Report on Stakeholder Workshop Results

Deliverable ID:	**D6.1**
Dissemination Level:	**PU**
Project Acronym:	**HAAWAII**
Grant:	**884287**
Call:	**H2020-SESAR-2020-2**
Topic:	**SESAR-ER4-18-2019**
Consortium Coordinator:	**DLR**
Edition date:	**11 November 2022**
Edition:	**01-00-00**
Template Edition:	**02.00.02**

SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Hartmut HELMKE (DLR)	Project Lead	10.11.2022
Johannes BLÜMEL (ACG)	D6.1 Leader	13.10.2022

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Hartmut HELMKE (DLR)	Project Lead	10.11.2022

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

Name/Beneficiary	Position/Title	Date
Hartmut HELMKE	Project Lead	10.11.2022
Pavel SMRZ (BUT)	WP4 Leader	By silent approval
Petr MOTLICEK (Idiap)	WP3 Leader	14.10.2022
Johannes BLÜMEL (ACG)	D6.1 Leader	13.10.2022
Teodor S. SIMIGANOSCHI (Isavia)	WP1 Leader	10.11.2022 via email to H. Helmke
Julia HARFMANN (NATS)	WP5 Leader	10.11.2022 via email to H. Helmke
Martina ERCEG (CCL)	Contact Point CCL	By silent approval

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date

Document History

Edition	Date	Status	Author	Justification
0.01	30.06.2021	Draft	H. Helmke	First Version with structure of document and description of WG “What is a readback error?”
0.02	13.09.2022	Draft	J. Blümel	Addition of 3.5 and 3.6; Deletion of 3.7.1, 3.7.2 and 3.7.3
0:03	13.10.2022	Draft	J. Blümel	Addition of 2nd Stakeholder Workshop
0.03	14.10.2022	Draft	P. Motlicek	Reading, minor edits done

0.03	18.10.2022	Draft	T. Simiganoschi	Reading and added the results of workshop “Application of Speech Recognition and Understanding the ATC environment”
0.03	31.10.2022	Draft	J. Harfmann	Addition of 4.5
	01.11.2022	Draft	H. Helmke	
	10.11.2022	Draft	H. Helmke	Adding feedback from S. Chen to workshop on adaptation to new airports
0.99	10.11.2022	Draft	J. Blümel	Finalisation of the document
	10.11.2022	Draft	H. Helmke	Set to final and minor formal changes
	11.11.2022	Final	H. Helmke	Approval of Isavia ANS added

Copyright Statement

HAAWAII

HIGHLY ADVANCED AIR TRAFFIC CONTROLLER WORKSTATIONS WITH ARTIFICIAL INTELLIGENCE INTEGRATION

This General document is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 884287 under European Union's Horizon 2020 research and innovation programme.

Abstract

Advanced automation support developed in Wave 1 of SESAR IR includes using of automatic speech recognition (ASR) to reduce the amount of manual data inputs by air-traffic controllers. Evaluation of controllers' feedback has been subdued due to the limited recognition performance of the commercial of the shell ASR engines that were used, even in laboratory conditions. Past exploratory research funded project MALORCA, however, has shown (on restricted use-cases) that satisfactory performance can be reached with novel data-driven machine learning approaches. The project builds on very large collection of data, organized with a minimum expert effort to develop a new set of models for complex environments of Icelandic en-route and London TMA. The deliverable is public.

The deliverable D6.1 summarizes the dissemination of the HAAWAII project by conducting Stakeholder workshops. This report summarizes the two Stakeholder workshops, first conducted end of June 2021 and the second one end of September 2022.

Table of Contents

List of Tables

1 Executive Summary

The deliverable D6.1 summarizes the dissemination of the HAAWAII project by conducting Stakeholder workshops. The current version of this report summarizes the first Stakeholder workshop conducted end of June 2021 and the second Stakeholder workshop in September 2022, which was also a demo day

Austro Control (ACG) together with the other HAAWAII project partners organized the first Stakeholder Workshop from 28th to 29th of June 2021. Unfortunately, due to the Covid-19 pandemic this workshop could not be held on ACG premises and had to be moved to a completely virtual meeting. To ensure a smooth workshop, NATS provided the proper support with their Microsoft Teams infrastructure.

The Stakeholder Workshop was divided into different parts. Each day started with presentations about the project and related content. Afterwards all participants were split up in different parallel working groups with 10 to 20 participants.

Summing both days up we had more than 60 participants each day, which over 120 in total.

First Day 28th of June

The Stakeholder Workshop was opened with some welcome notes from the ACG – ATM Department Director Christian Kern and the SESAR Joint Undertaking – Program Manager Cengiz Ari. Two presentations followed on the first day. Presentation slides of both days can be found on project website. https://www.haawaii.de

Six working groups were conducted in parallel after the presentation:

- Three parallel working groups on "Readback Error Detection Assistant: What is a readback error?"
- Two parallel working groups on "Standardizations related to Speech recognition"
- One working group on "Access to speech training data in Europe"

At the end of the day each working group presented their results to the audience and a short preview was given on the content of the next workshop day.

Second Day 29th of June

The day started again with round about one hour of presentations.

After the presentations four parallel working groups were offered:

- Two parallel working groups on "Readback Error Detection Assistant: HMI aspect of readback error detection"
- One working group on "Human Performance"
- One working group on "Workshop on legal aspects in ATM"

The Workshop ended with the presentation of the working group groups results and some closing words from the HAAWAII project lead Prof. Hartmut Helmke.

After the Stakeholder Workshop all partners started to evaluate the output of the different working groups. We got for sure a lot of good ideas, inputs and suggestions for the upcoming part of the project and much more we all learned a lot for the future of Automatic Speech Recognition and its challenges in ATM.

Summary of the working groups:

- Readback Error Detection Assistant: What is a readback error?
- Standardizations related to Speech recognition
- Access to speech training data in Europe
- Readback Error Detection Assistant: HMI aspect of readback error detection
- Human Performance
- Legal aspects in ATM

The second Stakeholder Workshop took place at NATS facilities from 27th to 28th of September 2022 and was integrated into a demo day. All project members from HAAWAII, as well as partners from industry and research attended at the workshop and demo day. The workshops were hold on the second day of the event.

Second Day 28th of September

Three working groups were conducted in parallel.

Workshops:

- Tuning of ASR Models – Training to new Environment and Data Privacy Aspects
- Application of Speech Recognition and Understanding in ATC Context (including CPDLC)
- Using Speech Recognition for Workload Prediction – How do we develop a useful Tool for Supervisors and Controllers?

2 Introduction

2.1 Purpose of the document

The HAAWAII project builds on a very large collection of data, organized with a minimum expert effort to develop a new set of models for complex environments of Icelandic en-route and London TMA.

This deliverable summarizes the dissemination of the first Stakeholder Workshop of the HAAWAII project. The main objective is to share the results with the public.

2.2 Scope

The HAAWAII project aims to research and develop a reliable, error resilient and adaptable solution to automatically transcribe voice commands issued by both air-traffic controllers and pilots.

One main objective of this research project is the dissemination of the results. For this purpose, different Stakeholder Workshops and Result demonstrations are planned in 2021 and in 2022.

