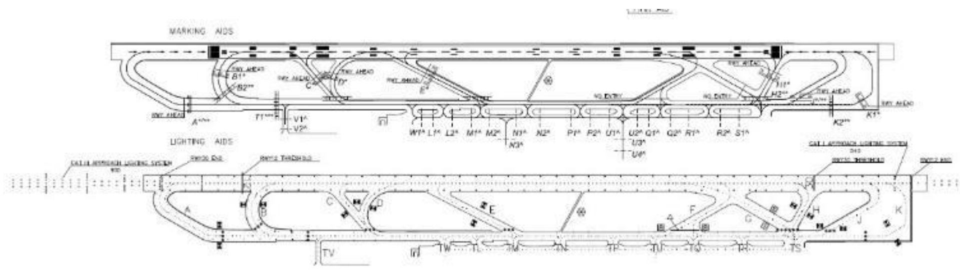


Figure B-16: Bologna airport layout, runway, taxiways

Figure B-15~~Figure B-15~~ reports Bologna Airport layout. The main taxiway T is parallel to the runway and it links all the aprons with the runway. Four aprons are available; Apron 1 in front of the terminal and the Control Tower, Apron 2 on left in front of the firefighting area and hangars, Apron 3 is the cargo area and Apron 4 for general aviation. Aprons 1, 2 and 3 are linked to taxiway T with a short taxiway TW, TL, TN, TM,TP, TU, TQ, TR, and TS; Apron 4 is separated from the other aprons and is linked to the main taxiway T with taxiway TV. The Runway and the main taxiway T are linked via the taxiways A, B, C, D, E, F, G, H, J and K (see Figure B-16~~Figure B-16~~).

Taxiway characteristics including width, surface and strength are reported in the following table.

2	Larghezza, superficie e resistenza delle TWY	TWY width, surface and strength
	A Larghezza: 23 M Superficie: ASPH Resisterza: PCN 50 F/B/X/T	**A** Width: 23 M Surface: ASPH Strength: PCN 50 F/B/X/T
	B Larghezza: 23 M Superficie: ASPH Resisterza: PCN 70 F/B/X/T	**B** Width: 23 M Surface: ASPH Strength: PCN 70 F/B/X/T
	C Larghezza: 23 M Superficie: ASPH Resisterza: PCN 50 F/A/X/T	**C** Width: 23 M Surface: ASPH Strength: PCN 50 F/A/X/T
	D Larghezza: 23 M Superficie: ASPH Resisterza: PCN 46 F/A/X/T	**D** Width: 23 M Surface: ASPH Strength: PCN 46 F/A/X/T
	E Larghezza: 23 M Superficie: ASPH Resisterza: PCN 42 F/A/X/T	**E** Width: 23 M Surface: ASPH Strength: PCN 42 F/A/X/T
	F Larghezza: 23 M Superficie: ASPH Resisterza: PCN 52 F/B/X/T	**F** Width: 23 M Surface: ASPH Strength: PCN 52 F/B/X/T
	G Larghezza: 23 M Superficie: ASPH Resisterza: PCN 85 F/B/X/T	**G** Width: 23 M Surface: ASPH Strength: PCN 85 F/B/X/T
	H Larghezza: 23 M Superficie: ASPH Resisterza: PCN 57 F/B/X/T	**H** Width: 23 M Surface: ASPH Strength: PCN 57 F/B/X/T
	J Larghezza: 23 M Superficie: ASPH Resisterza: PCN 60 F/A/X/T	**J** Width: 23 M Surface: ASPH Strength: PCN 60 F/A/X/T
	K Larghezza: 23 M Superficie: ASPH Resisterza: PCN 60 F/A/X/T	**K** Width: 23 M Surface: ASPH Strength: PCN 60 F/A/X/T
	T Larghezza: 23 M Superficie: ASPH Resisterza: PCN 70 F/A/X/T	**T** Width: 23 M Surface: ASPH Strength: PCN 70 F/A/X/T
	TL Larghezza: 45 M Superficie: ASPH Resisterza: PCN 120 F/A/W/T	**TL** Width: 45 M Surface: ASPH Strength: PCN 120 F/A/W/T
	TM Larghezza: 38 M Superficie: ASPH Resisterza: PCN 120 F/A/W/T	**TM** Width: 38 M Surface: ASPH Strength: PCN 120 F/A/W/T
	TN Larghezza: 41 M Superficie: ASPH Resisterza: PCN 14 F/C/W/T	**TN** Width: 41 M Surface: ASPH Strength: PCN 14 F/C/W/T
	TP Larghezza: 38 M Superficie: ASPH Resisterza: PCN 56 F/B/W/T	**TP** Width: 38 M Surface: ASPH Strength: PCN 56 F/B/W/T
	TQ Larghezza: 38 M Superficie: ASPH Resisterza: PCN 79 F/A/W/T	**TQ** Width: 38 M Surface: ASPH Strength: PCN 79 F/A/W/T
	TR Larghezza: 38 M Superficie: ASPH Resisterza: PCN 113 F/A/W/T	**TR** Width: 38 M Surface: ASPH Strength: PCN 113 F/A/W/T
	TS Larghezza: 75 M Superficie: ASPH Resisterza: PCN 111 F/A/W/T	**TS** Width: 75 M Surface: ASPH Strength: PCN 111 F/A/W/T
	TU Larghezza: 51 M Superficie: ASPH Resisterza: PCN 120 F/A/W/T	**TU** Width: 51 M Surface: ASPH Strength: PCN 120 F/A/W/T
	TV Larghezza: 19 M Superficie: ASPH Resisterza: PCN 87 F/A/W/T	**TV** Width: 19 M Surface: ASPH Strength: PCN 87 F/A/W/T
	TW Larghezza: 44 M Superficie: ASPH Resisterza: PCN 120 F/A/W/T	**TW** Width: 44 M Surface: ASPH Strength: PCN 120 F/A/W/T

Table B-7: Bologna airport taxiway features

The stands are grouped in blocks: all the stands belonging to a block have the same Apron Holding Position, i.e. a position where the a/c are pushed back and where they start up the engines.

B.4.1.2 Radio aids and surveillance systems

Bologna Airport is equipped with Primary and Secondary Surveillance RADAR and with Surface Movement RADAR (SMR). The PSR/SSR version is ATCR 33/S and it is Mode-S equipped. The range of the PSR covers about 65NM and the range of the SSR is about 110NM; the antennas are located together with a rotation every 4 seconds. Mode S information is displayed in a specific window of the CWP (controller working position) and includes several pieces of information such as a/c call-sign, Indicated Air Speed, Heading, Level, etc. The SMR provides a/c and vehicle positions on the manoeuvring area. Specific labelling is available on the CWP for identification. The SMR has a range of 3.5NM and also provides raw video information. The SMR is also able to detect foreign objects and flocks of birds on the runway.

Both runways are equipped with ILS; runway 12 until CAT IIIB and runway 30 until CAT1. The table below reports the main characteristics of the ILS for both runways.

Tipo di radioassistenza Type of aid CAT di/of ILS (VAR ILS/VOR)	ID	FREQ	Orario Operational hours	Coordinate antenna Antenna site coordinates (WGS84)	Elevazione antenna DME Elevation of DME antenna	Copertura operativa nominale Limitazioni Designated operational coverage Limitations	Note Remarks
1	2	3	4	5	6	7	8
GP	-	333.05 MHZ	H24	44°32'00.9''N 011°17'52.3''E	NIL	NIL	Slope 3.5° RDH:16.60 M
ILS RWY 12 LOC CAT IIIB (2° E-2010.0)	IBLN	108.90 MHZ	H24	44°31'45.2''N 011°18'21.4''E	NIL	NIL	1) Fascio posteriore non utilizzabile/ back beam not usable
DME	IBLN	CH 26X	H24	44°32'24.2''N 011°16'50.8''E	40 M AMSL	25 NM/10000 FT limitazioni a/limitations at 25 NM 120°/270° MRA 5000 FT 270°/120° MRA 2500 FT	NIL
GP	-	329.30 MHZ	H24	44°32'23.7''N 011°16'50.3''E	NIL	NIL	Slope 3° RDH:16.50 M
ILS RWY 30 LOC CAT I (1° E-2005.0)	IBBP	109.75 MHZ	H24	44°32'31.3''N 011°16'13.2''E	NIL	limitazioni oltre/limitations beyond 17 NM MRA 3000 FT	NIL

Table B-8: Bologna airport ILS features

B.4.1.3 Local traffic rules and Low Visibility Procedures

The use of the taxiways is regulated via some restrictions:

- TWY F and G shall be used only as an exit taxiway
- TWY B and D shall not be used to enter the runway 12 and perform backtrack
- TWY G is a rapid exit taxiway: max speed 93km/h
- minimum thrust requested to pilots on all taxiways/taxilanes.
- RWY 30 shall be used only if RVR (TDZ, MID and STOP/END) is equal or greater than 550m.

Moreover some restrictions applies depending on the ICAO code of the a/c:

- a/c with ICAO Code F shall use only taxiway A, J and K to enter the runway: A to enter runway 12, J as preferential to vacate runway 12 and K to enter runway 30.
- taxilane Z shall be used by a/c up to ICAO code C between TQ and TS
- a/c with ICAO code letter "D" are allowed to taxi on TWY T and on a/c stand taxilane Z only simultaneously with a/c with ICAO code letter "A"

- a/c with ICAO code letter "E" shall not taxi on a/c stand taxilane Z. Taxiing on TWY T and a/c stand taxilane Z simultaneously with any other a/c is forbidden
- a/c with ICAO code letter "F" shall not taxi on a/c stand taxilane Z. Taxiing on TWY T and a/c stand taxilane Z simultaneously with any other a/c is forbidden
- a/c with ICAO code D, E, F parked on stand 114 or 115 shall be pushed-back on TWY T through TWY TS
- a/c with ICAO code E, F parked on apron 3 shall be pushed-back on TWY T through TWY TW
- use of taxilane N allowed only for a/c up to ICAO code B included
- a/c with ICAO code letter "D" shall not taxi on the a/c stand taxilane Z between apron holding points Q2 and S1
- a/c with ICAO code letter "E" shall use TWY TU/TS/ TW as exit/entry TWY from/to aprons
- a/c with ICAO code letter "F" shall use TWY "TS" as exit/entry TWY from/to stands 114 and TWY "TW" as exit/entry from/to Apron 3.

Low visibility Procedures will be applied CATII/III approaches and to departure operations at following conditions:

- RVR TDZ is 550 m or below.
- Cloud base height/ceiling is below 200ft according to the meteorological local report.
- When the rapid deterioration of weather conditions recommends so.

Pilots will be informed by ATIS (Airline Travel Information System) and/or frequencies when LVP are in force. In case of poor visibility conditions, a reduced airport capacity can be expected due to the requirement of increased spacing between arriving a/c and/or restrictions applied to ground movements.

The ground movements and the separation between arriving a/c (arrival vs arrival) and between arriving and departing a/c (arrival vs departure) depends on the prevailing visibility conditions (CONDI VIS). As such, three visibility conditions are possible:

- CONDI VIS1: Visibility sufficient for the pilot to taxi and to avoid collision with other traffic on taxiways and at intersections by visual reference, and for personnel of control units to exercise control over all traffic on the basis of visual surveillance.
- CONDI VIS2: Visibility sufficient for the pilot to taxi and to avoid collision with other traffic on taxiways and at intersections by visual reference, but insufficient for personnel of control units to exercise control over all traffic on the basis of visual surveillance.
- CONDI VIS3: Visibility sufficient for the pilot to taxi but insufficient for the pilot to avoid collision with other traffic on taxiways and at intersections by visual reference, and insufficient for personnel of control units to exercise control over all traffic on the basis of visual surveillance. For taxiing, this is normally taken as visibilities equivalent to an RVR of less than 400 m but more than 75 m.
    - Local traffic regulation in CONDI VIS 2: Runway 12 is used preferentially and it is mandatory if RVR is less than 550m. Arriving a/c vacate runway 12 only via taxiway G,H and J and runway 30 only via B. Departing a/c enter runway 12 only via A and runway 30 via J. The stopbar at the Runway Holding point CAT II and III are activated. Minimum spacing between arriving a/c is 10NM if LVP are not in force, 12NM in case LVP in force, 15NM to permit departure between arrivals and LVP in force. In case of LVP, in order to ensure that the radio path of the ILS is free, the TWR controller will

clear for take-off a departure only if it will overfly the LOC antenna before the arriving a/c is 4NM on final.

- Local traffic regulation in CONDI VIS 3: Only runway 12 is used. Intermediate holding point (IHP) T1 on main taxiway is activated, the follow-me is positioned on the taxiway T abeam TS on TWR request in case of arrival. Departing a/c taxi to IHP T1 initially and then to RHP A. Further departures start taxi only once the previous one is between T1 and A RHP. Arriving a/c vacate the runway only via J and follow the follow-me until the parking. Push back operations are allowed only from stand belonging to not contiguous blocks. Minimum spacing between arriving a/c is 15NM in case of no departure and 16NM in case of departure. In order to ensure that the radio path of the ILS is free, the TWR controller will clear for take-off a departure only if it will overfly the LOC antenna before the arriving a/c is 4NM on final.

B.4.1.4 Controller Working Position

The TWR RWY controller is responsible to provide the Aerodrome Control Service, the Flight Information Service and the Alert Service to all traffic in the Aerodrome Traffic Zone (portion of airspace with radius of 5NM and 2000ft) and on the Runway. The TWR RWY CWP is depicted in the figure below.

Figure B-17: Bologna Airport TWR RWY CWP

The most important systems used by the ATCO in his/her tasks are the RADAR (air and ground), the Compunction system, the Light Control and the Strips.

The Air RADAR screen is in front of the ATCO and provide position and identification information of all traffic in the area of responsibility, i.e. ATZ (Aerodrome Traffic Zone) in particular information of the traffic on final. The SMR (Ground RADAR) screen is positioned in higher position (not visible in this picture) linked to the room ceiling in front of the ATCO. This position supports the ATCO in the RWY check operations performed before providing all the take-off and landing clearances.

The screen on the right is used by the ATCO to control the aerodrome lights, stopbar (RWY and intermediate) included. Specific buttons are available to set the light in accordance to the visibility conditions and to the approach category (CATII and III) in low visibility conditions. On this screen is also displayed to the ATCO a warning system that inform the ATCO of the aerodrome decategorization in case of system failure. Between the RADAR and the light screen the communication control panel is available. Via this panel, the ATCO manages the frequencies and the telephones. On the right strip printer prints the arrival strips 20 minutes before estimated landing time (departure strip are provided by GND ATCO to the RWY ATCO).

Figure B-18: Bologna airport COO working positions

The picture above shows the COO position where two screens are available. The screen on the right is the approach RADAR providing information of all inbound and outbound flight position within an area of about 100NM. The screen on the right provides the ATCO with the access to all the supporting systems (FDP, AOIS, ADM, see next section). A Communication panel is also available to manage frequencies and telephone.

Picture below reports the GND position (that includes also DEL function).

Figure B-19: Bologna airport GND working positions

The GND controller is responsible to provide the Aerodrome Control Service and the Flight Information Service on the manoeuvring area except the runway. Information from supporting systems are displayed in the screens on the left and on the right. The screen on the right displays AOIS (Aeronautical Operational Information system) information and the screen on the left displays ADM and FDP (see Supporting system section). The central screen is the SMR (ground RADAR) and provide the ATCO with the position information of all the traffic on the manoeuvring area.

B.4.1.5 Meteo systems

In Bologna Airport all the sensors required for CATIIIB operations are available. The meteo data available are:

- Wind (direction and intensity, both average and instant value)
- Pressure (QNH,QFE)
- Temperature, Dew Point

- Visibility general and RVR (in 3 points, i.e. TDZ, MID and STOP/END)
- Cloud base

The meteo info are provided to a/c via ATIS.

B.4.1.6 Supporting systems

Data supporting systems provide the controller with a set of information related to the scheduled times and to the route. The most important scheduled times are:

- EOBT/TOBT: Estimated (Target in case of A-CDM, Airport collaborative decision making) off-block time.
- ETOT/TTOT: Estimated (Target in case of A-CDM, Airport collaborative decision making) take off time.
- CTOT: Calculated take-off time that is provided by the *Network Manager Operations Centre* (*NMOC*)

In Bologna airport the data supporting system available are FDP (Flight Data Processing) and AOIS (Aeronautical Operational Information system). The FDP provides the Controller with route and clearance information for all IFR flights. The AOIS provides the Controller with a set of information, among them the scheduled time and the actual time, i.e. ALT (Actual Landing Time) and ATOT (Actual Take Off Time), and the NOTAM (NOTice to Air Man).

In Bologna airport the most used scheduled time are the EOBT and the CTOT.

B.4.1.7 Traffic main characteristics

Passenger flights are the most significant part of Bologna air traffic; about 90 % of these flights are operated by medium a/c like Boeing 737 or Airbus 320 and only few heavy a/c interest the airport. General aviation has dedicated terminal and only during fair periods this kind of traffic becomes significant. The major part of the movements in Bologna is during day, from 6 am to midnight. Night hours are used by cargo flights. Traffic is quite homogeneous during day but there are some peak hours as early morning during the so called “first wave” or late evening when all based a/c come back together. During peak hours radar approach sequences the inbound flights using vectoring and RNAV arrival route, normally about 10/12 flights arrive at the same time and are sequenced by the approach. In the same way, at morning during first wave about 20 flights schedule departure in the same time window of about 30-40 minutes. Control Tower manages and sequences the outbound traffic.

The airspace around the airport does not have significant restrictions and it is used mainly to protect approach and departure procedures on Bologna airport. Military flights do not interest significantly the airport: only limited traffic from/to the Italian Army maintenance center based at the airport.

In the aerodrome vicinity some airfields are located; they are used mainly from pilot schools and do not affect Bologna air traffic since training procedures in Bologna are forbidden.

B.4.2 Reference Scenarios

The following Reference scenarios have been planned:

- EXE-002-001 is performed with a reference equipment on a scenario where the visibility was progressively reduced from CONDIVIS1 to CONDIVIS2 to CONDIVIS3.
- EXE-002-004 is performed with a reference equipment on an unusual scenario where safety events were simulated.

B.4.3 Solution scenarios

In the solution scenario ATCOs have been operating a scenario that is comparable as the reference scenario, but in the solution scenario both the V/AR and the Air Gestures were available, in addition to conventional CWP. In fact, the subjects were in almost the same situation but in the solution scenario the controller was supported by the technical solution.

The main difference between reference and solution scenario is the additional presentation and supplementary possibility to interact offered by V/AR and Air Gestures respectively. Thus, if performance was influenced in either positive or negative way, this could be attributed to the solution offered. Moreover, the solution scenario included safety nets visualization in AR with specific reference to runway incursions and conflicting clearances.

Specific exercises were conducted in order to separately assess the influence of V/AR, Air Gestures and Safety Net visualization functions.

Specifically, the following solution scenarios were planned to be compared to reference scenario

- EXE-002-002 is a solution scenario that implements Virtual and Augmented Reality Tracking Labels and overlays depending on the specific visibility.

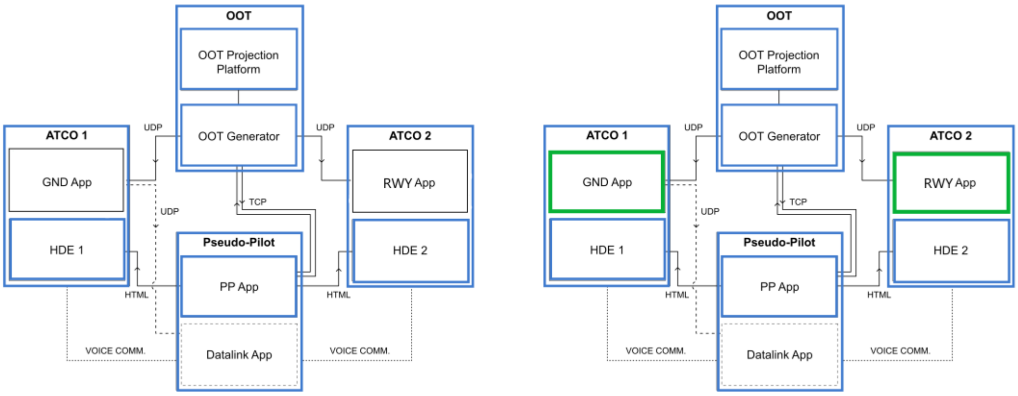


Figure B-20: Validation platform in Reference scenarios (left) and in EXE-002-002 (right)

- EXE-002-003 is a solution scenario where air gestures are implemented for the TWR GND controller to manage not-time-critical tasks (start-up and pushback).
- EXE-002-005 is a solution scenario on an unusual scenario where safety events are simulated and safety net are visualized by means of AR tools.

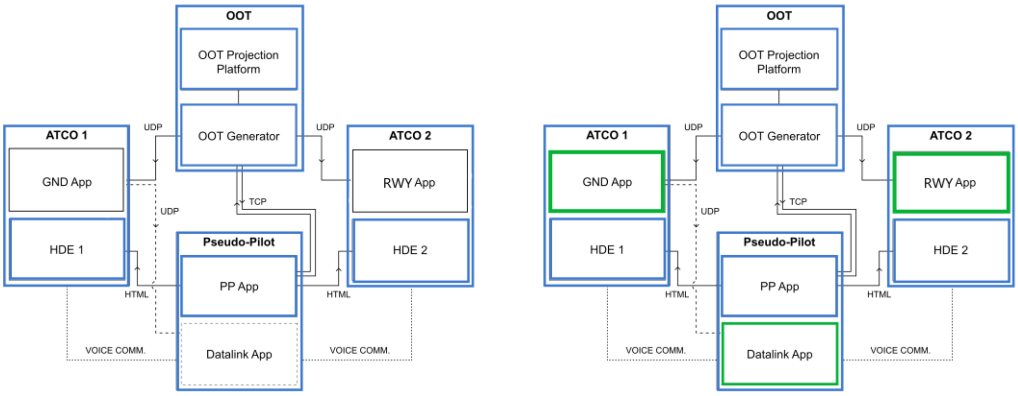


Figure B-21: Validation platform architecture in Reference scenarios (L) and EXE-002-003, EXE-002-005 (R)

The following table summarises all the solution scenarios main aspects and the comparison respect to reference scenarios:

	Reference	Solution
Scenario 1 **(40' VISIBILITY VARIATION:** **15' CONDIVIS1** **10' CONDIVIS2** **15' CONDIVIS3)**	EXE-002-001	EXE-002-002
		EXE-002-003 (VAR+TKL+AG) 15' CONDIVIS1
Scenario 2 **(30' CONDIVIS1 - UNSUAL)**	EXE-002-004	EXE-002-005 (VAR+TKL+SN)

Table B-9: List of reference and solution runs

B.5 Summary of EXE-002 assumptions

The main assumption for EXE-002 are summarized below.

The exercise simulated transitions from one visibility conditions to another in the following sequence CONDIVIS1 to CONDIVIS2 to CONDIVIS3, with a decreasing visibility and an increase in consequent restrictions. The opposite sequence was not evaluated.

The air gestures solution were assessed only in good visibility condition CONDIVIS1 and it was considered only for not-time-critical ground tasks, i.e. pushback and start-up.

The safety net visualization concerned runway incursions and conflicting clearances and was assessed through a technical test.

Identifier	Title	Description	Justification	Impact on Assessment

AS-EXE.002-01	Weather conditions transition	The sequence considered is CONDIVIS1 to CONDIVIS2 to CONDIVIS3	No impact	L
AS-EXE.002-02	Weather conditions for Air Gesture	The air gestures solution will be assessed only in good visibility condition CONDIVIS1	Impact is limited to CONDIVIS1	L
AS-EXE.002-03	Limited scope for Air Gestures	The air gestures solution will be considered only for not-time-critical ground tasks, i.e. pushback and start-up.	Impact is limited to the work of the Ground Controllers.	L
AS-EXE.002-04	Limited assessment for safety net visualization	The safety net visualization will concern runway incursions and conflicting clearances and it will be assessed through a technical test.	The assessment of safety net visualization is limited to a technical test.	L

Table B-10: EXE-002 Assumptions

B.6 Deviation from planned activities

No deviations from the planned activities.

B.7 EXE-002 validation results

B.7.1 Summary of Technological Validation Exercise EXE-002 results

EXE-002 Objective ID	EXE-002 Validation Objective Title	EXE-002 Success Criteria ID	EXE-002 Success Criteria Title	EXE-002 Results	EXE-002 Validation Objective Status
EX2-OBJ-05.971-TRL4-TVALP-FEAS.1010	To confirm the concept is operationally feasible when addressing the Use Case for Virtual or Augmented Reality, tracking labels, and Air Gestures	EX2-CRT-05.971-TLR4-TVALP-FEAS-1011	Identical to CRT-05.971-TLR4-TVALP-FEAS-1011	The experiment confirmed that the concept is operationally feasible when addressing Virtual or Augmented Reality under different visibility conditions with tracking labels.	OK
		EX2-CRT-05.971-TLR4-TVALP-FEAS-1012	Identical to CRT-05.971-TLR4-TVALP-FEAS-1012	The experiment confirmed that the concept is operationally feasible when addressing Air Gestures to issue not-time critical clearances on the TWR GND post. Suggestions were given on how to improve the air gesture functions.	OK

EUROPEAN PARTNERSHIP

Co-funded by the European Union

EX2-OBJ-05.971-TRL4-TVALP-FEAS.1020	Identical to OBJ-05.971-TRL4-TVALP-FEAS.1020.	EX2-CRT-05.971-TLR4-TVALP-FEAS-1021	Identical to CRT-05.971-TLR4-TVALP-FEAS-1021.	No technical showstoppers were experienced in the simulated tower environment. The AR overlays correctly collimated with the simulated out-of-the-tower view, including dynamic objects (i.e. a/c and ground vehicles). The exercise did not investigate the technical feasibility in the real tower environment.	OK
		EX2-CRT-05.971-TLR4-TVALP-FEAS-1022	Identical to CRT-05.971-TLR4-TVALP-FEAS-1022.	The system tested in the prototype did not need integration with other system enablers. The integration within the real environment is out of the scope of this validation	N/A

EX2-OBJ-05.971-TRL4-TVALP-H103.1010	Identical to OBJ-05.971-TRL4-TVALP-H103.1010	EX2-CRT-05.971-TLR4-TVALP-H103.1011	Feedback from controllers (at least 75%) shows that the prototype for V/AR supports controllers in maintain an acceptable level of workload.	Feedback from controllers (90%) shows that the prototype for V/AR supports controllers in maintaining an acceptable level of workload.	OK
		EX2-CRT-05.971-TLR4-TVALP-H103.1012	Identical to CRT-05.971-TLR4-TVALP-H103.1012	ATCOs (90%) provide positive feedback on quantity of information provided by V/A-R. However, only half of the ATCOs provided positive feedback on the quality of the information.	POK

		EX2-CRT-05.971-TLR4-TVALP-H103.1013	Identical to CRT-05.971-TLR4-TVALP-H103.1013	Majority of ATCOs (90%) responses is that V/A-R HMI supports ATCO in maintaining an adequate level of situation awareness.	OK
		EX2-CRT-05.971-TLR4-TVALP-H103.1014	Identical to CRT-05.971-TLR4-TVALP-H103.1014	Measured time spent in head up is increased in the solution scenario with respect to the reference scenario.	OK

		EX2-CRT-05.971-TLR4-TVALP-H103.1015	Identical to CRT-05.971-TLR4-TVALP-H103.1015	Majority of ATCOs (80%) agreed that they always had an adequate field of view when using the V/A-R system to perform their task. However, only half of the ATCOs agreed that the tracking label and the airport overlay provided by V/A-R were adequate and didn't generate confusion neither disturbance. This was due to the labels overlapping and covering the background and not always being aligned clearly with the corresponding A/C's	POK
		EX2-CRT-05.971-TLR4-TVALP-H103.1016	Identical to CRT-05.971-TLR4-TVALP-H103.1016	Six out of ten ATCOs agreed that the V/A-R system did not increase potential for human error compared to current operations.	POK

		EX2-CRT-05.971-TLR4-TVALP-H103.1017	Identical to CRT-05.971-TLR4-TVALP-H103.1017	ATCOs (90%) trust in the system's reliability is at an acceptable level.	OK
		EX2-CRT-05.971-TLR4-TVALP-H103.1018	Identical to. CRT-05.971-TLR4-TVALP-H103.1018	At least 75% of the ATCOs confirmed that the system is easy to use (80%), that the interface was clear and complete (80%) and that the device is physically comfortable (80%). However, less than 75% of the ATCOs confirmed that they would like to use the system frequently (50%), that they imagine most people can learn to use the system very quickly (70%), that they felt confident using the system (70%), that there was not too much inconsistency (40%) and the device did not cause any negative physical consequences like eyestrain (70%).	POK

		EX2-CRT-05.971-TLR4-TVALP-H103.1019	Majority of ATCOs (at least 75%) responses show that alarms and alerts in the prototype for V/AR are not too intrusive and support ATCOs in the early detection of ATC critical situations with respect to conflicting clearances and runway incursions.	All ATCOs responses show that alerts in the prototype for V/AR are effective and not intrusive and 90% of the responses show that the alerts support ATCOs in the early detection of ATC critical situations with respect to conflicting clearances and runway incursions.	OK
		EX2-CRT-05.971-TLR4-TVALP-H103.1020	Identical to CRT-05.971-TLR4-TVALP-H103.1020	ATCOs (80%) provide positive feedback on acceptance of V/A-R tool.	OK

		EX2-CRT-05.971-TLR4-TVALP-H103.1021	Identical to CRT-05.971-TLR4-TVALP-H103.1021	All ATCOs responses indicate that V/A-R HMI supports ATCO team (GND and TWR) in maintaining an acceptable level of situation awareness.	OK
EX2-OBJ-05.971-TRL4-TVALP-H103.1030	Identical to OBJ-05.971A-TLR4-TVALP-H103.1030	EX2-CRT-05.971-TLR4-TVALP-H103.1031	Identical to CRT-05.971-TLR4-TVALP-H103.1031	Majority of ATCOs (90%) responses indicate that ATCOs can apply operating methods in an accurate, efficient and timely manner.	OK

		EX2-CRT-05.971-TLR4-TVALP-H103.1032	Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in the investigated operating conditions.	Only half of the ATCOs indicated that operating methods are clearly identified and consistent in the investigated operating conditions.	NOK

EUROPEAN PARTNERSHIP

Co-funded by the European Union

EX2-OBJ-05.971-TRL4-TVALP-H103.1040	Identical to OBJ-05.971-TRL4-TVALP-H103.1040	EX2-CRT-05.971-TLR4-TVALP-H103.1041	Validation activities show that ATCOs (at least 75%) give positive feedback on job satisfaction and acceptance regarding the prototype for V/AR tracking labels and overlays.	Validation activities show that (85%) of the ATCOs answers with regard to job satisfaction and acceptance regarding the prototype for V/AR tracking labels and overlays are positive	OK
EX2-OBJ-05.971-TRL4-TVALP-H104.1010	To assess that the technical systems for V/A-R Air Gestures support the ATCOs in performing	EX2-CRT-05.971-TLR4-TVALP-H104.1011	Identical to CRT-05.971-TLR4-TVALP-H104.1011	Minority of ATCOs (40%) responses is that V/A-R Air Gestures supports ATCO in maintaining workload at acceptable level. Several ATCOs had difficulties using Air Gestures which increased their workload.	NOK

	the tasks under investigation.	EX2-CRT-05.971-TLR4-TVALP-H104.1012	ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of the interaction means supported by V/A-R Air Gestures.	Being a technical test only, this objective was addressed only through qualitative feedback collected during the debriefings. ATCOs mentioned it several times during the debriefings that they had difficulties using Air Gestures as the system did not always recognise their gestures.	NOK
		EX2-CRT-05.971-TLR4-TVALP-H104.1013	Identical to CRT-05.971-TLR4-TVALP-H104.1013	Only 60% (3 out of 5) of ATCOs responses is that V/A-R Air Gestures HMI supports ATCO in maintaining an adequate level of situation awareness. This was again due to the fact that they were not always able to give the clearance with the air gesture.	NOK

		EX2-CRT-05.971-TLR4-TVALP-H104.1014	Identical to CRT-05.971-TLR4-TVALP-H104.1014	Measured time spent in head up is increased in the Air Gesture solution scenario with respect to the reference scenario.	OK
		EX2-CRT-05.971-TLR4-TVALP-H104.1015	Identical to CRT-05.971-TLR4-TVALP-H104.1015	40% (2 out of 5) of the ATCOs responses it that the V/A-R Air Gestures increase the potential for human error due to usability issues.	NOK
		EX2-CRT-05.971-TLR4-TVALP-H104.1016	Laboratory tests show that ATCOs' (at least 75%) trust in the prototype for V/AR Air Gestures is at an acceptable level.	Laboratory tests show that ATCOs' (80%) trust in the prototype for V/AR Air Gestures is at an acceptable level.	OK

		EX2-CRT-05.971-TLR4-TVALP-H104.1017	Identical to. CRT-05.971-TLR4-TVALP-H104.1017	Only 20% (1 out of 5) of the ATCOs responded that the V/A-R Air Gestures have no impact on the usability whereas the other 80% (4 out of 5) percent believes that usability is negatively impacted, mostly from an ergonomic point of view.	NOK
		EX2-CRT-05.971-TLR4-TVALP-H104.1018	Identical to CRT-05.971-TLR4-TVALP-H104.1018	One ATCO thought the air gestures should be removed and two ATCOs pointed out that it should be avoided to use air gesture commands for runway authorizations/critical cases.	POK
EX2-OBJ-05.971-TRL4-TVALP-H104.1020	Identical to OBJ-05.971A-TLR4-TVALP-H104.1020	EX2-CRT-05.971-TLR4-TVALP-H104.1021	Identical to CRT-05.971-TLR4-TVALP-H104.1021	No specific impacts of Air Gesture on operation methods were mentioned by the ATCOs.	OK

		EX2-CRT-05.971-TLR4-TVALP-H104.1022	Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in the investigated operating conditions.	No specific impacts of Air Gesture on operation methods were mentioned by the ATCOs.	OK

EUROPEAN PARTNERSHIP

Co-funded by the European Union

EX2-OBJ-05.971-TRL4-TVALP-H104.1030	Identical to OBJ-05.971-TRL4-TVALP-H104.1030	EX2-CRT-05.971-TLR4-TVALP-H104.1031	Laboratory tests show that ATCOs (at least 75%) give positive feedback on job satisfaction and acceptance regarding the prototype for V/AR Air Gestures.	Laboratory tests show that ATCOs (80%) give positive feedback on job satisfaction and acceptance regarding the prototype for V/AR Air Gestures.	OK
EX2-OBJ-05.971-TLR4-TVALP-SAFE.1010	To give relevant input about safety issues when using V/AR with safety nets	EX2-CRT-05.971-TLR4-TVALP-SAFE-1011	Laboratory tests show that the prototype for V/AR with safety nets improves the safety performance by reducing human error.	Laboratory tests show that the prototype for V/AR with safety nets improves the perceived safety performance by reducing human error. See also results for EX2-CRT-05.971-TLR4-TVALP-H103.1016.	OK

		EX2-CRT-05.971-TLR4-TVALP-SAFE-1012	Identical to CRT-05.971-TLR4-TVALP-SAFE-1012	ATCO's workload with the implementation of Virtual/Augmented Reality applications is maintained at acceptable level and therefore not reducing safety levels. See also results for EX2-CRT-05.971-TLR4-TVALP-H103.1011.	OK
		EX2-CRT-05.971-TLR4-TVALP-SAFE-1013	Identical to CRT-05.971-TLR4-TVALP-SAFE-1013	ATCO's situational awareness with the implementation of Virtual/Augmented Reality applications is maintained at acceptable level and therefore not reducing safety levels. See also results for EX2-CRT-05.971-TLR4-TVALP-H103.1013.	OK
		EX2-CRT-05.971-TLR4-TVALP-SAFE-1014	Identical to CRT-05.971-TLR4-TVALP-SAFE-1014.	This will be addressed at solution level taking on board the EXE2 safety evidences.	N/A

EX2-OBJ-05.971-TLR4-TVALP-PERF.1010		EX2-CRT-05.971-TLR4-TVALP-PERF-1011	Laboratory tests show that the prototype for V/AR contributes to an assessment of Cost Efficiency performance by having a positive impact on situation awareness, workload and efficiency of ground operations.	Laboratory tests showed that the prototype for V/AR contributes to an assessment of Cost Efficiency performance by having a positive impact on situation awareness, workload and efficiency of ground operations, especially in low visibility conditions.	OK

		EX2-CRT-05.971-TLR4-TVALP-PERF-1012	Identical to CRT-05.971-TLR4-TVALP-PERF-1012	Laboratory tests show that the use of V/A reality applications improves Resilience by increasing situational awareness in Low visibility conditions while maintaining workload within acceptable limits.	OK

Table B-11: Technological Validation Results EXE-002

B.7.1.1 Results on technological feasibility

- V/A-R Tracking labels and airport overlays in all weather conditions (including LVC)
    - V/AR tracking labels correctly collimated to the simulated out-of-the-tower view in all weather conditions and within the transitions between different visibility conditions. All the other fixed overlays were also correctly collimated to the out-of-the-tower view. This was proven for both posts (i.e. TWR RWY and TWR GND) when the users were in fixed position as well as when they moved or stood up.
    - Each label was correctly tracked onto the associated flight and, since only simulated positioning data were considered, the labels movement were smooth even if no interpolation algorithms were considered for the positioning data. Overall, the labels and their content were always visible and readable, however a few suggestions were provided by ATCOs in order to improve the readability while reducing the clutter.
    - The adaptivity of both the tracking labels and the other overlays with respect to the visibility conditions, phase of flight, runway occupancies and LVP proved to work correctly and timely, with no latency and no showstoppers along the five days of simulations as well as during the TAT and OAT.
    - As far as the mixed reality hardware is concerned, most controllers considered it acceptable on an ergonomic perspective. However, a few operators had difficulties writing on paper strips while wearing MS HoloLens. This issue seems to be compounded by the use of glasses for far-sighted people which seems not to be manageable in one case. This aspect shall be further investigated.
    - The use of the mixed reality hardware microphone and headsets instead of the traditional tools proved to be feasible from a technical perspective.

- V/A-R Air Gesture
    - The use of V/A-R Air Gesture to interact with overlays proved to be feasible from a technical perspective. However, a few ATCOs had difficulties when using the air gestures to issue not-time critical clearances. This aspect shall be investigated considering the impact a specific training on air gestures might have onto the user's capability to correctly use the air gesture function.

B.7.1.2 Results per KPA

Method

In total, 10 ATCOs participated in the validation exercise. The average years of experience was 18,4 years, with a minimum of 6 and a maximum of 33 years. Three ATCOs were between 30 and 39 years old, four between 40 and 49 and three between 50 and 59. Six of the ATCOs are working at Bologna Airport, whereas the others are currently working at Torino, Forlì, Ancona and Rimini. For the purpose of the validation, each ATCO was either assigned to the ground controller or the runway controller position. There was no rotation of the ATCOs amongst the positions because it was important for the ATCO to experience the different technologies (tracking labels, air gesture, safety net) from the same position in order to make a good comparison between reference and solution. Although one ATCO only occupied one position, there were in total five ATCOs in each controller position, ensuring a comprehensive assessment of the concept from different perspectives and for different aspects.

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

- Head-Down Time
- Number of Switches Head Down/Head Up
- Situational Awareness
- Workload
- Acceptability & Job Satisfaction
- Trust
- Usability
- Teamwork & Communication
- Roles & Responsibilities
- Human Error
- Ergonomics
- Throughput

Head down time and the number of switches between head down and head up were measured on-line by the validation platform. The other 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-simulation Questionnaires) and interviews during debriefing at the end of the run and at the end of the exercise, while throughput was collected online during the simulation. See the table below for the daily agenda.

from	*To*	*Activity*
09:00	09:10	Welcome

09:10	09:30	Briefing
09:30	10:00	Platform familiarization (training)
10:00	10:40	RUN 001 (Reference)
10:40	11:00	Post-run questionnaire + debriefing
11:00	11:15	Coffee break
11:15	11:55	RUN 002 (V/A-Reality + Tracking Labels)
11:55	12:15	Post-run questionnaire + debriefing
12:15	12:30	RUN 003 (V/A-Reality + Tracking Labels + Air Gesture)
12:30	13:00	Debriefing
13:00	14:30	Lunch break
14:30	15:00	RUN 004 (Reference with safety event)
15:00	15:20	Post-run questionnaire + debriefing
15:20	15:35	Coffee break
15:35	15:55	RUN 005 (V/A-Reality + Tracking Labels + Safety Net)
15:55	16:15	Post-run questionnaire
16:15	16:45	Post-exercise questionnaire + final debriefing

Table B-12: Validation exercise planning

As can be seen from the Table, the ATCOs participated in five different runs:

- A baseline run with the reference scenario
- A solution scenario with Tracking Labels

- A solution scenario with Tracking Labels and Air Gesture (the Air Gesture was only applied to the ground controller position)
- A baseline run with the reference scenario including an unusual event
- A solution scenario with Tracking Labels and Safety Net including an unusual event

As can be seen in the Table, there was no post-run questionnaire for the run with Air Gesture. Due to its lower level of maturity, only qualitative data was collected. Moreover, the run with Air Gesture only lasted 15 minutes so it would not be reasonable to compare the post-run qualitative data to the data collected after the 40-minute reference scenario run.

The post-run questionnaire contained 8 questions, including the Bedford for workload, China Lake questionnaire for Situational Awareness, and the CARS for user acceptance. The Bedford Scale is a unidimensional rating scale designed to identify operator's spare mental capacity while completing a task. The single dimension is assessed using a hierarchical decision tree that guides the operator through a ten-point rating scale, each point of which is accompanied by a descriptor of the associated level of workload; task abandoned, extremely high workload with no spare capacity, very high workload with almost no spare capacity, very high workload with almost no spare capacity but no impact to the primary task, little spare capacity, reduced spare capacity, insufficient spare capacity for easy attention to additional tasks, enough spare capacity for all desirable additional tasks, workload low, workload insignificant. The China Lake is also a hierarchical decision tree, with the following ten-point rating scale: SA was too low, SA was very low, SA was low and unaware of most of the information required to perform the task effectively, SA was low and unaware of half of the information required to perform the task effectively, SA was insufficient, SA was reduced, SA was not complete, SA was good, SA was very good, SA was excellent. Similarly, the CARS consisted of a ten-point rating scale: Improvement mandatory, major flaws and considerable operator compensation needed to maintain safe operations, major flaws and some operator compensation needed to maintain safe operations, very annoying flaws, moderately annoying flaws, minor but annoying flaws, mildly unpleasant flaws, negligible flaws, flaws are rare.

The results analysis compared the collected data for the solution scenario with tracking labels (run 2) to that for the reference scenario (run 1), the solution scenario with safety event and safety net (run 5) to the reference scenario with safety event (run 4) and the solution scenario with air gesture (run 3) to the (first 15 minutes of the) reference scenario (run 1).

The post-exercise questionnaire contained 38 questions. Additionally, during the final debriefing, a Want/Have Matrix and a Human Performance Impact Matrix were used to collect data, see figures below. The purpose of the validation exercises is to move from TRL 2 to TRL 4 and the Want/Have Matrix and a Human Performance Impact Matrix were used to let ATCOs envisioning next needed development of the technologies and to review the benefit / impact mechanism at HP and SAF levels. 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 want to avoid (even if not experimented) for the technology. The Want/Have and the Human Performance Impact matrixes were judged as appropriate tools fitting well with the level of maturity. The templates of these matrixes can be found below, whereas the results collected by using the matrixes can be found in section B.8.3.5 (Final Debriefing evidences).

Figure B-22: Want/Have Matrix

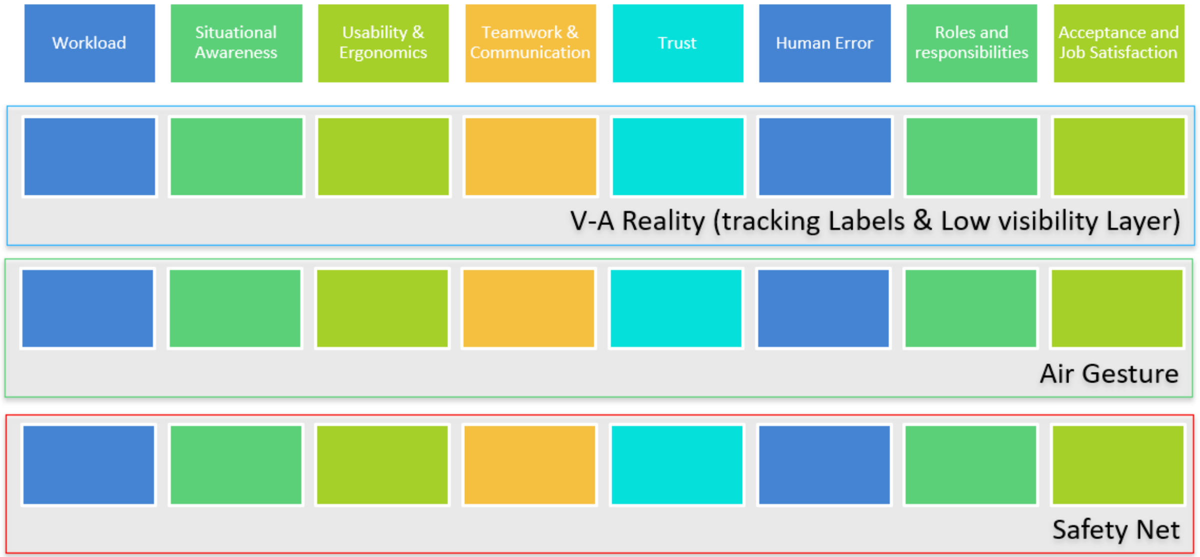


Figure B-23: Human Performance Impact Matrix

Performances

- V/A-R Tracking labels and airport overlay in LVC
    - Considering the results on Human Performance (see Human Performance Impact Matrix), the subjective workload was said not to be affected by the V/A-R Tracking labels and airport overlay. The lack of a decrease in subjective workload might be due to the label design, see the Human Performance section for more details. The Safety Net positively influenced subjective workload. Situational awareness was reported to increase thanks to the solutions.
- V/A-R Air Gesture
    - Workload was negatively affected by the Air Gesture solutions relative to the reference whereas the situational awareness seems unaffected. Thus, from the point of view of workload, the system's performance might be lower compared to the reference without Air Gesture. This might be due to the usability issues related to the use of air gestures.

Human Performance

- V/A-R Tracking labels and airport overlay in LVC
  - As it can be seen in the Want/Have Matrix, significant margin of improvements were related to the same aspect of tracking labels: Background and Overlapping (labels and external view). Six ATCOs wrote down that the overlapping of the labels should be removed and four ATCOs said that it should be avoided to have the labels blocking the view, for example by removing the background colour or rotating the labels. Similarly, one ATCO commented that there should be no labels blocking the final approach path. The problems related to the labels were a technical limitation of the system. When positioning the labels in the future, it should also be taken into account that labels should not be placed too far away from the A/C and that they should not cover AA/CC (including those AA/CC with which the ATCO is not in contact or AA/CC without a flight plan).
  - Further feedback were provided in relation with the provided information of TL. Two ATCOs indicated that the arrival altitude information does not need to be presented on the labels whereas one ATCO wrote down that distance and altitude data should be preserved. Additionally, one ATCO said that he would like to see the parking information for the A/C when it's in final approach, three ATCOs wrote down that they would like information on clearances and clearances in progress and another ATCO suggested to add the speed of a/c taxiing and landing. Four ATCOs warned that it should be avoided to present too much information and create visual noise.
  - Apart from the above-mentioned recommended improvements, the ATCOs were generally positive about the labels and the info they provided, which four of them specifically wrote down in the category of preserve. However, it was mentioned multiple times that fixed data should be removed, and replaced by customised and real-time data. For example, the weather data was found to be useful but should be constantly updated and show changes of the weather. The option to customise the data that is presented on the label could potentially address the disagreements amongst ATCOs as to which data they think should be presented on the labels. One ATCO even suggested that it should be possible to adjust the colour saturation and label size. Apart from providing data, the label was also said to be useful for highlighting where the plane is in low visibility conditions. In addition to the line between the label and the plane, labels could be indicated by a symbol in Low Visibility Conditions to further improve the highlight of the a/c. One ATCO suggested that there should be labels for other vehicles as well.
  - ATCOs mentioned that the HoloLens limits the augmented visual range and as a consequence they felt overstimulated to move their head in order to see the augmented information. When users move their head, their augmented field of view changes depending on the direction in which they look. As two ATCOs mentioned, there should be indicators that more information is available outside of the current visual field. Moreover, one ATCO said that the weather data should always be within the augmented field of view for the ground controller.

- ○ Whereas one ATCO complained about the low brightness of the HoloLens and one said that it should be avoided to further reduce the brightness of the HoloLens in the future, two other ATCOs said that it was too bright. One ATCO wrote down that the taxiway does not need to be illuminated, only the runway and the stop bars. Six ATCOs wrote down that the highlighting of the runway, holding points and taxiways in low visibility conditions should be preserved/achieved. Currently, the runway and main taxiway outlines appear during low visibility conditions (whenever the RVR value decreases below 1500m). The taxiway outline (T in Bologna) is coloured blue, while the runway changes colour based on the traffic currently occupying it: an approaching a/c in short final turns the outline colour yellow, while a departing traffic turns the contour cyan. In both cases, the colour of the outline turns back to white after the traffic clears the area.
- ○ ATCOs were very positive about the safety net tools, indeed five ATCO's wrote down that this feature should be persevered. The safety net tool was said to benefit Situational Awareness, as it helped the ATCOs to immediately recognise a hazard. The acoustic cues guide their attention and because both ATCOs (ground and runway) receive the notification, there is no need to communicate it and the team SA is enhanced.
- ○ According to one ATCO, it should be avoided to use the tool to increase capacity in Low Visibility Conditions, at least not without adding additional supporting equipment.
- ○ The ATCOs had a lot of suggestions on things that could be achieved in the future tool. First of all, they vowed for the integration with other systems and to remove old working methods such as the paper strip. Examples of systems that should be integrated are: Time-based separation tool to support the separation of Arrivals and Departures, complete radar, Automatic Dependence Surveillance and data-link. Two ATCOs also suggested to implement speech recognition. Finally, four ATCOs suggested to extend the (vertical) view (above the tower, the parking, etc.).
- ○ As can be seen from the Human Performance Matrix, Human Performance was generally said to be either not effected or positively affected by the V/A-R technology. Especially, in combination with the Safety Net, the tool seems to have a positive impact. The Human Performance element that was said to be most negatively affected is Usability & Ergonomics. This is in line with the comments related to the label design and the field of view found in the Want/Have matrix.

- V/A-R Air Gesture

  - ○ Whereas two ATCOs wrote down in the Want/Have Matrix that the air gestures should be preserved, one ATCO though it should be removed from the concept. In the future, ATCOs said it should be avoided to use air gesture commands for runway authorizations and critical cases.
  - ○ As can be seen from the Human Performance Matrix, Human Performance was said to be impacted negatively by the Air Gestures. Mainly Workload, Usability & Ergonomics, and Human error seem to be negatively affected by the tool.

EUROPEAN PARTNERSHIP
Co-funded by the European Union

Safety

- V/A-R Tracking labels and airport overlay in LVC
  - Considering the results on Human Performance (see Human Performance Impact Matrix), the perceived potential for Human Error decreased thanks to the V/A-R system for ground controllers, whereas runway controllers did not experience a clear decrease or increase. However, both controller positions benefited from an improvement in terms of Human Error thanks to the Safety Net.
- V/A-R Air Gesture
  - The Air Gesture did not clearly impact the perceived potential for Human Error. However, it was reported to increase workload, which could lead to a decrease of the level of safety.
- Safety Subjective Results against SAP identified HAZARDS

HZ ID	Hazard	Hazard Impact	Proposed evidence collection method/technique	EXE-002 (ENAV) Results
Hz1	Failure of the V/AR system (e.g. freezing of Tracking Labels or perceptual cues complete loss) prevents the AR device from being updated.	V/AR system is unresponsive potentially impacting trajectory management and associated safety nets.	Subjective: observation and feedback from ATCOs	No failure of the V/A-R system identified during the exercise. ATCOs only commented on the design, especially the position, of the Tracking Labels.
Hz2	TL is erroneously associated to a wrong a/c (wrong information)	ATCo may focus on the wrong a/c and issues the clearance intended for another a/c.	Subjective: observation and feedback from ATCOs	One ATCO mentioned that the TL was not always well aligned with the corresponding a/c. Moreover, ATCOs commented repeatedly that the TLs were covering each other as well as the background. However, these issues did not lead to errors in associating the right TL to the a/c.
Hz3	Presentation of TL information within the V/AR does not support ATCo in task execution	If the presentation of TL on the AR device is inadequate (e.g. TLs overlapping, size issues – e.g. depth).	Subjective: observation and feedback from ATCOs	ATCOs commented repeatedly that the TLs were covering each other as well as the background. However, ATCOs were able to execute their tasks without too much difficulty (subjective SA and workload were acceptable).

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Hz4	The responsiveness of the V/AR system as a result of ATCO's input/gesture or flight information changes) is inadequate for the accomplishment of operations.	If the use of V/AR introduces delays in the display of information, this may cause the ATCo to focus on V/AR until verified that the information has been correctly processed and displayed. This may have negative impact on ATCo situational awareness.	Subjective: observation and feedback from ATCOs	This hazard has not been identified in the exercise. ATCO subjective situational awareness was above the tolerable threshold.
Hz5	V/AR system fails to identify inputs (gestures) – no response	Unidentified inputs as a result of correct gestures may distract the ATCo from the primary task of ATS provision and results in temporary workload increase and reduction in situational awareness as failures in TL inputs may require increased concentration on the V/AR system.	Subjective: observation and feedback from ATCOs	Air gestures were not always recognised easily. This led to frustration, and in some cased decreased SA and increased workload, for the ATCO.
Hz6	V/AR system wrongly identifies input (gesture) – provides erroneous output	In isolated instances the erroneous recognition of an input has no significant safety impact as the ATCo is able to correct the input. If the TV/AR system fails to perform in accordance with the specified gesture recognition threshold this may have a human	Subjective: observation and feedback from ATCOs	Air gestures were not always recognised easily. This led to frustration, and in some cased decreased SA and increased workload, for the ATCO.

		performance impact causing disruption to the expected workflow and cognitive processes.		

Table B-13: Safety Subjective Results against SAP identified hazards

EUROPEAN PARTNERSHIP

Co-funded by the European Union

B.7.2 Analysis of EXE-002 Results per Technological Validation objective

B.7.2.1 EX2-OBJ-05.971-TRL4-TVALP-FEAS.1010 Results

To confirm the concept is operationally feasible when addressing the Use Case for Virtual or Augmented Reality, tracking labels, and Air Gestures

EX2-CRT-05.971-TLR4-TVALP-FEAS-1011

No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Virtual or Augmented Reality and tracking labels.

EX2-CRT-05.971-TLR4-TVALP-FEAS-1012

No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Air Gestures.

B.7.2.2 EX2-OBJ-05.971-TRL4-TVALP-FEAS.1020 Results

To identify possible technical feasibility issues and possible show stoppers.

EX2-CRT-05.971-TLR4-TVALP-FEAS-1021

Laboratory tests (based on a prototype) have verified the technical feasibility of the use of V/AR applications in the tower environment.

EX2-CRT-05.971-TLR4-TVALP- FEAS-1022

Laboratory tests have verified that the integration of the V/AR applications with other related system enablers is technically feasible.

B.7.2.3 EX2-OBJ-05.971-TRL4-TVALP-H103.1010

To assess that the technical systems for V/A-R Tracking Labels and overlays support the ATCOs in performing their tasks.

EX2-CRT-05.971-TLR4-TVALP-H103.1011

Feedback from controllers (at least 75%) shows that the prototype for V/AR supports controllers in maintaining an acceptable level of workload.

In the post-run questionnaire, controllers reported an average workload of 2,7 (on a 10-point scale) for runs involving augmented reality, compared to 2,6 for reference scenario runs. This means that workload was satisfactory without reduction, as the answers were within the acceptable level of workload (i.e., 5). The runway controller benefitted more than the ground controller from the solution, maybe because the solution is helping in the critical landing and departing phases and it might be interpreted that the solution was not so effective for the ground controller because of the overlap of the labels. One controller commented that his confidence with the technical environment improved in subsequent runs. See also the graphs below for a visual representation. The first figure shows the numbers of answers (on a 10-point Bedford scale) and the second one shows the average answers in the reference and solution scenarios.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

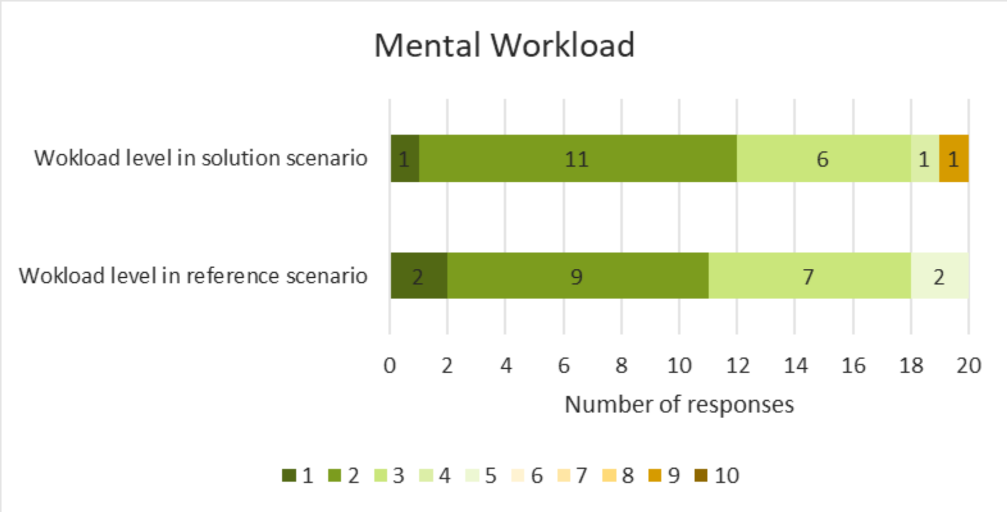


Figure B-24~~24~~: Post-Run questionnaire results related to mental workload

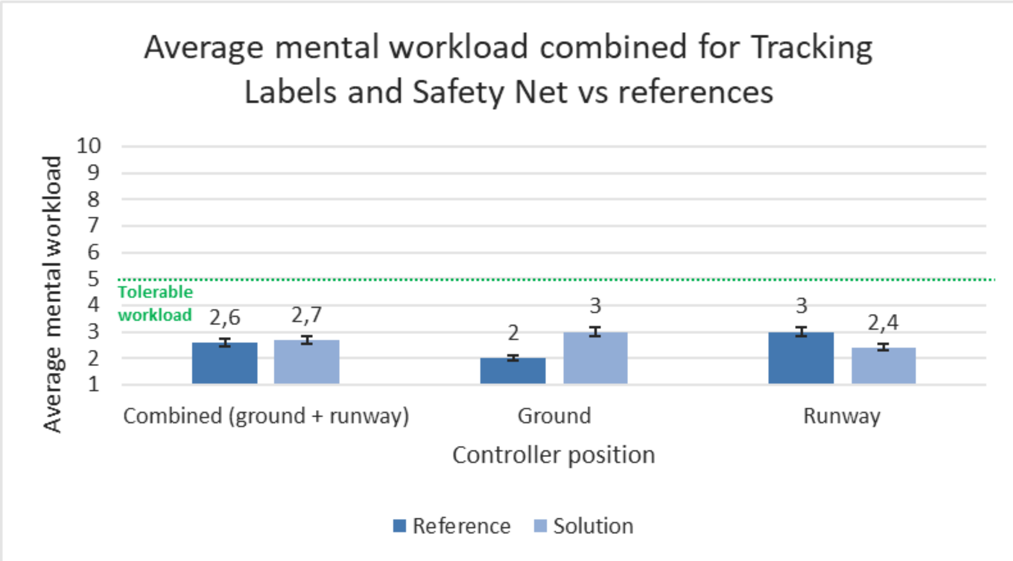


Figure B-25: Average mental workload in the reference and solution scenarios.

In the post-exercise questionnaire, only one controller reported ‘slightly heavy’ workload in the V/A-R scenarios compared to the reference scenarios, whereas the others all reported either very light (1), light (4), slightly light (1) or tolerable (5) workload levels. This is also represented in the figures below. Both figures show the distribution of answers among the different answer options. More specifically, the first graph shows the number of answers collected for each point of the 7-point Likert scale. The second graph shows the distribution of answers in percentage for both the ground and runway controller separately. Controllers’ comments indicated that the prototype for V/AR might provide a benefit in comparison to the reference situation in low visibility and unusual conditions.

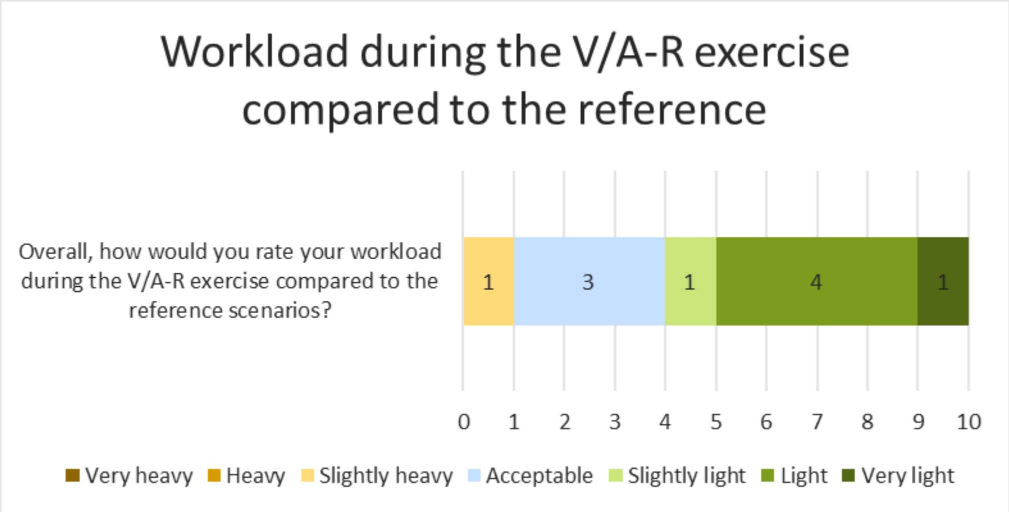


Figure B-26: Post-Exercise questionnaire results related to mental workload

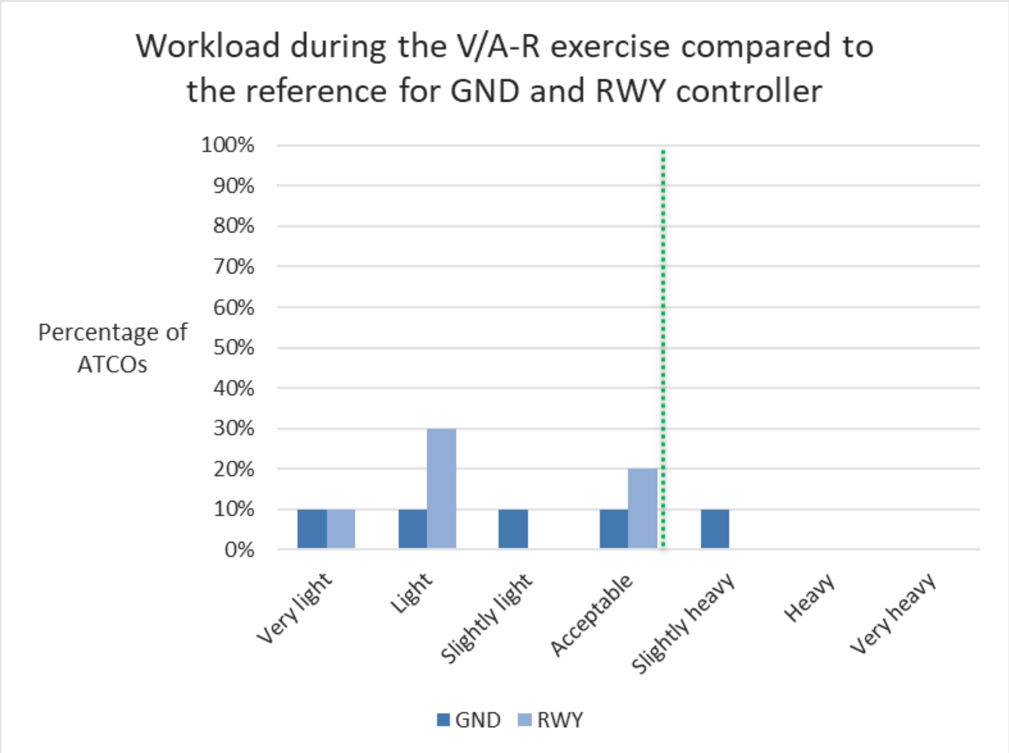


Figure B-27: Post-Exercise questionnaire results related to mental workload for ground (GND) and runway (RWY) controller position

During the debriefings, controllers mentioned that the technology made the tasks, such as push back, intuitive and fast. Workload related to communication is decreased because it is not necessary to speak as much compared to the reference scenario. Moreover, all information that is needed can be found in a concentrated place. However, controllers mentioned that this could lead to more fatigue at they are constantly monitoring the information in front of them. To not increase the workload, the amount and the position of the data should be taken into account.

Furthermore, 15 out of 20 responses on the post-run questionnaire indicated that the *physical* workload was acceptable, slightly light, light or very light in the solution scenarios (see figure below for the number of answers collected for each point of the 7-point Likert scale).

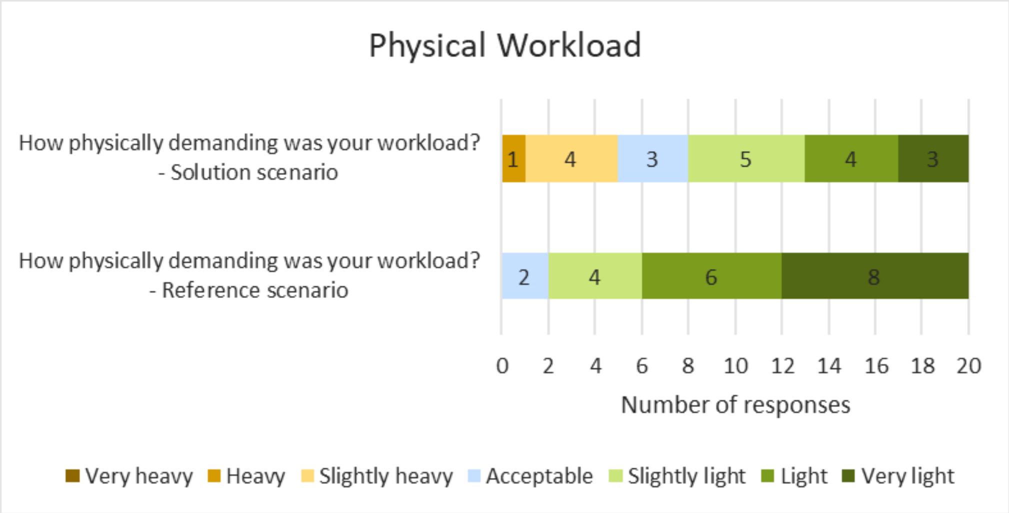


Figure B-28: Post-Run questionnaire results related to physical workload

After assigning numerical values to the answers (very light = 1, light = 2, and so on) the averages were calculated (see figures below). The average physical demand was 2,1 in the reference scenario and 3,4 in the solution scenario. Only the average physical workload of ground controllers in the traffic label solution scenario exceeded the acceptable workload level of 4. ATCOs had to get used to the helmet. Some said it was quite heavy and the position of the data forced them to rotate their head all the time.

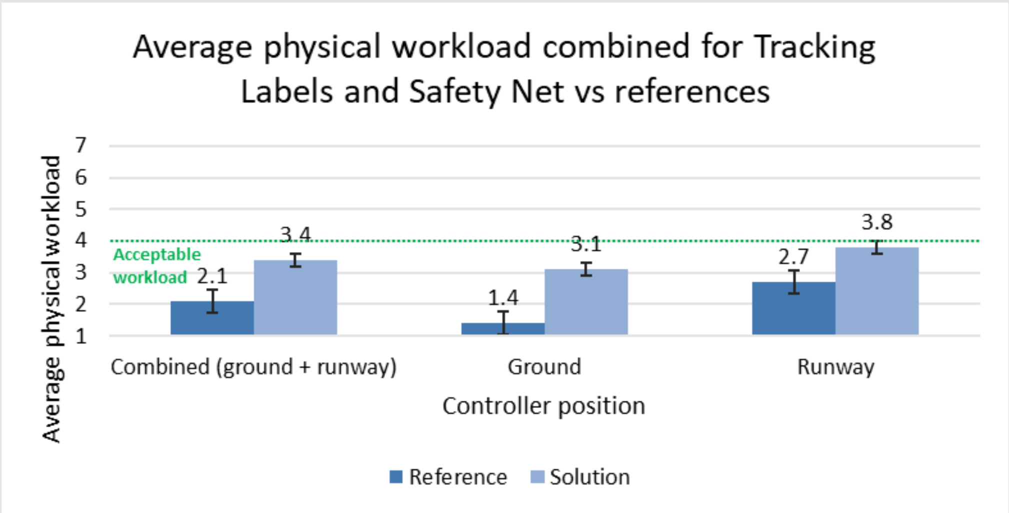


Figure B-29: Average physical workload in the reference and solution scenarios

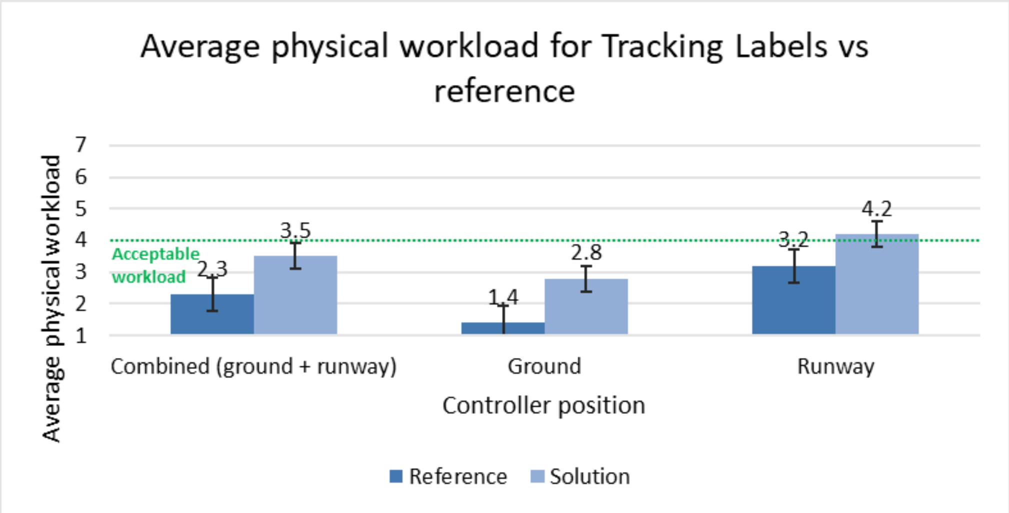
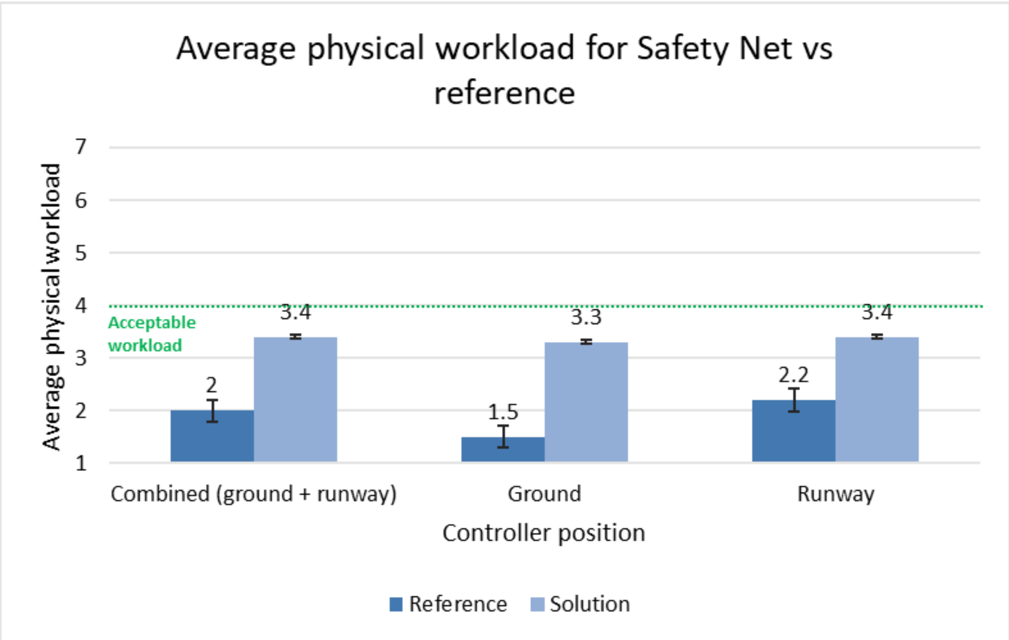


Figure B-30: Average physical workload in the reference and solution scenarios for tracking labels



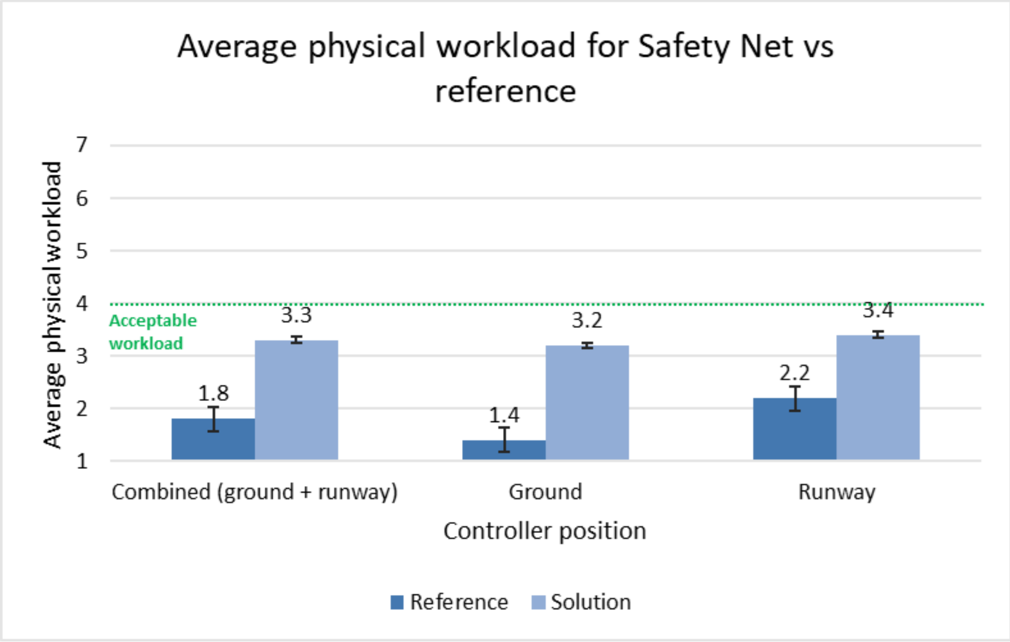


Figure B-31: Average physical workload in the reference and solution scenarios for safety net

EX2-CRT-05.971-TLR4-TVALP-H103.1012

ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R

- Adequacy: Whereas five out of ten ATCOs agreed in the post-exercise questionnaire that the tracking label and the airport overlay provided by V/A-R were adequate and didn't generate confusion neither disturbance, four of them disagreed and one neither agreed nor disagreed ATCOs commented that the labels were overlapping and covering the background and that they were sometimes badly aligned. Also during the debriefing, ATCOs pointed out that the background colour of the label should be removed and that it should be prevented to have labels overlapping each other or the final approach path.
- Quality: Five out of ten ATCOs agreed that the *quality* of the information provided by the V/A-R was adequate, clear and did not disturb them. Two ATCOs were undecided and two disagreed. Comments on the quality of the information provided were mostly about the overlapping of the labels which reduced the capability for observation. Additionally, there were some comments about the colour saturation of the label being too high.
- Quantity: Almost all ATCOs agreed that the *quantity* of the information provided by the V/A-R was adequate and did not disturb them. Only one ATCO disagreed. ATCOs suggested that more information should be provided regarding clearances and speed during taxi and take-off. They would also like to have parking info for the A/C when it is in final approach. It was also suggested to have a customisable set-up.
- See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately). Generally, the runway controllers answered more positively than the ground controllers. The aforementioned issues with regard to the system's adequacy, quality and quantity might particularly influence the tasks of the ground controller.

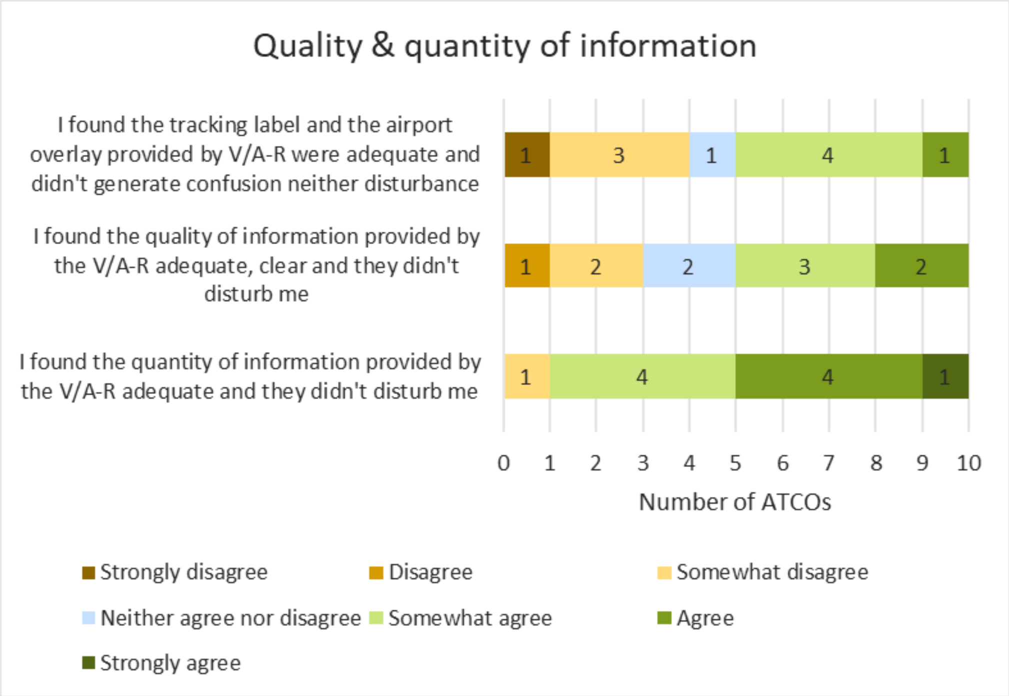


Figure B-32: Post-Exercise questionnaire results related to mental workload

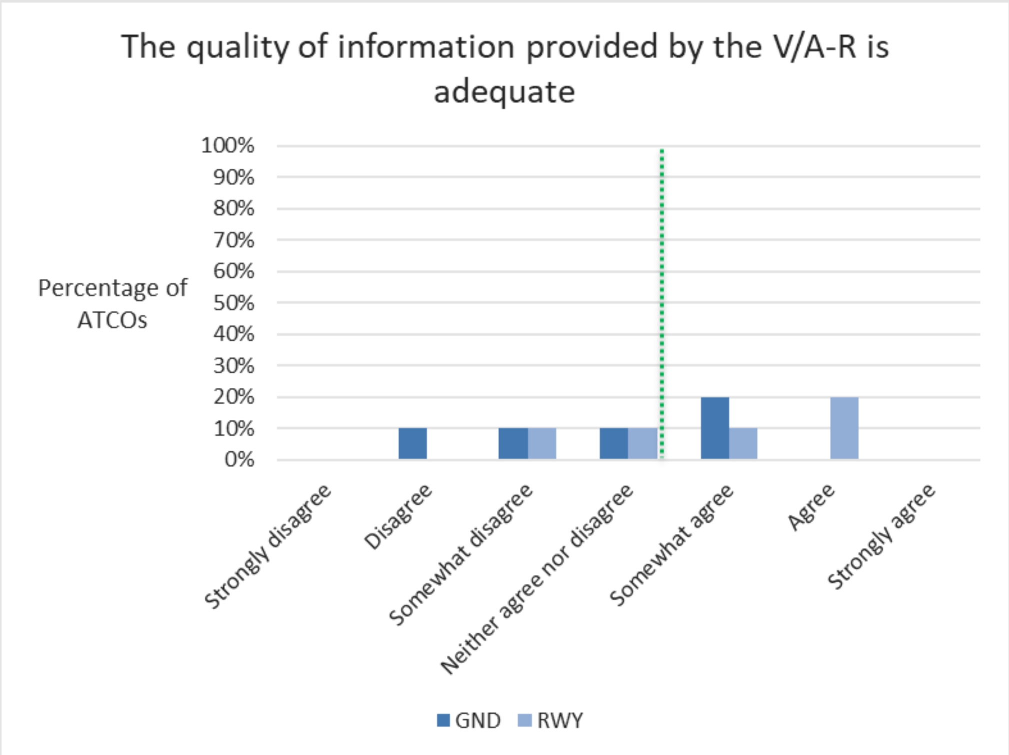


Figure B-33: Post-Exercise questionnaire results related to quality of information for ground (GND) and runway (RWY) controller position

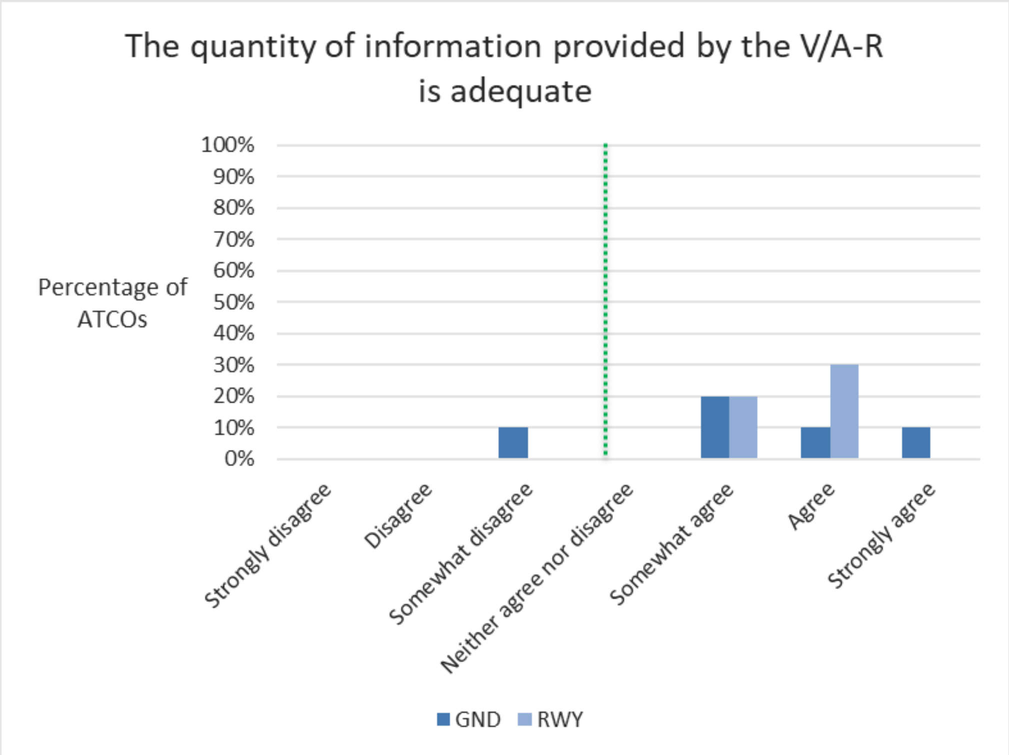


Figure B-34: Post-Exercise questionnaire results related to quantity of information for ground (GND) and runway (RWY) controller position

Further discussion during the debriefings revealed that there was a need for recent (rather than static) data (e.g., with regards to weather information) and changes/refresh of data should be indicated. Ideally, the A/V-R tool would be fully integrated with Automatic Dependence Surveillance in order to have data directly downloaded from the a/c. Furthermore, ATCOs experienced some difficulties to 'find' the information because sometimes the information is displayed outside of their current field of vision. They indicated that the information should be visible regardless of the direction in which they are looking or the visual range should contain cues pointing to information that is currently falling outside of the field of vision. One ATCO also indicated that information can be put at the top of the head up display because they do not need that part of the view.

