Stakeholder Workshop Report

D6-2

MALORCA

Grant:	**698824**
Call:	**H2020-SESAR-2015-1**
Topic 2:	**Data Science in ATM**
Consortium coordinator:	**DLR**
Edition date:	**26. February 2018**
Edition:	**1.00**

Founding Members
EUROPEAN UNION
EUROCONTROL
SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Hartmut Helmke	Project Leader	2018-02-07
Ajay Srinivasamurthy	Working Group 1 Leader	2017-04-20
Aneta Cerna	WP1 Leader, Workshop 1 leader	2017-05-11
Christian Windisch	WP2 Leader, Workshop 2 leader	2018-02-21
Christian Kern	Dissemination Manager	2018-02-21

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Hartmut Helmke	Project Leader	2018-02-26

Approved for submission to the SJU by — Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Hartmut Helmke	Project Leader	2018-02-26

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date

Document History

Edition	Date	Status	Author	Justification
0.01	2017-04-19		Hartmut Helmke	First Version with structure and working group for Proof-of-Concept-Trials
0.01	2017-04-20		Ajay Srinivasamurthy	Workshop Topic 1 summary added from Aneta's summary slides
	2017-05-11		Hartmut Helmke	Review and minor corrections
	2017-05-11		Aneta Cerna	Workshop feedback session created, minor changes in the chapter 3

	2018-02-07	Hartmut Helmke	Abstract added
	2018-02-22	Christian Kern	2nd Workshop summary incorporated (chapters 5., 6., 7.)
0.50	2018-02-22	Hartmut Helmke	Review and many improvements; acceptance of all previous changes of Hartmut, Ajay, Aneta and Matej
	2018-02-24	Christian Kern	Some additions, photographs, general conclusion, all previous changes of Hartmut accepted, review
0.99	2018-02-26	Hartmut Helmke	Final Review; acceptance of all changes of CK
1.00	2018-02-26	Christian Kern	Acceptance of all changes
	2018-02-26	Hartmut Helmke	Version number set to 1.00

MALORCA

MACHINE LEARNING OF SPEECH RECOGNITION MODELS FOR CONTROLLER ASSISTANCE

This project has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 698824.

Abstract / Executive Summary

Recently, the project AcListant® has achieved command error rates below 1.7 % for automatic speech recognition based on Assistant Based Speech Recognition (ABSR). It was validated that not only significant controller workload reductions were possible, but also significant improvement for ATM efficiency. One main issue to transfer ABSR from the laboratory to the ops-rooms is its costs of deployment. Currently each ABSR model must manually be adapted to the local environment due to e.g. different accents and deviations from standard phraseology. MALORCA proposes a general, cheap and effective solution to automate this re-learning, adaptation and customisation process to new environments, taking advantage of the large amount of speech data available in the ATM world.

The first stakeholder workshop was conducted in April 2017 with 58 participants. The second stakeholder workshop took place in Vienna in February 2018. The workshops consisted of presentation to the stakeholders and more important different and parallel working groups with a limited number of participants. The outputs of the working groups are in this document.

Table of Contents

1 Introduction

This document is the deliverable D6.2 of the MALORCA project to the SJU [1]. It summarizes the two stakeholder workshops of the MALOCA project. The first stakeholder workshop was hosted by Air Navigation Service Provider of Czech Republic (ANS CR) on the 12th April 2017 in Prague and the second workshop was hosted by Austro Control from 20th to 21st February 2018 in Vienna.

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

2 First Stakeholder Workshop, Prague April 2017

The first Stakeholder Workshop took place on April 12th 2017 at the facilities of the Czech air navigation service provider ANS CR. Different topics of the project have been presented and discussed during the workshop, see the agenda below:

- 09:00-09:30 Welcome Coffee
- 09:30-09:45 Workshop opening, The view of SJU
- 09:45– 10:15 MALORCA and its forerunners
- 10:15-11:00 ANSP perspective
- 11:00-11:30 Coffee break
- 11:30-12:30 Researcher perspective and challenges
- 12:30-13:30 Lunch
- 13:00-14:00 Demonstration corner
- 14:00-15:00 Workshop 1, 2, … running in parallel
- 15:00-15:30 Break
- 15:30 -16:15 Individual workshops findings
- 16:15- 16:30 Workshop Wrap-up, Q&A and Workshop End.

The following **Figure 1** shows one of the main outcomes of the Stakeholder Workshop. There is a clear distinction between the roadmap for introduction of speech recognition application in ATM on the one hand and machine learning for speech recognition applications on the other hand.

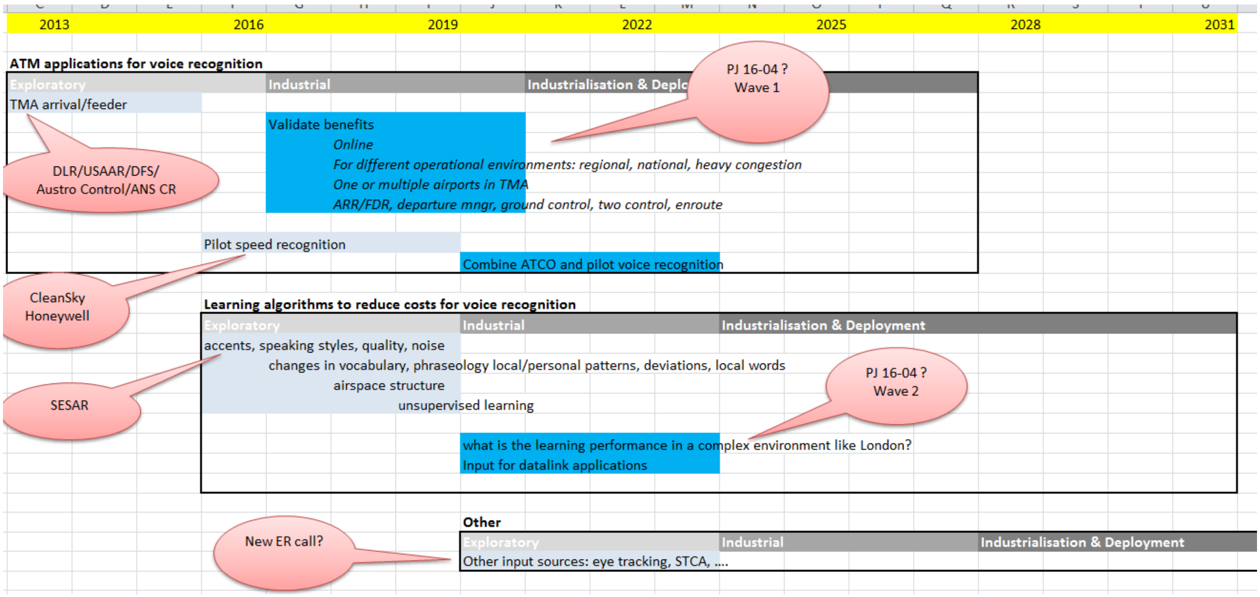


Figure 1: Roadmaps for Speech Recognition in ATM in general and Machine Learning for Speech Recognition

3 Working Groups of 1st Stakeholder Workshop

Four working groups (WG) were organized during the afternoon session.