2.3 Intended readership

This document is mainly intended for:

- **Participants** of the First Stakeholder Workshop end of June 2021,
- **Participants** of the Second Stakeholder Workshop and demo day in September 2022,
- **HAAWAII** consortium members in order to integrate the valuable feedback of workshop participants into the remaining months of project life time,
- **SESAR JOINT UNDERTAKING (SJU)** as Horizon 2020 Programme coordinator.

2.4 Background

During the AcListant® and AcListant®-Strips project funded by Helmholtz Validation Fund and DLR Technology Marketing two Stakeholder Workshops were conducted, one in 2013 and one in 2014. The SESAR2020 Exploratory Research funded project MALORCA (Machine Learning of Speech Recognition Models for Controller Assistance) conducted two workshops, one in April 2017 and Prague with 58 participants and one in February 2018 in Vienna with 40 participants.

The workshop conducted in June 2021 was the first fully virtual workshop, which was a consequence of the COVID-19 situation in Europe.

2.5 Structure of the document

The structure of this document is based on the Horizon 2020 template for project deliverables. It is organized as follows:

- **Chapter 1: Executive Summary**. Provides a summary of the key information and elements contained in the Technical Validation Report document.
- **Chapter 2: Introduction** (this chapter). Introduces the document.
- **Chapter 3:** Provides a detailed overview about the first Stakeholder Workshop.
- **Chapter 4:** Provides a detailed overview about the Second Stakeholder Workshop.
- **Chapter 5: References.** Provides a summary of the references.

2.6 Glossary of terms

The HAAWAII project has more than 20 different deliverables. Therefore, the HAAWAII project decides to have one separate document containing the glossary of terms, so that maintenance of the terms is eased and errors or misunderstandings only need to be changed in one place.

For simplifying the task of the readers, the contents of the master document are shown in the following table.

Term	Definition	Source of the definition
AcListant®	Venture Capital funded project Active Listening Assistant being conducted by DLR and Saarland University from 2013 to 2015.	PJ.16-04
Annotation	This task extracts the semantic concepts from the Transcription (i.e. text-to-concepts transformation), e.g., “DLH2BA DESCEND 80 FL, DLH2BA REDUCE 220 kt” and “AFR273 CORRECTION, AFR273 CONTACT VIENNA_RADAR, AFR273 CONTACT_FREQUENCY 129.500”.	D3.1
Assistant Based Speech Recognition (ABSR)	Special Instance of Automatic Speech Recognition which needs an assistant system to provide context in order to improve recognition rate and/or reduce error rate	See definition in [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. 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	PJ.16-04

Term	Definition	Source of the definition
	used for further applications such as visualization on an HMI.	
Callsign (Recognition) Error Rate	The number of callsign, which are wrongly recognized by ABSR and which are not rejected divided by the number of total given callsigns; in other words: the percentage of given callsigns wrongly shown on the controllers' HMI. "oscar kilo one" must be mapped to "OACK1" if this is the only "O..K1" in the air. Otherwise it is counted as an error.	in D1.2
Callsign Recognition Rate	The number of callsigns, which are correctly recognized by ABSR and are not rejected before divided by the number of total given callsigns; in other words: the percentage of given callsigns correctly shown on the controllers' HMI. "oscar kilo one" must be mapped to "OACK1" if this is the only "O..K1" in the air.	in D1.2
Callsign Rejection Rate	The number of callsigns, which are said by the ATCo, but mapped to NO_CALLSIGN divided by the number of total given callsigns; in other words: the percentage of given callsigns not shown at all on the controllers' HMI.	in D1.2
Chunk		D3.1
Clearance transmission identifier	The Clearance transmission identifier is part of the readback information and represents the Transmission unique identifier from the Transmission information. This will be used to trace and check a specific transmission from the multiple transmissions. See example in Table 1 Example of transmission information and identifiers	in D1.2
CoCoLoToCoCo	Controller Command Logging Tool for Context Comparison that provides a user-friendly interface to carry out transcriptions and various annotations for air traffic control voice commands.	D3.1
Command Prediction Error Rate	The number of controller commands which are given but not predicted (by the Command Hypotheses Predictor) divided by number of total given commands; in other words: the percentage of errors of the Command Hypotheses Predictor.	See definition in [1]
Command Recognition Rate	The number of controller commands which are correctly recognized by ASR and are not rejected before divided by number of total given commands; in other words: the percentage of given commands correctly shown on the controllers' HMI.	See definition in [1]

Term	Definition	Source of the definition
Command (Recognition) Error Rate	The number of controller commands which are wrongly recognized by ASR and which are not rejected divided by number of total given commands; in other words: the percentage of given commands wrongly shown on the controllers' HMI.	See definition in [1]
Communication group	Communication group is part of transmission information and it is a generated value or index that is used to identify and group multiple ATCo/Pilot transmissions that represent a single communication/dialogue. The single communication/dialogue is for example when pilot asks for higher flight level and the ATCo provides clearance for that flight level. See example of multiple transmissions grouped into communication groups in Table 1 Example of transmission information and identifiers.	in D1.2
Concept of Operations [ConOps]:	Concept of Operations [ConOps]: The ConOps is jointly elaborated by all ATM stakeholders, from the civil and military airspace users and service providers, to airports and the manufacturing industry to gain common understanding of the ATM system. It describes the operational targets, to move ATM towards trajectory-based operations whereby aircraft can fly their preferred trajectories, considering the matching between constraints and optimization. The ConOps allows all ATM stakeholders, from the civil and military airspace users and service providers, to airports and the manufacturing industry to gain common understanding of the ATM system. In this context, the ConOps is the operational answer to reach the ATM Performance improvements targeted by the ATM MP. Furthermore, the ConOps is an important reference for global interoperability and harmonization, as it has been adapted for Europe from the ICAO Global Air Traffic Management Operational Concept.	See definition in [2]
Controlling Working Position Identifier	The controlling working position identifier is part of the Transmission information and represents a name or index to identify the position that generated that specific transmission. See example in Table 1 Example of transmission information and identifiers.	in D1.2
Exploratory Research	The exploratory research investigates relevant scientific subjects (during the ATM Excellent Science & Outreach phase) and conducts feasibility studies looking for potential	See definition in [2]

Term	Definition	Source of the definition
	application areas in ATM (during the ATM application-oriented research phase).	
Horizon 2020	The EU Framework Programme for Research and Innovation.	SESAR 1, WP14, SESAR 2020
MALORCA	Machine Learning of Speech Recognition Models for Controller Assistance, Horizon 2020 funded project from 2016 to 2018	
PMP deliverable	Output produced by the projects that is submitted to the SJU via the SESAR 2020 collaborative platform and that is subject to quality assessment by the SJU. However, these deliverables do not appear in the grant agreement as contractual deliverables. The production of PMP deliverables is done in support of subsequent contractual deliverables and is described in the PMP.	See definition in [2]
Project Management Plan	Formal, approved document, provided by each SESAR 2020 Solution Project, used to manage its execution. It defines how the project is executed, monitored, controlled, and closed.	See definition in [2]
Read-back error detection rate	The number of correctly detected read-back errors (with or without correction) divided by the total number of read-back errors (with or without correction).	
Read-back error false alarm rate	The number of detected read-back errors, which are not a read-back error, divided by the total number of read-back errors (with or without correction).	
SESAR 2020	The SESAR 2020 (Single European Sky ATM Research) Research and Innovation (R&I) Programme will demonstrate the viability of the technological and operational solutions already developed within the SESAR R&I Programme (2008-2016) in larger and more operationally-integrated environments. At the same time, SESAR 2020 will prioritise research and innovation in a number of areas, namely integrated aircraft operations, high capacity airport operations, advanced airspace management and services, optimised network service performance and a shared ATM infrastructure of operations systems and services. SESAR 2020 will retain its founding members, the European Union and Eurocontrol.	SESAR 1, WP14, SESAR 2020, PJ.17-03

