Machine Learning of Speech Recognition Models for Controller Assistance

Hon. Prof. Dr. Hartmut Helmke

SJU Scientific Committee Meeting 06

Brussels, 30th May, 2018

The expensive part of
Automatic Speech Recognition (ASR)

Prague

tango papa, papa turn left heading three, ...

Word transcription

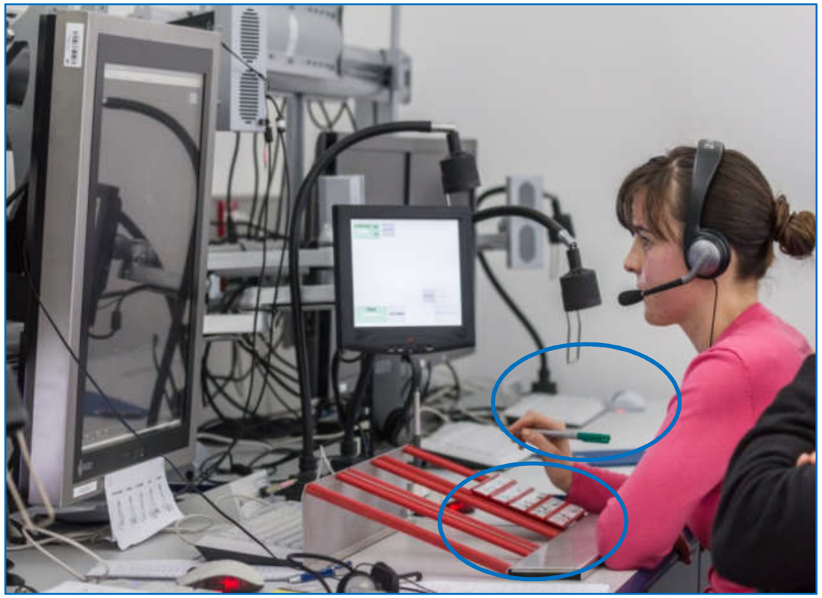
T7APP TURN_LEFT_HEADING 330
T7APP ..

Command transcription

Basic Recognizer is improved by Machine Learning from 8% command recognition error rate to 0.6%.

Contents

From Paper to Electronic Flight Strips with Speech Recognition Support



All flight information in digital form in the system (and on the radar screen).
This may results in higher controller workload.
Controllers have the additional workload. Others have the benefits.
ASR (= Automatic Speech Recognition) may be a solution

AcListant®-Strips: Validation at DLR in 2015

In 2014/2015 >20 controllers from DFS, Austro Control and ANS CR validated Speech Recognizer by USAAR in DLR labs for Dusseldorf Approach Area.

Goal:

Quantify the benefits for Automatic Speech Recognition wrt.

- controllers' workload and
- ATM efficiency.

Baseline:

- Commands entered by mouse into radar labels

Improved Mode

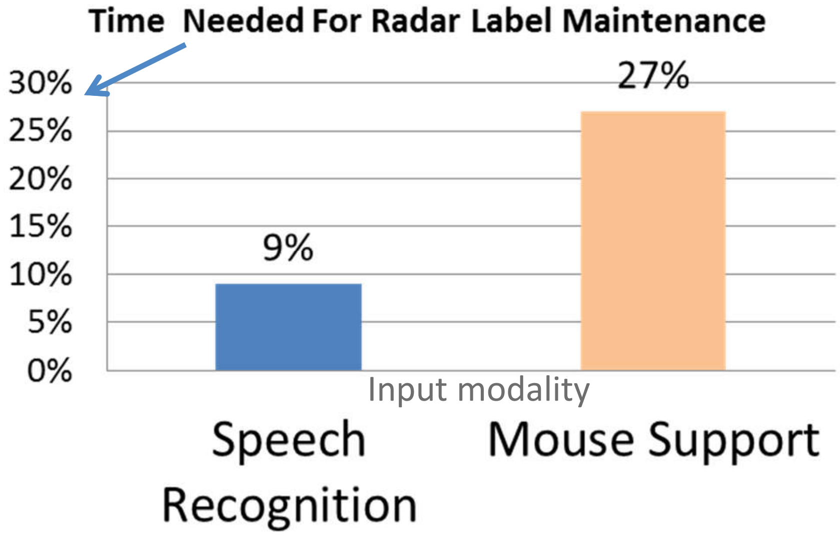
- Commands entered by ASR (**automatic speech recognition**), correction if necessary by mouse