- WG: Quality of audio signal and access to data
- WG: ASR-user Interaction and Phraseology deviations and locality variations
- WG: Proof-of-Concept Trials December 2017 / January 2018
- WG: ASR in ATC, what is next? How do we continue?

The results of each working group are presented in the following sections.

3.1 WG: Quality of audio signal and access to data

Moderator: Ajay Srinivasamurthy (IDIAP)

Assistant: Martin Jelinek (ANS CR)

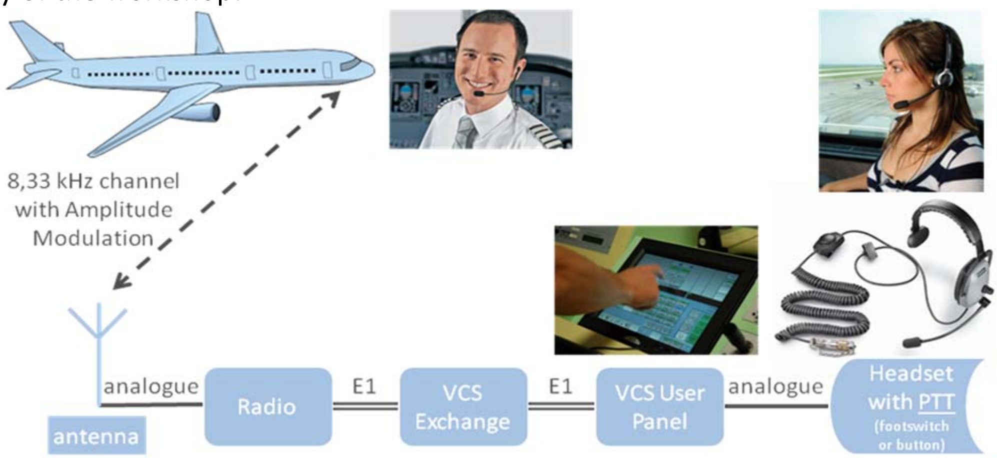
Aim: "ASR performance can be significantly improved if we have access to the signal before it is transmitted. This workshop should discuss how should that be done and to what extent it is feasible."

Agenda/Issues to be discussed:

1. Current VCS (Voice Communication Systems) possibilities:
    - line types used (analogue lines, E1),
    - situation on the market regarding migration to VoIP,
    - planned renewals.
2. Technology limitations:
    - 8,33 kHz Air-Ground channel with telephony voice signal bandwidth of 3,4 kHz,
    - sampling frequency of only 8 kHz, while speech recognition works significantly better with 16 kHz sampling.

3. Noise in the recordings (multiple controllers speaking simultaneously, channel noise, cockpit noise).
4. Integration of auxiliary microphones with better quality to the working positions / VCS panels.

Summary of the workshop:



Current voice chain scenario

Ad. 1. Current VCS (Voice Communication Systems) possibilities:
Current (E1/PCM) VCS user panels do not have interfaces capable of transmitting the voice data in any other interface and codec (coder-decoder) than E1/A-law – it is custom-made hardware, manufactured in small series and tuned for years, with no Ethernet/IP interfaces, expected today.
With future, VoIP VCS user panels, the chance of modification to include an output interface, sending the controller's voice using a higher sampling rate will be significantly easier, as these devices are internally typically Linux PCs, with standard Ethernet/IP interfaces.
VCS renewals towards VoIP are not foreseen to happen massively in the next few years, as the systems are not mature enough for broad deployment. Instead, ANSPs are doing mid-life renewals of currently operated E1 VCS systems.

Ad. 2. Technology limitations:
The entire voice communication chain is built and tuned to the available (telephone) quality of telephone lines and the limited bandwidth of Air-Ground radio channels:
The only codecs currently used in ATC (both E1 and VoIP) are:

- ITU-T G.711 PCM µ-law codec at 64kbps (default in North America & Japan),
- ITU-T G.711 PCM A-law codec at 64kbps (default elsewhere);

ED-137 (the standard for VoIP communications in ATC) allows two more options, both with even smaller bandwidth, i.e. worse quality:

- ITU-T G.728 LD-CELP (Code excited Linear Prediction) codec at 16kbps,
- ITU-T G.729 CS-ACELP (Algebraic CELP) codec at 8kbps.

Due to channel bandwidth (8,33 kHz), regulative limitations (ED-137 allows only A-law and u-law), there's no chance to get a higher sampling rate than 8 kHz in the direction from Air to Ground (i.e. including the pilot's readback).

The only chance to get a signal with better sampling rate than 8 kHz is for the direction from Ground to Air (controller's voice), forking the signal in the VCS user panel, before it gets digitized using one of the standard 8 kHz codecs. However, modifications of current VCS user panel firmware and addition of a dedicated network interface for distribution of voice data for speech recognizer will have to be done and such a modified VCS user panel will have to be recertified.

Ad. 3. Noise in the recordings:

Background noise in the recordings can significantly increase the error rate of speech recognition. However, there's no realistic chance to mitigate the noise neither in the cockpit, nor in the operations room.

Ad. 4. Integration of auxiliary microphones:

Frequency responses of Air Traffic Control – certified headsets:

Sennheiser SC 260 ATC/C3: 150 - 6800 Hz

Sennheiser HME 46-3S: 100 - 12000 Hz

Plantronics headsets for aviation: 300 - 5000 Hz

Most headsets are tuned to the telephony voice bandwidth, so there's a similar situation to VCS user panels. However, a chance of convincing headsets' manufacturers to develop and certify a modification of headsets with wider bandwidth microphones is significantly smaller than in the case of VCS.

A possible mitigation of the noise and low bandwidth issues is more data (voice recordings, metadata and transcripts). With additional data, we can build better ASR models that are robust to noise and poor quality. The amount of data depends on expected performance levels. Additional data is hence a quicker and cheaper mitigation strategy.

3.2 WG: ASR-user Interaction and Phraseology deviations and locality variations

Moderator: Christian Kern (ACG)

Assistant: Aneta Cerna (ANS CR)

Goal of the workshop is to open a discussion about possible ways to deploy the speech recognition output as well as discussion how to deal with phraseology deviations and locality variations in ATC world. More specifically, topics below are proposed to be discussed:

- Level of interaction between ATC controllers and speech recognition engine in operational room (foreground/background, active/passive interaction)
- HMI with minimum controller effort (intuitive vs. more complex/rich, ...)
- Strategies to update phraseology deviations within the ATCO's user-interface
- Minimally invasive approach for including deviations and variations

- Workflow to update known (automatically detected) deviations
- Standard phraseology training for ATCO´s

The group consists of representatives of the following enterprises: ACG, ANS CR, UdS, Harris Orthogon, Honeywell, LPS, NAVIAR, LFV, DLR, CCL, IAA, FAA

<u>Level of interaction between ATC controllers and speech recognition engine in operational room (foreground/background, active/passive interaction)</u>
No additional (new) task should add for the ATCO, no interaction between the speech recognition system and the ATCO should be required.