Term	Definition	Source of the definition
Transcription	This task involves the speech-to-text transformation, writing down word-by-word, what the ATCo has said. Examples are: “lufthansa two bravo alfa descend flight level eight zero and reduce speed two two zero knots” and “bonjour air_france two seven three [unk] confirm vien* correction contact vienna radar on one two nine decimal five”.	D3.1
Transmission Direction	This is either “ATCo” when the ATCo (ground) speaks to the pilot or “Pilot”, if the pilot (air) speaks to the ATCo.	D1.2
Transmission unique identifier	Transmission unique identifier is part of transmission information and represents a generated unique value or index that is used to distinguish one single transmission from either ATCo or Pilot.	D1.2
TRL 2 (V1)	Technology concept and/or application formulated: Applied research. Theory and scientific principles are focused on very specific application area(s) to perform the analysis to define the concept. Characteristics of the application are described. Analytical tools are developed for simulation or analysis of the application.	See definition in [2]
TRL 3	**Analytical and experimental critical function and/or characteristic proof-of concept:** Proof of concept validation. Active Research and Development (R&D) is initiated with analytical and laboratory studies including verification of technical feasibility using early prototype implementations that are exercised with representative data.	See definition in [2]
TRL 4 (V2)	Component/subsystem validation in laboratory environment: Standalone prototyping implementation and test with integration of technology elements and conducting experiments with full-scale problems or data sets.	See definition in [2]
True Positives (tp)	The total number of correctly predicted commands, i.e., the number of commands which were predicted which were actually given.	
False Positives (fp)	The total number of falsely predicted commands, i.e., the number of commands which were predicted but actually NOT given.	

Term	Definition	Source of the definition
False Negatives (fn)	The total number of commands which were falsely not predicted, i.e., the number of commands which were NOT predicted but were actually given.	
True Negatives (fn)	The total number of commands which were correctly not predicted, i.e., the number of commands which were NOT predicted and actually NOT given.	
Recall	Recall represents the percentage of actually given commands which were predicted. tp / (tp + fn)	
Precision	Precision represents the percentage of true predictions out of all the commands which were predicted. tp / (tp + fp)	
Accuracy	Accuracy represents the prediction rate. It also takes into account the number of commands which were correctly NOT predicted. (tp + tn) / (tp + fp + fn +tn)	
Segment	A part of the audio recording without any specific property	D3.1
Utterance	Segment of an audio file, which consists of a complete message by only one speaker to the other dialogue participants . In case of ATC it contains complete message of ATCo to one pilot or complete answer of pilot to ATCo. Utterance can contain one or more sentences e.g. "Good morning. Speed bird one three seven descend flight level eighty". Utterance segments can be automatically or manually created.	D3.1
SpokenData	A generic web based tool which allows to transcribe the speech recordings, while transcribers are supported by several functions to minimise their effort.	D3.1

Reference used in Glossary of terms

[1] H. Helmke, J. Rataj, T. Mühlhausen, O. Ohneiser, H. Ehr, M. Kleinert, Y. Oualil, and M. Schulder, "Assistant-Based Speech Recognition for ATM Applications," in 11th USA/ Europe Air Traffic Management Research and Development Seminar (ATM2015), Lisbon, Portugal, 2015.

[2] SESAR 2020 Execution guidance of ER4 projects : https://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/jtis/h2020-guide-project-handbook-er4-sesar-ju_en.pdf

Transmission unique identifier	ATCO/Pilot Transmission	Clearance transmission Identifier	Controlling Working Position Identifier	Communication Group
1	ATCO: XYZ descend flight level three one zero	LLAP	LLAP	1
2	Pilot: XYZ descending level three one zero	LLAP	LLAP	1
3	ATCO: hello london ABC two five nine altitude flight level two zero zero until xyz	TMASOUTH	TMASOUTH	2
4	Pilot:two zero zero xyz ABC two five nine	TMASOUTH	TMASOUTH	2

Table 1 Example of transmission information and identifiers

2.7 Acronyms and terminology

The HAAWAII project has more than 20 different deliverables. Therefore, the HAAWAII project decides to have one separate document containing these acronyms, so that maintenance of the acronyms is eased and errors or misunderstandings only need to be changed in one place.
For simplifying the task of the readers, the contents of the master document are shown in the following table.

Term	Definition
ABSR	Assistant Based Speech Recognition
ACC	Area Control Centre
ACG	Austro Control Österreichische Gesellschaft für Zivilluftfahrt (Austrian ANSP)
ADS-B	Automatic dependent surveillance–broadcast
AEC	Approach executive controller
AFIS	Aerodrome Flight Information Service
AG	Attention Guidance
AI	Artificial Intelligence
ANRIC	Aeronautical Radio Incorporated
ANSP	Air Navigation Service Provider
ANS-CR	Air Navigation Services of the Czech Republic
APC	Approach planning controller
APP	Approach
ARR	Arrival
ARTAS	ATM suRveillance Tracker And Server
ASR	Automatic Speech Recognition
ASTERIX	All Purpose Structured Eurocontrol Surveillance Information Exchange
ASW	Air situation window
ATC	Air Traffic Control
ATCo	Air Traffic Controller; also ATCO used, but ATCo preferred in HAAWAII project
ATM	Air Traffic Management
Avg	Average
BUT	Brno University of Technology
CBA	Cost Benefit Analysis
CER	Command or Context (Prediction) Error Rate, also used as CtxER

Term	Definition
Cmd	Command (files containing annotations)
CmDER	Command Error Rate
CmDRR	Command Recognition Rate
CoCoLoToCoCo	Controller Command Logging Tool for Context Comparison
Cor	Correct (files containing transcriptions)
COTS	Commercial of the shell
CPP	Context Portion Predicted
CONOPS	Concept of operations
CPDLC	Controller Pilot Data Link Communications
CTA	Control area
CTR	Controlled traffic region
CtxER	See CER
CV	Clearance verification
CWP	Controller Working Position
DASC	Digital Avionics Systems Conference
DEC	Departure executive controller
DEP	Departure
DFS	Deutsche Flugsicherung GmbH (German ANSP)
DLR	German Aerospace Center, Deutsches Zentrum für Luft- und Raumfahrt e.V.
DNN	Deep neural network
DPO	Data Protection Officer
DVI	Direct Voice Input
DVO	Direct Voice Output
EATMA	European Air Traffic Management Architecture, An architectural Model of European ATM for each SESAR Concept Story board step containing information relating to Operational activities.
EDR	Event Detection Rate
EML	European Media Laboratory
ENAIRE	Spanish ANSP
ER	En-Route
Err	Error (files containing errors)
EU	European Union