EX2-CRT-05.971-TLR4-TVALP-H103.1013

Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO in maintaining an adequate level of situation awareness

In the post-run questionnaire, ATCOs reported an average Situational Awareness of 8,2 (on a 10 point scale) for runs involving augmented reality, compared to 9.1 for reference scenario runs . This means that level of SA was satisfactory in both the reference and solution scenarios. See the figures below for the number of answers collected for each point of the 10-point China Lake scale and average situation awareness for different controller positions.

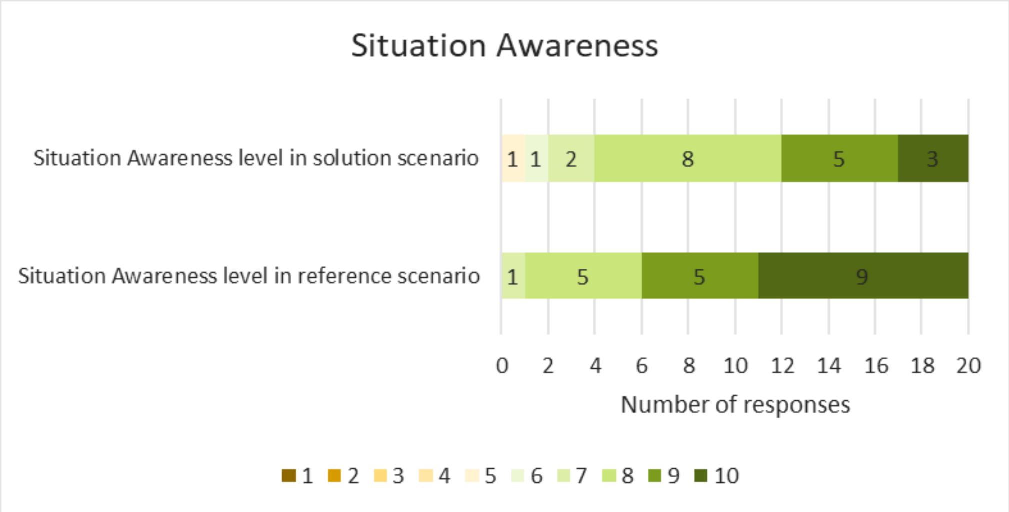


Figure B-35: Post-Run questionnaire results related to Situation Awareness

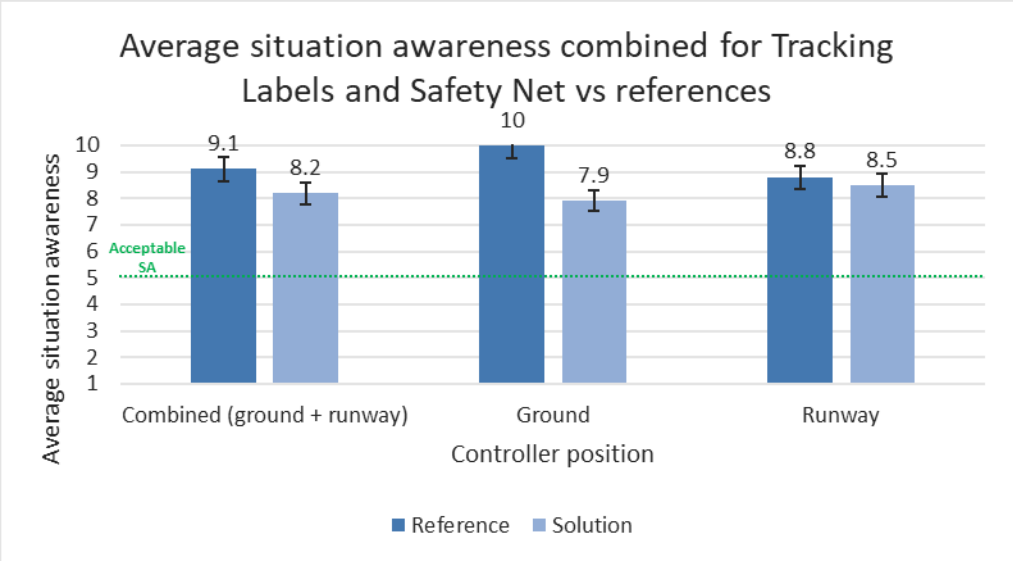


Figure B-36: Average situation awareness in the reference and solution scenarios

In the post-exercise questionnaire, ATCOs were asked to rate their Situational Awareness during the V/A-R exercise compared to the reference scenarios on a 7-point scale ranging from *Situational Awareness lost* to *Situational Awareness perfect*. Eight out of ten ATCOs rated their Situational Awareness during the V/A-R exercise execution compared to the reference scenarios as sufficient (1), high (6) or perfect (1). One ATCO responded that his SA was moderate and one ATCO said SA was degraded. See the figures below for the number of answers collected for each point of the 10-point scale and the distribution of answers in percentage for both the ground and runway controller separately. Generally, runway controllers reported higher SA.

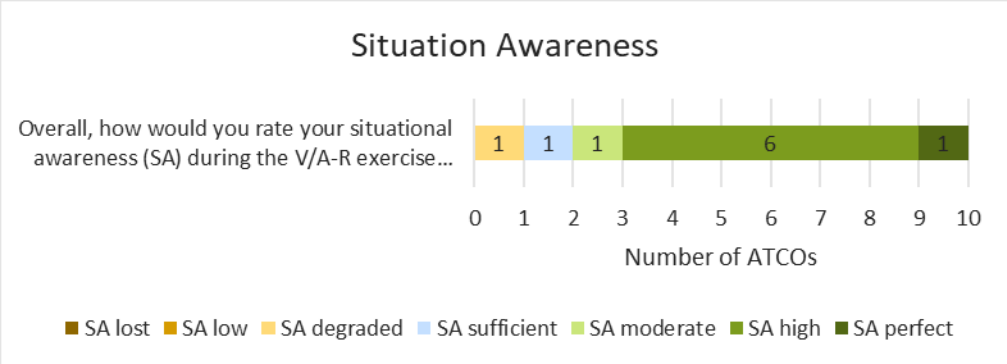


Figure B-37: Post-Exercise questionnaire results related to Situation Awareness.

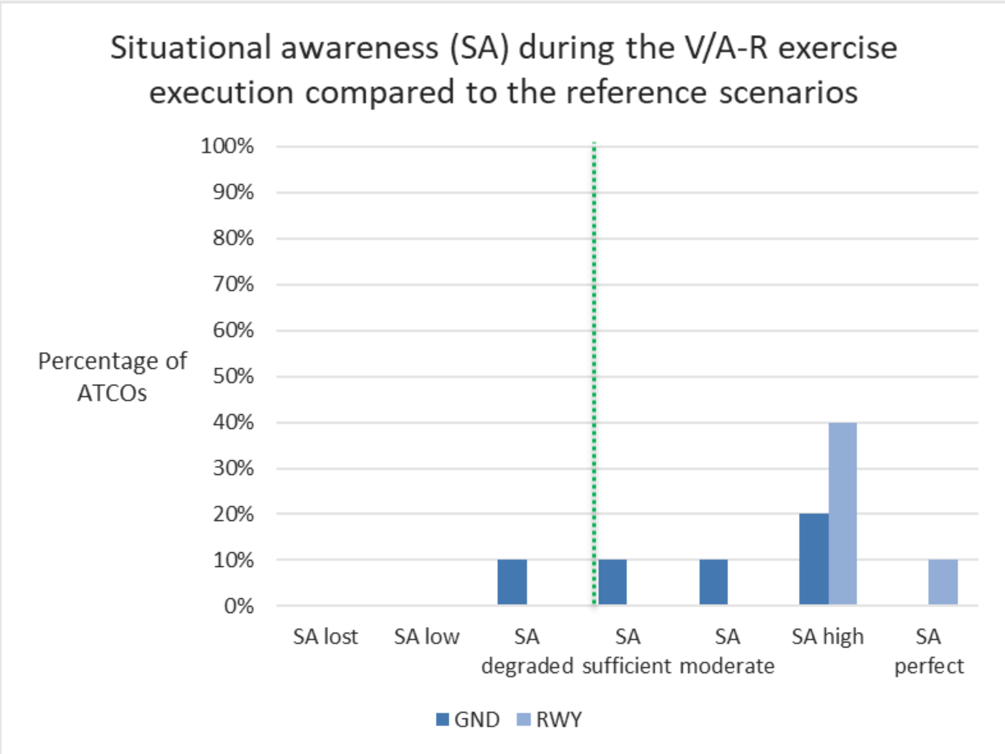


Figure B-38: Post-Exercise questionnaire results related to Situation Awareness for ground (GND) and runway (RWY) controller position

Furthermore, apart from one ATCO, all ATCOs agreed that the information provided by V/A-R improved the SA (during low and good visibility conditions) with respect to the reference situation (see the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately). Thus, it can be concluded that the majority of ATCOs thinks that V/A-R HMI supports ATCOs in maintaining an adequate level of SA. The comments in the questionnaire and the discussions during the debriefing made clear that the (potential) degradations in SA were due to the information representation. Specifically, the labels were said to overlap and cover the view. In addition, missing information and the need to turn the head towards the traffic in order to see the information were mentioned as causes for degraded SA. Thus, the feedback is strongly related to the quality and quantity of the presented information. Nevertheless, ATCOs pointed out that the technology can also benefit SA, especially during unusual situations. One ATCO mentioned that the colours helped him to maintain SA.

Co-funded by the European Union

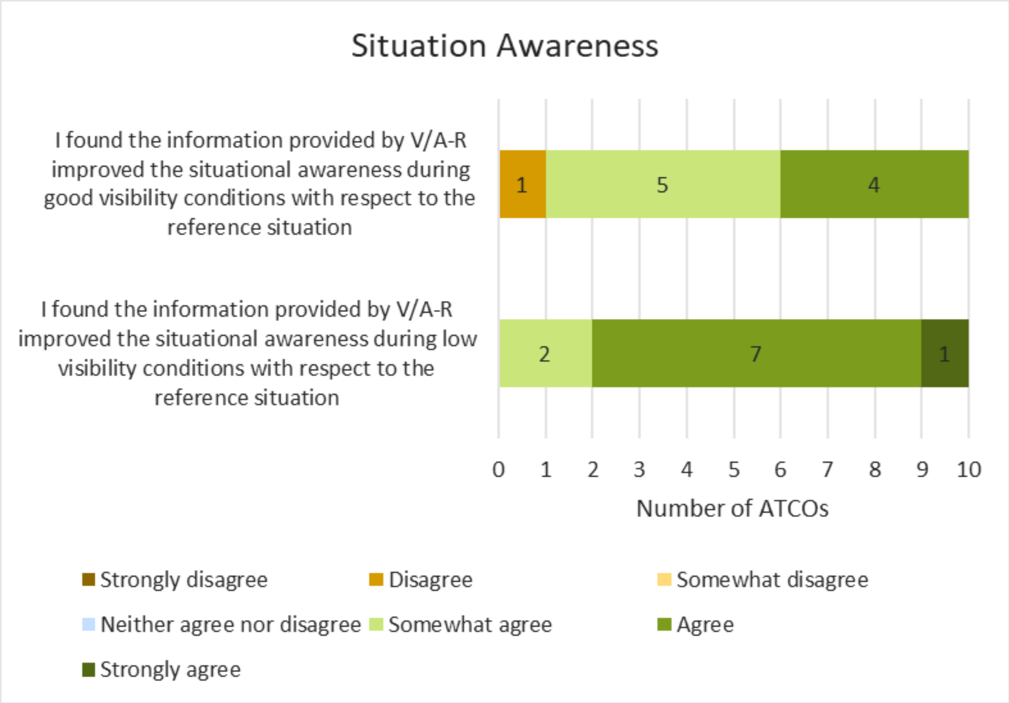


Figure B-39: Post-Exercise questionnaire results related to situation awareness

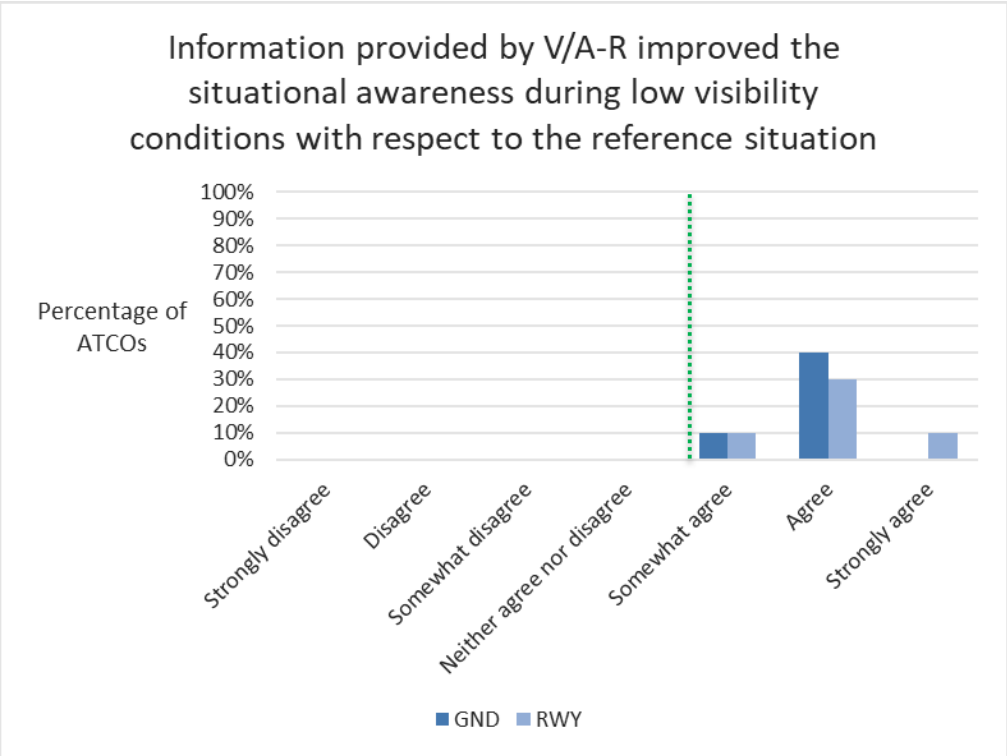


Figure B-40: Post-Exercise questionnaire results related to situation awareness for ground (GND) and runway (RWY) controller position

EX2-CRT-05.971-TLR4-TVALP-H103.1014

Measured time spent in head up is increased in the solution scenario with respect to the reference scenario

As can be seen from the figure below, head up time as a percentage of the total time is increased in the solution scenario with respect to the reference scenario. The bar chart graph compares the data in Run 1 + Run 4 ("Reference") with the data in Run 2 + 3 + 5 ("Solution"). Each data point indicates the percentage of Head Up time over the total time for the single run, for the single ATCO;

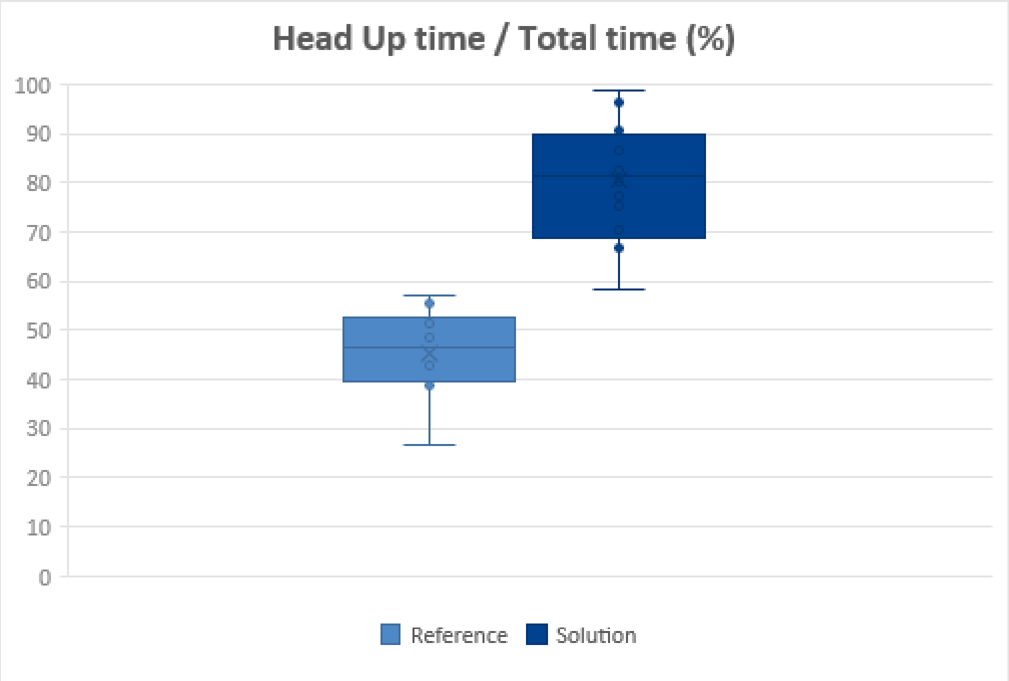


Figure B-41: Head up time as a percentage of the total time.

See the figures below for the improvement in head up time per ATCO for tracking labels and safety net respectively. Head up time increased relative to the reference scenario for all individual ATCOs and for both tracking labels and safety net.

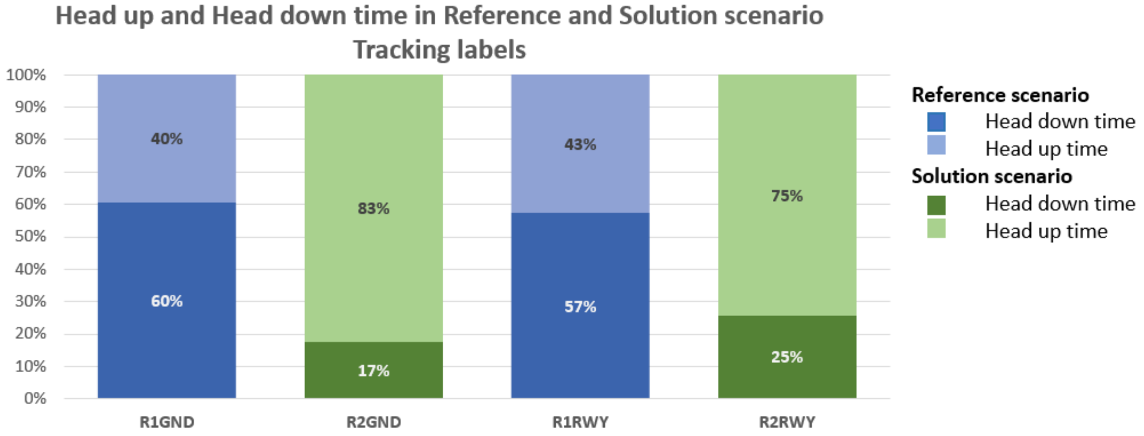


Figure B-42: Head Up time improvement per ATCO for Tracking Labels

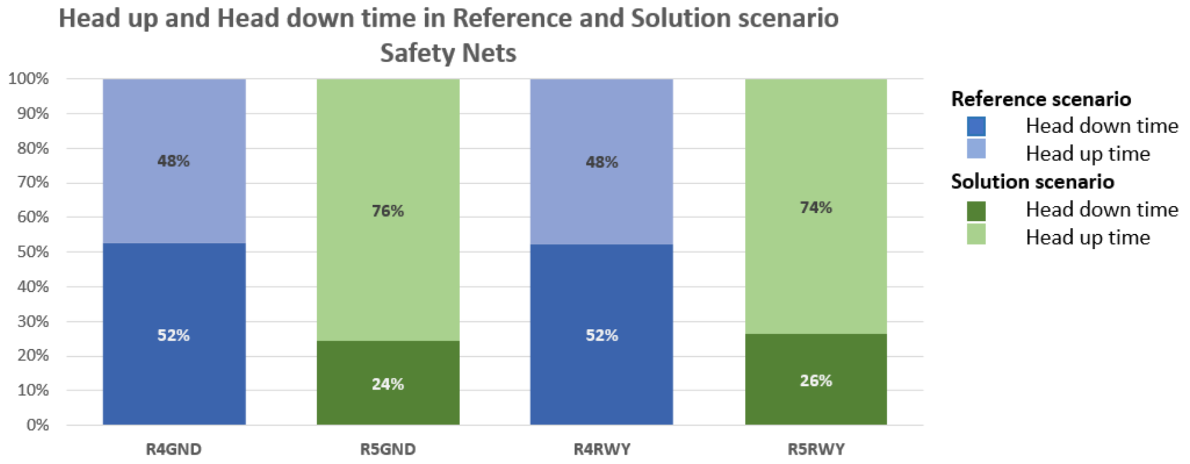


Figure B-43: Head Up time improvement per ATCO for Safety Net

Self-reported data confirms the head up improvements as all ATCOs agreed to the following statement in the post-exercise questionnaire: "I spent more time in head up when I used the V/A-R system with respect to the reference scenario" (see the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately).

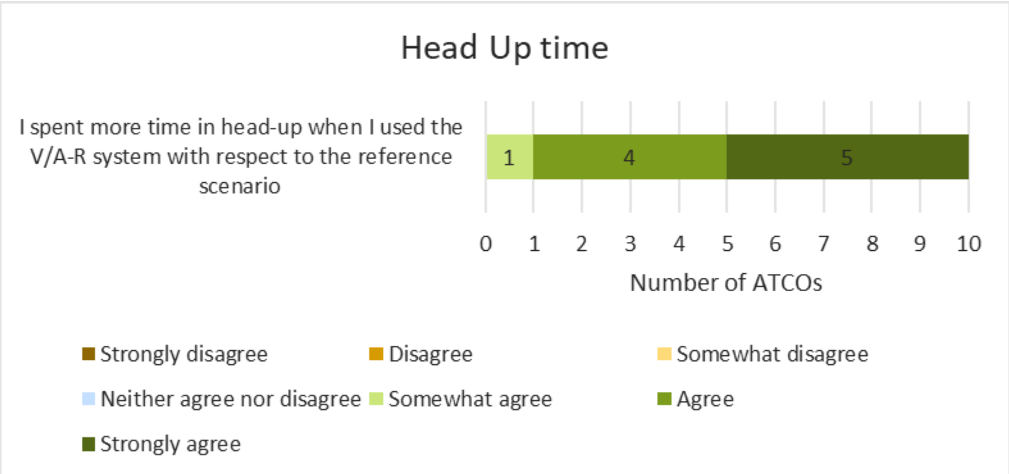


Figure B-44: Post-Exercise questionnaire results related to head up time

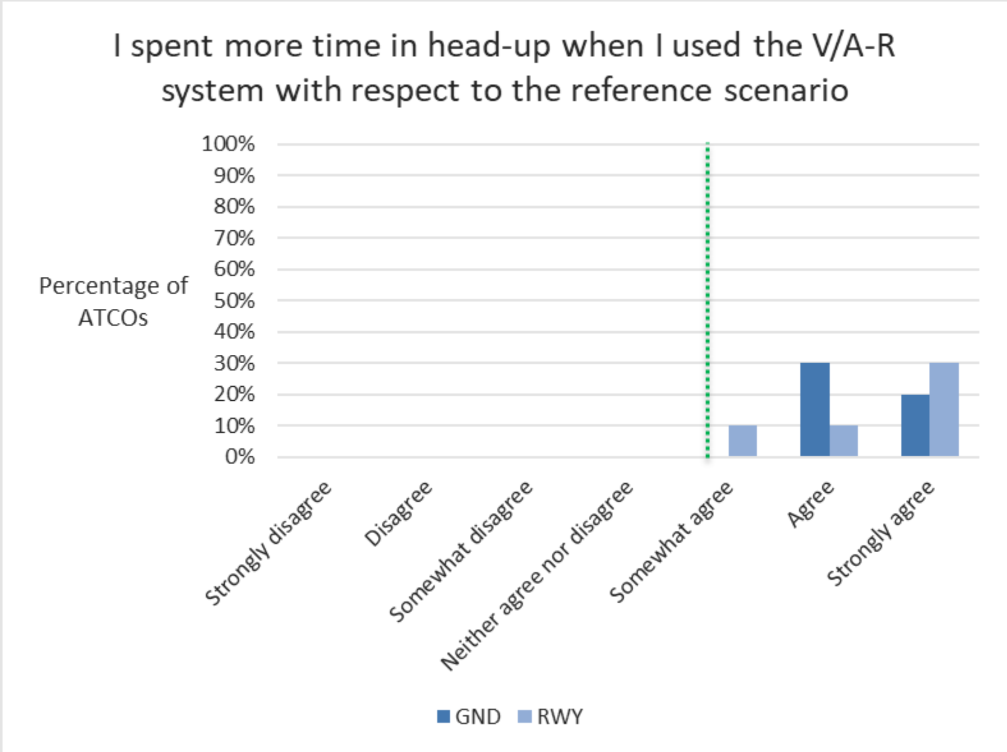


Figure B-45: Post-Exercise questionnaire results related to head up time for ground (GND) and runway (RWY) controller position

EX2-CRT-05.971-TLR4-TVALP-H103.1015

HMI of V/A-R tools does not overshadow the relevant information on the OTW view

Whereas five out of ten ATCOs agreed in the post-exercise questionnaire that the tracking label and the airport overlay provided by V/A-R were adequate and did not generate confusion neither disturbance, four of them disagreed (and one neither agreed nor disagreed; see Figure in section EX2-OBJ-05.971-TRL4-TVALP-H103.1010 on adequacy of information). ATCOs commented that the labels were overlapping and covering the background and that they were sometimes badly aligned. Also during the debriefing, ATCOs pointed out that the background colour of the label should be removed and that it should be prevented to have labels overlapping each other or the final approach path.

Apart from two ATCOs, all ATCOs agreed that they always had an adequate field of view when using the V/A-R system to perform their task. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately. Nevertheless, during the debriefing ATCOs mentioned that it would have been preferable to have a visual cue about the limited augmented reality field of view to be aware that the displayed information was available only within the augmented reality field of view.

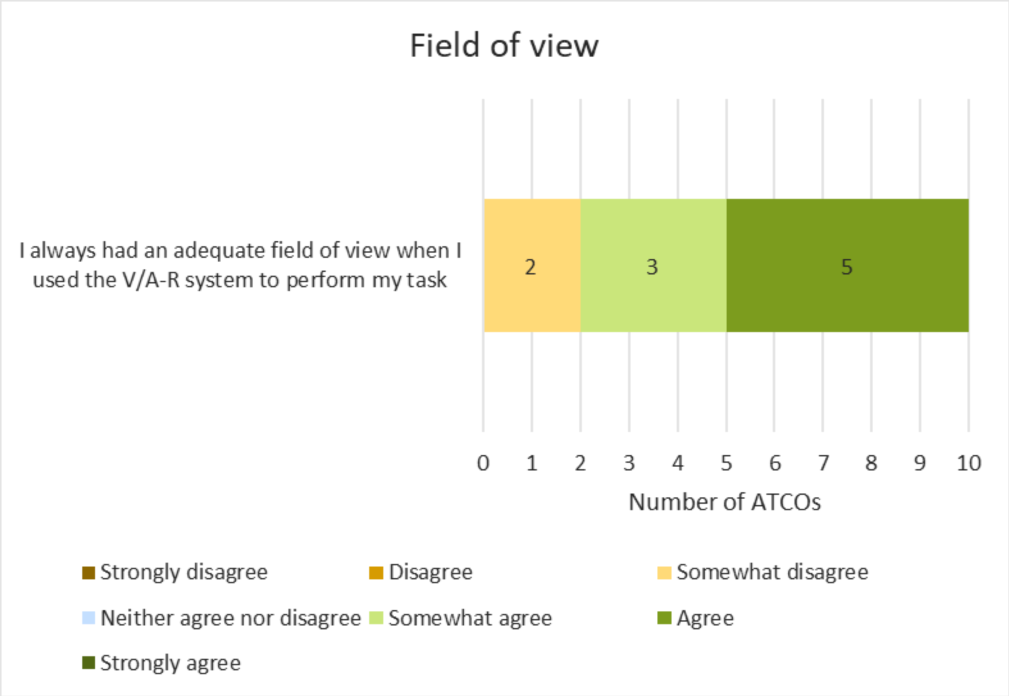


Figure B-46: Post-Exercise questionnaire results related to field of view

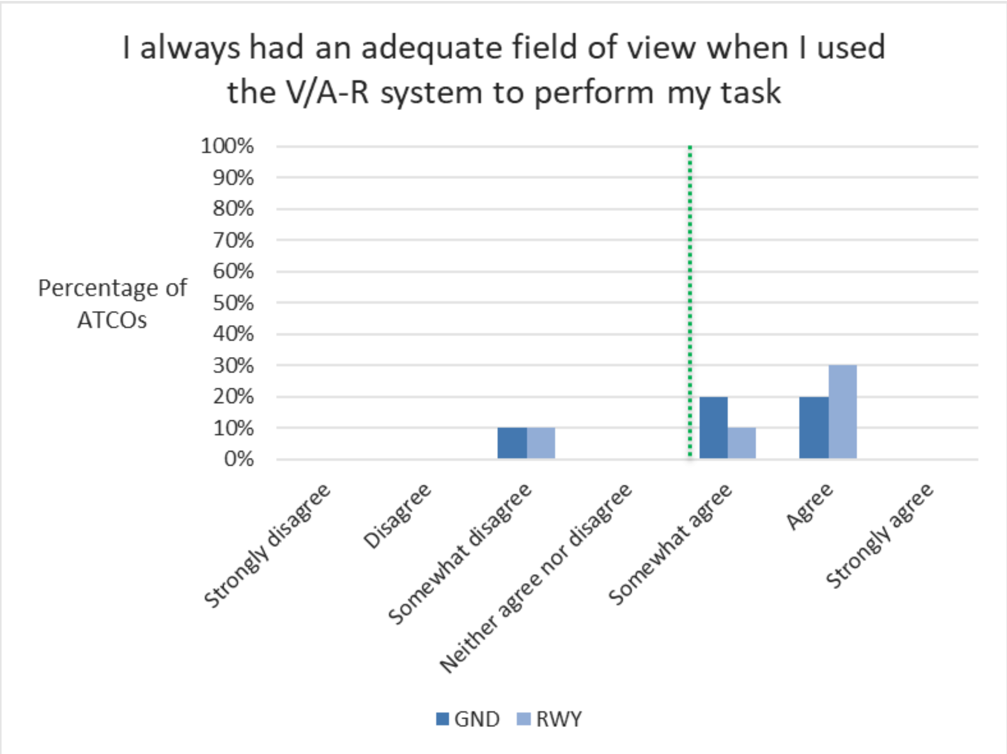


Figure B-47: Post-Exercise questionnaire results related to field of view for ground (GND) and runway (RWY) controller position

EX2-CRT-05.971-TLR4-TVALP-H103.1016

V/A-R HMI does not increase the potential for human error

Six out of ten ATCOs agreed that V/A-R systems design did not increase potential for human error compared to current operations, whereas two ATCOs somewhat disagreed and two were undivided. With the current interface design, the V/A-R systems design could lead to a potential for Human Error because the labels sometimes cover part of the manoeuvring area and the controller may not see an obstacle that is not detected by radar or GPS. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately.

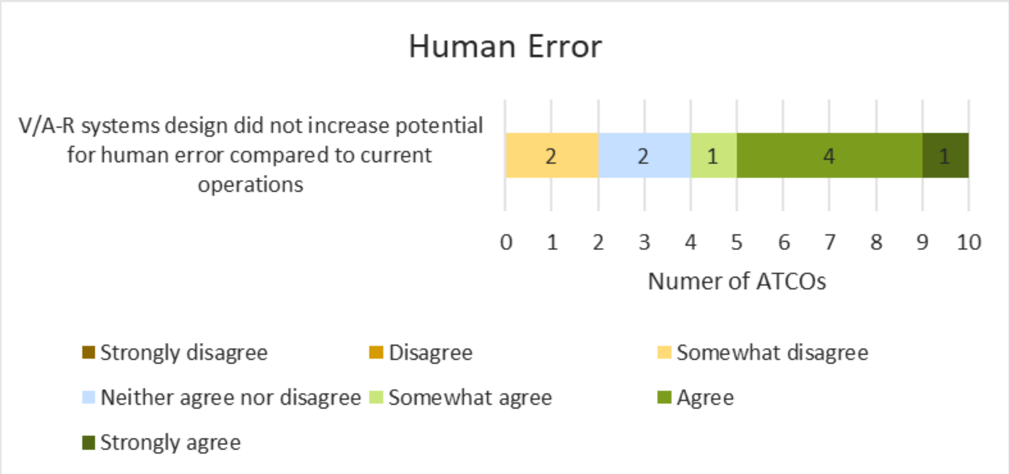


Figure B-48: Post-Exercise questionnaire results related to human error

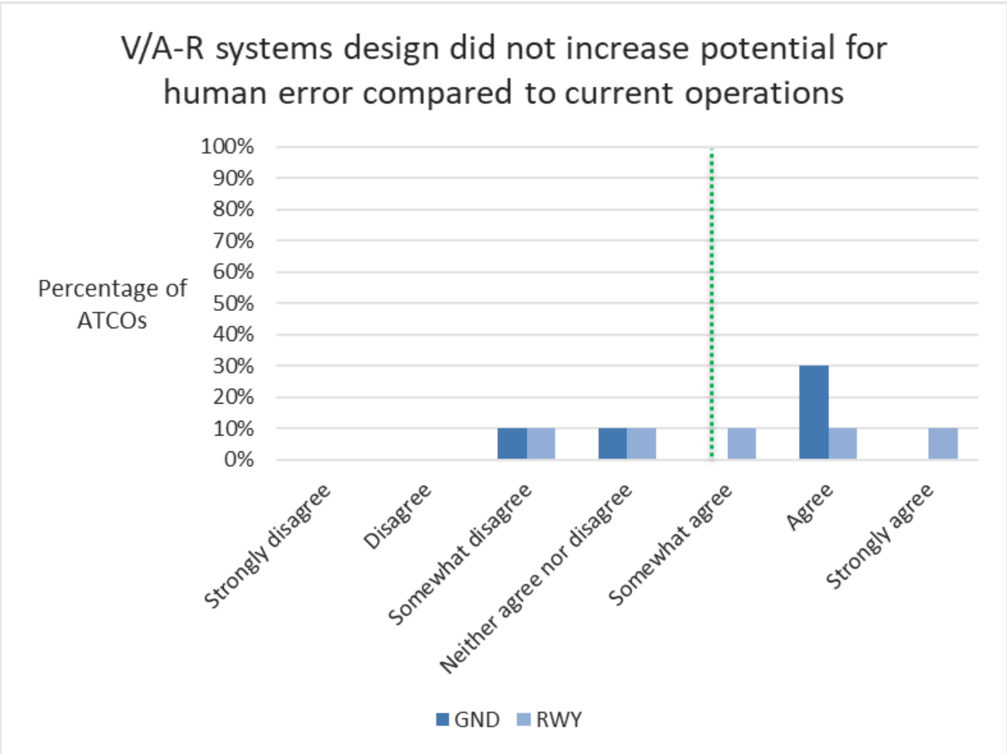


Figure B-49: Post-Exercise questionnaire results related to human error for ground (GND) and runway (RWY) controller position

During the debriefing, four ATCOs pointed out that the V/A-R technology can positively impact the potential for Human Error because of the availability of information on the screen. Thanks to the information on the screen, it is easier to see the differences and less likely to confuse airplanes. Normally, ATCOs have to look down on the radar in case of bad weather. In this sense, the glasses decrease the potential for human error.

EX2-CRT-05.971-TLR4-TVALP-H103.1017

ATCOs (at least 75%) trust in the system is at an acceptable level

In the post-exercise, 90% of the ATCOs agreed that they felt the V/A-R system was reliable. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately. Generally, the runway controllers provided more positive answers. This might again be related to the tracking labels covering the view.

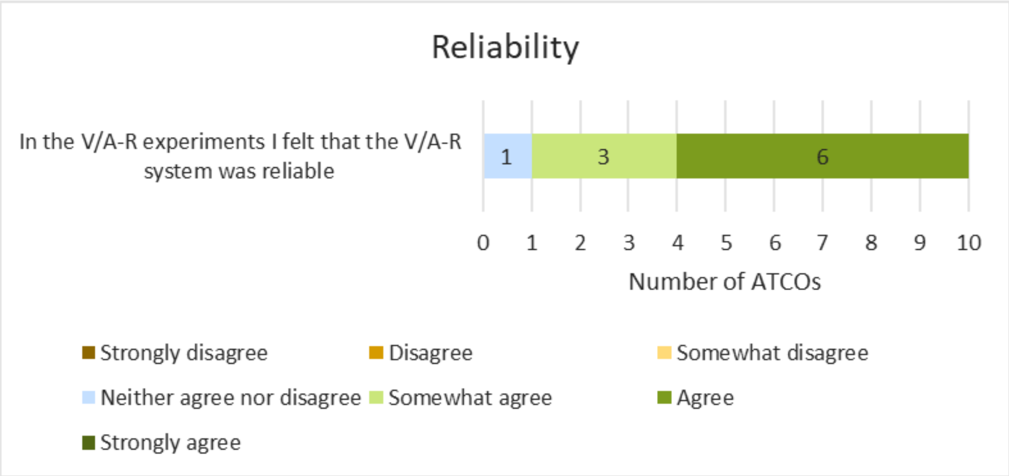


Figure B-50: Post-Exercise questionnaire results related to trust

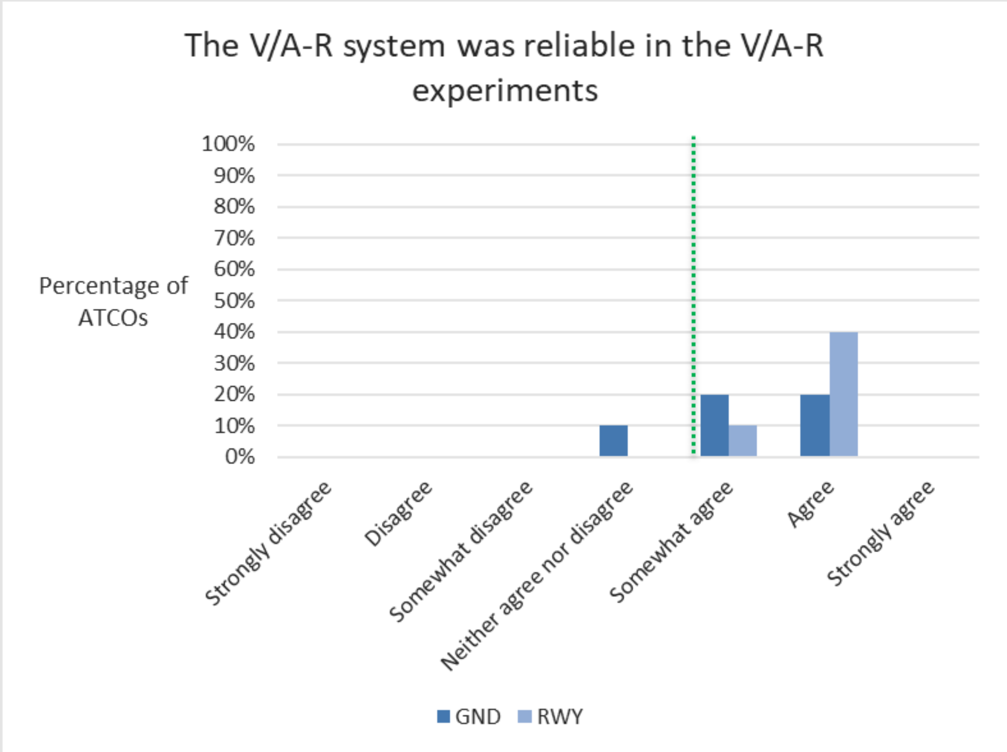


Figure B-51:Post-Exercise questionnaire results related to trust for ground (GND) and runway (RWY) controller position

During the debriefings, ATCOs claimed to trust the system. This can be confirmed by the answers in the HF impact matrix (see figure in section B.8.3.5 Final Debriefing evidences). Six out of ten ATCOs indicated that trust is neither negatively nor positively impacted by the V/A-R system according to them and the remaining four indicated a positive impact of the system on trust. As one ATCO mentioned, the preconditions for trust is that the data is equally confident as the data in the current system. One ATCO said that his trust in the system was temporarily decreased by the overlap of the labels.

EX2-CRT-05.971-TLR4-TVALP-H103.1018

Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R HMI

In terms of Usability & Ergonomics, there was no general agreement. Some ATCOs though that the HoloLens was comfortable, whereas other said that it disturbed them. First of all, it was mentioned that the glasses were relatively heavy. Secondly, two ATCOs mentioned that when they move their head, the vision becomes blurry. One ATCO also mentioned that it is a bit tiring to use the glasses in combination with their own prescription glasses. Additionally, some ATCOs noted that the physical comfort is impacted by the fact that they always have to look in the direction of the traffic to see the corresponding information. Other ATCOs mentioned that they like using the headset more than the headphone that they normally use. It made them feel more engaged. In the post-run questionnaire, 15 out of 20 responses testify that the physical workload was acceptable, slightly light, light or very light in the solution scenarios (see figures in section on workload).

In the HF impact matrix (see figure in section B.8.3.5 Final Debriefing evidences), 14 out of 25 arrow are pointed forwards or upwards whereas almost half of the arrows indicate a degradation of Usability

& Ergonomics. Negative ratings were collected mostly for the runway controller and for the use of air gestures.

In the post-exercise questionnaire, five out of ten ATCOs agreed that they would like to use the V/A-R system frequently, whereas two disagreed and two neither agreed nor disagreed. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately. Some ATCOs indicated that it would be preferable to use the system only during low visibility conditions. Furthermore, eight out of ten ATCOs agreed that they think the V/A-R system is easy to use and the other two ATCOs were neutral about this. One of them indicated that training is needed. Three ATCOs somewhat agreed that there was too much inconsistency in the V/A-R system while the remaining ATCOs disagreed (5) or were neutral (2). One ATCO commented that the colour use was not completely consistent. Seven ATCOs agree that they imaging most people to learn to use the V/A-R system very quickly (the remaining three ATCOs neither agreed nor disagreed). One ATCO commented that it could depend on the age of the user. Lastly, the majority of ATCOs, namely seven, agreed that they felt confident using the V/A-R system. The other three said they somewhat disagree. Again it was mentioned that training is needed. One ATCO also commented that some tools needs to be implemented for the device to be complete. This was also mentioned six times during the debriefing sessions in the Want/Have Matrix (see figure in section B.8.3.5 Final Debriefing evidences). ATCOs suggested the complete integration of the radar, Automatic Dependence Surveillance, automatic speech recognition and DATA-LINK. Specifically related to the interface, the post-exercise questionnaire showed that eight out of ten ATCOs agree that the interface was clear and complete. Some comments were made related to the limited visual range, the need to integrate other tools and information and the dimension of the labels. In terms of ergonomics, 70% of the ATCOs thinks that the V/A-R does not cause any negative physical consequences and 80% found the V/A-R device physically comfortable.

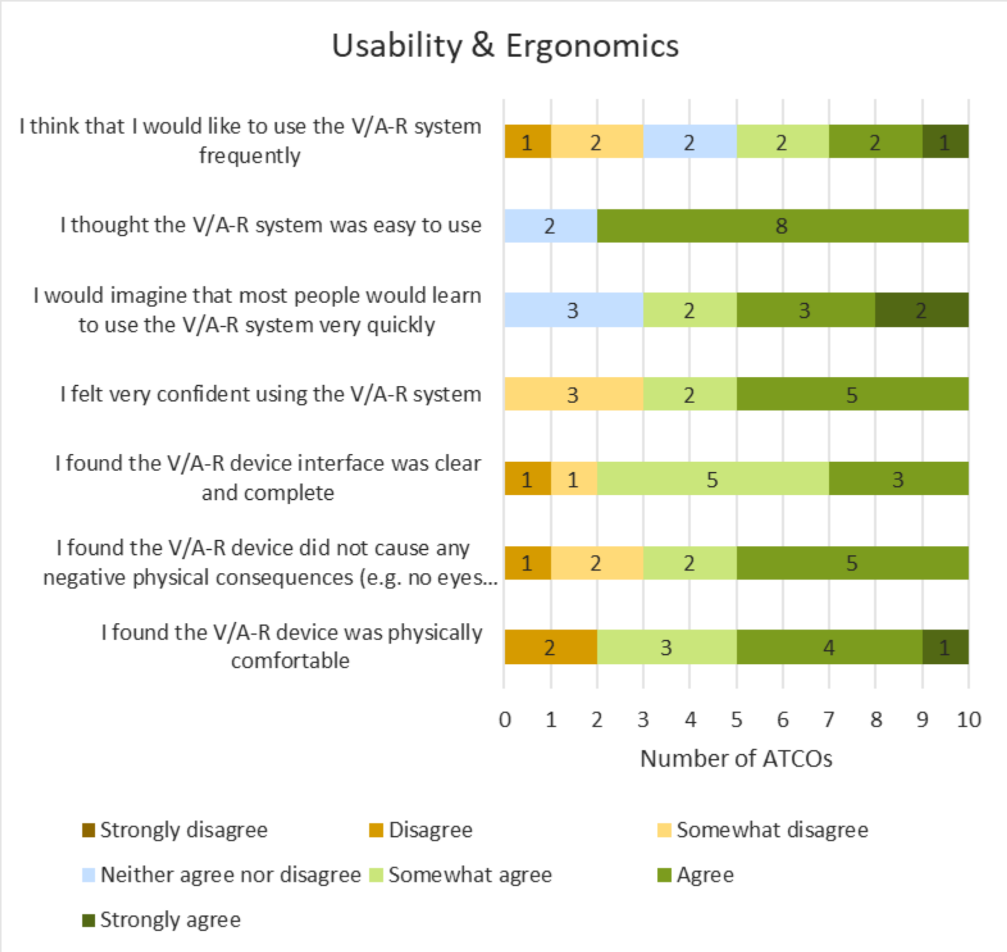


Figure B-52: Post-Exercise questionnaire answers related to usability & ergonomics

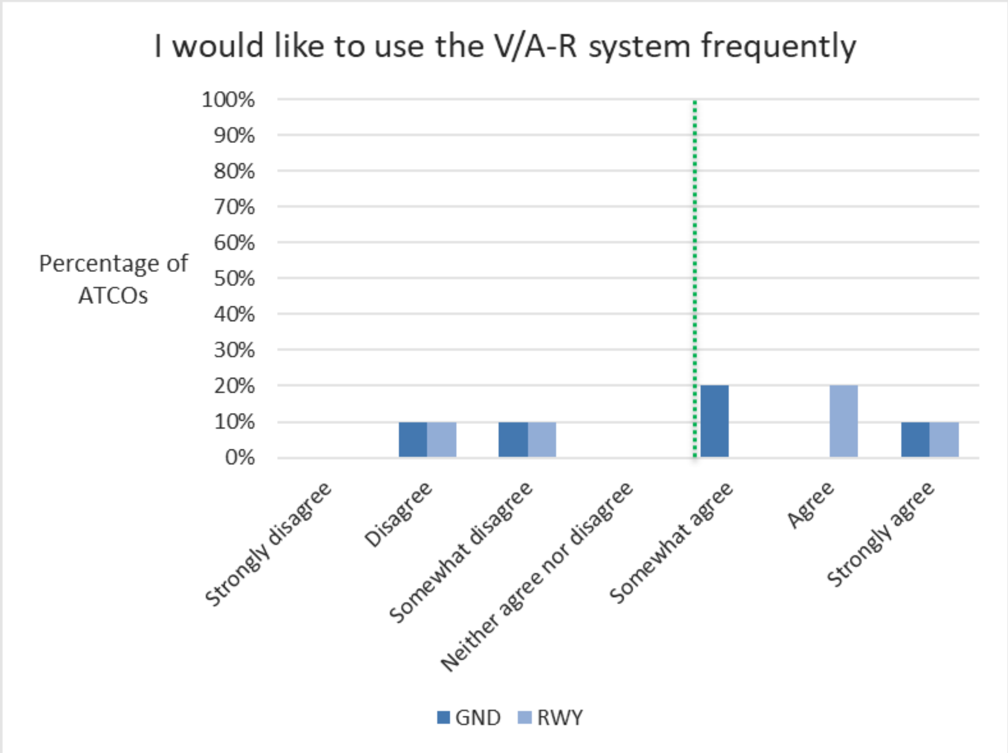


Figure B-53: Post-Exercise questionnaire answers related to frequency of use for ground (GND) and runway (RWY) controller position

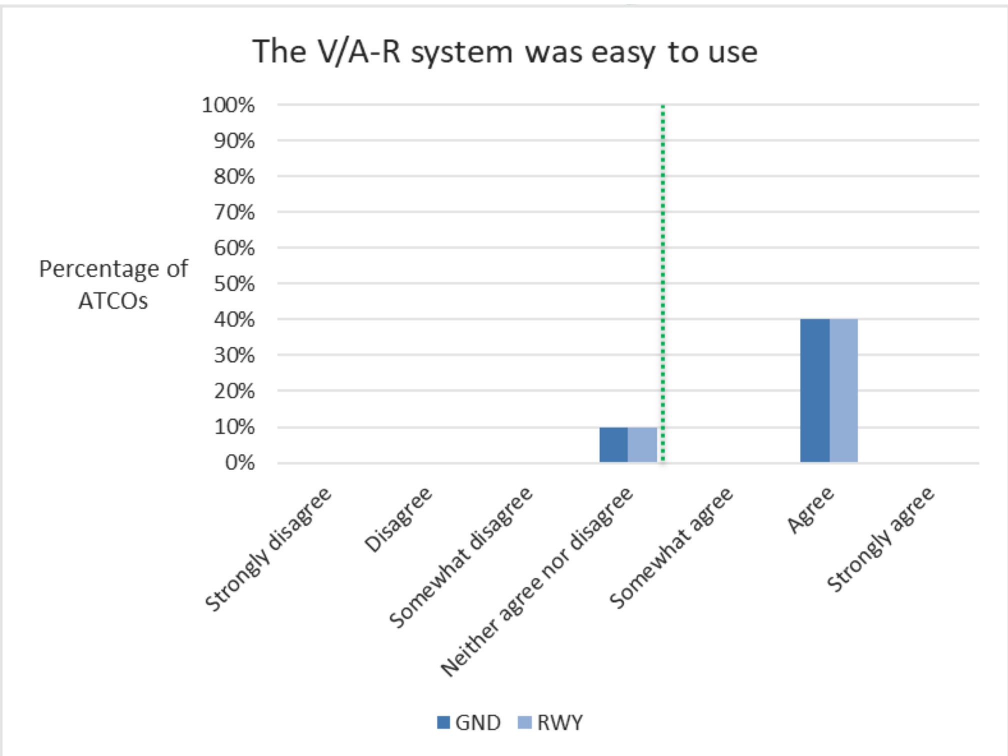


Figure B-54: Post-Exercise questionnaire answers related to ease of use for ground (GND) and runway (RWY) controller position

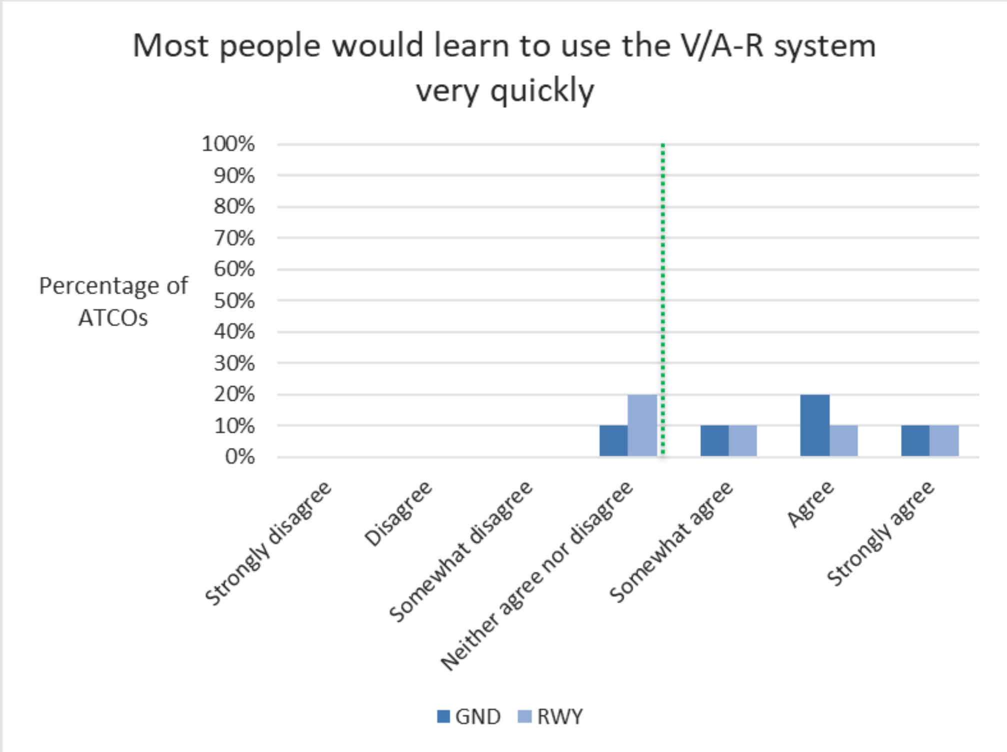


Figure B-55: Post-Exercise questionnaire answers related to learnability for ground (GND) and runway (RWY) controller position

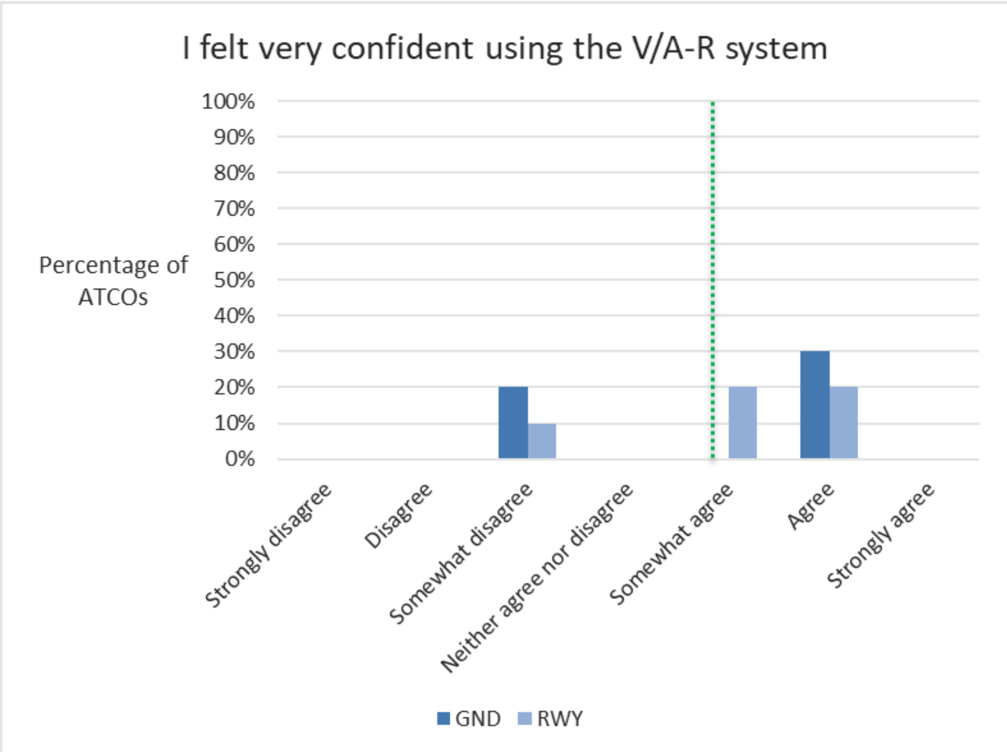


Figure B-56: Post-Exercise questionnaire answers related to confidence for ground (GND) and runway (RWY) controller position

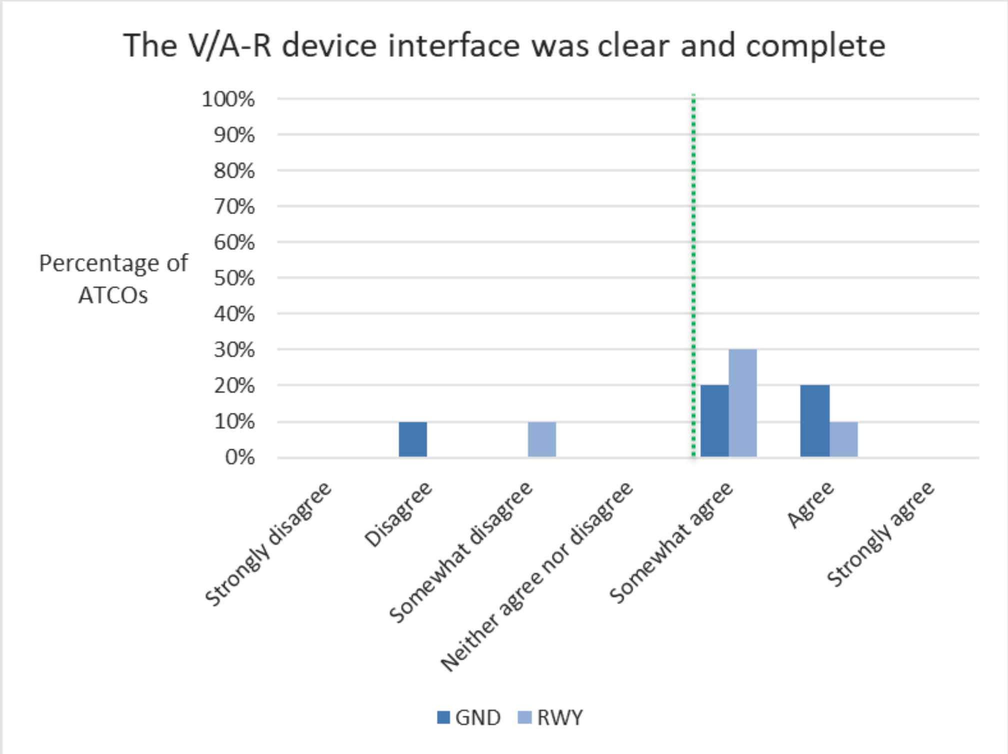


Figure B-57: Post-Exercise questionnaire answers related to interface clarity and completeness for ground (GND) and runway (RWY) controller position

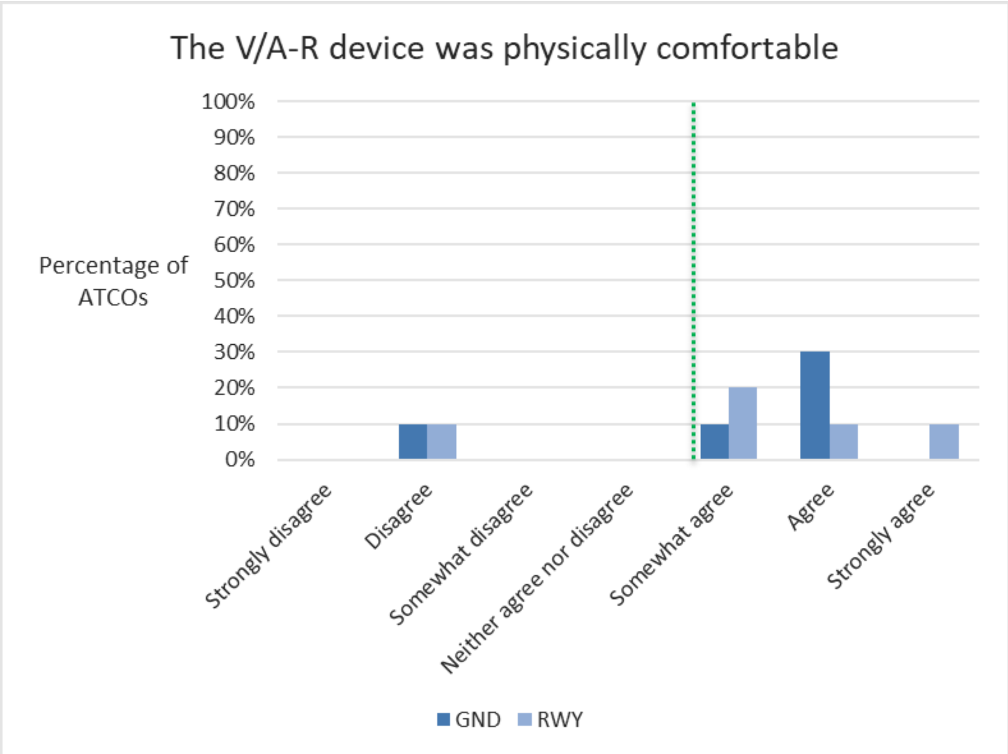


Figure B-58: Post-Exercise questionnaire answers related to physical comfort for ground (GND) and runway (RWY) controller position

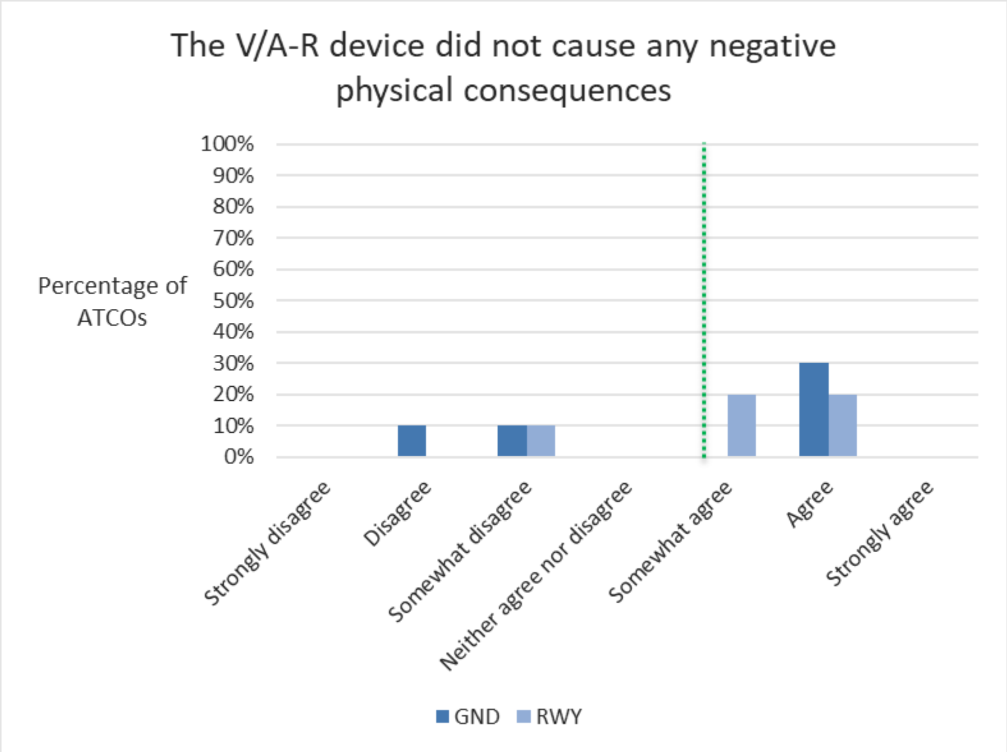


Figure B-59: Post-Exercise questionnaire answers related to physical consequences for ground (GND) and runway (RWY) controller position

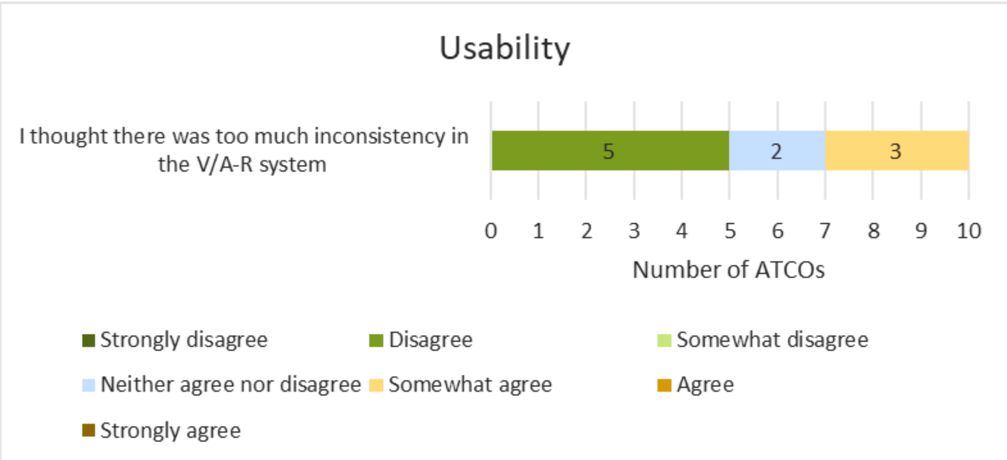


Figure B-60: Post-Exercise questionnaire answers related to usability

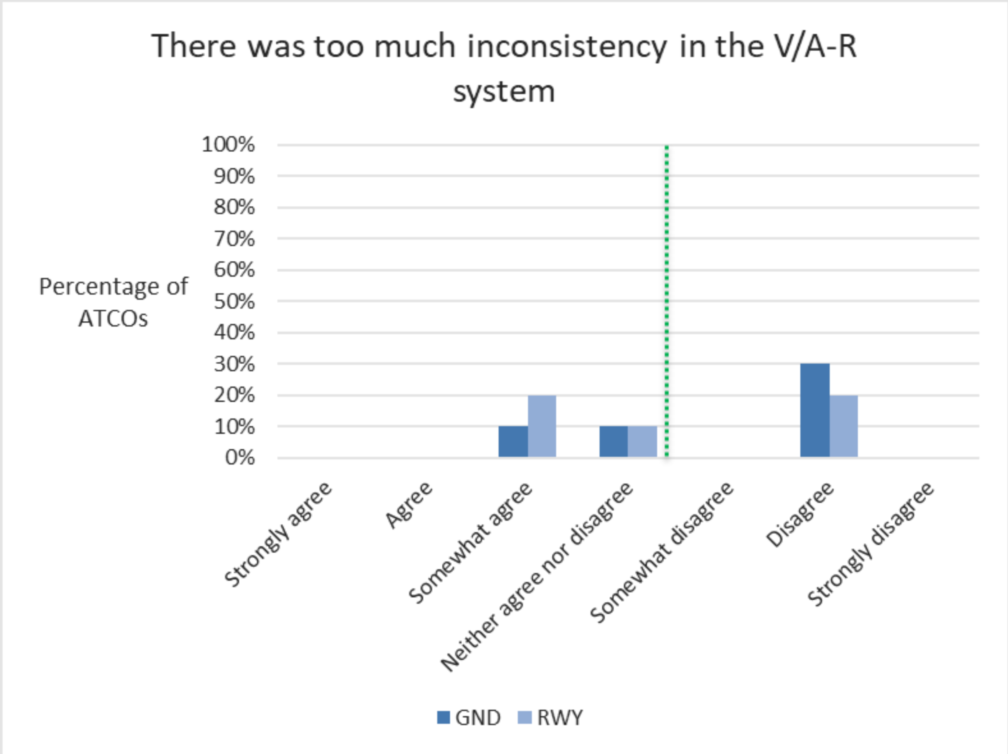


Figure B-61: Post-Exercise questionnaire answers related to inconsistency for ground (GND) and runway (RWY) controller position

EX2-CRT-05.971-TLR4-TVALP-H103.1019

Majority of ATCOs (at least 75%) responses show that alarms and alerts in the prototype for V/AR are not too intrusive and support ATCOs in the early detection of ATC critical situations with respect to conflicting clearances and runway incursions.

All ATCOs agreed that the V/A-R system alerts were effective and not intrusive. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately.

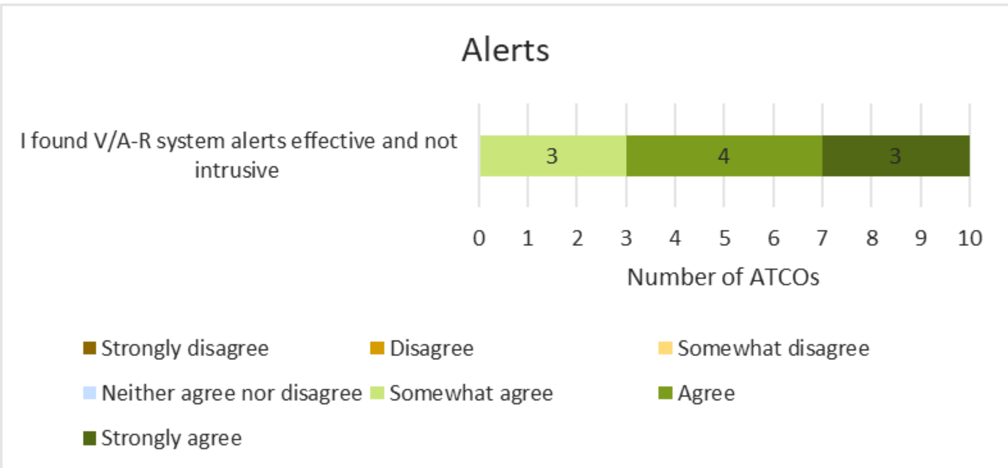


Figure B-62: Post-Exercise questionnaire answers related to alerts

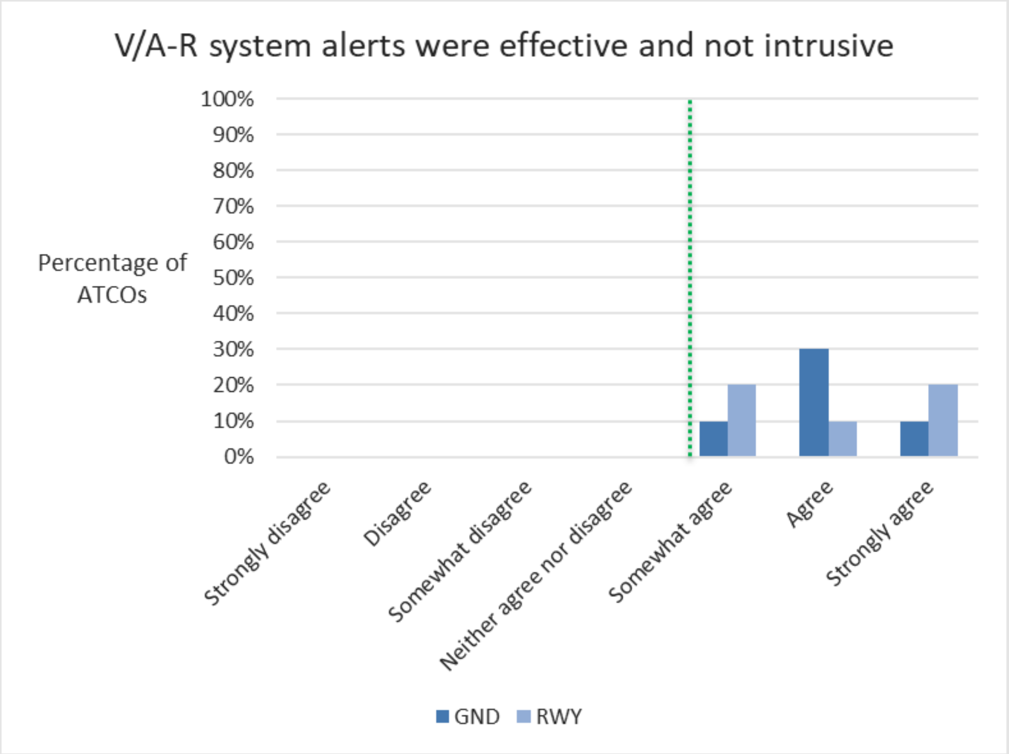


Figure B-63: Post-Exercise questionnaire answers related to alerts for ground (GND) and runway (RWY) controller position

EX2-CRT-05.971-TLR4-TVALP-H103.1020

ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R tool

In the post-run questionnaire, ATCOs reported an average Acceptance level of 6,7 for the V/A-R tool. See the figures below for the number of answers collected for each point of the 10-point CARS scale and the average Acceptance level for each solution and controller position. All mean values are above the acceptable minimum of 5.

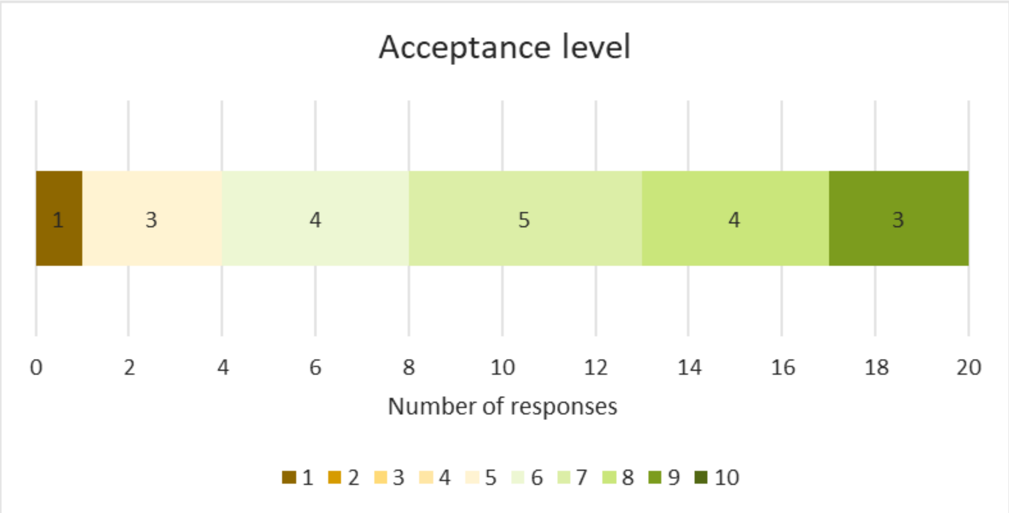


Figure B-64: Post-Run questionnaire results related to technology acceptance level

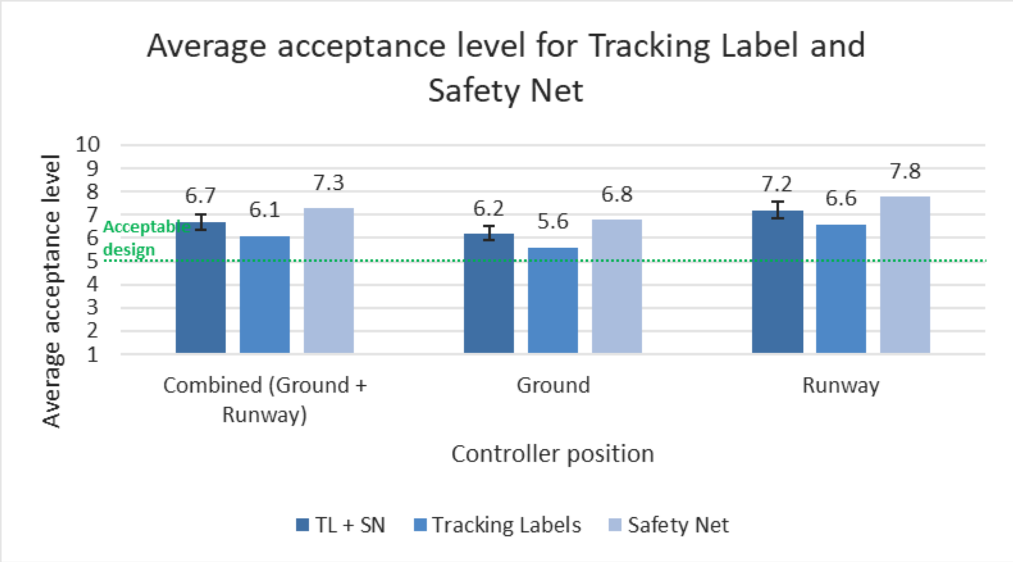


Figure B-65: Post-Run questionnaire results related to average technology acceptance level

In the post-exercise questionnaire, six out of ten ATCOs rated their acceptance of the V/A-R tool as high, two reported acceptance to be sufficient and two said their acceptance level was moderate. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately. Although the concept is highly acceptable, its implementation still needs to be further improved to be acceptable. Some ATCOs mentioned that the level of acceptance might depend on the age, experience and role of the ATCO. One ATCO thinks that the V/A-R tool might be more accepted in the future as similar devices become more common (also in daily life). Additionally, he pointed out that high traffic might be a problem and that the solution is mostly acceptable for bad vision conditions.

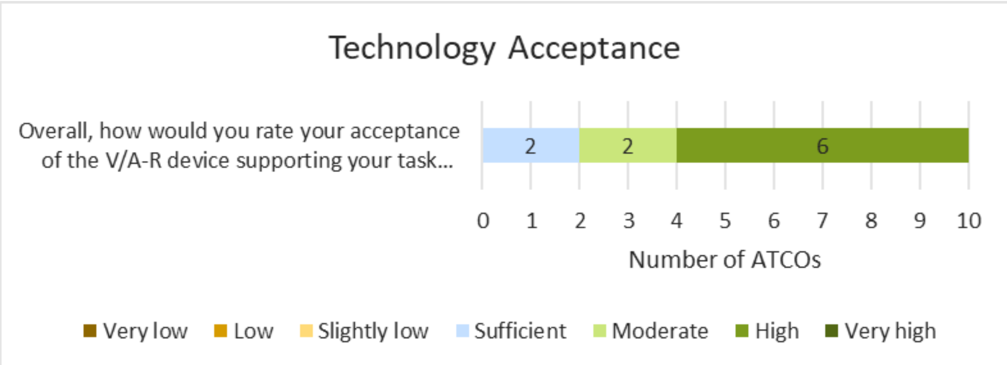


Figure B-66: Post-Exercise questionnaire results related to technology acceptance

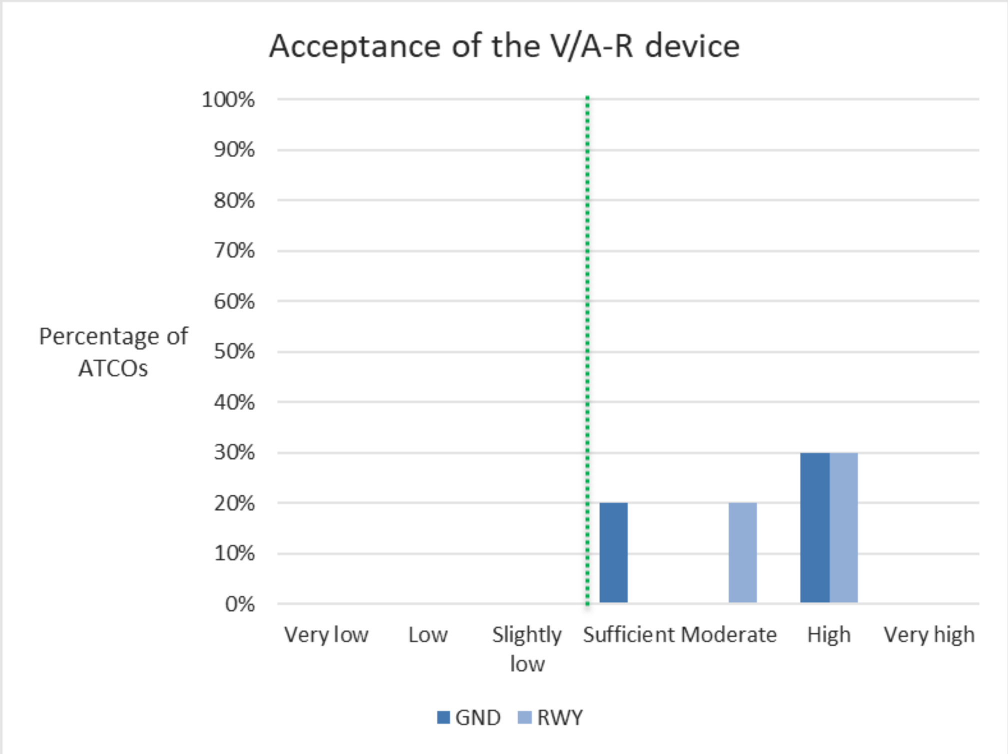


Figure B-67: Post-Exercise questionnaire results related to technology acceptance for ground (GND) and runway (RWY) controller position

EX2-CRT-05.971-TLR4-TVALP-H103.1021

Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO team (GND and TWR) in maintaining an acceptable level of situation awareness

Most ATCOs (7) rated the team situational awareness during the V/A-R exercise execution compared to the reference scenarios as high. The remaining three ATCOs rated the team SA to be either sufficient (2) or perfect(1). See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately.

As can be seen from the HF impact matrix (see figure in section B.8.3.5 Final Debriefing evidences), ATCOs indicated that the V/A-R tool supports similar or increased team SA compared to the reference scenario.

During the debriefing, one ATCO mentioned that team SA is affected by the fact that the ATCOs do not have the same vision. Another ATCO mentioned that it would be interesting to have the possibility to make something appear on the screen of the other ATCO.

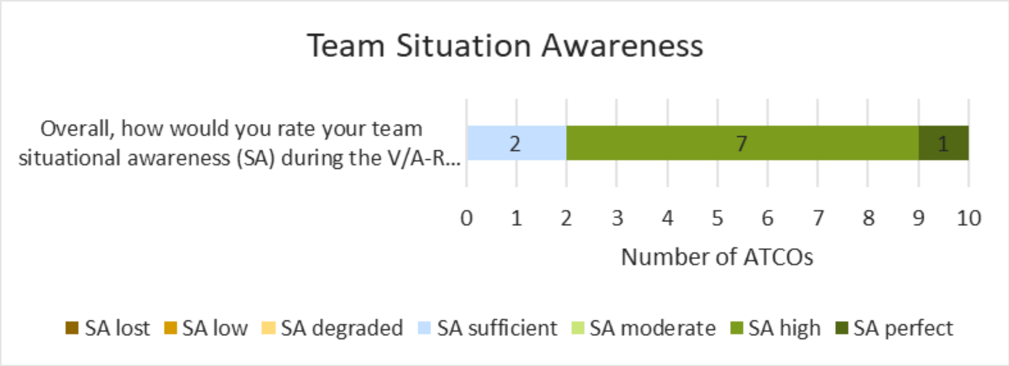


Figure B-68: Post-Exercise questionnaire results related to team situation awareness

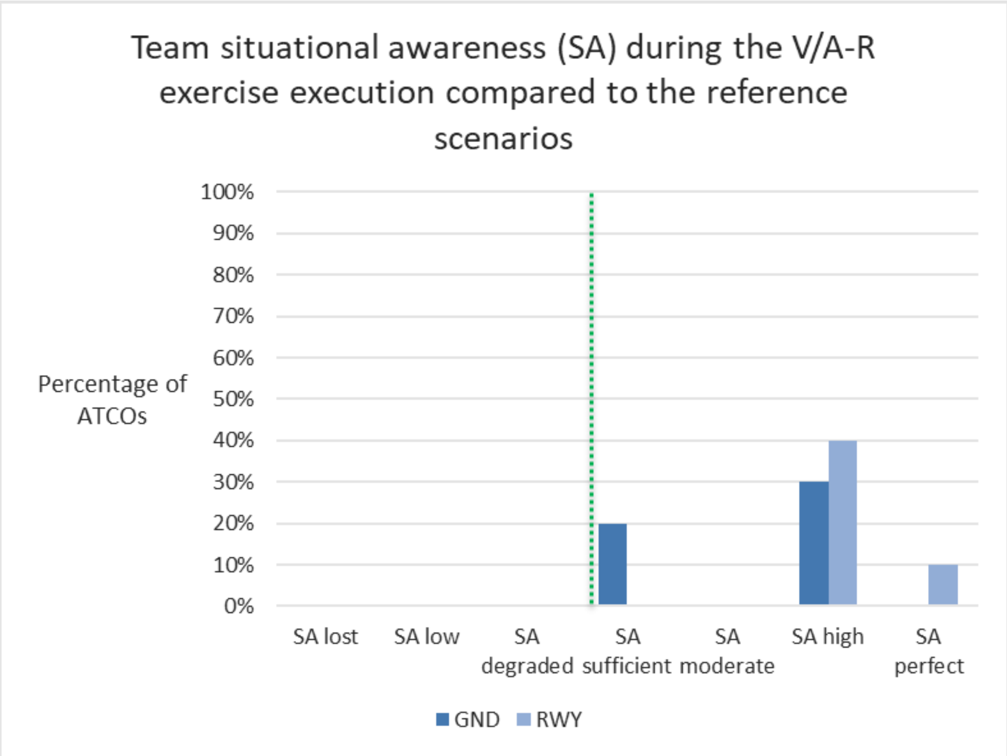


Figure B-69: Post-Exercise questionnaire results related to team situation awareness for ground (GND) and runway (RWY) controller position

B.7.2.4 EX2-OBJ-05.971-TRL4-TVALP- H103.1030

To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Tracking labels and overlays.