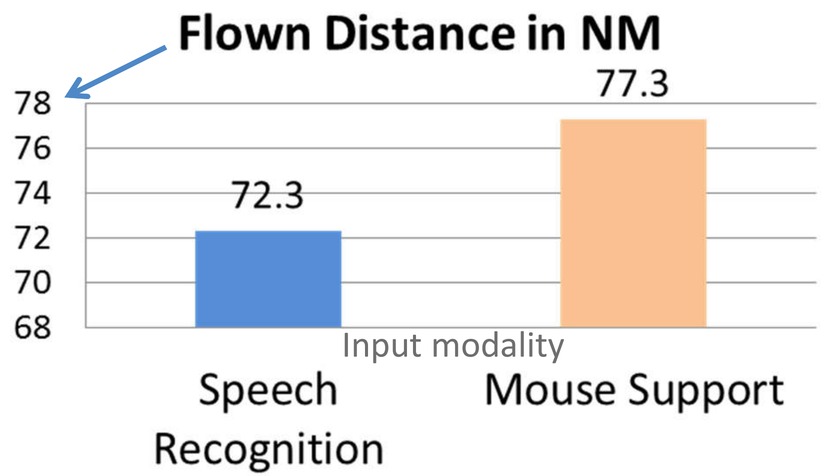
AcListant®-Strips: Validation at DLR in 2015 (2)

Recognition Rate

95.2%

Number of correctly recognized commands divided by number of given commands.
Correct means: callsign, type and value correct.





Speech Recognition increases flight efficiency.

Speech Recognition reduces workload.

Speech recognizer developed by USAAR and integrated with DLR arrival manager.

Validation Results of 2015 *

Airlines

save 50 to 65 liters of kerosene per flight

Airports

benefit from increase flow of 1 to 2 landings per hour

ANSPs and controllers

have reduced workload needed for clicking by a factor of 2 to 3

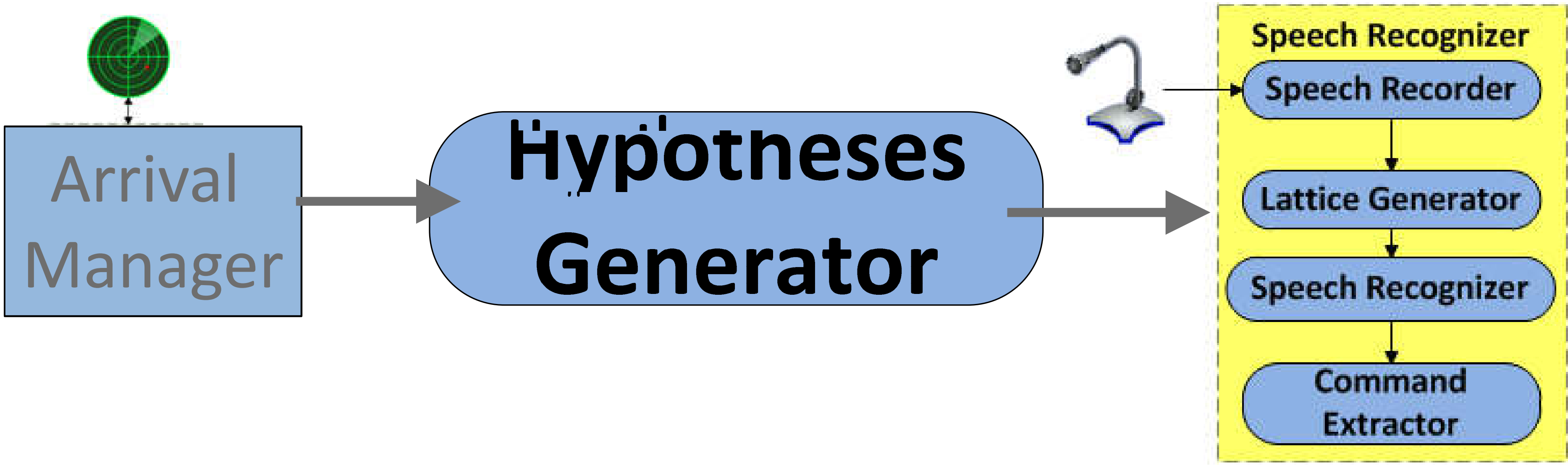
and benefit from reduced head down times which increases safety