Term	Definition
EXE	Exercise
FAA	Federal Aviation Administration
FANS	Future Air Navigation System
FDPS	Flight Data Processing System
FL	Flight level
FIR	Flight Information Region
ft	Feet
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HF	Human factors
HMI	Human Machine Interface
HUP	Human Performance
IB	Information Bottleneck
ICAO	International Civil Aviation Organization
ICE	Intelligent Communications Environment
ID	Identifier
Idiap	Idiap Research Institute
IEC	Information executive controller
ILS	Instrument landing system
IFR	Instrument Flight Rules
ISA	Instantaneous self assessment
JSON	JavaScript Object Notation
khz	Kilo hertz
KPA	Key Performance Area
kt	Knots
KWA	Keyword Spotting Algorithm, special implementation of callsign recognition
LAC	London Area Control
LTCC	London Terminal Control Centre
LTMA	London Terminal Manouvering Area
MALORCA	Horizon 2020 funded project MACHINE LEARNING OF SPEECH RECOGNITION MODELS FOR CONTROLLER ASSISTANCE

Term	Definition
MWM	Mental Workload Model
N/A	Not applicable
NASA TLX	NASA Task load index
NATS	United Kingdom ANSP
NAT OTS	NORTH ATLANTIC ORGANIZED TRACK SYSTEM
Nm	Nautical miles
No.	Number
NOK	Not Ok
NPR	Noise Preferential Route
OA	Open Access
Obj	Objective
OSED	Operational services and environment description
OTS	ORGANIZED TRACK SYSTEM
PC	Prestwick Centre
PEC	Director executive controller
PERF	Performance
PJ	Project
POK	Partly Ok
PST	Performance Stability
PSS	Paperless Strip System
PTT	Push to talk
R/T	Radio Telephony
RabbitMQ	is an open-source message-broker software (sometimes called message-oriented middleware)
REF	Reference
REQ	Requirement
ReTi	Reaction Time
RMA	Radar Manoeuvring Areas
RNAV	Area navigation
RTP	Real Time Protocol
RWY	Runway

Term	Definition
(S)VFR	(Special) Visual Flight Rules
S2T	Speech-To-Text
SA	Situation Awareness
SAD	Speech Activity Detection
SAF / SAFE	Safety
SAR	Safety assessment report
SASHA	Situation Awareness for SHAPE (Solutions for Human Automation Partnerships in European ATM)
SC APP	Approach Senior Controller
Scn	Scenario
SDK	Software Development Kit
SDDS	Surveillance Data Distribution
SESAR	Single European Sky ATM Research
SID	Standard instrument departure
SJU	SESAR Joint Undertaking
SME	Subject Matter Experts
SOL	Solution
STAR	Standard terminal arrival route
STCA	Short Term Conflict Alerting
T2C	Text-to-Concept
T2S	Text-to-Speech
TC	Terminal Control
TMA	Terminal Manoeuvring Area
TRL	Technology Readiness Level
TS	Technical Specification
TSWR	Tower
TTC	Text-to-Concept
TTS	Text-to-Speech
TVALP	Technical Validation Plan
TVALR	Technical Validation Report
V2T	Voice to Text

Term	Definition
V&V	Validation & Verification
VAD	Voice activity detection
VCS	Voice communication system
VFR	Visual flight rules
VieAPP	Vienna Approach
VRR	Voice Recognition and Response
VTT	Voice to Text
WDR	Word Detection Rate, approx.. 100% - WER
WER	Word Error Rate
WL	Workload
w.r.t.	with respect to
XML	eXtenable Markup Language

3 First Stakeholder, June 2021

3.1 Presentations Day 1

Agenda Day 1: Monday 2021-06-28

Start		End	Duration	TOP	Who
13:00	-	13:05	5	Welcome by Christian Kern/Austro Control	ACG
13:05	-	13:10	5	Welcome by SJU	SJU
13:10	-	13:35	25	Overview of the HAAWAII-Project	DLR
13:35	-	14:00	25	Speech Recognition in the ATC environment for Human Performance Evaluation, Radar Label Maintenance, Readback Error Detection	NATS / ISAVIA
14:00	**-**	**15:00**	**WORKSHOPS IN PARALLEL**		
- Readback Error Detection Assistant: What is a readback error? - Standardizations related to Speech recognition - Access to speech training data in Europe					
15:00	-	15:15	15	Coffee Break	
15:15	-	15:45	30	Presentation of Workshop Results (5-10 Minutes per Workshop)	All
15:45	-	16:00	15	Closing Words Day 1	DLR

3.2 Presentations Day 2

Agenda Day 2: Tuesday 2021-06-29

Start		End	Duration	TOP	Who
13:00	-	13:25	25	Looking behind the scenes of machine learning and AI - technical details of automatic speech recognition applied in ATC and first results	IDIAP
13:25	-	13:40	15	SJU expectations on ASR research	SJU
13:40	-	14:00	20	Legal challenges of Using Speech Data	University Bologna
14:00	-	15:00	WORKSHOPS IN PARALLEL		

3.3 Working Groups Related to Readback Error Detection

Two working groups with respect Readback Error Detection Assistant were conducted during the workshop. The first working group concentrate on finding out what is a readback error and what not. The results are presented in subsection 3.3.1 . The second working group tries to answer the question how and when to present a readback error candidate to the ATCo.

3.3.1 WG "What is a Readback Error?"

If the ATCo cleared a descent to flight level 100 and the pilot's read back is to flight level 90, the situation is quite clear. It is a readback error, which must be corrected very, very soon. If the ATCo wants "contact frequency one one nine decimal eight zero zero" and the pilot confirm with "nineteen eight bye" the situation with respect to ICAO phraseology is quite clear, but this readback is common practice. How to react? And what about "climb flight level one eight zero expedite passing flight level nine zero" and the readback "expedite nine to one eighty"? The workshop wants to brought together safety managers, air traffic controllers and other ATM experts.

The working group was splitted into three sub working groups, each with approximately 10 participants.

First 16 different use cases were presented to the audience. Each participant answered the questions anonymously. The answers were collected. Its main intention was to initiate a discussion within the three different groups. The answers of each group to the questions are presented in the appendix.

Results of the three sub working groups:

If the ATCo decides that a readback contains a readback error or may contain a readback error, (s)he will repeat her/his command. So, if a readback error assistant reports a readback error, this means that the assistant thinks that the ATCo should repeat, i.e. correct, the command. Unconventional, i.e. lousy, usage of phraseology is something different. This can also be reported, even in the context of readback error detection, because the results with respect to safety can be the same: ATCo and pilot interpret the given instruction differently.

The key question is, if a pilot's readback is worthy of the ATCo's further attention.

The ATCo has always the flexibility in operation, i.e. s(he) can interpret the context of a situation to decide whether it is important to correct or clarify a readback

- E.g., "I knew the pilot *meant* the correct thing even though the pilot did not *say* the correct thing or does not say it according to the book"
- E.g., "I know it was the correct pilot even though callsign was not read back resp. the callsign was even wrong"

Readback errors should be put into severity categories, e.g. critical, not important etc. Critical are altitude and heading commands in the TMA. Less critical are frequency changes or speed commands. The grade of deviation between the ATCo command and the pilot's readback also has an influence on the severity. If the ATCo says "flight level eight zero or below" and the pilot's readback is "below flight level eight zero", this is not so critical than a different flight level value itself.