EX2-CRT-05.971-TLR4-TVALP- H103.1031

Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner

In the post-run questionnaire, the ATCOs agreed that they were able to apply the operating methods for V/A-R in an accurate, efficient and timely manner for almost all runs. Only one ATCO somewhat disagreed after one of the runs. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the average answer for both solutions and controller position.

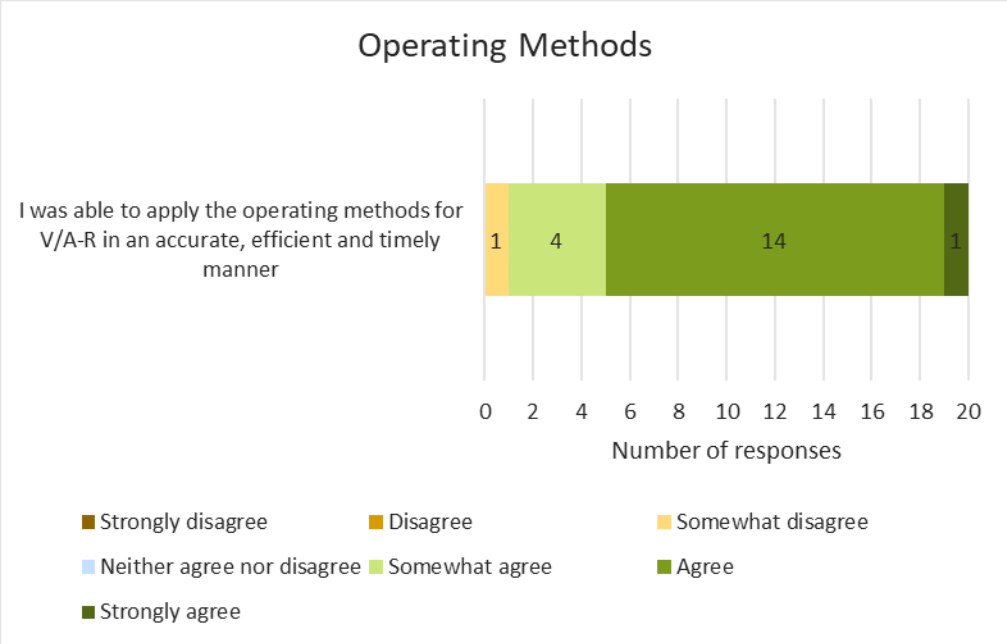


Figure B-70: Post-Run questionnaire answers related to operating methods

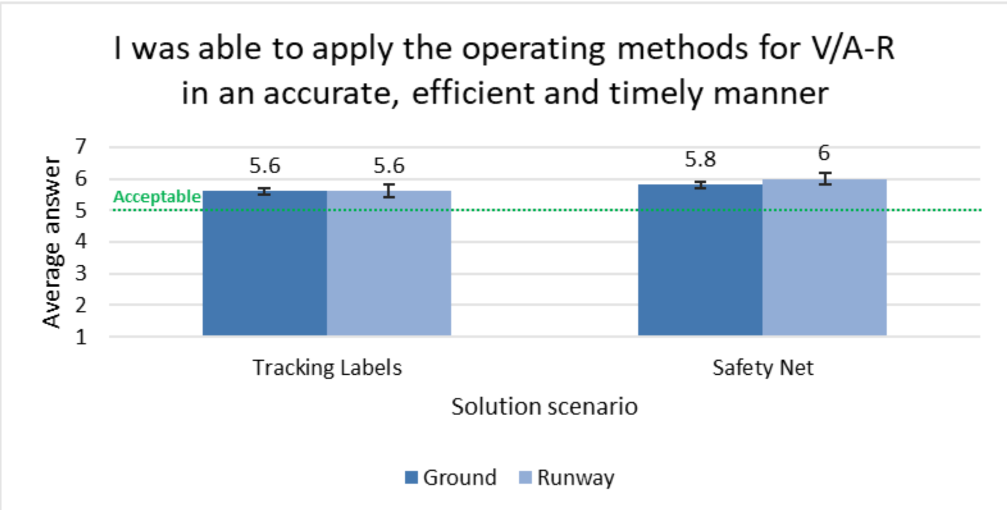


Figure B-71: Post-Run questionnaire answers related to application of operating methods for ground (GND) and runway (RWY) controller

EX2-CRT-05.971-TLR4-TVALP-H103.1032

Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in the investigated operating conditions.

In the post-run questionnaire, the ATCOs agreed in 70% of the cases that they found the operating methods for V/A-R to be clear, complete and exhaustive. The remaining 10% of the responses indicated that the ATCO neither agreed nor disagreed and 15% indicated that the ATCO disagreed. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the average answer for both solutions and controller position. The average value for the clarity and completeness of the tracking labels was less than the acceptable minimum of 5 for runway controllers. All the other average were above the acceptable minimum.

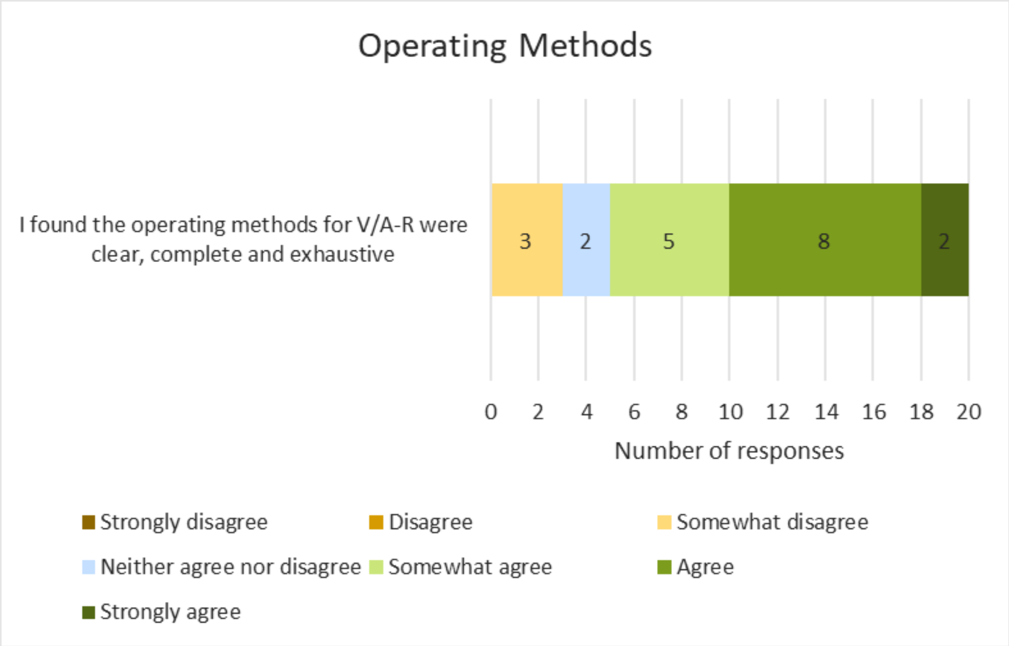


Figure B-72: Post-Run questionnaire answers related to operating methods

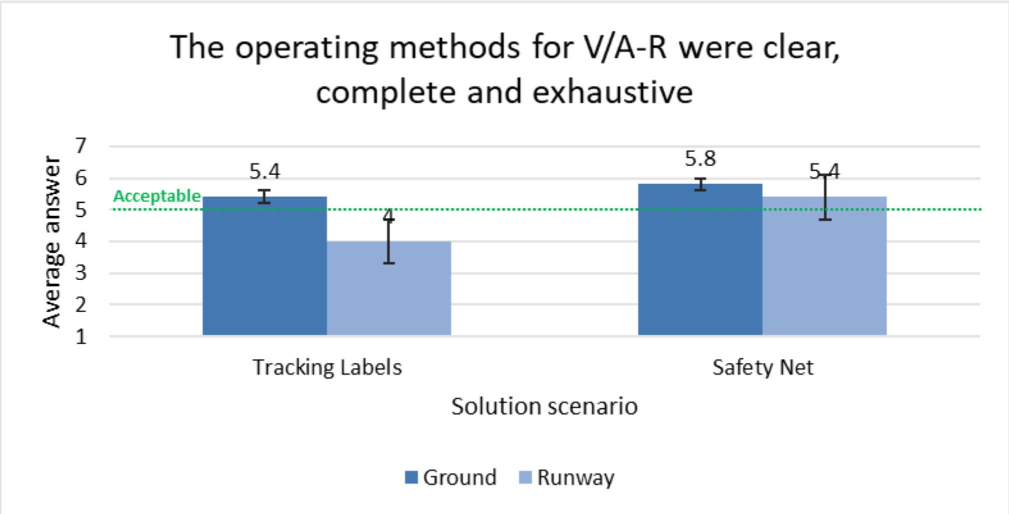


Figure B-73: Post-Run questionnaire answers related to clarity and completeness of operating methods for ground (GND) and runway (RWY) controller

B.7.2.5 EX2-OBJ-05.971-TRL4-TVALP- H103.1040

To assess job acceptance and satisfaction with the introduction of V/A-R tracking labels and overlays.

EX2-CRT-05.971-TLR4-TVALP-H103.1041

Validation activities show that ATCOs (at least 75%) give positive feedback on job satisfaction and acceptance regarding the prototype for V/AR tracking labels and overlays.

From the HF impact matrix (see figure in section B.8.3.5 Final Debriefing evidences), it can be seen that acceptance and job satisfaction is mostly said to not be impacted (20%) or to be impacted positively (65%). One ATCO commented that job acceptance & satisfaction depends on how the concept is introduced. He prefers the use of the V/A-R tool to be optional, so that he can use it whenever he wants, as an 'extra', in the same way he uses the headphones nowadays.

B.7.2.6 EX2-OBJ-05.971-TRL4-TVALP- H104.1010

To assess that the technical systems for V/A-R Air Gestures support the ATCOs in performing the tasks under investigation.

EX2-CRT-05.971-TLR4-TVALP-H104.1011

Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures supports ATCO in maintaining workload at acceptable level

As can be seen in Figure in section B.8.3.5 Final Debriefing evidences, 40% of the ATCOs (2) responded that the V/A-R Air Gestures either have no or a positive impact on the workload level. Several ATCOs had difficulties using Air Gestures. One ATCO said that he was not able to give the clearance with the air gesture (at least not easily) which made the task tiring.

EX2-CRT-05.971-TLR4-TVALP-H104.1012

ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of the interaction means supported by V/A-R Air Gestures.

ATCOs mentioned it several times during the debriefings that they had difficulties using Air Gestures as the system did not always recognise their gestures.

EX2-CRT-05.971-TLR4-TVALP-H104.1013

Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures HMI supports ATCO in maintaining an adequate level of situation awareness

As can be seen in Figure in section B.8.3.5 Final Debriefing evidences, 60% of the ATCOs (3) responded that the V/A-R Air Gestures either have no or a positive impact on the SA level. One ATCO pointed out that he lost SA because he was not able to give the clearance with the air gesture.

EX2-CRT-05.971-TLR4-TVALP-H104.1014

Measured time spent in head up is increased in the solution scenario with respect to the reference scenario

The line chart below provides a comparison between Run 1 'shortened' to run 3 length ("Reference") and Run 3 ("Air Gesture"). The "improvement" percentage is calculated by dividing the Head Up over Total time in the examined run by the corresponding time in the reference.

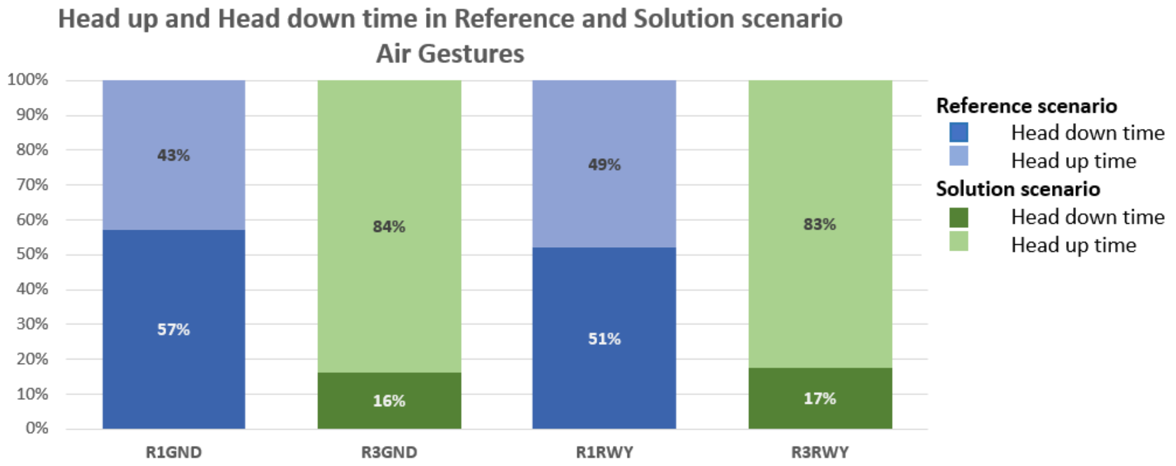


Figure B-74: Head Up time improvement per ATCO

EX2-CRT-05.971-TLR4-TVALP-H104.1015

V/A-R Air Gestures HMI does not increase the potential for human error

As can be seen in Figure in section B.8.3.5 Final Debriefing evidences, 40% of the ATCOs (2) responded that the V/A-R Air Gestures increase the potential for human error. On the other hand, one ATCO pointed out that the air gesture takes out the possibility of miscommunication (saying something wrong). Several ATCOs mentioned that there should be a voice communication back-up and that air gestures should not be used for critical cases.

EX2-CRT-05.971-TLR4-TVALP-H104.1016

Laboratory tests show that ATCOs' (at least 75%) trust in the prototype for V/AR Air Gestures is at an acceptable level.

As can be seen in Figure in section B.8.3.5 Final Debriefing evidences, 80% of the ATCOs (4) responded that the V/A-R Air Gestures either have no or a positive impact on the trust level.

EX2-CRT-05.971-TLR4-TVALP-H104.1017

Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R Air Gestures HMI

As can be seen in Figure in section B.8.3.5 Final Debriefing evidences, only 20% of the ATCOs (1) responded that the V/A-R Air Gestures have no impact on the usability whereas the other 80% percent believes that usability is negatively impacted. Two ATCOs reported that it was difficult to use air gestures from an ergonomic point of view.

EX2-CRT-05.971-TLR4-TVALP-H104.1018

ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R Air Gestures tool

As can be seen in Figure in section B.8.3.5 Final Debriefing evidences, 80% of the ATCOs (4) responded that the V/A-R Air Gestures either have no or a positive impact on the Acceptance & Job satisfaction level. One ATCO thought the air gestures should be removed and two ATCOs pointed out that it should be avoided to use air gesture commands for runway authorizations/critical cases.

B.7.2.7 EX2-OBJ-05.971-TRL4-TVALP- H104.1020

To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Air Gestures.

EX2-CRT-05.971-TLR4-TVALP- H104.1021

Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner

No specific impacts of Air Gesture on operation methods were mentioned by the ATCOs.

EX2-CRT-05.971-TLR4-TVALP-H104.1022

Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in the investigated operating conditions

No specific impacts of Air Gesture on operation methods were mentioned by the ATCOs.

B.7.2.8 EX2-OBJ-05.971-TRL4-TVALP- H104.1030

To assess job acceptance and satisfaction with the introduction of V/A-R Air Gestures.

EX2-CRT-05.971-TLR4-TVALP-H104.1031

Laboratory tests show that ATCOs (at least 75%) give positive feedback on job satisfaction and acceptance regarding the prototype for V/AR Air Gestures

As can be seen in Figure in section B.8.3.5 Final Debriefing evidences, 80% of the ATCOs (4) responded that the V/A-R Air Gestures either have no or a positive impact on the Acceptance & Job satisfaction level.

B.7.2.9 EX2-OBJ-05.971-TLR4-TVALP-SAFE.1010

To give relevant input about safety issues when using V/AR with safety nets

EX2-CRT-05.971-TLR4-TVALP-SAFE-1011

Laboratory tests show that the prototype for V/AR with safety nets improves the safety performance by reducing human error

Almost all ATCOs agreed that the V/A-R system helped them in the early detection of critical situations. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately. Moreover, all ATCOs indicated that the Safety Nets positively impacted the potential for Human Error (see Figure in section B.7.2.3).

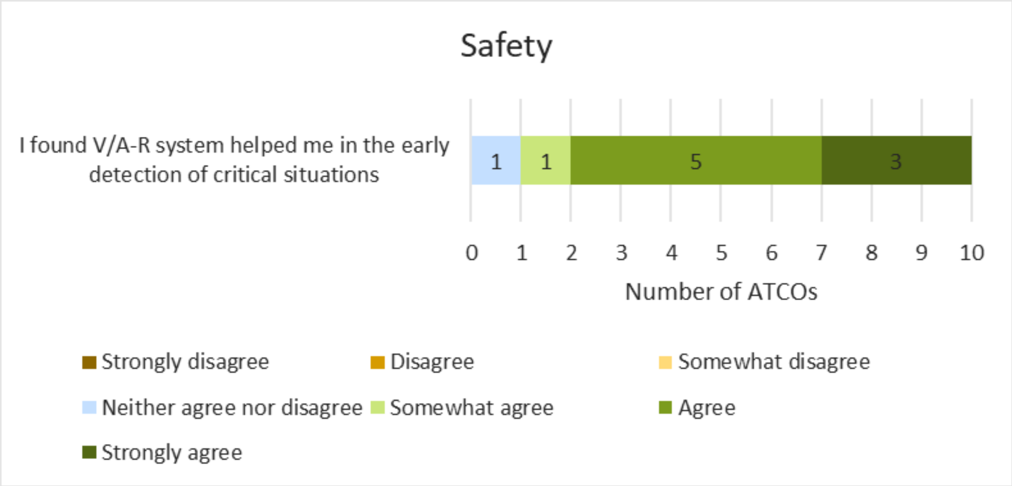


Figure B-75: Post-Exercise questionnaire answers related to safety

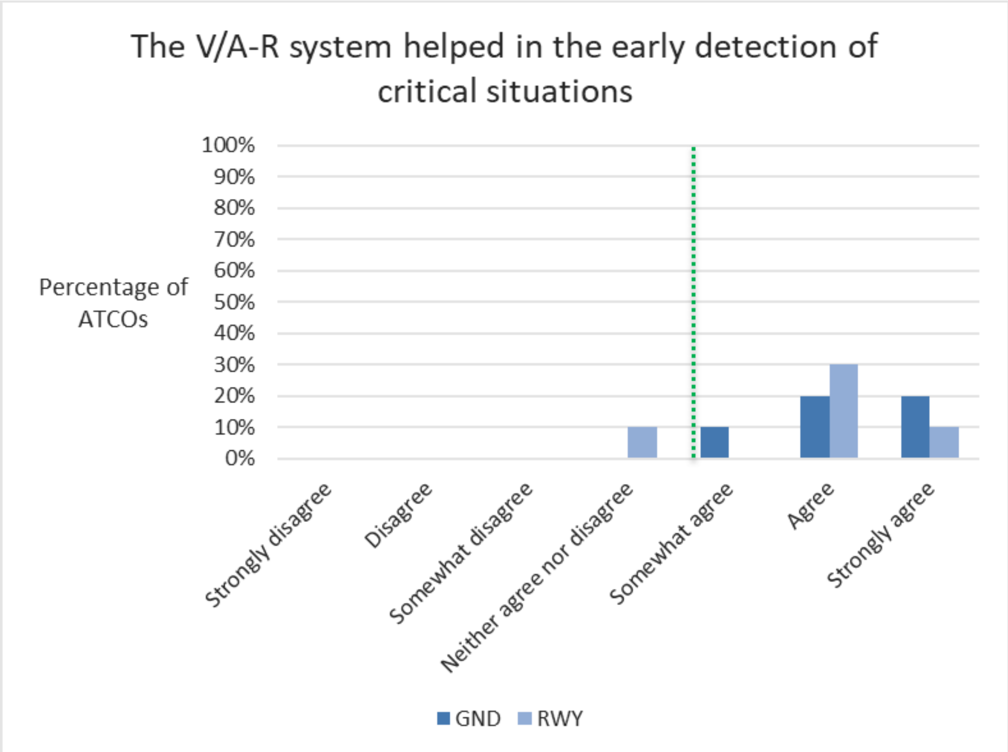


Figure B-76: Post-Exercise questionnaire answers related to early detection of critical situations for ground (GND) and runway (RWY) controller position

EX2-CRT-05.971-TLR4-TVALP-SAFE-1012

ATCO's workload with the implementation of Virtual/Augmented Reality applications is maintained at acceptable level and therefore not reducing safety levels.

In the post-run questionnaire, controllers reported an average workload of 2,7 (on a 10 point scale) for runs involving augmented reality, compared to 2,6 for reference scenario runs (see Figure in section EXE-OBJ-05.971-TRL4-TVALP-H103.1010 on workload). This means that workload was satisfactory without reduction. One controller mentioned that his confidence with the technical environment improved in subsequent runs. Indeed, average reported workload during run 5 (2,3) seems to be lower than average reported workload during run 2 (3,1). In the post-exercise questionnaire, only one controller reported 'slightly heavy' workload in the V/A-R scenarios compared to the reference scenarios, whereas the others all reported either very light (1), light (4), slightly light (1) or acceptable (3) workload levels (see Figure in EXE-OBJ-05.971-TRL4-TVALP-H103.1010 section on workload).

In the post-run questionnaire, 14 (70%) of the responses indicated that the level of safety during the preceding run was acceptable and no degradation or safety concern was raised. However, the average level of safety was lower in the solution scenario (5,1) compared to the reference scenario (6,3). This was due to the relatively low safety level reported by the ground controllers. Safety was said to be negatively impacted by the fact that clearance was not immediately visible, and the labels overlapping each other and covering the out of the window view. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the average level of safety for all scenarios and controller positions.

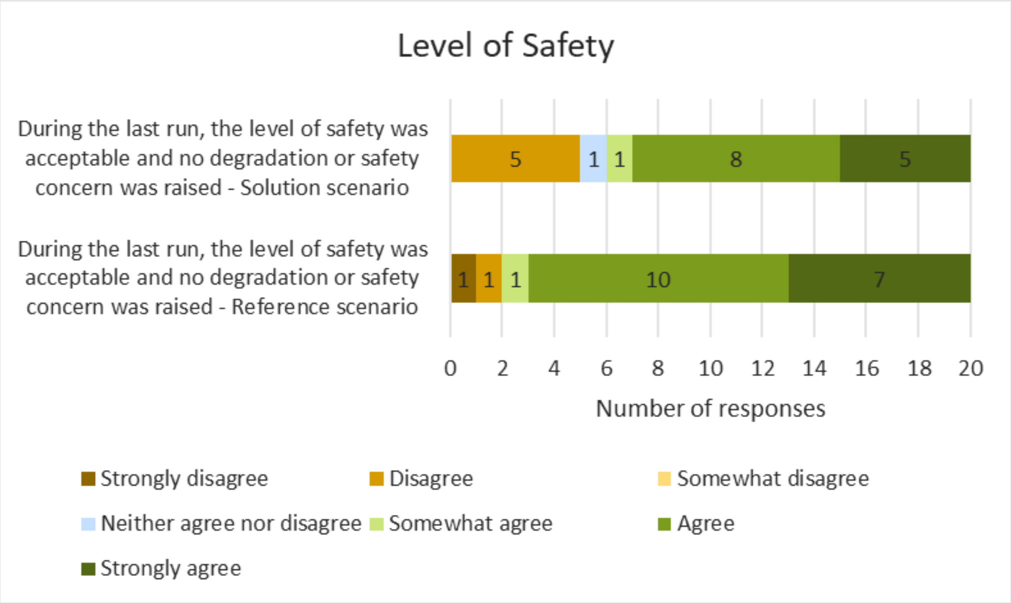
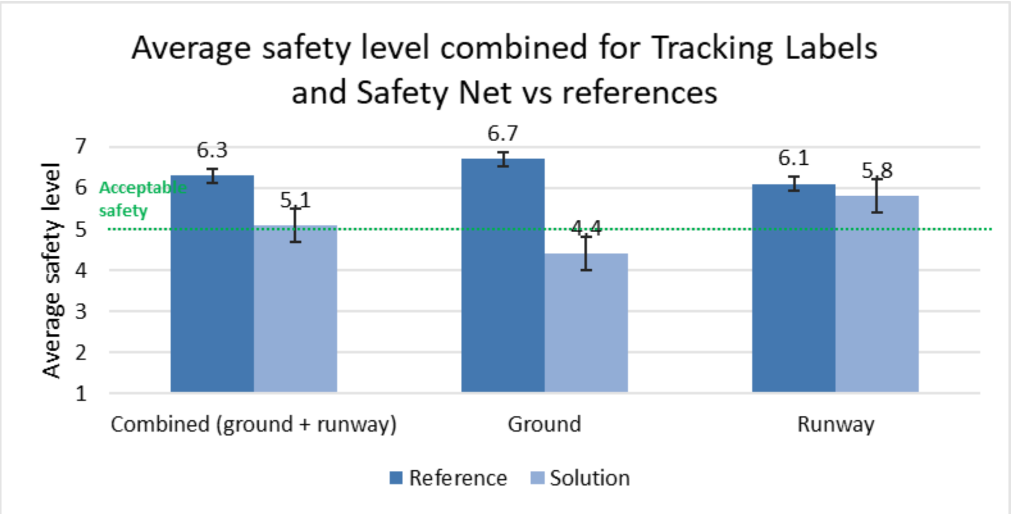


Figure B-77: Post-Run questionnaire answers related to level of safety



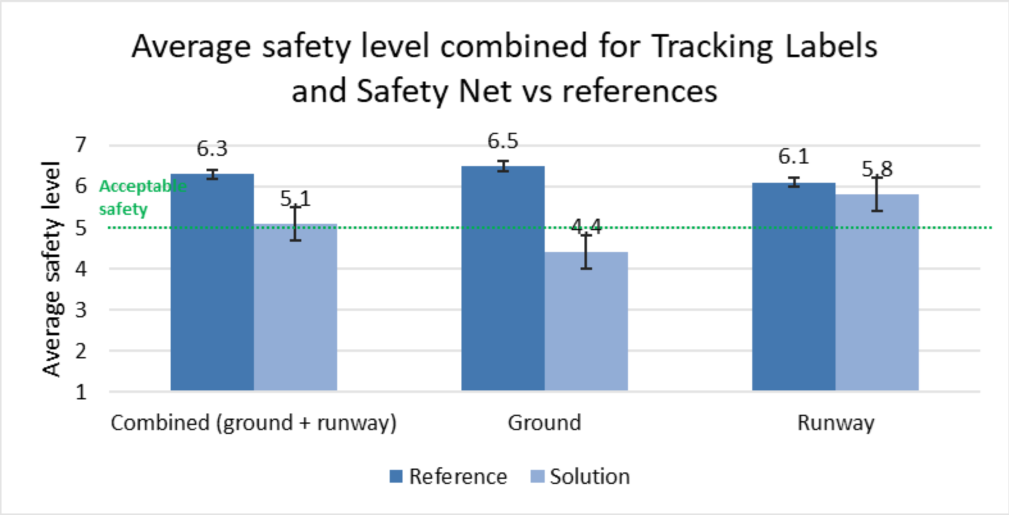


Figure B-78: Average level of safety in the reference and solution scenarios

The post-exercise questionnaire showed that eight out of ten ATCOs agreed that the overall subjective level of safety was at least as the today operations during the V/A-R exercise execution compared to the reference scenarios. See the figures below for the number of answers collected for each point of the 7-point Likert scale and the distribution of answers in percentage for both the ground and runway controller separately. In line with the other results, the runway controllers provided more positive responses to the question about overall safety.

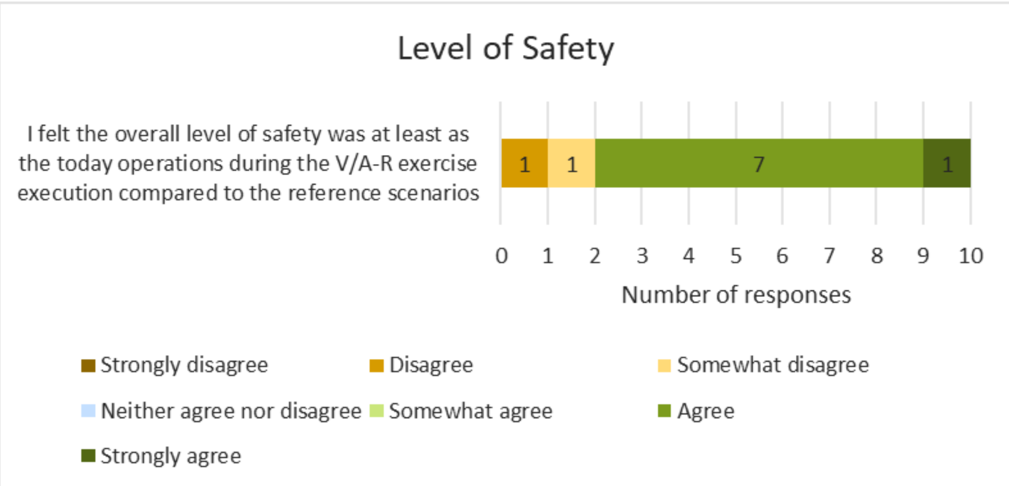


Figure B-79: Post-Exercise questionnaire answers related to level of safety

Co-funded by the European Union

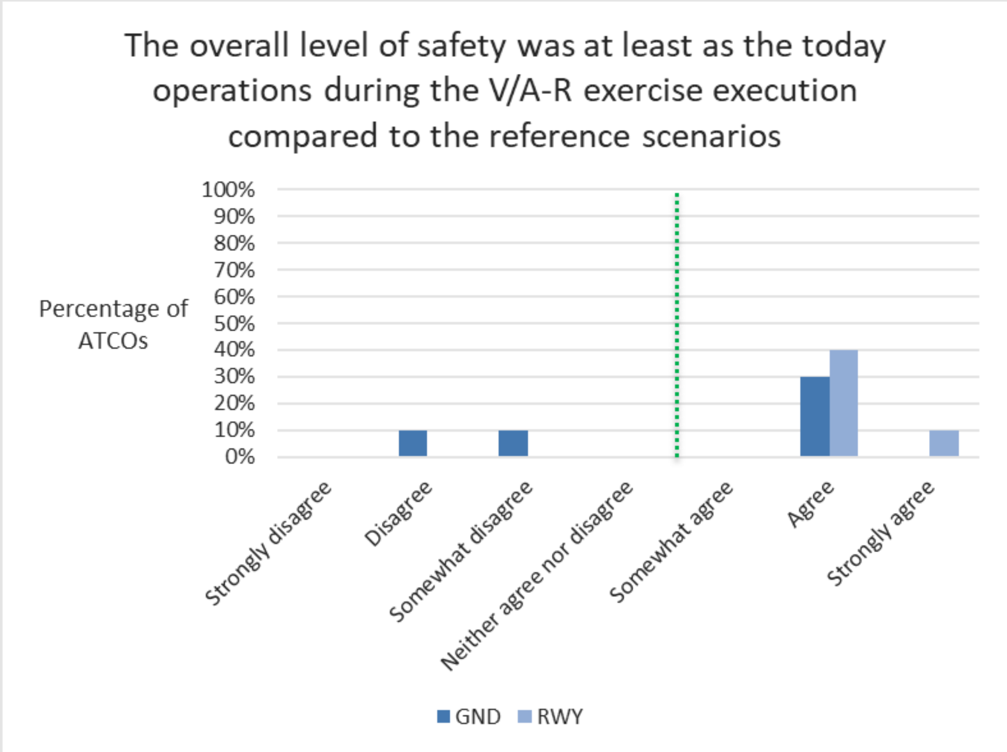


Figure B-80: Post-Exercise questionnaire answers related to level of safety for ground (GND) and runway (RWY) controller position

- The reaction time to the safety event decreased in the scenario with Safety Nets compared to the reference; from an average of 14 seconds to 9 seconds.

Figure B-81: Time to react to safety event in Reference and Solution scenario

EX2-CRT-05.971-TLR4-TVALP-SAFE-1013

ATCO's situational awareness with the implementation of Virtual/Augmented Reality applications is maintained at acceptable level and therefore not reducing safety levels.

In the post-run questionnaire, ATCOs reported an average Situational Awareness of 8,2 (on a 10 point scale) for runs involving augmented reality, compared to 9.1 for reference scenario runs (see Figure in EX2-OBJ-05.971-TRL4-TVALP-H103.1010 section on SA). This means that level of SA was satisfactory. In the post-exercise questionnaire, eight out of ten ATCOs rated their Situational Awareness during the V/A-R exercise execution compared to the reference scenarios as sufficient (1), high (6) or perfect (1) (see Figure in EX2-OBJ-05.971-TRL4-TVALP-H103.1010 section on SA). One ATCO responded that his SA was moderate and one ATCO said SA was degraded. Furthermore, apart from one ATCO, all ATCOs agreed that the information provided by V/A-R improved the SA during good visibility conditions with respect to the reference situation (see Figure in EX2-OBJ-05.971-TRL4-TVALP-H103.1010 section on SA). The comments in the questionnaire and the discussions during the debriefing made clear that the (potential) degradations in SA were due to the information representation. Specifically, the labels were said to overlap and cover the view. In addition, missing information and the need to turn the head towards the traffic in order to see the information were mentioned as causes for degraded SA.

In the post-run questionnaire, 14 (70%) of the responses indicated that the level of safety during the preceding run was acceptable and no degradation or safety concern was raised (see figure above). Safety was said to be negatively impacted by the fact that clearance was not immediately visible, and the labels overlapping each other and covering the out of the window view. The post-exercise questionnaire showed that eight out of ten ATCOs agreed that the overall subjective level of safety was at least as the today operations during the V/A-R exercise execution compared to the reference scenarios (see figure above).

B.7.2.10 EX2-OBJ-05.971-TLR4-TVALP-PERF.1010

To assess the performance benefits of equivalent visual operations for tower control through the use of applications for Virtual/Augmented Reality

EX2-CRT-05.971-TLR4-TVALP-PERF-1011

Laboratory tests show that the prototype for V/AR contributes to an assessment of Cost Efficiency performance by having a positive impact on situation awareness, workload and efficiency of ground operations.

In the post-exercise questionnaire, seven out of ten ATCOs rated their Situational Awareness during the V/A-R exercise execution compared to the reference scenarios as high (6) or perfect (1) (see Figure in EX2-OBJ-05.971-TRL4-TVALP-H103.1010 section on SA). Additionally, apart from one ATCO, all ATCOs agreed that the information provided by V/A-R improved the SA during good visibility conditions with respect to the reference situation (see Figure in EX2-OBJ-05.971-TRL4-TVALP-H103.1010 section on SA). ATCOs also pointed out that the technology can benefit SA, especially during unusual situations. One ATCO mentioned that the colours helped him to maintain SA.

Based on the post-run questionnaire results, the delta was calculated. The relative delta for the improvement of workload in the solution scenario in relation to the reference is -25% for runway controllers. For ground controllers, there was an increase of workload (relative delta ≈ 33%). In the post-exercise questionnaire, the majority of controllers reported very light (1), light (4), slightly light (1) workload levels in the V/A-R scenarios compared to the reference scenarios (see Figure in EXE-OBJ-05.971-TRL4-TVALP-H103.1010 section on workload). Controllers' comments indicated that the prototype for V/AR might provide a benefit in comparison to the reference situation in low visibility and unusual conditions. During the debriefings, controllers mentioned that the technology made the tasks, such as push back, intuitive and fast. Workload related to communication is decreased because it is not necessary to speak as much compared to the reference scenario. Moreover, all information that is need can be found in a concentrated place.

EX2-CRT-05.971-TLR4-TVALP-PERF-1012

Laboratory tests show that the use of V/A reality applications improves Resilience by increasing situational awareness in Low visibility conditions while maintaining workload within acceptable limits

In the post-exercise questionnaire, all ATCOs agreed that they found that the information provided by V/A-R improved the situational awareness during low visibility conditions with respect to the reference situation (see

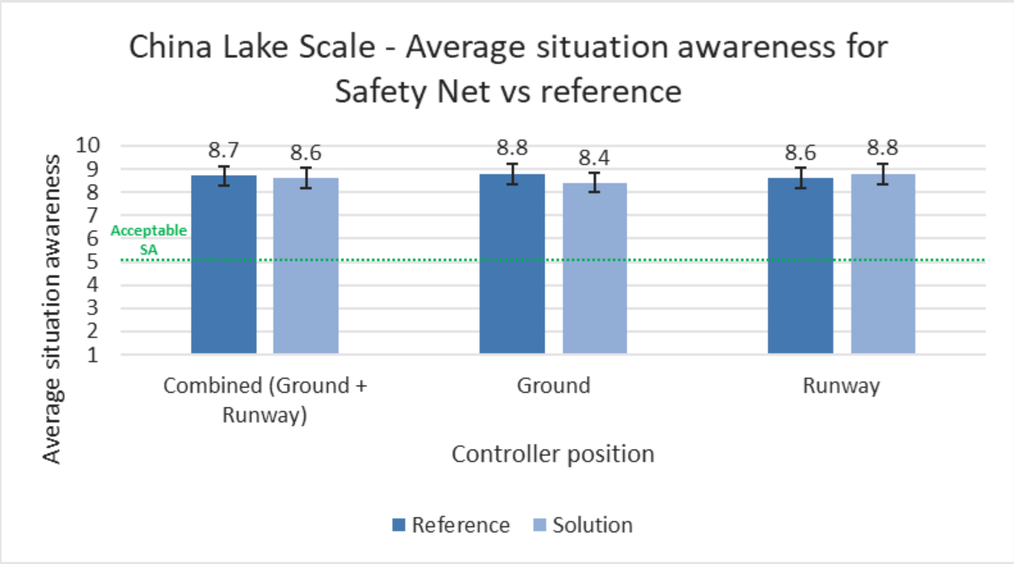


Figure B-87~~Figure B-87~~, previous section).

B.7.3 Unexpected behaviours/results

Although ATCOs agreed overall that workload and situational awareness improved compared to the reference scenario, the post-run results showed that there was no such improvement. This could potentially be due to the fact that the V/A-R tool is constantly presenting the ATCO with information in his field of view. This has both a positive and negative impact on human performance as the ATCO does not need to look for information in another place but can get tired because he is perceiving and processing the information. Furthermore, there is strong indication that workload would be lower and SA higher if the label design is improved and after some more familiarisation/training.

B.7.4 Confidence in results of EXE-002

B.7.4.1 Level of significance/limitations of Technological Validation Exercise Results

Simulation EXE2 has involved a wide range of Test subjects (10 ATCOs) with different background (Bologna airport mainly, but also ATCO from Torino, Forlì, Rimini and Ancona were involved) and expertise (6 years of experience to 33 years) with an age ranged from 30 years old to 59 years old in a simulation environment representing the Bologna operational environment with a high level of fidelity. Considering the simulation conditions, the results for V/A-R TL 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. This lack of training impact is anyway limited considering the very intuitive tools employed in the simulations.

The scope of V/A-R Air Gesture was reduced to a technical test. Indeed V/A-R Air Gesture run duration was only of 15 minutes and they were considered not representative to measure HP & SAF quantitative indicators. Indeed it was judged preferable to collect subjective feedback during the debriefing only.

B.7.4.2 Quality of EXE-002 results

One issue recorded for simulation day 2 was the temporary failure of HoloLens audio system that was overcome by the use of headsets. This issue anyway is judged not affecting the collection of data or the provided results, considering also the feedback provided by the ATCOs involved in the specific day.

Questionnaires have been used to collect ratings from the test subjects on the different aspects of V/A-R TL as explained in the method section (B.7.1.2): 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 TRL4 phase.

B.7.4.3 Significance of Technological Validation Exercises Results

The simulation exercise have been conducted on an experimental platform representing Bologna airport environment with a high degree of fidelity providing an operational significance adequate to support the TRL4 maturity assessment, of course with the limitations already mentioned in above sections 1 and 2.

A significant number of total run have been conducted among 5 simulation days (25 total number of runs) as well as a significant number of test subjects (10 ATCOs) have been involved to conclude that results are significant to support the TRL4 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.

B.8 Conclusions

B.8.1 Conclusions on technological feasibility

- V/A-R Tracking labels and airport overlay in LVC

Laboratory tests showed that the V/AR tracking labels and overlays can be effectively superimposed to the simulated out-of-the-tower view by means of a commercial see-through AR device. It was also demonstrated the feasibility of showing all needed information onto the AR overlays in a collimated head-up view for both TWR RWY and TWR GND operators at the same time.

Moreover, the possibility to use the visibility condition parameters (RVR) to trigger the activation of additional overlays was effectively implemented in a simulated environment. Also, the tests proved that safety nets can be displayed by means of AR overlays and safety warnings can be provided to the user via a mix of audio and visual cues exploiting the AR device hardware.

All tests were performed in a simulated environment, according to the maturity of the solution.

- V/A-R Air Gesture

Laboratory tests showed that, from a technical perspective, it is possible to use air gestures to interact with tracking labels in order to provide CDPLC messages to pseudopilots with specific reference to not time critical clearances (i.e. push-back and start-up). However, some ATCOs had difficulties when interacting with the system by means of air gestures. This aspect shall be further investigated.

B.8.2 Conclusions on performance assessments

B.8.2.1 Cost Efficiency Performance

- V/A-R Tracking labels and airport overlay in LVC
  - Laboratory tests showed that the prototype for V/AR contributes to an assessment of Cost Efficiency performance by having a positive impact on situation awareness, workload and efficiency of ground operations, especially in low visibility conditions.
- V/A-R Air Gesture
  - V/A-R Air Gesture was a pure technical test and it was not possible to have a complete HP assessment. According to the collected subjective feedback it cannot be concluded that cost efficiency performances were improved considering the difficulties related to its use that negatively impacted ATCOs' workload and efficiency of ground operations.

B.8.2.2 Human Performance

- V /A-R Tracking labels and airport overlay in LVC
  - The solution's impact on the following relevant topics were addressed and supported at V1/V2 level through questionnaires and debriefings: *mental and physical workload, quantity and quality of information, (team) situation awareness, head up time, human error, trust, usability, ergonomics, alarms, technology acceptance, job satisfaction and operating methods*. This was done in combination with a realistic simulation in which end-users performed realistic tasks both under normal and abnormal/degraded conditions.
  - This led to the confirmation of the solution's benefits for human performance as well as its open issues. The results indicated that the V/A-R tool provides a potential benefit for workload and situation awareness, especially in low visibility condition. However, in order to make a real impact on human performance, some things such as the label design need to be improved and the presented information might need to be adjusted. The Safety Net positively affected human error (amongst other things) and the head up time was improved in the solution scenarios.
  - The positive feedback from the ATCOs regarding acceptance indicate that the level of human performance needed to achieve the desired system performance for the proposed solution is consistent with human capabilities.
  - The outcomes of the validation exercise provide the necessary maturity to move to the next TRL4 validation phase.
- TRL4.V/A-R Air Gesture

- Air Gesture was addressed only at a qualitative level, in line with its maturity. Results indicated that the technology needs to be further improved in terms of usability in order to have a positive impact on human performance.

B.8.2.3 Safety

- V/A-R Tracking labels and airport overlay in LVC
  - Laboratory tests show that the prototype for V/AR with safety nets improves the perceived safety performance by reducing human error.
  - ATCO’s workload and situational awareness with the implementation of Virtual/Augmented Reality applications is maintained at acceptable level and therefore not reducing safety levels.
- V/A-R Air Gesture
  - ATCO’s workload and situational awareness with the implementation of Virtual/Augmented Reality Air Gesture application was not maintained at acceptable level and therefore potentially reducing safety levels.

B.8.3 Recommendations

B.8.3.1 Technological feasibility

- V/A-R Tracking labels and airport overlay in LVC
  - Refine the HMI design making the labels not overlapping each other nor shading any other objects.
  - Add a visual cue to show the borders of the augmented field of view to improve the ATCO’s awareness of the extension of the area/volume in which the overlays are shown with respect to the extension of the natural field of view.
  - Investigate the compatibility of AR see through device and prescription glasses, with specific reference to the ones used by farsighted subjects.
  - Further investigate the display of safety nets in head-up in non-nominal conditions.
  - Test the system in a real environment.
- V/A-R Air Gesture
  - Further investigation is needed to assess the impact a specific training on air gestures might have onto the user’s capability to correctly use the air gesture function.

B.8.3.2 Cost Efficiency Performance

- V/A-R Tracking labels and airport overlay in LVC
  - Cost Efficiency performance could be further improved by increasing the system’s positive impact on situation awareness and workload. This can be achieved by

EUROPEAN PARTNERSHIP
Co-funded by the European Union

improving the label design (i.e., position, overlapping and background colour of the labels).

- V/A-R Air Gesture
  - The V/A-R Air Gestures negatively impacted workload and SA as ATCOs had difficulties using them. The system should be further developed so that it recognises air gestures better. Potentially, also training and familiarisation will improve ATCOs performance using air gestures.

B.8.3.3 Human Performance

- V/A-R Tracking labels and airport overlay in LVC
  - Information requirements: To mitigate impacts on the workload, the amount and the position of the data should be taken into account.
    - Additional information should be provided regarding parking, clearances and speed during taxi and take-off.
    - Moreover, data should be recent (rather than static, e.g., with regards to weather information) and changes/refresh of data should be indicated.
    - Ideally, the A/V-R tool would be fully integrated with Automatic Dependence Surveillance in order to have data directly downloaded from the a/c. In addition, ATCOs suggested the complete integration of the radar, Automatic Dependence Surveillance and DATA-LINK.
    - The data of the V/A-R system should be equally confident as the data in the current system to ensure high levels of trust.
    - Furthermore, ATCOs experienced some difficulties to ‘find’ the information because sometimes the information is displayed outside of their current field of view. If feasible considering the capabilities of the technology, the information should be visible regardless of the direction in which the user is looking or the visual range should contain cues pointing to information that is currently falling outside of the field of vision. Additionally, more training is needed to get comfortable with the glasses.
    - Information could be put at the top of the head up display because they do not need that part of the view.
  - Design of the labels: Label design should be improved to maintain favourable levels with regards to workload, situation awareness, usability, trust and decrease the potential for Human Error.
    - The background colour of the label should be removed or changed according to the technology capabilities
    - It should be avoided to have labels overlapping each other or the final approach path and the labels should always be aligned.
    - The colour saturation of the label, its dimension, and possibly even the type of information presented, should be customisable.
  - Field of View:
    - Usability could be further improved by increasing the visual range. In order to provide optimal ergonomic comfort, the weight of the glasses should be lower.

        - In order to increase team SA, it could be interesting to provide the option for ATCOs to have the same view and/or to make something appear on the screen of the other ATCO.
    - It should be considered that acceptance might be higher in case the concept is introduced as optional, meaning that ATCOs can choose when to use it (e.g., only in low visibility conditions).

- V/A-R Air Gesture: In order to improve workload, SA, usability and level of acceptance, the air gestures should be less difficult to use. There should be a voice communication back-up and air gestures should not be used for critical cases to avoid potential for Human Error.

B.8.3.4 Safety

- V/A-R Tracking labels and airport overlay in LVC

    - Safety could be improved by making sure that the clearances are directly visible on the label and that the labels do not overlap each other or cover the out of the window view. In addition, some information (such as parking info) should be added and the information should be visible regardless of the direction in which the ATCO is looking. The Safety Net application should be kept as it positively impacted the potential for Human Error.

- V/A-R Air Gesture

    - Although, the concept of Air Gesture received positive feedback, they negatively impacted workload and SA as ATCOs had difficulties using them. The system should be further developed so that it recognises air gestures better. Potentially, also training and familiarisation will improve ATCOs performance using air gestures. Several ATCOs mentioned that there should be a voice communication back-up and that air gestures should not be used for critical cases.

- Safety results in relation to hazards

HZ ID	Hazard	Hazard Impact	Proposed evidence collection method/technique	EXE-002 (ENAV) Results
Hz1	Failure of the V/AR system (e.g. freezing of Tracking Labels or perceptual cues complete loss) prevents the AR device from being updated.	V/AR system is unresponsive potentially impacting trajectory management and associated safety nets.	Subjective: observation and feedback from ATCOs	No failure of the V/A-R system identified during the exercise. ATCOs only commented on the design, especially the position, of the Tracking Labels.
Hz2	TL is erroneously associated to a	ATCo may focus on the wrong a/c and issues the	Subjective: observation and	One ATCO mentioned that the TL was not always well aligned with the corresponding a/c. Moreover,

	wrong a/c (wrong information)	clearance intended for another a/c.	feedback from ATCOs	ATCOs commented repeatedly that the TLs were covering each other as well as the background. However, these issues did not lead to errors in associating the right TL to the a/c.
Hz3	Presentation of TL information within the V/AR does not support ATCo in task execution	If the presentation of TL on the AR device is inadequate (e.g. TLs overlapping, size issues – e.g. depth).	Subjective: observation and feedback from ATCOs	ATCOs commented repeatedly that the TLs were covering each other as well as the background. However, ATCOs were able to execute their tasks without too much difficulty (subjective SA and workload were acceptable).
Hz4	The responsiveness of the V/AR system as a result of ATCO's input/gesture or flight information changes) is inadequate for the accomplishment of operations.	If the use of V/AR introduces delays in the display of information, this may cause the ATCo to focus on V/AR until verified that the information has been correctly processed and displayed. This may have negative impact on ATCo situational awareness.	Subjective: observation and feedback from ATCOs	This hazard has not been identified in the exercise. ATCO subjective situational awareness was above the tolerable threshold.
Hz5	V/AR system fails to identify inputs (gestures) – no response	Unidentified inputs as a result of correct gestures may distract the ATCo from the primary task of ATS provision and results in temporary workload increase and reduction in situational awareness as	Subjective: observation and feedback from ATCOs	Air gestures were not always recognised easily. This led to frustration, and in some cased decreased SA and increased workload, for the ATCO.

		failures in TL inputs may require increased concentration on the V/AR system.		
Hz6	V/AR system wrongly identifies input (gesture) – provides erroneous output	In isolated instances the erroneous recognition of an input has no significant safety impact as the ATCo is able to correct the input. If the TV/AR system fails to perform in accordance with the specified gesture recognition threshold this may have a human performance impact causing disruption to the expected workflow and cognitive processes.	Subjective: observation and feedback from ATCOs	Air gestures were not always recognised easily. This led to frustration, and in some cased decreased SA and increased workload, for the ATCO.

Table B-14: EXE-002 Safety hazards

B.8.3.5 Final Debriefing evidences and separate results for Tracking Labels and Safety Nets

See below the tables with the results collected by using the Want/Have and the Human Performance Impact matrixes.

REMOVE (have but don't want)	PRESERVE (have and want)
• Visual range (Augmented Reality should not be limited to a portion of the view) • Air gesture • Low brightness of the lens	• Weather info • Flight info in the label • Air gesture (2) • Label (2)

• Paper strip • Background colour label (make labels transparent or rotate them) (3) • Arrival altitude info (2) • Presentation of fixed data (instead of customized data for each work station) • HoloLens helmet (limits freedom of movement) • Overlapping of labels (4) • Labels on final approach path • Fixed weather data • Very bright and flickering labels • Illumination of the taxiway (only highlight runway and stop bars)	• Essential label data • Safety net tools (warning about runway incursion) (5) • The functionality (of the label) for highlighting where the plane is • Highlighting of the runway and taxiways; runway colours in Low visibility Conditions (2) • Distance and altitude data
AVOID (don't have and don't want) • Visual noise • Air gesture commands for runway authorizations • Duplication of information in strips (paper + electronic); avoid using paper strip in addition to electronic strip • Label far away from A/C • Labels that cover A/C • Labels that hide VFR, AA/CC with which ATCO is not in contact or AA/CC without a flight plan • Too many items on label • Increased volume of information • To use the tool to increase capacity in Low Visibility Conditions (without adding additional supporting equipment) • Hindering direct observation (covering the view) • Reduction of brightness of the HoloLens • Air gesture for critical case	**ACHIEVE (don't have but want)** • Weather data that show changes of the weather (2) • Clean view (right amount of information) • Time-based separation tool to support the separation of Arrivals and Departures • Different colours "cyan" • Stripless • Field of view notices; Indicators that more info is available outside of the visual field (2) • Use of a less bright VR headset • Ability to adjust colour saturation and label size • Increase Field Of View augmented reality (2) • Improve the label alignment • Vehicles label • Parking info for the A/C when it is in final approach

	• Improve the presentation of the labels based on the operational structures • Complete radar integration • Full integration with Automatic Dependence Surveillance (= technology that downloads data directly from the a/c) • Integration of speech recognition (2) • Integration with systems in use, to remove old working methods • DATA-LINK integration • Labels that do not overlap • On Tower Ground Controller position weather always in the Field Of View • Highlight holding points (in Low Visibility Conditions) (3) • Highlight the runway in case of Visibility Condition 2 • Show arrival trajectory on runway when ATCO is authorising a/cs for take-off • Improved label design • Customisable set up • Label rotation • Smooth display of label (moves rather statically right now) • More flexibility on a/c labels • Information on clearances and clearances in progress (3) • Add speed of a/c taxiing and landing • Labels indicated by a symbol (not only by the line between the label and plane in Low Visibility Conditions) • Extend the (vertical) vision (above the tower, the parking, etc.) (2)

Table B-15: Want/Have Matrix results

Numbers in brackets indicate the number of ATCOs that repeated the comment.

	Virtual/Augmented Reality		Air Gesture		Safety Net	
	Ground	Runway	Ground	Runway	Ground	Runway
Workload	→→→↑↓	→↑↓→↓	↓→↑↓↓	N/a	↑→↑↑↑	↑↑↑→↑
Situational Awareness	→↓↑↑↑	→↑↑↑↑	→↓↑↓↑	N/a	↑↑↑↑↑	↑↑↑↑↑
Usability & Ergonomics	→↓↑↓↑	↓↓↑↓↓	↓↓↓↓→	N/a	↑↑→↑→	↑→→↓↑
Teamwork & Communication	→→↑→→	→→↑→↓	→→↑→↑	N/a	→↑↑→↑	↑→↑→↑
Trust	→→→↑↑	→→→↑↑	→→→↓↑	N/a	↑↑↑↑↑	↑→↑→↑
Human Error	↑↑↑↑↑	↓↑→↑↓	↓→↑↓→	N/a	↑↑↑↑↑	↑↑↑↑↑
Roles & Responsibilities	→→→→→	→→→→→	→→→→→	N/a	→→→↑↑	→→→→↑
Acceptance & Job Satisfaction	↓→↑↓↑	↓↑↑↑↑	→→↑↓↑	N/a	→→↑↑↑	↑→↑↑↑

Table B-16: HP Impact Matrix results

Upward arrows indicate an improvement whereas downward arrows indicated a degradation. A horizontal arrow indicates no chance in comparison to the reference scenario

As explained in the method section (B.7.1.2), there are two different reference scenarios: one without safety event and one with safety event. See below the workload, SA and safety level results separated for the scenarios including only Tracking Labels and for the scenarios including Safety Net (in relation to the corresponding reference scenarios).

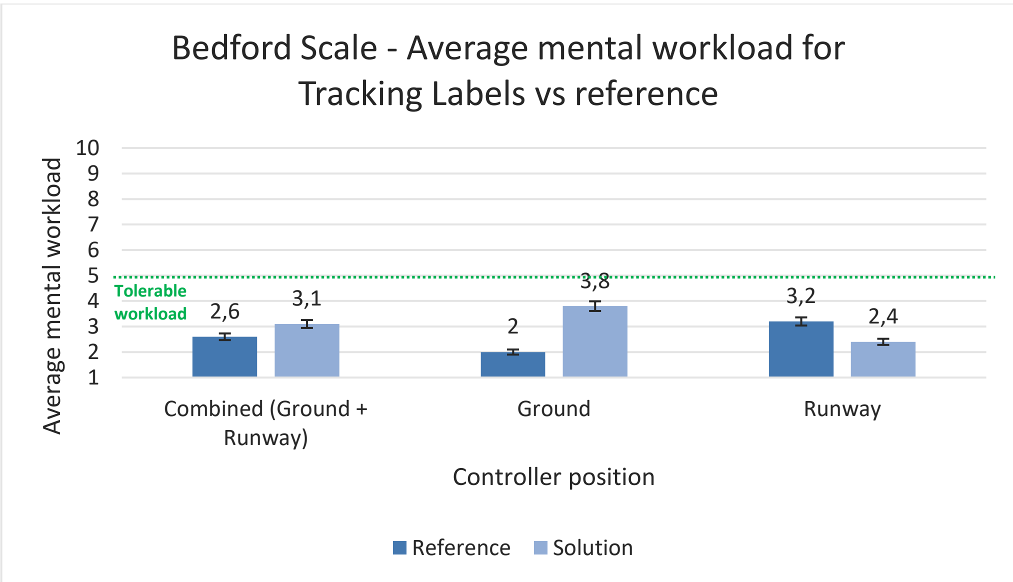
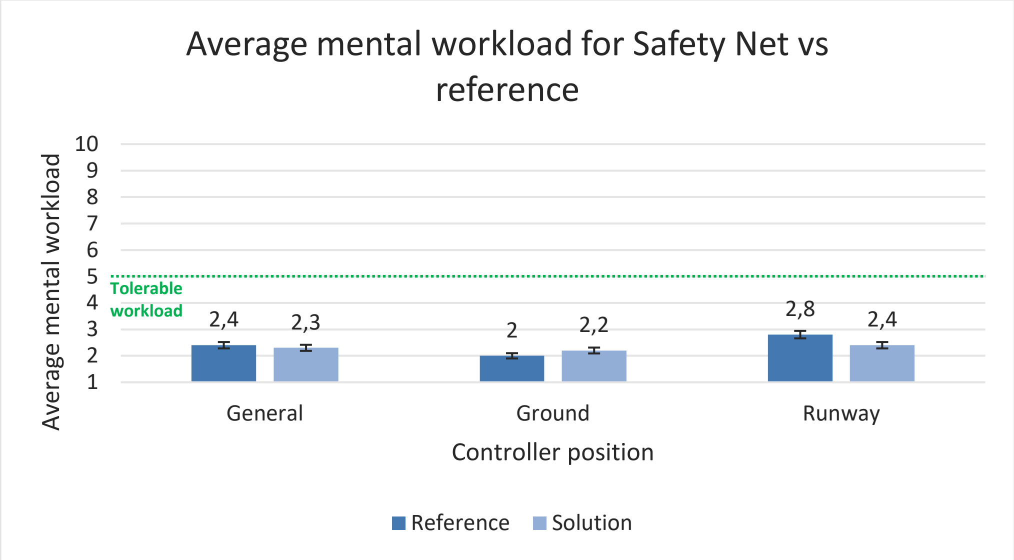


Figure B-82: Average mental workload in the reference and solution scenarios for the tracking labels

Co-funded by the European Union



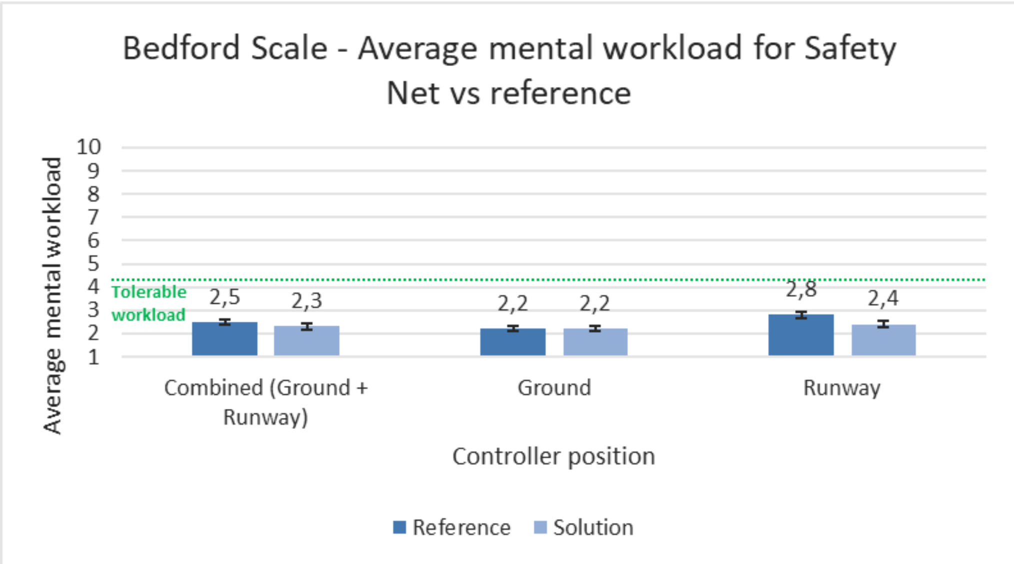


Figure B-83: Average mental workload in the reference and solution scenarios for safety net

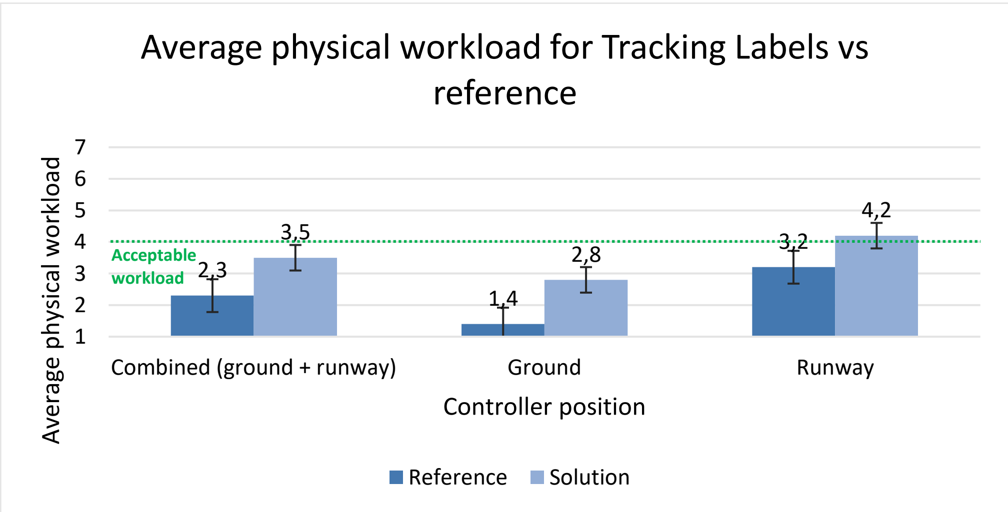


Figure B-84: Average physical workload in the reference and solution scenarios for tracking labels.

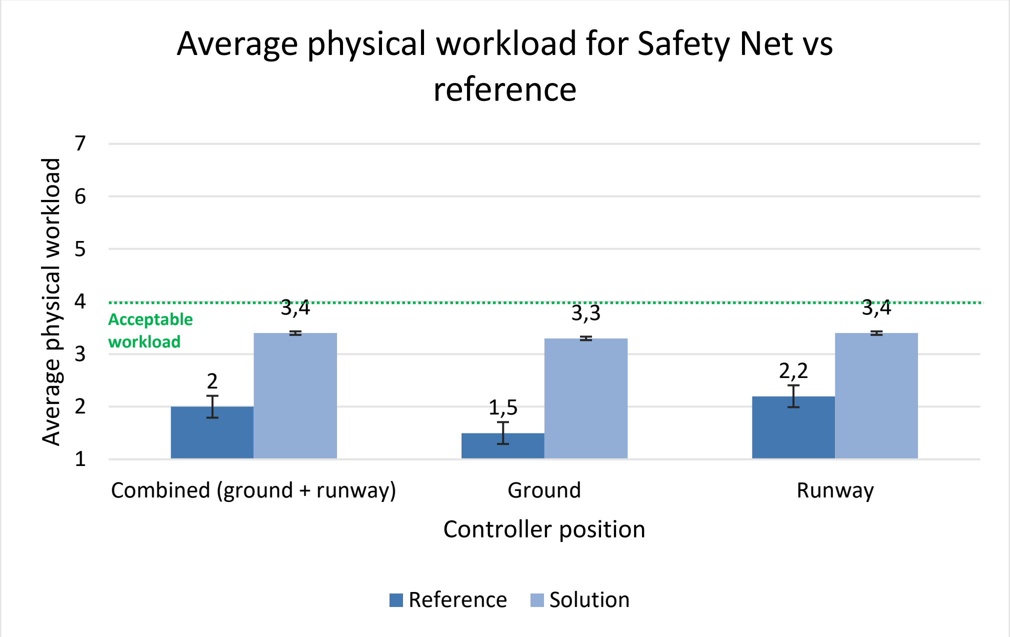


Figure B-85: Average physical workload in the reference and solution scenarios for safety net.

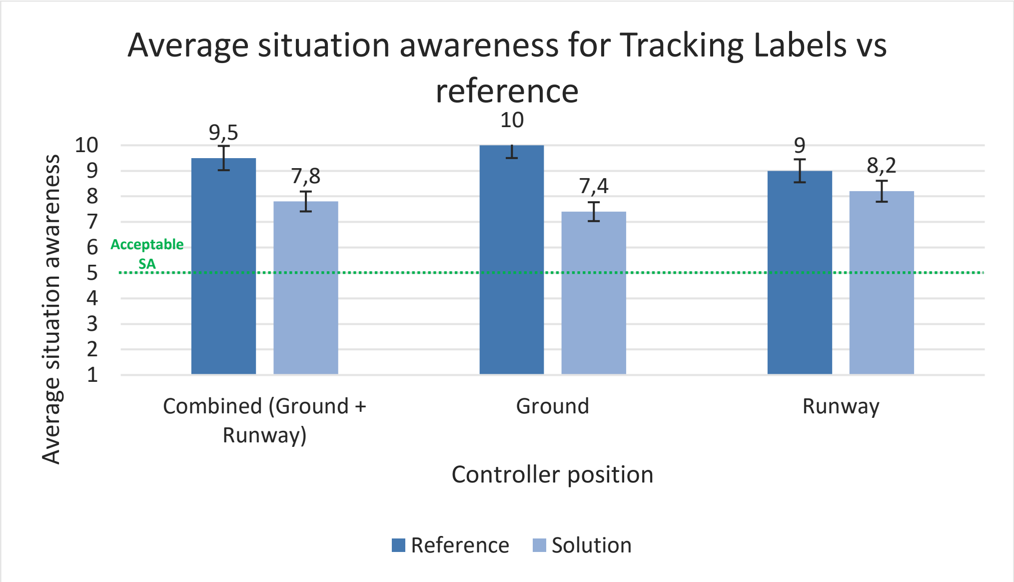
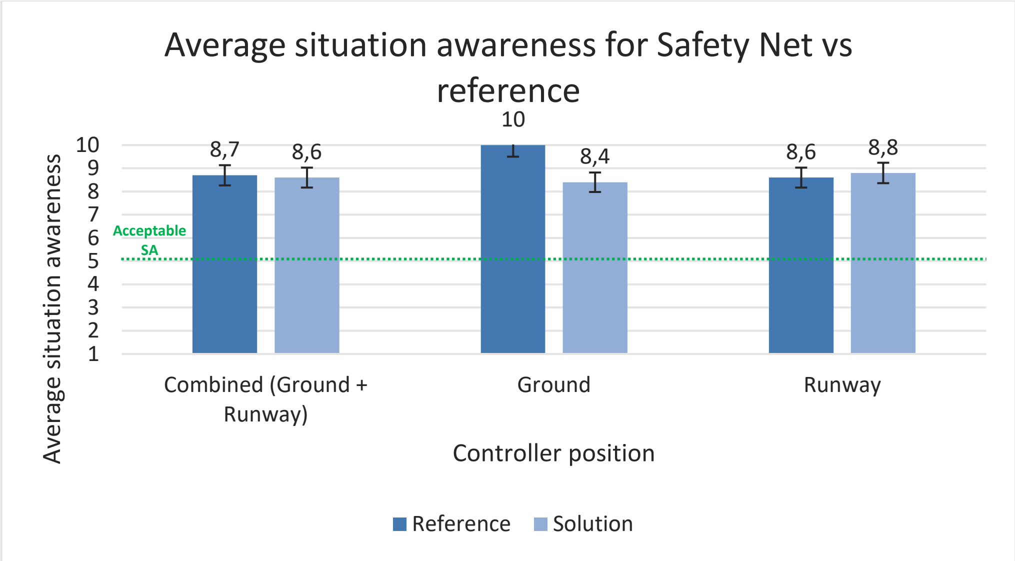


Figure B-86: Average situation awareness in the reference and solution scenarios for the tracking labels



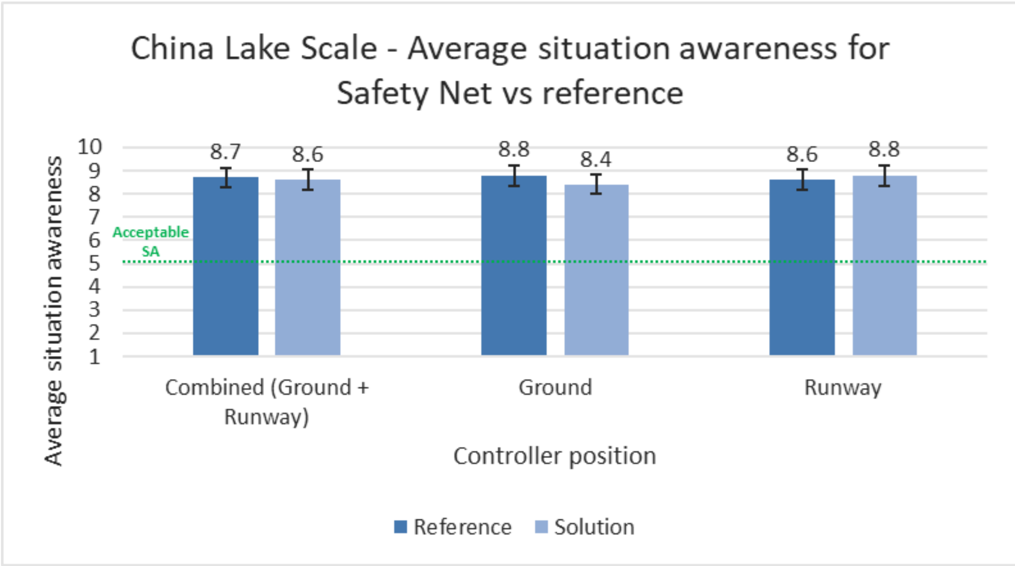


Figure B-87: Average situation awareness in the reference and solution scenarios for safety net.

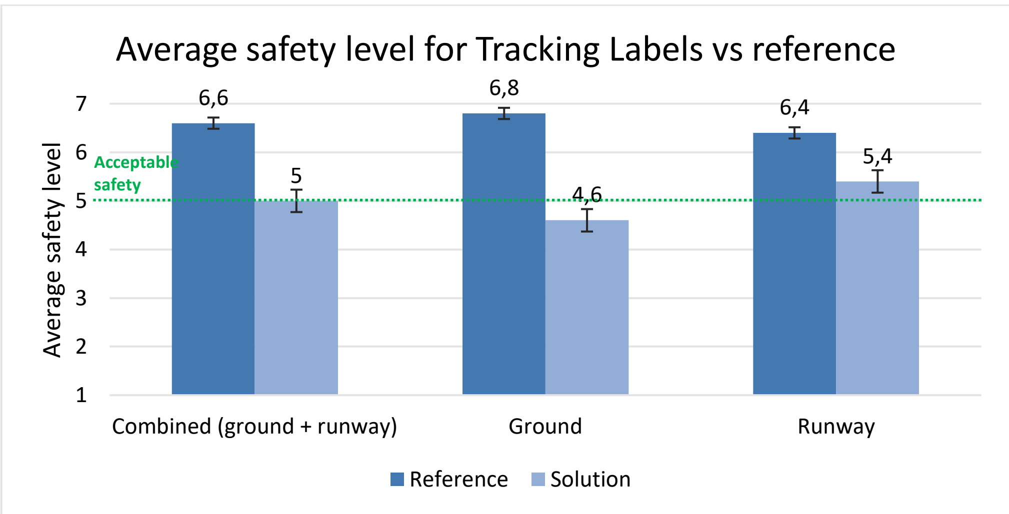
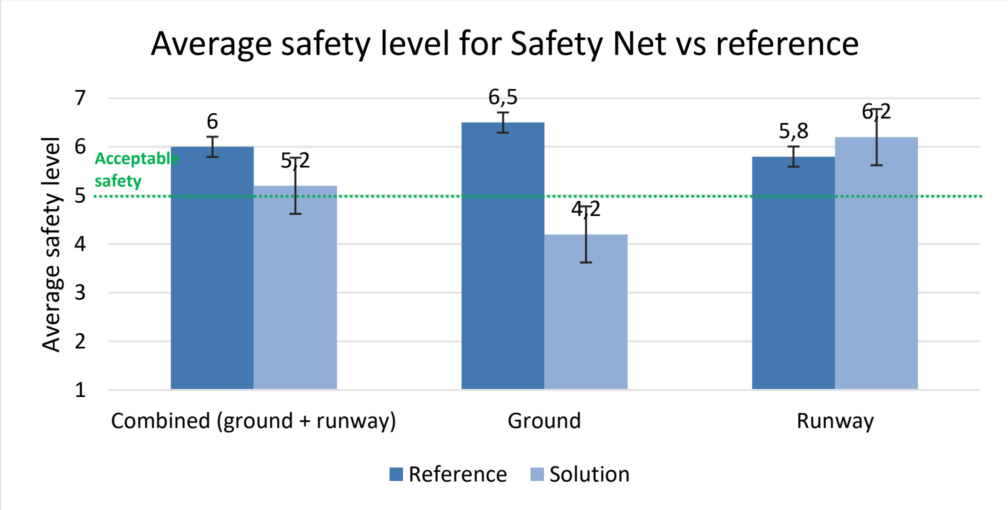


Figure B-88: Average level of safety in the reference and solution scenarios for tracking labels.



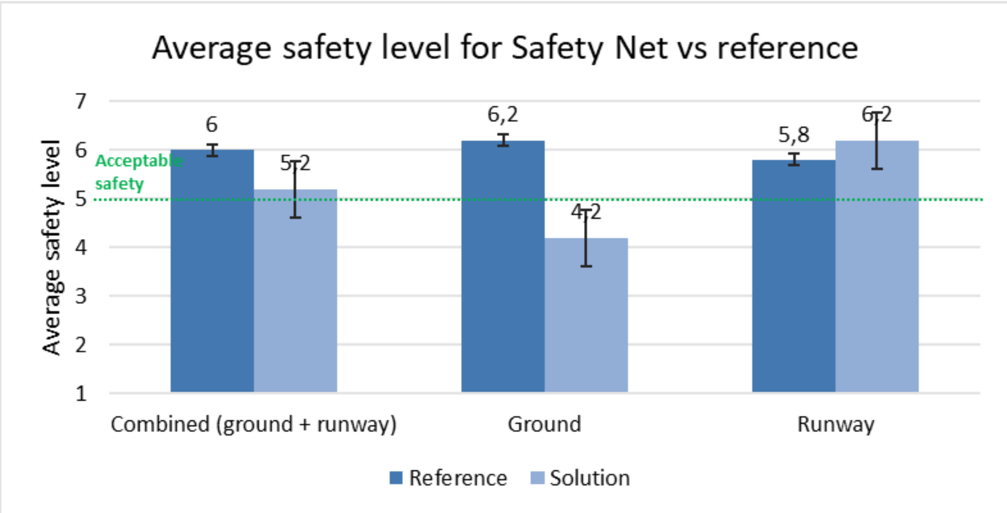


Figure B-89: Average level of safety in the reference and solution scenarios for safety net.

Appendix C Technological Validation Exercise 004 Report

C.1 Summary of EXE-004 plan

This chapter describes the technological validation exercise plan for EXE-05.97.1-TRL4-TVALP-ASR-004 carried out by INDRA and HungaroControl as part of Solution 97.2

C.2 EXE-004 description and scope

The validation exercise aimed to demonstrate the benefit of a speech recognition system with machine learning to support controllers' routine tasks in tower environment.

The exercise has been conducted jointly by Indra and HungaroControl in Asker, Norway, between 22nd November- 10th December 2021. The days on which the solution scenarios (with ASR) were performed were the followings:25-26th November and 9-10th December.

The exercise has been performed as a real-time simulation in remote tower environment. 6 ATCOs with active tower controller license participated in the validation exercise. Each ATCO worked with the ASR in three scenarios, each tailored to a different Norwegian aerodrome.

The high-level objective of this exercise was to be able to compare system with and without the ASR capability in series of human-in-the-loop simulation runs. The reference scenarios are considered to be the PJ05-35 Multiple Remote Tower exercises, which have been performed on the same week, right before the ASR validation. Therefore, ATCOs had a good idea of how they would manually operate the system if there was no ASR supporting them.

The industrial platform on which the validation exercise was performed is INDRA RTWR IBP platform. The platform consists of two Multi Remote Tower Modules (MRTM), but each MRTMs only had one (1) aerodrome operating in the ASR scenarios.

The validation exercise has demonstrated the following use cases:

Name	Description
UC-97-TRL4-TS-201	Highlighting of recognized callsign *The ASR recognizes a callsign from the verbal controller pilot communication and highlights it in the controllers' HMI for the relevant time duration. This helps the controller to check further information extracted from the complete utterance displayed in the HMI later on.*
UC-97-TRL4-TS-202	Showing full recognized utterance/command in HMI *The ASR system recognizes the complete utterance/command issued by the controller on the frequency and shows it on the controllers' HMI in a compact format. This includes recognition of the callsign, command types, command values, units, qualifiers and conditions if*

Name	Description
	applicable (e.g., DLH123 PUSHBACK, DLH123 TAXI VIA A B, DLH123 LINEUP RW23R).
UC-97-TRL4-TS-203	Manual manipulation of an ASR output *If the controller decides that the output of the ASR system is not correct, he/she has to correct the recognized command types and/or the recognized values or even the recognized callsign.*
UC-97-TRL4-TS-204	Automatic Acceptance of ASR output *When a command is recognized by the ASR system, it is then shown to the controller. If the controller does not reject the command within an (adjustable) time frame (e.g. 10 seconds), the recognition result is automatically accepted.*

The expected benefits were the followings:

- Human Performance
- Safety
- Cost Efficiency

C.3 Summary of EXE-004 objectives and success criteria

The Exercise will address all the objectives and success criteria defined in chapter 4.1, and all would be fully covered qualitative and quantitative assessment.

SESAR Technological Validation Objective ID/ /Description /Title	**SESAR Technological Validation Success Criteria ID/Description**	Exercise Validation Objective	Exercise Success Criteria
OBJ-05.972-TLR4-TVALP-FEAS.2010 To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS. TLR4 Operational feasibility	CRT-05.972-TLR4-TVALP- FEAS -2011 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Automatic speech recognition.	Same as solution	Same as solution

	CRT-05.972-TLR4-TVALP- FEAS -2012 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of AI suggestions.	Same as solution	Same as solution
OBJ-05.972-TLR4-TVALP- FEAS.2020 To identify possible technical feasibility issues and possible show stoppers TLR4 Pre-industrial feasibility	CRT-05.972-TLR4-TVALP- FEAS -2021 Laboratory tests (based on a prototype) have verified the technical feasibility of the use of ASR supported by AI/ML	Same as solution	Same as solution
	CRT-05.972-TLR4-TVALP- FEAS -2022 Laboratory tests have verified that the integration of the SESAR technological solution with other related system enablers is technically feasible	Same as solution	Same as solution
OBJ-05.972-TRL4-TVALP-H106.2010 To assess that the technical systems for ASR support the ATCOs in performing their tasks ASR impact on ATCO tasks	CRT-05.972-TLR4-TVALP-H106-2011 Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining workload at acceptable level	Same as solution	Same as solution
	CRT-05.972-TLR4-TVALP-H106-2012 Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining an adequate level of situation awareness	Same as solution	Same as solution
	CRT-05.972-TLR4-TVALP- H106-2013 ASR does not increase the potential for human error	Same as solution	Same as solution
	CRT-05.972-TLR4-TVALP- H106-2014 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by ASR	Same as solution	Same as solution
	CRT-05.972-TLR4-TVALP-H106-2015 Measured callsign recognition rate,	Same as solution	Same as solution

	command recognition rate, error rate and rejection rate of ASR system are at acceptable levels are considered within acceptable levels by the majority of ATCOS (at least 75%)		
	CRT-05.972-TLR4-TVALP- H106-2016 Majority of ATCOs (at least 75%) confirm adequate usability of ASR system	Same as solution	Same as solution
	CRT-05.972-TLR4-TVALP- H106-2017 Majority of ATCOs (at least 75%) provide positive feedback on acceptance of ASR tool	Same as solution	Same as solution
	CRT-05.972-TLR4-TVALP- H106-2018 ATCOs (at least 75%) trust in the system is at an acceptable level	Same as solution	Same as solution
OBJ-05.972-TRL4-TVALP-H106.2020 To assess the role of the ATCO is consistent with human capabilities and limitations with the introduction of ASR ASR impact on ATCO role	CRT-05.972-TLR4-TVALP- H106-2021 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient, and timely manner	Same as solution	Same as solution
	CRT-05.972-TLR4-TVALP-H106-1022 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions	Same as solution	Same as solution
OBJ-05.972-TRL4-TVALP-H106.2030 To assess job acceptance and satisfaction with the introduction of ASR ASR impact on job satisfaction	CRT-05.972-TLR4-TVALP-H106-2031 ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance	Same as solution	Same as solution
OBJ-05.972-TLR4-TVALP-SAFE.2010	CRT-05.972-TLR4-TVALP- SAFE -2011 The changes related to the implementation	Same as solution	Same as solution

To assess the impact of Automatic Speech Recognition on safety. Safety Impact	of Automatic Speech Recognition do not increase potential for human error and therefore not reducing safety levels.		
	CRT-05.972-TLR4-TVALP- SAFE -2012 ATCO's workload with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	Same as solution	Same as solution
	CRT-05.972-TLR4-TVALP- SAFE -2013 ATCO's situational awareness with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	Same as solution	Same as solution
	CRT-05.97B-TLR4-TVALP- SAFE -2014 Safety assessment activities and the results are documented and integrated in the overall solution validation results	Same as solution	Same as solution
OBJ-05.972-TLR4-TVALP-PERF.2010 To assess the performance benefits of Automatic Speech Recognition supported by AI/ML. TLR4 Performance Assessment	CRT-05.972-TLR4-TVALP-PERF-2011 Laboratory tests show that the SESAR technological solution improves Cost Efficiency performance by reducing cost per flight(PER.TRL4.3) (through e.g. reduction of workload, reduction of delay times,….)	Same as solution	Same as solution

C.4 Summary of validation scenarios addressed in EXE-004

C.4.1 Reference scenarios

In current tower operations environment, the controller issues ATC clearances or instructions and provides information to the flight crews by voice communications. The flight crew is expected to confirm the clearance by a readback or acknowledge the information – this means instant feedback to the ATCO.

For their effective operation, Tower systems need accurate data in timely manner. One of the necessary input data of the Tower systems are the Controller clearances and instructions. This input is done manually by the ATCO using mouse.

The reference scenarios represent the current way of working in the current operational environment in terms of technology (i.e. manual input). No direct comparison will be done, as the simulated operational environment was significantly different between the reference and solution scenario (multiple remote tower with 3 aerodromes vs single remote tower, respectively). Yet the reference scenario provided the opportunity to perform similar tasks on the same system, thus ATCOs could get a solid impression of the impact the ASR could have on task efficiency and effectiveness.

C.4.2 Solution scenarios

The technical validation exercise has been conducted in a (single) Remote Tower environment in Asker. Three airports have been selected and used during the validation. Each ATCO controlled only one aerodrome at the same time, but over the course of 3 exercises they worked on all the three aerodromes. Two controllers have participated in each run, providing ATC from two different CWP.

The solution scenarios address the conditions when the ASR is enabled. An ASR system is intended to support the ATCO performing manual work by automatically recognizing the verbal clearances and instructions.

In the Solution Scenarios:

- The ASR recognizes a callsign and highlights it in the controllers' HMI.
- The ASR system recognizes the complete instructions issued by the controller on the frequency and shows it on the controllers' HMI.
- When a command is recognized by the ASR system, it is then shown to the ATCO. The recognized command will be evaluated by a plausibility checker that uses machine learning and contextual information to indicate the plausibility of the command to the ATCO (ABSR).
- If the ATCO decides that the output of the ASR system is not correct, he/she must reject the recognized command and manually input the correct command into the system. If not rejected, the command is automatically accepted by the system after a configurable time-out.

