MALORCA 1st Stakeholder Meeting

# Working Groups

MALORCA Project Team

Prague, 12th of April 2017

SESAR JOINT UNDERTAKING

Founding Members

EUROPEAN UNION EUROCONTROL

# Topics – Working groups

Topic 1: **Quality of audio signal and access to data**

- Moderator: Ajay Srinivasamurthy (IDIAP)
- Assistant: Martin Jelinek (ANS CR)

Topic 2: **ASR-user Interaction and Phraseology deviations and locality variations**

- Moderator: Christian Kern (ACG)
- Assistant: Aneta Cerna (ANS CR)

Topic 3: **Proof-of-Concept Trials December 2017 / January 2018**

- Moderator: Hartmut Helmke (DLR)
- Assistant: Matej Nesvadba (ANS CR)

Topic 4: **ASR in ATC, what is next? How do we continue?**

- Moderator: Jurgen Rataj (DLR)
- Assistant: Christian Windisch (ACG)

# Topic 1 – Description
# Quality of audio signal and access to data

Motto:

"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."

Current VCS (Voice Communication Systems) possibilities:

- line types used (analogue lines, E1),
- situation on the market regarding migration to VoIP,
- planned renewals.

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.

Noise in the recordings (multiple controllers speaking simultaneously, channel noise, cockpit noise).

Integration of auxiliary microphones with better quality to the working positions / VCS panels.

# Topic 2 – Description

## ASR-user Interaction and Phraseology deviations and locality variations

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

# Topic 3 – Description

## Proof-of-Concept Trials December 2017/January 2018

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

The general topic of this WG is to discuss and refine ideas for the proof-of-concept trails performed in Prague and Vienna in December 2017/January 2018.

The basic idea is to compare versions of the speech recogniser:

- 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.

# Topic 4 – Description
# ASR in ATC, what is next? How do we continue?

**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?

# Topic 1 – Result 1/2
# Quality of audio signal and access to data

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 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:

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 panels 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.

# Topic 1 – Result 2/2
# Quality of audio signal and access to data

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 and metadata).

# Topic 2 – Result 1/3

## ASR-user Interaction and Phraseology deviations and locality variations

Group representatives: ACG, ANS CR, UdS, Harris  Orthogon, Honeywell, LPS, NAVIAR, LFV, DLR, CCL, IAA, FAA

No additional (new) task for the ATCO, no interaction btw. the speech recognition system and the ATCO

1. Need for the confirmation:
    - Click on the label / anywhere on the screen / separate button
    - Time parameter as automatic confirmation, only in case it is wrong and the confirmation is needed
    - Highlight of the label (shaded out)
    - Table of commands somewhere on the screen ( x too many lists already there)
2. Need for reading the value anyway

# Topic 2 – Result 2/3

## ASR-user Interaction and Phraseology deviations and locality variations

During the heavy traffic – the importance of updated information is pretty high and desired (flight information – flight level, heading), tendency to use standard phraseology.

Active click for datalink application (CPDLC) might be designed.

Support for coordination

Future use - List of the instructions for the readback monitoring, no need to go to the VCS system and checking the instruction.

# Topic 2 – Result 3/3

## ASR-user Interaction and Phraseology deviations and locality variations

Strategies to update phraseology deviations within the ATCO's user-interface

- 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

Workflow to update known (automatically detected) deviations

- 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 neccesary)
- Some updates could be done in the SIM environment already

# Topic 3 – Result 1/5

## Proof-of-Concept Trials December 2017/January 2018

What is the concept?

- It is **NOT** to show that adequate recognition rates are possible
  → AcListant® shows that with ABSR >95% are possible
- It is **NOT** to show that adequate Speech Recognitions enables business cases
- → AcListant®-Strips shows that with ABSR 50-70 litres fuel saving plus reduced controller workload are possible

What is then the concept?

AcListant®/AcListant®-Strips spent 1.3 Mio €.