3.3.2 WG "How and when to present a readback error to the ATCo"

If the ATCo cleared a descent to flight level 90 and the pilot's read back is to flight level 80, this is a readback error. Should the ATCo get immediately a warning or should a readback error assistant first check the following utterance of the ATCo? Maybe the ATCo will correct the readback error without being bothered by the assistant. How to react if the pilot just says "going down to nine" or if the readback error assistant just understands "descending flight level nine [coughing]" or "descending flight level [noise] zero" or when a Japanese pilot after twelve hours of flight time tries to approach Charles de Gaulle. The workshop brought together human factors experts air traffic controllers other ATM experts, but also speech recognition and understanding experts.

The working group was splitted into two sub working groups, each with approximately 15 participants.

First eight different use cases were presented to the audience. Each participant answered the questions anonymously. The answers were collected. Its main intention was to initiate a discussion within the three different groups. The answers of each group to the questions are presented in the appendix.

3.4 Working Groups Related to Data Management

This discussion was a part of the discussion group related to "access to speech training data in Europe" (lead by Idiap).

Google uses more than two hundred thousand of training data for its speech recognizers. Submissions to speech recognition conference report how many bi-words are used for model training and validation. A bi-Words stands for "one billion words". MALORCA project just benefits from 4 hours of training data for Prague airspace and nine hours for Vienna airspace. Airbus has offered 40 hours of training for its Speech Recognition Challenge in 2018. MITRE has access to 150,000 hours of recorded voice utterances – every month.

How can we improve the situation in Europe?

What are the legal challenges? What are business challenges? Data is the oil of the 21st century, so why sharing data with a possible competitor? What are the technical challenges

The workshop will run in parallel sessions to discuss different challenges. It brings together legal experts for data privacy and data ownership issues, data mining and machine learning experts, but also users who are producing the data and whose personal rights are at least touched.

3.5 Working Groups Related to ATCo Workload Prediction

On the second day of the first HAAWAII stakeholder workshop (29th June 2021) the Human Performance (HP) working group took place. The session was facilitated jointly by HP specialists from both Austro Control and NATS. The aim of the working group was to share the insights gained so far in speech recognition as a valuable HP metric and more importantly to get the participants' feedback and ideas on the way forward.

The first half of the session was used to present participants with HP concepts as well as opportunities and challenges of using speech recognition output as a proxy for overall human performance. The big question that we put in front of participants was "How would you quantify workload based on speech recognition data?".

The second half of the session was used to discuss this question as well as participants' views on the presented material overall.

The main findings were:

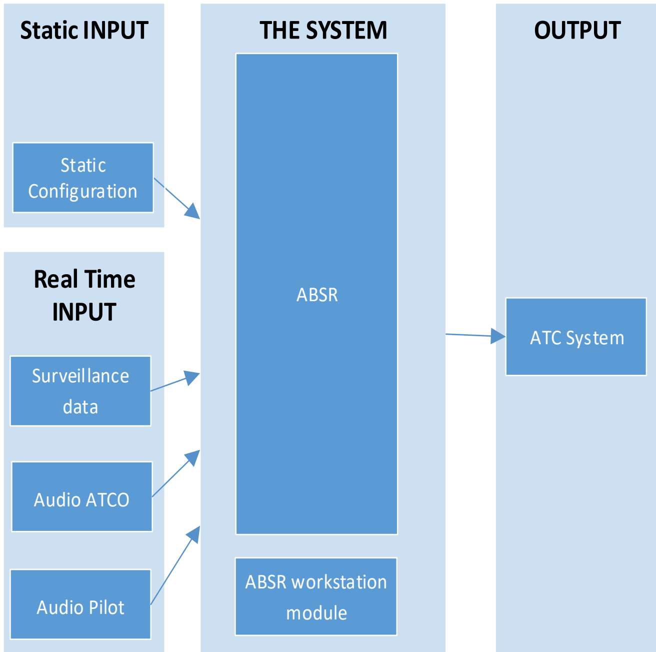
- Acknowledgement that a joint approach between quantitative and qualitative data gathering and analysis is needed to produce accurate insights into HP
- Recognition that extensive validation is needed with both traditional and new tools
- Assurance will be required before speech recognition data can be used for decision making by users
- Work on target audience is needed – data needs to be presented to supervisors and network planners and developed with a user centered design philosophy in mind
- Speech recognition data should be interwoven with other decision data such as weather or complexity
- Desire to use speech recognition data for off-line design like airspace or technology developments and change management
- Preference for simple visual displays of data rather than numbers or graphs
- Integration obstacles remain; especially convincing ANSPs or users to have unified data strategies
- Need to train supervisors and other users on ***how to trust the use*** of big data for decision making

In summary, the working group was well received and confirmed as well as broadened the HAAWAII team's understanding of how to progress the use of speech recognition data as another puzzle piece in measuring overall human performance.

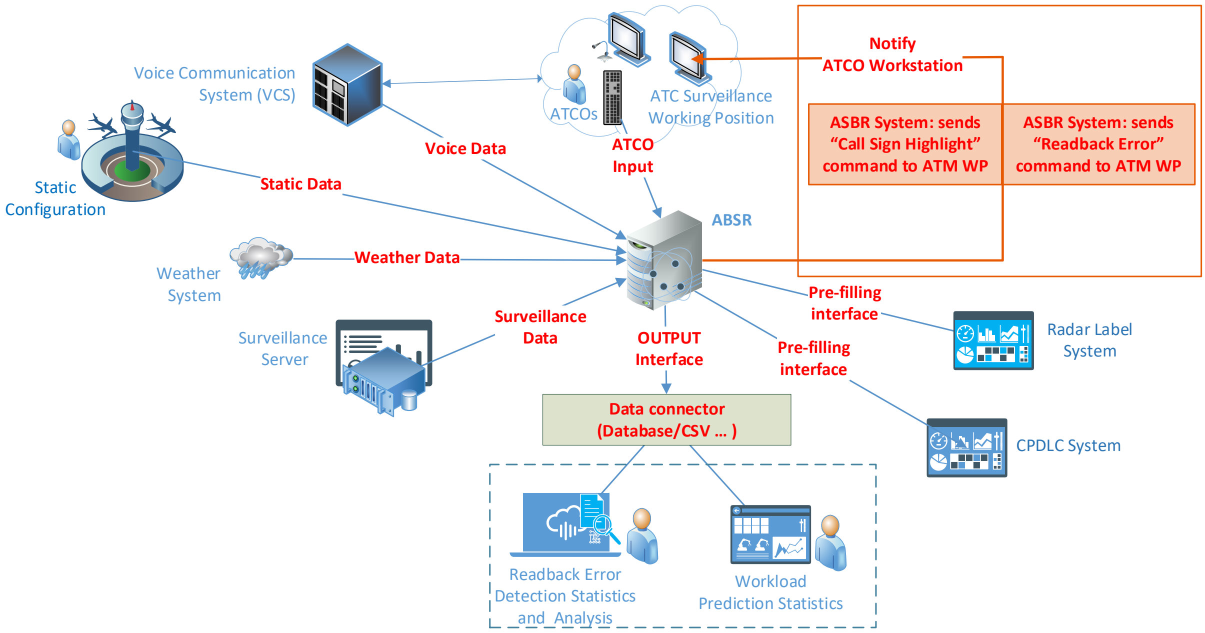
3.6 Working Groups Related to Standardization of Interfaces to and from Function Block "Automatic Speech Recognition"

The "Standardization related to speech recognition" workshops were split into 2 groups, each group with around eight participants. The coordinator of the group presented the workshop power point presentation and challenged the participants to answer different related questions related to integration of the ABSR applications into ATM environment.