- Is there a need for confirmation, how should confirmation be given in case yes?

  Four options:
  - Click on the label / anywhere on the screen / separate button
  - Automatic confirmation after a specified timeframe, manual input only in case of wrong recognition or if manual confirmation is particularly required (needs specification)
  - Highlight of the label (shaded out)
  - Table of commands somewhere on the screen (x too many lists already there, not enough space on support screens, controllers need to change focus)

<u>HMI with minimum controller effort (intuitive vs. more complex/rich, ...)</u>

- There is a need for reading the value anyway, to prove the correctness of a certain input.
- During heavy traffic the importance of updated information is pretty high and desired (flight information – flight level, heading), tendency to use standard phraseology is increased.
- Active click for datalink application (CPDLC) might be designed.
- Support for coordination
- Future use - List of the instructions for the read back monitoring, no need to go to the VCS system and checking the instruction.

<u>Strategies to update phraseology deviations within the ATCO's user-interface</u>
<u>Minimally invasive approach for including deviations and variations</u>

- Deviations must be allowed and should be modelled
- The goal – user specific model
- If the user has the positive feedback and see clearly the benefits, then he might tend to follow the standard phraseology

<u>Workflow to update known (automatically detected) deviations</u>

- The changes need to be identified, modelled and learned
- Dataset maintenance
- Online learning vs. deployment of the changes – some delay in the update will be there
- Online updates of statistic model – no immediate impact (some period of time is necessary)
- Some updates could be done in the SIM environment already

Standard phraseology training for ATCO´s

- Speech recognition could give trainees feedback whether or not they use standard (respectively correct) phraseology and specific training can be developed.

3.3 WG: Proof-of-Concept Trials December 2017 / January 2018

Moderator: Hartmut Helmke (DLR)

Assistant: Matej Nesvadba (ANS CR)

The general topic of this WG is to discuss and refine ideas for the proof-of-concept trials performed in Prague and Vienna in December 2017/January 2018, which results in the following questions:

- How to directly involve ATCOs in the Proof-of-Concept Trials?
- How to create a demo an ATCO can touch and play with?

The basic idea is to compare different versions of the speech recogniser, i.e.

- the baseline system
- the improved system, which benefits from machine learning

The goal is to validate that on the same set of input data, the MALORCA developed speech recogniser has better performance than baseline ABSR.

Results of working group:

First the working group discussed that controllers prefer to validate an Automatic Speech Recognition Application in an ATM scenario. That, however, is not the focus of the MALORCA project. This was addressed in the AcListant® and AcListant®-Strips project of DLR and USAAR. MALORCA needs to show in proof of concept trials that MALORCA can adapt a **basic ABSR system** much **cheaper** with machine learning techniques to new approach areas and controller acceptability (with respect to recognition rate) is still given, see also **Figure 1**.

Two different proof-of-concept experiments which included ATCOs interaction were discussed.

Figure 2 shows the basic setup of the first idea.

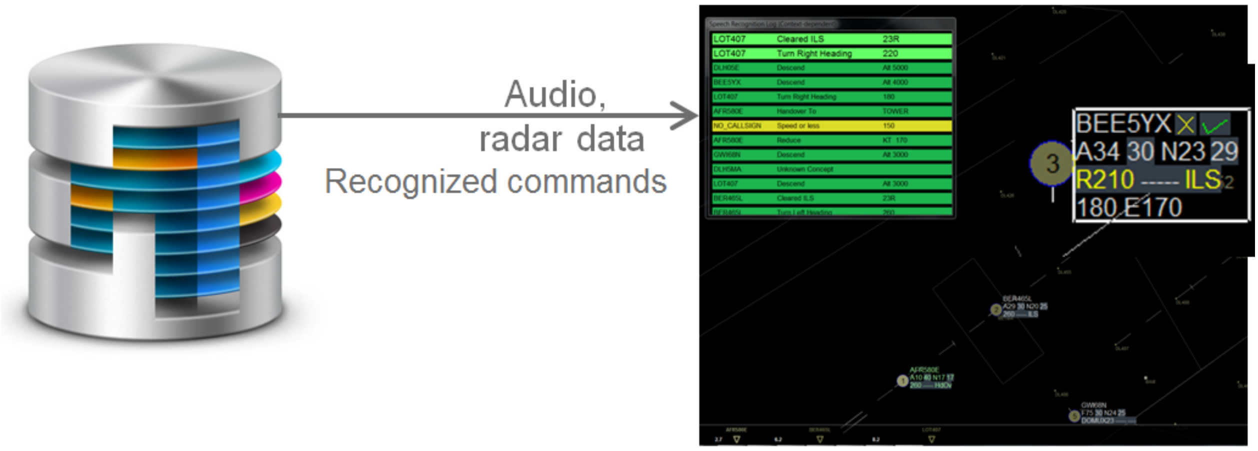


Figure 2: Basic Setup of Proof-of-Concept-Trial with Replay of Radar and Voice Data

Speech data and radar data are recorded and transcribed. After the recording phase, the voice utterances are presented to the controller together with the online recognition results of the trained ABSR system. The controller listens to the voice recordings (his/her own or those of a colleague) and sees the corresponding radar data. The speech recognizer tries to recognize the voice recordings in real time, i.e. no recorded recognitions are presented. As shown in **Figure 2** the output of the speech recognizer is shown in the radar label in yellow colour (BEE5YX REDUCE 210, and BEE5YX CLEARED_ILS 23L in the example in **Figure 2**). The speech log (in upper left corner in **Figure 2**) is not shown to the controller by default, but can be shown on request.

If the speech recognizer fails (false recognition or recognizes nothing) the controller has to input the command sequence manually. This may include a rejection of the false recognition. Normally the controller will use mouse and keyboard for correction. However – not discussed in the working group – also voice could be used for that task.[2] The scenario should last approx. 60 minutes.

[2] The challenge might be that two input channels exist for the online speech recognizer, the recorded wave files and the live input from the controller. In seldom case there might be two inputs nearly at the same time.