The following scenarios were played:

Week 1			
	Aerodrome 1	Aerodrome 2	Aerodrome 3
1/1	ATCO 1	ATCO 2	
1/2	ATCO 3		
2/1	ATCO 2		ATCO 1
2/2		ATCO 3	ATCO 2
3/1		ATCO 1	

3/2			ATCO 3

Week 2			
	Aerodrome 1	Aerodrome 2	Aerodrome 3
1/1	ATCO 4	ATCO 5	
1/2	ATCO 6		
2/1	ATCO 5		ATCO 4
2/2		ATCO 6	ATCO 5
3/1		ATCO 4	
3/2			ATCO 6

Table C-17: Scenarios in EXE-ASR-004

C.5 Summary of EXE-004 Assumptions

Identifier	Title	Description	Justification	Impact on Assessment

AS-EXE.004-01	Limited Simulation Scope – Hybrid type of validation	Due to the nature of the validation the technical side will be emphasized, even if it impacts the scenario realism	ASR implementation in the tower system was not mature enough for operational use. The validation was in between a technical and an operational validation. It took place in a relatively operational environment and not in a laboratory, and during the validation the ATCOs were following a validation scenario with predefined traffic, but the ASR implementation was not designed for any operational use and by running a validation scenario instead of a technical laboratory testing with planned and consistent input, it had influence on the results. Therefore, the RTS is more like a usability test with some use cases that the ATCOs have to focus on. Measures of e.g. workload and situational awareness are therefore limited.	High

C.6 Deviation from planned activities

Whilst it was not necessarily a deviation from the plan, the team got re-assured that this simulation could not be an operational test in the sense that the solution was not sufficiently mature to support the usual type of ATCO involvement.

Human performance assessment was therefore minimal, limited to usability (i.e. system performance, HMI design, trust).

C.7 EXE-004 validation results

C.7.1 Summary of Technological Validation Exercise EXE-004 results

SESAR Technological Validation Objective ID/ /Description /Title	SESAR Technological Validation Success Criteria ID/Description	SESAR Technological Validation Results	SESAR Technological Validation Objective Status
OBJ-05.972-TLR4-TVALP-FEAS.2010 To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS. TLR4 Operational feasibility	CRT-05.972-TLR4-TVALP- FEAS - 2011 No operational showstoppers have been identified during laboratory tests (based on a prototype) related to the use of Automatic speech recognition.	The validation exercise did not reveal any operational showstoppers	OK
	CRT-05.972-TLR4-TVALP- FEAS - 2012 No operational showstoppers have been identified during laboratory tests (based on a prototype) related to the use of AI suggestions.	The validation exercise did not reveal any operational showstoppers	OK
OBJ-05.972-TLR4-TVALP- FEAS.2020 To identify possible technical feasibility issues and possible showstoppers TLR4 Pre-industrial feasibility	CRT-05.972-TLR4-TVALP- FEAS - 2021 Laboratory tests (based on a prototype) have verified the technical feasibility of the use of ASR supported by AI/ML	It is technically feasible to use ASR supported by AI/ML to assist, or automate, selected Aerodrome ATC system inputs.	OK
	CRT-05.972-TLR4-TVALP- FEAS - 2022 Laboratory tests have verified that the integration of the SESAR technological solution with other related system enablers is technically feasible	The validation exercise verified the integration between an automatic speech recognition module and an Aerodrome ATC system.	OK

OBJ-05.972-TRL4-TVALP-H106.2010 To assess that the technical systems for ASR support the ATCOs in performing their tasks ASR impact on ATCO tasks	CRT-05.972-TLR4-TVALP-H106-2011 Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining workload at acceptable level	Not applicable as workload could not be measured during the tests	N/A
	CRT-05.972-TLR4-TVALP-H106-2012 Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining an adequate level of situation awareness	Not applicable as situational awareness could not be measured during the tests	N/A
	CRT-05.972-TLR4-TVALP- H106-2013 ASR does not increase the potential for human error	Not applicable as human error could not be measured during the tests	N/A
	CRT-05.972-TLR4-TVALP- H106-2014 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by ASR	Only 66% of the ATCOs agreed that the level of information provided by the ASR is adequate.	POK
	CRT-05.972-TLR4-TVALP-H106-2015 Measured callsign recognition rate, command recognition rate, error rate and rejection rate of ASR system are at acceptable levels are considered within acceptable levels by the majority of ATCOS (at least 75%)	The subjective perception of callsign recognition, clearance recognition and understanding other parameters were below the cut-off line (at least 75%). However, the speed and accuracy of the callsign recognition was highly regarded.	POK

Co-funded by the European Union

	CRT-05.972-TLR4-TVALP- H106-2016 Majority of ATCOs (at least 75%) confirm adequate usability of ASR system	The usability in terms of ASR performance was not optimal, and the HMI related questions did not meet the cut-off score criteria either as the ASR HMI was not designed for an operational validation. However, the design of the ASR module was overall acceptable for the ATCOs and also many improvement ideas have been gathered to further enhance the system.	POK
	CRT-05.972-TLR4-TVALP- H106-2017 Majority of ATCOs (at least 75%) provide positive feedback on acceptance of ASR tool	The usability in terms of ASR performance was not optimal, and the HMI related questions did not meet the cut-off score criteria either as the ASR HMI was not designed for an operational validation. However, the design of the ASR module was overall acceptable for the ATCOs and also many improvement ideas have been gathered to further enhance the system.	POK
	CRT-05.972-TLR4-TVALP- H106-2018 ATCOs (at least 75%) trust in the system is at an acceptable level	The usability in terms of ASR performance was far from optimal, and the SATI results did not meet the cut-off score criteria either.	NOK
OBJ-05.972-TRL4-TVALP-H106.2020 To assess the role of the ATCO is consistent with human capabilities and limitations with the introduction of ASR ASR impact on ATCO role	CRT-05.972-TLR4-TVALP- H106-2021 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient, and timely manner	Not applicable as it was not a fully operational validation therefore this objective could not be measured during the tests	N/A

	CRT-05.972-TLR4-TVALP-H106-1022 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions	Not applicable as it was not a fully operational validation therefore this objective could not be measured during the tests	N/A
OBJ-05.972-TRL4-TVALP-H106.2030 To assess job acceptance and satisfaction with the introduction of ASR ASR impact on job satisfaction	CRT-05.972-TLR4-TVALP-H106-2031 ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance	Not applicable as it was not a fully operational validation therefore this objective could not be measured during the tests	N/A
OBJ-05.972-TLR4-TVALP-SAFE.2010 To assess the impact of Automatic Speech Recognition on safety. Safety Impact	CRT-05.972-TLR4-TVALP- SAFE - 2011 The changes related to the implementation of Automatic Speech Recognition do not increase potential for human error and therefore not reducing safety levels.	Not applicable as human error could not be measured during the tests.	N/A
	CRT-05.972-TLR4-TVALP- SAFE - 2012 ATCO's workload with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	Not applicable as workload could not be measured during the tests	N/A
	CRT-05.972-TLR4-TVALP- SAFE - 2013 ATCO's situational awareness with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	Not applicable as situational awareness could not be measured during the tests	N/A

	CRT-05.97B-TLR4-TVALP- SAFE - 2014 Safety assessment activities and the results are documented and integrated in the overall solution validation results	Safety assessment activities, conducted according to SESAR SRM, are documented in TSAR.	OK
OBJ-05.972-TLR4-TVALP-PERF.2010 To assess the performance benefits of Automatic Speech Recognition supported by AI/ML. TLR4 Performance Assessment	CRT-05.972-TLR4-TVALP-PERF-2011 Laboratory tests show that the SESAR technological solution improves Cost Efficiency performance by reducing cost per flight (PER.TRL4.3) (through e.g. reduction of workload, reduction of delay times,....)	The exercise was not designed as an operational use of ASR and to measure workload	N/A

C.7.1.1 Results on technological feasibility

C.7.1.2 Results per KPA

C.7.2 Analysis of EXE-004 Results per Technological Validation objective

C.7.2.1 EX4- OBJ-05.972-TRL4-TVALP- FEAS.2010

CRT-05.972-TLR4-TVALP- FEAS -2011

Success criteria	Result
No operational showstoppers have been identified during laboratory tests (based on a prototype) related to the use of Automatic speech recognition.	OK. No operational showstoppers were identified

CRT-05.972-TLR4-TVALP- FEAS -2012

Success criteria	Result
No operational showstoppers have been identified during laboratory tests (based on a prototype) related to the use of AI suggestions.	OK. No operational showstoppers were identified related to the use of AI suggestions.

C.7.2.2 EX4- OBJ-05.972-TRL4-TVALP- FEAS.2020

CRT-05.972-TLR4-TVALP- FEAS -2021

Success criteria	Result
Laboratory tests (based on a prototype) have verified the technical feasibility of the use of ASR supported by AI/ML	OK. Technically it should be feasible to automate some manual inputs in an ATC system based on the clearances the ATCO issue by voice to pilots. The exercise only explored recognizing the speech of the ATCO. The speech of the flight crew received by radio would be of lower quality, but is also something that could be automatically recognized for simple system inputs – e.g. highlighting the flight on a situational display or highlighting an electronic flight strip.

CRT-05.972-TLR4-TVALP- FEAS -2022

Success criteria	Result
Laboratory tests have verified that the integration of the SESAR technological solution with other related system enablers is technically feasible	OK. The integration of an ASR module with an ATC system to aid input of clearances into the system based on clearances issued to pilots by radiotelephony is feasible.

C.7.2.3 EX4- OBJ-05.972-TRL4-TVALP-H106.2010

CRT-05.972-TLR4-TVALP-H106-2011

Success criteria	Result
Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining workload at acceptable level.	Not applicable as workload could not be measured during the tests

Although workload could not be measured during the tests, ATCOs mentioned that the strength of the ASR lies in the potential to reduce workload. If the system could recognize all kinds of accents and clearances with high confidence, it could truly positively affect workload and efficient task performance.

CRT-05.972-TLR4-TVALP-H106-2012

Success criteria	Result
Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining an adequate level of situation awareness	Not applicable as situational awareness could not be measured during the tests

Although situational awareness could not be addressed in the tests, ATCOs expressed their concerns that this solution could negatively impact their SA by not following the current status of the system due to the automatic system updates.

1. During the reference scenario they felt that they had the opportunity for "self-check". However, with the ASR scenario the feeling of checking themselves was lost as the system took over the manual input after they provided the clearance or instruction. But if they have to continuously check the system, according to the ATCOs, no actual progress has been made in terms of workload reduction.

CRT-05.972-TLR4-TVALP- H106-2013

Success criteria	Result
ASR does not increase the potential for human error	Not applicable as human error could not be measured during the tests

Although human error could not be explicitly measured, the feedback received with regards to situational awareness applies here as well. ATCOs expressed their concerns that they may not realise if the pilot readback is not aligned with the given clearance.

CRT-05.972-TLR4-TVALP- H106-2014

Success criteria	Result
ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by ASR	POK: Only 66% of the ATCOs agreed that the level of information provided by the ASR is adequate.

The pie chart below shows that 66% of the ATCOs agreed that the level of information displayed by the ASR was sufficient. Two ATCOs disagreed with this statement. On the left side the chart reveals that 83% of the ATCOs were aware with the confidence level of the recognition. Admittedly, an additional window was opened for them on the HMI and they could check the actual percentages of the recognition as a part of the ASR module – not intended as an operational HMI.

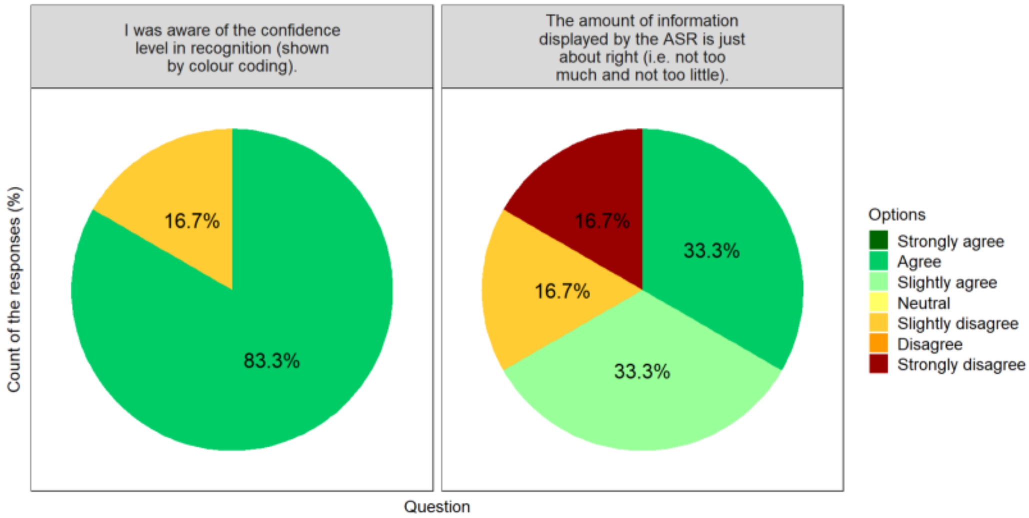


Figure C-90: Feedback on the level and quality of information provided by the system.

CRT-05.972-TLR4-TVALP- H106-2015

Success criteria	Result
Measured callsign recognition rate, command recognition rate, error rate and rejection rate of ASR system are at acceptable levels are considered within acceptable levels by the majority of ATCOS (at least 75%)	POK, as the subjective perception of callsign recognition, clearance recognition and understanding other parameters were below the cut-off line (at least 75%). However, the speed and accuracy of the callsign recognition was highly regarded.

The figure below shows user perception of successful 1) callsign, 2) clearance, 3) other parameter understanding. According to the results, the system performed best with the callsign recognition (66.7% positive feedback), and the other parameters (66.7% positive feedback), followed by the clearance recognition (55.5%).

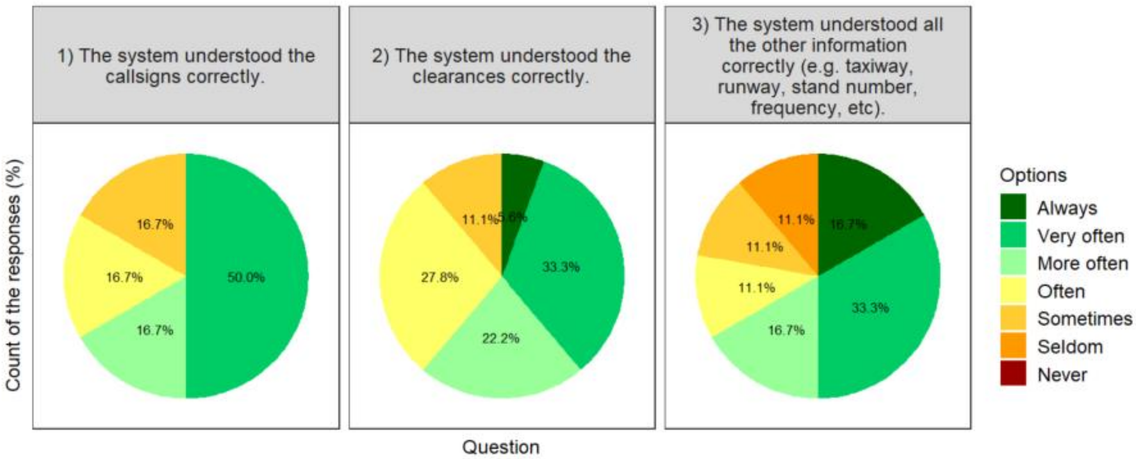


Figure C-91: Feedback on the adequacy of callsign recognition, clearance recognition and other parameters

ATCO suggestion was to use headset instead of microphone to increase fidelity. It seemed that the system recognized the instructions more robustly with the headset ("My point rankings were less than in the previous run, but I was able to deliver a constant standard. So, the system was able to recognize and correct my sentences continuously and offer mostly correct clearances.")

One of the positives ATCOs highlighted from system performance point of view was the fast and accurate callsign recognition.

CRT-05.972-TLR4-TVALP- H106-2016

Success criteria	Result
Majority of ATCOs (at least 75%) confirm adequate usability of ASR system	POK. The usability in terms of ASR performance was not optimal, and the HMI related questions did not meet the cut-off score criteria either as the ASR HMI was not designed for an operational validation. However, the design of the ASR module was overall acceptable for the ATCOs and also many improvement ideas have been gathered to further enhance the system.

The first questionnaire is a standardised one with 6 items addressing trust in ATC automation tools (EUROCONTROL SATI). Although the outcomes are more related to CRT-, some items are also relevant to utility and usability (e.g. the ASR was useful; the ASR worked accurately; the HMI was understandable). As the figure below shows, the overall feedback is far from the cut-off line defined in the success criteria. Besides the HMI being understandable, it seems like the reliability, robustness and accuracy did not meet expectations. This result is not a surprise taking into account the maturity of the solution and that the ASR module was a prototype and not designed for an operational validation.

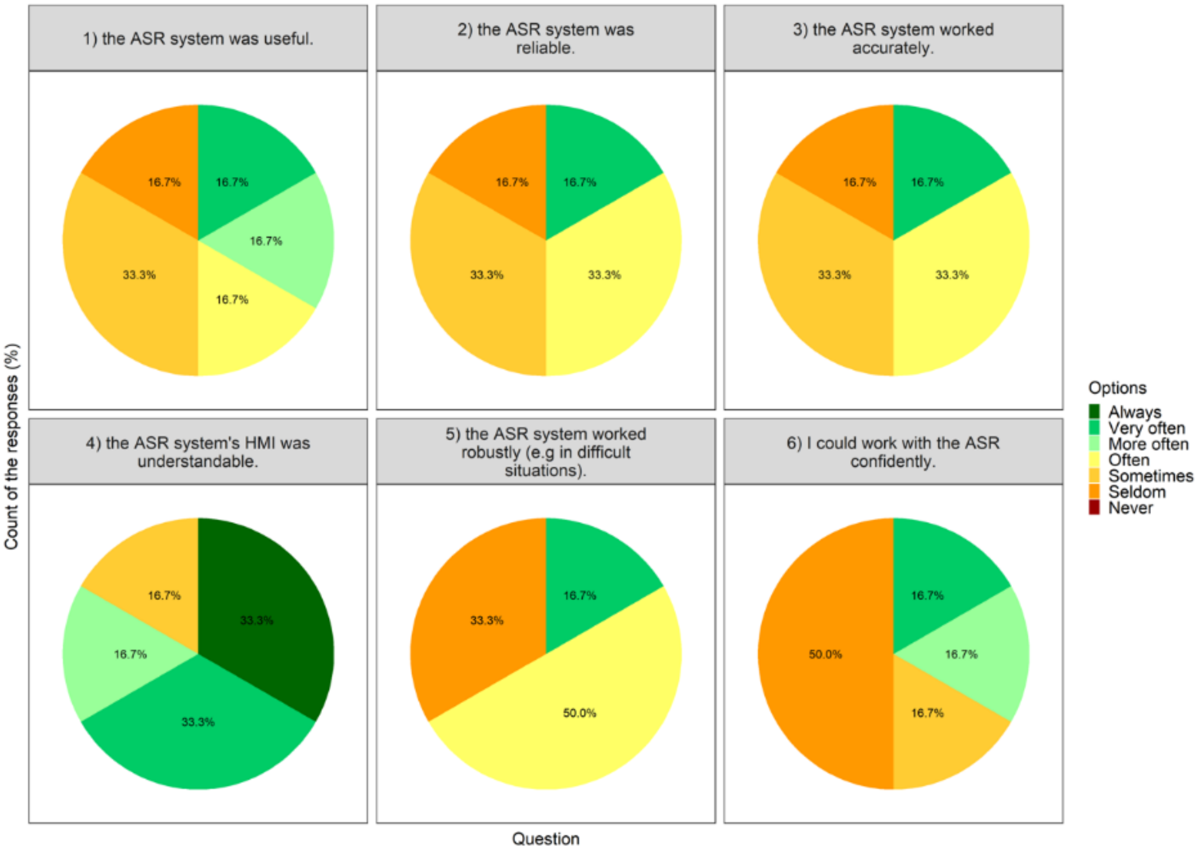


Figure C-92: Feedback on the SATI questionnaire

In terms of HMI, the feedback indicates 66.7% of the ATCOs regarded the HMI as user-friendly (e.g. click on accept/reject). Many of them liked the design of the HMI (i.e. callsign highlight) and also agree that by using ASR there would be less manual input. In fact, the callsign highlight function was one of the positive functions that has been brought up in the discussions by the end-users.

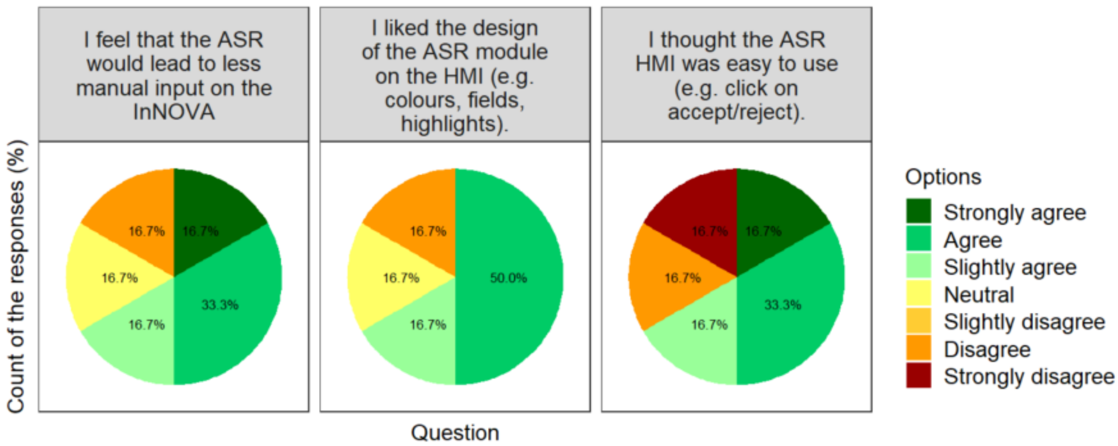


Figure C-93: Feedback on the adequacy of the HMI design

One of the weaknesses of the system was its responsiveness. The figure below shows that the latency was not appropriate- only 50% of the ATCOs gave positive feedback. ATCOs also emphasized that the system must become much faster because they do not want to continuously check the system during work. If they need to accept or reject something by clicking, they could input the command manually as well. By being this slow the system was not regarded as a real help.

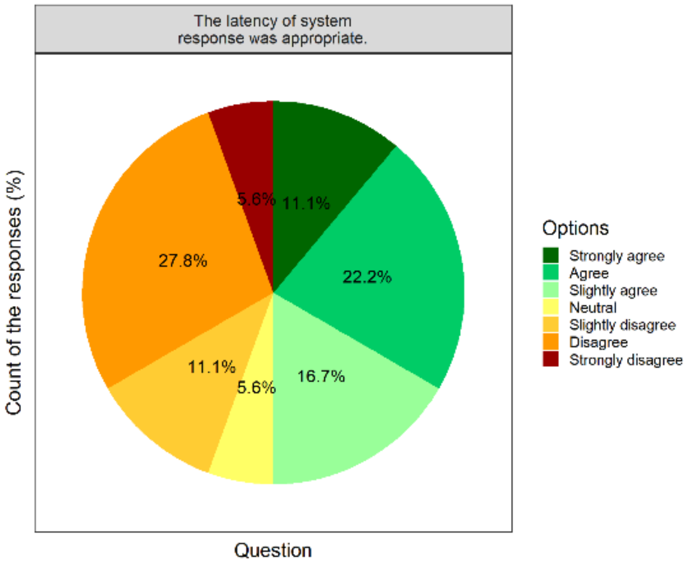


Figure C-94: Feedback on the perceived latency of the system

CRT-05.972-TLR4-TVALP- H106-2017

Success criteria	Result
Majority of ATCOs (at least 75%) provide positive feedback on acceptance of ASR tool	POK. The usability in terms of ASR performance was not optimal, and the HMI related questions did not meet the cut-off score criteria either as the ASR HMI was not designed for an operational validation. However, the design of the ASR module was overall acceptable for the ATCOs and also many improvement ideas have been gathered to further enhance the system.

Refer to feedback to success criteria CRT-05.972-TLR4-TVALP- H106-2016.

By looking at the usability questions it is evident that ATCOs do not regard the ASR tool as fully acceptable in its current state. However, judging by the maturity level of the solution this does not come as a surprise.

In addition, the system mixed up some of the letters and callsigns (WIF vs WIZ) and many of the ATCOs felt that they had to change their pronunciation to get the recognition right (see figure below). This also influence the acceptance of the system.

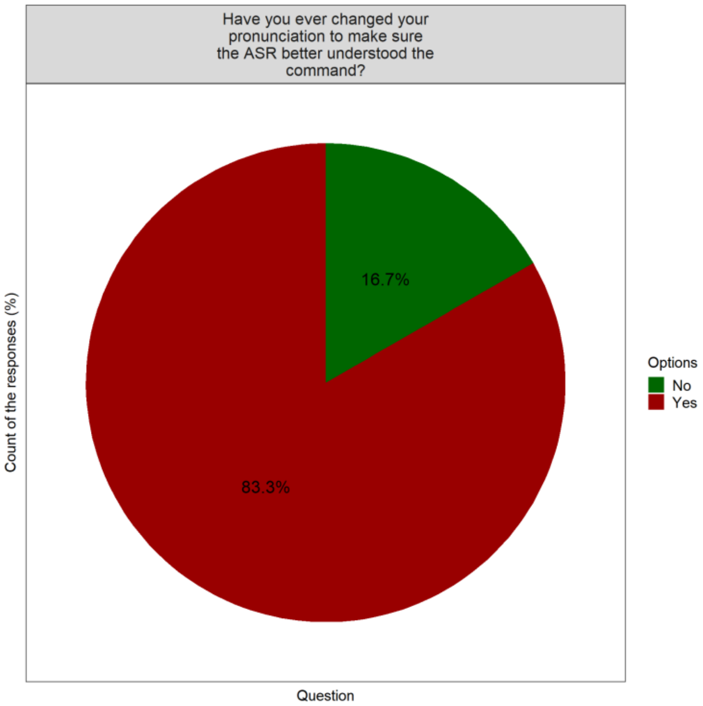


Figure C-95: Feedback on the perceived need of changing the pronunciation for successful recognition

CRT-05.972-TLR4-TVALP- H106-2018

Success criteria	Result
ATCOs (at least 75%) trust in the system is at an acceptable level	NOK. The level of trust did not meet the cut-off score criteria.

The SATI questionnaire used to address Trust in ATC automation tools has been already reported in CRT-05.972-TLR4-TVALP- H106-2016. According to the results (see figure below), ATCOs do not regard the reliability, robustness and accuracy as sufficient, but the ASR HMI was well accepted.

At this stage of maturity level, the trust would not be expected to be as high as the success criteria.

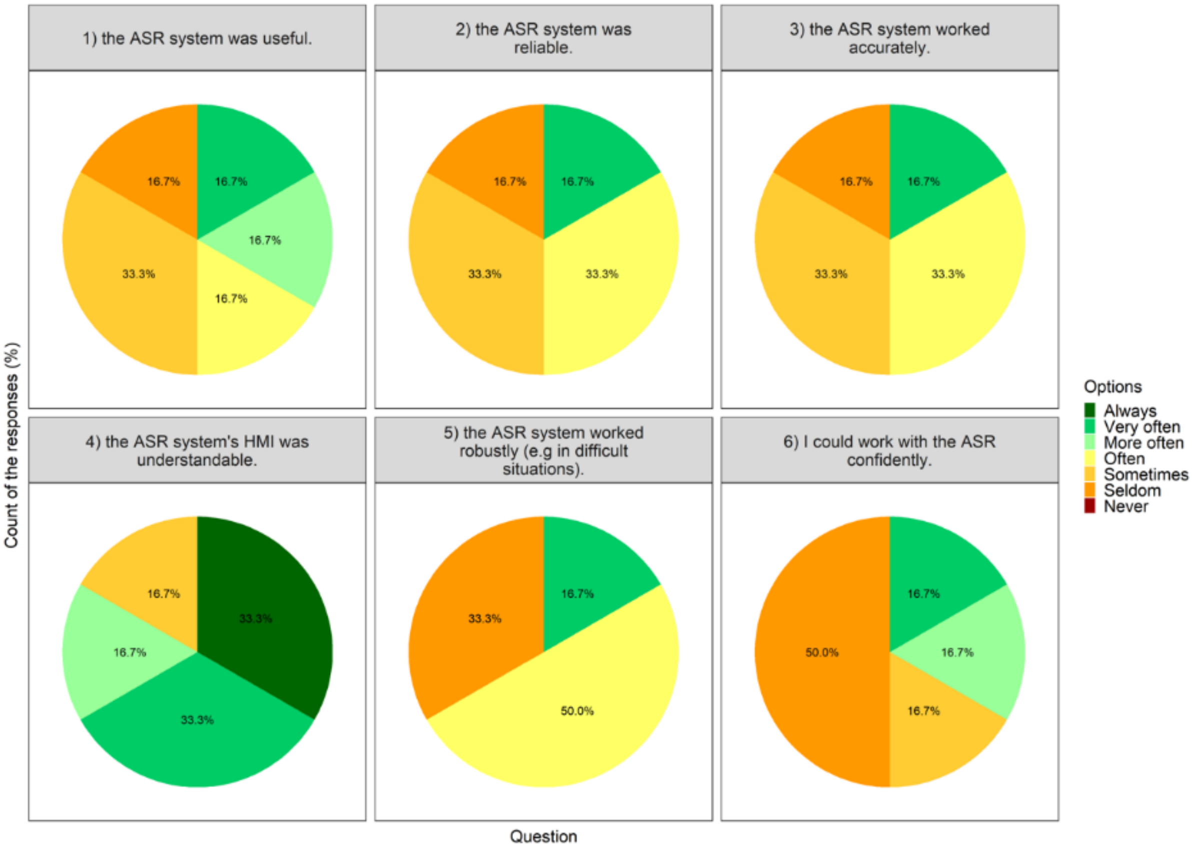


Figure C-96: Feedback on the SATI questionnaire

Interestingly, another idea came up in the debriefing session. Instead of using the speech recognition to update the strips automatically, the solution could be used to check whether the pilot provided the correct readback and notify the ATCO in case of a mismatch. The ATCOs suggested that the system should recognise if an incorrect information is being transferred to the pilot or vica versa and this "confirmation"/ "error prevention" function could have a positive impact on safety as well. This would allow the ATCOs a second chance to check their work before proceeding with the next actions.

C.7.2.4 EX4- OBJ-05.972-TRL4-TVALP-H106.2020

CRT-05.972-TLR4-TVALP- H106-2021

Success criteria	Result
Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient, and timely manner	Not applicable as it was not a fully operational validation therefore this objective could not be measured during the tests

CRT-05.972-TLR4-TVALP-H106-2022

Success criteria	Result

Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions	Not applicable as it was not a fully operational validation therefore this objective could not be measured during the tests

C.7.2.5 EX4- OBJ-05.972-TRL4-TVALP-H106.2030

CRT-05.972-TLR4-TVALP- H106-2031

Success criteria	Result
ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance	Not applicable as it was not a fully operational validation therefore this objective could not be measured during the tests

Transition factors were difficult to address at such a low maturity level, but the comments related to CRT-05.972-TLR4-TVALP- H106-2017 apply.

C.7.2.6 EX4- OBJ-05.972-TRL4-TVALP-SAFE.2010

CRT-05.972-TLR4-TVALP- SAFE -2011

Success criteria	Result
The changes related to the implementation of Automatic Speech Recognition do not increase potential for human error and therefore not reducing safety levels.	Not applicable as human error could not be measured during the tests.

Although human error could not be explicitly measured, the feedback received with regards to situational awareness applies here (ref: CRT-05.972-TLR4-TVALP-H106-2012). If we take out the only opportunity to self-check, ATCOs may not realise that the pilot readback was not aligned with the clearance (see CRT-05.972-TLR4-TVALP-H106-2013).

CRT-05.972-TLR4-TVALP- SAFE -2012

Success criteria	Result
ATCO's workload with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	Not applicable as workload could not be measured during the tests

Although workload could not be measured during the tests, ATCOs mentioned that the strength of the ASR lies in the potential to reduce workload. If the system could recognize all kinds of accents and clearances with high confidence, it could truly positively affect workload and efficient task performance (see CRT-05.972-TLR4-TVALP-H106-2011).

EUROPEAN PARTNERSHIP

Co-funded by the European Union

CRT-05.972-TLR4-TVALP- SAFE -2013

Success criteria	Result
ATCO's situational awareness with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	Not applicable as situational awareness could not be measured during the tests

Although situational awareness could not be addressed in the tests, ATCOs expressed their concerns that this solution would negatively impact their SA (see CRT-05.972-TLR4-TVALP-H106-2012). Therefore it

CRT-05.972-TLR4-TVALP- SAFE -2014

Success criteria	Result
Safety assessment activities and the results are documented and integrated in the overall solution validation results	Safety assessment activities, conducted according to SESAR SRM, are documented in TSAR.

In order to acquire objective feedback on the concept, several indicators had been identified, and data was collected during the validation sessions accordingly.

The sample contains over 900 cases collected through the different validation scenarios. It's important to remember that several of the cases are a result of practising and experimenting with the system by pronouncing differently or saying the wrong callsign intentionally. This is coming from the type of the validation, and potentially degrades the results.

- Callsign recognition rate – the proper callsign was recognized and highlighted to the ATCO in 81.16% of the cases (758/934)
- Callsign recognition error rate – improper callsign was recognized and highlighted to the ATCO in 7.82% of the cases (73/934)
- Callsign recognition rejection rate – no callsign was recognized and highlighted to the ATCO in 11.03% of the cases (103/934)
- Command recognition rate – the proper command was recognized and presented to the ATCO in 93.15% of the cases (925/993), while the callsign and the command were both correct in 76.03% of the cases (755/993)
- Command recognition error rate –improper command was recognized and presented to the ATCO in 1.81% of the cases (18/993)
- Command recognition rejection rate – no command was presented to the ATCO in 5.04% of the cases (50/993)

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Although callsign and command recognition rates are promising, it is important to highlight that further development shall concentrate on the improvement of callsign and command recognition error rates, because these indicators are the most important from safety point of view. Acceptable values shall be defined in risk assessment sessions.

Delay is 0.482 sec on average (sd = 0.3 sec) for command recognition, and 0.494 sec on average (sd = 0.302 sec) for callsign recognition. As the HMI polled/refreshed the ASR HMI once a second, the delay is biased. Performance should be improved by increasing the polling/refresh rate.

C.7.2.7 EX4- OBJ-05.972-TRL4-TVALP- PERF.2010

CRT-05.972-TLR4-TVALP- PERF -2011

Success criteria	Result
Laboratory tests show that the SESAR technological solution improves Cost Efficiency performance by reducing cost per flight (PER.TRL4.3) (through e.g. reduction of workload, reduction of delay times,….)	Not applicable. The exercise was not designed as an operational use of ASR and to measure workload.

Not possible through laboratory test to show any influence on Cost Efficiency. Workload could not be measured during the tests.

A properly implemented ASR could lead to reduction of workload but this was a technical validation, and the maturity did not reach the necessary level to measure Cost Efficiency.

C.7.3 Unexpected behaviours/results

C.7.4 Confidence in results of EXE-004

C.7.4.1 Level of significance/limitations of Technological Validation Exercise Results

The validation of ASR was added to a validation of remote tower operation. All instructions/clearances transmitted on radio to flight crews were translated to text structured to be used as system input and presented to the ATCO in a dedicated new window on the ATC system HMI. This gave a good indication of the ASR ability to convert speech into structured text, and the feasibility of ASR to be used to automate inputs in an ATC system. At the same time this added tasks to the controllers in the exercise, to monitor the result provided by the ASR module. Even though the exercise was set up with additional tasks for the ATCO to check the technical performance and feasibility of ASR to be used to automate ATC system inputs, it was assumed that ASR would have the potential to reduce tasks/system input and workload if implemented to automate ATCO HMI inputs in the ATC system.

The validation exercise was neither a pure operational validation nor a pure technical validation. The validation exercise used ATCOs for collecting data, but still not using something that could be regarded as an operational ASR based input of clearances/instructions into an TWR ATC system. The validation exercise was more a hybrid between the two validation types. Expectations, and validation objectives included success criteria that could be suitable for an operational validation, but the present validation platform was created more for validating the technical feasibility of using ASR to automate inputs in a TWR ATC system.

C.7.4.2 Quality of EXE-004 results

The quality of the results in the area of recognizing commands using ARS technology was good. This was also the focus of the exercise and validation platform set-up.

The quality of the results relating to using the recognized commands to automate ATC system HMI input was not that good, as the focus was more on trying to automate as many inputs as possible. The validation platform setup was also missing some inputs that often is performed for every flight, but those were not in the predefined list of clearances and did not have any impact on strip status (i.e. “backtrack” or “joint traffic circuit” were frequently used but those instructions did not create an event to update the flight strip)

The measurements of recognizing callsigns and clearances and data collection were performed during all the simulation runs.

During the runs, apart from the predefined scenario, there were also practice sessions and other experimental activities with the system, free tests, and on occasion some end-users deliberately pronounced callsigns incorrectly, to see how the system responds. This definitely influenced the quality of the result. To perform a pure laboratory test with scripted callsigns and clearances could show a result with better quality of exercise result.

C.7.4.3 Significance of EXE-004 results

As this technical validation was executed in a quite operational environment and closely related to an operational validation of remote tower operations, it gave an impression of an “unfinished” functionality, compared to other functionality validated in the remote tower validation. The results however demonstrate the feasibility of automating ATC system HMI input based on ASR technology. Validating ASR in this hybrid environment also highlights that the set of inputs required by the ATC system shall be explicitly defined, as there will be instructions that will not affect the automatic EFS update process but could have an impact on user acceptance.

The validation exercise was based on assumption that it is the spoken instructions and clearances by the ATCO that can be recognized and used as automated system input. There could also be ASR of the requests, readback and other utterance from the flight crew transmitted by radio.

C.8 Conclusions

C.8.1 Conclusions on technological feasibility

The exercise demonstrated that it is feasible to use ASR technology to capture Aerodrome ATC instructions and clearances transmitted by radio to flight crews. It also demonstrated that it is feasible to use captured instructions and clearances to automate ATC system inputs. ATC systems tend to require more and more inputs by the ATCO. Inputs are typically performed for every flight and can be simple inputs where no data is input, only pushing/clicking a button. ASR technology could also allow for more ATC system inputs that would otherwise be regarded as not being justifiable due to the extra ATCO workload to perform the system input. So, from a technological viewpoint it should also be recognized that ASR technology could open up for more system inputs (not only automate current inputs) that create improvement in areas like efficiency, safety and other areas.

C.8.2 Conclusions on (human) performance assessments

In general, most of ATCOs saw the potential in applying speech recognition in a TWR environment. However, the ASR system was not at such level of maturity to gather very positive feedback in its current state. There were functions that were highly regarded, e.g. callsign highlight, which could support situational awareness. However, there were a number of issues that have to be corrected or further improved, e.g. the size of the vocabulary and the system latency to recognise a variety of commands. In addition, some concerns regarding situational awareness were brought up in debriefing sessions.

As an alternative use for voice recognition, it came up that instead of automatically updating the EFS, the system could be used to check whether the pilot provided the correct readback and notify the ATCO in case of a mismatch. Such an “error prevention” functionality could have a positive impact on safety and overall end-user acceptance.

C.8.3 Recommendations

Some of the system-related recommendations were the followings:

- Extend the vocabulary by considering the benefit to automate an input
- Take voice samples from the whole (Hungarian) TWR ATCO staff during simulations to optimise ASR performance
- Improve the use of Machine Learning to provide better interpretation of ASR
- Implement ASR for pilot side to highlight callsign when a pilot is calling in or, instead of automatically updating the EFS, the system could function in a more “preventative” manner and check whether the pilot provided the correct readback and notify the ATCO in case of a mismatch.
- Use headset instead of microphone. The system will recognize the instructions more robustly.
- Strip highlight in compact mode should be more visible (*note: it has been improved between the two group’s simulation, and got much better for Group 2*)

Appendix D Technological Validation Exercise 005 Report

D.1 Summary of EXE-005 plan

The plan for the validation of Sol 97.1 within exercise EXE-VAR-005 (in the following EXE-005) can be found in section 5.5 of the SESAR 2020 - PJ05-W2 Sol 97 D3.1.033 - Technical Validation Plan (TVALP) Final version [28].

D.2 EXE-005 description and scope

The exercise investigates the use of augmented reality applications in a conventional tower environment. The selected environment is the Spanish airport Vitoria (LEVT). The airport is classified as "other" " (<15000 mov/yr) according to EATMA operational environment [1], but it could also be classified a small airport as it is an IFR airport dedicated mainly to cargo and the number of European and intercontinental flights are high.

Augmented Reality in ATC Tower environment supports the Air Traffic Controller by blending real world images with computer-generated data (virtual reality) in real-time, so that visual information can be enhanced to improve identification and tracking of a/c (or vehicles) on the airport surface.

A head mounted display is used to present information via augmented reality to the controller. The information displayed are labels with a/c identification associated to the position of the a/c, location of the runway and taxiway, and location of key surface reference buildings/objects.

The technology supports the controller in good visibility conditions by decreasing the head-down time or reduced visibility by physical obstacles, but it is expected to be especially useful under low visibility conditions (LVC) or in light shortage such as sunrise, sunset or at night.

The information displayed is configurable and some of the elements are expected to be displayed only in LVC. Air gestures will be used to adapt the information displayed to the user preferences. Being the target airport of this exercise in the other/small category, [1][1] radio and paper flight strips are the main support tools for the controller and no electronic interaction related to a/c control (e.g. push-back authorization) will be performed thorough air gestures.

The roles involved are the tower ground controller and the runway controller but due to the dimensions of the airport, one actor performs both roles.

The airport does not have surface surveillance radar but it is expected that an ADS-B antenna is installed during 2022.

Next use cases described in the Technical Specification [27] have been validated in the exercise:

- UC-97-TRL4-TS-103: Retrieve of information by means of V/AR. *Use V/AR (including tracking labels) to retrieve all relevant positioning, identification, flight status, weather information needed for the specific RWY and GND tasks in the specific condition.*
- UC-97-TRL4-TS-104: Tracking labels in Augmented Reality for landing/departing a/cs. *Use tracking labels in AR device so that a clear distinction between different a/c lined up for landing can be made.*

EXE-005 addresses the following KPA:

- Human Performance

- Safety
- Resilience

V/AR in a control tower environment will support the achievement of the following top-level performance expected benefits:

- Increased safety
- Reduction of controller workload;
- Increased situational awareness, the CWP is simpler and all information is integrated.
- Increased ATCO efficiency;
- Improved HMI and usability and performance of interactions
- Increased resilience (LVC)

D.3 Summary of EXE-005 objectives and success criteria

The objectives and validation success criteria are covered by Exercise 005 are listed in the table below. They correspond to the V/A-R and air gestures technology.

SESAR Solution Validation Objective	SESAR Solution Success Criteria	Coverage and comments on the coverage of SESAR Solution Validation Objective in EXE-005	Exercise Validation Objective	Exercise Success Criteria
OBJ-05.971-TRL4-TVALP-FEAS.1010 To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS.	CRT-05.971-TLR4-TVALP-FEAS-1011 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Virtual or Augmented Reality and tracking labels.	Fully	EX5-OBJ-05.971-TRL4-TVALP-FEAS.1010 To confirm the concept is operationally feasible when addressing the Use Case for Virtual or Augmented Reality, tracking labels, and Air Gestures	EX5-CRT-05.971-TLR4-TVALP-FEAS-1011 Identical to CRT-05.971-TLR4-TVALP-FEAS-1011
	CRT-05.971-TLR4-TVALP-FEAS-1012 No operational show-stoppers have been identified during	Fully		EX5-CRT-05.971-TLR4-TVALP-FEAS-1012 Identical to CRT-05.971-TLR4-TVALP-FEAS-1012

Co-funded by the European Union

	laboratory tests (based on a prototype) related to the use of Air Gestures.			
	CRT-05.971-TLR4-TVALP-FEAS-1013 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Attention Guidance.	N/A		N/A
OBJ-05.971-TRL4-TVALP-FEAS.1020 To identify possible technical feasibility issues and possible show stoppers.	CRT-05.971-TLR4-TVALP-FEAS-1021 Laboratory tests (based on a prototype) have verified the technical feasibility of the use of V/A-R applications in the tower environment.	Fully	EX5-OBJ-05.971-TRL4-TVALP-FEAS.1020 Identical to OBJ-05.971-TRL4-TVALP-FEAS.1020.	EX5-CRT-05.971-TLR4-TVALP-FEAS-1021 Identical to CRT-05.971-TLR4-TVALP-FEAS-1021.
	CRT-05.971-TLR4-TVALP-FEAS -1022 Laboratory tests have verified that the integration of the V/A-R applications with other related system enablers is technically feasible.			EX5-CRT-05.971-TLR4-TVALP- FEAS-1022 Identical to CRT-05.971-TLR4-TVALP-FEAS-1022.
OBJ-05.971-TRL4-TVALP-H103.1010 To assess that the technical systems for V/A-R Tracking Labels and overlays support the ATCOs in performing their tasks.	CRT-05.971-TLR4-TVALP-H103.1011 Majority of ATCOs (at least 75%) responses is that V/A-R supports ATCO in maintaining workload at acceptable level.	Fully	EX5-OBJ-05.971-TRL4-TVALP-H103.1010 Identical to OBJ-05.971-TRL4-TVALP-H103.1010	EX5-CRT-05.971-TLR4-TVALP-H103.1011 Identical to CRT-05.971-TLR4-TVALP-H103.1011
	CRT-05.971-TLR4-TVALP-H103.1012 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R			EX5-CRT-05.971-TLR4-TVALP-H103.1012 Identical to CRT-05.971-TLR4-TVALP-H103.1012

	CRT-05.971-TLR4-TVALP-H103.1013 Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO in maintaining an adequate level of situation awareness.			EX5-CRT-05.971-TLR4-TVALP-H103.1013 Identical to CRT-05.971-TLR4-TVALP-H103.1013
	CRT-05.971-TLR4-TVALP-H103.1014 Measured time spent in head up is increased in the solution scenario with respect to the reference scenario.			EX5-CRT-05.971-TLR4-TVALP-H103.1014 Identical to CRT-05.971-TLR4-TVALP-H103.1014
	CRT-05.971-TLR4-TVALP-H103.1015 HMI of V/A-R tools does not overshadow the relevant information on the OTW view.			EX5-CRT-05.971-TLR4-TVALP-H103.1015 Identical to CRT-05.971-TLR4-TVALP-H103.1015
	CRT-05.971-TLR4-TVALP-H103.1016 V/A-R HMI does not increase the potential for human error.			EX5-CRT-05.971-TLR4-TVALP-H103.1016 Identical to CRT-05.971-TLR4-TVALP-H103.1016
	CRT-05.971-TLR4-TVALP-H103.1017 ATCOs' (at least 75%) trust in the system is at an acceptable level.			EX5-CRT-05.971-TLR4-TVALP-H103.1017 Identical to CRT-05.971-TLR4-TVALP-H103.1017
	CRT-05.971-TLR4-TVALP-H103.1018 Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R HMI.			EX5-CRT-05.971-TLR4-TVALP-H103.1018 Identical to. CRT-05.971-TLR4-TVALP-H103.1018
	CRT-05.971-TLR4-TVALP-H103.1019			N/A

	Majority of ATCOs (at least 75%) responses is that alarms and alerts are not too intrusive and support ATCOs in the early detection of ATC critical situations.			
	CRT-05.971-TLR4-TVALP-H103.1020 ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R tool.			EX5-CRT-05.971-TLR4-TVALP-H103.1020 Identical to CRT-05.971-TLR4-TVALP-H103.1020
	CRT-05.971-TLR4-TVALP-H103.1021 Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO team (GND and TWR) in maintaining an acceptable level of situation awareness.			EX5-CRT-05.971-TLR4-TVALP-H103.1021 N/A
OBJ-05.971A-TLR4-TVALP- H103.1030 To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Tracking labels and overlays	CRT-05.971-TLR4-TVALP-H103.1031 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner.	Fully	EX5-OBJ-05.971-TRL4-TVALP-H103.1030 Identical to OBJ-05.971A-TLR4-TVALP-H103.1030	EX5-CRT-05.971-TLR4-TVALP- H103.1031 Identical to CRT-05.971-TLR4-TVALP-H103.1031
	CRT-05.971-TLR4-TVALP-H103.1032 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions.			EX5-CRT-05.971-TLR4-TVALP-H103.1032 Identical to CRT-05.971-TLR4-TVALP-H103.1032

OBJ-05.971-TRL4-TVALP- H103.1040 To assess job acceptance and satisfaction with the introduction of V/A-R tracking labels and overlays	CRT-05.971-TLR4-TVALP-H103.1041 ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance.	Fully	EX5-OBJ-05.971-TRL4-TVALP-H103.1040 Identical to OBJ-05.971-TRL4-TVALP-H103.1040	EX5-CRT-05.971-TLR4-TVALP-H103.1041 Identical to CRT-05.971-TLR4-TVALP-H103.1041
OBJ-05.971-TRL4-TVALP-H104.1010 To assess that the technical systems for V/A-R Air Gestures support the ATCOs in performing their tasks.	CRT-05.971-TLR4-TVALP-H104.1011 Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures support ATCO in maintaining workload at acceptable level.	Fully	EX5-OBJ-05.971-TRL4-TVALP-H104.1010 To assess that the technical systems for V/A-R Air Gestures support the ATCOs.	EX5-CRT-05.971-TLR4-TVALP-H104.1011 Identical to CRT-05.971-TLR4-TVALP-H104.1011
	CRT-05.971-TLR4-TVALP-H104.1012 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R Air Gestures.			EX5-CRT-05.971-TLR4-TVALP-H104.1012 Identical to CRT-05.971-TLR4-TVALP-H104.1012
	CRT-05.971-TLR4-TVALP-H104.1013 Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures HMI supports ATCO in maintaining an adequate level of situation awareness.			EX5-CRT-05.971-TLR4-TVALP-H104.1013 Identical to CRT-05.971-TLR4-TVALP-H104.1013
	CRT-05.971-TLR4-TVALP-H104.1014 Measured time spent in head up is increased in the solution scenario with respect to the reference scenario.			EX5-CRT-05.971-TLR4-TVALP-H104.1014 Identical to CRT-05.971-TLR4-TVALP-H104.1014

EUROPEAN PARTNERSHIP

Co-funded by the European Union

	CRT-05.971-TLR4-TVALP-H104.1015 V/A-R Air Gestures HMI does not increase the potential for human error.			EX5-CRT-05.971-TLR4-TVALP-H104.1015 Identical to CRT-05.971-TLR4-TVALP-H104.1015
	CRT-05.971-TLR4-TVALP-H104.1016 ATCOs' (at least 75%) trust in the system is at an acceptable level.			EX5-CRT-05.971-TLR4-TVALP-H104.1016 Identical to CRT-05.971-TLR4-TVALP-H104.1016
	CRT-05.971-TLR4-TVALP-H104.1017 Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R Air Gestures HMI.			EX5-CRT-05.971-TLR4-TVALP-H104.1017 Identical to. CRT-05.971-TLR4-TVALP-H104.1017
	CRT-05.971-TLR4-TVALP-H104.1018 ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R Air Gestures tool.			EX5-CRT-05.971-TLR4-TVALP-H104.1018 Identical to CRT-05.971-TLR4-TVALP-H104.1018
OBJ-05.971A-TLR4-TVALP- H104.1020 To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Air Gestures	CRT-05.971-TLR4-TVALP-H104.1021 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner.	Fully	EX5-OBJ-05.971-TRL4-TVALP-H104.1020 Identical to OBJ-05.971A-TLR4-TVALP-H104.1020	EX5-CRT-05.971-TLR4-TVALP- H104.1021 Identical to CRT-05.971-TLR4-TVALP-H104.1021
	CRT-05.971-TLR4-TVALP-H104.1022 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions.			EX5-CRT-05.971-TLR4-TVALP-H104.1022 Identical to CRT-05.971-TLR4-TVALP-H104.1022

OBJ-05.971-TRL4-TVALP- H104.1030 To assess job acceptance and satisfaction with the introduction of V/A-R Air Gestures	CRT-05.971-TLR4-TVALP-H104.1031 ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance.	Fully	EX5-OBJ-05.971-TRL4-TVALP-H104.1030 Identical to OBJ-05.971-TRL4-TVALP-H104.1030	EX5-CRT-05.971-TLR4-TVALP-H104.1031 Identical to CRT-05.971-TLR4-TVALP-H104.1031 .
OBJ-05.971-TLR4-TVALP-SAFE.1010 To assess the impact of Virtual/Augmented Reality applications on safety.	CRT-05.971-TLR4-TVALP-SAFE-1011 Laboratory tests show that the Virtual/Augmented Reality applications improve the safety performance by reducing human error.	Fully	EX5-OBJ-05.971-TLR4-TVALP-SAFE.1010 Identical to OBJ-05.971-TRL4-TVALP-SAFE.1010	EX5-CRT-05.971-TLR4-TVALP-SAFE-1011 Identical to CRT-05.971-TLR4-TVALP-SAFE-1011 .
	CRT-05.971-TLR4-TVALP-SAFE-1012 Laboratory tests show that the Virtual/Augmented Reality applications improve the safety performance by reducing ATCO workload.			EX5-CRT-05.971-TLR4-TVALP-SAFE-1012 Identical to CRT-05.971-TLR4-TVALP-SAFE-1012
	CRT-05.971-TLR4-TVALP-SAFE-1013 Laboratory tests show that the use of Virtual/Augmented Reality applications improves the safety performance by increasing situational awareness.			EX5-CRT-05.971-TLR4-TVALP-SAFE-1013 Identical to CRT-05.971-TLR4-TVALP-SAFE-1013
	CRT-05.971-TLR4-TVALP-SAFE-1014 Safety assessment activities and the results are documented and			EX5-CRT-05.971-TLR4-TVALP-SAFE-1014 Identical to CRT-05.971-TLR4-TVALP-SAFE-1014.

	integrated in the overall solution validation results.			
OBJ-05.971-TLR4-TVALP-PERF.1010 To assess the performance benefits of equivalent visual operations for tower control through the use of applications for Virtual/Augmented Reality.	CRT-05.971-TLR4-TVALP-PERF-1011 Laboratory tests show that the use of V/A applications improves Cost Efficiency performance by reducing the cost per flight (through e.g. reduction of workload, reduction of delay times).	Fully	EX5-OBJ-05.971-TLR4-TVALP-PERF.1010	EX5-CRT-05.971-TLR4-TVALP-PERF-1011 Identical to CRT-05.971-TLR4-TVALP-PERF-1011
	CRT-05.971-TLR4-TVALP-PERF-1012 Laboratory tests show that the use of V/A reality applications improves Resilience by increasing situational awareness in low visibility conditions while maintaining workload within acceptable limits.			EX5-CRT-05.971-TLR4-TVALP-PERF-1012 Identical to CRT-05.971-TLR4-TVALP-PERF-1012

Table D-18: Validation Objectives addressed in EXE-005

D.4 Summary of EXE-005 validation scenarios

D.4.1 Reference scenario(s)

The reference scenario is current operations at the Vitoria ATC control room.

Vitoria airport (LEVT) is located at the north of Spain. According to EATMA operational environments [1], Vitoria is classified as “other” ” (<15000 mov/yr) size airport, but it could also be classified a small airport as it is an IFR airport dedicated mainly to cargo and the number of European and intercontinental flights are high. In 2019 it managed 174K passengers, 10,8K movements, and 64,5 M cargo. Most of the movements are related to cargo, although it also has a passenger terminal. The airport has one H-24 runway. The airport layout is shown in the figure below. The runway 04/22 is Category II/III.

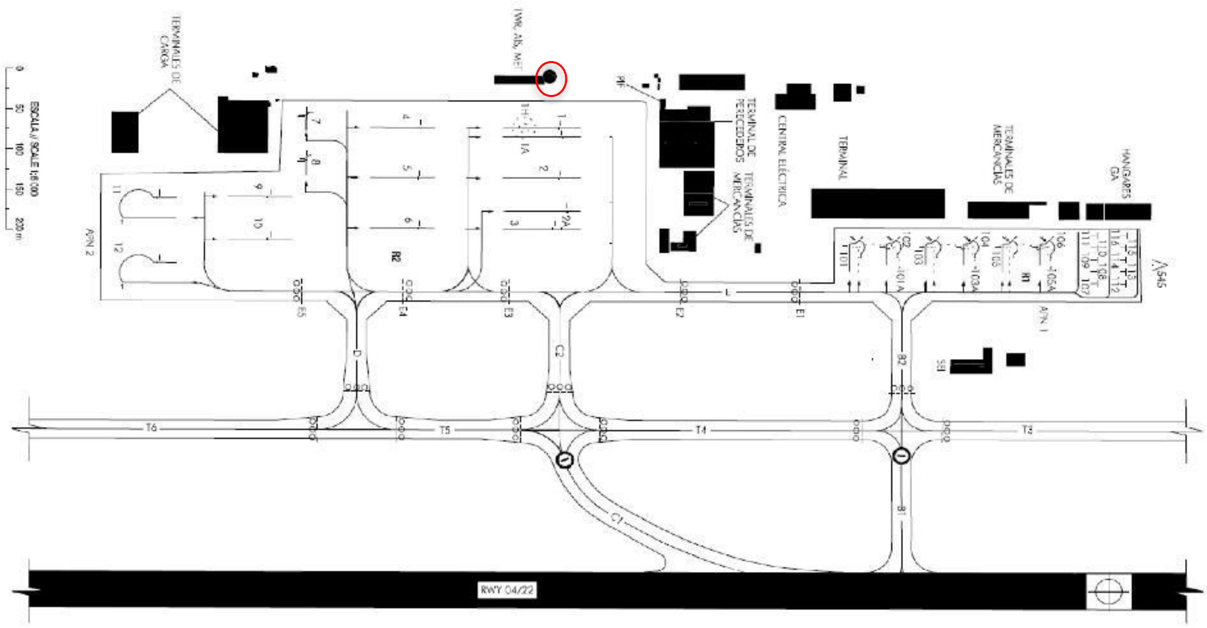


Figure D-97: Vitoria airport layout

D.4.1.1 Main airport infrastructure

Vitoria airport has an ATC TWR that is located near the main parking area, the red circle in Figure D-97~~Figure D-97~~. Vitoria airport has a SMGCS that includes:

- Stops bars and antirunway incursion bars;
- Centre taxiway lights.
- The airport does not have surface radar.

D.4.1.2 Main ATC TWR characteristics

The airport provides landing, departure, and approach service from the ATC tower. Controllers in the tower also provides approach control service for landing and departures to/from Burgos airport. The operative configuration is a monoposition (one controller performs the role of ground and runway controller). The working shifts are of 8 hours during the day and 4 hours by night.

The airspace volumes under the controller responsibility are the LEVT CTA, two CTA from Burgos and the ATZ and CTR from LEVT. Burgos airport has and AFIS (Aerodrome Flight Information Service) during the public attendance schedule and no ATS service during the restricted schedule.

The landing and departure service is provided only at Vitoria airport where there is barely radar coverture in the ATZ (limited by the higher layer) nor in the CTR. In the CTA under responsibility, the coverture is complete and conventional approach service is provided.

The ATC Tower has an approach SACTA CWP with paper strips.

D.4.1.3 Low Visibility Procedures

The airport has three different low visibility procedures:

- LVTO in runway 04
- LVTO in runway 22
- CAT II in runway 04

There are four phases in the LVP: monitoring when visibility starts to deteriorate, stand-by where vehicles movements are restricted, implementation, where vehicles and a/c movements are restricted, and cancellation when the visibility conditions improve.

Monitoring

Weather conditions are closely monitored and close coordination between the aeronautical meteorological service and the tower is established. All the necessary units are informed on the LVP initiation and phase.

Stand-by

It is a transition phase where the impacted services and users will prepare the means and perform the necessary tasks to ensure the procedures can be applied if necessary. Once everything is in place, they will stand by until the implementation or cancellation phases are activated.

Vehicles movements are restricted in the manoeuvre area. TWR will contact vehicles on the taxiway and platform via radio frequency with the following phrase:

“Low visibility procedures in stand-by phase. Clear manoeuvring area”

Implementation

Once the meteorological LVC starts, all the services and users will proceed accordingly with their responsibilities and tasks. Flight crew will be informed about the LVP implementation via radio.

Taxiway TA6 will be closed.

Vehicles in the manoeuvring area will only be allowed through taxiways leading to the runway guided by an A/c Rescue and Fire Fighting, ARFF, vehicle. They will only use authorized service roads. Only on vehicle from the ARFF is authorized to provide the guide service. If more than one a/c/vehicle needs guidance, it will have to wait until the ARFF vehicle is free.

Arriving a/c:

- RVR < 550m: Mandatory guidance from platform entry until the parking stand;
- RVR < 550m: Under demand from the taxiway to the parking stand;
- RVR < 185m and centre taxiway lights are not available: Mandatory guidance from the taxiway until the stand.

Departing a/c:

- RVR < 550m: Mandatory guidance from parking stand until intermediate waiting point;
- RVR < 550m: Under demand from parking stand to runway waiting point;

Co-funded by the European Union

- RVR < 185m and centre taxiway lights are not available: Mandatory guidance from the stand to runway waiting point.

If the meteorological conditions improve it is possible to change to Stand-by phase or to the cancellation phase depending on the weather forecast.

In the implementation phase, the movement area is divided in two blocks:

Vitoria LVP indicate that the movement area is divided in two blocks as presented in Figure D-98~~Figure D-98~~

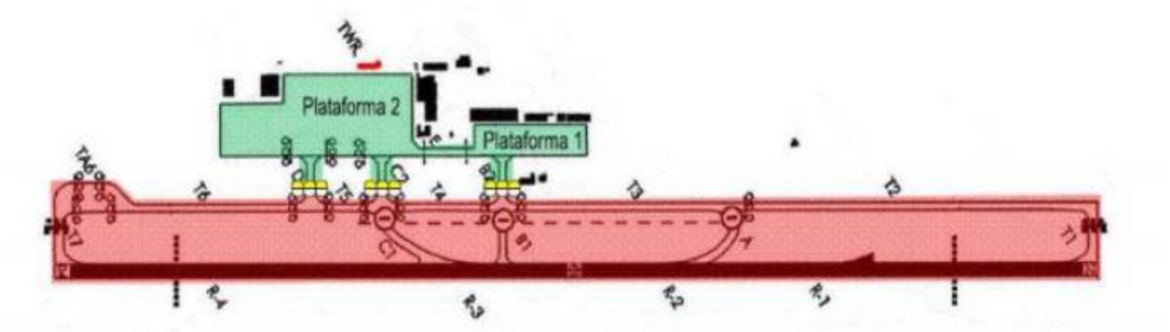


Figure D-98: LVP Blocks at Vitoria airport

The possibility to establish different blocks at Vitoria airport depending on the visibility conditions is currently under study. The most restricted phase would be when the Runway Visual Range, RVR, is below 185m, where the movement area would be again two blocks as now a days. In the table and figure below the approach under study is presented.

PHASE	OPERABILITY	RESTRICCIÓN	RVR
FASE III - IMPLEMENTATION	RESTRICTED	• TWY TA6 is closed • Vehicles in manoeuvre area are guided by SSEI.	1.300m ≥ RVR ≥ 550m
	BLOCK RESTRICTED	• Block operation III.a	• RVR<550m Block operation without free intermediate block
		• Block operation III.b	• RVR<400m Block operation with free intermediate block
		• Block operation III.c	• RVR<185m manoeuvre area is one block

Table D-19: LVP at Vitoria airport - summary

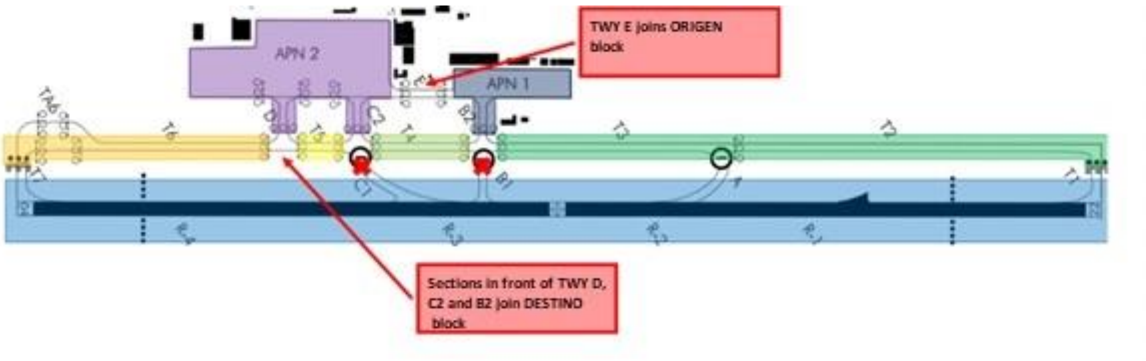


Figure D-99: LVP Blocks at Vitoria airport - 2

Cancellation

LVP are deactivated and the airport resumes normal operation.

D.4.2 Solution Scenario(s)

The solution scenario is the same as the reference scenario but controllers have a V/A R device – Hololens - available - to support them in their operations. The V/A R device is fed with processed information from a/c's ADS-B and the airport layout.

It is expected that the landing and departure services benefits from the augmented reality support service fed by an ADS-B signal due to the following characteristics of the airport:

- Low radar coverture;
- Bad weather with high number of fog banks;
- High number of operations in light shortage: sunrise, sunset and night.
- The V/A R device supports them in the next use cases:

By providing information about a/c on the manoeuvre area. A label is linked to each a/c on the area indicating its callsign. The label moves with the real a/c. Further information such as horizontal velocity and altitude is also displayed.

By providing information on the vehicle on the manoeuvre area. A label with the identification of the vehicle is displayed to the controllers through the device.

By providing information on the landing /departing a/c. A label with the a/c callsign, horizontal speed, and altitude is displayed to the controller. By proving the main elements of the airport layout to be displayed in LVC. Figure D-100~~Figure D-100~~ presents the view from Vitoria TWR of the cargo platform with the virtual labels displayed.

Figure D-100: V/A-R label on Vitoria airport

The same V/A R device is also able to detect and identify different types of air gestures. The controller is able to interact with different virtual elements in this exercise through air gestures. The interaction is not linked to operational authorisations.

The air gestures allows the controller to:

- Calibrate the device;
- Pick, drag and drop AR elements;
- Navigate between a list of a/c displayed on the V/A R;
- Increase/decrease the number of a/c displayed. Filter by altitude and/or distance;
- Enable/disable LVC display.

Figure D-101~~Figure D-101~~ presents some of the interactive menus.

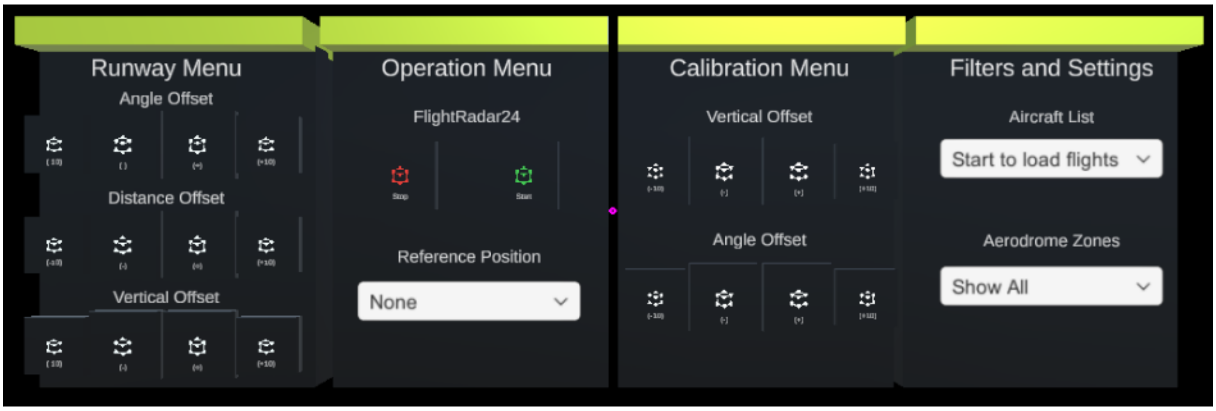


Figure D-101: Interactive menu

Figure D-102~~Figure D-102~~ below presents the LCV block over the runway to indicate if it was occupied (in red) or not (light blue). It should be noted that the colours in the photo are more solid than the ones that were presented to controllers on the V/A R device. The occupancy or not of the runway (R) was also displayed on a square that was always in the controller line of vision. The squares regarding taxiway (T) and platform (P) were displayed for demonstration purpose but the LVP software was not developed.

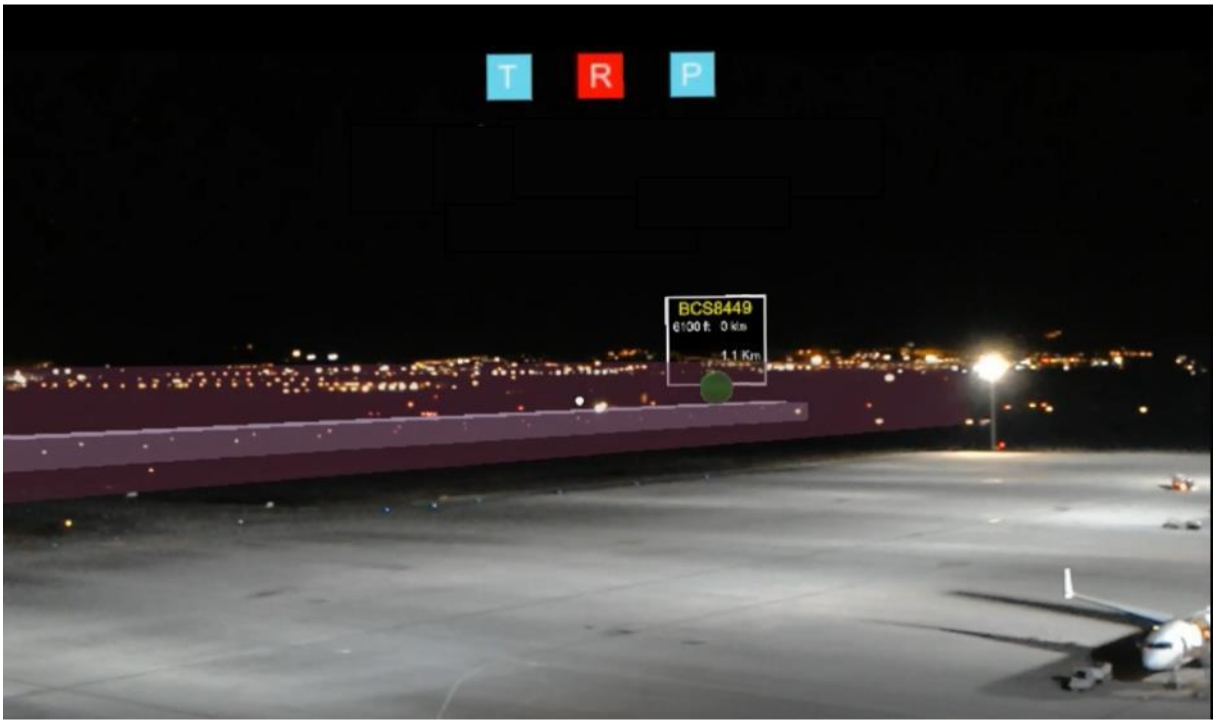


Figure D-102: LVP runway block

The exercise performed was a shadow mode in the TWR ATC break area located on floor below the control room. The simulation runs were spread along several days and tuned to represent different illumination environments:

- By day;
- By night;
- During sunset;

- Although initially planned, there were no LVC during the simulation days.

Different exercises were performed with a vehicle enabled with an ADS-B transmitter. These exercises took into account the several blocks configuration, the restrictions, and RVR, identified. For testing and improvement of the technical solution, an ADS-B-like Mobile software has been developed.

D.5 Summary of EXE-005 assumptions

Identifier	Title	Description	Justification	Impact on Assessment
AS-EXE.005-01	Monoposition	Simulation will be focused on the work of one tower controller with the roles of runway and ground controller	Due to traffic load the monoposition is used in some conditions at Vitoria airports	Medium
AS-EXE.005-02	Responsibilities	Roles and responsibilities do not change due to the use of V/A- R device	The prototype support controllers on their task but does not change the responsibilities	Low
AS-EXE.005-03	ADS-B	All a/c and vehicles under analysis will use ADS-B	The prototype analyses the ADS-B signal. Vehicles and a/c without ADS-B and are out of scope of the exercise	Low

Table D-20: EXE-005 Assumptions

D.6 Deviation from the planned activities

There were no deviations from the Plan.

D.7 EXE-005 validation results

This section presents the final status of the objectives and success criteria. It starts with a summary table, and afterwards provides the individual details.

D.7.1 Summary of EXE-005 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
EX5-OBJ-05.971-TRL4-TVALP-FEAS.1010	To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS.	EX5-CRT-05.971-TLR4-TVALP-FEAS-1011	No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Virtual or Augmented Reality and tracking labels.	OK	Successfully Validated
		EX5-CRT-05.971-TLR4-TVALP-FEAS-1012	No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Air Gestures.	OK	

EX5-OBJ-05.971-TRL4-TVALP- FEAS.1020	To identify possible technical feasibility issues and possible show stoppers.	EX5-CRT-05.971-TLR4-TVALP-FEAS-1021	Laboratory tests (based on a prototype) have verified the technical feasibility of the use of V/A-R applications in the tower environment.	OK	Successfully Validated
		EX5-CRT-05.971-TLR4-TVALP-FEAS-1022	Laboratory tests have verified that the integration of the V/A-R applications with other related system enablers is technically feasible.	OK	
EX5-OBJ-05.971-TRL4-TVALP- H103.1010	To assess that the technical systems for V/A-R Tracking Labels and overlays support the ATCOs in performing their tasks.	EX5-CRT-05.971-TLR4-TVALP-H103.1011	Majority of ATCOs (at least 75%) responses is that V/A-R supports ATCO in maintaining workload at acceptable level.	OK	Partially OK
		EX5-CRT-05.971-TLR4-TVALP-H103.1012	ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R	OK	
		EX5-CRT-05.971-TLR4-TVALP-H103.1013	Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO in maintaining an adequate level of situation awareness.	Not OK	

		EX5-CRT-05.971-TLR4-TVALP-H103.1014	Measured time spent in head up is increased in the solution scenario with respect to the reference scenario.	OK		
		EX5-CRT-05.971-TLR4-TVALP-H103.1015	HMI of V/A-R tools does not overshadow the relevant information on the OTW view.	OK		
		EX5-CRT-05.971-TLR4-TVALP-H103.1016	V/A-R HMI does not increase the potential for human error.	OK		
		EX5-CRT-05.971-TLR4-TVALP-H103.1017	ATCOs' (at least 75%) trust in the system is at an acceptable level.	OK		
		EX5-CRT-05.971-TLR4-TVALP-H103.1018	Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R HMI.	OK		
		CRT-05.971-TLR4-TVALP- H103.1020	ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R tool.	OK		
EX5-OBJ-05.971-TRL4-TVALP- H103.1030	To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R	EX5-CRT-05.971-TLR4-TVALP- H103.1031	Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner.	OK		Successfully Validated

EUROPEAN PARTNERSHIP

Co-funded by the European Union

	Tracking labels and overlays				
		EX5-CRT-05.971-TLR4-TVALP-H103.1032	Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions.	OK	
EX5-OBJ-05.971-TRL4-TVALP- H103.1040	To assess job acceptance and satisfaction with the introduction of V/A-R tracking labels and overlays	EX5-CRT-05.971-TLR4-TVALP-H103.1041	ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance.	OK	Successfully Validated
EX5-OBJ-05.971-TRL4-TVALP- H104.1010	To assess that the technical systems for V/A-R Air Gestures support the ATCOs in performing their tasks.	EX5-CRT-05.971-TLR4-TVALP-H104.1011	Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures support ATCO in maintaining workload at acceptable level.	OK	Partially OK
		EX5-CRT-05.971-TLR4-TVALP-H104.1012	ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R Air Gestures.	OK	
		EX5-CRT-05.971-TLR4-TVALP-H104.1013	Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures HMI supports ATCO in	Not OK	

			maintaining an adequate level of situation awareness.		
		EX5-CRT-05.971-TLR4-TVALP-H104.1014	Measured time spent in head up is increased in the solution scenario with respect to the reference scenario.	OK	
		EX5-CRT-05.971-TLR4-TVALP-H104.1015	V/A-R Air Gestures HMI does not increase the potential for human error	OK	
		EX5-CRT-05.971-TLR4-TVALP-H104.1016	ATCOs' (at least 75%) trust in the system is at an acceptable level.	OK	
		EX5-CRT-05.971-TLR4-TVALP-H104.1017	Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R Air Gestures HMI.	OK	
		EX5-CRT-05.971-TLR4-TVALP-H104.1018	ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R Air Gestures tool.	OK	
EX5-OBJ-05.971-TRL4-TVALP- H104.1020	To assess that the role of the ATCO is consistent with human capabilities and	EX5-CRT-05.971-TLR4-TVALP- H104.1021	Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an	Not OK	Partially OK

	limitations with the introduction of V/A-R Air Gestures		accurate, efficient and timely manner.		
		EX5-CRT-05.971-TLR4-TVALP- H104.1022	Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions.	OK	
OBJ-05.971-TRL4-TVALP- H104.1030	To assess job acceptance and satisfaction with the introduction of V/A-R Air Gestures	EX5-CRT-05.971-TLR4-TVALP-H104.1031	ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance.	OK	Successfully Validated
OBJ-05.971-TLR4-TVALP-SAFE.1010	To assess the impact of Virtual/Augmented Reality applications on safety.	EX5-CRT-05.971-TLR4-TVALP-SAFE-1011	Laboratory tests show that the Virtual/Augmented Reality applications improve the safety performance by reducing human error.	OK	Successfully Validated
		EX5-CRT-05.971-TLR4-TVALP-SAFE-1012	Laboratory tests show that the Virtual/Augmented Reality applications improve the safety performance by reducing ATCO workload.	OK	

		EX5-CRT-05.971-TLR4-TVALP-SAFE-1013	Laboratory tests show that the use of Virtual/Augmented Reality applications improves the safety performance by increasing situational awareness.	OK
		EX5-CRT-05.971-TLR4-TVALP-SAFE-1014	Safety assessment activities and the results are documented and integrated in the overall solution validation results.	OK
EX5-OBJ-05.971-TLR4-TVALP-PERF.1010	To assess the performance benefits of equivalent visual operations for tower control through the use of applications for Virtual/Augmented Reality.	EX5-CRT-05.971-TLR4-TVALP-PERF-1011	Laboratory tests show that the use of V/A applications improves Cost Efficiency performance by reducing the cost per flight (through e.g. reduction of workload, reduction of delay times).	OK
		EX5-CRT-05.971-TLR4-TVALP-PERF-1012	Laboratory tests show that the use of V/A reality applications improves Resilience by increasing situational awareness in low visibility conditions while maintaining	OK

	workload within acceptable limits.

Table D-21: Technological Validation Results EXE-005

D.7.1.1 Results on technological feasibility

EXE-005 successfully integrated live ADS-B data from a/c on Vitoria airport in an augmented reality device and presented the information blended with out of the window view. Linked virtual and real information was presented to controllers in different light conditions, such as day, sunset and night.

The V/A R device was able to understand two gestures that were used by controllers to navigate/filtrate the information. The gestures were used by technical people to prepare the functionalities.

Next UC described in the TS [27] were validated in the exercise. Feedback on each UC:

- [NSV-4][UC-103] Retrieve of information by means of V/AR. THE UC was successfully validated. No changes to the TS description have been identified.

- [NSV-4][UC-104] Tracking labels in Augmented Reality for landing/departing a/cs. The anti-overlapping functionality was finally not implemented. Due to the amount of simultaneous a/c at Vitoria airport, this was not considered as a drawback, but controllers consulted indicated that they would like to have it available. No changes to the TS description have been identified.

The interaction using the air gestures did not correspond to any of the UCs described in the intermediate TS. Coordination with the TS is needed to incorporate this interaction in an existing or new UC.

New requirements were identified:

- The ADS-B information was updated following the ADS-B rate. This resulted on information jumps in some cases. A smooth algorithm to present information should be investigated in TRL6. Members of the Solution has pointed out that remote towers already use this kind of algorithms and could be useful to investigate them.
- ADS-B large and short-range antennas were tested in the airport to gather the a/c information. The long-range antenna was good to detect en route and TMA a/c but bas for final approach and on surface a/c. The short-range antenna was good to detect final approach and surface movements but did not cover the complete runway (Vitoria runway has 3500 m length). An study of coverture regarding the a/c information source should be performed before deployment.

EXE-005 covered requirements from Sol.97 TS/IRS [27] are listed in the table below, which presents the requirements that have been validated and any comments detected/received during the technical validation preparation or execution.

Requirement Id	Requirement	Comments
REQ-05-W2-97.1-TS-VAR01.0001	The V/AR system shall depict conformal information as overlapped to the real object it is associated to	Validated
REQ-05-W2-97.1-TS-VAR01.0002	The V/AR system shall not obstruct the natural field of view of the ATCO with augmented reality elements.	Validated

REQ-05-W2-97.1-TS-VAR01.0003	The V/AR system shall be able to avoid cluttering of synthetic overlays that may obstruct the real view or overlap with other information.	Validated. In fact the transparency of the objects was an issue that controllers considered as important. Modifying the V/AR device brightness level helped to increase/decrease the opacity of the objects
REQ-05-W2-97.1-TS-AIRG01.0001	The air gesture system shall be able to recognize at least two different gestures and assign each of them to a specific function.	Validated
REQ-05-W2-97.1-TS-AIRG01.0002	The air gesture system shall be able to point at an object by means of gaze interaction and command it by means of air gesture.	Validated
REQ-05-W2-97.1-TS-AIRG01.0003	The air gesture system shall be able to provide contactless interaction (as opposed to multi touch interaction)	Validated
REQ-05-W2-97.1-TS-AIRG01.0004	The air gesture system shall be able to provide markerless tracking of the hand/fingers.	Validated
REQ-05-W2-97.1-TS-VAR01.0004	The V/AR system shall have a 30° x 15° minimum field of view for the augmented viewing port.	Validated
REQ-05-W2-97.1-TS-PERF.0001	Implementation of the V/AR system shall support the operational timing requirements expressed by the end-users	Confirmed. The prototype had this refresh rate, but it resulted in “jumps” in information presentation. A new requirement regarding algorithm smoothing has been introduced.
REQ-05-W2-97.1-TS-PERF.0002	The introduction of V/AR system shall preserve the level of performance of the current systems	Validated
REQ-05-W2-97.1-TS-PERF.0003	The latency and validation of surveillance position data for a/c and vehicles in the V/AR system should not exceed 1.5 seconds	Not covered
REQ-05-W2-97.1-TS-PERF.0004	The latency and validation of identification data for a/c and vehicles in the V/AR system should not exceed 3.5 seconds	Confirmed. The prototype had this refresh rate, but it resulted in “jumps” in information presentation. A new requirement regarding algorithm smoothing has been introduced.
REQ-05-W2-97.1-TS-PERF.0005	The refresh rate of information in the V/AR system should not be lower than 2.5 seconds 85% of the times	Validated
REQ-05-W2-97.1-TS-SEC0.0001	The V/AR system linked to the controllers’ working positions shall operate within a segregated network.	Validated

REQ-05-W2-97.1-TS-SEC0.0002	For data stored in V/AR system linked to the controllers' working positions there shall be a periodic backup procedure in place in order to guarantee recovery of corrupted or lost data.	Not covered
REQ-05-W2-97.1-TS-SEC0.0003	The V/AR system linked to the controllers' working positions shall be protected with appropriate Anti-Malware software or policies to avoid installation of malicious software.	Not covered
REQ-05-W2-97.1-TS-SEC0.0004	Data stored in the V/AR system linked to the controllers' working positions shall be protected through encryption procedures.	Not covered
REQ-05-W2-97.1-TS-SEC0.0005	Communication between involved actors, through radio and data links, shall be always ensured to prevent ground conflicts	Not covered

Table D-22: EXE-005 validated requirements

D.7.1.2 Results per KPA

The objectives are depicted per KPA. Please refer to the next section.

D.7.2 Analysis of EXE-005 Results per Technological Validation objective

This section provides a consolidated analysis of EXE-005 results per validation objective, subdivided per success criteria for each validation objective. For each Success criteria, the source of the analysis will be given, the results shown, and a determination of whether the criteria has proved successful or not. Then, combining all the success criteria results, an overall determination of whether the validation objective was achieved or not will be determined.