The ABSR should receive data from the existing ATM systems (e.g., surveillance data, voice data, weather data etc.) and send the analyzed results or commands towards the controller working position or/and towards a technical workstation.



A more detailed diagram was presented in the workshop in order for participants to visually understand the challenges of integrating the ABSR into the operational environment:



The following questions were addressed in the workshop, the answers below are summarized from each workshop:

Question 1: What possible issues we might encounter, e.g., delays, defining 3 recording servers in the VCS, synchronization of Voice/Signalling, not everybody supports ED137 with PTT/SQU signalling indication?

Answer Q1:

- To minimize processing delays the ABSR needs to know who the pilot is and who the ATCO is so the PTT and SQU signaling must be send towards the ABSR (Voice Data needs to be splitted).
- Good idea to use ED137 (PTT and SQU are send with the protocol). We need to define some acceptable delays, but the recording is sent in terms of milliseconds. Not possible to define 3 recordings (ED137 defines only 2) but it can be added to the standard. Based on processing of VCS Jitter might appear so keep it in mind.

Question 2: What other options can we use to receive the Voice Data from VCS into the ABSR?

Answer Q2:

- Connect to the Headset of ATCo or RJ45 analogue recording interface with PTT and SQU signals directly from the Controller Working Position to the ABSR.
- One option would be direct integration into the VCS application and not as a recorder but as a specific protocol or fully integrated, also for other applications: Speech to text, Text to speech.

Question 3. What possible issues we might encounter, e.g., delay, defining more surveillance destinations, not everybody supports CAT062, not enough information's in CAT062?

Answer Q3:

- ASTERIX Category 062 is the right connection for surveillance data (ARTAS is the main tracker in Europe and outputs CAT062). The HAAWAII applications should clearly indicate what fields within CAT062 is supported, provide ICD.
- Does HAAWAII support Downlinked Aircraft Parameters (DAPS) or Mode-S ?
- If tracker at the ANSP does not support I062/390 field in the CAT062 message (is not connected to ATM Systems). The HAAWAII should also require mini flight plan, i.e. Mini flight plan ASTERIX Category 032.
- The ABSR application should also understand some static data like Environmental data, Sector definitions – Dynamic sectors.

Question 4: What protocol would be suitable to send commands towards the radar screen? ASTERIX message directly to the Controller working position with extra fields? ABSR stores the indications in a Database or CSV and the application will extract the necessary data from the Database or CSV

Answer Q4:

- To avoid the ABSR development and the ABSR to be just an ABSR, some options:
    - Add results in SQL or text file and the 3rd party to read the results
    - ABSR integration with the ANSP mater Control System (API or other way) that centralizes and manages the messages towards the Radar System, Flight strips, CPDLC or other systems. This Control System centralizes and manage all the messages in the ANSP environment.
    - Possible integrate only with electronic flight strips
- It is not sure whether the ABSR sends directly to the surveillance screen maybe to a server within the ATM system.
- For Readback Error indications the ABSR should send similar message as ASTERIX Category 004.
    - Maybe the objective is to add to CAT004 in long term
    - Discussion within the group whether that is correct for ABSR to develop its own standard
    - ASTERIX Category not suitable for free text for ontology?
- Pre-filling of radar labels

- No standard message input to ATM system.
- No open industry interface
- In the short term ABSR will have to use non-standardaized interfaces of the ATM companies

- Highlight the radar labels,
  - Idea, **investigate whether ABSR can use Squawk-ident feature**, i.e. Special Purpose Indicator (SPI) I048/020 or I062/080.
  - Configure ABSR as a surveillance source or tracker
  - ABSR would then require specific SIC/SAC.

Question 5: How can the ABSR get integrated with the existing flight strips systems, are there vendor specific protocols, can it be done?

Answer Q5:

Pre-Fill of the flight strips

- No standard message input to ATM system.
- No open industry interfaces
- In the short term ABSR will have to use non-standardized interfaces of the ATM companies

Question 6: How can the ABSR get integrated with the CPDLC systems, are there vendor specific protocols, can it be done?

Answer Q6:

- To avoid having the ATCo with 2 microphones (one for CPDLC and one for VCS) the ATCo can have the same Controller Working Position (Voice Com) and dial a CPDLC G/G button and ABSR answers automatically and takes the command and sends it to the Radar screen and the ATCo accepts it and sends it to the CPDLC application (same Control System as above).
- Adaptations of CPDLC messages needed for each ANSP, in the short term ABSR will have to use non-standardized interfaces of the ATM companies.
- For the applications Pre-filling of radar labels, Pre-fill of CPDLC message, Pre-Fill of the flight strips the same solutions might be used:
  - Use already vendor specific standards provided by ATM System providers.
  - Long term build up your own standard (Idea, look into FIXM or SWIM)

Question 7: Accepted delays for Notifications and commands

Answer Q7:

- In the scenarios described when the ABSR sends a notification or command towards Radar, Flight strips or CPDLC, what delays would be acceptable for operations:
  - Readback Error indication (ATCo/Pilot communication loop ended with a readback error, the ATCo would like to see the readback error indication in **1 second**)
  - Pre-fill the Radar labels (since the communication has started the ATCo would like to see the radar labels pre-filled in **1 second**)
  - Highlight the radar labels (since the communication has started the ATCo would like to see the radar labels highlighted in **1 second**)
  - Pre-fill of flight strips (after the ATCo and Pilot communication has started the ATCo would like to see the flight strips pre-filled in **1 second**)
  - Pre-fill the CPDLC message (after the ATCo speaks the message the ATCo would like to see the CPDLC message on the screen in **1 second**)
- The majority or participants agreed that the timers should be less than 1 second (to keep in mind the sync with the update of Radar screen data not to add delay). MLAT is faster so we can go lower. The other participants agreed with 1 to 2 seconds but no more than 1.5 seconds

Other questions were addressed to the audience but they have been removed and not discussed due to lack of interest:

- What other options can we use to receive the Voice Data from VCS into the ABSR?
- What other weather data received from the weather system do you think is useful for the ABSR?
- How can the ABSR get integrated with the Weather Systems, are there vendor specific protocols, can it be done?

4 Second Stakeholder Workshop September 2022

The second Stakeholder Workshop took place at NATS facilities from 27th to 28th of September 2022 and was integrated into a demo day. All project members from HAAWAII, as well as partners from industry and research attended at the workshop and demo day. The workshops were hold on the second day of the event.

Presentations Day 2

Agenda Day 2: Wednesday 28-09-2022

UK local time (GMT +01:00).