society

saves approx. 130 kg of CO_2 per flight

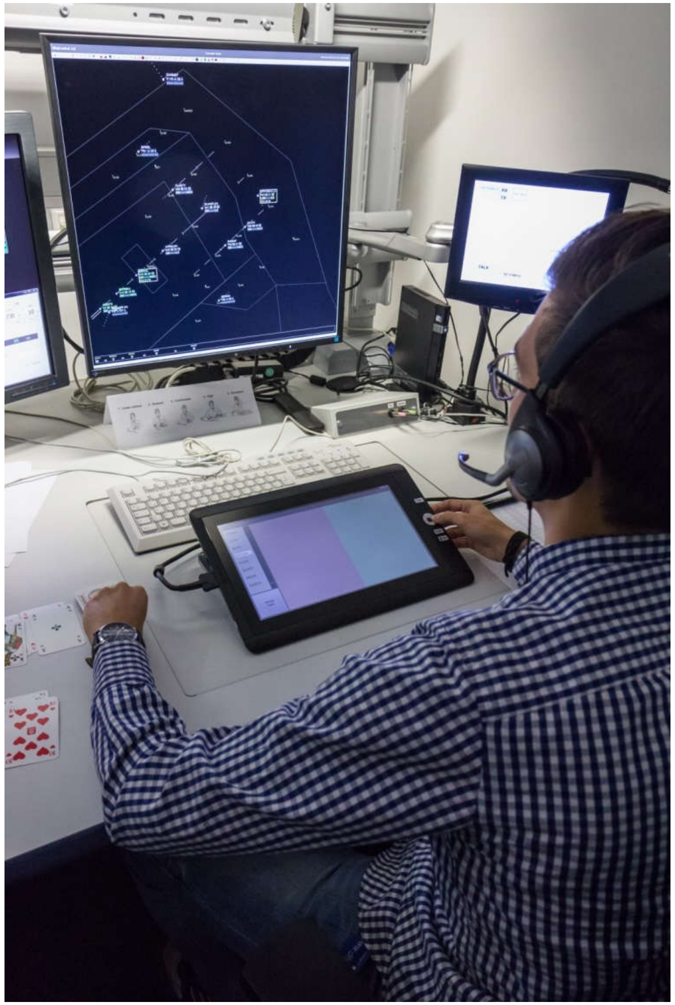
*
A320,
0.8 kg / l, 1 kg kerosene results in 3.15 kg C02; 35 landings per hour
extrapolation of results of 60 minutes scenarios for 23R,
8 controllers…, see papers at DASC 2016 and FAA/Eurocontrol ATM Seminar 2017

Assistant Based Speech Recognition (ABSR)



Contents

- Motivation for Speech Recognition in Air Traffic Control
- MALORCA-Project
- Adaptation to Prague and Vienna and Approach Area by Machine Learning
    - Experiment
    - Results of Technical Experiment
    - Feedback of Controllers (Operation Validation)
- Conclusions