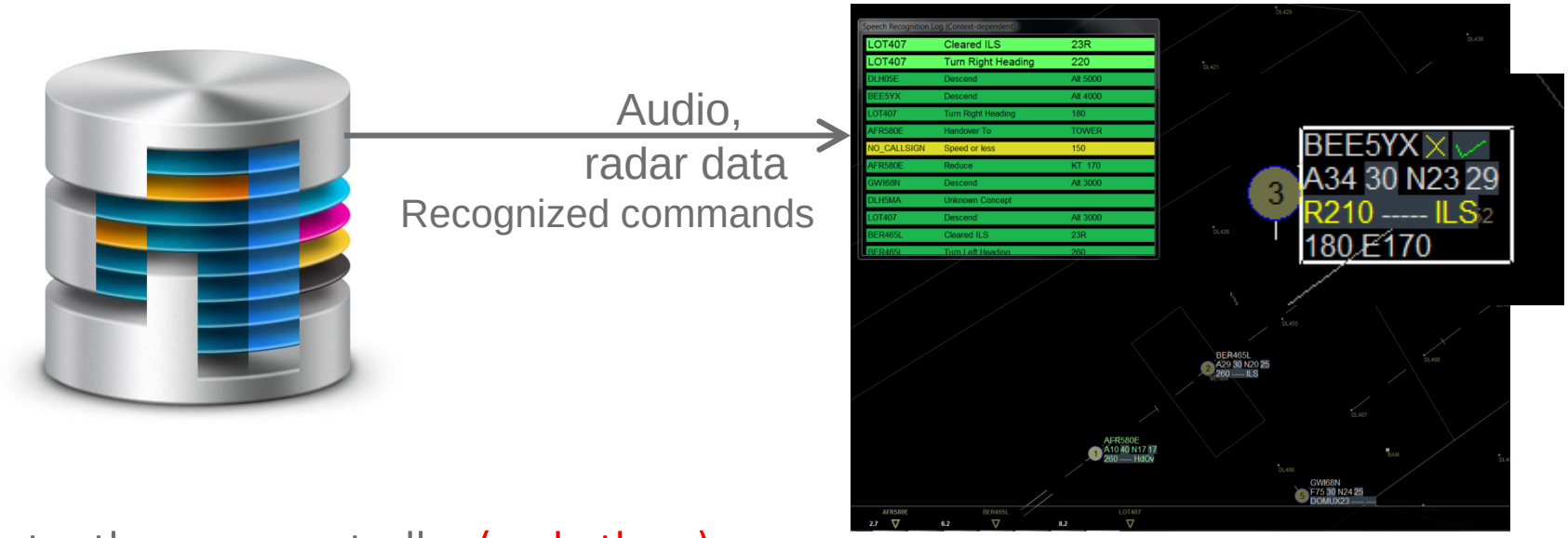
The research question is:

**Can we adapt a basic ABSR system much cheaper with machine learning techniques to new approach areas and performance is still acceptable by controllers.**

# Topic 3 – Result 2/5

## Proof-of-Concept Trials December 2017/January 2018

After the recording phase, the voice utterances are presented to the controller together with the online recognition results of the trained system.




© DLR

Later the same controller (and others) can see
a replay of their work together with the recognition results:
Recognized commands shown in radar label and speech recognition log.
Observer gives feedback concerning errors and recognitions.

# Topic 3 – Result 3/5
## Proof-of-Concept Trials December 2017/January 2018

With this approach we do not show that machine learning goes into the right direction.

- Comparison experiment might be the better option.
    - Running baseline system, running trained system (by unsupervised learning)
    - Randomly select N (e.g. 200) difference of both systems, knowing the correct transcription not necessary
    - Replay the N selected different situations
        - Controllers sees radar situation plus recognition1 plus recognition2
        - Controller judges which recognition is the best one (often none is completely correct)
        - The controller does not know which recognizer is which one

# Topic 3 – Result 4/5
## Proof-of-Concept Trials December 2017/January 2018

**Which recognition is better?**

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

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

AUA392P EXPECT_ILS 34,
AUA392P TRANSITION BALAD_3N,
AUA392P REDUCE 220

NO_CALLSIGN EXPECT_ILS 34,
NO_CALLSIGN TRANSITION BALAD_3N,
NO_CALLSIGN REDUCE 220

We only have a sunturk 392P in the air!
Same decision?

03\BALAD_20160621_A5_2\2016-06-21__09-51-09-41

# Topic 3 – Result 5/5

## Proof-of-Concept Trials December 2017/January 2018

We need more money (if simulation trials are needed).

Participants were “forced” to vote for Replay OR Comparison

3 for Replay, 4 for Comparison

Comparison Experiment could mean: Present twenty 5-minutes scenarios to a controller (10 with basic, 10 with trained system) and give controller questionnaire with respect to e.g.

- Callsign recognition
- Recognition time
- Command value recognition…

We do that in August/September with one controller and decide how to continue, i.e. doing replay and comparison or only one or nothing or …

# Topic 4 – Result

In preparation..

# MALORCA

Machine Learning of Speech Recognition Models for Controller Assistance

DLR

UNIVERSITÄT DES SAARLANDES

idiap RESEARCH INSTITUTE

austro CONTROL

Air Navigation Services of the Czech Republic

Covering the sky...

# Thank you very much for your attention!

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.

SESAR JOINT UNDERTAKING

Founding Members

EUROPEAN UNION EUROCONTROL