Start	Duration	TOP	How
09:00	10	Welcome back	Talk
09:10	30	Using Automatic Speech Recognition for Workload Metrics Extraction – Analysis and Results	Presentation
09:40	75	Workshops: • Tuning of ASR Models – Training to new Environment and Data Privacy Aspects • Application of Speech Recognition and Understanding in ATC Context (including CPDLC) • Using Speech Recognition for Workload Prediction – How do we develop a useful Tool for Supervisors and Controllers?	Break Out Rooms – both virtual and on site
10:55	30	Longer Coffee Break	-
11:25	20	Workshop Summaries and Output	Presentation
11:45	15	Demo Day Feedback Session Closing Remarks- Day 2	Talk
12:00	Open Ending	Demonstration (hands on session): • See live speech recognition on Heathrow Approach	Attendees trying prototypes developed by HAAWAII

		• Try it yourself – challenge the speech recognition software to understand you in the role of a pilot/controller • Guess the workload – subjective and objective workload scores	

4.2 Tuning of ASR Models – Training to new Environment considering the Data Privacy Aspects

The main question from the audience was always, "How many voice recordings we need to adapt the HAAWAII approach to our intended environment". There was no clear, answer like "if you add 10 hours you can expect a WER of x % and a command recognition rate of y %". It always depends on the noise level, how different it is from already modelled accents, the phraseology deviations etc. There is no guarantee that a certain number of silence-reduced hours for finetuning a large pretrained speech recognition model will lead to a guaranteed WER above a certain percentage.

The best approach is here to record two hours of voice recordings together with the corresponding surveillance data., transcribe and annotate the data and solve the data privacy issues. Then a first test with automatic transcription and annotation can be done. Everybody knows then the current status and then it can be decided, which steps have the most potential to get the best results for the available budget.

It can, however, be stated, that the availability of surveillance data dramatically influences the recognition performance on semantic level, i.e. an improvement of 10-20% relative on command level can be expected (i.e. only 75% instead of 90%). On word level the availability of surveillance data is also helpful, but the influence is not so important, maybe a relative reduction of the word error rate (WER) by 5% relative can be expected, i.e. 4.8% instead of 5%.

Experience especially from the US participants in the workshop shows that once a certain diversity in the in-domain data is achieved, e.g. 30 facilities with 10 from each domain (tower, approach, etc.), then the fine-tuned model can generalize better and requires much less labelled, i.e., manually transcribed and annotated, data from a particular facility to achieve good WER. Toward this point of data robustness, MITRE e.g. found a model finetuned with 60 hours from one facility achieves good WER on that facility, but the WER doubled when applying that model to other facilities. In the opposite direction a model finetuned with 100 hours from over 80 facilities achieved good WER on many facilities, with smaller variations by facility.

Another thing to add is the importance of the optional language model used during decoding with large pretrained and finetuned speech recognition models, especially in low resource situations. MITRE has found that the decoding language model was important when the model was finetuned with less labelled audio data. When more than 100 hours of labelled training data are available, the weight of the language model could be decreased or even set to zero, The availability of push-to-talk also has a big effect on word level and therefore also on semantic level, i.e. on command level. Without push-to-talk the recognition time is slower, expect one second more, because you never know, when the utterance ends. You need to implement voice activity detection (VAT). The word error rate increases also by 20 to 30% relative.

Summary:

Start with 2 hours of test data and then iteratively improve (or stop the project). Provide surveillance data and push-to-talk.

4.3 Application of Speech Recognition and Understanding the ATC environment

4.3.1 HAAWAII ASRU applications

The HAAWAII project has addressed the following ASRU applications

- Callsign Highlighting
- Pre-filling radar labels
- Integration of CPDLC and Automatic Speech Recognition and Understanding
- Workload prediction
- Readback Error Detection Assistant

4.3.2 Workshop questions

- Introduction of each participant.
- What is your organization planning with a speech recognition application?
- What application is important from the list above?
- Are you using any speech recognizer currently, what speech recognizer are you using?
- How do you see the CPDLC integrated into the Operational environment?

4.3.3 Workshop results

New Possible applications beside the ones that were on the list:

- Phraseology checker for ATCOs and Pilots
- ASR transcription of the ATC environment voice from the communication room
- Use Voice automation to reduce number of clicks:
- Example with keywords for CPDLC: Data Speedbird .....
- Example with keywords for Flight strips: Strips Speedbird ...
- Pilot training, integration with the pilot simulator (not Pseudo Pilot)

Important applications that have a chance to go operational:

- Call Sign Highlight

- Replacing Pseudo Pilot in the Simulator and reducing the number of Pseudo Pilots (cost effectiveness)
- Prefilling of CPDLC
- Prefilling of Flight strips

The participants had no speech recognizer in Operational, Leonardo started testing one in 2020.

Most of participants are interested to have an ABSR but the integration costs and integration are quite complicated.

4.4 Using Speech Recognition for Workload Prediction – How do we develop a useful tool for supervisors and also for air traffic controllers?

On the second day of the demonstration days the results of the workload metrics prediction work was presented to the audience. One of the workshop sessions then intended to g into a more detailed discussion following this presentation to understand what participants think should be the next steps to develop this application further.

The following points were discussed and considered valuable further steps for development and application:

- Comparison of different sector types in terms of command combinations to understand baselines for those different types
- Analyse trends across a unit to aim for standardisation of R/T use
- Understand individual R/T "habits" to identify potentially unnecessary workload drivers
- Focus on analysis of ATCO R/T rather than both ATCO and pilots. ATCOs are within ANSP gift to influence, pilots are not
- Post incident and overload analysis – understand build ups to incidents
- Develop individual baselines – normal performance envelope
- Comparisons between ATCOs who use CPDLC a lot and those who do not – help understand workload benefits of new technology
- Understand how "no concept" links to individual workload

5 References

[1] H. Helmke; A. Prasad, T. S. Simiganoschi, J. Harfmann: HAAWAII project: D2.3 One Month of Surveillance and Voice Data from Isavia, version 1.00, 4th March 2021.

[2] H. ARILÍUSSON, T. SIMIGANOSCHI, H. HELMKE, J. HARFMANN: HAAWAII project: D1.1: Operational Concept Document, version 01.00.00, 19. August 2020.

[3] H. ARILÍUSSON, T. SIMIGANOSCHI, H. HELMKE, J. HARFMANN: HAAWAII project: D6.2: Updated Operational Concept Document, version 01.50.00, intermediate version, 16. May 2021.

[4] H. HELMKE, S. Shetty.: HAAWAII project: D3.3: Four Hours of Silence Reduced Annotated Controller and Pilot Utterances for Iceland, version 01.00.00, 25. May 2021.

[5] H. Helmke, M. Kleinert, O. Ohneiser, H. Ehr, S. Shetty, “Machine Learning of Air Traffic Controller Command Extraction Models for Speech Recognition Applications,” IEEE/AIAA 39th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2020.

[6] H. Helmke, M. Slotty, M. Poiger, D. F. Herrer, O. Ohneiser et al., “Ontology for transcription of ATC speech commands of SESAR 2020 solution PJ.16-04,” in IEEE/AIAA 37th Digital Avionics Systems Conference (DASC). London, United Kingdom, 2018.

[7] Zuluaga-Gomez, J., Sarfjoo, S. S., Prasad, A., Nigmatulina, I., Motlicek, P., Ohneiser, O., and Helmke, H. "BERTraffic: BERT-based Joint Speaker Role and Speaker Change Detection for Air Traffic Control Communications", 2023 IEEE Spoken Language Technology Workshop (SLT), Doha, Qatar, 2023.