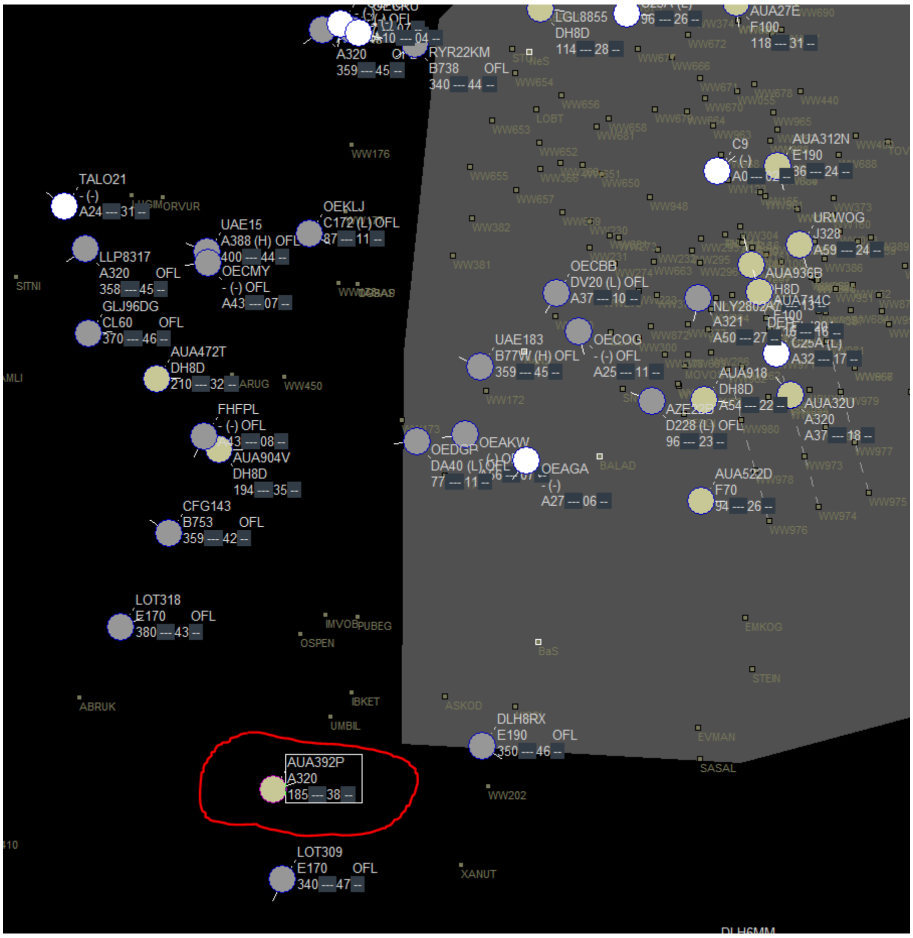


Figure 3: Traffic Situation when speaking to AUA392P in the comparison experiment

The second idea for the proof-of-concept trials is based on pairwise comparison of the output of two different speech recognizers, one could be the basic ABSR system developed in WP3 and the second the trained one improved in WP4. Both speech recognizers run on the available recorded voice and radar data sets. Recognitions resulting in the same command sequences (from both speech recognizers) are ignored in the further analysis. Only recordings which result in different outputs for both recognizers are further analysed. N (e.g. 200) differences of both systems are randomly selected[3] (knowing the correct transcription is not necessary). Each voice utterance is played to the controller. The controller can listen to it and also replay it (as often as he/she likes). The corresponding radar situation plus the output of recognizer one is shown to the controller.
Figure 3 shows an example.
Recognizer 1 understands the following word sequence

[3] The selection process must be (nearly) random, otherwise we might get a bias in the experiment outcome, i.e. the selections may favour the trained ABSR system.

niki three nine two papa servus ils approach runway three four cleared the balad three november transition reduce speed two twenty no level restriction

resulting in the following word sequence

NLY392P EXPECT_ILS 34, NLY392P TRANSITION BALAD_3N, NLY392P REDUCE 220

Recognizer 2 may understand the following word sequence

niki austrian nine two papa servus ils approach runway three four cleared balad three november transition reduce speed two twenty no level correction

resulting in the following word sequence

NO_CALLSIGN EXPECT_ILS 34, NO_CALLSIGN TRANSITION BALAD_3N, NO_CALLSIGN REDUCE 220

Now the controller has to judge which command sequence output he/she prefers. The output of this experiment is how often recognizer 1 is preferred to recognizer 2. The controller does not know in advance which recognizer is in which case the basic one and which one is the trained one.
The characteristics and also the pro and cons of both approaches were discussed by the working group participants.

Replay of Voice and Radar Data	Comparison Experiment
Recognition Speech is measurable	
Absolute Values concerning recognition rate are observable by controller	Comparison of two Speech Recognizers possible
Controller is more involved in experiment, i.e. doing is normal work	Controller can concentrate on ASR output
Indirect workload measurements are possible (time for radar label maintenance)	

Table 1: Characteristics of both Proof-of-Concept Ideas

At the end of the working group the eight participants were asked to make a final decision: Either selecting “Replay” or “Comparison” Experiment. The secret voting ends in three votes for “Replay” and four votes for “Comparison”.

After the working group it was decided (not in the working group) to present both ideas to one Austro Control controller and one ANS CR controller in Sep/Oct. then improving the experiment setup due to their feedback and finally deciding in Oct/Nov whether the Replay or Comparison setup or none of them should be used for the final Proof-of-Concept trials.
It must be emphasized that the two controllers participating in the discussion highly preferred a real traffic simulation setup as it was performed during AcListant® trials in Braunschweig. This approach, however, is not possible due to MALORCA budget restrictions and is not the focus of the MALORCA project.

3.4 WG: ASR in ATC, what is next? How do we continue?

Moderator: Jürgen Rataj (DLR)

Assistant: Christian Windisch (ACG)

Comment: Applications of ASR in ATC

First part

- What is the vision of the ANSP about ASR in:
  - 10 years
  - 20 years
- What are the benefits the ANSP expect from ASR in those time frames?
- What is the vision of the suppliers about ASR?

Second part

- Is learning from data only a hype in the research?
- Is there a useful successor of MALORCA exploring self-adapting and self-configuring ATC-systems?
- Which step is then next most beneficial step?

What is the vision of the ANSP about ASR in the future:

- ASR will be a part of the future even if CDPLC is coming, because the controller may use voice as a beneficial input media
- ASR has the potential to create additional benefits on other working positions, e.g. planner positions at London airspace
- ASR will also be an enrichment for training not only by substituting the pseudo pilots, but also by self-education of the novice

What are the benefits the ANSPs expect from ASR?

- Lower workload of the controller
- Higher acceptance for support tools

What is the vision of the suppliers about ASR?

- ASR will be a valuable, but complicate (to maintain) input mean

Is learning from data only a hype in research?

- Learning will be an important mean for future system developments

Is there a useful successor of MALORCA exploring self-adapting and self-configuring ATC-systems?

- no clear answer was found

Which step is then the most beneficial step?