The questions asked of the participants are listed in Table D-23 below. As the question scales are not uniform (some are 1-5, others 0-7, etc.) all the ranges of responses have be normalized from 0-10, and have all been adjusted so that the range moves from a 0 signifying a very bad response, to a 10 as a very good response. This makes the conglomeration of various questions into an average response per success criteria possible.

97.1 EXE05 Questionnaire questions	
Situational Awareness	
In the previous working period,….	
1	I was ahead of the traffic
2	I started to focus on a single problem or a specific area of the airport surface.
3	there was a risk of forgetting something important (like providing a clearance on time or communicating a change to an a/c).
4	I was able to plan and organise my work as I wanted.
5	I was surprised by an event I did not expect (like an a/c call).
6	I had to search for an item of information.
7	Do you think that with the objects offered by the new functionality you have a **complete and reliable** representation of the traffic picture?
8	The new HMI (AR Goggles) **improves the capacity to focus attention** when and where appropriate.

9	Compared to your current situation, how do you think the new tools affected your situational awareness?
	Please provide additional comments about Situational Awareness
Workload	
In the past working period, how much effort did it take to…?	
10	Mental Demand
11	Physical Demand
12	Temporal Demand
13	Performance
14	Effort
15	Frustration
Workload II	
In the past working period, how much effort did it take to…?	
16	Identify the a/c in my operational area
17	Access relevant a/c or flight information
18	Integrate information from various sources to form a mental picture.
19	Recall necessary information
20	Recognise a mismatch of available data with the traffic picture
21	Would the weight of the AR goggles increase your perceived workload during an entire shift?
22	Would the comfort of the AR goggles increase your perceived workload during an entire shift?
23	Compared to your current situation, how do you think the new tools available affected your perceived workload?
	Please provide additional comments about workload perceived during this run
Trust and Acceptability	
24	the AR goggles were useful
25	the AR goggles were reliable
26	the AR goggles worked accurately
27	the AR goggles were understandable.
28	the AR goggles worked robustly (in difficult situations, with invalid inputs, etc.).
29	I was confident when working with the AR goggles .
Trust and Acceptability II	
30	The AR goggles increased the heads-up, out the window time compared to current operations.
31	The information provided by the AR goggles enabled me to identify the traffic whatever the callsign format was (radio-name, spelling, numbers,..)
32	The accuracy of the information provided by the AR goggles is adequate for the accomplishment of operations (number of **Callsign** Recognition, Error, & no recognition)
33	The information provided by the AR goggles was timely.
34	The accuracy of the information provided by the AR goggles is adequate for the accomplishment of operations (**Command** Recognition, command Error)
35	After a system error I experienced to have insufficient recovery time in order to manage my traffic safely.
	The problem/s was/were due to
	information presented but wrong
	information not timely presented
	information not presented
	misunderstanding of displayed information
36	I liked using the system
37	The system was easy to use
38	The AR goggles's operating methods are clearly identified and consistent in all operating conditions.
39	I had full trust (reliable and useful) in the system using AR goggles
40	I would like more / different information (or presentation) in the AR goggles.

	If you answered yes, elaborate your answer here:
	Please provide additional comments about Trust and Acceptability percieved during this run:
HMI – Callsign Tags	
41	The windows, fields, colors, fonts, eetc. Are consistent with the HMI displayed in other systems and subsystems within the CWP.
42	The new HMI of the AR goggles adequately highlights significant flights clearly and immediately
43	The a/c label HMI enabled me to maintain a degree of situational awareness that I deemed necessary
44	Callsigns & flight tags are presented in the most suitable place and do not overshadow the relevant information on the OTW view.
45	Callsign colour is adequate
46	Callsign size is adequate
47	Interaction with the graphical objects is quick and easy.
48	The information displayed in the AR goggles is concise and complete.
49	I would like more / different information displayed in the AR goggles.
	…if necessary, please elaborate
	…if necessary, please elaborate
HMI – Menus	
50	The Menus displayed are appropriate (accurate data at the appropriate site).
51	The fields shown in the Menus are sufficient
52	The size of the Menus were appropriate
53	Fonts and colors in the Menus are easily readable
54	The search for information in the Menus is intuitive
55	I generally appreciate the new features introduced by the AR goggles
	Please provide additional comments about HMI:
	Please provide additional comments about HMI:
56	Do you have any concerns regarding safety from your experience using the goggles? If so, please describe the situation(s) that occurred in the exercise or that might have occurred in the exercise or in reality. How was it detected and recovered? What factors contributed to that occurrence (e.g. Erroneous information provided by the system? What could have made it worse?
Safety	
57	The AR-system does not interfere with the availability and/or reliability of other systems and components installed at the CWP
58	The use of the AR system does not compromise ATCo task execution (slow-down, lag, freezes) and as a result, safety.
59	The inputs provided by the AR system are consistent and non-conflicting with the existing visual indications of flight progress displayed on the CWP.
60	Do you perceive the AR goggles as helpful in low-visibility conditions?
61	Do you think that the AR goggles could help change the low-visibility procedures to allow more movements?
62	Finally, do you think the functionality evaluated has helped you develop your work more safely and efficiently than today (during this run)?
	Please include any other comments/remarks

Table D-23: EXE-005 Questionnaire Questions

A success criteria is deemed successful if the normalized scores area above 50%. This seems sufficient for a TRL4 maturity validation. If the results are exactly 50%, they are deemed inconclusive.

D.7.2.1 EX5-OBJ-05.971-TRL4-TVALP-FEAS.1010

To confirm the concept is operationally feasible when addressing the Use Case for Virtual or Augmented Reality, tracking labels, and Air Gestures

EX5-CRT-05.971-TLR4-TVALP-FEAS-1011

No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Virtual or Augmented Reality and tracking labels.

This success criteria is answered by looking at the debrief comments and comments in the questionnaires. While there are some comments that need to be addressed in future iterations of the prototype regarding data stability, size of the markers in the HMI and the weight of the glasses, nothing was noticed that would lead us to believe that the technology should not go forward in development.

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-FEAS-1012

No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of Air Gestures.

This success criteria is also answered by looking at the debrief comments and comments in the questionnaires. While there are some comments that need to be addressed in future iterations of the prototype regarding training and the use of the air gestures, nothing was noticed that would lead us to believe that the technology should not go forward in development. In addition, the glasses used were the first generation Hololens glasses. The 3rd generation glasses are said to include air gestures that are more intuitive to learn, so future tests should take this into account.

The results for this success criteria are **OK.**

D.7.2.2 EX5-OBJ-05.971-TRL4-TVALP-FEAS.1020

To identify possible technical feasibility issues and possible show stoppers.

EX5-CRT-05.971-TLR4-TVALP-FEAS-1021

Laboratory tests (based on a prototype) have verified the technical feasibility of the use of V/A-R applications in the tower environment.

This success criteria is answered by looking at questions 24-38, shown in Figure D-103~~Figure D-103~~ below. The average of the normalized responses shows a 65% favourable response rate.

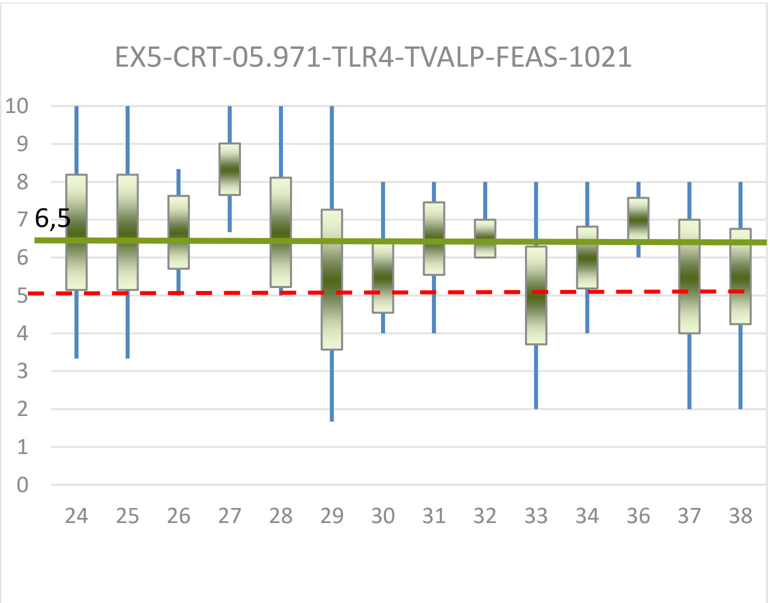


Figure D-103: CRT-05.97-TLR4-TVLAP-FEAS-1021 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-FEAS-1022

Laboratory tests have verified that the integration of the V/A-R applications with other related system enablers is technically feasible.

This success criteria is answered by looking at question 40, shown in Figure D-104~~Figure D-104~~ below. The average of the normalized responses shows a 62% favourable response rate.

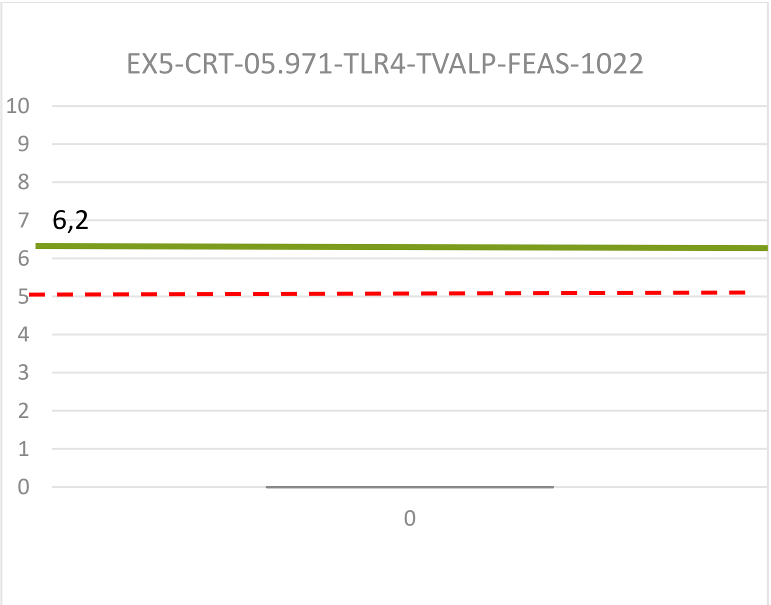


Figure D-104: CRT-05.97-TLR4-TVLAP-FEAS-1022 results

The results for this success criteria are **OK.**

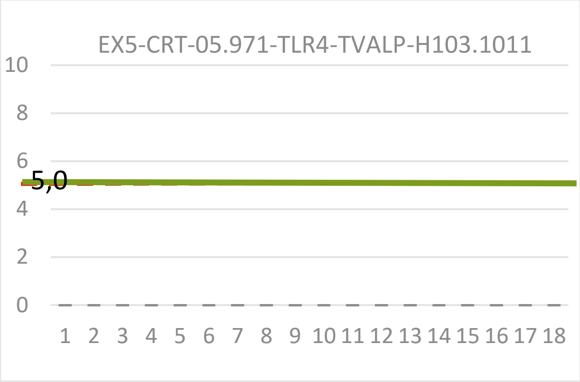
D.7.2.3 EX5-OBJ-05.971-TRL4-TVALP- H103.1010

To assess that the technical systems for V/A-R Tracking Labels and overlays support the ATCOs in performing their tasks.

EX5-CRT-05.971-TLR4-TVALP-H103.1011

Majority of ATCOs (at least 75%) responses is that V/A-R supports ATCO in maintaining workload at acceptable level.

This success criteria is answered by looking at the questions related to workload (questions 10-20, 23) shown in, shown in Figure D-105~~Figure D-105~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses show a 75% favourable response rate.



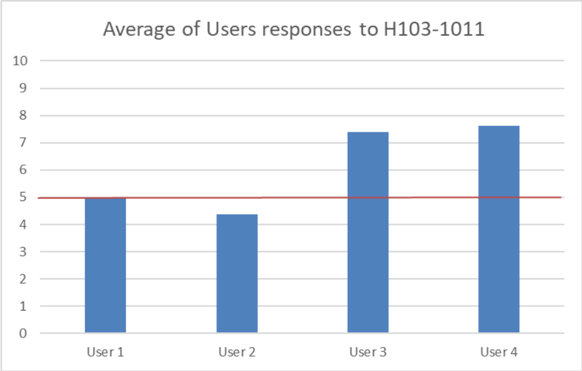


Figure D-105: CRT-05.97-TLR4-TVLAP-H103-1011 results

The results for this success criteria are **OK.**

Controllers consider that the weight and conform of the AR device influences greatly in the perceived workload (questions 21 and 22)

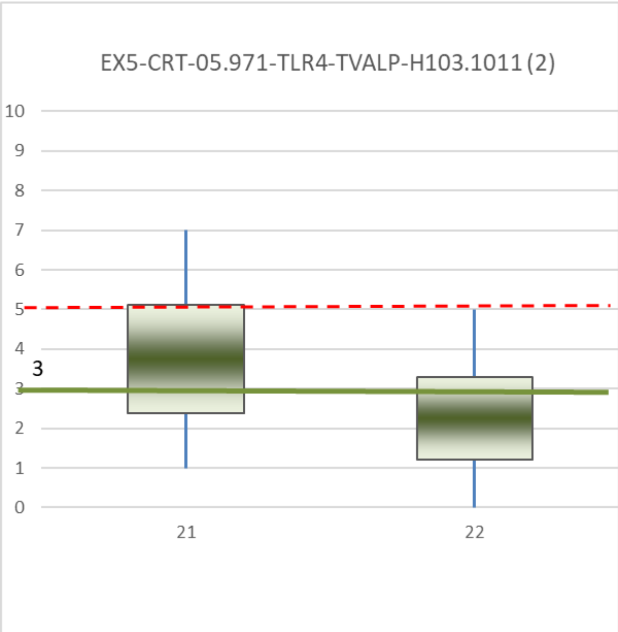


Figure D-106: CRT-05.97-TLR4-TVLAP-H103-1011 results - 2

EX5-CRT-05.971-TLR4-TVALP-H103.1012

ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R

This success criteria is answered by looking at questions 31-34, 40, 42, 48 and 49, shown in Figure D-107~~Figure D-107~~. It can be seen that 100% of the users responded favourably, and an average of the normalized responses show a 63% favourable response rate.

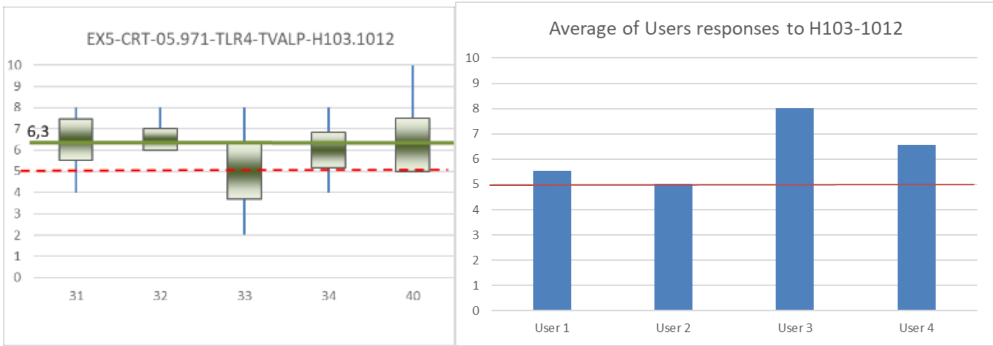


Figure D-107: CRT-05.97-TLR4-TVLAP-H103-1021 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H103.1013

Majority of ATCOs (at least 75%) responses is that V/A-R HMI supports ATCO in maintaining an adequate level of situation awareness.

This success criteria is answered by looking at questions 1-9 shown in Figure D-108~~Figure D-108~~. It can be seen that only 50% of the users responded favourably, but an average of the normalized responses show a 69% favourable response rate.

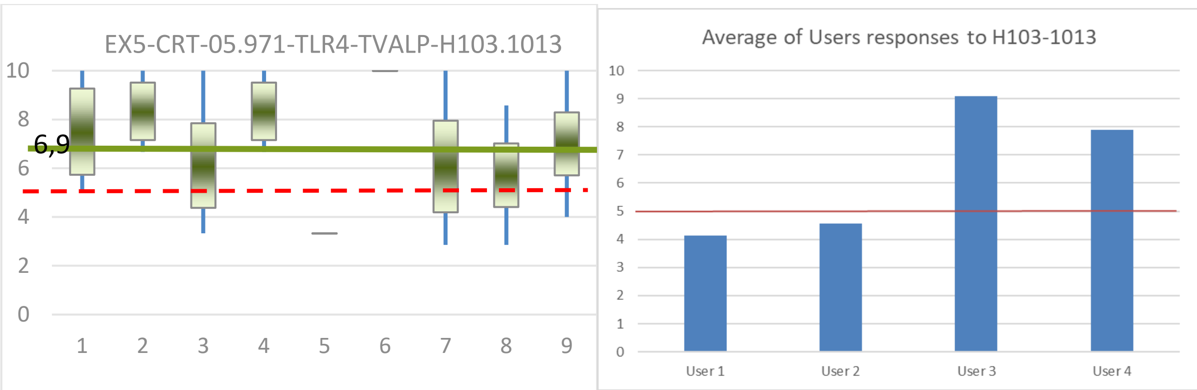


Figure D-108: CRT-05.97-TLR4-TVLAP-H103-1013 results

The results for this success criteria are **Not OK.**

EX5-CRT-05.971-TLR4-TVALP-H103.1014

Measured time spent in head up is increased in the solution scenario with respect to the reference scenario.

This success criteria is answered by looking at question 30, shown in Figure D-109~~Figure D-109~~. The average of the normalized responses shows a 55% favourable response rate.

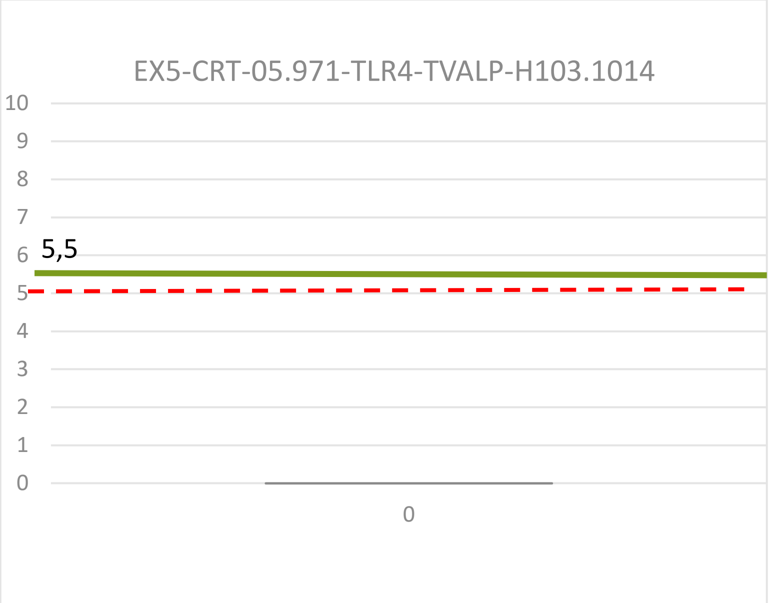


Figure D-109: CRT-05.97-TLR4-TVLAP-H103-1014 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H103.1015

HMI of V/A-R tools does not overshadow the relevant information on the OTW view.

This success criteria is answered by looking at questions 44 and 57, shown in Figure D-110~~Figure D-110~~. The average of the normalized responses shows a 63% favourable response rate.

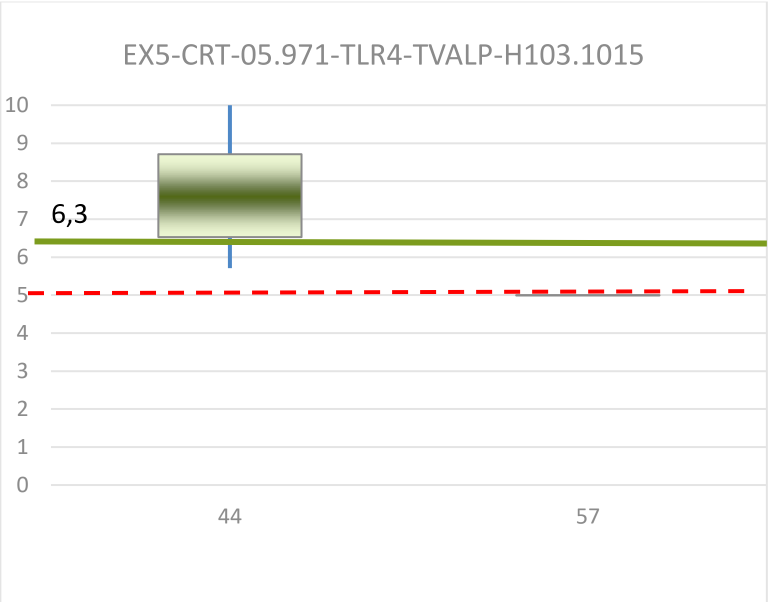


Figure D-110: CRT-05.97-TLR4-TVLAP-H103-1015 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H103.1016

V/A-R HMI does not increase the potential for human error.

This success criteria is answered by looking at questions 1-11, shown in Figure D-111~~Figure D-111~~. The average of the normalized responses show a 64% favourable response rate.

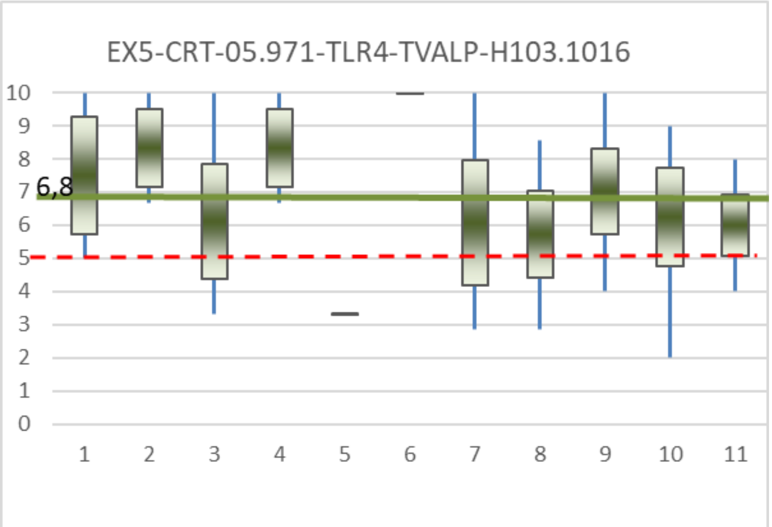


Figure D-111: CRT-05.97-TLR4-TVLAP-H103-1016 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H103.1017

ATCOs' (at least 75%) trust in the system is at an acceptable level.

This success criteria is answered by looking at question 39, shown in Figure D-112~~Figure D-112~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses show a 60% favourable response rate.

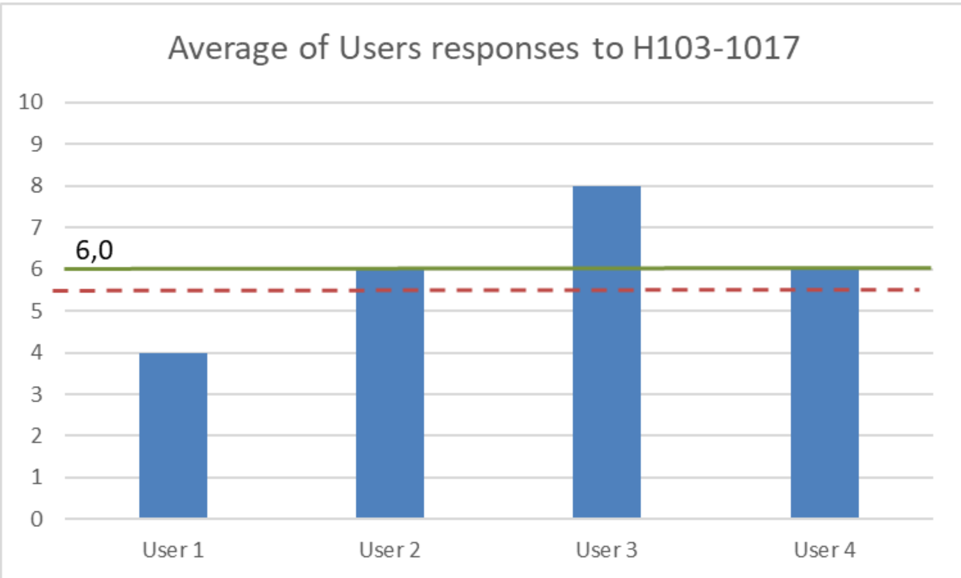


Figure D-112: CRT-05.97-TLR4-TVLAP-H103-1017 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H103.1018

Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R HMI.

This success criteria is answered by looking at questions 24, 51-55, and 58, shown in Figure D-113~~Figure D-113~~. It can be seen that 100% of the users responded favourably, and an average of the normalized responses show a 73% favourable response rate.

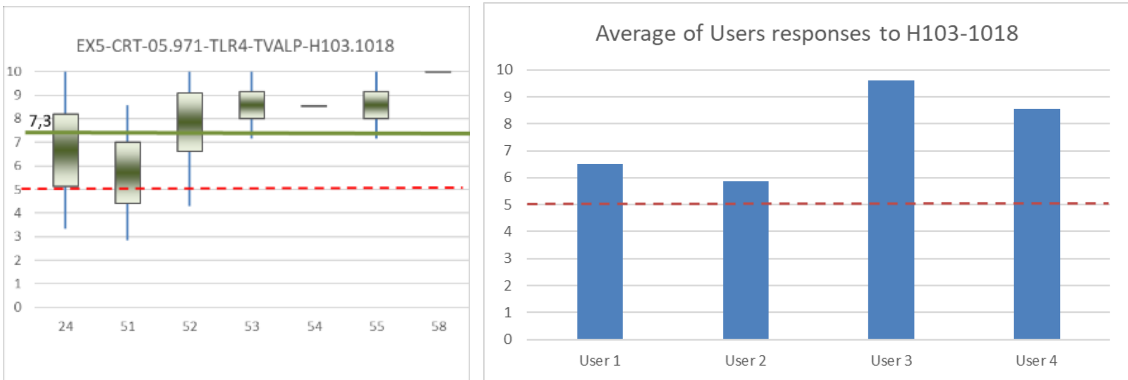


Figure D-113: CRT-05.97-TLR4-TVLAP-H103-1018 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H103.1020

ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R tool.

This success criteria is answered by looking at questions 24-40, shown in Figure D-114~~Figure D-114~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses show a 65% favourable response rate.

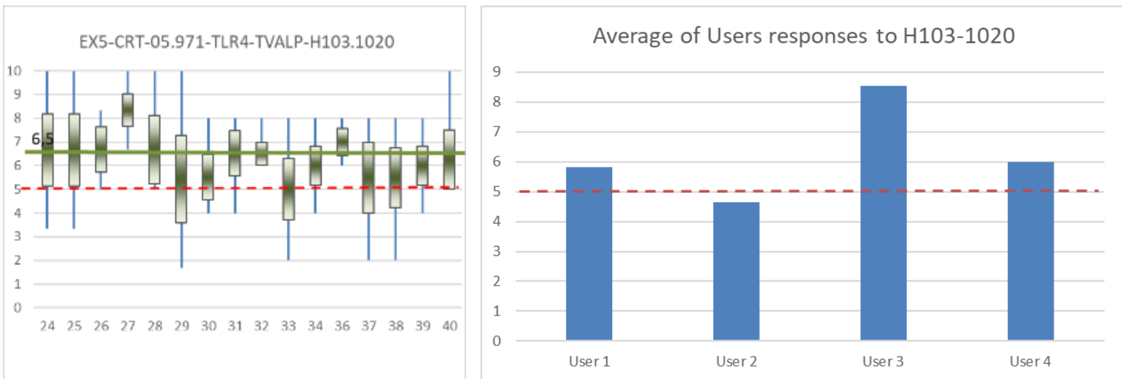


Figure D-114: CRT-05.97-TLR4-TVLAP-H103-1020 results

The results for this success criteria are **OK.**

D.7.2.4 EX5-OBJ-05.971-TRL4-TVALP- H103.1030

To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Tracking labels and overlays.

EX5-CRT-05.971-TLR4-TVALP- H103.1031

Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner.

This success criteria is answered by looking at questions 33, 38, and 50, shown on Figure D-115~~Figure D-115~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses show a 59% favourable response rate.

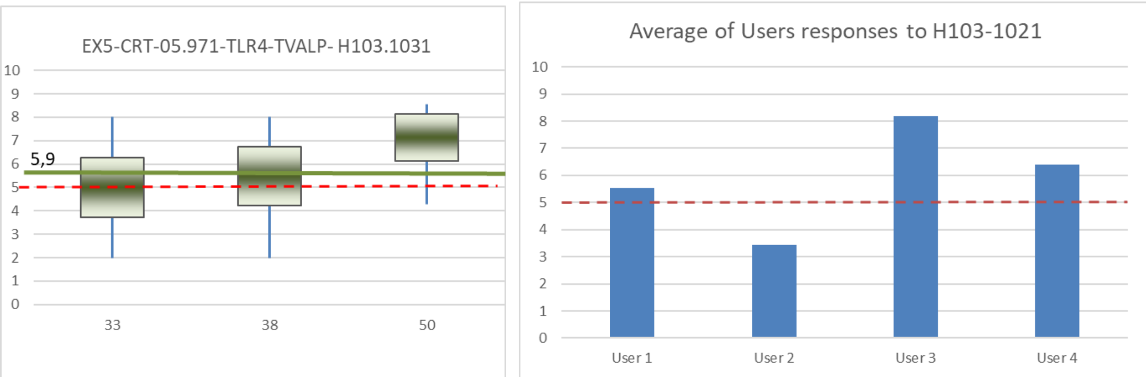


Figure D-115: CRT-05.97-TLR4-TVLAP-H103-1031 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP- H103.1032

Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions.

This success criteria is answered by looking at question 38, as shown in Figure D-116~~Figure D-116~~. It can be seen that 75% of the users responded favourably, and an average response of the four was a 55% favourable score.

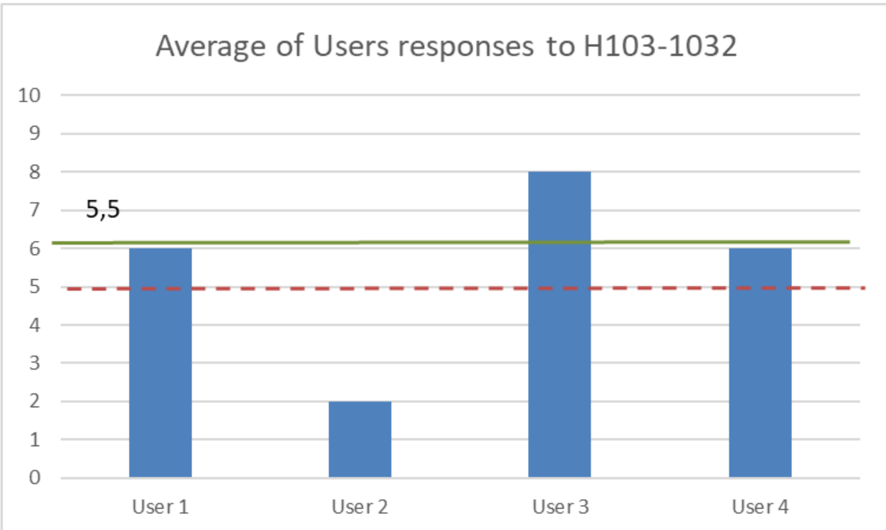


Figure D-116: CRT-05.97-TLR4-TVLAP-H103-1032 results

The results for this success criteria are **OK.**

D.7.2.5 EX5-OBJ-05.971-TRL4-TVALP- H103.1040

To assess job acceptance and satisfaction with the introduction of V/A-R tracking labels and overlays

EX5-CRT-05.971-TLR4-TVALP-H103.1041

ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance.

This success criteria is answered by looking at questions 36-40, shown in Figure D-117~~Figure D-117~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses show a 65% favourable response rate.

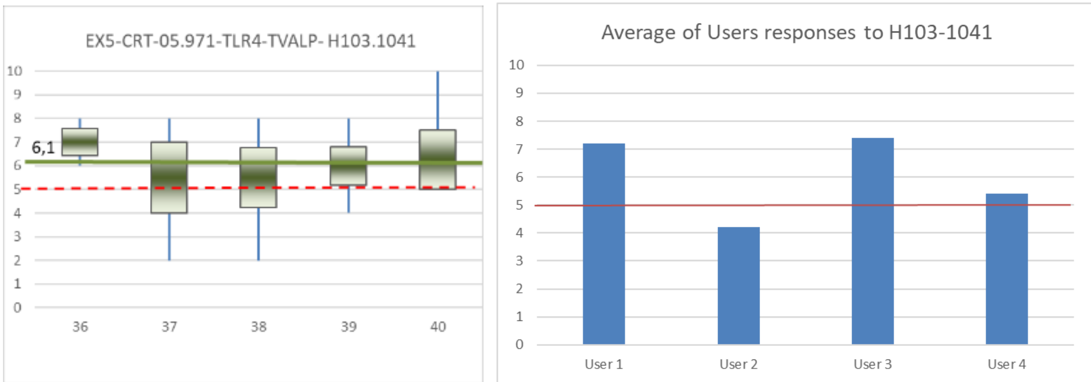


Figure D-117: CRT-05.97-TLR4-TVLAP-H103-1041 results

The results for this success criteria are **OK.**

D.7.2.6 EX5-OBJ-05.971-TRL4-TVALP- H104.1010

To assess that the technical systems for V/A-R Air Gestures support the ATCOs in performing their tasks.

EX5-CRT-05.971-TLR4-TVALP-H104.1011

Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures support ATCO in maintaining workload at acceptable level.

This success criteria is answered by looking at questions 10-20 and 23, shown in Figure D-118~~Figure D-118~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses show a 66% favourable response rate.

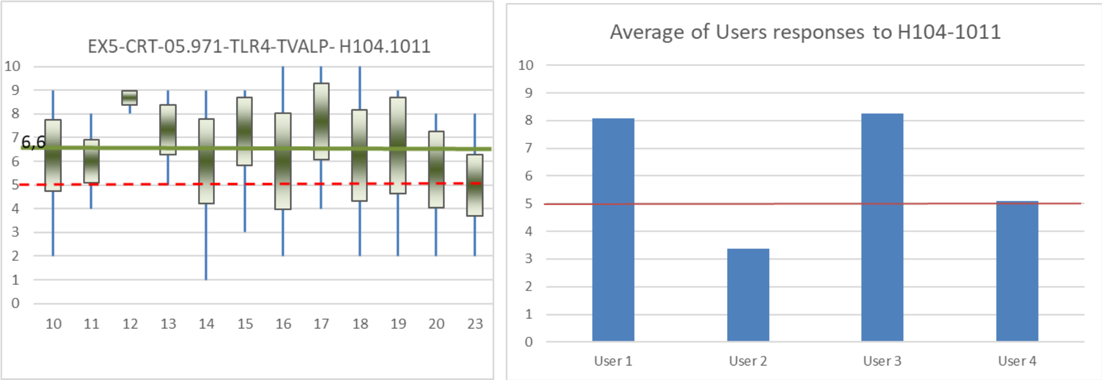


Figure D-118: CRT-05.97-TLR4-TVLAP-H104-1011 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H104.1012

ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by V/A-R Air Gestures.

This success criteria is answered by looking at question 34, shown in Figure D-119~~Figure D-119~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses show a 60% favourable response rate.

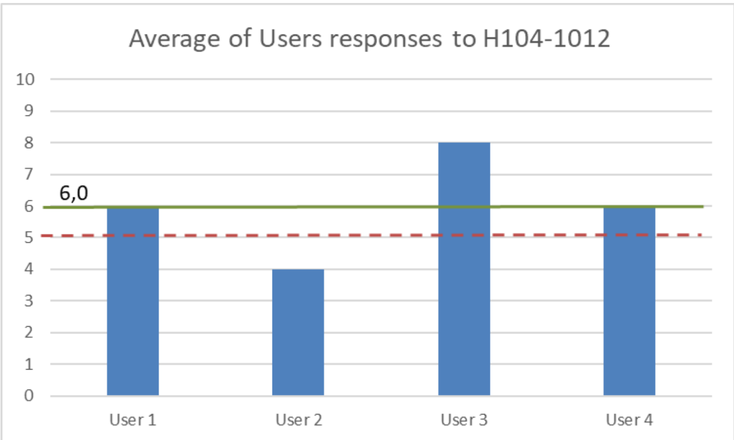


Figure D-119: CRT-05.97-TLR4-TVLAP-H104-1012 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H104.1013

Majority of ATCOs (at least 75%) responses is that V/A-R Air Gestures HMI supports ATCO in maintaining an adequate level of situation awareness.

This success criteria is answered by looking at questions 1-9, shown Figure D-120~~Figure D-120~~. As in H103-1013, it can be seen that only 50% of the users responded favourably, but an average of the normalized responses show a 69% favourable response rate.

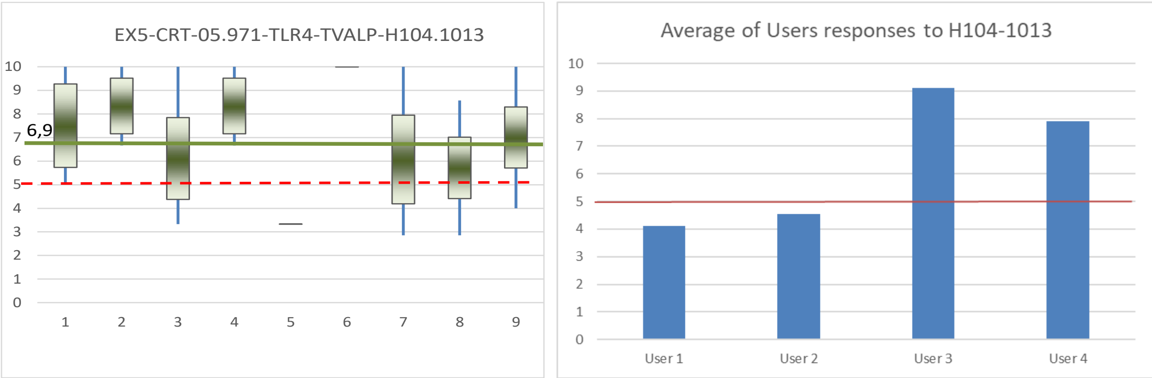


Figure D-120: CRT-05.97-TLR4-TVLAP-H104-1013 results

The results for this success criteria are **Not OK.**

EX5-CRT-05.971-TLR4-TVALP-H104.1014

Measured time spent in head up is increased in the solution scenario with respect to the reference scenario.

This success criteria is answered by looking at question 30, shown in Figure D-121~~Figure D-121~~. An average of the normalized responses show a 55% favourable response rate.

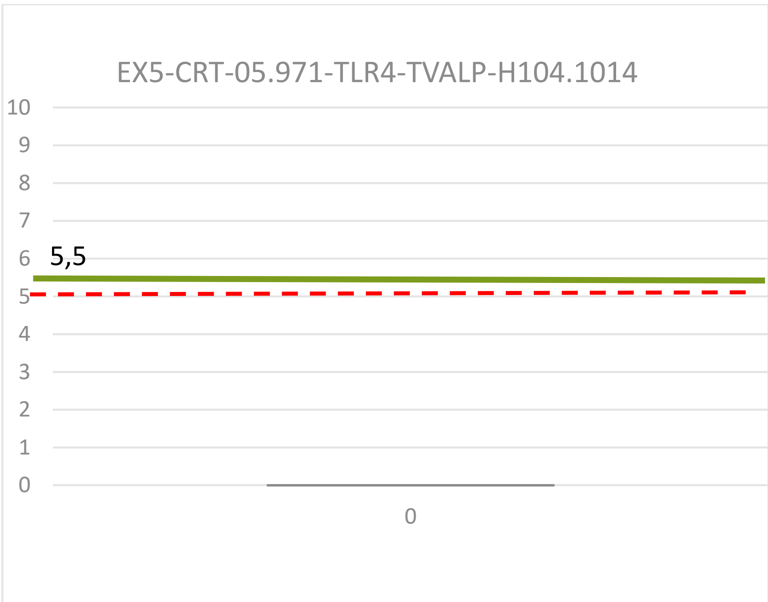


Figure D-121: CRT-05.97-TLR4-TVLAP-H104-1014 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H104.1015

V/A-R Air Gestures HMI does not increase the potential for human error. This success criteria is answered by looking at situational awareness questions 1-9, shown in Figure D-122~~Figure D-122~~. An average of the normalized responses show a 69% favourable response rate.

The increase in situational awareness reduces the likelihood of human error such as allowing two simultaneous a/c at the same low visibility block.

The human error is also assessed through observations. At the initialization of the runs controllers experienced some difficulties to be recognized by the system. This problem was solved after some minutes and can be associated to deficient training.

No human errors were observed during the execution of the experiment.

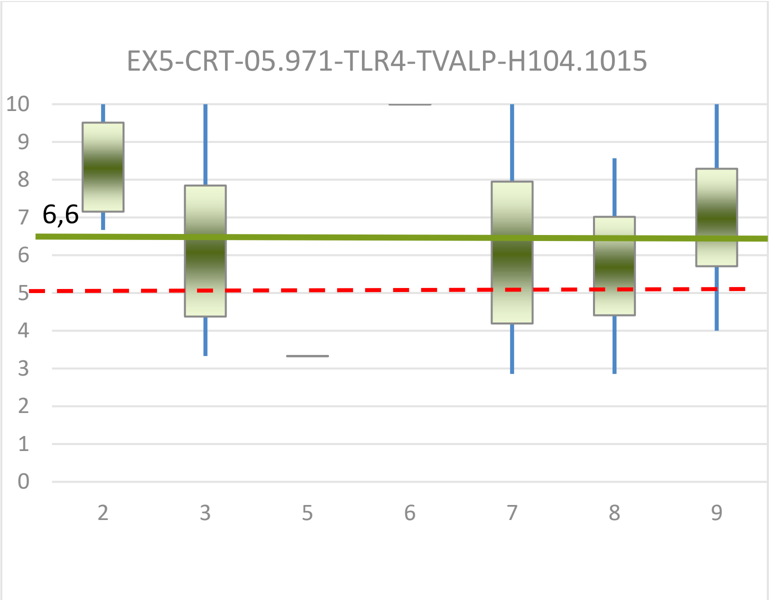


Figure D-122: CRT-05.97-TLR4-TVLAP-H104-1015 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H104.1016

ATCOs' (at least 75%) trust in the system is at an acceptable level.

This success criteria is answered by looking at question 39, shown in Figure D-123~~Figure D-123~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses shows a 60% favourable response rate.

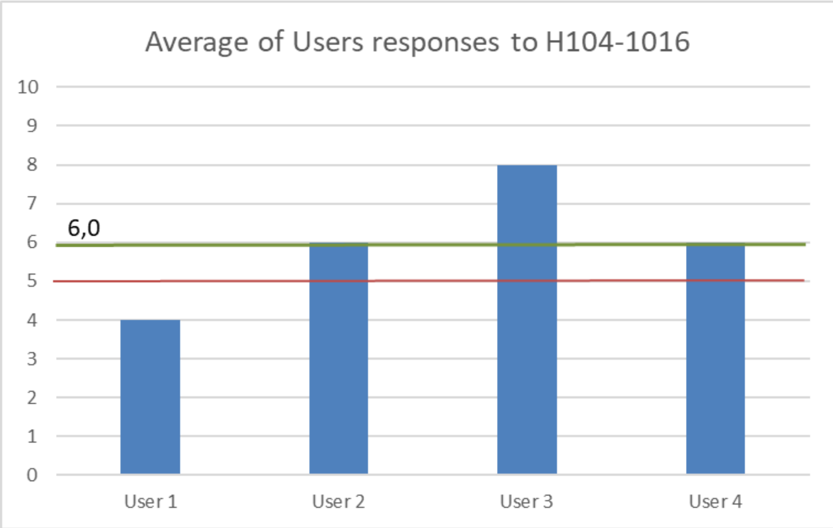


Figure D-123: CRT-05.97-TLR4-TVLAP-H104-1016 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H104.1017

Majority of ATCOs (at least 75%) confirm an adequate level of usability of V/A-R Air Gestures HMI.

This success criteria is answered by looking at questions 45-47, shown in Figure D-124~~Figure D-124~~. It can be seen that all the users responded favourably, and an average of the normalized responses shows a 67% favourable response rate.

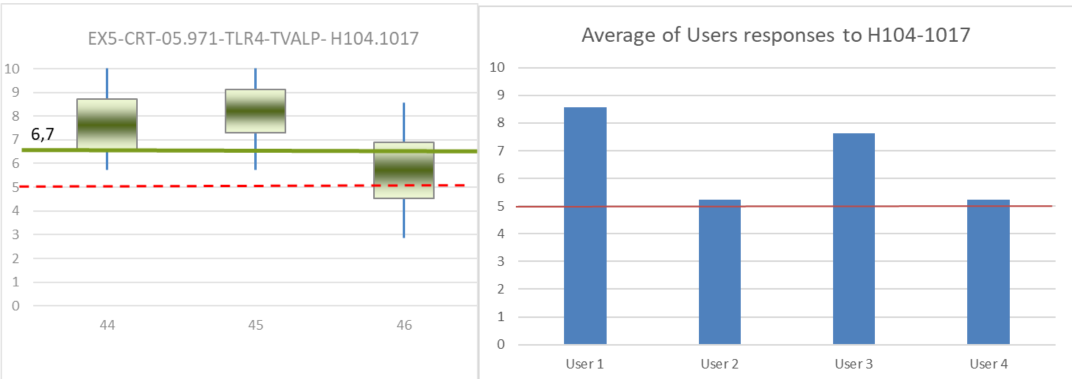


Figure D-124: CRT-05.97-TLR4-TVLAP-H104-1017 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-H104.1018

ATCOs (at least 75%) provide positive feedback on acceptance of V/A-R Air Gestures tool.

This success criteria is answered by looking at questions 48 and 49, shown in Figure D-125~~Figure D-125~~. It can be seen that all the users responded favourably, and an average of the normalized responses shows a 63% favourable response rate.

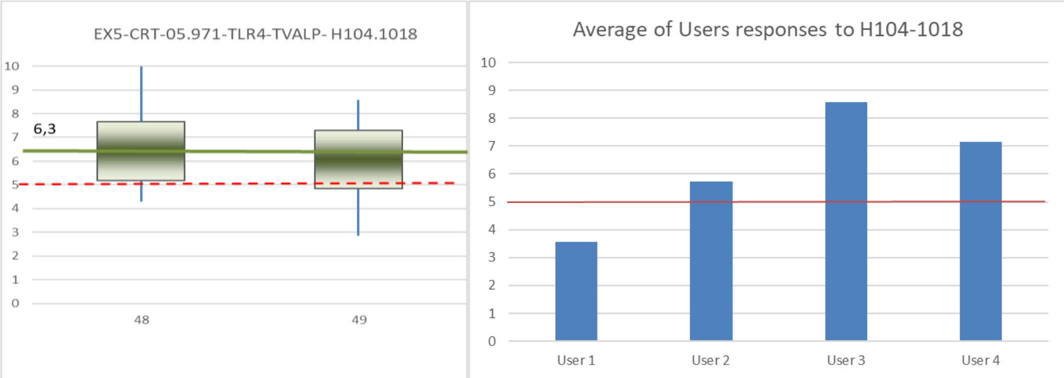


Figure D-125: CRT-05.97-TLR4-TVLAP-H104-1018 results

The results for this success criteria are **OK.**

D.7.2.7 EX5-OBJ-05.971-TRL4-TVALP- H104.1020

To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of V/A-R Air Gestures

EX5-CRT-05.971-TLR4-TVALP- H104.1021

Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient and timely manner.

This success criteria is answered by looking at questions 33, 38,50, shown in Figure D-126~~Figure D-126~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses show a 59% favourable response rate.

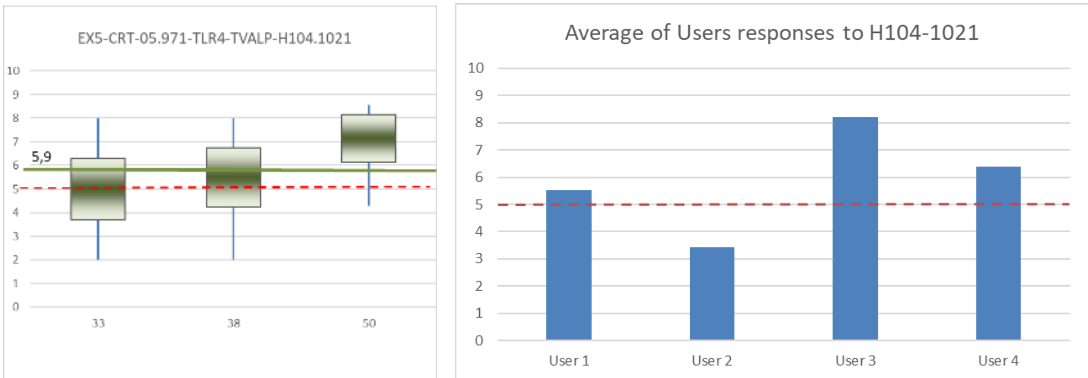


Figure D-126: CRT-05.97-TLR4-TVLAP-H104-1021 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP- H104.1022

Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions.

 Co-funded by the European Union

This success criteria is answered by looking at question 38, shown on Figure D-127~~Figure D-127~~. It can be seen that 75% of the users responded favourably, and an average of the normalized responses show a 55% favourable response rate.

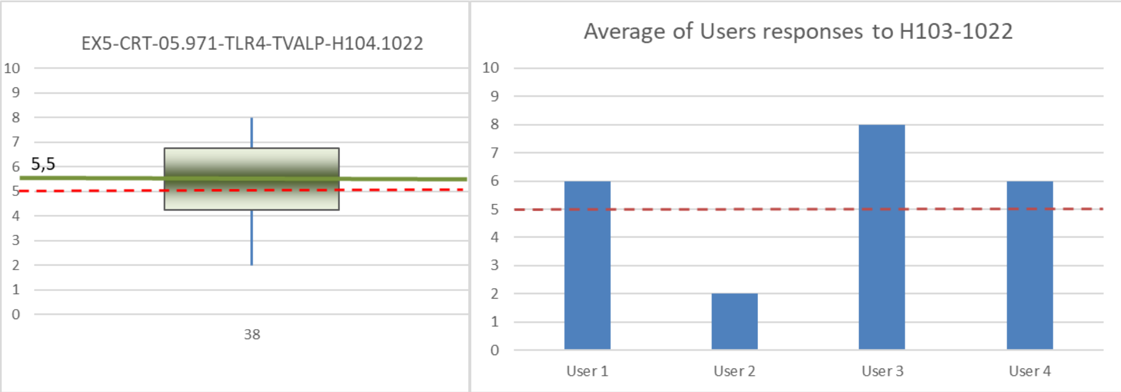


Figure D-127: CRT-05.97-TLR4-TVLAP-H104-1022 results

The results for this success criteria are **OK.**

D.7.2.8 EX5-OBJ-05.971-TRL4-TVALP- H104.1030

To assess job acceptance and satisfaction with the introduction of V/A-R Air Gestures

EX5-CRT-05.971-TLR4-TVALP-H104.1031

ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance.

This success criteria is answered by looking at all the questions, since every aspect of the questionnaire is a part of job satisfaction and acceptance of the new tool. An average of the normalized responses in Figure D-128~~Figure D-128~~ shows a 63% favourable response rate.

The results for this success criteria are **OK.**

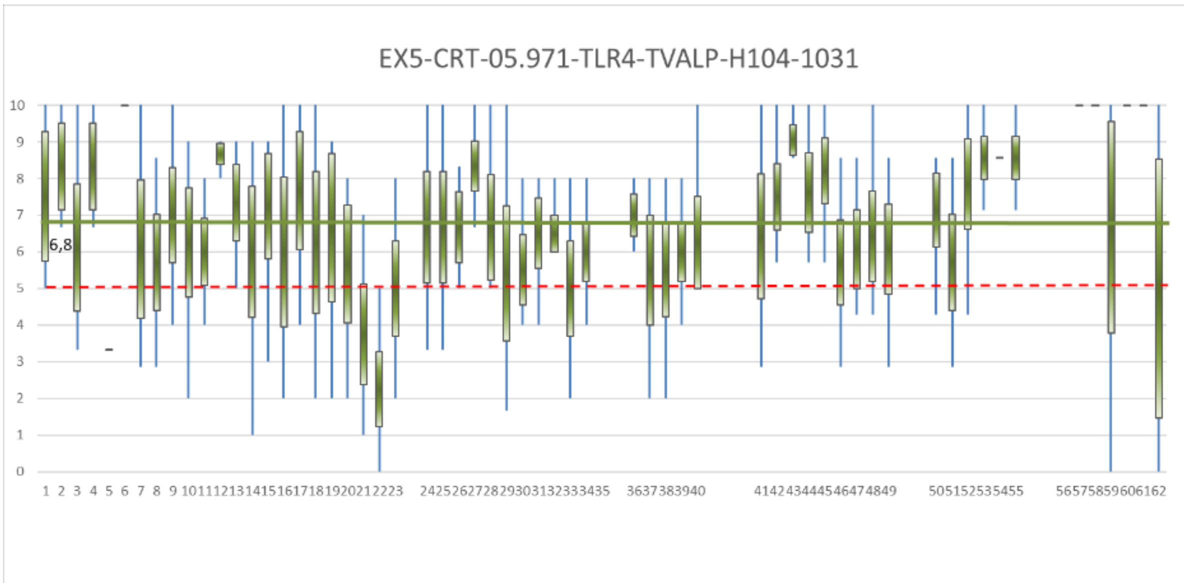


Figure D-128: CRT-05.97-TLR4-TVLAP-SAFE-1031 results

D.7.2.9 EX5-OBJ-05.971-TLR4-TVALP-SAFE.1010

To assess the impact of Virtual/Augmented Reality applications on safety.

EX5-CRT-05.971-TLR4-TVALP-SAFE-1011

Laboratory tests show that the Virtual/Augmented Reality applications improve the safety performance by reducing human error.

This success criteria is answered by looking at situational awareness questions 1-9, shown on Figure D-129~~Figure D-129~~. An average of the normalized responses show a 69% favourable response rate.

The increase in situational awareness reduces the likelihood of human error such as allowing two simultaneous a/c at the same low visibility block. The human error is also assessed through observations. No human errors were observed during the execution of the experiment.

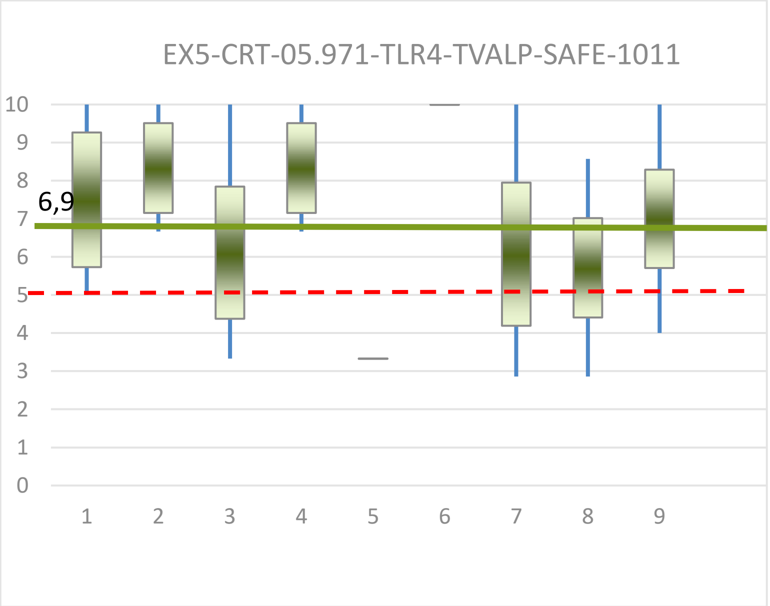


Figure D-129: CRT-05.97-TLR4-TVLAP-SAFE-1011 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-SAFE-1012

Laboratory tests show that the Virtual/Augmented Reality applications improve the safety performance by reducing ATCO workload.

This success criteria is answered by looking at the questions related to workload (questions 10-20 & 23, shown below).

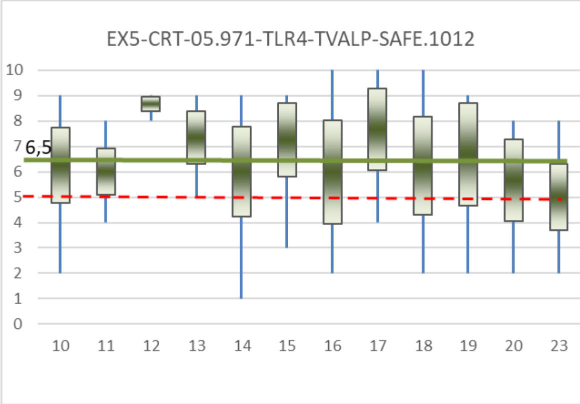


Figure D-130: CRT-05.97-TLR4-TVLAP-SAFE-1012 results

An average of the normalized responses show a 69% favourable response rate.

The results for this safety success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-SAFE-1013

Laboratory tests show that the use of Virtual/Augmented Reality applications improves the safety performance by increasing situational awareness.

This success criteria is answered by looking at questions 9 and 43, shown in Figure D-131~~Figure D-131~~. An average of the normalized responses show a 80% favourable response rate.

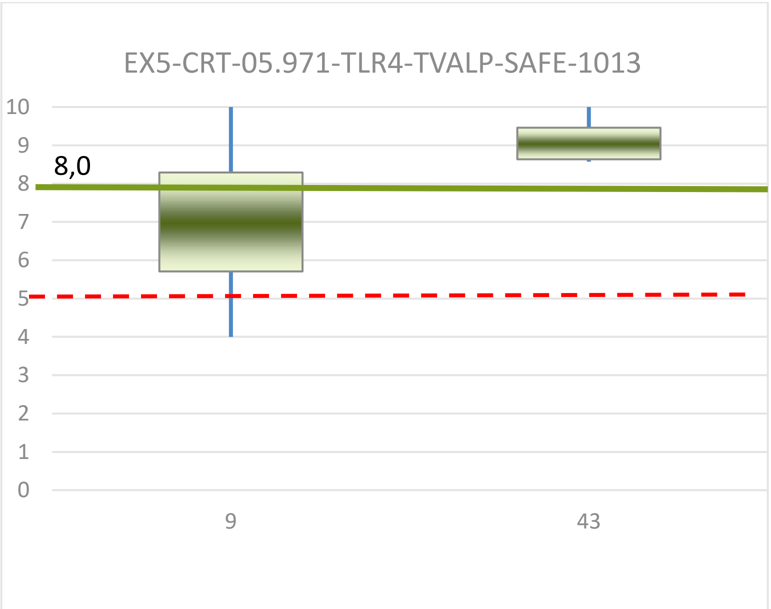


Figure D-131: CRT-05.97-TLR4-TVLAP-SAFE-1013 results

The results for this success criteria are **OK.**

EX5-CRT-05.971-TLR4-TVALP-SAFE-1014

Safety assessment activities and the results are documented and integrated in the overall solution validation results. Coordination regarding safety documentation to take into account the outcome of this exercise is in place.

Relevant safety comments are related to questions 56 through 62 regarding perceived safety, shown in Figure D-132~~Figure D-132~~. The only comment regarding Safety was related to some data dropouts of the altitude and speed indicators on the flight tags. An average of the normalized responses show a 57% favourable response rate.

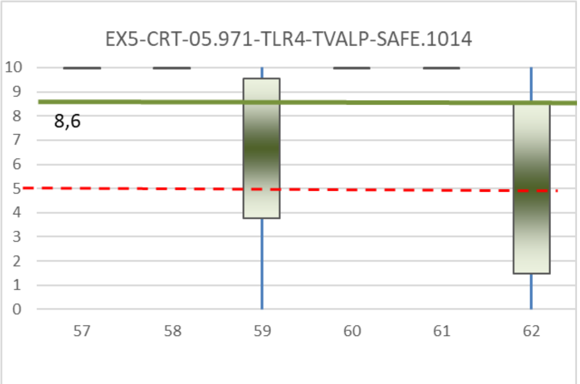


Figure D-132: CRT-05.97-TLR4-TVLAP-SAFE-1014 results

The results for this success criteria are **OK.**

D.7.2.10 EX5-OBJ-05.971-TRL4-TVALP-PERF.1010

To assess the performance benefits of equivalent visual operations for tower control through the use of applications for Virtual/Augmented Reality.

EX5-CRT-05.971-TLR4-TVALP-PERF-1011

Laboratory tests show that the use of V/A applications improves Cost Efficiency performance by reducing the cost per flight (through e.g. reduction of workload, reduction of delay times).

This success criteria is answered by looking at question 61, shown in Figure D-133~~Figure D-133~~. An average of the normalized responses show a 100% favourable response rate.

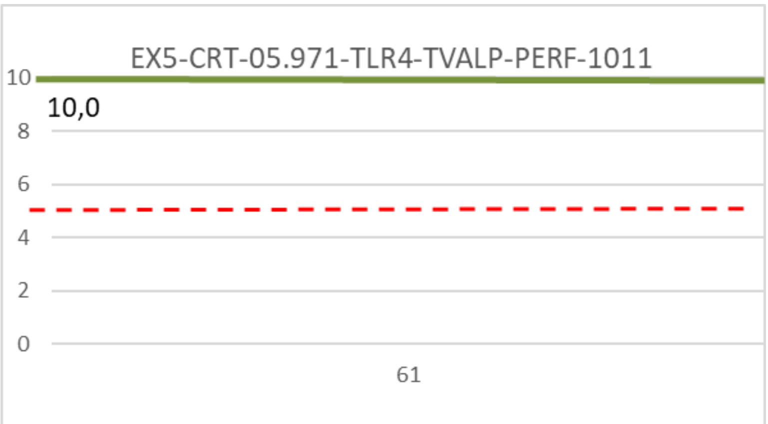


Figure D-133: CRT-05.97-TLR4-TVLAP-PERF-1011 results

The results for Cost Efficiency are **OK**. Nevertheless this outcome needs to be confirmed by the CBA taking into account all the parameters.

EX5-CRT-05.971-TLR4-TVALP-PERF-1012

Laboratory tests show that the use of V/A reality applications improves Resilience by increasing situational awareness in low visibility conditions while maintaining workload within acceptable limits.

This success criteria is answered by looking at question 60, shown in Figure D-134~~Figure D-134~~. An average of the normalized responses show a 100% favourable response rate.

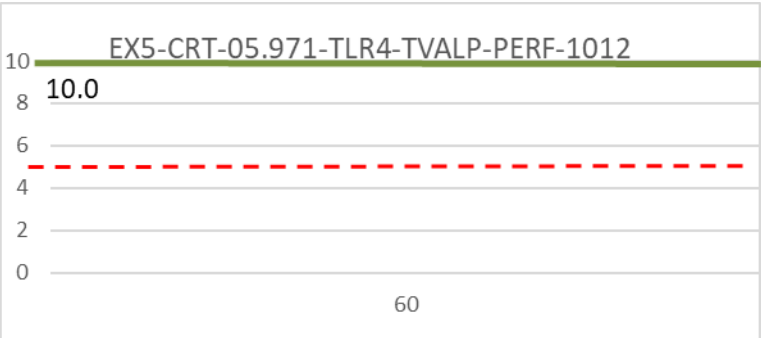


Figure D-134: CRT-05.97-TLR4-TVLAP-PERF-1012 results

The results for Resilience are **OK**. Nevertheless this outcome needs to be confirmed by the as the outcome from Situational awareness was not as good as expected.

D.7.3 Unexpected behaviours/results

When setting up the experiment at the Vitoria airport, there was a certain amount of trouble with acquiring the ADS-B Signal (antenna placement/directionality) below a certain altitude. A/c in the air above a few hundred feet could be acquired with ease, but the signal from a/c below that altitude or on the ground could not be acquired. It was later understood that this was due to the directionality of the antenna. Once the antenna was placed in a position of about 45º from horizontal, the ground traffic could be acquired. Troubleshooting this problem reduced the time available for the controllers to test the glasses from 5 days down to two. This reduction in the number of controllers accounts for the medium confidence in the performance and HP results.

Another unexpected behaviour was some data dropouts during final approach. As the a/c would approach the threshold, the altitude and speed data would sometimes drop out to zero. This was caused by an intermittent loss of signal that was interpreted by the HMI engine as a parked a/c. It was a problem that could easily be fixed in later exercises and doesn't pose a long term issue with safety, but the development of the solution needs to be monitored.

D.7.4 Confidence in results of EXE-005

D.7.4.1 Level of significance/limitations of Technological Validation Exercise Results

Technologically speaking, given the fact that it was a shadow mode exercise, using the actual ADS-B data, performed in an actual tower environment, and overlaying the HMI on the real traffic, the confidence was very high being the target maturity TRL4.

D.7.4.2 Quality of Technological Validation Exercises Results

The quality of the results for a TLR4 exercise is considered high but the fact that no antenna coverage analysis was performed prior to the execution of the exercise should still be taken into consideration

D.7.4.3 Significance of Technological Validation Exercises Results

When looking at the questionnaire results, the fact that there were only 4 participants can lead to one outlier score skewing the results. However, since the results broadly follow previous results from RETINA, a medium confidence in the results of this exercise can be assumed.

D.8 Conclusions

Overall, the controllers were very interested in the solution, especially for use in night or LVC operations. The presentation of the information was deemed satisfactory, with some mentions of improvements for future versions regarding the size of the information presented in the overlays.

The weight of the glasses is still an issue for use during an entire shift. The final solution should have a considerable lighter device.

D.8.1 Conclusions on technological feasibility

All signs point to a feasible implementation. The ADS-B data, once the antenna placement problems were resolved, was processed and presented to the controller is a seamless manner. The air gestures used to control the menu could be a bit more intuitive, but after some time practicing them, were easily understood and accomplished.

D.8.2 Conclusions on performance assessments

Safety

The use of the glasses was shown to be beneficial to safety in night or LVC operations in airports that do not have surface surveillance systems such as ground radar. There were no showstoppers, but the data dropout and the tag jumps issues need to be resolved.

Other areas of performance - controller workload, resilience, etc.

Results in the questionnaires that were favourable need further confirmation from the other S97 exercise. Situational awareness feedback was not as good as expected. Controllers related this feedback to the dropout and tag jumps.

D.8.3 Recommendations

Even though EXE-005 greatly progressed the validation of the use of Augmented Reality glasses in the tower environment, there are certain things that need further testing and validation before the technology can be deemed ready for operational deployment.

- **Data smoothing** – The jumps that the flight tag makes from one update of the ADS-B position information to the next is still a bit jarring and could benefit from a data smoothing interpretation algorithm like is used in other ground surveillance displays. This would help the flight tag follow the actual a/c as it moves. This could easily be added to the display processes.

- **Antenna coverage study** – In addition to the angle of the ADS-B antenna with respect to the horizon, the location of the antenna with respect to other obstacles such as airport structures could interfere with continuous coverage while the a/c move around the airport surface. Since Vitoria airport has a lot of cargo traffic, it has a long runway and the coverage at the far end of the runway was blocked by some buildings. Before deployment, a coverage analysis to determine if one or more antennas are required for full coverage should be completed.
- **Glasses Ergonomics** – The glasses used in this exercise were a 1st generation model. One of the complaints was that they were too heavy and cumbersome for use during an entire shift. More study using a 3rd generation device should be done to study how lower weight and a wider angle of view would be sufficient for use during an entire shift, or at least for longer durations.
- **HMI** – Further study should be taken on to determine the proper size of the flight tags and tracking ball that balances sufficient information, ease of viewing, and occupies the minimal field of view. The new air gesture functionality for the 3rd generation devices should also be investigated to see if the more intuitive gestures improve usability and reduce training time.
- **Filters** - The addition of an altitude filter to allow the controller to filter out a/c that are either flyovers or outside the scope of their control.

Appendix E Technological Validation Exercise 006 Report

E.1 Summary of EXE-006 Plan

This section reports about various aspects of the technological validation exercise for EXE-05.97.2-TRL4-TVALP-ASR-006 (abbreviated as EXE-006). There were no major deviations from the technical validation plan [28]. Details about the conduction of the exercise and result analysis are given in the following.

E.2 EXE-006 description and scope

EXE-006 investigated the benefits of an Assistant Based Speech Recognition (ABSR) system coupled with an electronic flight strip system for air traffic controllers working within a simulated Multiple Remote Tower environment. The hardware setup can be seen in Figure E-135~~Figure E-135~~.

Figure E-135: Multiple Remote Tower environment with a row of monitors per airports under ATCo control

The ATCo had to control air traffic and ground vehicles on three remote airports (named Vilnius, Kaunas, and Palanga) with rather simple airport topology at the same time. A radar display per each of the airports visualized the air traffic in the airport's vicinity. The biggest airport (Vilnius) also had a ground radar display (as it is in a real working environment). The electronic flight strip system consisted of one column per airport and four bays per column. The electronic flight strips changed their bays with further progress of the flight status when arriving or departing. All displays are prototypic DLR developments to enable usage of ATCos from many different countries that are used to different systems at their own controller working position within their air navigation service provider. During the simulations ATCos mainly needed to give the ATC clearance, allow for startup and pushback, instruct taxi, lineup and takeoff/landing clearances for the single runway in use next to handling special situations on ground. Three simulation pilots (one for each airport) in another room communicated

EUROPEAN PARTNERSHIP

Co-funded by the European Union

with the ATCo to run air and ground traffic with the support of a pseudo-pilot interface (see Figure E-136~~Figure E-136~~).

Figure E-136: Pseudo-pilot interface for Palanga airport with flight strips and radar view

The developed prototypic ABSR system used radar data, flight plan data, and meteorological data to predict a/c callsigns and controller commands for the next ATCo utterances using machine learning algorithms. Those forecasted callsign and command hypotheses supported an automatic speech recognition engine to automatically recognize word sequences. The engine's recognized word sequences from the ATCo utterance (speech-to-text, transcription) were then in turn automatically analysed to extract meaningful content, i.e., ATC concepts such as commands with callsigns, command types, values, units, etc. (text-to-concept, annotation). The outcome of the ABSR system was only shown to the ATCos in solution scenarios (not in the reference scenario where a side laptop for ABSR output visualization was evaluated by the technical supervisor from time to time (see Figure E-137~~Figure E-137~~).

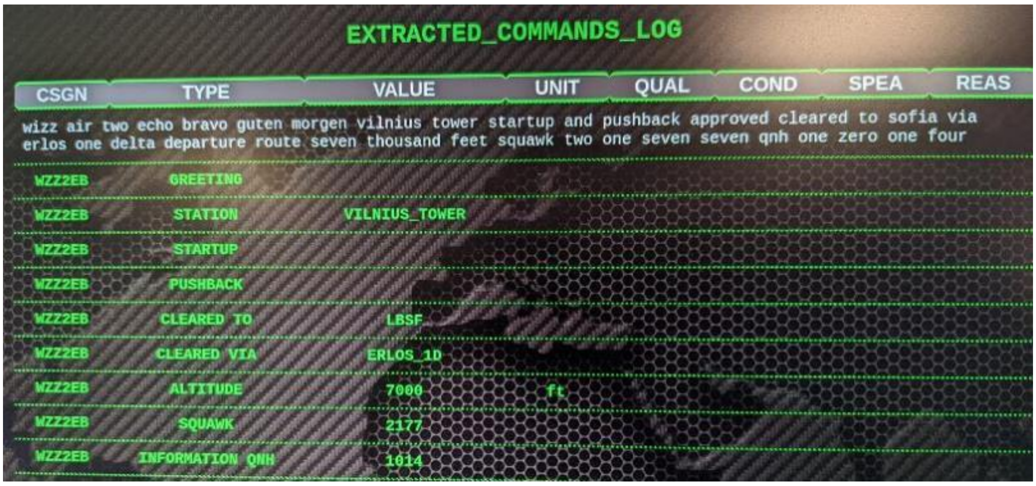


Figure E-137: Output Log for Debugging Purposes of Transcription and Annotation of ATCo utterances

The dedicated ABSR use cases UC-97-TRL4-TS-201 until 204 for solution scenarios have been applied as defined in the TS/IRS document [27]. The extracted callsign has been highlighted in the electronic flight strip (UC-97-TRL4-TS-201) directly after recognizing and extracting even before the ATCo finished the utterance. The command relevant ATC concepts were used to maintain the flight strips, i.e., through highlighting that a startup clearance has been given as extracted from the recognized word sequence. The complete relevant transcription and annotation (full recognized command in agreed ontology format) has been displayed in the outside view of the human machine interface as shown in Figure E-138~~Figure E-138~~ (UC-97-TRL4-TS-202). The ATCo needed to check if the automatically highlighted icons (representing issued commands and thus changes in the a/c status) were correct. The ATCo needed to modify the ABSR output in case of errors (UC-97-TRL4-TS-203). If the ABSR output remained unchanged for 10 seconds, it has been automatically accepted (UC-97-TRL4-TS-204).

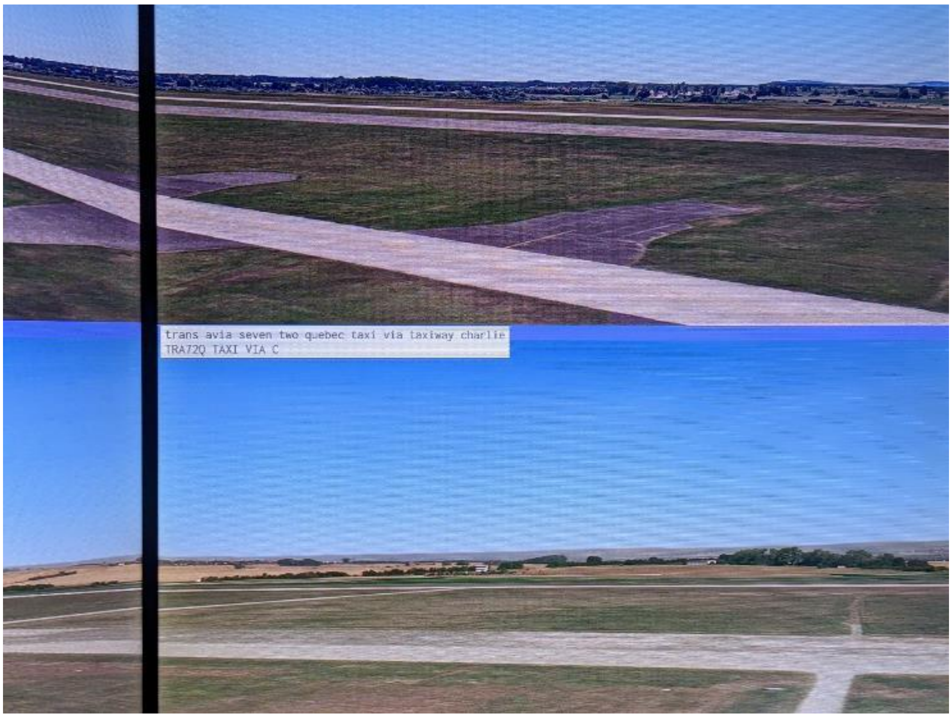


Figure E-138: Automatic transcription and annotation for an ATCo utterance in the outside view

Each validation day with an ATCo began with organizational tasks, a briefing, and a demographics questionnaire. It was followed by 60 minutes training run with low to medium traffic (30 minutes each with reference and solution condition, i.e., without ABSR and with ABSR support). Then, two simulation runs of 60 minutes each with reference and solution condition, respectively, and medium traffic were carried out. One run included a bird strike, the other run included a sick passenger in an a/c as special situations that the ATCos needed to handle and coordinate with ground vehicles. In order to minimize the influence of a learning effect, reference and solution scenario have been alternated for ATCos throughout the validation campaign. After each run, the ATCos needed to fill a questionnaire regarding workload, situation awareness, etc. and give comments and answers in a debriefing. Finally, ATCos needed to fill an overall tailor-made questionnaire on the ABSR system after a final debriefing. The very last task of ATCos was to check some automatic transcriptions of the solution scenario runs that they did earlier – being supported by DLR's transcription and annotation tool. The flow of using speech recognition data in the flight strips can be traced in Figure E-139~~Figure E-139~~.

Figure E-139: ATCo in front of electronic flight strip with highlighted callsign, outside view, ABSR output

Thus, the validation exercise was used to quantify any ATCo productivity enhancements in terms of workload, acceptance, usability, etc. through the advanced support functionalities in the controller working position with automatic flight strip maintenance and highlighting features. The validation campaign took place at DLR TowerLab in Braunschweig, Germany from February 14 to March 3, 2022 with five tower ATCos from Oro Navigacija (Lithuania) and five tower ATCos from AustroControl (Austria) – despite the heavy challenges of pandemic hygiene concept and travelling to Germany in extreme storm conditions. The nine male and one female ATCo had an arithmetic mean age of 31.9 years (standard deviation, SD: 5.5 years). The ATCos had 7.4 years of professional working experience as an ATCo (SD: 5.8 years), while ON ATCos were already longer on duty (9 years, SD: 7.3 years) compared to ACG ATCos (5.7 years, SD: 3.9 years).

E.3 Summary of EXE-006 objectives and success criteria

The following summary presents the table of validation objectives as originally defined in the TVALP [28]. Exercise validation objectives and success criteria are generally identical to the overall objectives and success criteria for Solution 97.2 (ASR), but focus on the tested ABSR prototype.

SESAR Solution Validation Objective	SESAR Solution Success Criteria	Exercise Validation Objective	Exercise Success Criteria

OBJ-05.972-TRL4-TVALP-FEAS.2010 Operational Feasibility	CRT-05.972-TLR4-TVALP-FEAS-2011 No showstoppers related to ASR.	EX6-OBJ-05.972-TRL4-TVALP-FEAS.2010 Idem	EX6-CRT-05.972-TLR4-TVALP-FEAS-2011 Idem
	CRT-05.972-TLR4-TVALP-FEAS-2012 No showstoppers related to AI.		EX6-CRT-05.972-TLR4-TVALP-FEAS-2012 Idem
OBJ-05.972-TRL4-TVALP-FEAS.2020 Technical Feasibility	CRT-05.972-TLR4-TVALP-FEAS-2021 Technical feasibility of ASR.	EX6-OBJ-05.972-TRL4-TVALP-FEAS.2020 Idem	EX6-CRT-05.972-TLR4-TVALP-FEAS-2021 Idem
	CRT-05.972-TLR4-TVALP- FEAS -2022 Technical integration.		EX6-CRT-05.972-TLR4-TVALP- FEAS-2022 Idem
OBJ-05.972-TRL4-TVALP-H106.2010 ATCo Task Support by ASR	CRT-05.972-TLR4-TVALP-H106.2011 Workload maintained at acceptable level.	EX6-OBJ-05.972-TRL4-TVALP-H106.2010 Idem	EX6-CRT-05.972-TLR4-TVALP-H106.2011 Idem
	CRT-05.972-TLR4-TVALP-H106.2012 Situation awareness maintained at adequate level.		EX6-CRT-05.972-TLR4-TVALP-H106.2012 Idem
	CRT-05.972-TLR4-TVALP-H106.2013 Human error not increased.		EX6-CRT-05.972-TLR4-TVALP-H106.2013 Idem
	CRT-05.972-TLR4-TVALP-H106.2014 ASR adequacy.		EX6-CRT-05.972-TLR4-TVALP-H106.2014 Idem
	CRT-05.972-TLR4-TVALP-H106.2015 ASR rates acceptable.		EX6-CRT-05.972-TLR4-TVALP-H106.2015 Idem
	CRT-05.972-TLR4-TVALP-H106.2016 ASR usability.		EX6-CRT-05.972-TLR4-TVALP-H106.2016 Idem

	CRT-05.972-TLR4-TVALP-H106.2017 ASR acceptance.		EX6-CRT-05.972-TLR4-TVALP-H106.2017 Idem
	CRT-05.972-TLR4-TVALP-H106.2018 ASR trust.		EX6-CRT-05.972-TLR4-TVALP-H106.2018 Idem
OBJ-05.972-TLR4-TVALP-H106.2020 ASR impact on ATCo role	CRT-05.972-TLR4-TVALP-H106.2021 Operating methods applied accurately, efficiently and in timely manner.	EX6-OBJ-05.972-TRL4-TVALP-H106.2020 Idem	EX6-CRT-05.972-TLR4-TVALP-H106.2021 Idem
	CRT-05.972-TLR4-TVALP-H106.2022 Operating methods clearly identified and consistent in all operating conditions.		EX6-CRT-05.972-TLR4-TVALP-H105.2022 Idem
OBJ-05.972-TRL4-TVALP-H106.2030 Job Acceptance and Satisfaction	CRT-05.972-TLR4-TVALP-H106.2031 Positive feedback on job satisfaction and acceptance.	EX6-OBJ-05.972-TRL4-TVALP-H106.2030 Idem	EX6-CRT-05.972-TLR4-TVALP-H106.2031 Idem
OBJ-05.972-TLR4-TVALP-SAFE.2010 Safety Impact	CRT-05.972-TLR4-TVALP-SAFE-2011 Improvement of safety performance by reducing human error.	EX6-OBJ-05.972-TLR4-TVALP-SAFE.2010 Idem	EX6-CRT-05.972-TLR4-TVALP-SAFE-2011 Idem
	CRT-05.972-TLR4-TVALP-SAFE-2012 Improvement of safety performance by reducing ATCo workload.		EX6-CRT-05.972-TLR4-TVALP-SAFE-2012 Idem
	CRT-05.972-TLR4-TVALP-SAFE-2013 Improvement of safety performance by increasing SA.		EX6-CRT-05.972-TLR4-TVALP-SAFE-2013 Idem
	CRT-05.972-TLR4-TVALP-SAFE-2014		EX6-CRT-05.972-TLR4-TVALP-SAFE-2014

	Safety assessment activities and results are documented and integrated in overall solution validation results.		Idem
OBJ-05.972-TLR4-TVALP-PERF.2010 Performance Benefits	CRT-05.972-TLR4-TVALP-PERF-2011 Improvement of Cost Efficiency by reducing cost per flight.	EX6-OBJ-05.972-TLR4-TVALP-PERF.2010 Idem	EX6-CRT-05.972-TLR4-TVALP-PERF-2011 Idem

Table E-24: Summary of Validation Objectives addressed in EXE-006

E.4 Summary of EXE-006 validation scenarios

E.4.1 Reference Scenarios

In the reference scenario (also called "baseline"), ATCos had the hardware setup as described in section E.2. After being trained in the reference scenario working condition for 30 minutes, the simulation run with reference scenario took one hour. During the run, tower ATCos needed to control the air traffic at three remote airports using outside view, radar displays, and the electronic flight strip system. All status changes of a/c needed to be documented in the electronic flight strip system using an electronic pen. Thus, the ATCo needed to click on the status icons on the right side of each strip (see Figure E-175) to indicate, e.g., STARTUP, TAXI, CLEARED TO LAND, etc.

Figure E-140: Electronic flight strips in different bays (air, runway, ground, stand)

Every five minutes, the ATCo needed to rate his/her workload on a displayed *instantaneous self-assessment of workload* (ISA) scale from 1 to 5 (Figure E-141~~Figure E-141~~).

Figure E-141: Instantaneous self-assessment of workload (ISA) scale to be responded to

Furthermore, the ATCos were asked to perform a secondary task next to their primary ATC task. After 10 and 40 minutes in the scenario, ATCos needed to sort a deck of 48 cards and name one to four randomly missing cards (Figure E-142~~Figure E-142~~). This sorting of cards was repeated three times each or maximum 15 minutes (after 10 minutes) or 13 minutes (after 40 minutes), respectively. This secondary task shall give a more objective impression about workload when comparing the time needed to sort and identify missing cards between reference and solution scenario.

Figure E-142: ATCo interrupts card sorting as secondary task to check outside view

The air traffic input (valid for both scenarios) comprised 12 flights in Vilnius (plus two ground vehicles), 6 flights in Kaunas (plus one ground vehicle), and 5 in Palanga, so 23 flights plus 3 ground vehicles (11.5% of relevant traffic) in total. For later evaluation, the results refer to all 26 traffic vehicles (flights plus ground vehicles) as ATC communication took place between ATCos and pilots or ground vehicle drivers, respectively.