MALORCA project in Admin numbers

- MALORCA = Machine Learning of Speech Recognition Models for Controller Assistance
- 24 months duration (Apr. 2016 to March 2018)
- 37 deliverables (5 public ones)
- 5 partners with funding of 538 k€ plus 267 kCHF
- Total: 83.1 PM

- www.malorca-project.de

Motivation of MALORCA project

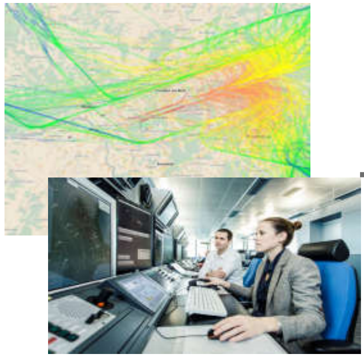
- We have high command recognition rates (> 90%) and low command recognition error rates (< 2%)
- Users (Air traffic controllers) want the system, because Automatic Speech Recognition reduces controllers' workload.
- We even have a business case (less fuel consumption, more landings per hour, …)

However,

DLR and USAAR spent more than 1.4 million Euros for AcListant® and AcListant®-Strips just for adaptation to Dusseldorf Approach Area for landing direction 23R.

Motivation – The MALORCA Project

AcListant



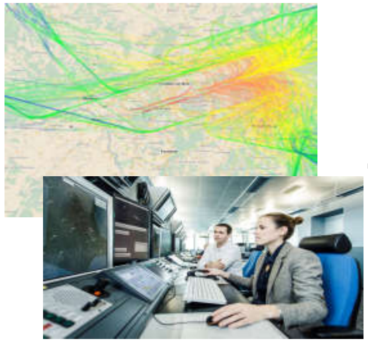
Radar Data

Voice Recordings

Group of experts

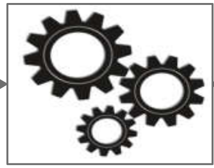
Assistant Based Speech Recognition

MALORCA



Radar Data

Voice Recordings

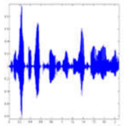


Automatic Learning Process

Assistant Based Speech Recognition

Instead of (highly skilled and paid) experts, machine learning is used.

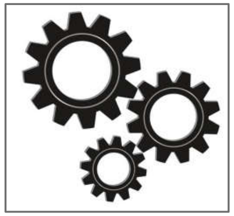
MALORCA objectives



Speech Recognition related objectives

- Provide speech recognition tools for different deployment areas
- Improvement of command recognition error rate by machine learning (ML)

MALORCA objectives (2)



Machine learning related objectives

- Develop a multi-modal, state-of-the-art, automatic learning system
  - To reduce costs of data
  - To speed up development
  - To reduce manual adaptation effort

Two different roadmaps exists (at least):

1. Speech Recognition Roadmap
2. Machine Learning Roadmap

MALORCA concentrates on Machine Learning,
i.e. learning instead of programming/configuration.

MALORCA objectives (3)

Bringing together **experts from multiple disciplines**, i.e. experts

- from data science, machine learning,
- speech processing and recognition and
- air traffic control