- answer possible if MALORCA has shown its potential

4 First Stakeholder Workshop in Prague, Feedback

All participants of the workshop were asked to give their feedback in the form of several questions, see below:

1. Would you attend the final stakeholder workshop in Vienna yes/no?
2. Would you prefer to have a two half days workshop (from lunch day 1 to lunch day 2)?
3. Would you prefer to have a two days' workshop (from morning day one to afternoon day two), yes/no?
4. Would you prefer more presentation sessions instead of the working groups yes/no?
5. General comments …?

The collected answers are in the following table. The number of respondents was not high, however, it needs to be understand that one received feedback represented several participants from the same company. We expect that it will guide us in the right direction when preparing the final stakeholder workshop at the end of the project.

Workshop feedback (12 respondents)	**Positive answer**
1. Would you attend *the final stakeholder workshop in Vienna?*	**92%**
2. Would you prefer to have a 2-half days workshop (from lunch day 1 to lunch day 2)?	**67%**
3. Would you prefer to have a two days' workshop (from morning day one to afternoon day two)?	**33%**
4. Would you prefer more presentation sessions instead of the working groups yes/no?	**25%**

Table 1 - Workshop feedback

General comments were kind of positive feedback on the workshop organization itself. No objections.

5 Second Stakeholder Workshop, Vienna February 2018

The second and final Stakeholder Workshop during the MALORCA project took place on February 20th and 21st 2018 at the facilities of Austro Control. The overall title of the workshop was "Machine Learning Benefits in Speech Recognition Systems for Air Traffic Control" and the invitation outlined the following topics to be covered by this workshop:

- Presentation of the final project outcomes
- Live Proof of Concept Demonstration on ATC Operational Data
- Discussions in Working Groups

The invitation to the workshop was intended for air navigation service providers, experts in data science, machine learning, speech processing and recognition, and for industry partners. 35 experts followed the invitation to Vienna representing among others NATS, Airbus, DFS, FRAPORT, Honeywell, Frequentis and different European universities.

What has been presented and discussed during the workshop in detail states the agenda (with abstracts) below:

Agenda Day 1: 13:00 to 17:00

13:00 to 13:10 Welcome

13:10 to 13:40 MALORCA brings together Automatic Speech Recognition and Machine Learning. The presentation gives an overview of MALORCA project and its partners. The different roadmaps of Automatic Speech Recognition on the one hand and Machine Learning on the other hand are presented. Furthermore, the agenda of the two days including logistic information is presented.

13:40 to 14:10 Components of Machine Learning for Assistant Based Speech Recognition. The presentation presents the different components, which are needed to train an Assistant Based Speech Recognition System by Machine Learning. This includes e.g. acoustic, language and command prediction models, but also radar data and tool-supported controller utterance transcription.

14:10 to 14:45 Proof-of-Concept Trials Proof-of-concept of MALORCA project is split into two technical (T1, T2) and two operational (O1, O2) activities.

- T1 is a workshop with technical experts to evaluate the ABSR prototype implementation against the technical requirements.
- T2 is an offline evaluation to quantify the improvements of the ABSR system with respect to the amount of available training data.
- O1 involves controllers who concentrate only on the different outputs of a baseline ABSR system and on an ABSR system trained with all the available MALORCA training data.
- O2 puts the trained ABSR system in a simulation environment with a replay of historic radar data and controller voice recordings from real Prague and Vienna in- and out-bound traffic. ABSR is used here to support the controllers in maintaining radar labels. Quantitative and qualitative results and feedback are presented.

14:45 to 15:30 Coffee break including prototype demonstrations.
The speech recognition prototypes which were used for technical and operation proof-of-concept trials are demonstrated on real life traffic from Vienna and Prague (from 2016) and a microphone is available for workshop participants to evaluate performance and limits of Assistance Based Speech Recognition.

Figure 4 below gives an impression of the demo site installed during the workshop and showing a licensensed controller of the Vienna Approach Control Unit issuing different clearances to aircraft to live challenge the ABSR-system:

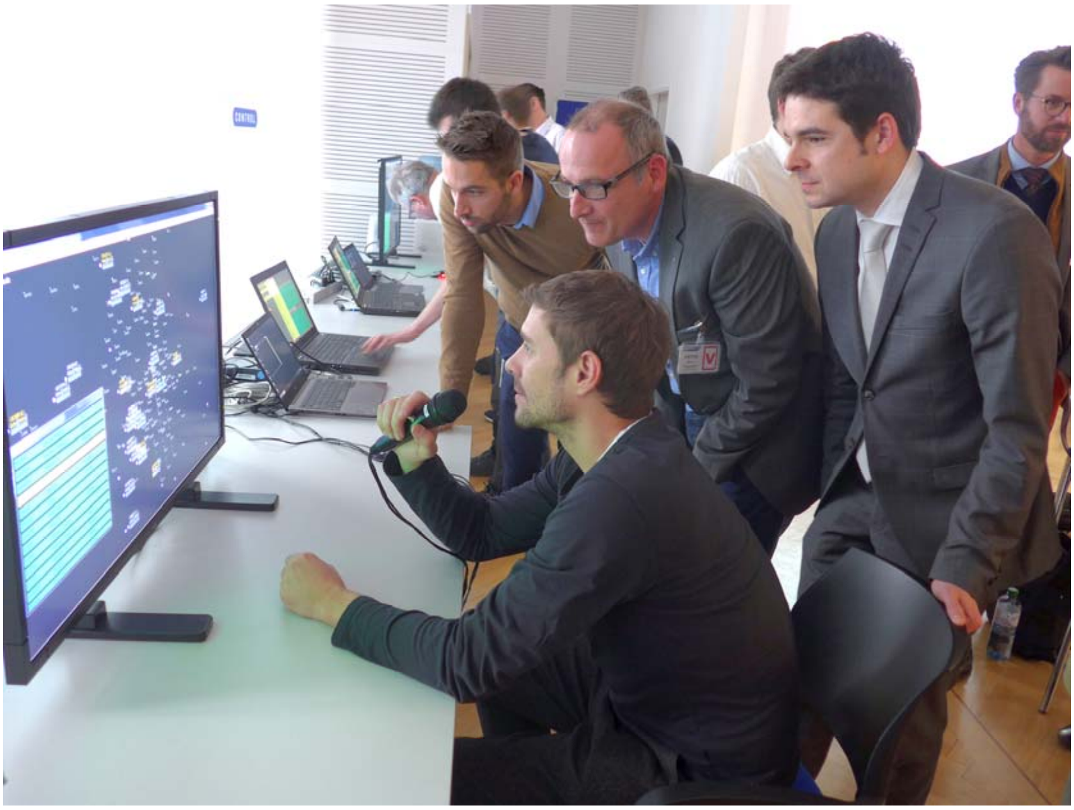


Figure 4: Austro Control Controller together with Dissemination Manager Chr. Klein explaining the Proof-of-Concept Trials and the benefits of the Assistant Based Speech Recognizer to to the Working Shop particiapants

15:30 to 16:30 MALORCA Working Groups Part 1. Three parallel working groups discussed different topics.