E.4.2 Solution scenarios

In the solution scenario, ATCos had the same hardware setup as in the reference scenario. Furthermore, the ISA response and secondary task remain identical. The only difference was the support of the ABSR system. ATCos could majorly resign from using the electronic pen to maintain flight strips and benefit from automatic maintenance through the ABSR system, i.e., the ABSR output was used to highlight the status icons and callsigns in electronic flight strips automatically (see Figure E-143~~Figure E-143~~). The ATCos only needed to check the output and correct if needed (as can be seen in this video: https://www.youtube.com/watch?v=Y76kQmo_ANU&cbrd=1).

Figure E-143: Electronic flight strips with highlighted callsign recognized from an ATCo utterance (DLH4TN)

E.5 Summary of EXE-006 assumptions

There were two assumptions for EXE-006 that concerned simulation scope and weather conditions. The primary target was to investigate the effect of machine learning supported speech recognition for tower ATCos in TRL4. Hence, there was no focus on changing runway directions or weather conditions in the multiple remote tower environment to only make the new ATC experience harder for ATCos.

Identifier	Title	Description	Justification	Impact on Assessment
AS-EXE.006-01	Limited Simulation Scope	Simulation focused on the work of one tower ATCo	The choice of having one ATCo carrying out the tower 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
AS-EXE.006-02	Weather Conditions	Good weather conditions throughout the simulation	No impact	Low

Table E-25: EXE-006 assumptions overview

E.6 Deviation from planned activities

There were no deviations from the planned activities.

However, it has to be noted, that the technical team of the validation campaign replaced a laptop and made a software update regarding allowed CPU load for the ASR engine after the eighth ATCo in the simulation campaign.

E.7 EXE-006 validation results

E.7.1 Summary of EXE-006 results

The ABSR technology solution as tested on TRL4 compared to the reference scenario delivered many different positive results and aspects for further refinement. An overview is given through the following table of validation objectives and success criteria.

Technological Validation Exercise #006 Validation Objective ID and Title	Technological Validation Exercise #006 Success Criterion ID and Title	Technological Validation Exercise #006 Results	Technological Validation Exercise #006 Validation Objective Status
EX6-OBJ-05.972-TRL4-TVALP-FEAS.2010 Operational Feasibility	EX6-CRT-05.972-TLR4-TVALP-FEAS-2011 No showstoppers related to ASR.	Almost all ATCos were positive (or very positive) about the potential for support through an ABSR system. No operational showstopper has been identified. Furthermore, the mean value of a questionnaire item regarding ASR show-stoppers was in the acceptable range. The experiment confirmed that the concept is operationally feasible when addressing the ASR use cases. Several suggestions for further improvement especially on the electronic flight strip system (which was not the core on investigation, but the visible part of the human machine interface) were given.	OK
	EX6-CRT-05.972-TLR4-TVALP-FEAS-2012 No showstoppers related to AI.	The mean value of a questionnaire item regarding operational AI show-stoppers was in the acceptable range. However, the AI itself was not transparent to the ATCo while working. The ATCos could just judge the visible effects of the ABSR system (being trained with AI techniques).	OK
EX6-OBJ-05.972-TRL4-TVALP-FEAS.2020	EX6-CRT-05.972-TLR4-TVALP-FEAS-2021	The mean value of a questionnaire item regarding	OK

Technical Feasibility	Technical feasibility of ASR.	technical ASR/AI show-stoppers was in the acceptable range.	
	EX6-CRT-05.972-TLR4-TVALP-FEAS-2022 Technical integration.	The ABSR system was perceived as well integrated. The ABSR system was supporting ATCos throughout all solution runs.	OK
EX6-OBJ-05.972-TRL4-TVALP- H106.2010 ATCo Task Support by ASR	EX6-CRT-05.972-TLR4-TVALP-H106.2011 Workload maintained at acceptable level.	No significant differences in workload were found between reference and solution scenario. While the secondary task (sorting cards) and the ISA tend to show a workload reduction of ATCos when being supported by ASR, NASA-TLX and Bedford Workload Scale tend into the opposite way. However, the statement “I think that ASR supports me in maintaining workload at acceptable level” was rated with 7.8 on a 10 point scale (90% of ATCos rated this item with 7 or above).	OK
	EX6-CRT-05.972-TLR4-TVALP-H106.2012 Situation awareness maintained at adequate level.	No significant differences in SASHA score (situation awareness assessment) were found between reference and solution scenario. The statement “I think that ASR supports me in maintaining an adequate level of situation awareness” was rated with 7.7 on a 10 point scale (90% of ATCos rated this item with 7 or above).	OK
	EX6-CRT-05.972-TLR4-TVALP-H106.2013 Human error not increased.	Due to the ATCos questionnaire ratings, the potential for human errors have not been increased.	OK
	EX6-CRT-05.972-TLR4-TVALP-H106.2014 ASR adequacy.	Accuracy of ASR was rated to be good (callsign highlighting 8.9/10, other values for e.g., commands around 7/10).	OK
	EX6-CRT-05.972-TLR4-TVALP- H106.2015 ASR rates acceptable.	This ASR adequacy result is supported by the analysis of word error rates, callsign recognition error rates, and	OK

		command recognition error rates, and callsign prediction rates. However, it has to be noted, that the system performance was even worse that theoretically possible due to a software problem (see numbers below for online and offline recognition in detail, i.e., 80% command recognition rate compared to 91% possible; 88% command recognition for commands that changed the a/c status in flight strips compared to 93% possible; 92% callsign recognition rate compared to 98% possible).	
	EX6-CRT-05.972-TLR4-TVALP-H106.2016 ASR usability.	The system usability scale (SUS) score was 75 for solution (with ASR) compared to 71 for baseline (without ASR).	OK
	EX6-CRT-05.972-TLR4-TVALP-H106.2017 ASR acceptance.	80% of ATCos stated with 8/10 or more points that they would accept such an ASR system in their normal CWP.	OK
	EX6-CRT-05.972-TLR4-TVALP-H106.2018 ASR trust.	80% of ATCos stated with 6/10 or more points (so above scale mean) that they had trust in the ASR system.	OK
EX6-OBJ-05.972-TRL4-TVALP-H106.2020 ASR impact on ATCo role.	EX6-CRT-05.972-TLR4-TVALP-H106.2021 Operating methods applied accurately, efficiently and in timely manner.	80% of ATCos stated with 8/10 or more points that they could apply operating methods in a timely manner.	OK
	EX6-CRT-05.972-TLR4-TVALP-H106.2022 Operating methods clearly identified and consistent in all operating conditions.	80% of ATCos stated with 6/10 or more points (so above scale mean) that operating methods were clearly identified and consistent in all operating conditions.	OK
EX6-OBJ-05.972-TRL4-TVALP-H106.2030 Job Acceptance and Satisfaction	EX6-CRT-05.972-TLR4-TVALP-H106.2031	The ATCos were satisfied working with the system overall - 100% of ATCos stated this with 6/10 or more points (so above scale mean).	OK

	Positive feedback on job satisfaction and acceptance.		
EX6-OBJ-05.972-TLR4-TVALP-SAFE.2010 Safety Impact	EX6-CRT-05.972-TLR4-TVALP-SAFE-2011 Improvement of safety performance by reducing human error.	See EX6-CRT-05.972-TLR4-TVALP-H106.2013	OK
	EX6-CRT-05.972-TLR4-TVALP-SAFE-2012 Improvement of safety performance by reducing ATCo workload.	See EX6-CRT-05.972-TLR4-TVALP-H106.2011	OK
	EX6-CRT-05.972-TLR4-TVALP-SAFE-2013 Improvement of safety performance by increasing SA.	See EX6-CRT-05.972-TLR4-TVALP-H106.2012	OK
	EX6-CRT-05.972-TLR4-TVALP-SAFE-2014 Safety assessment activities and results are documented and integrated in overall solution validation results.	Safety assessment activities and results are documented and integrated in overall solution validation results.	OK
EX6-OBJ-05.972-TLR4-TVALP-PERF.2010 Performance Benefits	EX6-CRT-05.972-TLR4-TVALP-PERF-2011 Improvement of Cost Efficiency by reducing cost per flight.	A significant reduction of workload or delay times could not be shown in the laboratory trials. Due to the nature of the multiple remote tower setup (no dense traffic at neither of the three airports), an improvement in cost efficiency could not be shown. However, due to the verbal feedback of ATCos, a support of ASR at a tower CWP can be assumed.	POK

Table E-26: EXE-006 results

E.7.1.1 Results on technological feasibility

As summarized in Table E-29, the ABSR system was technologically feasible.

E.7.1.2 Results per KPA

Feasibility

As summarized in Table E-29, the ABSR system was operationally feasible.

Human performance

As summarized in Table E-29, the ABSR system supported to keep human performance at an adequate level (workload, situation awareness, etc.).

Safety

As summarized in Table E-29, safety should not be impacted in any significant way.

E.7.2 Analysis of EXE-006 results Results per Technological Validation objective

The following sections explain details of the results of EXE-006 per Technological Validation Objective. All listed objectives were addressed (see Table E-29). Each of the ten ATCos took part in a baseline run and a solution run (in alternate order to control for learning effects), i.e., twenty simulation runs are analysed in the following.

E.7.2.1 Questionnaire types

Three types of questionnaires have been used per ATCo on each validation day.

- Pre-Run Questionnaire (executed once in the morning) containing
  - Demographics (Age, Professional Experience, Gender)
- Post-Run Questionnaire (executed twice, i.e., identical procedure after baseline runs and solution runs) containing three questionnaires from SHAPE project (Solutions for Human Automation Partnerships in European Air Traffic Management) and other well-established questionnaires
  - NASA-TLX
  - Bedford Workload Scale and Workload Explanation
  - SASHA ATCo (Situational Awareness for SHAPE)
  - AIM (Assessing the Impact of Automation on Mental Workload)
  - SATI (SHAPE Automation Trust Index)
  - SUS (System Usability Score)
  - CARS (Controller Acceptance Rating Scale)
- Post-Validation Questionnaire (executed once in the afternoon after all runs) containing
  - 28 statements to be rated regarding validation objectives for safety, human performance, cost efficiency, and technical feasibility

If answers on the post-validation questionnaire of the ten ATCos are reported in the following, the scale ranges from 1 (fully disagree) to 10 (fully agree), i.e., the scale mean is 5.5. The short version legend identifier is given in square brackets after the statement that ATCos needed to rate [Example Short ID].

For many of the analyses, a differentiation between the five Lithuanian (ON) and the five Austrian (ACG) ATCos is made as there are some deviations worth to report compared to all ATCos (all). Furthermore, the results might in addition be differentiated for the male and the female ATCo.

E.7.2.2 ABSR output analysis

Verbal utterances of ATCos that were triggered with the push-to-talk button during twenty hours of simulations runs (radar data duration) have been recorded as wav-files. For each wav-file of the ten times two simulation runs (baseline and solution) exists an automatic transcription and an automatic annotation. All wav-files with a net speech duration of almost 4.5h (average duration per wav file between 6 and 7 s) have been manually transcribed and annotated ("gold") with DLR's Controller Command Logging Tool for Context Comparison (CoCoLoToCoCo) to enable comparison and calculations about recognition and error rates on word level and semantic level (see Figure E-144~~Figure E-144~~).

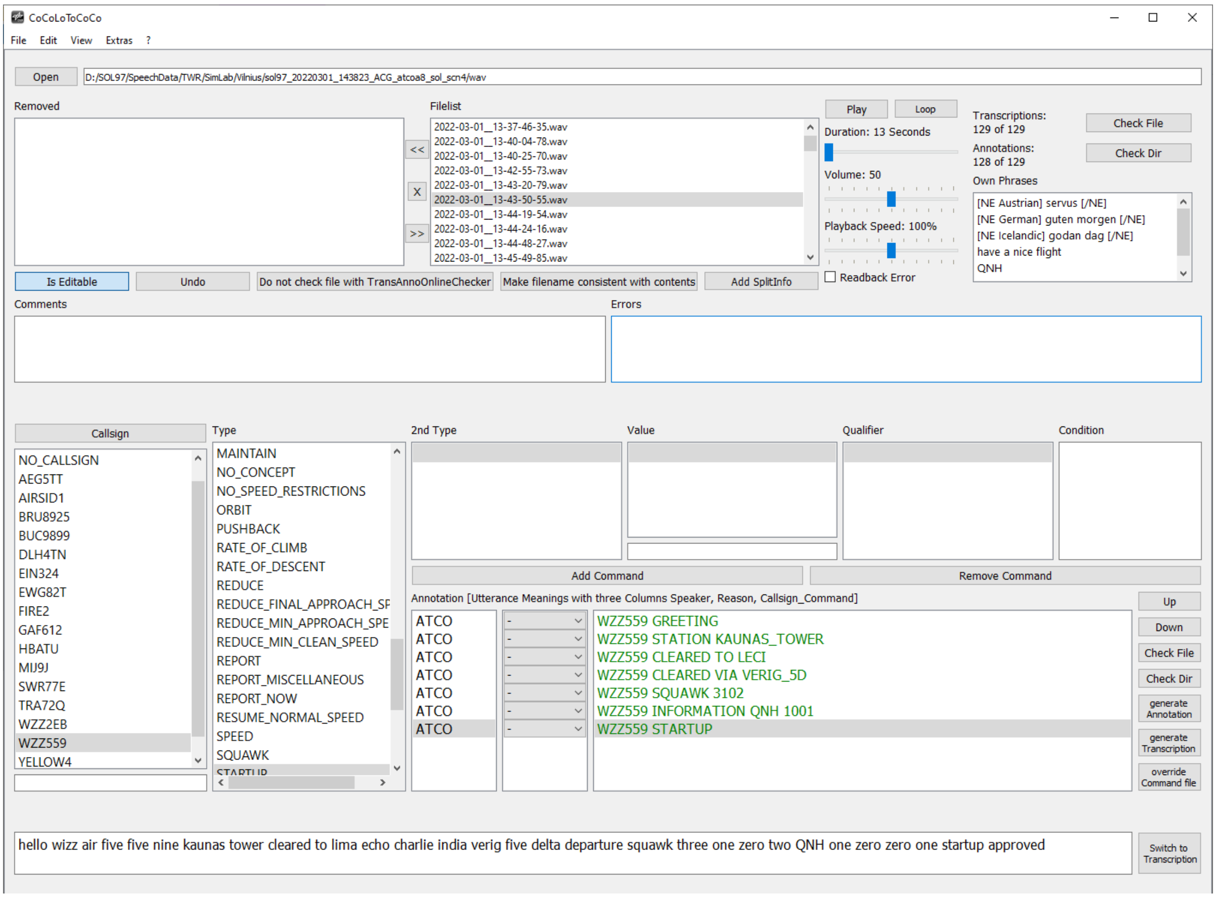


Figure E-144: Software tool CoCoLoToCoCo to support transcription and annotation of ATC utterances

Some abbreviation used in the following shall be introduced:

- cmds = commands
- Onl = online (as experienced by ATCos)
- Off = offline (analysis of audio files afterwards)
- OffBst = offline with callsign boosting
- Trans = Transcriptions
- WER = Word Error Rate

E.7.2.3 Speech-To-Text quality

The following table shows the speech-to-text quality for OffBst mode.

Speech-To-Text Accuracy	Offline recogn. with callsign boosting on audio files					
Directory	# Words	LevenDist	# Subst	# Del	# Ins	% WER
0214_ON_atcol1_sol_scn3	1908	25	7	10	8	1.3
0214_ON_atcol1_bas_scn4	1874	104	43	51	10	5.5
0215_ON_atcol2_bas_scn3	2185	81	24	44	13	3.7
0215_ON_atcol2_sol_scn4	1989	93	41	35	17	4.7
0216_ON_atcol3_sol_scn3	1920	55	20	10	25	2.9
0216_ON_atcol3_bas_scn4	1977	70	30	6	34	3.5
0217_ON_atcol4_bas_scn3	1369	157	62	78	17	11.5
0217_ON_atcol4_sol_scn4	2028	159	74	71	14	7.8
0218_ON_atcol5_sol_scn3	2359	109	37	23	49	4.6
0218_ON_atcol5_bas_scn4	2052	86	44	34	8	4.2
0224_ACG_atcoa6_bas_scn3	1944	151	72	62	17	7.8
0224_ACG_atcoa6_sol_scn4	1978	97	49	35	13	4.9
0228_ACG_atcoa7_sol_scn3	1858	58	22	26	10	3.1
0228_ACG_atcoa7_bas_scn4	1897	91	45	34	12	4.8
0301_ACG_atcoa8_bas_scn3	1829	119	46	65	8	6.5
0301_ACG_atcoa8_sol_scn4	1898	116	55	51	10	6.1
0302_ACG_atcoa9_sol_scn3	1802	69	39	19	11	3.8
0302_ACG_atcoa9_bas_scn4	1827	60	34	16	10	3.3
0303_ACG_atcoa10_bas_scn3	2068	123	59	42	22	5.9
0303_ACG_atcoa10_sol_scn4	2109	108	56	40	12	5.1
MEAN all	1944	97	43	38	16	**5.1**
MEAN ON	1966	94	38	36	20	**5.0**
MEAN ACG	1921	99	48	39	13	**5.1**
MEAN male	1971	90	40	34	16	**4.5**
MEAN BAS all	1902	104	46	43	15	**5.7**
MEAN BAS ON	1891	100	41	43	16	**5.7**
MEAN BAS ACG	1913	109	51	44	14	**5.7**
MEAN BAS male	1961	98	44	39	15	**5.0**
MEAN SOL all	1985	89	40	32	17	**4.4**
MEAN SOL ON	2041	88	36	30	23	**4.3**
MEAN SOL ACG	1929	90	44	34	11	**4.6**
MEAN SOL male	1980	81	36	28	17	**4.1**

There were some technical problems with the ASR engine in the baseline run of the female ATCo (loss of data). The low amount of training data with female ATCo voices led to a worse performance of the ABSR system in the validation runs (atcoa4) compared to the male ATCos.

The following table shows the speech-to-text quality for Off mode.

Speech-To-Text Accuracy		Offline recogn. on audio files				
Directory	# Words	LevenDist	# Subst	# Del	# Ins	% WER
0214_ON_atcol1_sol_scn3	1908	27	8	10	9	1.4
0214_ON_atcol1_bas_scn4	1874	111	40	61	10	5.9
0215_ON_atcol2_bas_scn3	2185	82	24	45	13	3.8
0215_ON_atcol2_sol_scn4	1989	93	41	35	17	4.7
0216_ON_atcol3_sol_scn3	1920	59	20	13	26	3.1
0216_ON_atcol3_bas_scn4	1977	67	29	6	32	3.4
0217_ON_atcol4_bas_scn3	1369	172	65	91	16	12.6
0217_ON_atcol4_sol_scn4	2028	166	78	78	10	8.2
0218_ON_atcol5_sol_scn3	2359	111	37	27	47	4.7
0218_ON_atcol5_bas_scn4	2052	94	45	40	9	4.6
0224_ACG_atcoa6_bas_scn3	1944	155	72	72	11	8.0
0224_ACG_atcoa6_sol_scn4	1978	96	45	39	12	4.9
0228_ACG_atcoa7_sol_scn3	1858	64	23	32	9	3.4
0228_ACG_atcoa7_bas_scn4	1897	92	45	34	13	4.8
0301_ACG_atcoa8_bas_scn3	1829	121	45	68	8	6.6
0301_ACG_atcoa8_sol_scn4	1898	111	52	53	6	5.8
0302_ACG_atcoa9_sol_scn3	1802	71	38	23	10	3.9
0302_ACG_atcoa9_bas_scn4	1827	58	32	17	9	3.2
0303_ACG_atcoa10_bas_scn3	2068	133	61	53	19	6.4
0303_ACG_atcoa10_sol_scn4	2109	114	55	47	12	5.4
MEAN all	1944	100	43	42	15	5.2
MEAN ON	1966	98	39	41	19	5.2
MEAN ACG	1921	102	47	44	11	5.3
MEAN male	1971	92	40	38	15	4.7
MEAN BAS all	1902	109	46	49	14	5.9
MEAN BAS ON	1891	105	41	49	16	6.0
MEAN BAS ACG	1913	112	51	49	12	5.8
MEAN BAS male	1961	101	44	44	14	5.2
MEAN SOL all	1985	91	40	36	16	4.6
MEAN SOL ON	2041	91	37	33	22	4.4
MEAN SOL ACG	1929	91	43	39	10	4.7
MEAN SOL male	1980	83	35	31	16	4.2

The following table shows the speech-to-text quality for Onl mode.

Speech-To-Text Accuracy **Online recogn. with callsign boosting from stream**

Directory	# Words	LevenDist	# Subst	# Del	# Ins	% WER
0214_ON_atcol1_sol_scn3	1906	53	12	26	15	2.8
0214_ON_atcol1_bas_scn4	1869	232	33	190	9	12.4
0215_ON_atcol2_bas_scn3	2183	193	25	147	21	8.8
0215_ON_atcol2_sol_scn4	1976	175	24	145	6	8.9
0216_ON_atcol3_sol_scn3	1920	40	18	19	3	2.1
0216_ON_atcol3_bas_scn4	1977	50	21	20	9	2.5
0217_ON_atcol4_bas_scn3	1281	717	93	558	66	56.0
0217_ON_atcol4_sol_scn4	2022	356	92	196	68	17.6
0218_ON_atcol5_sol_scn3	2359	58	23	31	4	2.5
0218_ON_atcol5_bas_scn4	2044	114	38	67	9	5.6
0224_ACG_atcoa6_bas_scn3	1944	579	102	455	22	29.8
0224_ACG_atcoa6_sol_scn4	1978	133	43	80	10	6.7
0228_ACG_atcoa7_sol_scn3	1837	400	20	332	48	21.8
0228_ACG_atcoa7_bas_scn4	1889	364	77	242	45	19.3
0301_ACG_atcoa8_bas_scn3	1829	479	60	349	70	26.2
0301_ACG_atcoa8_sol_scn4	1896	428	57	340	31	22.6
0302_ACG_atcoa9_sol_scn3	1801	110	39	67	4	6.1
0302_ACG_atcoa9_bas_scn4	1827	114	42	67	5	6.2
0303_ACG_atcoa10_bas_scn3	2067	157	50	90	17	7.6
0303_ACG_atcoa10_sol_scn4	2109	140	52	71	17	6.6
MEAN all	1936	245	46	175	24	**13.6**
MEAN ON	1954	199	38	140	21	**11.9**
MEAN ACG	1918	290	54	209	27	**15.3**
MEAN male	1967	212	41	152	19	**11.0**
MEAN BAS all	1891	300	54	219	27	**17.4**
MEAN BAS ON	1871	261	42	196	23	**17.1**
MEAN BAS ACG	1911	339	66	241	32	**17.8**
MEAN BAS male	1959	254	50	181	23	**13.2**
MEAN SOL all	1980	189	38	131	21	**9.8**
MEAN SOL ON	2037	136	34	83	19	**6.8**
MEAN SOL ACG	1924	242	42	178	22	**12.8**
MEAN SOL male	1976	171	32	123	15	**8.9**

The color highlighting has the following meaning (emphasizes some interesting results):

1.3	Training data already contained speech samples from this ATCo
2.9	Offline WER worse than online WER
11.5	Highest WER due to audio device software problem
7.8	Still highest WER
3.1	Fourth highest WER online; lowest WER offline

We recorded 2,437 wav files with net speech time of 16,114s (4.48h net talking during 20h radar simulation ~22%) in twenty simulation runs. Online results summary:

- four of twenty runs with WER <3%
- twelve of twenty runs with WER <9%
- three of twenty runs with WER >23% (technical problems affected those)

Two technical problems: One with audio device continuously disconnecting for atcol4_bas and partly CPU overload for at least the first eight ATCos (maybe even for all). Offline results summary:

- all runs <8% WER.

E.7.2.4 Text-To-Concept quality

The callsign prediction error rate was 0.09% (0.13% for baseline; 0.05% for solution).

The following subsections present recognition and error rates on callsign and command level as well as the portion of words from the utterances (Unknown Classified Rate) that have not been used for ATC concept extraction (i.e., neither for callsigns or command parts).

Furthermore, the rates are shown on more detailed level for:

- Baseline/solution runs
- ON/ACG ATCos
- Male/female ATCos
- Command types (“All”; “Relevant” if appearing more than 25 times in all 20 runs; “EFS” having a visible effect in the electronic flight strips; “Status” that changed the a/c status in the electronic flight strips; “Outside” as just be shown on the monitors for the outside view; “Hypo-EFS” that could have been highlighted in the flight strips, but have not been during the trials such as recognizing the active runway in an utterance).

The following tables show the text-to-concept quality for Gold transcriptions (assumed to be 100% correct).

all 20 runs 7,560 cmds Gold Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	95.9%	2.4%	99.8%	0.2%	13.3%	100.0%
ON ATCos	97.1%	1.5%	99.7%	0.2%	12.5%	49.9%
ACG ATCos	94.8%	3.2%	99.9%	0.1%	14.2%	50.1%
male ATCos	95.8%	2.5%	99.8%	0.2%	13.2%	91.8%
female ATCos	97.6%	1.1%	99.8%	0.2%	13.4%	8.2%

10 BAS runs 3,701 cmds Gold Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	95.9%	2.4%	99.7%	0.3%	13.8%	49.0%
ON ATCos	97.6%	1.3%	99.7%	0.3%	13.0%	24.1%
ACG ATCos	94.1%	3.5%	99.8%	0.2%	14.7%	24.8%

10 SOL runs 3,859 cmds Gold Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	96.0%	2.3%	99.8%	0.1%	12.8%	51.0%
ON ATCos	96.6%	1.8%	99.7%	0.2%	12.0%	25.8%
ACG ATCos	95.4%	2.9%	100.0%	0.0%	13.7%	25.3%

18.3% of all problematic annotations go back to the three ground vehicles that just make 11.5% of all relevant traffic. Further 7.3% of problematic annotations go back to the emergency a/c even if this just makes 3.8% of the flights.

Command Type Group	# Command Types	Command Recognition Rate	Command Error Rate
All	63	97.3%	2.0%
Relevant	34	97.8%	1.5%
EFS	21	97.6%	1.4%
Status	18	96.9%	1.9%
Outside	3	99.7%	0.1%
Hypo-EFS	4	98.3%	1.6%

The following tables show the text-to-concept quality for OffBst transcriptions (current best word error rates of automatic speech-to-text with callsign boosting).

all 20 runs 7,560 cmds Off Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	91.4%	4.5%	98.4%	0.9%	14.0%	100.0%
ON ATCos	92.7%	3.9%	98.6%	0.6%	12.8%	49.9%
ACG ATCos	90.1%	5.1%	98.2%	1.2%	15.2%	50.1%
male ATCos	91.7%	4.4%	98.7%	0.9%	13.9%	91.8%

10 BAS runs 3,701 cmds Off Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	91.0%	4.6%	98.6%	0.8%	14.5%	49.0%
ON ATCos	92.8%	3.6%	99.0%	0.3%	13.2%	24.1%
ACG ATCos	89.3%	5.5%	98.1%	1.2%	15.8%	24.8%

10 SOL runs 3,859 cmds Off Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	91.8%	4.5%	98.2%	1.1%	13.6%	51.0%
ON ATCos	92.7%	4.1%	98.1%	0.9%	12.6%	25.8%
ACG ATCos	90.9%	4.8%	98.3%	1.2%	14.6%	25.3%

Command Type Group	# Command Types	Command Recognition Rate	Command Error Rate
All	62	94.1%	4.0%
Relevant	31	94.6%	3.5%
EFS	21	94.1%	4.2%
Status	18	93.0%	4.8%
Outside	3	97.5%	1.9%
Hypo-EFS	4	96.3%	2.8%

Furthermore, we calculated the recognition rates for instruction sub-parts that are (not to be mixed up with the command recognition rate where all sub-parts need to be correct):

- command type recognition rate: 94%
- command second type recognition rate: 94.7%
- value recognition rate: 94.1%
- value (further part) recognition rate: 96%
- unit recognition rate: 96.1%
- qualifier recognition rate: 95.7%
- condition recognition rate: 96%
- callsign recognition rate (per command): 95.4% (if there are multiple commands in an utterance, the callsign is counted as wrong for each of the commands; the above reported callsign recognition rate per utterance of 98% for all ATCos is only calculated once per utterance independent of the number of commands)

Those eight numbers lead to the overall command recognition rate of slightly more than 91% (see table above). However, those numbers are of limited expression as the command extraction algorithm does not independently extract the instruction sub-parts, but extracts sub-parts in case hints for certain command types are found. Nevertheless, for some ABSR applications, the numbers for sub-part extraction might be of interest . The following tables show the text-to-concept quality for Off transcriptions (offline recognition from recorded audio files).

all 20 runs 7,560 cmds OffBst Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	91.0%	4.6%	98.0%	0.9%	14.0%	100.0%
ON ATCos	92.6%	3.8%	98.5%	0.5%	12.9%	49.9%
ACG ATCos	89.5%	5.3%	97.4%	1.3%	15.1%	50.1%
male ATCos	91.3%	4.5%	98.3%	0.8%	13.8%	91.8%

10 BAS runs 3,701 cmds OffBst Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	90.5%	4.5%	97.8%	0.8%	14.4%	49.0%
ON ATCos	92.5%	3.5%	98.5%	0.3%	13.3%	24.1%
ACG ATCos	88.6%	5.5%	97.1%	1.4%	15.6%	24.8%

10 SOL runs 3,859 cmds OffBst Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	91.6%	4.6%	98.1%	0.9%	13.6%	51.0%
ON ATCos	92.7%	4.2%	98.4%	0.6%	12.5%	25.8%
ACG ATCos	90.4%	5.1%	97.8%	1.2%	14.7%	25.3%

16.2% of all problematic annotations go back to the three ground vehicles that just make 11.5% of all relevant traffic.

Command Type Group	# Command Types	Command Recognition Rate	Command Error Rate
All	62	93.7%	4.1%
Relevant	31	94.2%	3.5%
EFS	21	93.9%	4.3%
Status	18	92.7%	4.9%
Outside	3	96.8%	2.2%
Hypo-EFS	4	96.0%	2.8%

The following tables show the text-to-concept quality for Onl transcriptions (online recognition as "experienced" by ATCos during simulation runs including technical issues).

all 20 runs 7,560 cmds Onl Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	79.4%	7.0%	91.7%	3.1%	15.4%	100.0%
ON ATCos	84.2%	5.5%	92.1%	2.4%	13.8%	49.9%
ACG ATCos	74.6%	8.6%	91.3%	3.9%	17.0%	50.1%
male ATCos	81.2%	6.6%	94.0%	2.5%	14.9%	91.8%

10 BAS runs 3,701 cmds Onl Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	75.7%	7.5%	89.1%	3.8%	16.2%	49.0%
ON ATCos	80.1%	5.6%	88.9%	2.8%	14.6%	24.1%
ACG ATCos	71.4%	9.3%	89.3%	4.8%	17.9%	24.8%

10 SOL runs 3,859 cmds Onl Trans	Command Recognition Rate	Command Error Rate	Callsign Recognition Rate	Callsign Error Rate	Unknown Classified Rate	Amount of Data
all ATCos	82.9%	6.6%	94.2%	2.4%	14.5%	51.0%
ON ATCos	88.0%	5.4%	95.2%	2.0%	13.2%	25.8%
ACG ATCos	77.7%	7.9%	93.2%	2.9%	16.1%	25.3%

Command Type Group	# Command Types	Command Recognition Rate	Command Error Rate
All	61	87.7%	6.5%
Relevant	31	88.3%	6.0%
EFS	21	89.4%	6.4%
Status	18	88.2%	7.2%
Outside	3	93.8%	3.9%
Hypo-EFS	4	89.7%	6.5%

E.7.2.5 EX6-OBJ-05.972-TRL4-TVALP-FEAS.2010

Operational use of ASR

The post-validation questionnaire contained two statements about operational feasibility with the ASR system:

1) *Procedures and operating methods are acceptable when using the ASR tool. [ProcOKwASR]*
2) *There are no changes needed to current working methods/procedures to fully support the use of ASR tool. [NoChgNeed]*
3) *The ASR tool would be operationally acceptable under either nominal or non-nominal conditions. [OpAccAllCond]*

The results are shown in Figure E-145~~Figure E-145~~. Procedures and operating methods seem to be completely ok with a mean value of 8.5 and a standard deviation of only 1.0. There are probably some changes of current working methods needed to fully support the use of ASR tool as the mean value equals the scale mean value of 5.5. However, ON ATCos rated this statement with almost 7, while ACG ATCos rated with slightly above 4 points. The ASR seems to be operationally acceptable under different conditions, most probably under the majority of nominal and a few non-nominal conditions as the ATCo rating was just slightly beyond the scale mean value.

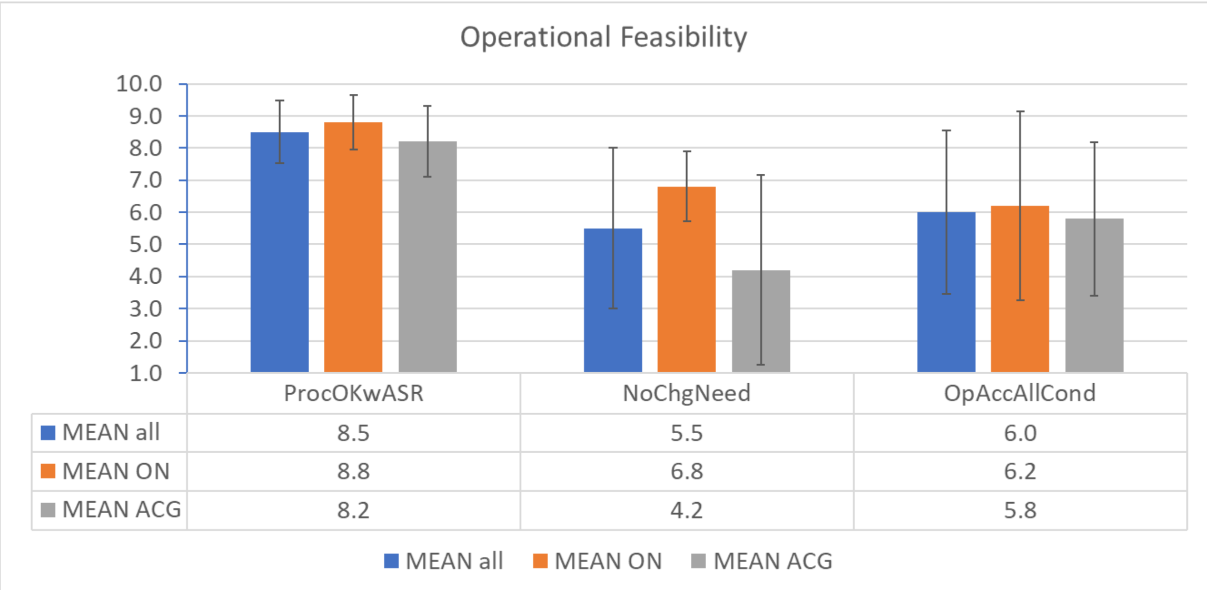


Figure E-145: ATCo ratings on operational feasibility

Operational show-stoppers

The post-validation questionnaire contained two statements about operational show-stoppers regarding ASR and regarding AI, respectively:

4) *With Automatic Speech Recognition (ASR) I found operational show-stoppers regarding used speech recognition applications. [ASR]*

5) With Automatic Speech Recognition (ASR) I found operational show-stoppers regarding artificial intelligence used to build the speech recognition system. [AI]

The results are shown in Figure E-181. The average result (below the red dotted scale mean line) indicates that there were rather no show-stoppers found during the simulation runs. However, the answers were very inhomogeneous as the standard deviation of 2.8 (for both values of "all") represent. In tendency, the ATCos that experienced worse recognition rates of ABSR output, also rated the statements for show-stoppers higher, e.g., when just analysing the male ATCos for which ABSR performed better, the average drops even further to the acceptable range below 5.5.

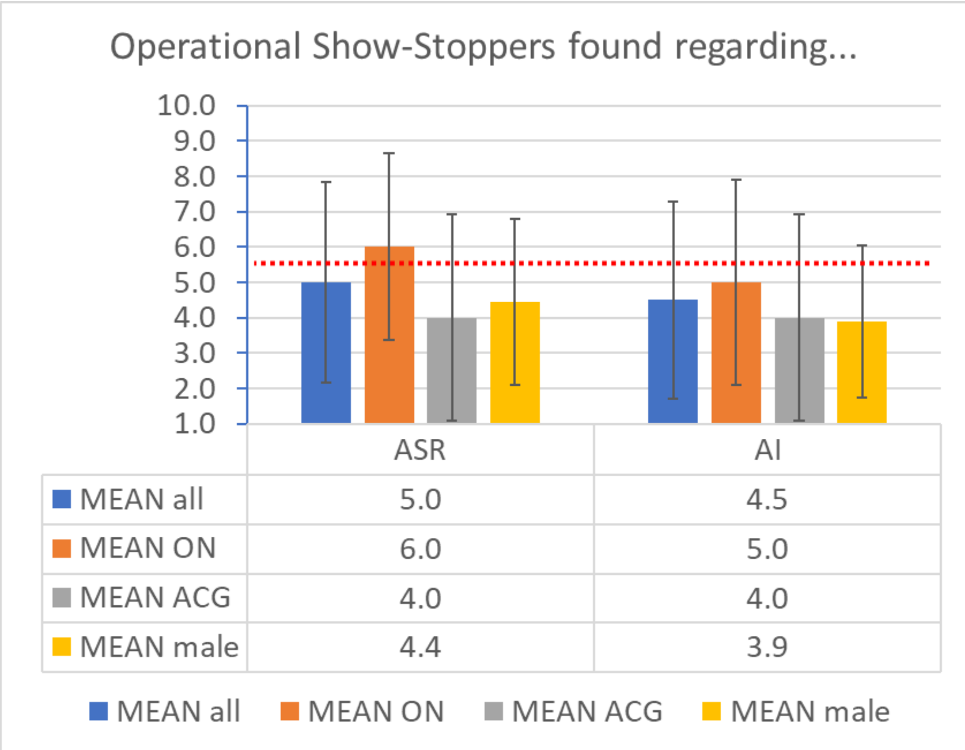


Figure E-146: ATCo ratings on operational show-stoppers

System usability

The post-run questionnaire contained the ten statements of the System Usability Scale (SUS). The results[3] are shown in Figure E-147~~Figure E-147~~. Considering all ATCos, the SUS score was 4 points (5.65%) higher in the solution condition (SOL) with ASR support compared to the baseline condition (BAS) without ASR support. The difference of 4 points remains when just analysing ON score or ACG score independently. However, the score itself is 14.5 points higher for ON than for ACG.

This is probably due to the fact that ON really liked the electronic flight strip display (also in the baseline version) whereas ACG ATCos needed to adapt themselves more to the strip system due to the difference to their daily-life system.

While the standard deviation for all ATCos in BAS condition was 16.3 with 15.9 for ON and 16.6 for ACG, the standard deviation was 16.9 in SOL condition with 9.7 for ON and 20.5 for ACG, respectively.

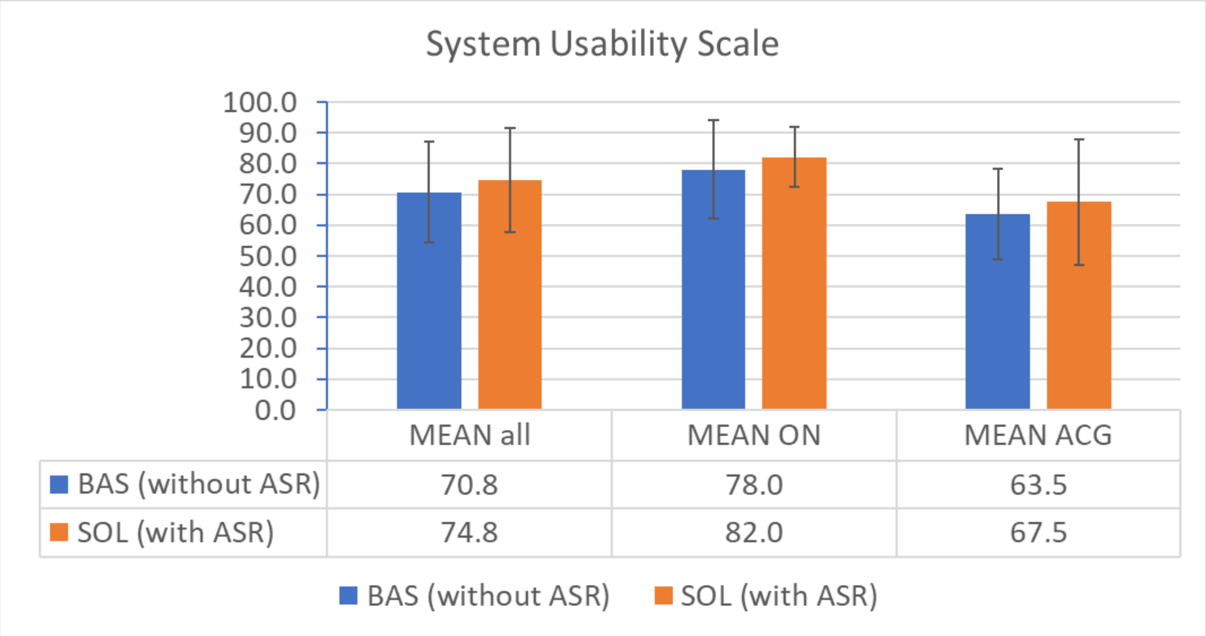


Figure E-147: ATCo ratings on system usability

[3] One of the 100 items remained unanswered both in baseline (without ASR) and solution (with ASR) condition. Therefore, the scale mean "3" ((5-1)/2) was chosen as a replacement to not heavily influence the overall result.

E.7.2.6 EX6-OBJ-05.972-TRL4-TVALP-FEAS.2020

ASR accuracy

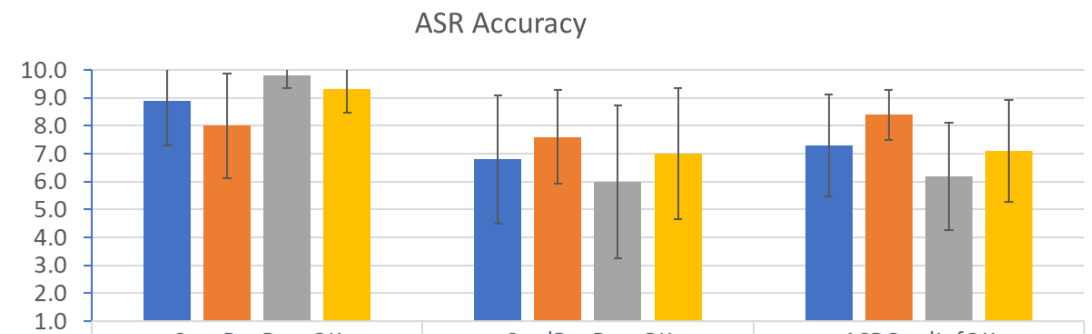
The post-validation questionnaire contained three statements about technical feasibility with respect to recognition and error rate of callsigns and commands:

1) *The recognition rate and recognition error rates for callsigns by ASR were at an acceptable level. [CsgnRecRateOK]*

2) *The recognition rates and recognition error rates for commands by ASR were at an acceptable level. [CmdRecRateOK]*

3) *Overall, the level and quality of information provided by ASR were an acceptable level. [ASRQualInfOK]*

The results are shown in Figure Figure E-148~~Figure E-148~~. ATCos rated the recognition of callsigns as almost perfect with a mean value around 9. Considering only male ATCos or only the ACG ATCos leads to mean ratings of close to scale maximum value 10. The recognition rates of ATC commands were also perceived as good with a mean vale around 7. The general quality level of information presentation from ASR was also rated to be at an acceptable level with a mean value of slightly beyond 7. It has to be noted that the command recognition and overall ASR information displayed were rated much higher from ON than from ACG ATCos.



	CsgnRecRateOK	CmdRecRateOK	ASRQualInfOK
MEAN all	8.9	6.8	7.3
MEAN ON	8.0	7.6	8.4
MEAN ACG	9.8	6.0	6.2
MEAN male	9.3	7.0	7.1

MEAN all MEAN ON MEAN ACG MEAN male

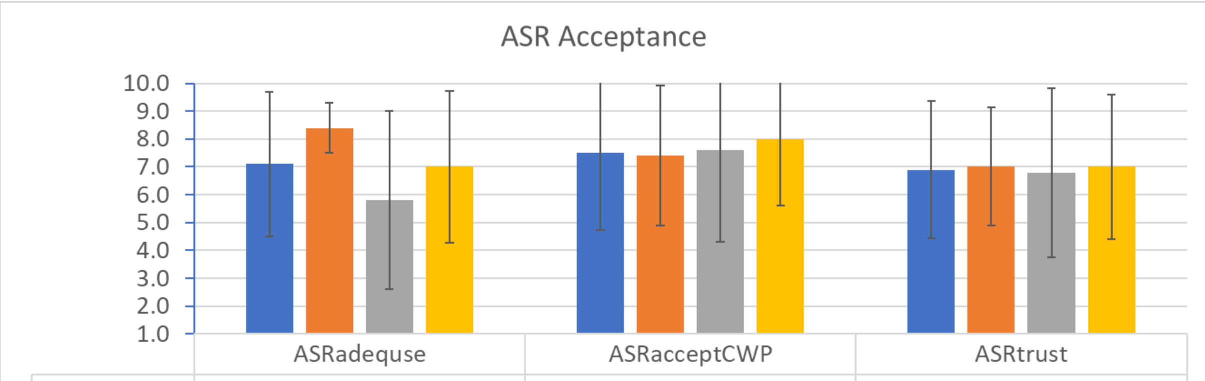
Figure E-148: ATCo ratings on technical ASR accuracy

ASR acceptance

The post-validation questionnaire contained three statements about acceptance and trust of ASR system:

1) *I think that the ASR system is adequately usable. [ASRadequse]*
2) *I would accept such an ASR system in my future tower CWP. [ASRacceptCWP]*
3) *My trust in the ASR system is at an acceptable level. [ASRtrust]*

The results are shown in Figure E-149~~Figure E-149~~. ATCos rated the adequate usage of ASR with a mean value around 7. However, it has to be noted that it was rated much higher from ON than from ACG ATCos. All ATCos would accept such an ASR system in their future tower CWP with a mean value of 7.5 (even 8 if only considering male ATCos). They also trusted the ASR system with a mean value around 7.



	ASRadequse	ASRacceptCWP	ASRtrust
MEAN all	7.1	7.5	6.9
MEAN ON	8.4	7.4	7.0
MEAN ACG	5.8	7.6	6.8
MEAN male	7.0	8.0	7.0

Figure E-149: ATCo ratings on technical ASR acceptance

ASR interface

The post-validation questionnaire contained four statements about the ASR interface:

1) *The ASR tool interface (HMI) provides suitable access to relevant information in all situations. [ASRrelevInfo]*
2) *The ASR tool interface (HMI) does not display any non-essential information (clutter). [ASRessentInfo]*
3) *The ASR tool display is both comprehensible and acceptable. [ASRcomprehaccep]*
4) *The timeliness of the ASR tool display is within acceptable limits. [ASRtimeliness]*

The results are shown in Figure E-150~~Figure E-150~~. Relevant information of the ABSR system can be assessed (mean value 7.4, but more than 1.5 points rated higher by ON than by ACG). The ASR tool seems to only present essential information with a mean value of 8.2 (again ON rated almost 1.5 points higher than ACG). The ASR visualization is perceived as comprehensible with a mean value of 7.7 (again ON rated almost 2 points higher than ACG). Finally, the output of the ABSR system was shown quite timely (mean value 7.5) due to the ATCo feedback.

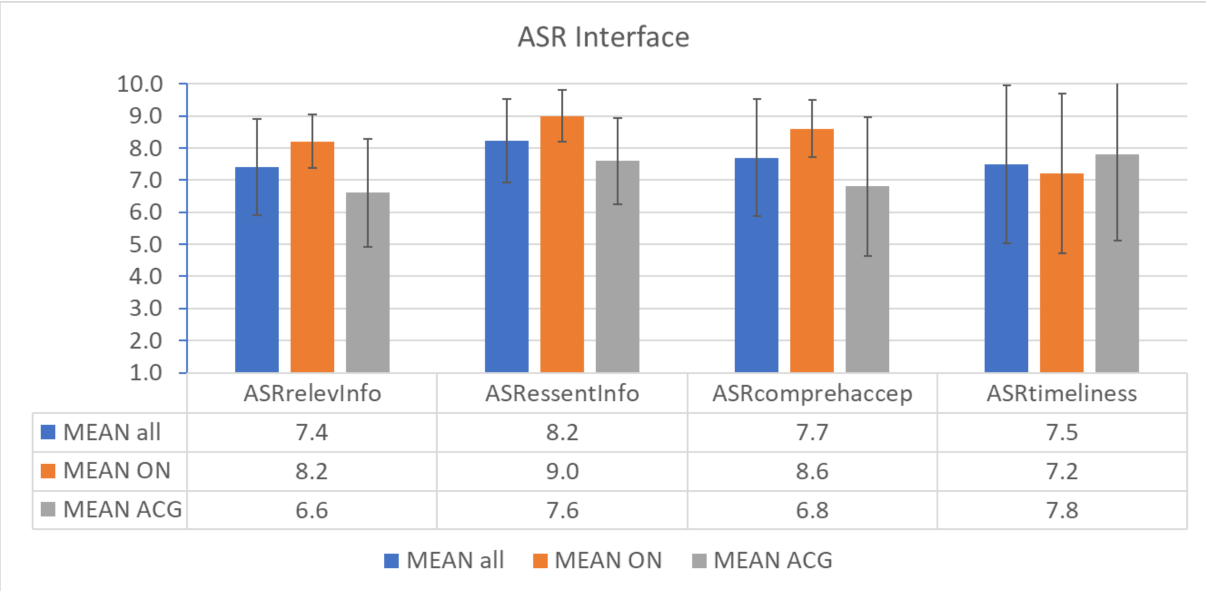


Figure E-150: ATCo ratings on ASR interface

Callsign highlighting

The post-validation questionnaire contained two statements about technical feasibility of callsign highlighting:

5) *With Automatic Speech Recognition (ASR) highlighting of a/c callsigns in the electronic flight strip display technically worked well. [HIGHL-CSGN]*

6) *With Automatic Speech Recognition (ASR) highlighting of a/c callsigns in the electronic flight strip display supported me in recognizing which a/c callsign has been (speech) recognized quickly. [RECOG-CSGN]*

The results are shown in Figure E-151~~Figure E-151~~. The highlighting of callsigns in the electronic flight strip display was perceived as working technically very well with a mean of 9.7 on a 10-point scale and a low standard deviation of 0.5. This functionality also helped the ATCos to recognize which a/c callsign has been recognized by the ABSR system with a mean value of 8.1, i.e., where all the following recognized ATC commands will be highlighted.

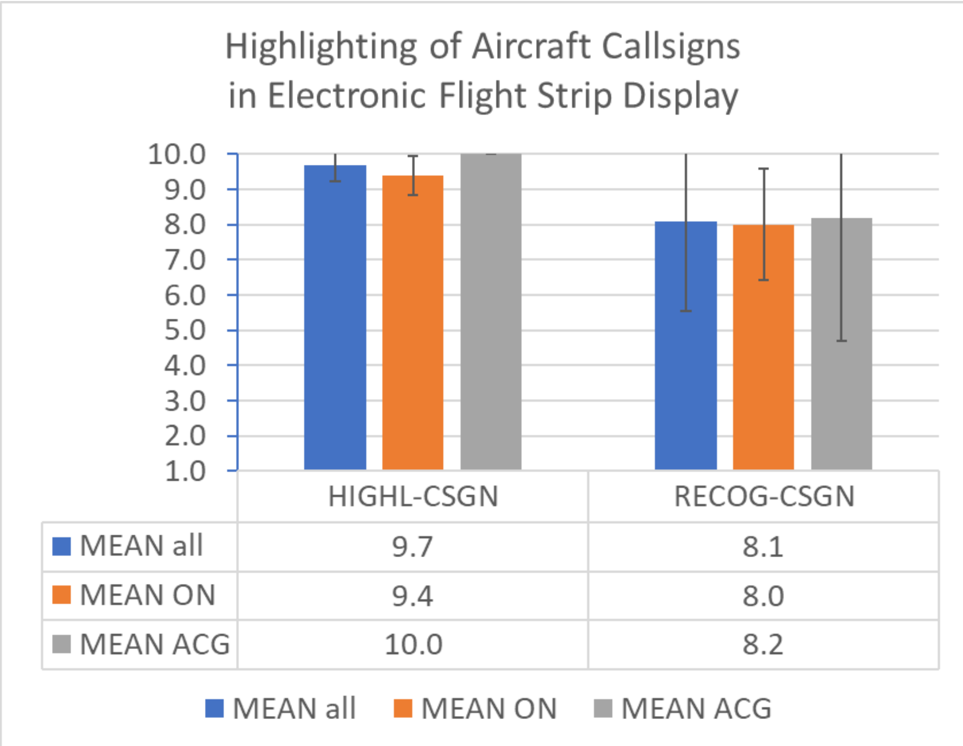


Figure E-151: ATCo ratings on callsign highlighting

Controller Acceptance Rating Scale (CARS)

The post-run questionnaires contained the CARS statement:

"Please read the descriptors and score your overall level of user acceptance experienced during the run. Please check the appropriate number.

1) Improvement mandatory. Safe operation could not be maintained.
2) Major Deficiencies. Safety not compromised, but system is barely controllable and only with extreme controller compensation.
3) Major Deficiencies. Safety not compromised but system is marginally controllable. Considerable compensation is needed by the controller.
4) Major Deficiencies. System is controllable. Some compensation is needed to maintain safe operations.
5) Very Objectionable Deficiencies. Maintaining adequate performance requires extensive controller compensation.
6) Moderately Objectionable Deficiencies. Considerable controller compensation to achieve adequate performance.
7) Minor but Annoying Deficiencies. Desired performance requires moderate controller compensation.
8) Mildly unpleasant Deficiencies. System is acceptable and minimal compensation is needed to meet desired performance.
9) Negligible Deficiencies. System is acceptable and compensation is not a factor to achieve desired performance.
10) Deficiencies are rare. System is acceptable and controller doesn't have to compensate to achieve desired performance."

The results of the CARS questionnaire are shown in Figure E-152~~Figure E-152~~. The acceptance was in average 0.6 points higher on the CRS scale for the baseline condition compared to solution. The absolute value was 6.8 versus 6.2 (0.8 points higher for ON in average and 0.8 points lower for ACG in average).

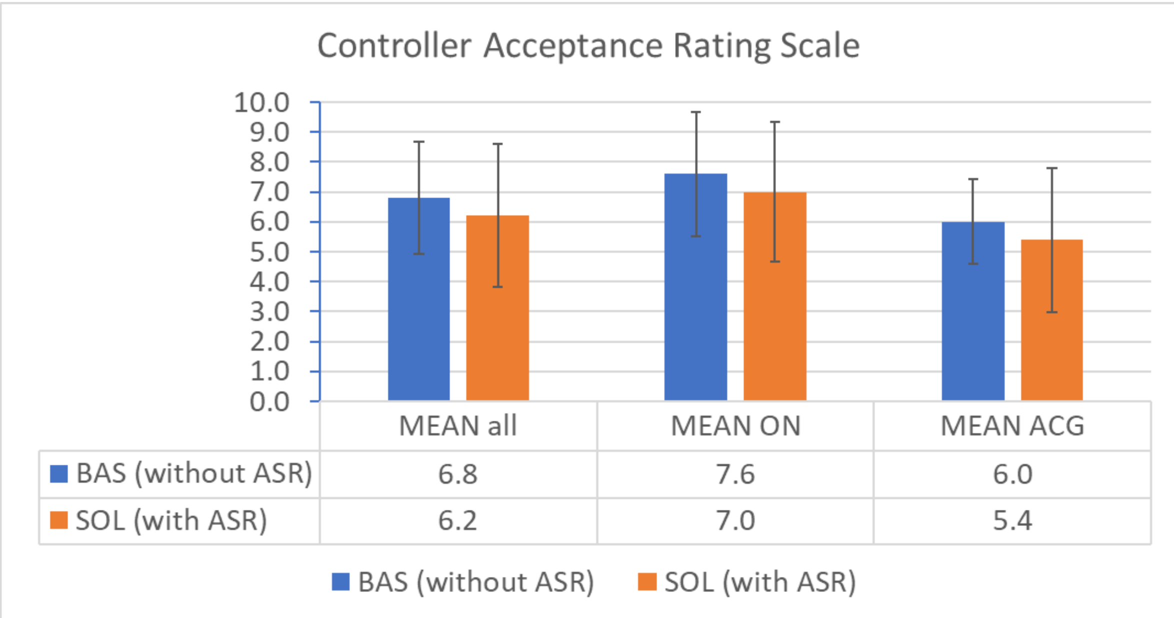


Figure E-152: ATCo ratings on CARS

Technical show-stoppers

The post-validation questionnaire contained two statements about technical show-stoppers:

1) *With Automatic Speech Recognition (ASR) I found technical show-stoppers regarding used speech recognition applications. [ASR]*

2) *With Automatic Speech Recognition (ASR) I found technical show-stoppers regarding artificial intelligence (AI) used in the speech recognition system. [AI]*

The results are shown in Figure E-153~~Figure E-153~~. The average result (below the red dotted scale mean line) of 4.2 and 3.8, respectively, indicate that there were rather no show-stoppers found during the simulation runs. However, the answers were very inhomogeneous as the standard deviation of 2.7 and 2.6, respectively (for values of "all") represent. In tendency, the ATCos that experienced worse recognition rates of ABSR output, also rated the statements for show-stoppers higher, e.g., when just analysing the male ATCos for which ABSR performed better, the average drops even further in the acceptable range below 5.5. Furthermore, mean values for technical show-stoppers are roughly one point below mean values for operational show-stoppers.

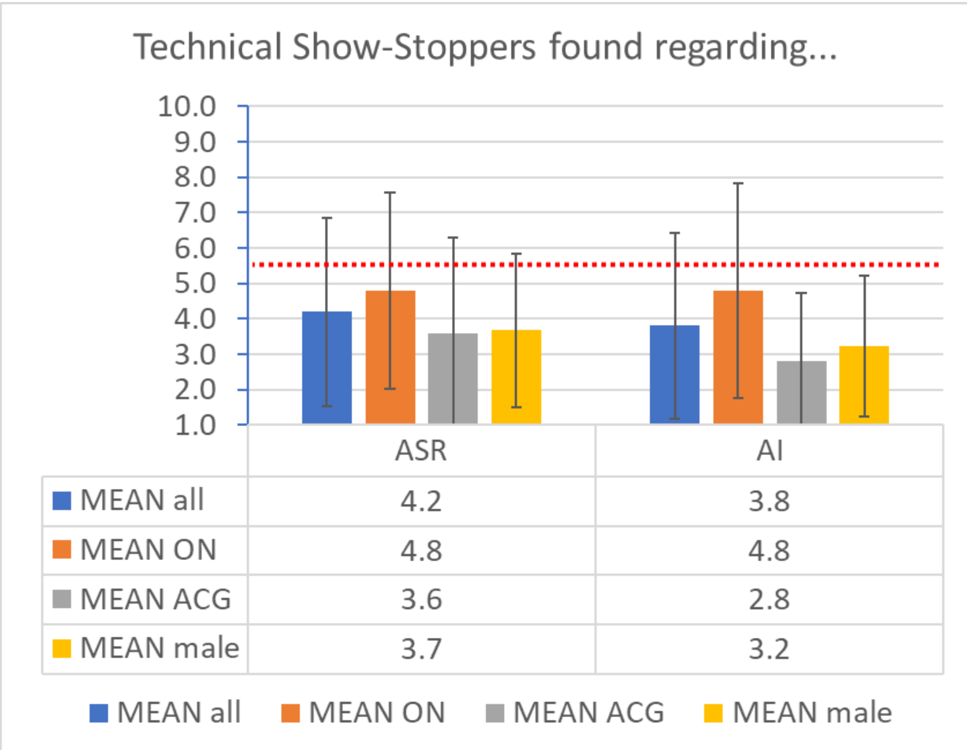


Figure E-153: ATCo ratings on technical show-stoppers

E.7.2.7 EX6-OBJ-05.972-TRL4-TVALP- H106.2010

NASA-TLX

The post-run questionnaires contained the six statements of the NASA-TLX (National Aeronautics and Space Administration - Task Load Index) questionnaire. These six statements are:

1) How mentally demanding was the task? [Mental Demand, MD]
2) How physically demanding was the task? [Physical Demand, PD]
3) How hurried or rushed was the pace of the task? [Temporal Demand, TD]
4) How successful were you in accomplishing what you were asked to do? [Operational Performance, OP]
5) How hard did you have to work to accomplish your level of performance? [Effort, EF]
6) How insecure, discouraged, irritated, stressed, and annoyed were you? [Frustration, FR]

In addition, the 15 pair-wise comparisons of workload contributing factors (as other part of NASA-TLX questionnaire) were assessed with ATCos once. The results of the weighted NASA-TLX overall and per each of the six dimensions are shown in Figure E-154~~Figure E-154~~ and Figure E-155~~Figure E-155~~.

The overall workload (OW) due to NASA-TLX was higher for solution than for baseline condition (43.1 and 38.9, respectively with huge standard deviations around 17.5). However, the difference was only induced by the ON ATCo ratings as the OW for ACG remained identical in baseline and solution.

Furthermore, a clear learning effect during the validation day in terms of NASA-TLX OW can be seen. Those five ATCos who started with baseline, rated the baseline with an OW of 41.9; those five ATCos who started with solution, rated the baseline (their second run) with an OW of 32. Those five ATCos who started with baseline, rated the solution with an OW of 48.9; those five ATCos who started with solution, rated the solution (their second run) with an OW of 37.2.

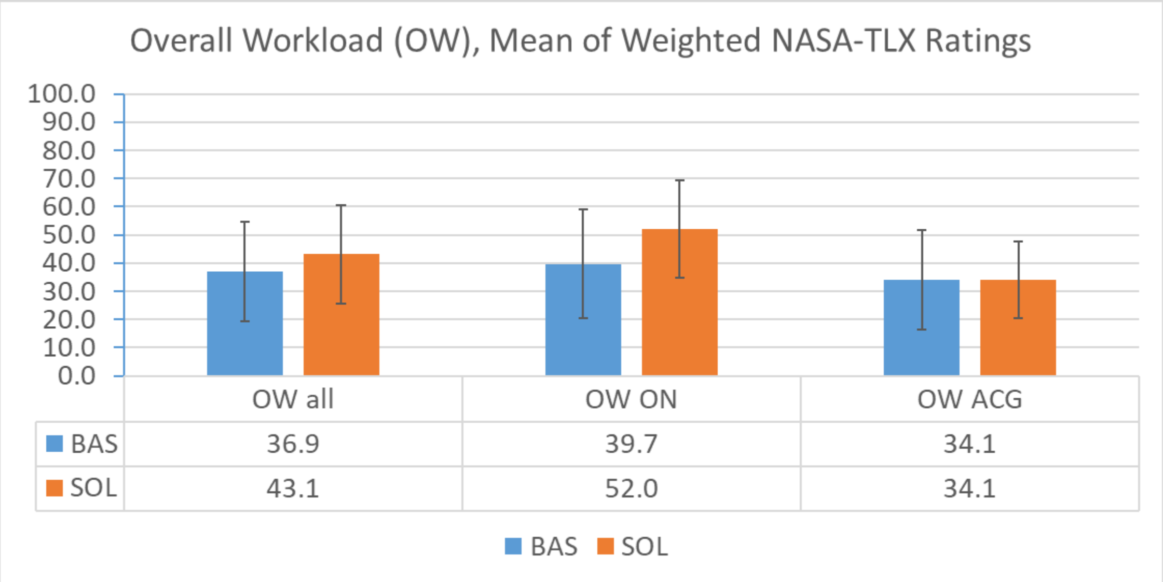


Figure E-154: ATCo ratings on NASA-TLX (Weighted Overall Workload)

When looking at the subscores for all six NASA-TLX dimensions, half of them (three) were rated equal or better in SOL compared to BAS (PD, EF, FR), the other half was rated vice versa (MD, TD, OP). In general, physical demand (PD, 3.3%) was rated being a less important contributor to workload, and mental demand (MD, 23.3%) being the most important contributor to workload. The other four dimensions were rather equally important contributors to overall workload (TD 22%, OP 18%, EF 16.7%, FR 16.7%). The horizontal axis in Figure E-155~~Figure E-155~~ shows the weight, the area shows the contribution of this very dimension to the OW of BAS and SOL condition, respectively.

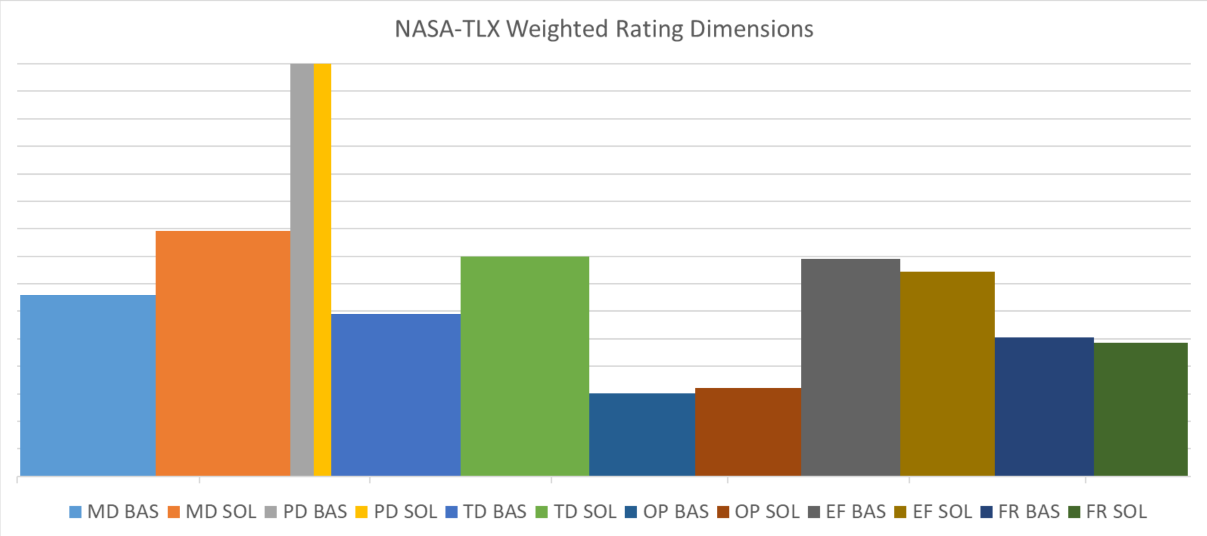
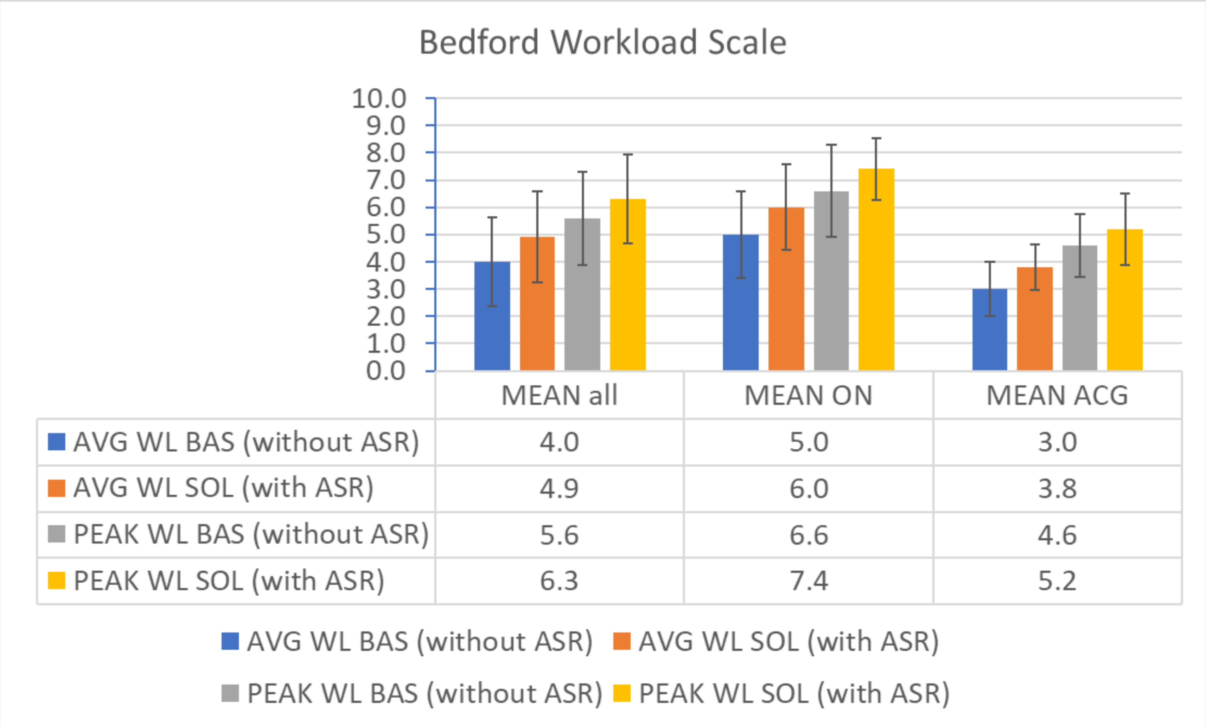


Figure E-155: ATCo ratings on NASA-TLX (Weighted Workload Factors)

Bedford workload scale

The post-run questionnaires contained the two statements of the Bedford Workload Scale questionnaire to rate the average workload (AVG WL) and peak workload (PEAK WL) during the last run on a scale from 1 to 10 with 10 being the highest workload. The results are shown in Figure E-191.

The average and peak workload were 0.9 and 0.7 points higher, respectively, in the solution condition compared to baseline condition. The peak workload was roughly 1.5 points higher than the average workload. However, the workload level in general was roughly two points lower for ACG than for ON ATCos.



	MEAN all	MEAN ON	MEAN ACG
AVG WL BAS (without ASR)	4.0	5.0	3.0
AVG WL SOL (with ASR)	4.9	6.0	3.8
PEAK WL BAS (without ASR)	5.6	6.6	4.6
PEAK WL SOL (with ASR)	6.3	7.4	5.2

Figure E-156: ATCo ratings on Bedford Workload Scale

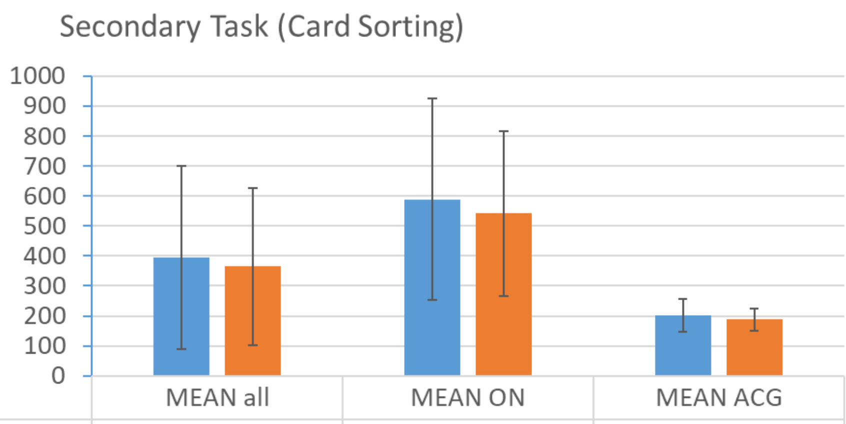
Secondary task (card sorting)

The ATCos always needed to make sure that their primary task of doing ATC remains safe and efficient. However, if they have time for a secondary task, they should sort cards. This method has already been used in older ASR projects to generate a more objective measure for workload than just via self-ratings.

ATCos needed to sort 48 cards of a German Doppelkopf deck into six decks (Aces, Kings, Queens, Jacks, Tens, Nines). At the beginning all 48 cards are on one stack with picture side of cards looking downwards. Each card needed to be turned around in a single move with just one hand to put it onto the correct of the six decks. After sorting, ATCos should name one to four randomly missing cards (that the supervisor took out of the 48 cards deck prior to start sorting). If there was an error in naming the missing cards (e.g., not all missing cards named), ATCos must try again until all missing cards are named correctly. The time measurement in seconds started when the deck of 48 cards was put next to the electronic flight strip display. The time measurement was ended when all missing cards were named correctly.

Sorting cards was trained once in each of the thirty minutes training runs. Card sorting in the baseline and solution runs started after 10 min (for at least 15 min or at least three rounds) and again after 40 min (for at least 13 min or at least three rounds). Those time frames comprised of higher traffic density to measure any difference in workload through ASR support.

The results are shown in Figure E-157~~Figure E-157~~. ATCos finished their secondary task 8.3% slower in baseline runs when not being supported by ASR (395s vs. 364s with SD of 305s and 262s). This difference was 8.8% for ON and 6.9% for ACG. When translating the result into workload, again, ON ATCos experienced a higher workload level (around 9 min sorting average) than ACG ATCos (around 3 min sorting average), but workload in solution condition seems to be lower than in baseline regarding the secondary task of card sorting. Also, the secondary task showed a great learning curve, i.e., ATCos were 18.9% slower in sorting the cards in their first simulation run compared to their second simulation run (baseline and solution alternated).



	MEAN all	MEAN ON	MEAN ACG
Sort-Time [s] BAS (without ASR)	395	588	201
Sort-Time [s] SOL (with ASR)	364	541	188

Sort-Time [s] BAS (without ASR) Sort-Time [s] SOL (with ASR)

Figure E-157: ATCo performance in secondary task (card sorting)

Instantaneous Self-Assessment of Workload (ISA)

During each simulation run, ATCos needed to rate their workload of the recent five minutes on a scale from 1 (bored) to 5 (almost overloaded). The results are shown in Figure E-193.

The average ISA workload was almost 0.1 points less in solution condition with ASR support compared to baseline condition (2.1 and 2.0 points, respectively). The ISA of ON ATCos was on a higher level with 2.6 and 2.4, respectively, but had a much lower standard deviation of below 0.3. The ISA score of ACG ATCos was around 1.6 with a standard deviation more than twice as much as of ON ATCos.

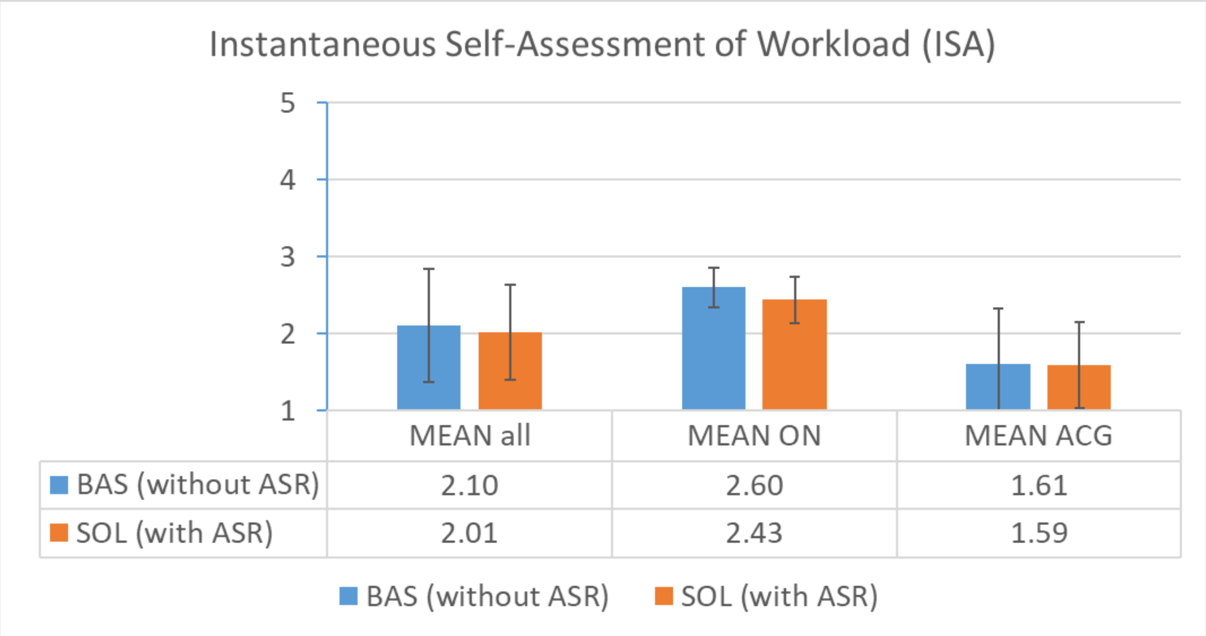


Figure E-158: ATCo workload self-assessment (ISA)

High workload contribution

The post-run questionnaires contained a free-text question about high workload: “Which factors/events/conditions have contributed to potentially high workload?”.

The structured answers and the number of ATCos noting this after the twenty conducted simulation runs (multiple notions in one questionnaire answer possible) were as follows:

- New/unknown airspace/airport layout (especially multiple remote tower): 15
- New/unknown equipment/hardware/software/electronic flight strips: 7
- Checking of ABSR output (only in solution condition): 4
- Unexpected/unusual air traffic situations: 3
- Other: Secondary task (2), tower view/runway perspective (2), slightly different phraseology to always name the calling tower (2), miscommunication, system errors

Interpreting the above results, 15 of 20 ATCo answers stated that the unknown multiple remote tower environment with unknown airport layouts induced higher workload. Furthermore, many ATCos remarked that the flight strip handling was difficult (as some details were different from “home”). This means that the majority of workload increasing factors can be assigned to environment aspects that should normally not be tested in the ABSR validation trials. The checking of ABSR output as well as unexpected situations and some further aspects were only a minor factor for higher workload.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

SASHA ATCo (Situational Awareness for SHAPE)

The post-run questionnaires contained the six statements of the SASHA ATCo as follows: In the previous run…

1) ...I was ahead of the traffic. [AHEAD]
2) ...I started to focus on a single problem or a specific airport. [FOCUS]
3) ...there was a risk of forgetting something important (such as communicating on other airport frequencies). [FORGET]
4) ...I was able to plan and organise my work as wanted. [PLAN]
5) ...I was surprised by an event I did not expect (such as an a/c call). [SURPRISE]
6) ...I had to search for an item of information. [SEARCH]

The seven-item answer scale ranged "Never, Seldom, Sometimes, Often, More Often, Very Often, Always". To present the results in a bar diagram, "Never" is translated to 0%, "Seldom" to 1/7 %"…"Very Often" to "6/7 %" until "Always" to 100%. The results are shown in Figure E-159~~Figure E-159~~. The mean values of the first two items AHEAD and FOCUS are better for BAS than for SOL condition. The mean values of the last four items FORGET, PLAN, SURPRISE, SEARCH are equal or better for SOL condition compared to BAS condition without analysing standard deviations as differences of mean values are rather small.

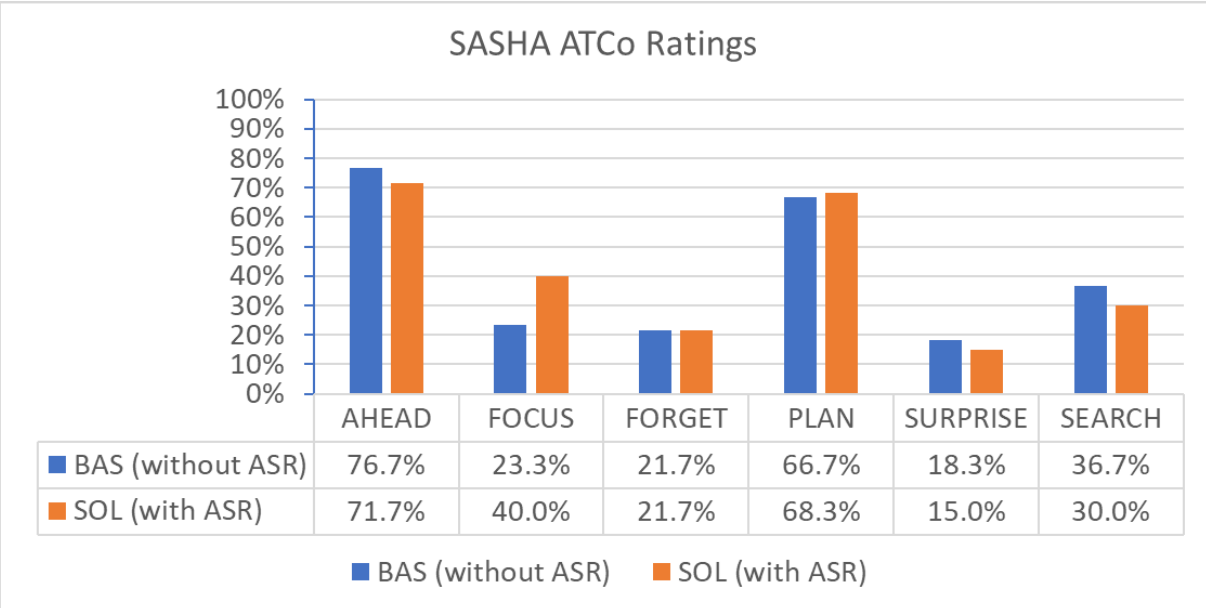


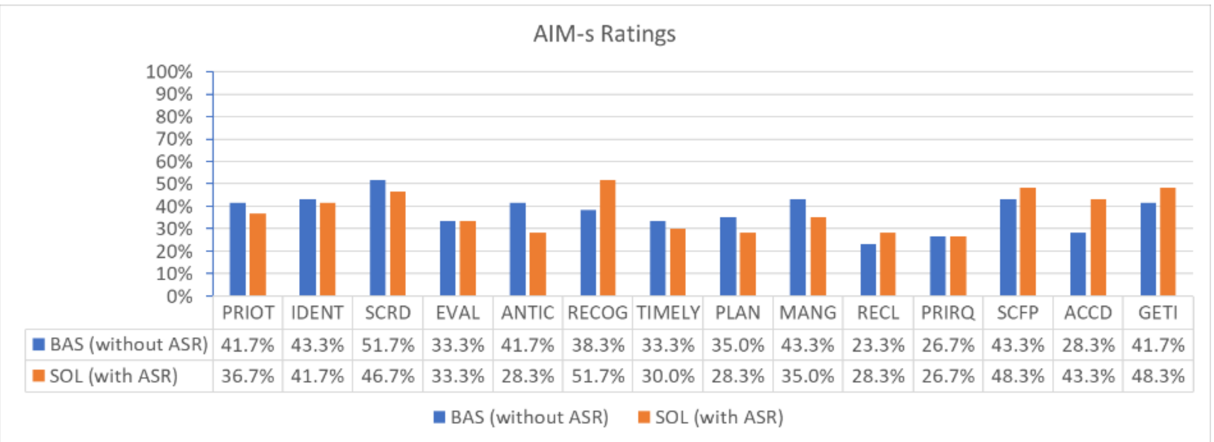
Figure E-159: ATCo ratings on SASHA ATCo questionnaire

AIM (Assessing the Impact of Automation on Mental Workload)

The post-run questionnaires contained the sixteen statements of AIM-s as follows: In the previous run, how much effort did it take to...

1) ...prioritise tasks? [PRIOT]
2) ...identify potential conflicts? [IDENT]
3) ...scan radar or any display? [SCRD]
4) ...evaluate conflict resolution options against the traffic situation and conditions? [EVAL]
5) ...anticipate the future traffic situation? [ANTIC]
6) ...recognise a mismatch of available data with the traffic picture? [RECOG]
7) ...issue timely commands? [TIMELY]
8) ...evaluate the consequences of a plan? [PLAN]
9) ...manage flight data information? [MANG]
10) ...share information with team members? [SHARE]
11) ...recall necessary information? [RECL]
12) ...anticipate team members' needs? [TMN]
13) ...prioritise requests? [PRIRQ]
14) ...scan flight progress data? [SCFP]
15) ...access relevant a/c or flight information? [ACCD]
16) ...gather and interpret information? [GETI]

The seven-item answer scale ranged “None, Very Little, Little, Some, Much, Very Much, Extreme”. To present the results in a bar diagram, “None” is translated to 0%, “Very Little” to 1/7 %”…”Very Much” to “6/7 %” until “Extreme” to 100%. The statements SHARE and TMN are not analysed further as there were no team members during the simulation runs (so fourteen statements remain). The results are shown in Figure E-160~~Figure E-160~~.
Nine of the fourteen statements have been rated better in average (less) for SOL condition than for BAS condition. Only the five statements related to information RECOG, RECL, SCFP, ACCD, GETI have been rated worse for SOL condition compared to BAS condition.