Bringing together experts from multiple disciplines

Bringing together experts from multiple disciplines

Bringing together experts from multiple disciplines

Participants from

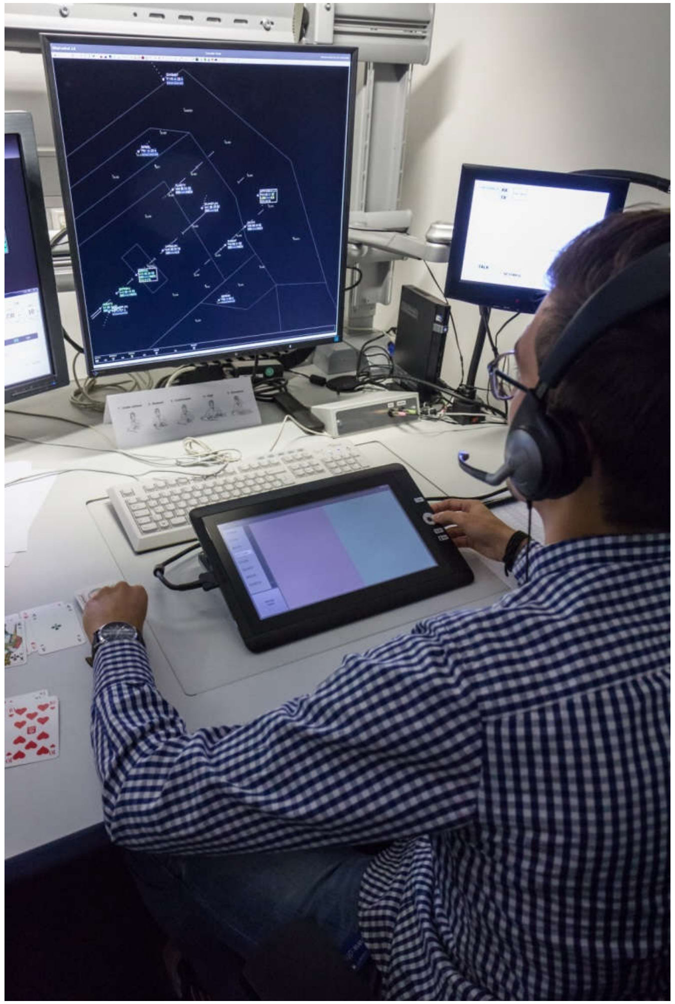
- **12 ANSPs** (DFS, Coopans (IAA, ACG, LFV, CROC, NUAC HB) , **FAA,** NATS, ANS CR, Avinor, LPS)
- System Suppliers (**Thales**, Frequentis, CS-Soft, ZCU, CVU, Harris)
- Speech Recognition (UFA, E-Sigma)
- Academia (Linköping University , Idiap, Saarland University, Univerity of Munich)
- Airport (FRAPORT)
- **Airbus,** Honeywell, Aeroholding
- **SJU**

Contents

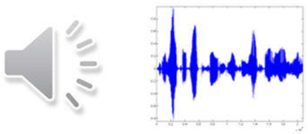
MALORCA
Machine Learning of Speech Recognition Models for Controller Assistance

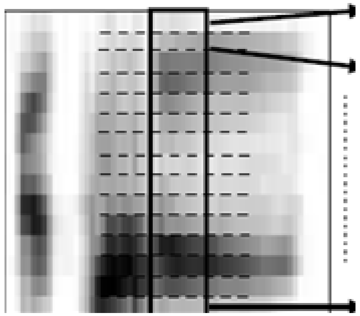
SESAR JOINT UNDERTAKING

- Motivation for Speech Recognition in Air Traffic Control
- **Building Blocks of a Basic Speech Recognizer**
- Adaptation to Prague and Vienna and Approach Area by Machine Learning
    - Experiment
    - Results
- PJ.16-04 and Speech Recognition
- Next Steps
- Conclusions



From Speech Signal to HMI





turkish five kilo juliet maintain two two zero knots or greater descent three thousand feet

AM

LM

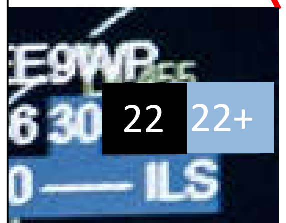
Lex

AM = Acoustic Model (Deep Neural Networks)
LM = Language Model (e.g. grammar)
Lex = Lexicon