16:30 to 17:00 MALORCA Working group result presentation to workshop participants

Agenda Day 2: 9:00 to 13:00

9:00 – 09:15 Introduction, Summary of Day 1, MALORCA video. During proof-of-concept trails MALORCA team has created a video, which will be presented to audience. Furthermore, logistic information is provided.

09:15 – 09:45 Gap Analysis and Next Challenges. Although MALORCA has shown that machine learning techniques enable the cost-efficient transfer of Assistant Based Speech Recognition to new approach areas, many challenges remain. The presentation focuses on these challenges to encourage brainstorming for solutions in the following working groups. The presentation of problems includes a more detailed presentation of MALORCA results which were not covered by Proof-of-Concept trials.

09:45 – 10:15 PM Transfer of Assistant Based Speech Recognition to New Approach Areas. AcListant® project has shown that command error rates below 2% are possible with Assistant Based Speech Recognition (ABSR) for Dusseldorf Area. MALORCA develops Machine Learning techniques to enable transfer of Assistant Based Speech Recognition to new Approach Areas. The presentation outlines which steps are needed to adapt ABSR to a new airport with the support of machine learning and without.

10:15 to 10:45 Coffee break including prototype demonstrations. The speech recognition prototypes which were used for technical and operation proof-of-concept trials are demonstrated on real life traffic from Vienna and Prague (from 2016) and a microphone is available for workshop participants to evaluate performance and limits of Assistance Based Speech Recognition.

10:45 to 11:15 DLR Cooperation of MALORCA and PJ 16-04. Both the MALORCA project and the SESAR solution PJ16-04 are led by DLR. The presentation focuses on speech recognition activities of PJ16-04 and outlines how MALORCA results have already influenced the speech recognition activity in PJ16-04. This includes the development of an ontology for controller command transcription and also focussing on safety aspect which were neither considered in MALORCA nor in AcListant®.

11:15 to 12:15 MALORCA Working Groups Part 2. Three parallel working groups discussed different topics.

12:15 to 12:45 MALORCA Working group presentation to workshop participants.
Each working group will select a speaker and a keeper of the minutes. The speaker will present the output of his/her working group to plenum to encourage discussion for come together in Vienna downtown.

12:45 to 13:00 Plenum speaker Summary.
Critical review of MALORCA objectives and MALORCA achievements with respect to limited SJU resources and lessons learned from perspective of MALORCA, SJU and selected workshop speaker.

6 Working Group Results of 2nd Stakeholder Workshop

Six working groups (WG) were organized during the 2nd Stakeholder Workshop (3 per day):

- WG: Which recognition rates are operationally sufficient and affordable for operational needs? How to integrate even more context information?
- WG: ABSR in the ops room at Easter 2020?
- WG: Active Online-Learning, learning from feedback, using available mouse inputs of the controller or mode-S output or …
- WG: How to exploit MALORCA results (especially with you) PJ.16-04 wave-2, Horizon 2020 …
- WG: Detailed evaluation of Prague prototype
- WG: Challenges and drawbacks of local phraseology deviations? How to handle them

The results of each working group are presented in the following sections.

6.1 WG: Which recognition rates are operationally sufficient and affordable for operational needs? How to integrate even more context information?

Moderator: Christian Windisch (ACG)

Context integration was not discussed. It was as a pre-condition for achieving acceptable error rates.
It was stated that the problem consists of (undetected) errors and not of rejections. The ATCOs are able to handle rejections easily by doing the required input via mouse, just as today.

Statistically the ATCOs themselves have an error rate as well and do wrong mouse inputs. This rate is between 1 and 2 percent. As this error rate should not increase in order to keep the level of safety either with mouse or voice input, the targeted error rate of a speech recognizer should be less than 1%. Such good recognition rate would make speech recognition useable for the ATCOs and would not urge controllers to invest too much effort to correct errors or doing the data input manually respectively.

An impression of the working group at work gives Figure 5 below:

Figure 5: Working Group discussing needed recognition and error rates

6.2 WG: ABSR in the ops room at Easter 2020?

Moderator: Dietrich Klakow (UdS)

The working group identified three major topics that need to be addressed to make it happen:

- implement the system
- increase end user buy in
- get management support

With respect to the implementation there are two basic approaches related to the interfaces (data input and output). The first one would be going via Thales who provides the primary systems for many ANSPs. Given that Thales is involved in PJ16-04, this seems to be possible in principle, however, there was some substantial scepticism, that Thales would actually provide the necessary software interfaces. The workshop identified an alternative via the providers of back-up/secondary systems. These are expected to be more open, because they could use this as an opportunity to strengthen their position. Regarding the radar data using AD-B might be a suitable alternative. In case getting access to the primary screen is a problem an additional screen (e.g. on a tablet) is possible. Regarding bringing the MALORCA software up to the necessary software maturity level required for certification seems possible and specific industry partners have been identified.

To address the second issue, that is the increase of the end user buy in, it was stated that this is very critical and needs to be started as soon as possible. It is important that controllers from the ops room can use the MALORCA prototype such that say can directly convince themselves of the benefits. Standardized user experience feedback needs to be elicited in order to further improve the system. When the actual enrolment in the ops room happens, the benefit has to be clear from the very beginning.

It was agreed that the two above mentioned steps are no show-stoppers when done right. More critical is getting the necessary management support. There was scepticism that it would end up too low on the list of priorities to get actually done. Whether this is really the case or not would need to be explored. However, the working group participants agreed that hooking up with some bigger initiatives (e.g. going to electronic strips) might ease it. Other arguments would be that the system can be used and rolled out for more than just approach even though there - due to its complexity - the benefit is largest.

6.3 WG: Active Online-Learning, learning from feedback, using available mouse inputs of the controller or mode-S output or …

Moderator: Petr Motlicek (IDIAP)

The working group raised many new questions:

- How can we learn from scratch without needing transcriptions?
- How long to improve?
- When learning sufficient with respect to budget and quality?
- Do we need a good recognition rate on average or a minimum recognition rate (to be prepared for the extreme cases)?

Error rate should be below 1%

Also a nearly perfect ABSR system will make errors.
Use the different data modalities (radar, flight strip information, mode-S…)
Run the ABSR system in parallel to collect more data without telling the controller.

6.4 WG: How to exploit MALORCA results (especially with you) PJ.16-04 wave-2, Horizon 2020 …

Moderator: Jürgen Rataj (DLR)

MALORCA addresses several aspects:

- Exploit speech recognition of high quality
- Exploit learning for speech recognition configuration and maintenance
- Exploit learning for assistance system configuration and maintenance
- Exploit speech recognition of high quality (i.e. low error rate)

What is necessary for exploitation?