[8] Zuluaga-Gomez, J., Prasad, A., Nigmatulina, I., Sarfjoo, S., Motlicek, P., Kleinert, M., Helmke, H., Ohneiser, O. and Zhan, Q. "How Does Pre-trained Wav2Vec2. 0 Perform on Domain Shifted ASR? An Extensive Benchmark on Air Traffic Control Communications", 2023 IEEE Spoken Language Technology Workshop (SLT), Doha, Qatar, 2023.

Appendix A June 2021 Stakeholder Workshop Working Groups

The following appendix contains the detailed feedback and results of each working group….

A.1 Working Group "What is a Readback Error?"

A.1.1 Sub WG 1 led by Hartmut Helmke

What's a readback error? It depends…

- By the book – easier to decide
    - But is 100% by the book realistic from ATCO and pilot perspective?
- Key question: an error worthy of ATCO attention?
- ATCO has flexibility in operation – can interpret context of situation to decide whether it is important to correct readback
    - E.g., "I knew the pilot *meant* the correct thing even though the pilot did not *say* the correct thing or say it by the book"
    - E.g., "I know it was the correct pilot even though callsign was not read back"
- Perhaps provide readback statistics to inform ATCOs about possible frequency of notifications
- Future (tomorrow?) discussions:
    - Important to consider how ATCO will be notified
    - How to handle the possibility of speech recognition error?

A.1.2 Sub WG 2 led by Matthias Kleinert

The summary of this working group was:

- Readback error should be put into different categories, e.g. critical, not important etc.
    - Check of the ICAO Phraseology could be valuable to find the category, but maybe this is another application
    - Missing digit in frequencies which occurs quiet commonly might not be as critical as other things
- System should notice whether unit is not repeated by the pilot and figure out from the value or the context that a flight level or speed was meant. If this is possible no reaction of the ATCo is necessary.
- Context could be important – time difference between communication or if somebody spoke in between
- Not all elements of the ontology are equally important from commands, e.g. maybe unit not as important if system can detect from context what was meant
- Qualifier: Qualifier is always important to be repeated. However, in most cases missing qualifiers are not corrected by ATCos

- If the turn direction is not important, then ATCO would not mention it in the command, but if the ATCO mentions it, it means it is very IMPORTANT.

A.1.3 Sub WG 3 led by Jules Harfmann

What is a read back error?

1. ***Any*** transmission that deviates from what the controller has said to the pilot or the pilot has said to the controller when it is read back to them by the receiver. (***i.e. all errors notified***)
2. A discrepancy between what the transmitter has said and the receiver has read back that meets certain safety or performance criteria and ***should be corrected***. (i.e. a graduated set of errors according to an extra algorithm based on environmental and contextual requirements)
3. What the transmitter ***wants to be notified as being wrong*** when read back by a receiver. (i.e. individual choice of error rates)

Detailed feedback to each question

Question 1 – feedback says that it is quite straight forward.

Question 2 - straight forward and no feedback

Question 3 – 2 yes, 5 no and 2 not sure:

a. It is an error if you go straight to ICAO phraseology but the data is correct as read back so it is not a read back error but perhaps an error of phraseology
b. Agreement on this from around the table
c. Julia asks if we would want to see this on HMI
d. Another commenter suggests that any error should always be corrected
e. It was then pointed out busy controllers do not want to correct something that has been understood by all. So, the type of error really matters. Too many alerts can also desensitize controllers.

Question 4 – mixture of answers:

a. Does it matter which units are missed?
b. The qualifier disambiguates. So because he said ‘Speed’ first it means that he does not necessarily need knots attached.
c. Missing double qualifier ‘speed” and “Unit” missing so therefore it should be a read back error for safety matters
d. With very short windows to detect, we need to be accurate with the judgement. It was stated several times though that this one is quite obvious when both the qualifier and unit
e. It was raised though that because the heading was correct, therefore there are only two options left and if radar confirmed the speed slowed down then you would not necessarily need an alert of read-back error

Question 5 – almost everyone said it was a read back error, but it was raised – at what point do you want the detection or notification from the system?

a. The table agreed that the error should really be notified at the end because that's when it is most crucial
b. It depends what system the ANSP has. Some have the ability to play back the read back quite quickly and others do not
c. At what point does the system actual do the checking? It was raised that in this case, the error happened the second time it was stated but that the system might not actually know this is related to the previous one

Question 6 – mixed results here

a. Can the ATCO tell the voice of the pilot? If they can, does it matter? When you err on the side of bothering the pilot less
b. Can the algorithm be tailored to an environment? In the case of the US for example, a British call sign might mean the read back errors are triggered less than the more common call signs. This could be a requirement stated by the customer

Question 7 –

a. In the US the leading figure is not mandatory which means that it is not necessary there.
b. European answers differently though saying that this is definitely a blatant error
c. In Austria it was noted that controllers require three figures for a heading and therefore this would be a read back error and is requested to be notified as one

Question 9 – read back without condition

a. Without the condition it would be expected that the Pilot turns. Therefore, this should be a read back error – because its not certain if the pilot got the message about when to turn. The conditional clearance is mandatory and therefore this is clearly a read back error

Question 10 – readback without qualifier for RIGHT

a. It was discussed whether if it was obvious not to turn 340 degrees (e.g. a long turn) then would it matter?
b. It was answered that this does happen sometimes and so the qualifier is absolutely needed.

Question 11 and 12 – similar call signs in the air

a. The algorithm should be able to tell that there are no other flights in the air. Can the algorithm also detect if it's the first contact or not? If it is not then it might not matter
b. With 12 – it was obvious that when another known flight is in the air the table agreed that there needs to be a read back error

Question 13 missing letter in pilots call sign

a. If the algorithm is not so finely tuned then it was recommended that the algorithm should err on the side of accuracy
b. It was asked to the US representative what the feedback has been from controllers about too many alerts. It was noted that too many alerts cause more errors and also annoys controllers more than they used to. If you over alert the controller you could potentially cause them to make more errors.

Question 14 – there was clear feedback that everyone thought it was a read back error.

a. Would you still want it to be notified immediately? Or should the system wait?
b. ACG suggests that the error should be notified immediately. Due to speed of things changing in the air. Nobody disagreed.

Question 15 – when to bother the ATCO?

a. Uniform agreement that it was a read back error and the controller should be notified.

Question 16 – cultural differences identified again

a. In the US flight levels stop at 100 and then go to feet. So, this is different to Europe and hence why there was a little disagreement with whether this was a read back error or not.

Overall – the context and environment were a theme throughout, but generally there was agreement that read-back errors need to be notified to controllers.

For consideration:

1. Qualifiers and units? If you have one, do you need the other?
2. When the system alerts? Immediately or waits?
3. There should be a level of 'importance' of notifications
4. When does the system actually make a judgement?
5. Standard phraseology rules? How much can you bend them?
6. Is it more dangerous to stick to phraseology and over-alert controllers?
7. What about interruptions to conversations? How does the system deal with this?
8. Can voice recognition consider other data from context (e.g. surveillance?)
9. Do controllers want to know every error? – when is the system more accurate than the person?