	PRIOT	IDENT	SCRD	EVAL	ANTIC	RECOG	TIMELY	PLAN	MANG	RECL	PRIRQ	SCFP	ACCD	GETI
BAS (without ASR)	41.7%	43.3%	51.7%	33.3%	41.7%	38.3%	33.3%	35.0%	43.3%	23.3%	26.7%	43.3%	28.3%	41.7%
SOL (with ASR)	36.7%	41.7%	46.7%	33.3%	28.3%	51.7%	30.0%	28.3%	35.0%	28.3%	26.7%	48.3%	43.3%	48.3%

Figure E-160: ATCo ratings on AIM-s questionnaire

SATI (SHAPE Automation Trust Index)

The post-run questionnaires contained the six statements of the SATI as follows: In the previous working period, I felt that…

1) ...the system was useful. [USEFUL]
2) ...the system was reliable. [RELIABLE]
3) ...the system worked accurately. [ACCURACY]
4) ...the system was understandable. [UNDERSTAND]
5) ...the system worked robustly (in difficult situations, with invalid inputs, etc.). [ROBUST]
6) ...I was confident when working with the system. [CONFIDENT]

The seven-item answer scale ranged “Never, Seldom, Sometimes, Often, More Often, Very Often, Always”. To present the results in a bar diagram, “Never” is translated to 0%, “Seldom” to 1/7 %”…”Very Often” to “6/7 %” until “Always” to 100%. The results are shown in Figure E-161~~Figure E-161~~.

The mean values of the first item USEFUL is clearly better for SOL than for BAS condition. However, the other five mean values are better for BAS than for SOL condition. It is noteworthy, that the four statements RELIABLE, ACCURACY, UNDERSTAND, ROBUST of ON ATCos have better ratings for SOL than for BAS condition in average.

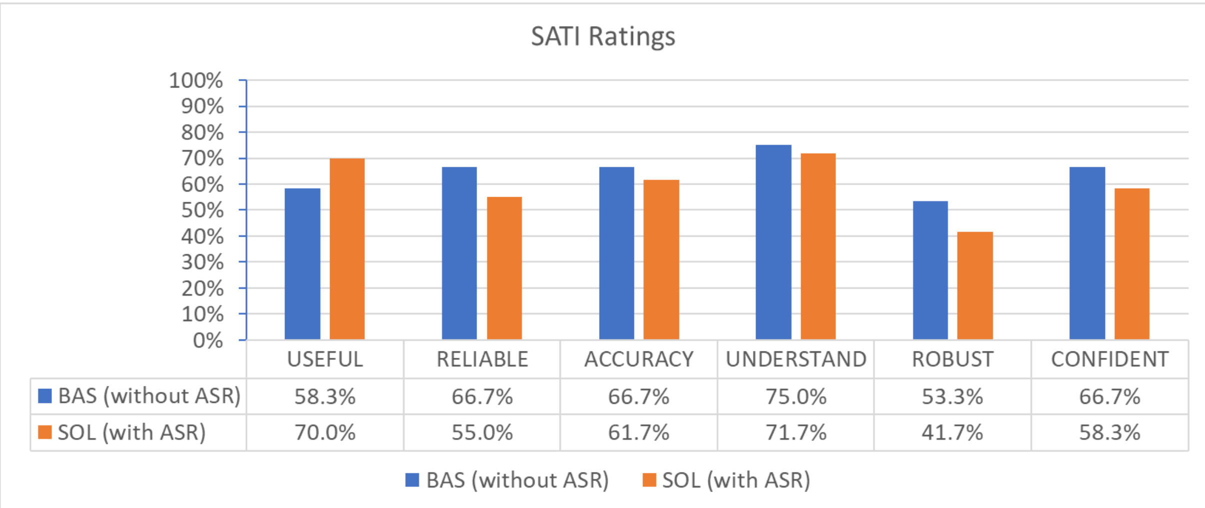


Figure E-161: ATCo ratings on SATI questionnaire

Debriefing

The debriefing was conducted as a semi-structured interview with some pre-defined questions and some options for further thoughts and inputs. The feedback of ATCos is semantically reported per category in the following.

Study Preparation and Conduction

1) Briefing slides via e-mail two weeks before the trials and briefing at DLR was very good.
2) All ATCos felt well-trained for the purpose of the validation after one hour of training.
3) Simulation pilots performed well comparable to simulation pilots at the air navigation service providers of the ATCos themselves.
4) Air traffic scenarios were rated to be fine for the study purpose.
5) On the one hand, baseline condition (manual work) was like every day work, so performance might be better therefore (2 ATCos). On the other hand, ASR in solution condition was good because it supports me in using a flight strip system that I am not used to.

Flight Strip System (feedback not related to study purpose ‘ABSR technology’)

1) Runway bay handling needs to be improved (sorting, highlighting, timing, etc.).
2) Drag-and-drop functionality over borders of flight strip bays for individual planning purposes was needed.
3) Handling of training flights (touch-and-go/low approach) that do not switch from an arrival flight strip to a departure flight strip were slightly difficult.
4) Strip handling for a/c that cross the control zone was difficult due to status icon options.
5) Visual flagging of strips (left/right) would be beneficial.
6) Hide some non-frequent status icons.
7) Deleting of manual free text pen input needed.
8) “Takeoff” status should include “lineup”-status (if not given explicitly).
9) Combination of selection of taxi status and taxiway would be easier.
10) Some suggestions for other colors, e.g., of ground vehicles, and color consistency with other systems.
11) One ATCo loved the flight strip system, the majority of ATCos were basically ok with it.
12) Many ATCos liked the fade-away functionality of flight strips.

Outside View (feedback not related to study purpose ‘ABSR technology’)

1) ABSR output (transcription and annotation in solution condition) was just checked for curiosity by all ATCos.
2) Portion of gazes at the three areas ‘outside view’, ‘radar view’, ‘flight strip display’: too much on flight strips and too few on outside view where one can hardly identify small objects.

ABSR Functionality (also related to electronic flight strip display)

1) ABSR concept and implementation was found to be good by many ATCos.
2) Checking ABSR output in the flight strip display slows some ATCos, because in the baseline mode, ATCos tick while speaking.
3) Some ATCos judged the speed of ABSR output while speaking as sufficient, two ATCos wanted to have faster output.
4) Non-standard situations should be covered well by ASR.
5) Speech understanding (annotation process) was good to cover errors of speech recognition (transcription process).
6) Highlighting of callsigns and status icons (in green) and 10s-highlighting mechanism in electronic flight strips was fine for all ATCos.

7) When ASR worked fine, tendency to over-rely on automatism existed.
8) In case of non-recognition, double effort to manually recognize the error and correct it compared to pen input (2 ATCos).
9) When I am home in Lithuania/Austria, I tell my colleagues that working with DLR's speech recognition was…
    a. interesting (said by all ON ATCos)
    b. worked pretty well (2 ATCos)
    c. positively surprising (even when speaking fast)
    d. very good even if I am no early adaptor of new technologies and being very safety critical.
10) If you would use it tomorrow in your tower controller working position (not multiple remote tower), would ASR help?
    a. Yes, would be great (3)
    b. Nothing to be changed to be used tomorrow (1)
    c. Great support possible if some/many aspects are improved (4)
11) Did you speak differently in baseline and solution condition?
    a. In baseline less carefully spoken, because only pilots needed to understand (3 ATCos).
    b. Spoken closer to phraseology in solution as being better supported then (2 ATCos).
    c. Some stated that there was no difference in speaking.
12) ATCos automatically become more phraseology conform: That is one of the greatest advantages of such a technology.

Further Applications/Ideas/Things to be changed

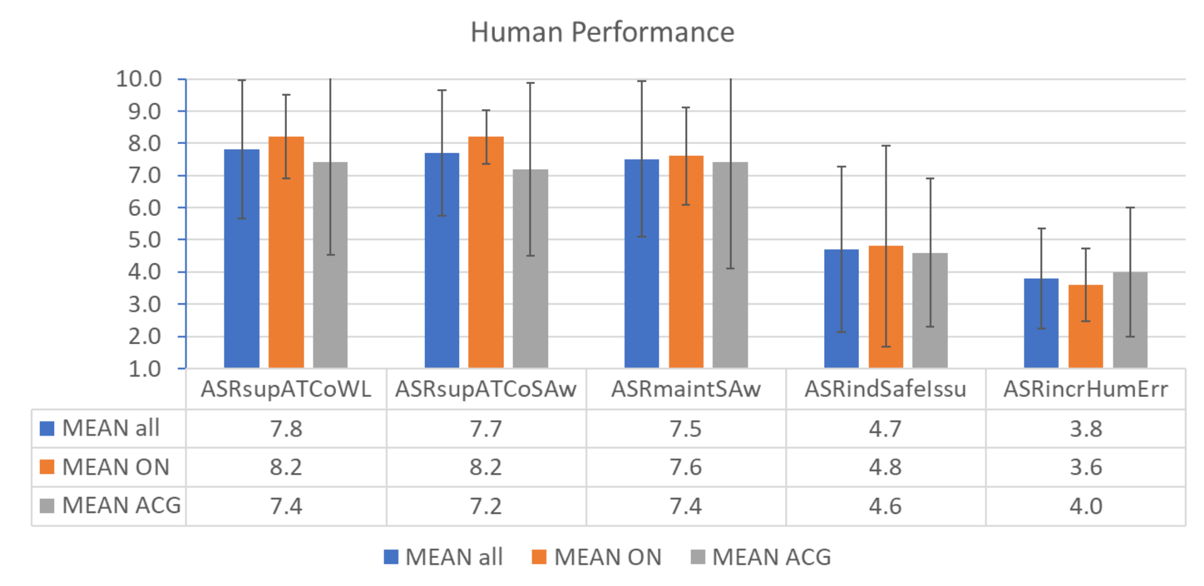
1) Callsign highlighting in flight strip display from pilot utterance would help to identify the communication partner.
2) Speech log for pilot utterances (esp. in emergency situations) anywhere on the controller screen.
3) Connect ABSR output with
    a. radar information for automatic setting of landed/departed status.
    b. lighting system to turn off stop bar lights in case of lineup clearance.
    c. follow-the-greens for correct lighting.
    d. airport phone conversation to automatically extract and include stand numbers given by the airport.
4) Safety net functionality for dedicated aspects in case of good error rates, e.g., readback error detection.
5) Transcription for incident analysis and searching for callsigns; other analysis on transcribed data.
6) Great technology for on-the-job-training (OJT).

Human performance questions

The post-validation questionnaire contained five statements about human performance:

1) *I think that ASR supports me in maintaining workload at acceptable level. [ASRsupATCoWL]*
2) *I think that ASR supports me in maintaining an adequate level of situation awareness. [ASRsupATCoSAw]*
3) *My situational awareness is maintained at acceptable level with Automated Speech Recognition (ASR). [ASRmaintSAw]*
4) *I see many safety related issues to be solved regarding automatic speech recognition implementation. [ASRindSafeIssu]*
5) I think that ASR did increase the potential for human errors. [ASRincrHumErr]

The results are shown in Figure E-162~~Figure E-162~~. ASR seems to support maintaining situation awareness and workload of ATCos at an acceptable level with mean values of 7.5 and beyond on a 10-point scale. In addition, ASR was not found to induce safety issues or to increase the potential for human errors with mean values below the scale mean of 5.5.



	ASRsupATCoWL	ASRsupATCoSAw	ASRmaintSAw	ASRindSafeIssu	ASRincrHumErr
MEAN all	7.8	7.7	7.5	4.7	3.8
MEAN ON	8.2	8.2	7.6	4.8	3.6
MEAN ACG	7.4	7.2	7.4	4.6	4.0

Figure E-162: ATCo ratings on human performance

E.7.2.8 EX6-OBJ-05.972-TRL4-TVALP-H106.2020

The post-validation questionnaire contained two statements about operating methods:

1) *I can apply operating methods in an accurate, efficient, and timely manner with ASR. [AccOpMeth]*

2) *I think that operating methods are clearly identified and consistent in all operating conditions. [OpMethConsis]*

The results are shown in Figure E-163~~Figure E-163~~. The operating methods with ASR seem to be accurate, efficient, timely and consistent in different conditions with mean values of 8 and 7.4 respectively.

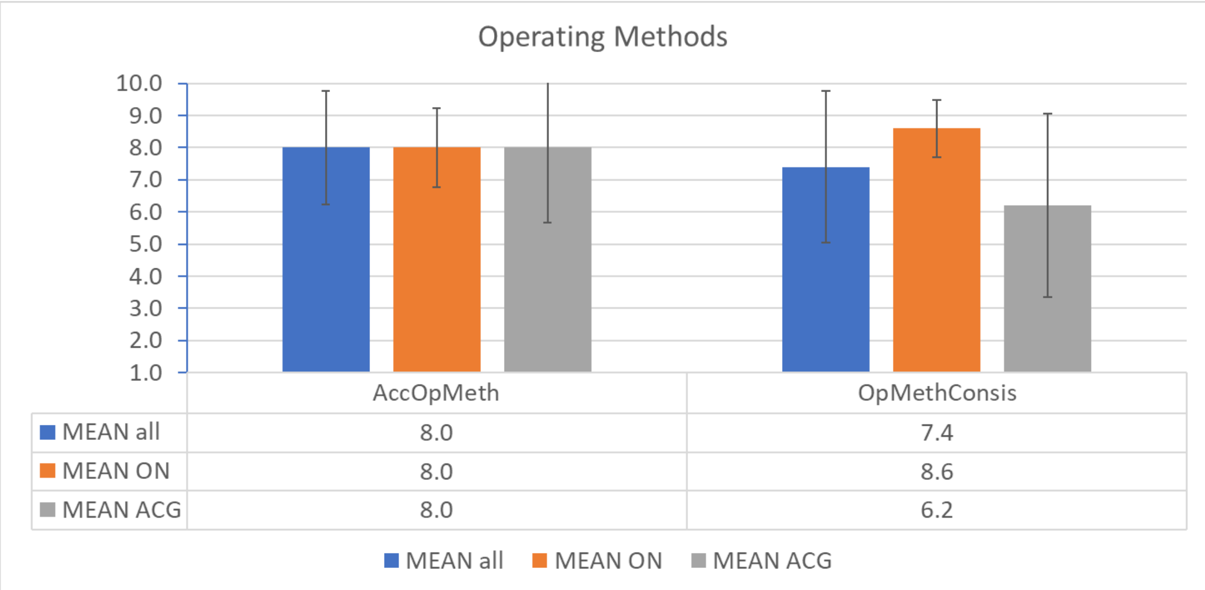


Figure E-163: ATCo ratings on operation methods

E.7.2.9 EX6-OBJ-05.972-TRL4-TVALP-H106.2030

The post-validation questionnaire contained one statement about job satisfaction:

1) *Overall, I was satisfied performing my task with ASR. [JobSatisf]*

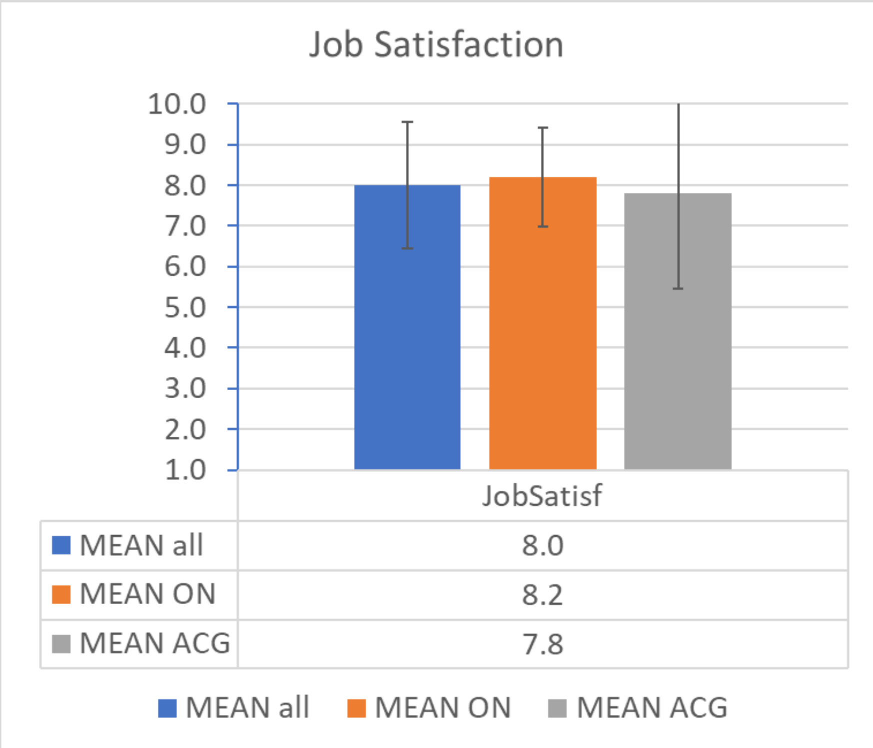


Figure E-164: ATCo ratings on job satisfaction

The results are shown in Figure E-164~~Figure E-164~~. ATCos rated their job satisfaction with using ASR high (mean value of 8).

E.7.2.10 EX6-OBJ-05.972-TLR4-TVALP-SAFE.2010

For evaluating the aspects of human errors, workload, and situation awareness with post-validation questionnaire statements refer to Human Performance Questions above.

E.7.2.11 EX6-OBJ-05.972-TLR4-TVALP-PERF.2010

The post-validation questionnaire contained one statement about workload in connection with flight delays:

1) *I think ASR support can lead to reduce/balanced ATCo workload and reduced flight delays. [WL_RedFlightDelay]*

The results are shown in Figure E-165~~Figure E-165~~. The mean value is just very slightly above the scale mean of 5.5, but has the highest standard deviation of all statements in the post-validation questionnaire with 3.0.

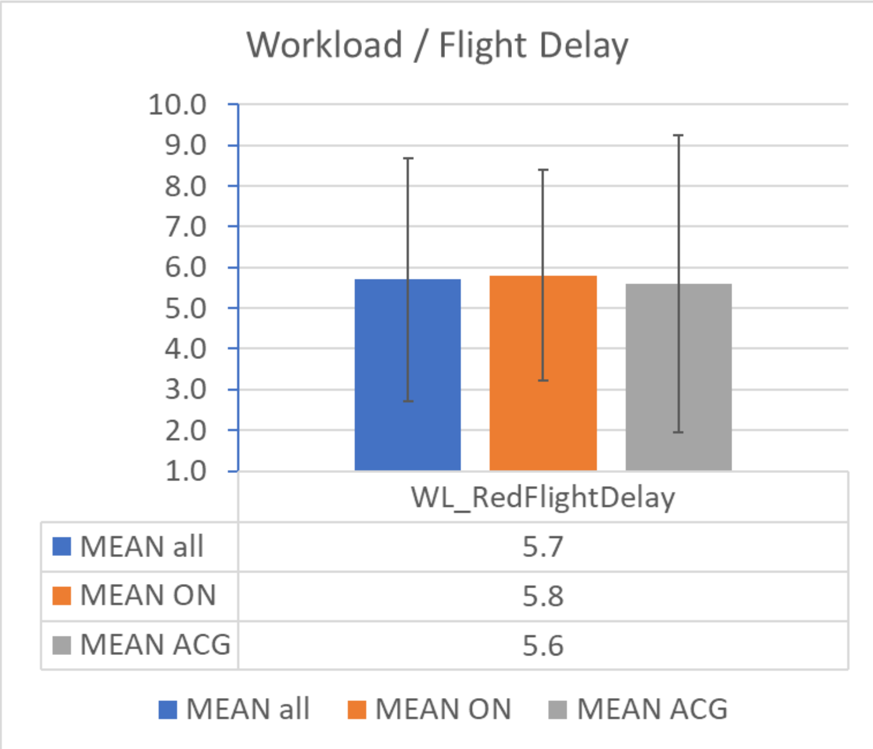


Figure E-165: ATCo ratings on workload/flight delays

E.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 much 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). This was most probably caused by CPU overload leading to “hick-ups” in recorded wav-files and delay in ASR word recognition output.

E.7.4 Confidence in results of EXE-006

E.7.4.1 Level of significance/limitations of EXE-006 Technological Validation Exercise Results

When considering the results of the TRL6 validation exercise EXE-05.97.2-TRL4-ASR-006 several factors should be considered:

- The validation is based on real-time simulation environment addressing speech recognition in a multiple remote tower environment – three towers named Vilnius, Kaunas and Palanga. 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. For example, the ATCos think:
    - Label handling is appropriate,
    - All ATCos were familiar with the new system and some prefer the new system than the one they currently using in their existing workplaces,
    - Vienna ATCos had difficulties with the new colors of the display because it is very different from the one they use every day. This affected their situational awareness and speed of reaction.
    - DLR system is acceptable and generic and its usage is widely applicable,
    - There were no electronic or any other coordination between sectors in the day to day operations. There was only coordination with ground vehicles in case of abnormal situations (bird strike, emergency landing).

Provided such functions were fully replicated the ATCos would be in more familiar environment, in terms of learned hand-motor functions and acquired skills, even the expectation of how certain HMI will perform etc.

- During the validation, differences from the current operational environment and methods in the day to day operations were present. Different variables related to the new system functionalities are affecting the result.
- 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 TWR 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.
- Nominal situations and a few emergency and abnormal situations (bird strike and emergency landing) were addressed.
- TWR ATCos participated in the TRL4 validation exercise. Therefore, it can be assumed that the results will be valid for all TWR ATCos.
- The participants with pseudo-pilot role were ATC experts from DLR - which in turn had to learn the new role of work.

E.7.4.2 Quality of EXE-006 results

The quality of the validation results is determined as medium 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.

E.7.4.3 Significance of EXE-006 results

Each of the ten ATCos did two runs per day: one reference and one solution run (alternating order to avoid learning effect in the data). Each ATCO was working simultaneously with three airports and corresponding displays.

As each ATCo was at DLR from 8:30 to 16:30, we also did exactly 10 validation days, i.e., we had some days in between where there was no ATCo at DLR, e.g., because PANSA cancelled their participation. Start was Feb 14, end was Mar 3. Hence, in sum ten ATCo feedbacks were collected for the whole validation. For operational significance the existing airspace and applicable procedures and corresponding letters of agreement were applied. All participants were holders of an active tower ATCo licence.

Statistical significance was rather absent in the quantitative questionnaire results of ATCos due to high standard deviations. However, the recognition and error rates of the ABSR system base on a lot of utterances. Hence, these numbers haver higher significance.

E.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 a (multiple) aerodrome control tower environment.

E.8.1 Conclusions on technological feasibility

The ASR technology has shown to be feasible in an ATC tower 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 91% with error rates for callsigns of below 1% and for commands below 5% are possible to achieve. The quantitative and qualitative feedback of ATCos was good and motivating to go beyond TRL4 and would have been even better if the full potential of ABSR accuracy have been offered to them.

E.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 ASR support. The positive results for system usability, job satisfaction and some workload measurements show the potential of ABSR in a (multiple remote) tower environment – even if a row of other measurements do not show any significant differences between baseline and solution.

The data shows that ATCos speak differently, i.e., closer to phraseology if being supported by ABSR (i.e., solution runs have higher command recognition rates that baseline runs; in the latter, the speech was analysed as well, but the output was not shown to the ATCo). On the one hand, this might be, because they get better support if recognition rates are higher, on the other hand, it might be due to the pure awareness of working with speech recognition in the background. If ATCos are sticking closer to ICAO phraseology just by pure presence of an ABSR system, that could already be a safety feature.

To summarize, EXE-006 has shown great potential of using the output of an ABSR system. However, it also revealed relevant aspects to be considered when moving forward from TRL4 to TRL6 in a future ASR activity.

E.8.3 Recommendations

Recommendations for further development related to the SUT:

- ASR-related:
  - EXE006.TRL4.REC001: To be further developed as a potential on job training help.
  - EXE006.TRL4.REC002: To be further developed as potential help for incident analysis.
  - EXE006.TRL4.REC003: To provide appropriate training of the ATCos to the new system functionalities in order to achieve better understanding and to build up trust into the new functionalities.
  - EXE006.TRL4.REC004: Provide appropriate training and exposure of the ATCos to the new system functionalities to achieve better understanding critical for building up trust into the new functionalities.
- Not relevant for ASR, but for electronic flight strip system:
- EXE006.TRL4.REC005: Runway bay handling to be improved.

- EXE006.TRL4.REC006: Improvement of display colours.
- EXE006.TRL4.REC007: Drag-and-drop functionality over borders of flight strip bays for individual planning purposes to be used.
- EXE006.TRL4.REC008: Further improvement of strip system with visual flagging of callsign.

General 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 (also leading to worse results of recognition rates as shown above). Furthermore, a big amount of data must be recorded from operations rooms (not from labs), because ATCos speak different in simulations. Finally, the recording configuration for training and validation should be the same, e.g., in both recordings with a face mask or in both without (which was not the case for EXE-006, but worsened the results slightly).

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 (TRL6 and beyond) 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 F Technological Validation Exercise 007 Report

F.1 Summary of EXE-007 Plan

The present section is a report of the technological validation exercise EXE-05.97.2-TRL4-TVALP-ASR-007 (in the following EXE-007), run by Leonardo, on its Rome premises, situated on Via Tiburtina. Test activities ran as detailed in the Technical Validation Plan [28] document.

F.2 EXE-007 description and scope

Leonardo validation Exercise 007 demonstrated the benefits of introducing an automatic speech recognition system in Air Traffic Management to support a set of ATCOs routine tasks in the Tower and Ground environment.

The ASR system (ASR4ATC) was integrated in Leonardo Lead In Sky CWP, interacting via A-SMGCS 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, asking ATCOs to approve it once it is prefilled. To this purpose, a set of valid ATCO clearances, instructions and routine tasks were identified, along with their corresponding HMI masks/features.

Simulations were run in Rome, at Leonardo Tiburtina site, at the beginning of May 2022 (04 - 06 May). Validation took place in the shape of a human-in-the-loop real-time simulation in Ground and Tower environments, simulating scenarios at Sofia Airport. Leonardo Lead In Sky Controller Working Position was used, using ASMGC-S as opposed to Electronic Flight Strips. The remaining SW elements of the platform all belong to the Lead In Sky suite. Tests were run simulating traffic in and around Sofia airport, with the support of six ATCOs all working in turn as tower/ground 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 air traffic into/out of Sofia, with flights arriving and leaving, with Ground operations until take-off clearance or starting from landing clearance stages. Sofia can be considered to be a medium size airport, with a single runway and not overly complex taxiways.

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 Ground ATC clearances, such as startup, pushback, taxi, lineup and takeoff/land, and ground traffic was light. 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 (see Technical Validation Plan [28] for reference):

- UC-97-TRL4-TS-201 Highlighting of recognized callsign
- UC-97-TRL4-TS-202 Showing full recognized utterance/command in HMI
- UC-97-TRL4-TS-203 Manual manipulation or rejection of an ASR output

F.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 a Tower sector, while the other to a Ground 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 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 F-166~~Figure F-166~~ below

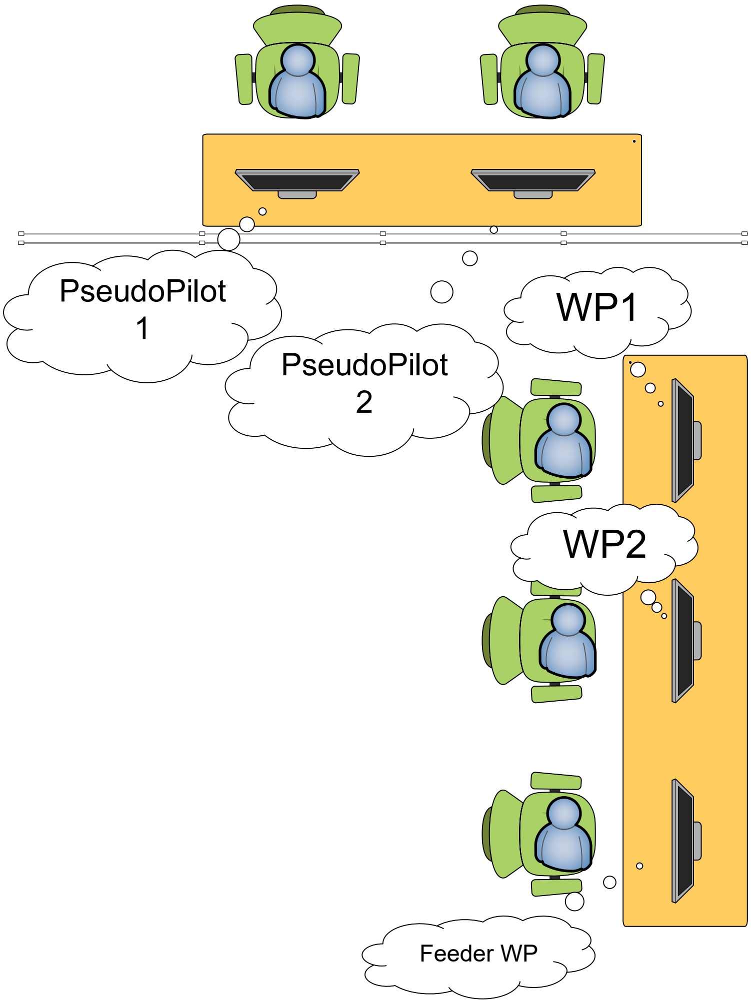


Figure F-166: Simulation room layout for EXE-007

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

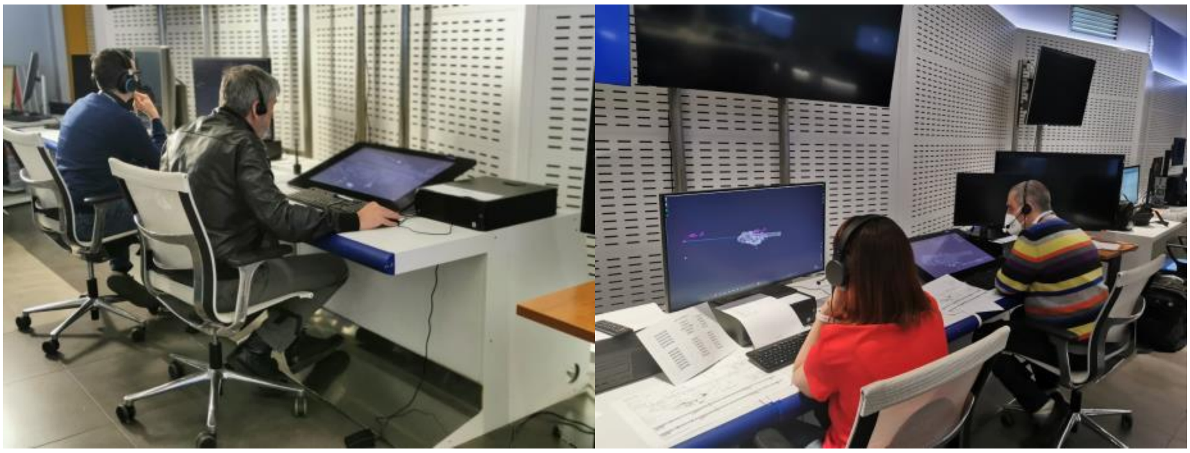


Figure F-167: ATCOs WPs for TWR(L) and GND(R) for EXE-007

The simulation was run on the platform as shown in Figure F-168~~Figure F-168~~ below.

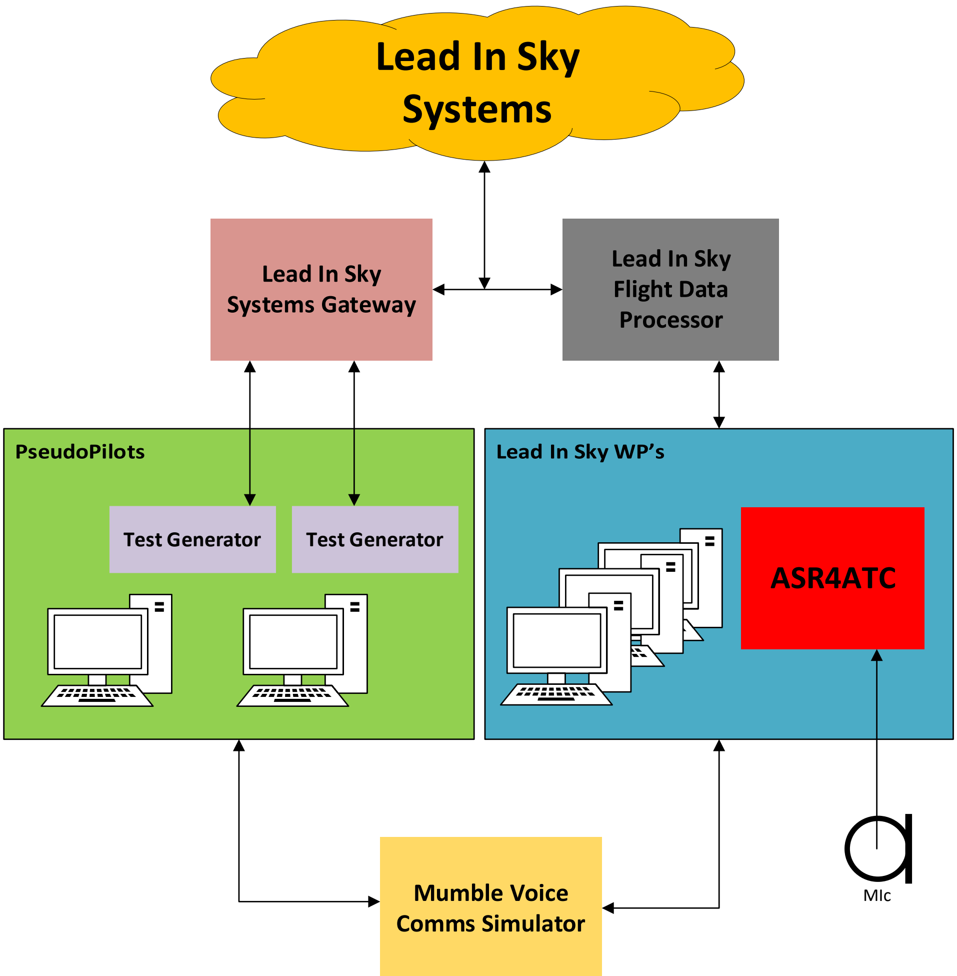


Figure F-168: Simulation platform for EXE-007

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, airport. The airport, Sofia (LBSF) qualifies as a medium size, 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.

F.2.1.1 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 of [38], aptly reduced for the validation scope

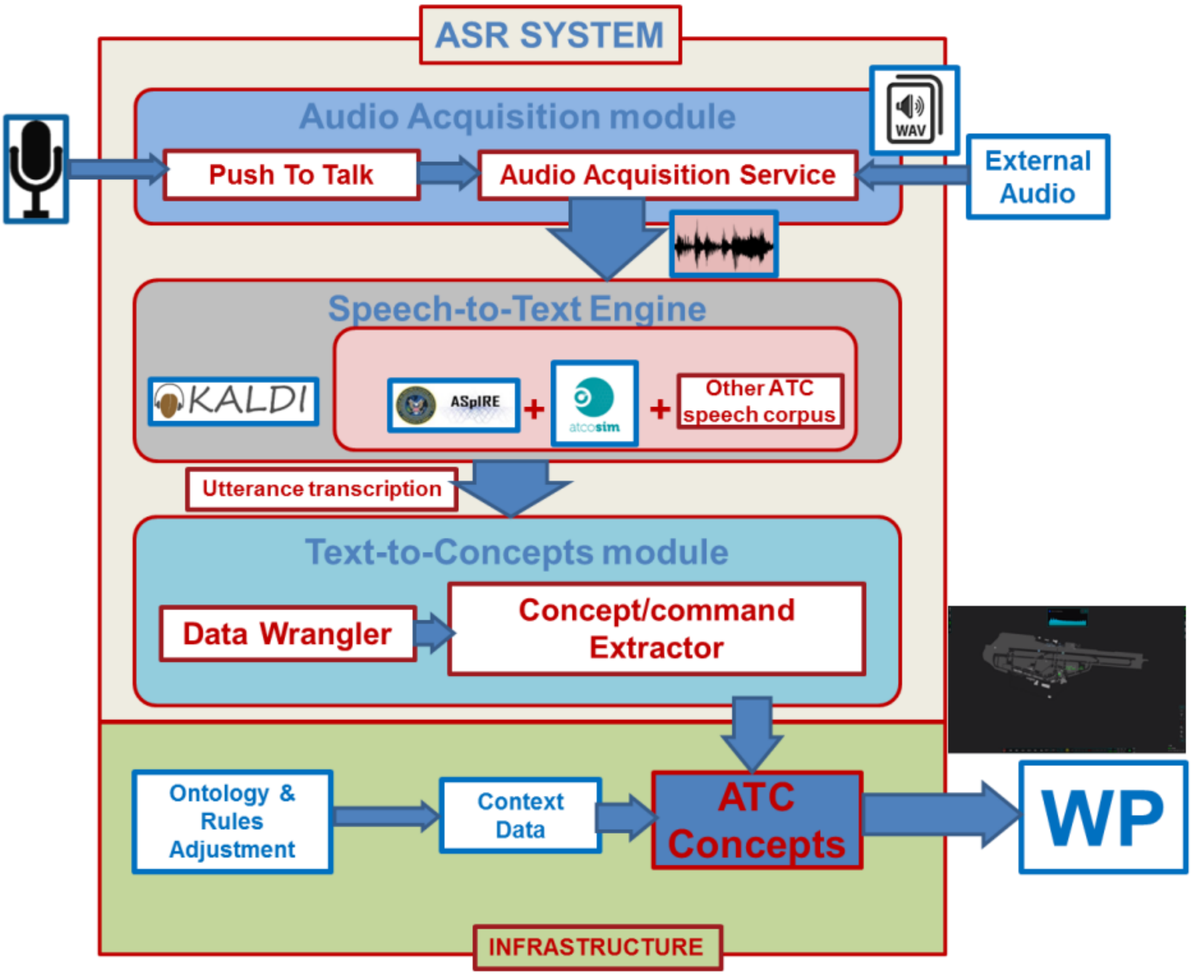


Figure F-169: A block diagram of the Leonardo ASR4ATC module, used for EXE-007

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.

F.2.1.2 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 F-170~~Figure F-170~~ 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 F-170: A screenshot of WP showing the audio recording widget while recording

F.2.1.3 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.

F.2.1.4 Context-based data generator

Context-based data are crucial in order to improve ASR performance. In the EXE-007 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.

F.2.1.5 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 F-171: View of a Ground GTG simulation scenario

Figure F-172: 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 aerodrome area.

F.2.1.6 Aerodrome Layout

Sofia (ICAO Code LBSF) airport is, as it shows from the picture below, of medium complexity and traffic volumes. There is one 09-27 runway, of which only 09 was used due to the simulation environment, and about 30 taxiways all leading to an apron which is situated, along with terminal buildings, all to one side of the runway, therefore not requiring frequent runway crossings.

Figure F-173: Sofia aerodrome operational view via Leonardo WP

F.2.1.7 Controller Working Position

The Controller Working Position utilized was a purpose custom version of the Leonardo Lead In Sky product, with a characterization on Ground and Tower operations. Maps and geographical information were all based on Bulgaria and Sofia AIP.

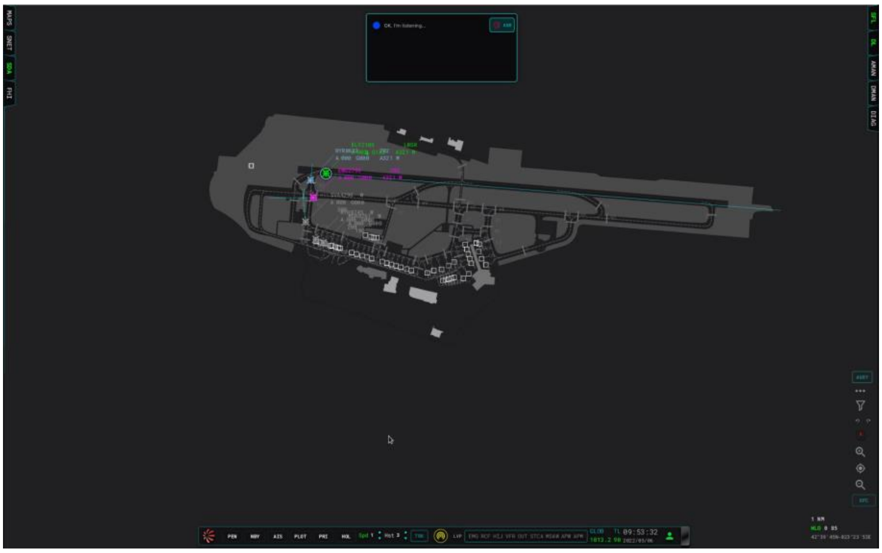


Figure F-174: Sofia aerodrome on Leonardo WP showing ASR recording widget

F.2.1.8 Traffic main characteristics

The simulation traffic was designed having in mind reliable airport 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.

F.3 Summary of EXE-007 objectives and success criteria

All SESAR Solution Validation Objectives with their respective Success criteria for Solution 97.2 as defined in section 4.3.2 of the TVALP [28], are covered in Exercise 007 without modified Exercise Validation Objectives and/or Exercise Success Criteria.

Please refer to the TVALP [28] section 4.3.2 for specific validation objectives and associated Success Criteria.

F.4 Summary of EXE-007 validation scenarios

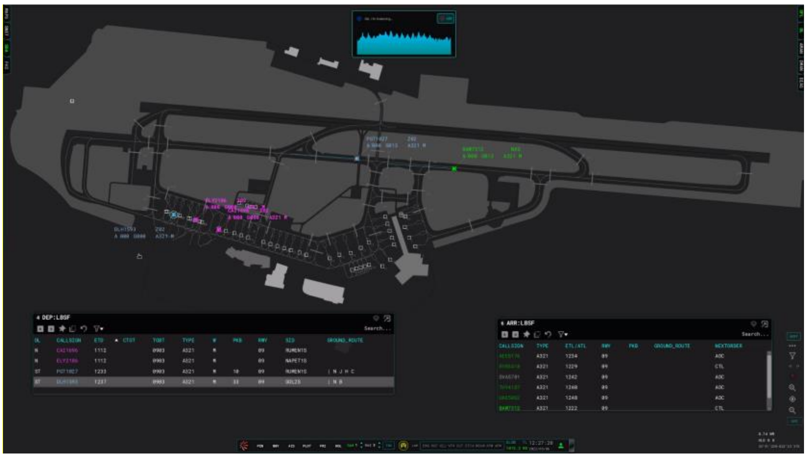


Figure F-175: Sofia Airport Ground Operations simulated on Leonardo WP with ASR

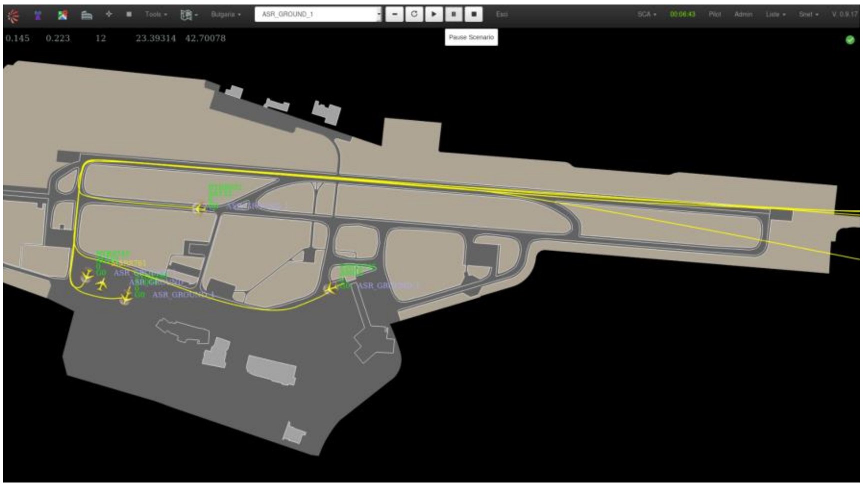


Figure F-176: Leonardo GTG track generator and simulator running a scenario simulating Sofia Airport

F.4.1 Reference Scenario

The reference scenario addresses the current tower and ground operational environment, at Sofia Airport, as shown in the figure below.

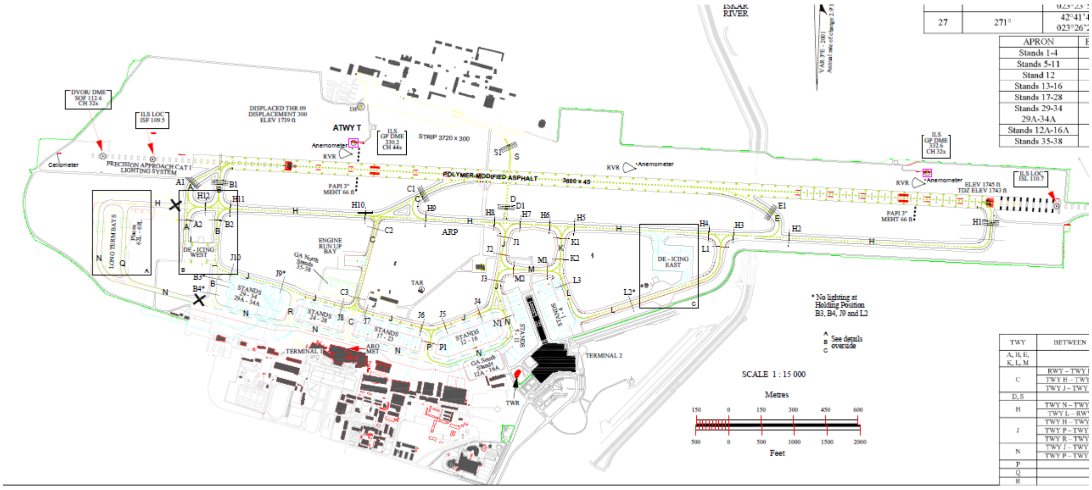


Figure F-177: Sofia Airport Map

In such an operational context, controllers issue ATC clearances and instructions to the flight crews by R/T communications. The flight crew is expected to confirm each clearance by read-back, in an accurate and timely manner. As soon as they receive the read-back, ATCOs manually update the system (using mouse and keyboard) in order to input the clearances issued and align the CWP data. Flights are subdivided into departing and arriving, in an even proportion.

Each scenario integrates ASMGC-S tool for Ground traffic management, available as part of the Leonardo Lead In Sky CWP, while ATCOs manually manage traffic.

Such a reference setting was used as a baseline against which the solution scenarios - implementing the ASR module - were compared.

F.4.2 Solution scenarios

Solution scenarios address experimental conditions in which ASR support is utilized, triggered by ATCOs pressing a pedal under their working position desk, or a hotkey on the keyboard.

The ASR system is intended to support and expedite ATCOs performing routinely tasks (e.g. updating ground routes on the a/c label) by automatically recognizing a set of verbal clearances/values extracted from R/T communication and by prefilling applicable masks. It is the ATCOs' task to select whether and when ASR is being used, while using R/T communication with pseudo pilots. Leonardo implementation of ASR gives visual feedback of its operation while ATCOs can activate or deactivate it at their own convenience. Then in turn ASR will generate the appropriate masks corresponding to the recognized order. If the contents of a mask or a popup window generated by ASR are correct, the ATCO will simply acknowledge them with a mouse click. The command is then implemented in the system according to the ATCO clearance in exactly the same way it would normally be during normal use of WP.

Figure F-178: Leonardo Lead In Sky WP with ASR recording widget

If the ATCO finds that the ASR output is not representing what they just said, he/she will correct manually the recognized command/values within the relevant mask(s), or they will just reject it, filling the mask by hand, or retrying to activate the command by voice.

Clearances and commands in such a scenario are issued by ATCOs also by means of ASMGC-S using ASR, following suit from the previous example. In the current WP implementation, some clearances such as STARTUP, PUSHBACK or TAXI were issued with popups, with no qualifiers or arguments other than the time the order is expected to be fulfilled at (and for Taxi it was an undesired limitation).

Scenarios were named: SOL1, SOL2, SOL3 where SOL1 is the reference scenario run using ASR, SOL2 is a different scenario, and SOL3 puts an accent on pushing the ASR tool to its limit, testing also the performance of context-based data, to improve callsign recognition rate.

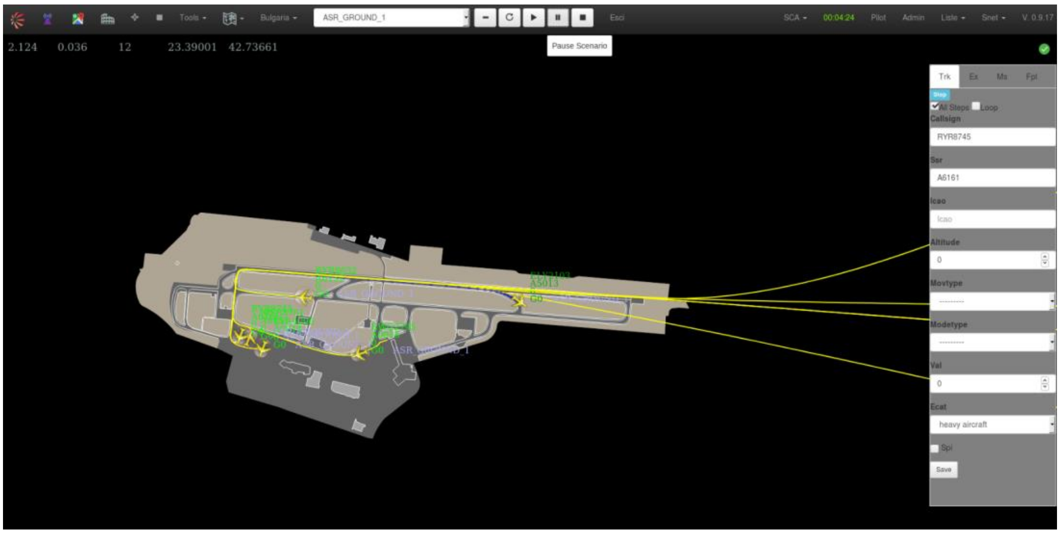


Figure F-179: GTG simulator running a Solution scenario

F.5 Summary of EXE-007 assumptions

Identifier	Title	Description	Justification	Impact on Assessment
AS.EXE-007.01	Simulation traffic Scope	The extent of simulation will be limited to medium size single runway airport	The underlying choice of Sofia airport sets, with its size, the amount of traffic to handle, and therefore the amount data for ASR to interpret and process	High
AS.EXE-007.02	Traffic conditions	Traffic conditions will be regular in terms of flow and amount	The test has to reflect a “normal” setting	Low
AS.EXE-007.03	Weather	Normal/good	No impact analysis of abnormal weather conditions	Low
AS.EXE-007.04	Ambient noise	Noise generated by conversation in a control tower Room	Ambient noise will be limited to what can be generated by conversation of other ATCOs present in a control room. No simulation of a/c generated noise is foreseen	Medium

Table F-27: Technological Validation Results EXE-007

Due to platform integration issues with WP (also regarding ground route instructions), the number and the extent of clearances was limited, especially for taxi clearances, in the implementation used for the simulation.

F.6 Deviation from planned activities

A few deviations from the initial exercise plan are to be reported:

- The last solution run of the simulation was performed with only one controller (TWR).
- As a result of the first deviation, one post-run questionnaire response is missing, therefore 17 solution-run related questionnaires were analysed, instead of 18.

- Manual correction rate - data logs related to this indicator are not available. ATCOs were asked to report a qualitative estimation (in %) of the manual correction required during the validation exercise.
- The final debriefing on the third day of the simulation could not be held due to time constraints.

F.7 EXE-007 validation results

F.7.1 Summary of EXE-007 results

Validation Objective	Success Criteria	Result	VO Status
OBJ-05.972-TRL4-TVALP-FEAS.2010 To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS.	CRT-05.972-TLR4-TVALP-FEAS-2011 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use of ASR.	No malfunctions related to ASR occurred; ASR performance during the validation was deemed acceptable.	OK
	CRT-05.972-TLR4-TVALP-FEAS-2012 No operational show-stoppers have been identified during laboratory tests (based on a prototype) related to the use AI suggestions.	No hiccups occurred during the runs and all minor issues were due to the WP and simulation platform integration rather than ASR.	
OBJ-05.972-TRL4-TVALP-FEAS.2020 To identify possible technical feasibility issues and possible show stoppers.	CRT-05.972-TLR4-TVALP-FEAS-2021 Laboratory tests (based on a prototype) have verified the technical feasibility of the use of ASR supported by AI/ML.	AI/ML techniques were successfully employed to generate an ASR model which functioned to expectations. Usage of context-based data has significantly improved performance, at no reliability cost.	OK
	CRT-05.972-TLR4-TVALP-FEAS-2022 Laboratory tests have verified that the integration of the SESAR technological solution with other related system	Context-based data greatly improved performance, with no stability concerns. Integration with WP and in	

Co-funded by the European Union

Validation Objective	Success Criteria	Result	VO Status
	enablers is technically feasible.	general LIS suite did not show blocking issues.	
OBJ-05.972-TRL4-TVALP-H106.2010 To assess that the technical systems for ASR support the ATCOs in performing their tasks.	CRT-05.972-TLR4-TVALP-H106-2011 Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining workload at acceptable level.	100% ATCOs declare that ASR supports controllers in maintaining an acceptable level of workload.	OK
	CRT-05.972-TLR4-TVALP-H106-2012 Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining an adequate level of situation awareness.	100% ATCOs responded that ASR supports ATCO in maintaining an adequate level of situation awareness. Enhancement of the platform was suggested via changing the background colours of the ASR pop-up window or/and a more evident highlighting of the 'Hooked' a/c on the HMI, which would increase situational awareness.	OK
	CRT-05.972-TLR4-TVALP-H106-2013 ASR does not increase the potential for human error	66% of ATCOs agreed that the ASR system did not increase potential for human error compared to current operations.	OK
	CRT-05.972-TLR4-TVALP-H106-2014 ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by ASR	100% of ATCOs provided positive feedback on adequacy of ASR feedback with some improvements suggested, as shown above	OK
	CRT-05.972-TLR4-TVALP-H106-2015 Measured callsign recognition rate, command	100% of ATCOs provided positive feedback on callsign and command recognition rate. 66% of ATCOs gave	OK

Validation Objective	Success Criteria	Result	VO Status
	recognition rate, error rate and rejection rate of ASR system are considered within acceptable levels by the majority of ATCOS (at least 75%)	favourable feedback on callsign and command rejection rates.	
	CRT-05.972-TLR4-TVALP-H106-2016 Majority of ATCOs (at least 75%) confirm adequate usability of ASR system	100% of ATCOs responded positively regarding usability of the ASR system. Some concerns due to lack of familiarity in how to use the pedal for ASR activation also were reported.	OK
	CRT-05.972-TLR4-TVALP-H106-2017 Majority of ATCOs (at least 75%) provide positive feedback on acceptance of ASR tool	94% of ATCOs passed positive feedback on the acceptance of the ASR tool. Moreover, 100% of ATCOs said the frequency of wrongly highlighted call signs and of incorrectly recognised ASR commands was acceptable, as well as the tool's latency, considering the traffic samples under evaluation in the scenarios.	OK
	CRT-05.972-TLR4-TVALP-H106-2018 ATCOs (at least 75%) trust in the system is at an acceptable level	83% ATCOs provided affirmative response being asked regarding trust in the ASR tool.	OK
OBJ-05.972-TRL4-TVALP-H106.2020 To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of ASR.	CRT-05.972-TLR4-TVALP-H106-2021 Majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient, and timely manner	The majority of ATCOs mentioned several time during the debriefings that the tool's latency and feedback allowed them to apply operating methods in an accurate, efficient, and timely manner. They also	OK

Validation Objective	Success Criteria	Result	VO Status
		stated no change arose in operating methods.	
	CRT-05.972-TLR4-TVALP-H106-2022 Majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions	100% of ATCOs responded that operating methods working with ASR were clear and consistent.	OK
OBJ-05.972-TRL4-TVALP-H106.2030 To assess job acceptance and satisfaction with the introduction of ASR.	CRT-05.972-TLR4-TVALP-H106-2031 The majority of ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance	All ATCOs provided positive feedback on job satisfaction and acceptance.	OK
OBJ-05.972-TLR4-TVALP-SAFE.2010 To assess the impact of Automatic Speech Recognition on safety.	CRT-05.972-TLR4-TVALP-SAFE -2011 The changes related to the implementation of Automatic Speech Recognition do not increase potential for human error and therefore not reducing safety levels.	66% of ATCOs responded that ASR does not increase the potential for human error compared to current operations.	OK
	CRT-05.972-TLR4-TVALP-SAFE -2012 ATCO's workload with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	ATCO's workload with ASR is maintained at acceptable levels. 83% of ATCOs agreed that the level of safety was at least as in today's operations and not affected negatively by ASR. See also CRT-05.972-TLR4-TVALP-H106-2011	OK

Co-funded by the European Union

Validation Objective	Success Criteria	Result	VO Status
	CRT-05.972-TLR4-TVALP-SAFE -2013 ATCO's situational awareness with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.	ATCO's situational awareness with ASR is maintained at an acceptable level, therefore not reducing safety levels. See also CRT-05.972-TLR4-TVALP-H106-2012.	OK
	CRT-05.97B-TLR4-TVALP-SAFE -2014 Safety assessment activities and the results are documented and integrated in the overall solution validation results	No dedicated safety event/scenario was tested. However, safety effects related to the ASR were investigated across the whole validation exercise, verifying tool performance and the effectiveness of context data support usage in recognising a/c call signs. No specific safety related issues were identified.	OK
OBJ-05.972-TLR4-TVALP-PERF.2010 To assess the performance benefits of Automatic Speech Recognition supported by AI/ML.	CRT-05.972-TLR4-TVALP-PERF-2011 Laboratory tests show that the SESAR technological solution improves Cost Efficiency performance by reducing cost per flight (PER.TRL4.3) (through e.g. reduction of workload, reduction of delay times)	Qualitative feedback from ATCOs indicates at least no detrimental effect of ASR on performance. A more detailed description can be found in the KPA analysis (section F.7.1.2)	POK

Table F-28: Technological Validation Results EXE-007

F.7.1.1 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 tower 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. WP clearances, due to LIS platform development, validation and integration, were reduced to simplified ones, especially for taxi orders. Conversely, also usual orders such as backtrack, shutdown, hold short, cross, vacate and several others, while implemented in the ASR engine, were not be tested since not available as part of the utilized WP instruction set.

Some commands such as “continue” “contact”, “proceed”, “assume” 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).

F.7.1.2 Results per KPA

Method

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

For the purpose of the validation, each ATCO was assigned either to the ground controller or to the tower controller position and swapped role after 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 A-SMGCS
- 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-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:

Time frame	04.05.2022	05.05.2022	06.05.2022
10:00 -10:05	Briefing	Briefing	Briefing
10:05 - 10:50	REF1_97	REF1_97	REF1_97
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_97	SOL1_97	SOL1_97
12:10-12:30	Questionnaire & Debriefing	Questionnaire & Debriefing	Questionnaire & Debriefing
12:30-12:35	Briefing	Briefing	Briefing
12:35-13:20	SOL2_97	SOL2_97	SOL2_97
13:20-13:40	Questionnaire & Debriefing	Questionnaire & Debriefing	Questionnaire & Debriefing
13:40-14:45	Lunch	Lunch	Lunch
14:45-15:30	SOL3_97	SOL3_97	SOL3_97
15:30-16:05	Questionnaire & Final Debriefing	Questionnaire & Final Debriefing	Questionnaire & Final Debriefing

Table F-29: RTS agenda

As Table F-29, 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 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.

Figure F-180: 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

- Workload

The average level of workload reported for the solution scenarios (3.5 out of 10) was below the maximum tolerable WL level (5) and identical to the average workload level calculated for the reference scenario. Moreover, all ATCOs agreed that the level of workload during the solution scenarios was acceptable and the 'hook' function was reported as the main contributor to this judgment.

- Situational Awareness

Results collected reveal no decline in situational awareness when using ASR. In fact, it was indicated that when using the system, ATCOs feel like they are provided with increased required information, compared with the amount they have normally available and which they should look for on their own; as a result they were able to better plan their work. The 'Hook' function was reported to help improving situational awareness. All ATCOs rated situational awareness as either 'high' or 'perfect' during solution scenarios, which suggests that controllers were generally satisfied about their situational awareness levels when using ASR. Some improvements were also identified as described in section F.8.3.

- Usability

Results in the usability area were also positive. Overall, ATCOs were confident about using ASR and would like to use it frequently. They found ASR easy to use and its functions well integrated. Half the ATCOs agreed that they would need initial support in order to be able of effectively using the system, and most of them agreed that some training on the system would be required, to understand how the tool "behaves" and also to learn how to proactively *adapt* their speech to ASR. Adaptation was particularly relevant in the exercise due to the limited number of utterances/commands implemented in the current ASR4ATC version. Generally, the system wasn't found complex, and no inconsistencies were reported. Additionally, the data collected indicated that ATCOs use the 'hook' function in different ways, which might be due to different roles or tasks they are involved in when using it.

More specifically, when asked if they used the function when there was R/T communication with another sector, there was no agreement between the ATCOs. Answers were also divided when ATCOs were asked if they would use the function to manage inbound traffic and deviating traffic. However, the majority of ATCOs agreed that they use 'hook' to identify flights in dense traffic, when entering their sector, or when arriving traffic calls for info or requests. During debriefings it was suggested that ASR could also be used to improve interaction with the eFlight strips by automatically filling in the information on the strips in order for ATCOs to cross-check information.

Opinions were divided on the ASR activation means as some ATCOs prefer the use of a pedal and others the use of keys.

- Acceptance & job satisfaction

Most ATCOs provided a high rating for acceptance. Both the frequency of incorrectly highlighted ASR call signs and the frequency of inaccurately recognised ASR commands was acceptable for ATCOs, as well as the system latency considering the traffic scenarios evaluated (medium load, single runway). No issues were reported on the system's interaction with A-SMGCS, however ATCOs suggested that tighter integration with A-SMGCS functionalities (e.g. ASR being able to display the a/c taxi route) be considered. All ATCOs agreed that job satisfaction increases when using ASR.

- Trust

Overall trust in the system was mostly reported by ATCOs between 'high' and 'very 'high', especially as a result of the good recognition rates and latency of the system.

Table F-30 shown below indicates the results collected during the final debriefings by using Want/Have matrices.

REMOVE (have but don't want) • Background colours of the ASR pop-up window • The use of the two keys for: ASR activation and for frequency communication	**PRESERVE (have and want)** • Rate of recognition • Latency • Hook function • The automatic update of the ARR/DEP list of the system • The final ATCO approval of the command by clicking ENTER on the HMI or pressing it with the keyboard • All current functionalities
AVOID (don't have and don't want) • n/a	**ACHIEVE (don't have but want)** • Integrate the whole ATC phraseology or extend other relevant command (e.g. TOC, taxi to stand etc..) (2) • Automatic recognition and execution of commands (e.g assuming traffic, activation/ de-activation of stop bars) (2) • ASR to highlight an Occupied/ Closed RWY/ Taxiway (2) • Simultaneous use of frequency and ASR • ASR to be activated with 'Push to talk', without pressing a pedal or a key • Better integration with A-SMGCS functionalities (e.g ASR displays the a/c taxi routing)

Table F-30: Want/Have Matrix results

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

Want-Have Matrix

- Preserve

ATCOs were generally positive about all current functionalities of the ASR tool, its recognition rate and latency in response, which they specifically mentioned in the 'preserve' category of the matrix. In addition, the 'Hook' function was said to be very useful for determining a certain a/c's position and to maintain situational awareness.

One ATCO particularly liked the fact that ASR displays the command for final ATCO approval by clicking ENTER on the HMI or pressing a key, as they considered it a useful safety barrier. Another ATCO mentioned they would preserve the automatic update of the ARR/DEP list of the system, which could eventually be integrated with electronic flight strips in future ASR developments.

- Achieve

In terms of ASR improvements, two ATCOs wrote down that it would be useful for ASR to recognise the whole ATC ontology and include other relevant commands e.g 'taxi to stand, turn left/right, vacate etc.' as this would significantly reduce workload. Moreover, for further Workload reduction, two ATCOs suggested that ASR should automatically recognise and execute some of the commands such as assuming traffic or activation/de-activation of stop bars.

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.

- Remove

One ATCO indicated that the use of the two keys, one for the activation of the ASR and the other for communicating via R/T should be removed as it can become a source of confusion, increasing workload and impacting situational awareness. Another ATCO suggested that the background colours of the ASR pop-up window should be more visible, as they also might impact situational awareness.

- Avoid

No feedback was provided by any of the controllers in this section.

Safety

- Considering the results on Human Performance, the perceived potential for Human Error did not increase when using ASR system for controllers. ASR did not clearly impact the perceived potential for Human Error.
- The overall safety level was perceived by most of the controllers the same as in today's operations.

F.7.2 Analysis of EXE-007 results Results per Technological Validation objective

F.7.2.1 OBJ-05.972-TRL4-TVALP-FEAS.2010

To confirm the concept is operationally feasible when addressing the identified Use Cases in the TS.

CRT-05.972-TLR4-TVALP-FEAS-2011

No operational show-stoppers were identified during laboratory tests (based on a prototype) related to the use of ASR.

CRT-05.972-TLR4-TVALP-FEAS-2012

No operational show-stoppers were identified during laboratory tests (based on a prototype) related to the use of AI suggestions. Given the usage of context-based data, actual show stoppers would just imply high rates of incorrect recognition as opposed to no recognition at all, always preferable in the event of an error. The AI model used had been gathered using several generic corpora and focused training on the validation exercise.

F.7.2.2 OBJ-05.972-TRL4-TVALP-FEAS.2020

To identify possible technical feasibility issues and possible show stoppers.

CRT-05.972-TLR4-TVALP-FEAS-2021

Laboratory tests (based on a prototype) have verified the technical feasibility of the use of ASR supported by AI/ML.

Other than platform inconsistencies which could lead to repeated failures, as well as network malfunctions, which belong more to the ICT realm rather than to the solution itself, the ASR module, made up of its functional blocks, has been found to work reliably with no performance or availability issues. On the other hand, the WP integration turned out to be a limiting factor, due to limited prior testing and availability, also causing limitations in the range and scope of some commands.

CRT-05.972-TLR4-TVALP-FEAS-2022

Laboratory tests have verified that the integration of the SESAR technological solution with other related system enablers is technically feasible.

F.7.2.3 OBJ-05.972-TRL4-TVALP-H106.2010

To assess that the technical systems for ASR support the ATCOs in performing their tasks.

CRT-05.972-TLR4-TVALP-H106-2011

The majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining workload at acceptable levels.

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 reference scenario, the average level of workload was estimated at 3.5, while for the solution scenarios, no significant difference in the average workload level is observed in all three scenarios tested (3.8 in Sol.1 as compared to 3.5 in Sol. 2 and 3.2 in Sol.3). Furthermore, no significant standard deviation difference was observed calculating it for each scenario sample (1.64 for ref. scenario, 1.47 for Sol.1, 1.38 for Sol2 and 0.84 for Sol 3).

As a result, the three solution scenarios can be considered homogeneous, with no significant peaks in workload.

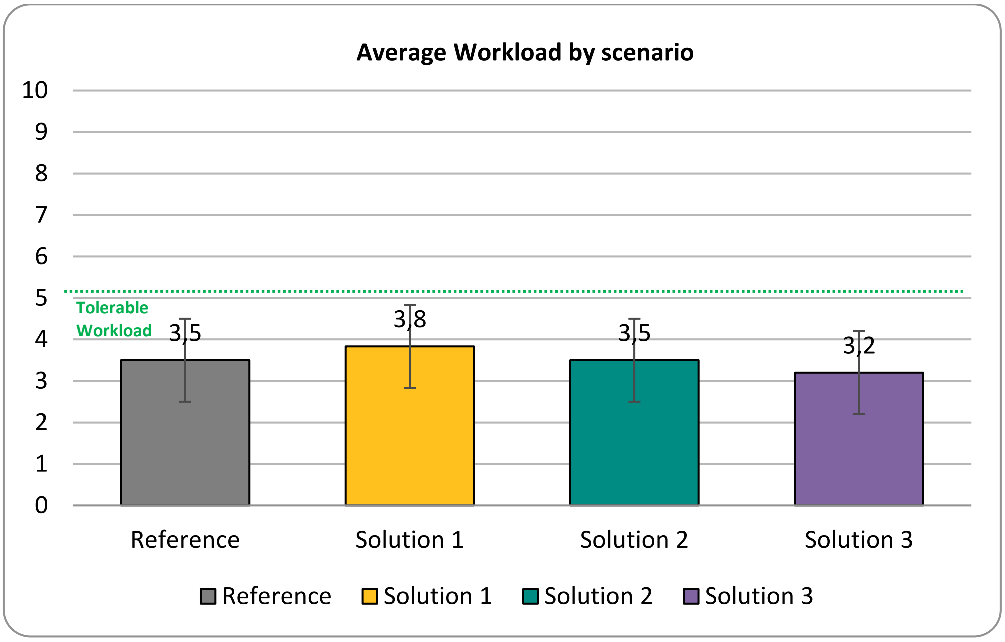


Figure F-181: Post-run questionnaire: Average Workload by reference and solution scenarios

To compare Workload in the reference scenario to the solution scenarios, an average workload level was calculated for the reference scenarios related responses. Such a workload level which resulted identical to the average workload for all solution scenarios (3.5). Although the data sample for the reference scenarios is smaller than the one for all solution scenarios, after applying Welch's t-test, it resulted that the mean values of the two quantities are not statistically different, since the p value was 0.96 (above 0.05). An identical Workload level for both scenario types indicates that WL does not increase when using ASR. Please see below Figure F-182~~Figure F-182~~ for a visual representation.

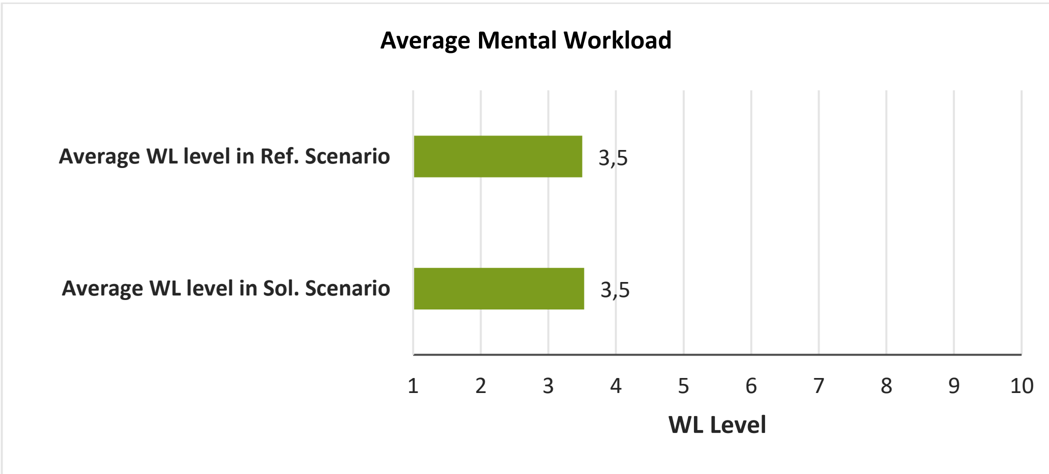


Figure F-182: Post-run questionnaire: Average Mental Workload in reference and solution scenarios on the Bedford scale

For the solution scenarios, most ATCO responses (40%) indicated an acceptable level of WL, below the maximum tolerable WL level (5), while 12% of responses were just below the threshold (6). Figure F-183~~Figure F-183~~ shows the distribution of answers (on a 10-point Bedford scale), expressed also in percentage. Moreover, in the comments section of the Post-run questionnaire on workload, one ATCO commented that it was very easy to cope with the tool. Another ATCO added that the workload level allowed them to accomplish the task while two of the ATCOs mentioned that controlling the traffic was not demanding during the solution runs.

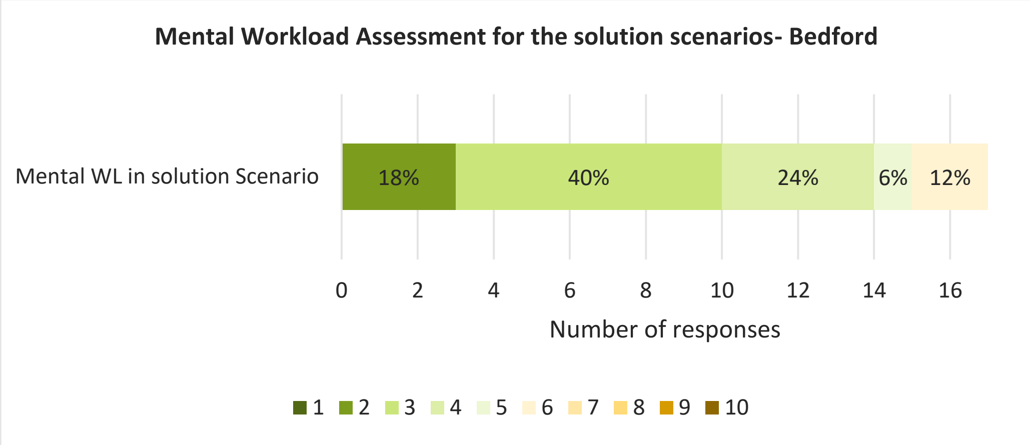


Figure F-183: Post-run questionnaire: Number of responses for Mental Workload in the Solution scenario

Figure F-184 below shows the average mental workload for solution vs. reference scenarios according to controller role. Although all mean values are below the maximum tolerable workload level (5), in the ground position, the average workload reported during the solution scenario (3.75) is slightly higher than the one in the reference scenario (2.33). In the debriefings, it was revealed that ASR support was less evident for the GND position, as some instructions were missing from the ASR phraseology such as stand numbers, which might have contributed to a higher workload during the solution scenario. However, for the TWR position, the workload level during the reference scenario has a much higher average (4.66) as compared to the scenario with ASR (3.33). This might be due to the fact that the instructions integrated into the ASR system seem more useful for the TWR role, therefore a more evident support of ASR can be observed. Also, according to the controllers, the 'Hook' function supported with maintaining a low level of workload and improved situational awareness (see also situational awareness graphs under CRT-05.972-TLR4-TVALP-H106-2012).

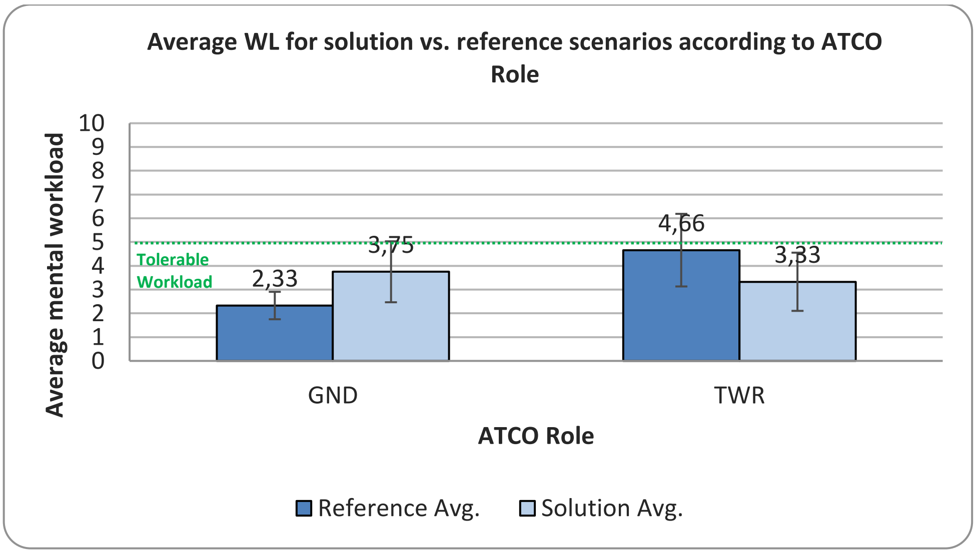


Figure F-184: Post-run questionnaire: Average mental workload for solution vs reference scenarios according to ATCO role

During the post-run questionnaires, ATCOs were also asked to fill in a NASA TLX questionnaire, for which a scale ranging from 1 to 20 (Very Low (1) to Very High (20)) was used. The responses to the questionnaire (See Figure F-185~~Figure F-185~~) also indicate homogeneity between scenarios for a majority of elements on the questionnaire, similar to the Bedford questionnaire responses. For the 'frustration' element, ATCOs reported low values, ranging on average from 2 to 2.5 for all scenarios. The 'effort' element of the questionnaire also shows low average values, ranging from 3.5 to 4.8.

For the 'Performance' element, the questionnaire scale was inverted and ranged from Perfect (1) to Failure (20), in order to match the rest of the elements, meaning that the lower the value assigned, the better the results. Most average reported values were generally positive, however a drop in performance can be observed in Sol. 1 (8.3) as compared to Sol. 2 (2.5) and Sol. 3 (5.5). This might be due to Sol.1 being the first ASR scenario for each simulation day and as such, ATCOs needed some extra time to adjust to the ASR system. By comparison, a slightly less positive performance value can be observed for the Reference scenario (5.5) which might be due to the benefits that the 'hook' function brings in the scenarios with ASR. In terms of the 'temporal demand', the average values are also low for all scenarios, ranging from 3.2 to 4.8. By comparison, a higher 'physical demand' is observed for Sol. 3, which could be motivated by the fact that ATCOs were asked to push the ASR tool to its limit, testing also the performance of context-based data. A higher 'Mental demand' is observed for the Ref. scenario as compared to the rest of the scenarios, which might also be motivated by the 'Hook' function supporting ATCOs in the solution scenarios.

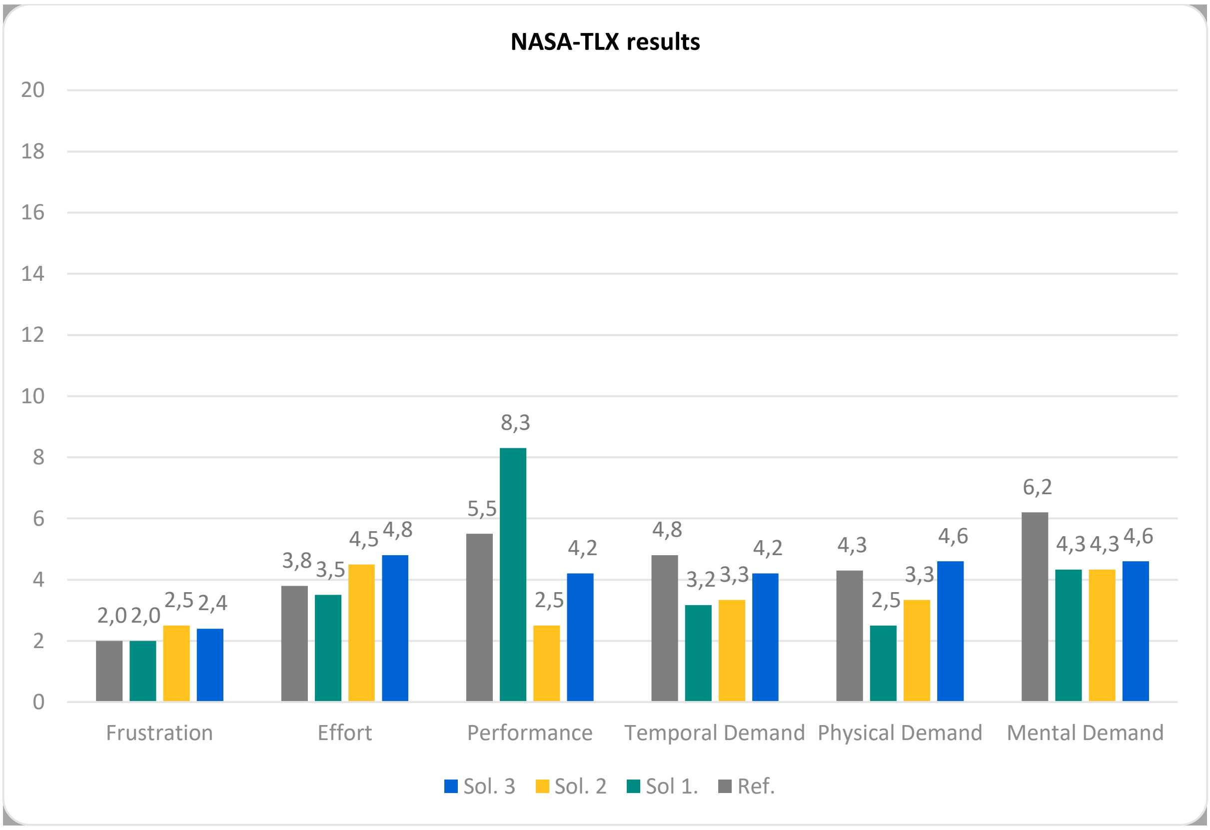


Figure F-185: Post-run questionnaire: Mental workload with NASA TLX

The overall NASA-TLX average value for the reference versus solution scenarios can be seen in the graph below. It can be observed that the average WL for the scenarios with ASR (3.7) is lower that the average WL calculated for the Reference scenario (4.4). There is no significant difference between the standard deviation values for the two data sets (SD for the reference scenario is 5.9 while for the solution scenario is 4.8).

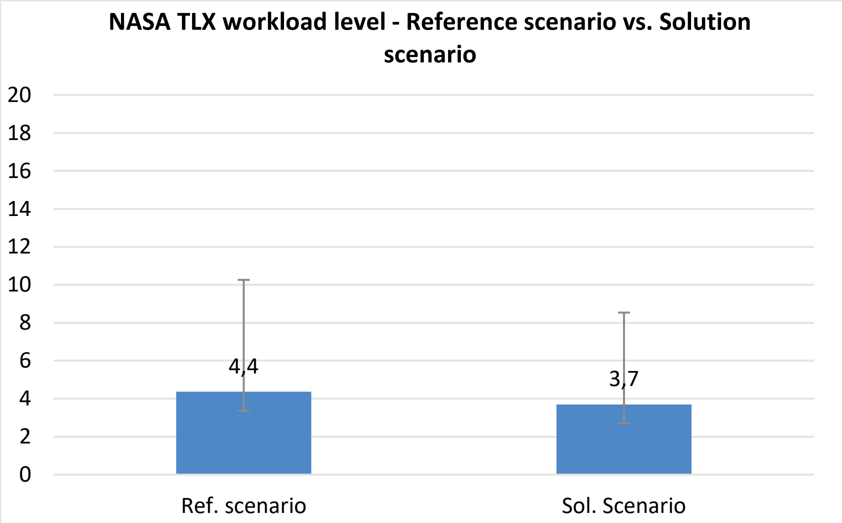


Figure F-186 - Post Run Questionnaire: Average WL per reference versus solution scenarios

In the post-exercise questionnaire, 33% of ATCOs reported 'acceptable' workload in the ASR scenarios compared to the reference scenarios, while 67% of them reported light workload levels. Such a finding is also represented in Figure F-187~~Figure F-187~~ below which shows the distribution of answers on the 7-point Likert scale, expressed in percentage.

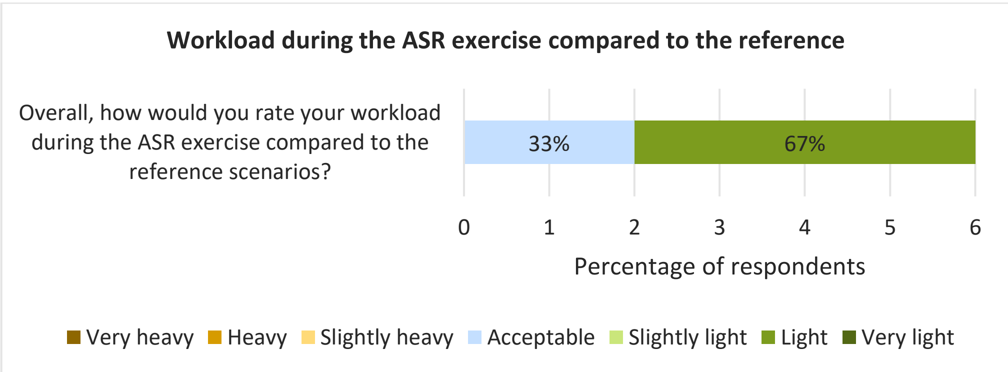


Figure F-187: Post-exercise questionnaire: workload during the ASR exercise

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. All controllers were in agreement, with answers ranging from 'somewhat agree' to 'strongly agree'. This is also represented in Figure F-188~~Figure F-188~~ below, which shows the distribution of answers on the 7-point Likert scale, also expressed in percentage.

During the post-run debriefings for the scenarios with ASR, several controllers mentioned that the level of workload was deemed low and that factors that contributed to that were the support provided by the 'Hook' function, as well as the recognition rate of the tool and the low latency.