- Validation of benefits of speech recognition, (PJ 16-04 in a limited way)
- inform many other ANSPs about great results of MALORCA, (presentations, papers)
- Add new applications to rise benefits like
    - recognize read back errors,
    - extend to other working positions,
    - critical word (keyword) spotting,
    - controller workload assessment,
- post processing of speech data for different performance evaluation
- Introduce an expert group for speech recognition,
- Start regulation, standardisation

Who is interested to support this?

- other working position
    - Fraport,

  - NATS,
  - Avinor
- recognize read back errors
  - NATS

Who is interested to use it?

- Fraport,
- NATS,
- Avinor

Exploit learning for speech recognition configuration and maintenance:

What is necessary?

- Going more into the depth of the technology,
- Shorten the implementation time for such systems
- Solve operational problems
  - Controller acceptance

Who is interested to support this?

- NATS,
- Thales,
- (maybe also) Frequentis

Who is interested to use it?

- Due to lack of time the aspect was not discussed.

Exploit learning for assistance system configuration and maintenance

- Not discussed

Conclusions:

- Speech recognition is a technology which creates more benefits as could be shown by the Workshop demonstrations and presentations.
- It is an enabler for additional functions.
- An introduction of an expert group will be useful e.g. concerning standardisation and regulation.
- Cost savings generated for suppliers is difficult to be supported by ANSPs
- Supplies should be interested in such a work.
- It is until today difficult to get a significant amount of data for learning.

6.5 WG: Detailed evaluation of prototypes

Moderator: Matej Nesvadba (ANS CR)

The general block diagram in Figure 6 below gives a better understanding of how the MALORCA System basically works:

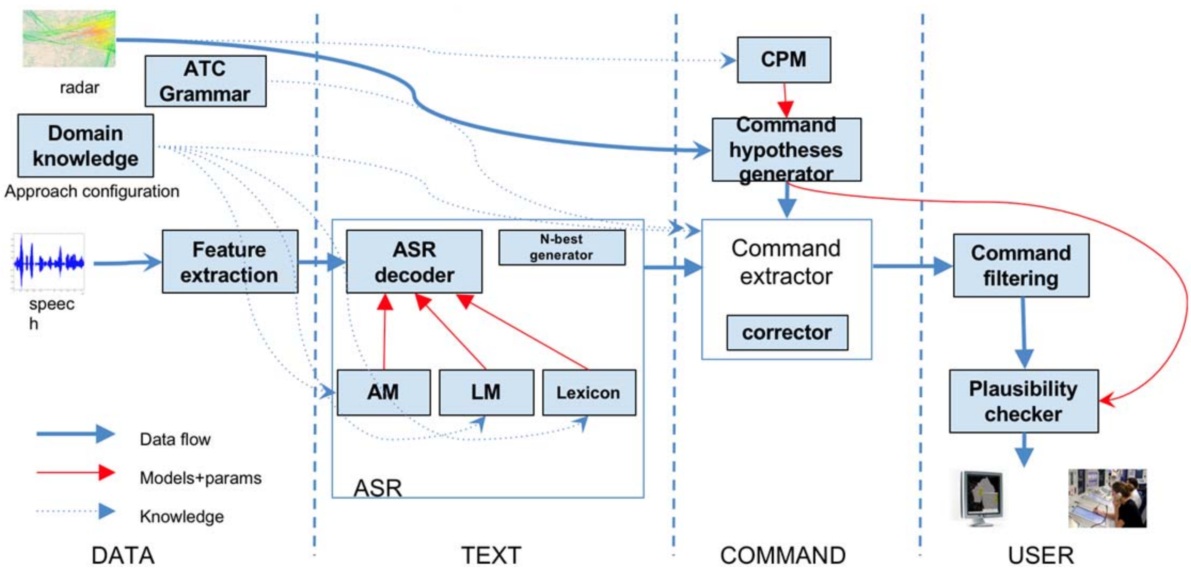


Figure 6: Basic building blocks of an Assistant Based Speech Recognition (ABSR) system

With this in mind the participants of this working group as the users of the speech recognizer challenged the prototypes by giving certain commands. The team discussed all experiences and results and concluded that no significant or even systematic problems were observed. The general feedback of the working group members was that the system works.

Derived from the intensive discussions the working group finally focussed on some particular examples of high relevance in daily practise. Also the results of this exercises were satisfying.

6.6 WG: Challenges and drawbacks of local phraseology deviations? How to handle them

Moderator: Ajay Srinivasamurthy (IDIAP)

Different challenges exist in the working group. Some suppliers of ATC training simulators just see phraseology deviations as not acceptable. They use ASR systems for training purposes, i.e. to train correct phraseology. They have no need to address the problem. They are interested in seeing if ASR can be used to detect incorrect phraseology and provide feedback to trainees.
Phraseology deviations in ops rooms are not exceptions, and hence they need to be addressed. The problem has not been formally addressed. MALORCA's approach is to model phraseology deviations in the grammar if they are common, because we need the buy-in of the controllers instead of blaming them.

MALORCA currently adapts the grammar, but the desired solutions that grammar resp. language models are automatically updated by learning. The grammar contains already more elements for modelling the exceptionally behaviour than for the normal phraseology, although of course most of the utterances are still covered by the normal grammar. This approach involves significant continued manual effort and hence we need to explore ways to automatically learn deviations. Some ideas for automatic learning of phraseology that came up during the discussion: to use multiple ABSR systems to detect and correct phraseology, using end to end systems to learn grammar automatically, learn phraseology automatically with controller feedback and correction.

7 Summary to the 2nd Stakeholder Workshop in Vienna

The workshop has “selected” a speaker in the morning of the second day. It was Nathan Vink from NATS. At the end of the workshop he gave a short summary, which is repeated here again.

The trend from “Less controllers to more traffic” is obvious. Therefore, taking away the additional workload from controller will give back the time.

- How could we use ASR for alternative purposes, i.e. to get more out of it if it cannot deliver the black and white productivity claims (due to other factors – e.g. cannot get more planes into Heathrow etc.)?

- How can it be used for systems where ‘mouse clicks’ are not the primary method of operation?
  For example, the way controllers enter data may not be the limiting problem. Accents may not be the biggest problem – filling the ‘silence’ with chatter may be more of a problem – processing power. Voice over IP vs. Analogue channels and accessing the raw voice data. Will Voice over IP change things?

Challenges:

1. Traffic is increasing and number of controllers (and pilots) decreasing. We need to improve human performance, airspace and technology to have any chance of managing this.
2. From my perspective it requires a real step up in understanding the workloads of controllers now, and how we can help them with that.
3. ASR demonstrates a recognisable step forward in that. It has some challenges and a few gaps still to achieve, but these gaps may prove to be more opportunities. For example, assisting Group Supervisors in understanding their controller’s workloads so they know when to intervene…also the save in time could off-set the problem of delay induced by VOIP.
4. Recognition rate vs error rate = beat the controller and its game on! If the computer has a smaller error rate than controllers then I can get technology assurance for it.