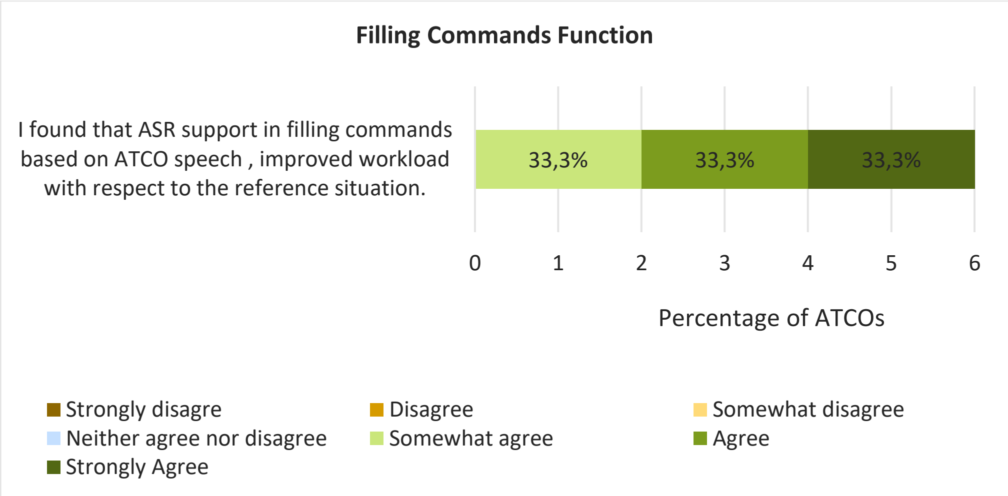


Figure F-188: Post-exercise questionnaire: workload improvement due to the 'filling commands' function

CRT-05.972-TLR4-TVALP-H106-2012

Majority of ATCOs (at least 75%) responses is that ASR supports ATCO in maintaining an adequate level of situation awareness.

In the post-run questionnaire, 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[4]) that better represents their experience during each run. No significant difference was observed between the reference and the three scenarios, in terms of average experienced frequency, for any of the situational awareness statements, as shown in Figure F-189~~Figure F-189~~ below. This means that all three scenarios can be considered consistent, with no significant drops in Situational Awareness compared to the reference scenario.

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

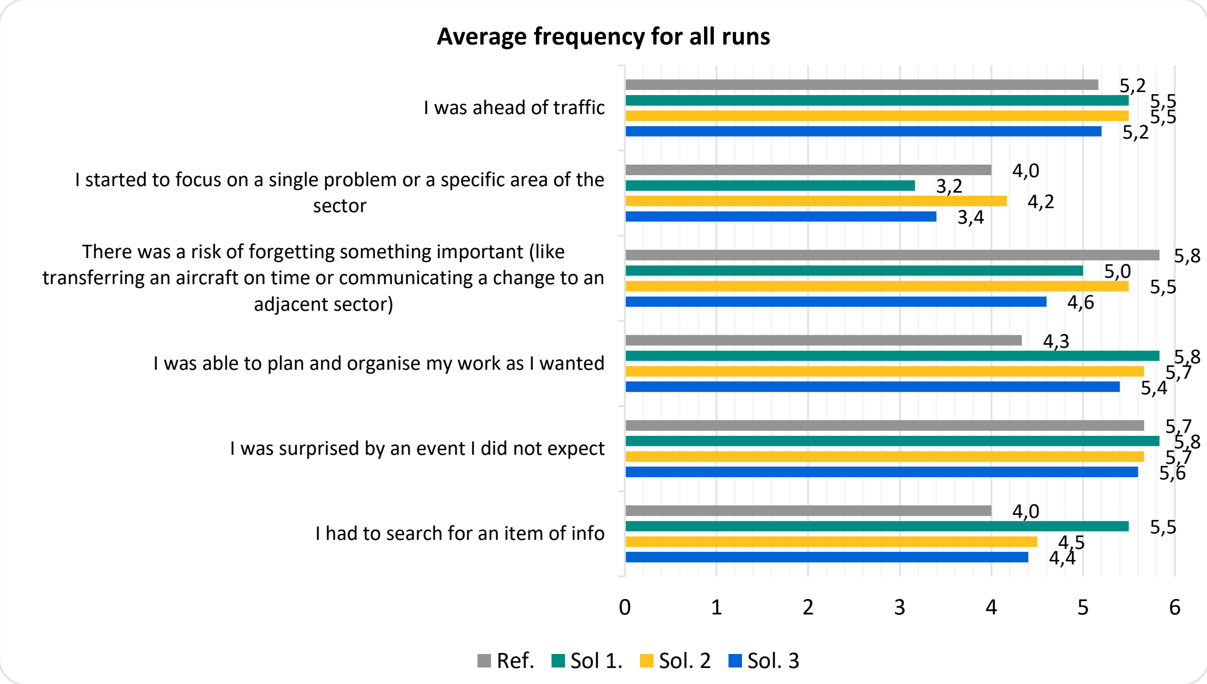


Figure F-189: Post-run questionnaire: Average experienced frequency per type of scenario for each statement on the SASHA questionnaire

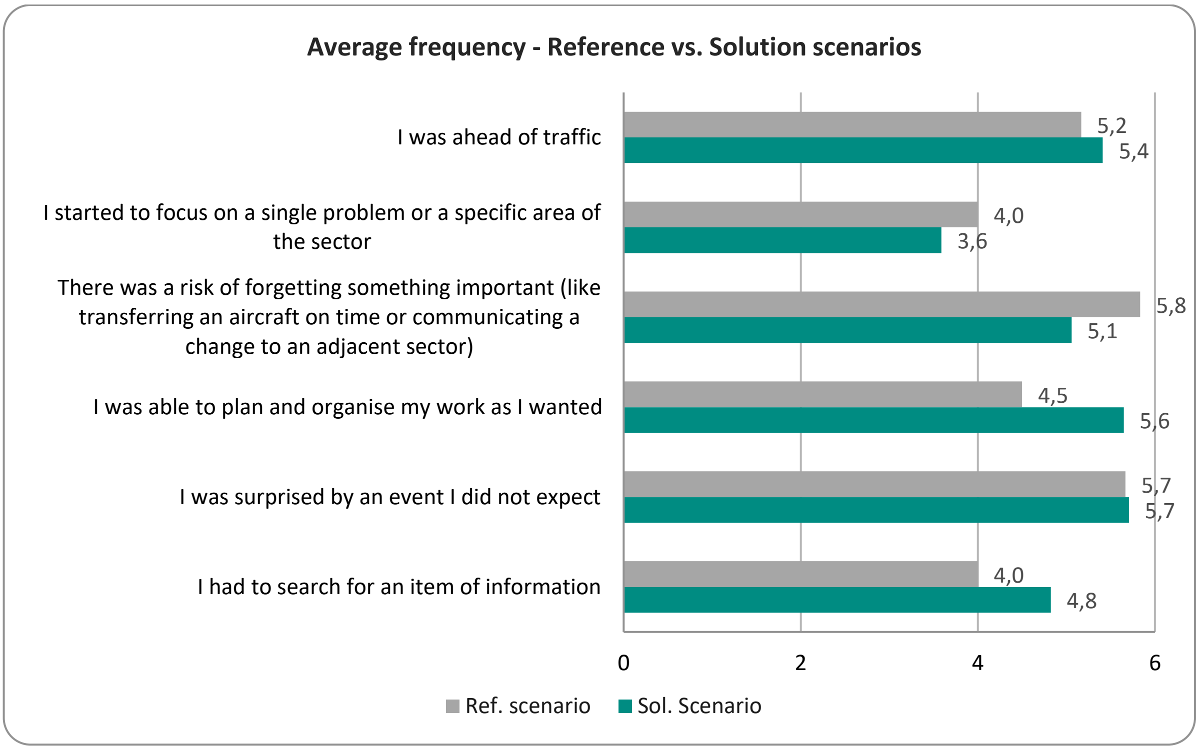


Figure F-190: Post-run questionnaire: Average reported frequency for Situational Awareness Statements comparing Reference and Solution scenarios

The figure below shows the average situational awareness calculated for the reference scenarios, in comparison with the solution scenarios altogether. An increase in situational awareness is observed for the solution scenario (5 out of 6) as compared to the average SA in the reference scenario (4.8 out of 6). There is no significant difference between the standard deviation values for the two data sets (SD 1.5 for the reference scenario compared to SD 1.4 to the solution scenarios).

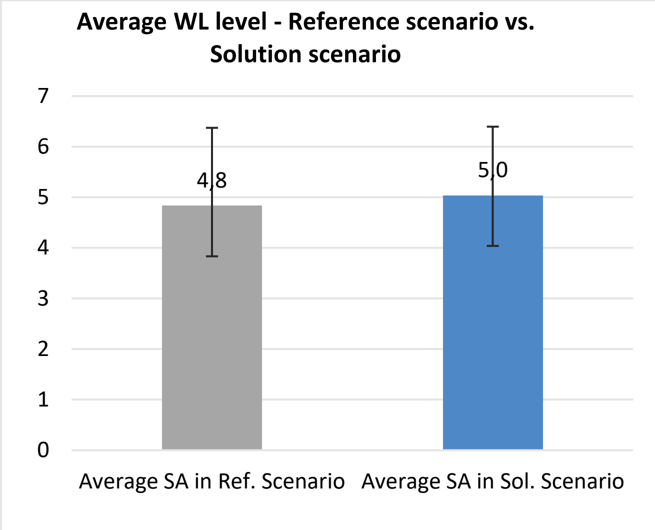


Figure F-191 - Average WL level - Reference scenario vs. Solution scenario

When comparing the reference to solution scenarios, on average, ATCOs said that they were very often ahead of traffic, with little difference between reference (5.2) and solution scenarios (5.4). This means that using ASR does not negatively impact the ATCOs ability to foresee future traffic situations.

There was little difference between the reference and solution scenarios in ATCOs responses, with an average of (4) and (3.6) respectively when asked if during the run, they started focusing on a single problem or a specific area of the sector. This means that ASR does not particularly distract the ATCOs as compared to when controlling traffic without using ASR.

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. In the solution scenarios, the 5.1 value indicates that ATCOs feel they would rarely 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 4.5 is obtained. However, a much higher value was reported for the solution scenario (5.6), which means that when using ASR, ATCOs feel like they can better plan and organise their work.

When asked if they were surprised by an unexpected event, an average frequency of 5.7 is reported by ATCOs for both reference and solution scenarios which indicates that such a situation never occurs and that there is no difference between the two scenarios.

An average frequency of 4 was reported when ATCOs were asked if they had to search for an item of information during the Reference scenario; however, a higher value can be observed for the solution scenario (4.8). This means that when using ASR, ATCOs are provided with more required information which they would normally have to look for. See Figure F-192~~Figure F-191~~ below for the average frequency reported in each one of the questions in the SASHA questionnaire, for both reference and solution scenarios joined together. The same figure below shows 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[5]

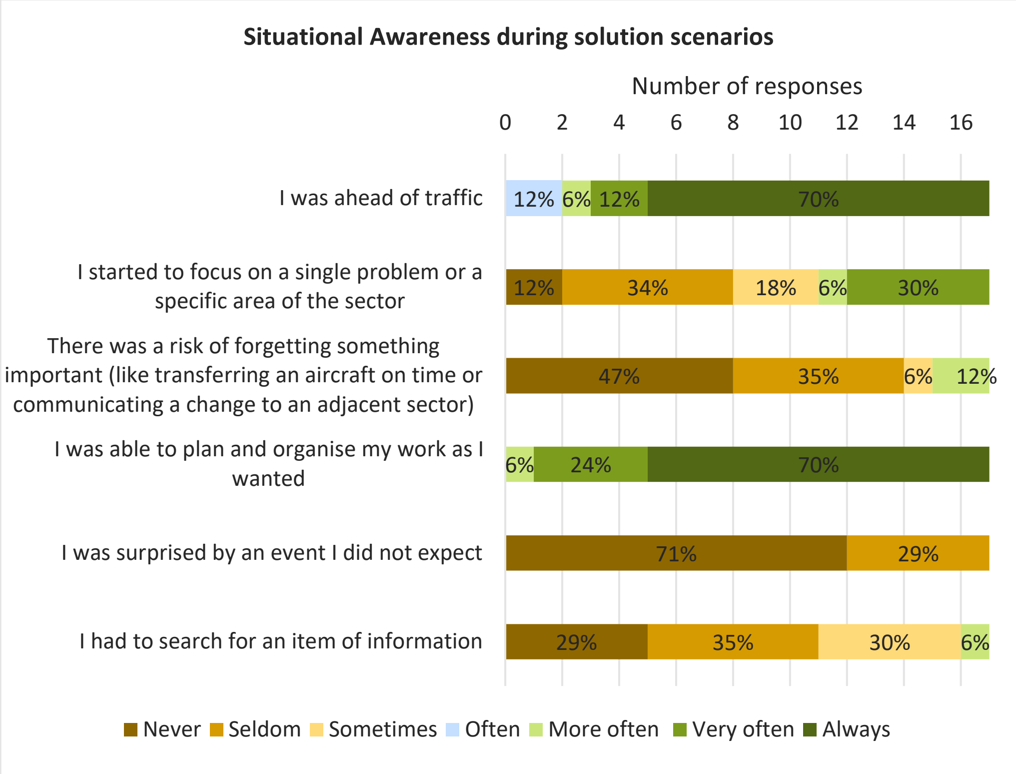


Figure F-192: Post-run questionnaire: Situational Awareness statements during solution scenarios

When looking at situational awareness related average frequency experienced by the controllers according to each ATCO role in the solution scenario in the figure below, it can be observed that there is no significant difference in experienced situational awareness, for any of the six statements in the SASHA questionnaires.

[5] Please note that the percentage calculation is an approximated value, in order to avoid cluttering the chart

Fehler! Verweisquelle konnte nicht gefunden werden.

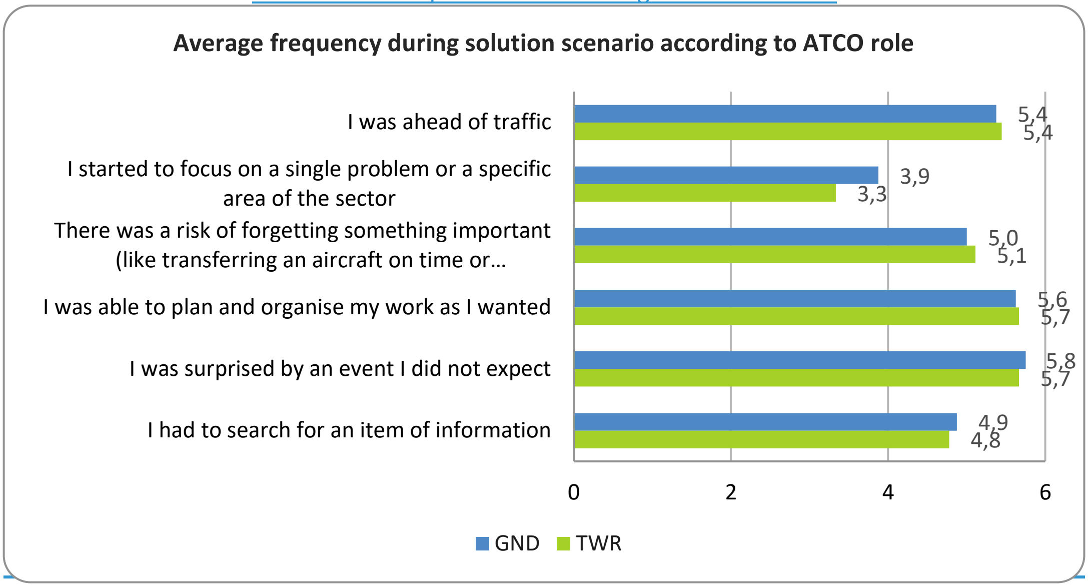


Figure F-193: Post-run questionnaire: Average reported frequency for Situational Awareness statements per ATCO role

In the post-exercise questionnaire, ATCOs were also asked if the ASR "Hook" function used to highlight a desired call sign improves situational awareness with respect to the reference situation: answers ranged from ‘agree’(2) to ‘strongly agree’ (4). This means that all ATCOs were in agreement that the ‘Hook’ function improves situational awareness. See Figure F-194~~Figure F-194~~ below for the number of answers collected for each point of the 7-Point Likert Scale, also expressed in percentage.

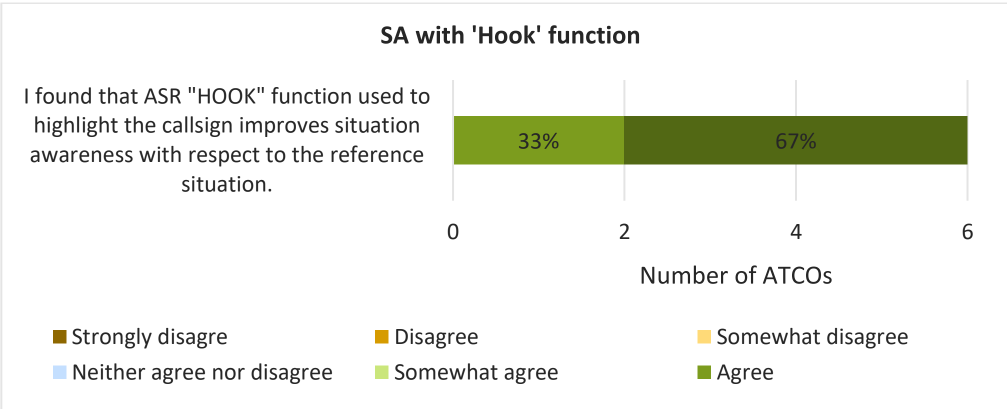


Figure F-194: Post-exercise questionnaire: Situational awareness improvement with ‘Hook’ function

In the post-exercise questionnaire, out of the six ATCOs participating in the simulation, four rated the overall Situational Awareness during the ASR exercise as *High*, while the other two rated it as *Perfect*. This means that Situational awareness was satisfactory for all ATCOs when using ASR. See Figure F-195~~Figure F-194~~ below for the number of answers collected for each point of the 7-Point Likert Scale, as well as their percentage.

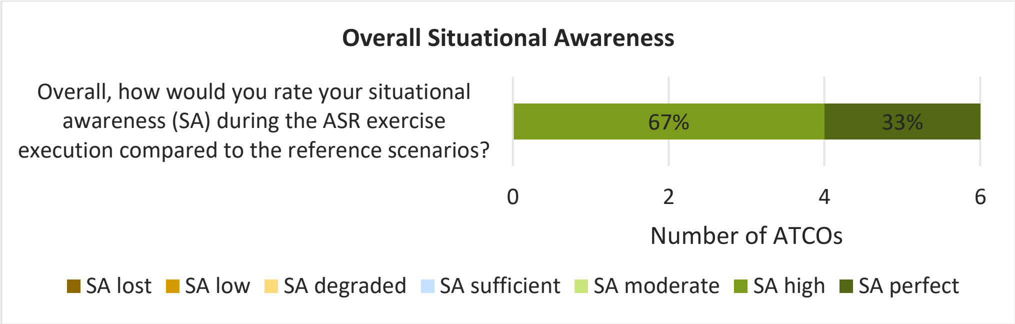


Figure F-195~~194~~: Post-exercise questionnaire: overall situational awareness during the ASR exercise

During the debriefing sessions, ATCOs provided positive feedback with respect to situational awareness when using ASR. Some of them mentioned that the "hook" function was very reliable and that it supports them maintaining a high situational awareness level and a low level of workload. However, one ATCO mentioned that a more evident highlighting of the a/c on the HMI would be useful, when using the function. Another ATCO commented that this function could also be useful also in identifying a/c by their a/c types.

CRT-05.972-TLR4-TVALP-H106-2013

Four out of six ATCOs agreed that ASR did not increase potential for human error compared to current operations, whereas one ATCOs somewhat disagreed, and one strongly disagreed. See Figure F-196~~Figure F-195~~ below for the number of answers collected for each point of the 7-point Likert scale, also expressed in percentage.

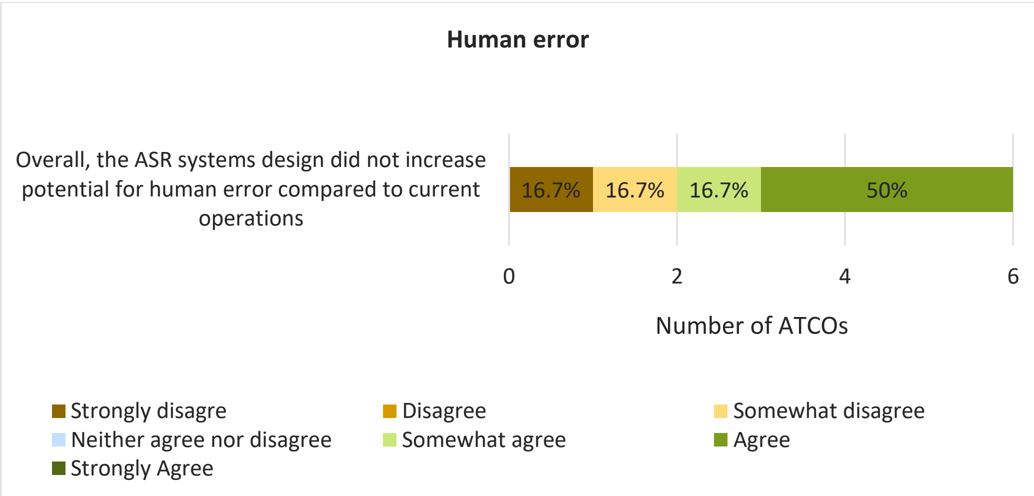


Figure F-196~~195~~: Post-exercise questionnaire: responses for potential increase of Human Error

During the debriefings, ATCOs provided positive general feedback for the ASR system, and no negative comments were given when asked about the system's potential to increase human error.

CRT-05.972-TLR4-TVALP- H106-2014

ATCOs (at least 75%) provide positive feedback on adequacy (level and quality) of information provided by ASR.

In the post-exercise questionnaire, five out of six ATCOs agreed that the level of feedback support provided by ASR was adequate and clear, and did not disturb them, while one of them strongly agreed. See Figure F-197~~Figure F-196~~ below for the number of answers collected for each point of the 7-point Likert scale, also expressed in percentage.

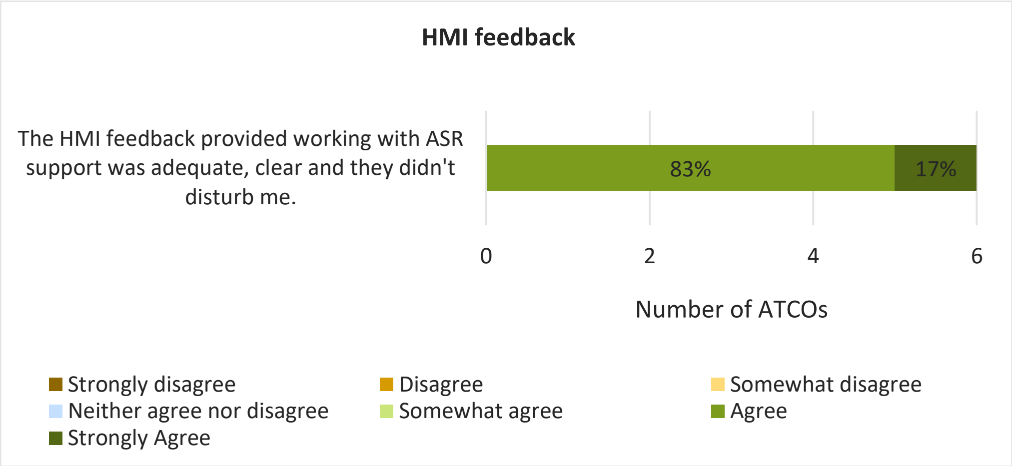


Figure F-197~~196~~: Post-exercise questionnaire: Adequacy of HMI feedback provided

Further discussion during the debriefings revealed that the ASR pop-up window was not obvious to see and could sometimes require the controllers to look for it, partly because of its colour coding, partly because it would not appear consistently in the same place, but in the last place where the mouse cursor was left. As a result, some controllers expressed the desire to have an 'ASR' pop-up window which would display useful information situated always in the same place, similar to a chat window.

CRT-05.972-TLR4-TVALP- H106-2015

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 Figure F-198~~Figure F-197~~ below. Generally, positive results can be observed for all statements, with acceptable ASR callsign and command recognition rate by all ATCOs.

Most ATCOs found the callsign and command rejection rates as acceptable (4), while (2) ATCOs were neutral about it. 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 the comments section, one ATCO mentioned that they almost didn't have to correct any commands while using ASR.

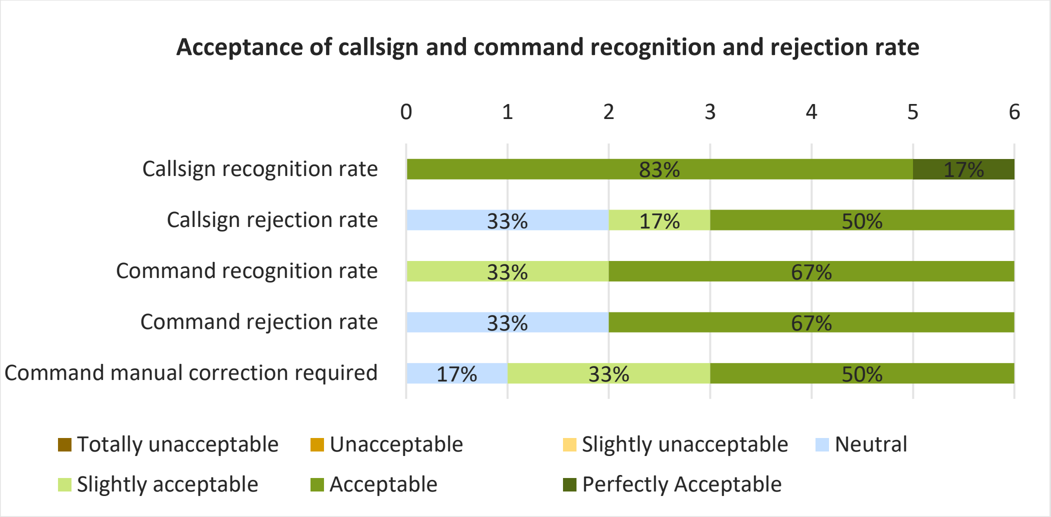


Figure F-198~~197~~: Acceptance of callsign and command recognition and rejection rate

During the debriefings, controllers mentioned they were happy with the ASR recognition rate.

The following table shows the text-to-concept recognition rates for all utterances used for the duration of the exercise:

COMMAND (all 9 solution runs)	COMMANDS ISSUED	RECOGNISED COMMANDS	COMMAND RECOGNITION RATE
hook	83	71	87%
taxi	136	95	70%
startup	55	37	67%
pushback	9	8	88%
lineup	45	39	87%
takeoff	59	38	64%
clear to land	67	57	85%

Table F-31: Command type recognition rate per command type

A pictorial representation of Table **F-31**

~~Table F-31~~ above can be seen in Figure F-199~~Figure F-198~~ below:

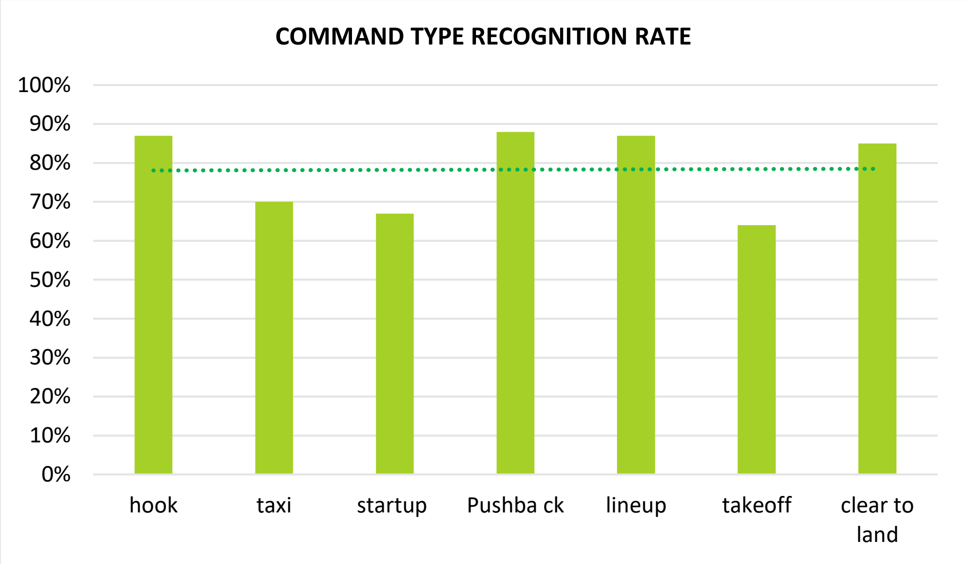


Figure F-199~~198~~: Recognition rate per command type

The table below shows the total number of commands issued throughout the validation week (454), and the total number of recognised commands (345), which is also expressed in percentage (overall command recognition rate is 76%). The overall command recognition rejection rate is also listed in the table (24%), referring to every time ATCOs uttered a command into the ASR system and no command was presented on the screen or track was highlighted.

Command recognition error rate is also listed in the table (2.9%), referring to the percentage of occurrences of a command incorrectly recognized by ASR system yet displayed on the HMI pop-up window.

TOTAL # COMMANDS ISSUED	TOTAL # RECOGNISED COMMANDS	OVERALL COMMAND RECOGNITION RATE	COMMAND RECOGNITION ERROR RATE	COMMAND RECOGNITION REJECTION RATE
454	345	64.6%	5.1%	35.4%

Table F-32: Overall command recognition, error and rejection rates

The table below shows the number of commands issued per each role, together with the percentage of recognised commands. No significant difference is seen in command recognition between the two ATCO roles:

ATCO ROLE	# COMMANDS ISSUED	# COMMAND TYPE RECOGNISED	RECOGNITION RATE
GND	284	222	77%
TWR	167	124	74%

Table F-33: Number of commands issued per ATCO role and their type recognition rates

Resulting in an overall **Command Type Recognition Rate** of **75.9%.**

The table below presents the command recognition and rejection rates, per solution. A high overall recognition rate can be observed, however an significant increase in performance is seen from Sol. 1 to Sol. 2, which might be due to ATCOs increased familiarity with the ASR system through the simulation day. However, a slight decrease is noticed for Sol. 3 which could be related to the fact that ATCOs were asked to put stress on the system in this scenario, to test the context data.

SOLUTION	OK	NOK
SOL 1	68%	32%
SOL 2	83%	18%
SOL 3	73%	27%

Table F-34: Command recognition/rejection rate per solution

Please see the table below for average callsign recognition rates per solution. Again, the callsign recognition rate increases throughout the simulation day, potentially as a result of ATCOs increased familiarity. Only a slight decrease is noticed for Sol. 3, despite extra stress being put on the system. Overall callsign recognition rate (90%) is much higher than overall command recognition rate seen in Table F-32 (76%).

CALLSIGN RECOGNITION RATE	
SOL 1	85.5%
SOL 2	93.8%
SOL 3	90.2%
OVERALL RECOGNITION RATE	89.8%

Table F-35: Average callsign recognition rate

Delay

The average delay for all recognised commands was calculated as (1.3 ± 0.8) s while for the rejected commands was (2.1 ± 1.4) s; on top of which (18 ± 5) ms should be added, due to average network delay.

CRT-05.972-TLR4-TVALP-H106-2016

Majority of ATCOs (at least 75%) confirm adequate usability of ASR system.

In the Post-Exercise Questionnaire, all ATCOs agreed that they would like to use ASR frequently. See Figure F-200~~Figure F-199~~ below for the number of answers collected for each point of the 7-point Likert scale also expressed as a percentage. Five out of six ATCOs disagreed that the system was unnecessarily complex, while one somewhat agreed to it. All ATCOs agreed that ASR was easy to use, its functions were well integrated, and that people would learn quickly how to use the system. Three ATCOs indicated that they would need the support of technical personnel to be able to use the system and only one out of six agreed that there were too many inconsistencies in the ASR system, while the remaining ATCOs disagreed (4) or were neutral (1). One of the ATCOs found the system cumbersome to use, while the rest (3) disagreed or were neutral (2). All ATCOs were confident in using the system. When asked if they needed to learn a lot of things before using ASR, most ATCOs disagreed (5) and only one agreed.

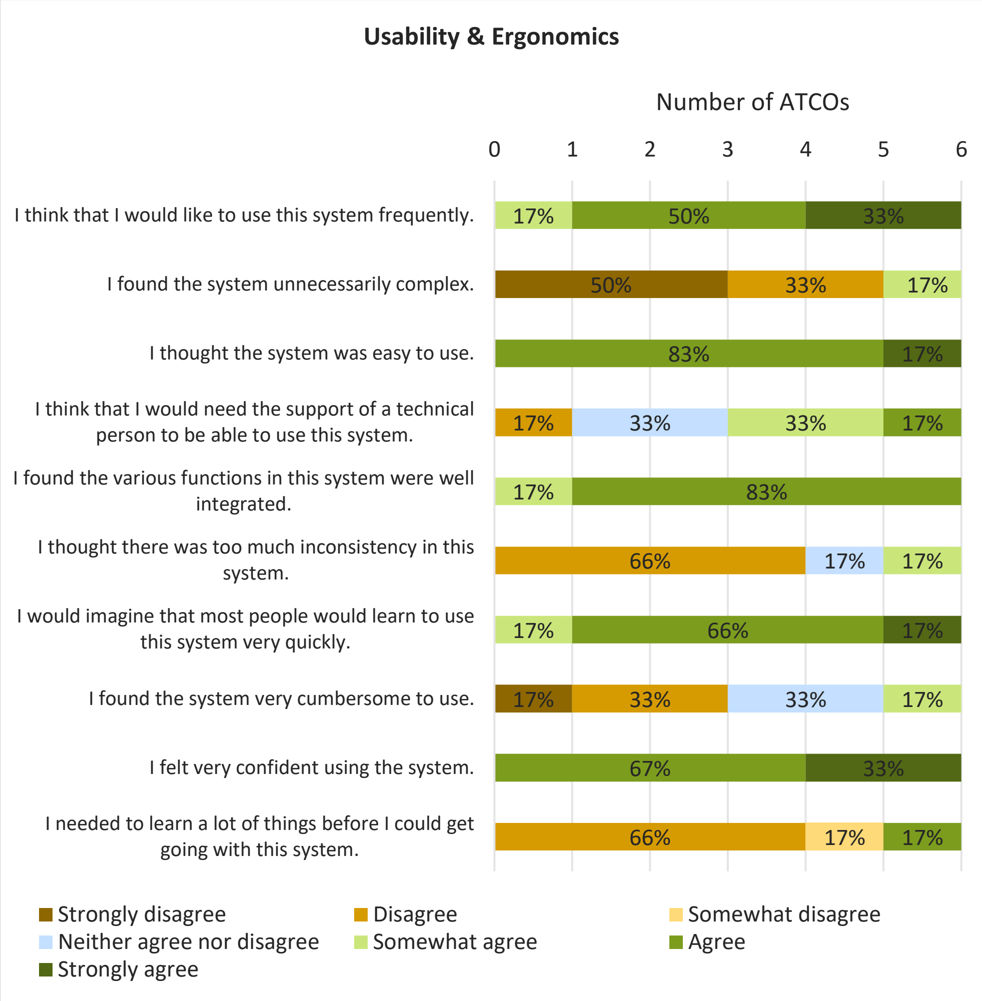


Figure F-200~~199~~: Post-exercise questionnaire: usability (SUS) questionnaire responses distribution per statement

In terms of the use of the “Hook” function, all ATCOs agreed that they use it when arriving traffic called for info or request (e.g. a runway change). Most ATCOs (5) agreed that they use the function with identifying flights in complex and dense traffic and when they needed to manage inbound traffic and deviating traffic while one ATCO disagreed in both cases, which suggests they are not using the function in that way. Five out of six ATCOs agreed to use the “Hook” function to identify flights when entering their sector while the remaining one is neutral (neither agree nor disagree). When asked if they used the function in case there was a voice communication exchange with another sector, there was no agreement between the ATCOs. Two ATCOs strongly agreed, one agreed, two of the ATCOs neither agreed or disagreed and the remaining controller disagreed. The difference in responses might be due to the different controller positions having different roles and tasks which might require the same function in different ways.

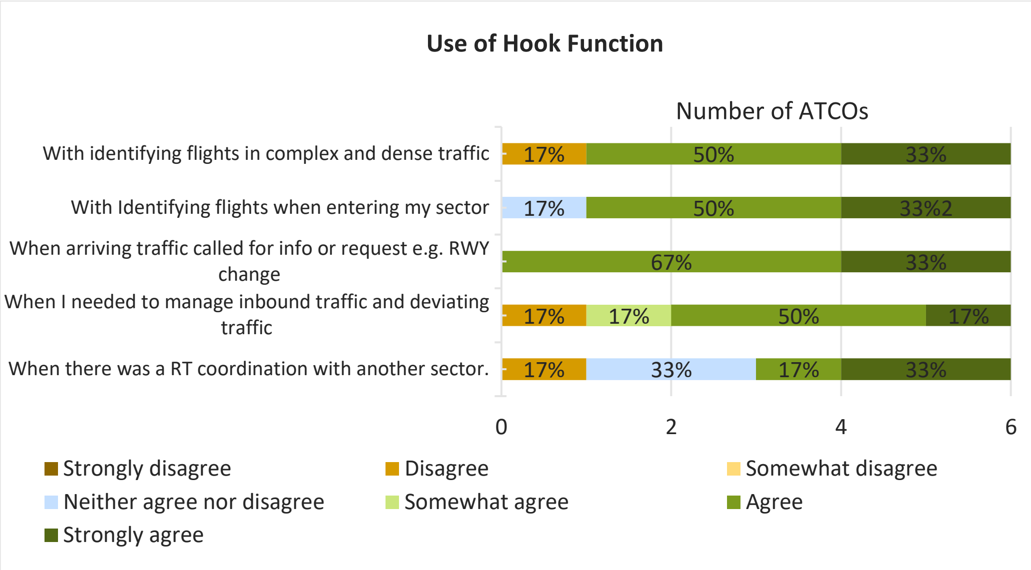


Figure F-201~~200~~: Post-exercise questionnaire: use of "Hook" function

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 routinely ATC tasks, without impacting ASMGC-S. As it can be seen in the diagram below, there was no general agreement between ATCOs. 50% of them prefer to use ASR only in routinely tasks, while the other 50% would use it for more complex tasks. Two of the controllers said that they would only use ASR for callsign recognition, one was neutral about it and the remaining 50% of ATCOs either disagreed (2) or strongly disagreed (1) which means they would also like to use ASR for more complex tasks. One of the ATCOs commented that ASR is very useful in finding potential conflicts when issuing clearances.

The divided opinions related to ASR use could be related to ATCOs experience or roles covered, which might involve different operational needs. During the debriefing, one ATCO mentioned they would like to use ASR for coordination between ground and tower and as a safety barrier by means of automatically filling in the information on the eFlight strips (landing, take-off) while doing a cross check of it at the same time.

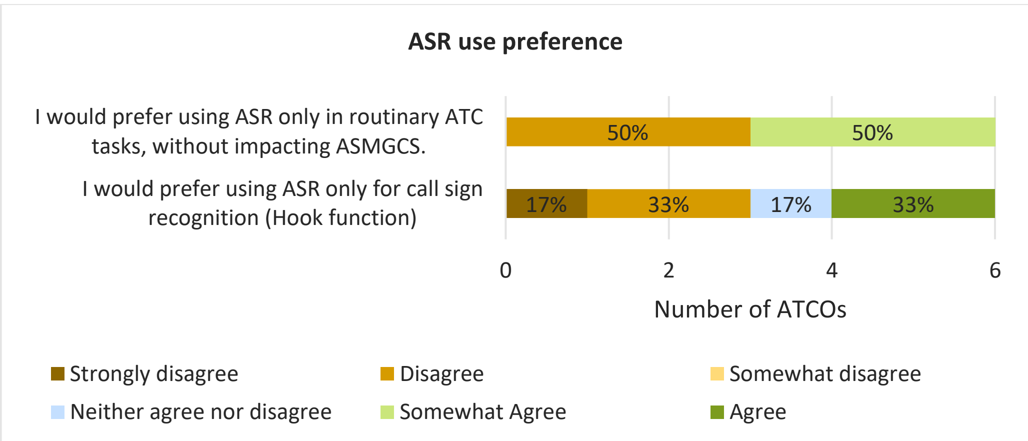


Figure F-202~~201~~: Post-exercise questionnaire: ASR use preference

When asked to rate the overall usability of the system, good agreement can be observed between ATCOs. One ATCO rated the usability as *Very high*, four ATCOs rated it as *High* and one rated it as *Moderate*. This means that on the whole, ATCOs were happy with ASR usability. Figure F-203~~Figure F-202~~ below shows the distribution of responses with regard to overall usability.

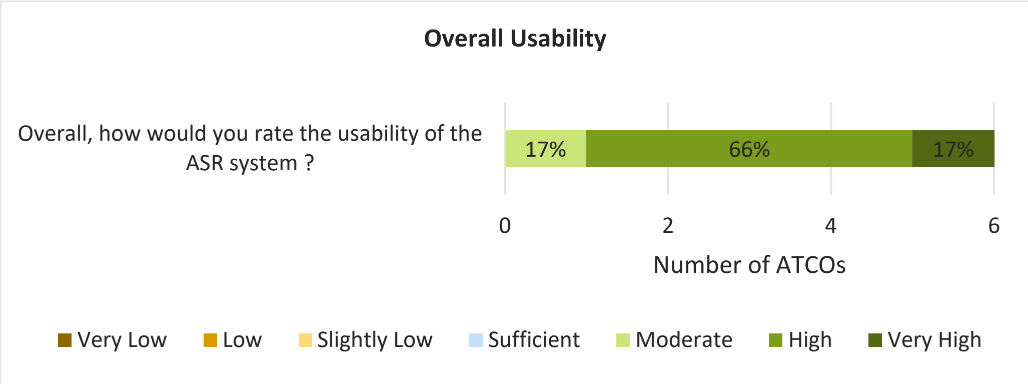


Figure F-203~~202~~: Post-exercise questionnaire: overall usability rating

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, (5) ATCOs strongly agreed on the need of dedicated training while the remaining ATCO agreed.

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. taxi to stand, taxi right/left, vacate, etc.

Moreover, 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.

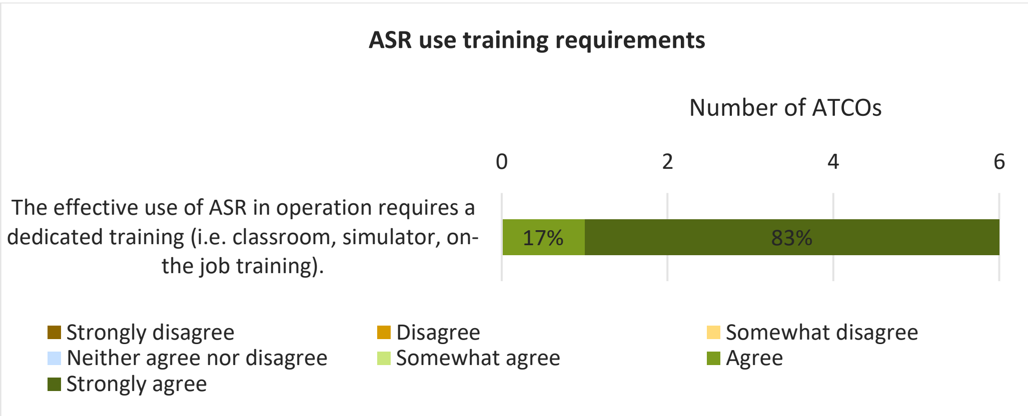


Figure F-204~~203~~: Post-exercise questionnaire: ASR use training requirements

In the debriefings, ATCOs mentioned that training on the system would be useful, especially on how to better use the standard phraseology for a higher recognition rate and on HMI aspects, such as information displayed on the potential ASR dedicated pop-up window they initially suggested to implement (as also shown in the results from CRT-05.972-TLR4-TVALP- H106-2014).

CRT-05.972-TLR4-TVALP- H106-2017

Majority of ATCOs (at least 75%) provide positive feedback on acceptance of ASR tool.

In the CARS post-run questionnaire, ATCOs reported an average Acceptance level of 8.1 for the ASR tool.

See the figures below for the number of answers collected for each point of the 10-point CARS scale and the average acceptance level for each controller position. All mean values are above the acceptable minimum of 5.

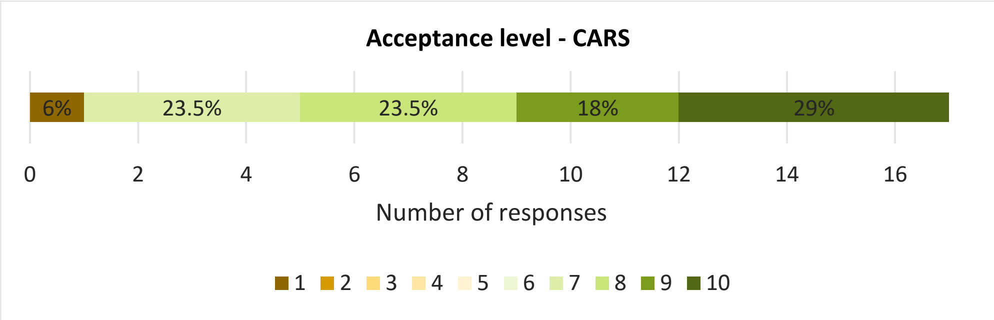


Figure F-205~~204~~: Post-exercise questionnaire: acceptance (CARS) questionnaire responses

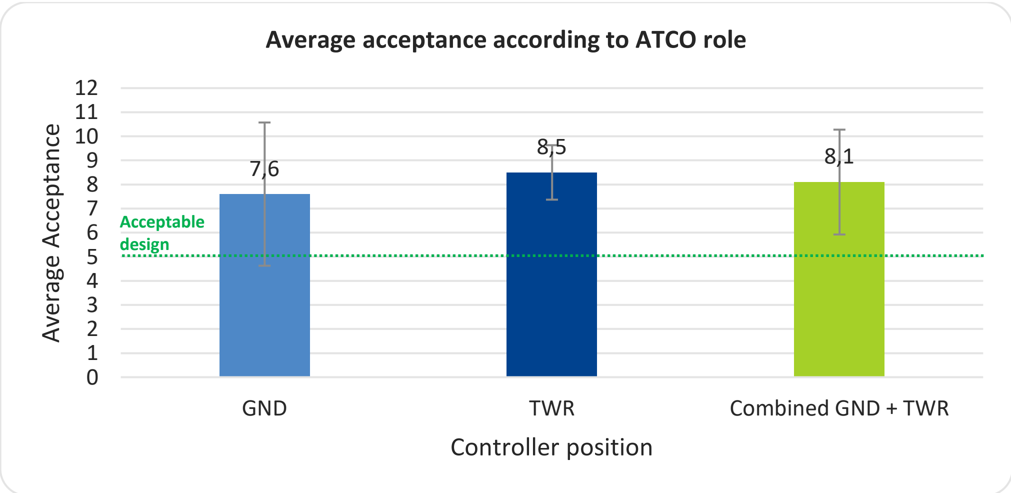


Figure F-206~~205~~: Post-exercise questionnaire: average acceptance rate according to ATCO role

In the post-exercise questionnaire, ATCOs were also asked to rate the acceptance of ASR incorrect call signs and commands frequency. As shown in Figure F-207~~Figure F-206~~ below, a general agreement can be observed between ATCOs that both the frequency of erroneously highlighted ASR call signs and the frequency of mistakenly recognised ASR commands was acceptable.

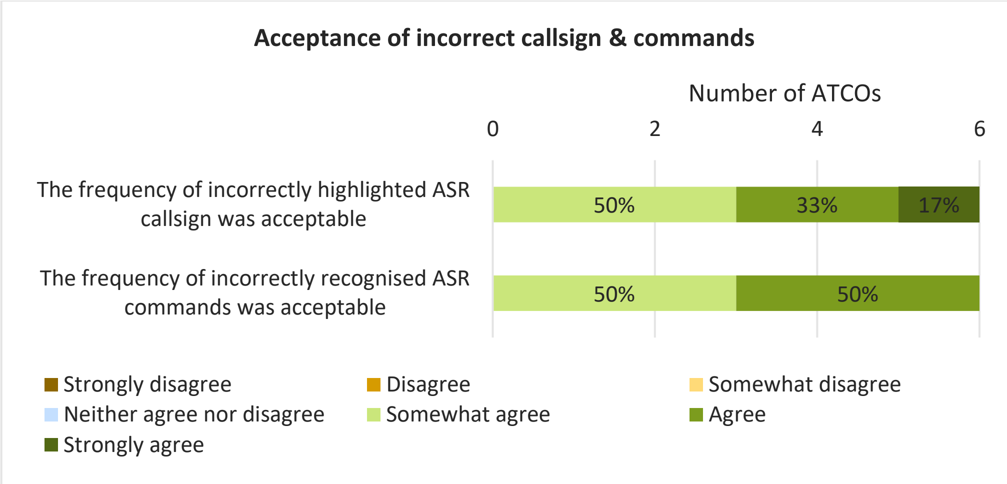


Figure F-207~~206~~: Post-exercise questionnaire: acceptance of incorrect callsign & commands frequency

ASR system latency was also explored in another post-exercise questionnaire. As it can be observed in Figure F-208~~Figure F-207~~, all ATCOs were satisfied with latency from different perspectives: command recognition, "hook" function and overall system feedback latency. Also during the debriefings ATCOs mentioned that the latency of ASR was acceptable.

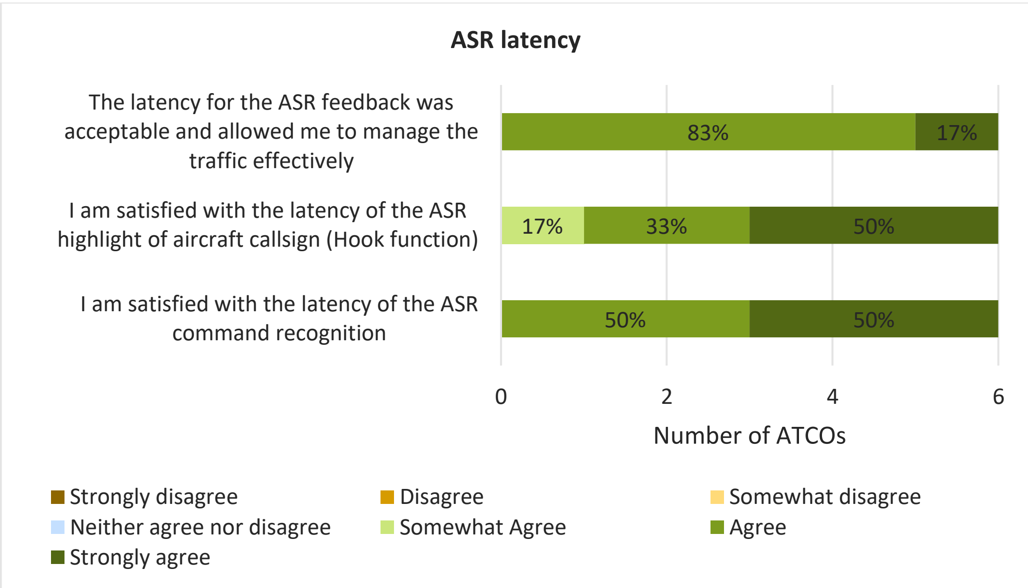


Figure F-208~~207~~: Post-exercise questionnaire: Latency of ASR response

The ATCOs were also asked about the overall interaction between ASR and the A-SMGCS system. All controllers were generally satisfied by the interaction and no issues were reported. However, some suggestions on the interaction of ASR with A-SMGCS were provided by some ATCO's as described in the WANT/HAVE matrix used for the final debriefing (Also to see F.7.1.2 - Human Performance).

Figure F-209~~Figure F-208~~ below shows the number of answers collected for each point of the 7-point scale.

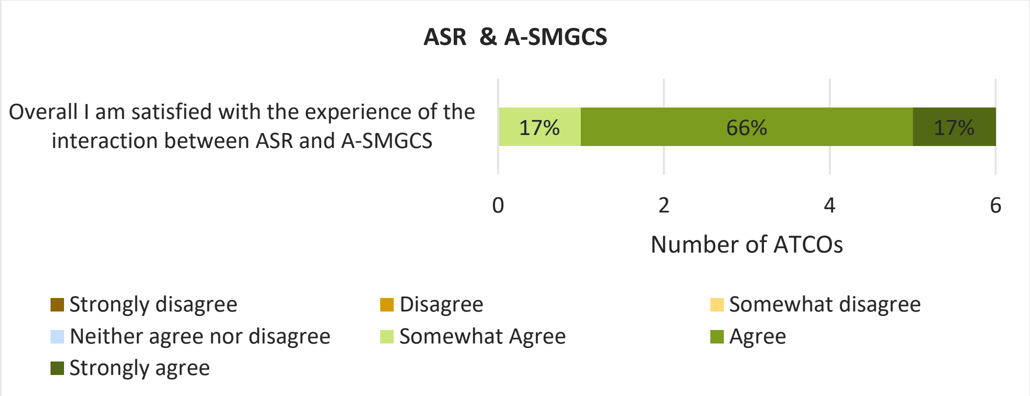


Figure F-209~~208~~: Post-exercise questionnaire: Interaction of ASR with A-SMGCS

When asked to rate the overall acceptance of the system in the post-exercise questionnaire, a general agreement can be observed between ATCOs. (5) ATCOs rated the acceptance as High, and one rated it as 'Moderate'. This means that overall, ATCOs were happy with the ASR system.

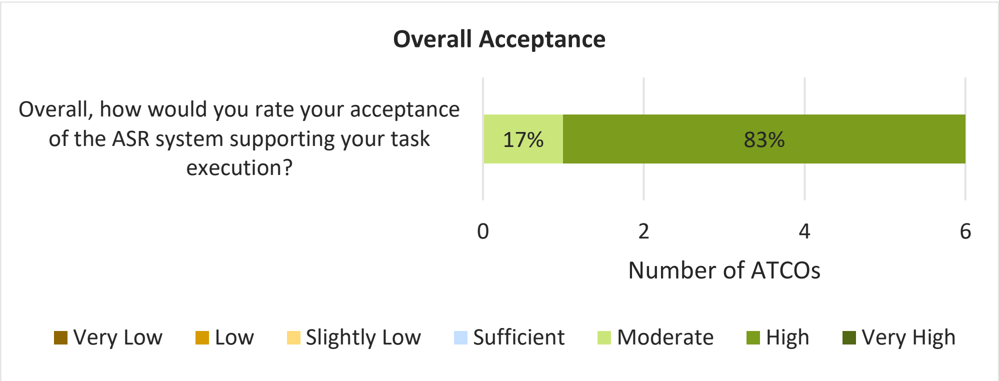


Figure F-210~~209~~: Post-exercise questionnaire: Overall Acceptance of the ASR system

The debriefing sessions revealed a high ATCO acceptance of the system, in particular due to its high recognition rate and low latency.

CRT-05.972-TLR4-TVALP- H106-2018

ATCOs (at least 75%) trust in the system is at an acceptable level.

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 frequency they assessed 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.

No significant difference was observed between the reference and the three solution scenarios, in terms of average experienced frequency, for any of the trust statements, as seen in the figure below. Therefore, the three solution scenarios can be considered homogeneous, with no significant drops in Trust as compared to the reference scenario.

Figure F-211~~Figure F-210~~ below illustrates the average for each SATI statement, for the reference scenario responses against solution scenarios responses joined together. Although the data sample for the reference scenario is smaller than the one for solution scenarios joined together, applying Welch's t-test shows that the mean values are not statistically different.

On average, ATCOs found the system useful with little difference between reference (4.5) and solution scenarios (4.8). A slightly higher value can be observed for the solution scenario, meaning that the ASR system was found useful more frequently as compared to the reference scenario. A similar difference was noticed for the reliability of the system, where the average frequency for the solution scenario was 4.5 as compared to the reference scenario, rated with 4.2. Similarly, this means that the ASR system was found reliable more often as compared to the reference scenario.

When asked if the system worked accurately, the average ATCO responses obtained were very similar for both reference (4.1) and solution scenarios (4.0), with a frequency of 'more often' on the seven-point Likert scale. An average frequency of 3.8 was reported when ATCOs were asked if the system was understandable in the reference scenario. A slight increase was noticed for the solution scenario (4.3) which means that ARS4ATC is easily understood more frequently than the current systems not using it.

When asked if the system worked robustly, the average ATCOs responses were 3.2 (often) in the reference scenarios. However, in the solution scenarios, the average value of 3.8 (more often) indicate that ATCOs feel that the system with ASR works more robustly as compared to controlling traffic without it.

Most ATCOs were very often confident in working with the system, in both reference and solution scenarios, with an average of 4.8 and 4.7 respectively. See Figure F-212~~Figure F-211~~ below for the average frequency for both reference and solution scenarios, for each one of the six statements in the SATI questionnaire.

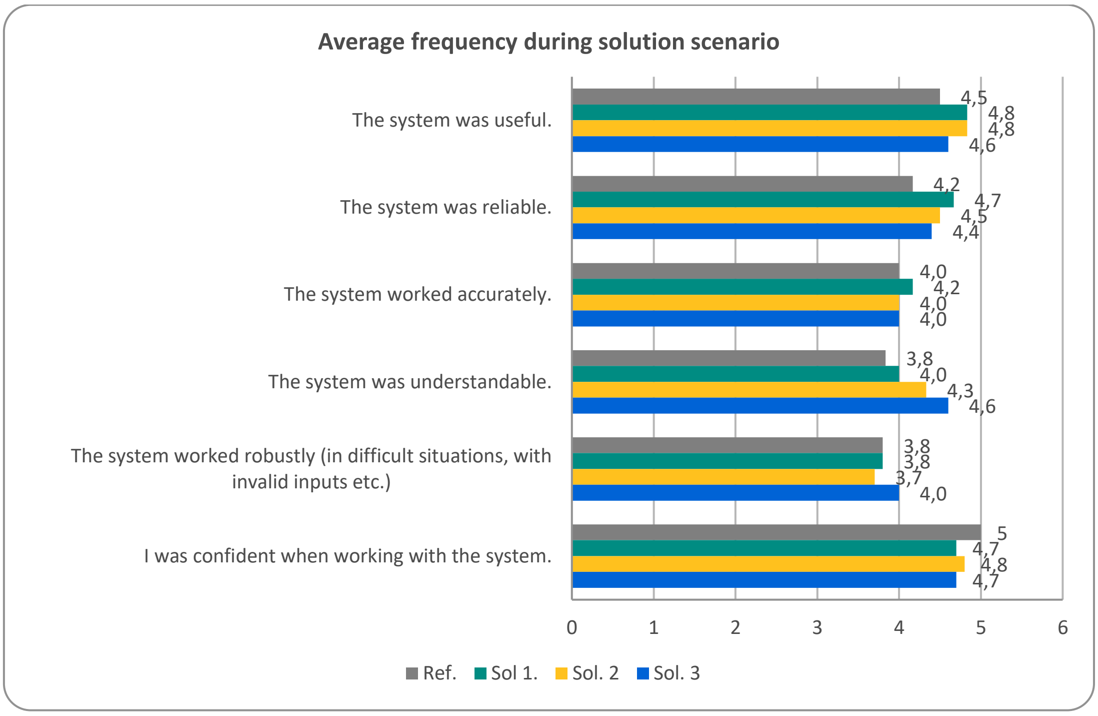


Figure F-211~~210~~: Post-run questionnaire: Average frequency per scenario type of each statement on the SATI standard questionnaire

When looking at the average frequency experienced by the controllers according to each controller working position in the solution scenario, (Figure F-212~~Figure F-211~~), it shows no significant difference in experienced trust, for most statements in the SASHA questionnaire.

However, the average frequency for 'the system was understandable' for GND controllers was reported as 4.6 (very often) as compared to TWR controllers for which the average was calculated as 4.0 (often). Similarly, when asked if the system worked robustly, an average of 4.3 (more often) was obtained for GND controllers, as opposed to TWR controllers' average of 3.4 (often). Such differences might be related to the fact that controllers had to cover different roles and therefore used ASR in a different manner, according to each role.

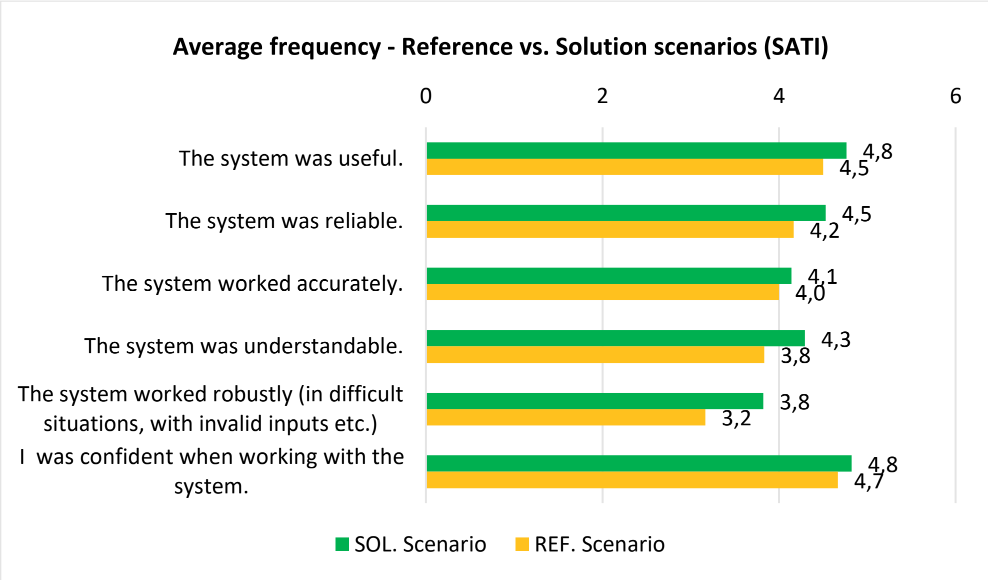


Figure F-212~~211~~: Post-run questionnaire: average frequency per SATI questionnaire statement comparing reference and solution scenarios

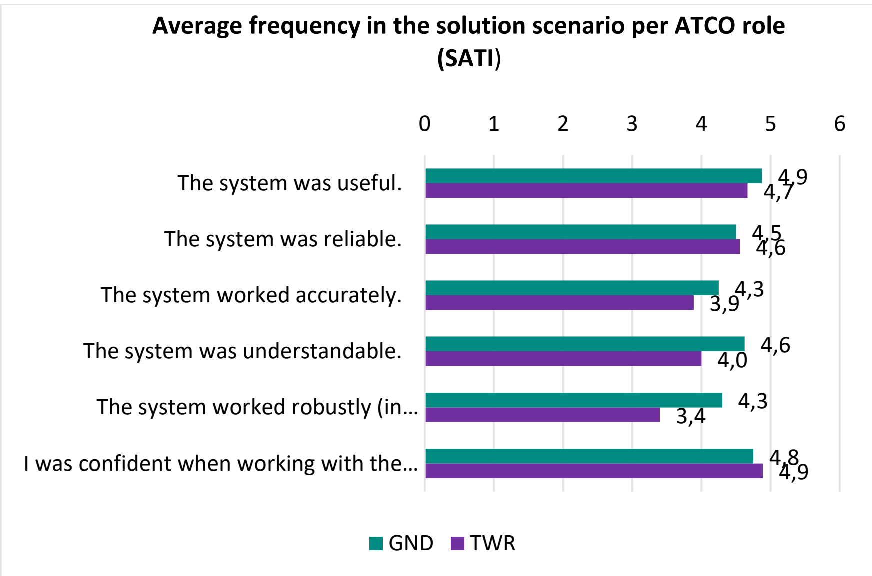


Figure F-213~~212~~: Post-run questionnaire: average frequency in the solution scenario for the SATI questionnaire statements per controller working position

In the post-exercise questionnaire, out of six ATCOs participating in the simulation, five of them rated Trust during the ASR exercise in a positive manner – one rated it *'Very High'*, three rated it as 'High', one rated it as *'Moderate'* and one as *'Sufficient'*. This indicates that the majority of ATCOs trust the ASR system. See Figure F-214~~Figure F-213~~ below for the number of answers collected for each point of the 7-Point Likert Scale.

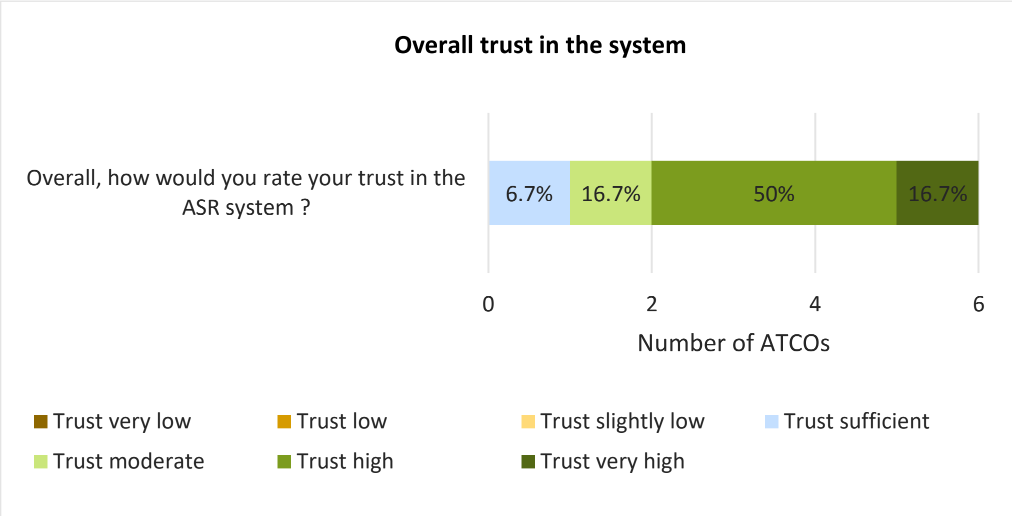


Figure F-214~~213~~: Post-exercise questionnaire: Overall trust in the ASR system

Moreover, during the debriefings, ATCOs showed a positive attitude when asked about their trust in the system. The recognition rate and the latency of the ASR were mentioned as contributing factors to the high trust levels identified during the post-run and post-exercise questionnaires.

F.7.2.4 OBJ-05.972-TRL4-TVALP-H106.2020

To assess that the role of the ATCO is consistent with human capabilities and limitations with the introduction of ASR.

CRT-05.972-TLR4-TVALP-H106-2021

The majority of ATCOs (at least 75%) responses is that ATCOs can apply operating methods in an accurate, efficient, and timely manner

During the post-run and post-exercise questionnaires and debriefings, ATCOs and were generally satisfied about the tool's latency and feedback provided (as shown in Figure F-208~~Figure F-207~~), which allowed them to apply operating methods in an accurate, efficient, and timely manner. Moreover, during the debriefings, ATCOs also mentioned they experienced no change in operating methods during the solution scenarios. This means that controllers were generally satisfied about applying operating methods when using ASR (Figure F-215~~Figure F-214~~).

CRT-05.972-TLR4-TVALP-H106-2022

The majority of ATCOs (at least 75%) responses is that operating methods are clearly identified and consistent in all operating conditions.

In the post-exercise questionnaire, ATCOs agreed in 100% of the cases that they found operating methods for ASR to be clear, complete and exhaustive, under all operating conditions. During the debriefings, ATCOs mentioned they experienced no change in operating methods when using ASR.

Figure F-215~~Figure F-214~~ below shows the number of answers collected for each point of the 7-point Likert.

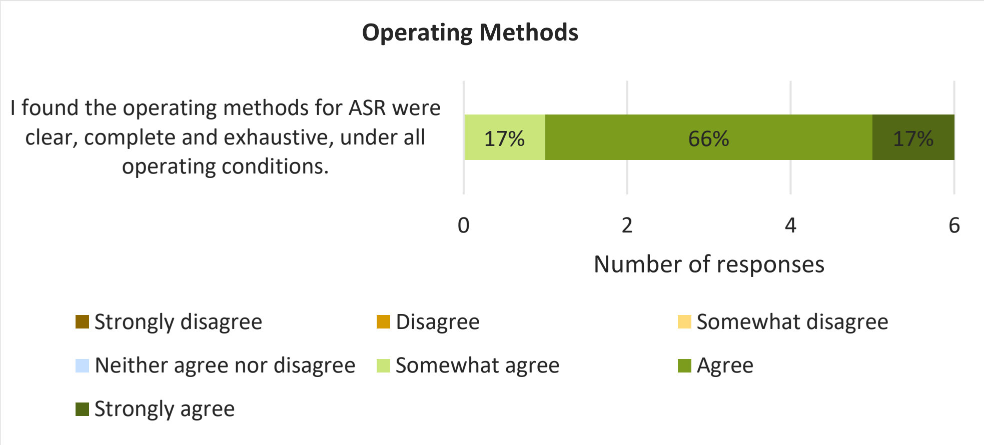


Figure F-215~~214~~: Post-exercise questionnaire: Operating Methods with ASR

F.7.2.5 OBJ-05.971-TRL4-TVALP-H103.1040

To assess job acceptance and satisfaction with the introduction of ASR.

CRT-05.972-TLR4-TVALP-H106-2031

The majority of ATCOs (at least 75%) provide positive feedback on job satisfaction and acceptance.

In the post-exercise questionnaire, ATCOs agreed in 100% of instances on a job satisfaction increase when using ASR. Figure F-216~~Figure F-215~~ below illustrates the number of answers collected for each point of the 7-point Likert scale.

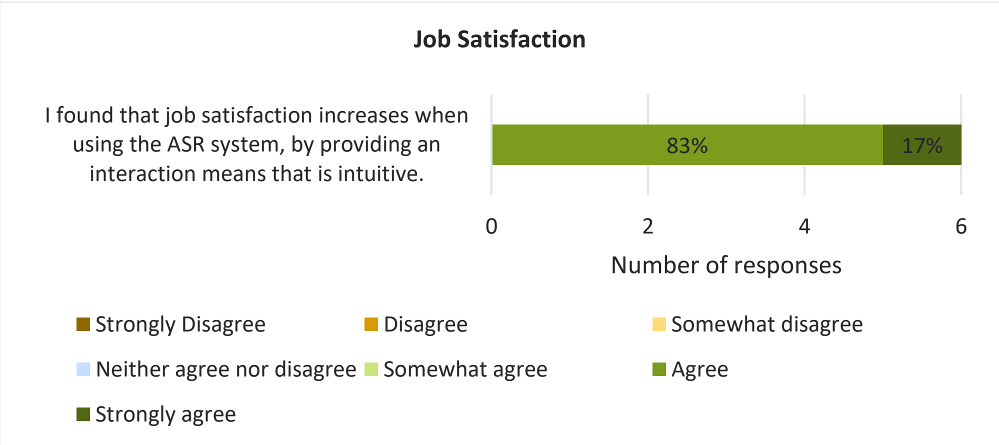


Figure F-216~~215~~: Post-exercise questionnaire: Job Satisfaction with ASR

F.7.2.6 OBJ-05.972-TLR4-TVALP-SAFE.2010

To assess the impact of Automatic Speech Recognition on safety.

CRT-05.972-TLR4-TVALP-SAFE-2011

The changes related to the implementation of Automatic Speech Recognition do not increase potential for human error and therefore are not reducing safety levels.

As can be seen in Figure F-196~~Figure F-195~~ in section F.7.2.3, 70% of ATCOs (4) responded that the ASR does not increase the potential for human error compared to current operations. During the debriefing, no concerns were expressed in terms of a potential increase in human error when using ASR.

CRT-05.972-TLR4-TVALP-SAFE-2012

ATCO's workload with the implementation of Automatic Speech Recognition is kept at acceptable levels and therefore does not reduce safety levels.

In the post-run questionnaire, controllers reported an average Workload of 3.5 on a 10-point scale for both reference and solution scenarios. This means that workload was satisfactory without reduction. (see Figure F-185~~Figure F-185~~). The average workload level did not change significantly according to ATCO role; however, in the ground position, the average workload reported during solution scenarios (3.75) is slightly higher than in the reference scenario (2.33). Such an indication might be due to some essential instructions missing from the ASR phraseology. For the tower position, the average workload level during the reference scenario is higher (4.66) as compared to the solution (3.33) which might be explainable with support coming from the 'Hook' function.

In the post-exercise questionnaire, most controllers (4) reported a light level of workload during the solution scenarios while the other two rated the workload level as 'acceptable', as shown in Figure F-187~~Figure F-187~~. Moreover, all controllers agreed that the level of workload improved during the solution scenarios thanks to the ASR command filling function as per Figure F-188~~Figure F-188~~.

In the post-run questionnaire, controllers were asked whether the level of safety was acceptable, and no degradation or safety concerns were raised during ASR-assisted scenarios. Most responses show a positive attitude regarding the level of safety. On 41% of occasions ATCOs 'strongly agreed', on 47% percent of occasions they 'agreed', while the rest of the responses (12%) were neutral (neither agree nor disagree). One controller even mentioned that they felt very safe in using this system while another one commented that ASR has a great potential in operational environment.

There was no general agreement 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. Although the majority of responses (approx. 65%) were positive, 29% of answers were neutral and on 6% of the occasions, ATCOs 'strongly disagreed'. Such a difference in ATCOs opinions could be explained by some of their comments, stating that in their experience, ASR does not recognise most of standard phraseology, but only a few keywords and that the system misunderstood communications several times. In conclusion, the level of safety was generally acceptable during the ASR runs.

Figure F-217~~Figure F-216~~ below details the number of answers collected for each of the 7-point Likert scale, for the two questions mentioned above.

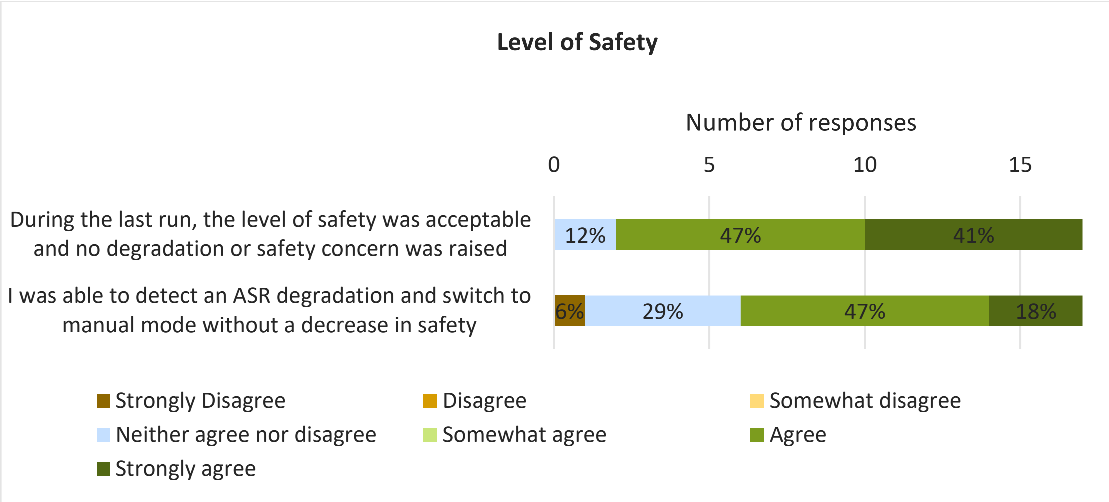


Figure F-217~~216~~: Post-run questionnaire: Level of Safety during Solution Scenarios

In the Post-exercise questionnaire, ATCOs were asked if the overall level of safety was at least equal to today's operations during the ASR exercise execution. Five out of six ATCOs agreed, while only one somewhat disagreed, as Figure F-218~~Figure F-217~~ below shows. This indicates the overall level of safety was considered as satisfactory for most of the controllers.

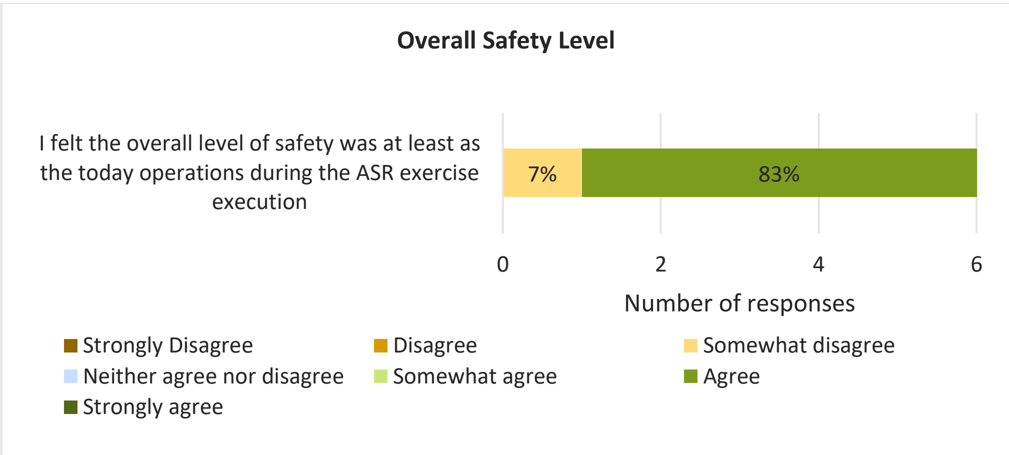


Figure F-218~~217~~: Post-exercise questionnaire: Overall Safety level with ASR

During debriefings, no safety concerns were expressed in reference to the ASR tool. However, as it can be seen in the WANT/HAVE matrix, ATCOs would like to keep the ASR feature requiring a final ATCO approval of the command by clicking ENTER on the HMI or pressing a key on the keyboard, as they consider it to be a useful safety barrier.

CRT-05.972-TLR4-TVALP-SAFE-2013

ATCO's situational awareness with the implementation of Automatic Speech Recognition is maintained at acceptable level and therefore not reducing safety levels.

In the SASHA post-run questionnaire, ATCO reported that they managed to be ahead of traffic during the ASR runs, with no significant difference from the reference runs. With ASR, ATCO were able to better plan their work and they spent less time looking for information as compared to current operations. ASR did not cause any unexpected events or did not particularly drive the ATCOs to focus on a single problem as compared to the reference scenario. However, some ATCOs thought that when using ASR they might be rarely exposed to the risk of forgetting something important.

The post-exercise questionnaire showed that in brief, situational awareness was satisfactory for all ATCOs when using ASR, as 83% of ATCOs rated it as 'High' and 17% as 'Perfect', as Figure F-195~~Figure F-194~~ displays. Also during the debriefings ATCOs provided positive feedback with respect to situational awareness, particularly with regard to the 'hook' function. Besides, in order to obtain an increase in situational awareness, a more evident highlight of the a/c track on the HMI was suggested by ATCOs.

In general, from the post-run and post-exercise questionnaires, it is reasonable to conclude that ATCOs were generally satisfied with the level of Situational awareness when using ASR.

CRT-05.97B-TLR4-TVALP-SAFE-2014
Safety assessment activities and their results are documented and integrated in the overall solution validation results.

The safety aspect related to ASR was investigated across the whole validation exercise. No specific safety related issues were identified.

F.7.2.7 OBJ-05.972-TLR4-TVALP-PERF.2010

To assess the performance benefits of Automatic Speech Recognition supported by AI/ML.

CRT-05.972-TLR4-TVALP-PERF-2011
Laboratory tests show that the SESAR technological solution improves Cost Efficiency performance by reducing cost per flight (PER.TRL4.3) (through e.g. reduction of workload, reduction of delay times), as also shown in Chapter 4.

F.7.3 Unexpected behaviours/results

None were found, both in the validation platform and in the operation of ASR4ATC per se.

F.7.4 Confidence in results of EXE-007

F.7.4.1 Level of significance/limitations of Technological Validation Exercise Results

Simulation EXE-007 has involved a range of Test subjects (6 ATCOs) with different backgrounds and expertise levels in a simulation environment representing Sofia Airport 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. Such a lack of training effect is anyway limited considering how intuitive tools employed in the simulations were.

F.7.4.2 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 in section F.7: both accuracy and confidence in the collected results as well as measured indicators are judged to be of satisfactory quality to support the maturity assessment in a TRL4 phase.

F.7.4.3 Significance of Technological Validation Exercises Results

The simulation exercise has been conducted on an experimental platform representing Sofia Airport environment with a high degree of fidelity providing an operational significance adequate to support the TRL4 maturity assessment, of course with limitations already mentioned in Sections F.7.1 and F.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 TRL4 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.

F.8 Conclusions

F.8.1 Conclusions on technological feasibility

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 inevitable limitations, results of the Validation Exercise indicate good performance and positive results of the assessment of the ASR tool made by ATCOs. 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.

F.8.2 Conclusions on performance assessments

F.8.2.1 Cost Efficiency Performance

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

F.8.2.2 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 A-SMGCS 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 ASR4ATC has no negative impact in terms of workload and situation awareness. Beneficial effects arising from the support offered by the ‘Hook’ function on situational awareness resulted from simulations. However, in order to further enhance efficiency of ASR support, some improvements were identified: a need to enrich the phraseology recognised by ASR4ATC, as well as tighter integration with WP and some specific A-SMGCS functions. 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.

- Positive feedback from ATCOs regarding acceptance and trust in the system indicates that the level of ASR technical performance was acceptable and consistent with human capabilities.

- Favourable reactions from ATCOs in terms of usability suggest a high quality of user experience when interacting with ASR4ATC and its related functions. Nevertheless, 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.

F.8.2.3 Safety

- ATCO workload and situational awareness remain at acceptable levels and therefore do not appear to reduce safety levels integrating ASR4ATC into the LIS suite

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

F.8.3 Recommendations

F.8.3.1 Technological feasibility

The ASR technology has shown to be feasible in an ATC tower 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
- 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
- Incorporate the entire SESAR Ontology, in order to enlarge recognition capabilities

in the view of a wider choice of airport in which for ASR4ATC to operate. Tighter WP integration would also be beneficial in order to cover a broader range of operational scenarios.

F.8.3.2 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.

F.8.3.3 Human Performance

Human performance could be improved from the following perspectives:

Workload:

- ATCOs recommend enriching the type and number of ATC commands to be integrated in ASR4ATC, by automatically recognising and executing commands such as assuming traffic or activation/de-activation of stop bars. Voice operated commands were considered as

beneficial, in the view of a broader ASR adoption and usage, expected to provide further support to workload and situational awareness.
- Finding an alternative means for the use of the two keys, one for the activation of the ASR and the other for communicating on the radio frequency would improve workload, as an initial step. Always-on ASR voice operation is a long term goal, overlapping r/t communication.

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
- Allowing controllers to search a/c also by their type would significantly improve the 'Hook' function effectiveness.

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 pop-up window' which would display logs and transcripts, always in the same place, similar to a chat window, in order for ATCOs to inspect ASR operation when and if necessary.

Acceptance & Job Satisfaction

- Further and tighter integration with A-SMGCS functionalities would be recommended. For example, it would be useful if ASR could recognise and display the taxi route assigned to an a/c by a controller or if it could display a runway as 'occupied' when recognising that a vehicle using that runway is in contact with the tower. Also, ASR could highlight a closed taxiway on the WP HMI.
- 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 in the tower environment by foreseeing interaction between ASR and the eFlight Strips, which would allow ATCOs to activate, via ASR, safety barriers.
- Another recommendation received from ATCOs is to improve ASR so as to activate safety barriers such as stop bars.

F.8.3.4 Safety

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

Appendix G SESAR Technological Solution 97.1 Maturity Assessment

The following table was answered according to "SESAR 2020 - Execution Framework – Project Handbook – SESAR Maturity Criteria.xls" (V2-TRL4)

SESAR Maturity Criteria_1 (1_5)_Air (

SESAR Maturity Criteria_1 (1_5)_Atte

SESAR Maturity Criteria_1 (1_5)_VR_

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Appendix H SESAR Technological Solution 97.2 Maturity Assessment

The following table was answered according to “SESAR 2020 - Execution Framework – Project Handbook – SESAR Maturity Criteria.xls” (V2-TRL4)

SESAR Maturity Criteria_1 (1_5)_ASR

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Logos of project partners in Sol 97.1 and Sol 97.2

Air Navigation Services of the Czech Republic	ALMA MATER STUDIORUM A.D. 1088
austro CONTROL	CRIDA I+D+i - ATM
CROATIA CONTROL	deepblue
DLR	ENAIRE
enav	HungaroControl
indra	integra
LEONARDO	nlr
ON ORO NAVIGACIJA	

EUROPEAN PARTNERSHIP

Co-funded by the European Union