5. I think ASR can be expanded to the 'more complex' areas, but some further understanding of the varying cognitive tasks across European ANSPs may be required – e.g. 'how controllers do their tasks' for example what happens for systems which are not mouse driven?
6. What is exciting though is the opportunity to develop 'new ways of doing business. That may help us achieve increase Human performance. ASR could even be used to help build airspace – perhaps through using it as an indicator of pressure points.
7. Psychology impact of neural networks on understanding antecedents of performance both of the human and the AI interaction – i.e. what is the effect on reduced cognitive feedback that could happen when the system automatically does the work for you.
8. Could ASR perhaps change the way we think about the three mechanisms of communication for controllers?
    a. Verbal – previously all controllers had to go on
    b. Non-verbal – could new and clever HMI represent meaning better to improve the communication methods
    c. Paralinguistic – improving grammar and meaning through clearer representation of communication

8 Conclusion after two Stakeholder Workshops

After the stakeholder workshops being held in Prague in April 2017 and Vienna in February 2018 the project team concludes that by sharing all knowledge we gathered in the MALORCA project with that many experts in this domain a very good basis is created for future development and the implementation of ABSR in ATM-systems. Instantly SESAR2020 PJ16-04 is pursuing this under the participation of 23 organizations from 16 European countries (Figure 7).

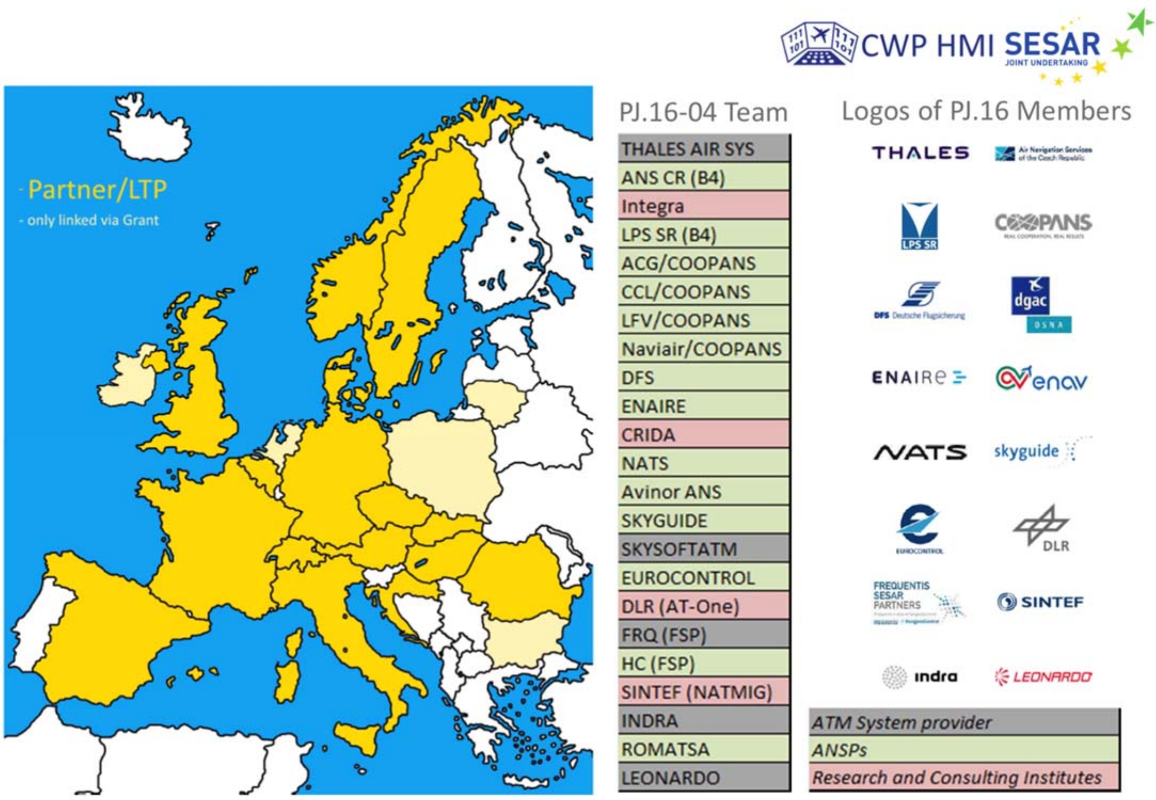


Figure 7: Participants of SESAR 2020 PJ.16-04 solution

The photographs (Figure 8) below show all participants of the 1st Stakeholder Workshop in Prague:

Figure 8: Workshop Participant in Prague April 2017

and of the 2nd Stakeholder Workshop in Vienna (Figure 9):

Figure 9: Workshop Participant in Vienna February 2018

Appendix

Appendix A Abbreviations

ABSR	Assistant Based Speech Recognition
ACC	Area Control Centre
ACELP	Algebraic Code-excited Linear Prediction
Acoustic model	Used in ASR to represent relationship between an audio signal and the linguistic units
ACG	Austro Control Österreichische Gesellschaft für Zivilluftfahrt mit beschränkter Haftung
AcListant	Active Listening Assistant
ANS CR	Air Navigation Services of the Czech Republic
ANSP	Air Navigation Service Provider
APP	Approach Control Unit
ASR	Automatic Speech Recognition
ATCO	Air Traffic Controller
CELP	Code-excited Linear Prediction
COOPANS	COOPeration between ANSProvider
CTX	Ctx-file = context file automatically generated from radar data
Concept generator	Extraction of semantic concept relevant to the task
Context integrator	Combination of ASR hypotheses and context information
DLR	Deutsches Zentrum für Luft- und Raumfahrt (German Aerospace Centre)
DM	Dissemination Manager
DoD	Definition of Done
Hypothesis rule generator	Rule generator
Idiap	Idiap Research Institute
ITU-T	International Telecommunication Union – Telecommunication Standardization Sector
Language model	It represents a probability distribution over sequences of words
MALORCA	Machine Learning of Speech Recognition Models or Controller Assistance
MLS	MALORCA Learning System
LOWW	Vienna Airport
NTP	Network Time Protocol
PIC	Pilot in command
PL	Project Leader
PCM	Pulse Code Modulation
PMP	Project Management Plan
POC	Point of Contact
PRG	Prague
SES	Single European Sky
SID	SESAR Innovation Days
SJU	SESAR Joint Undertaking
tbd	To be defined
TMA	Terminal Manoeuvring Area
TWR	Aerodrome Control Tower
UdS	See USAAR
USAAR	Saarland University
VCS	Voice Communication Systems
VoIP	Voice over Internet Protocol
WP	Work Package