Language model assigns probabilities to sequence of words

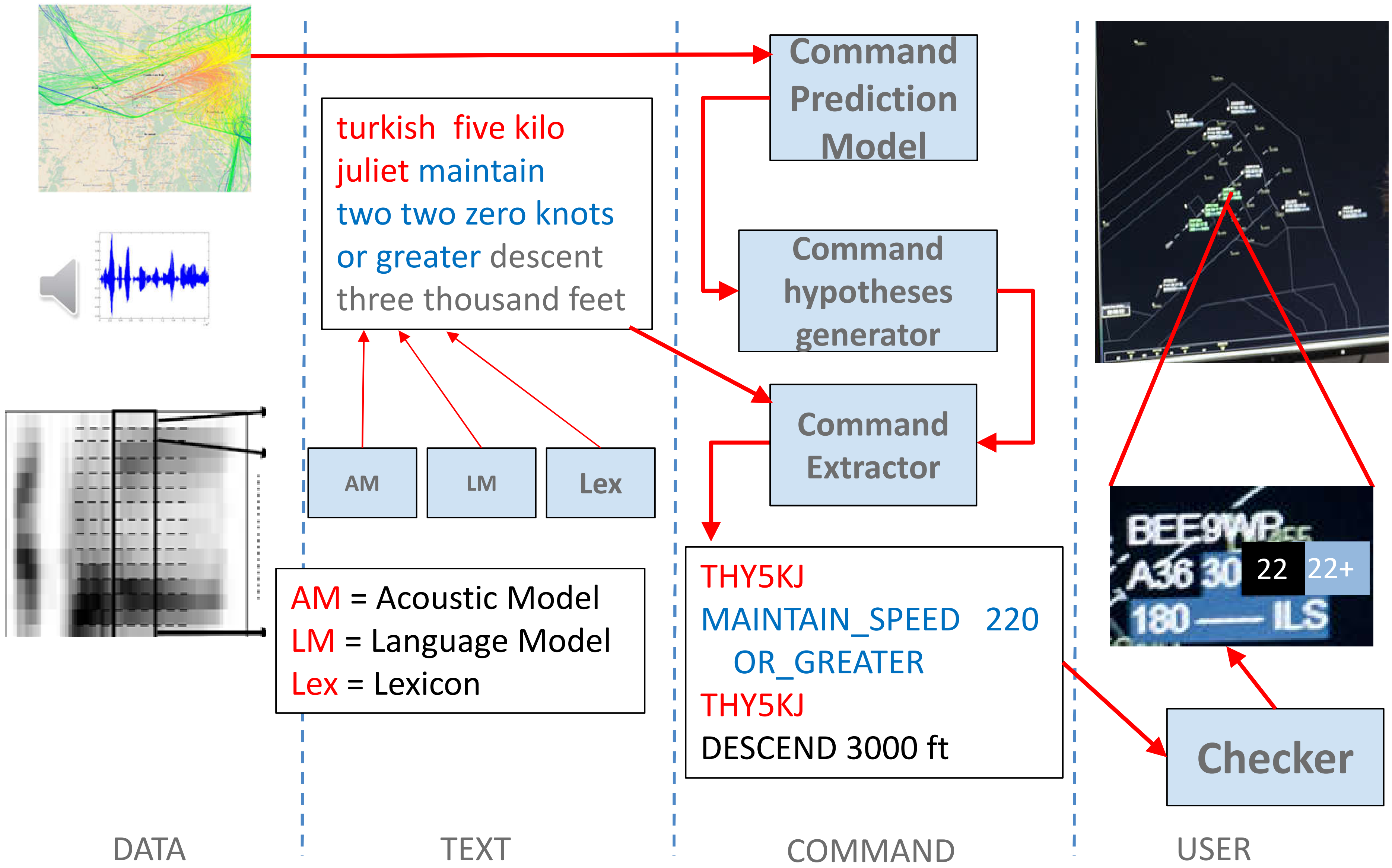
GO FOR A TEA BREAK, GOOD BYE

GO FOR A T-BRAKE, GOOD BUY



DATA | TEXT | COMMAND | USER

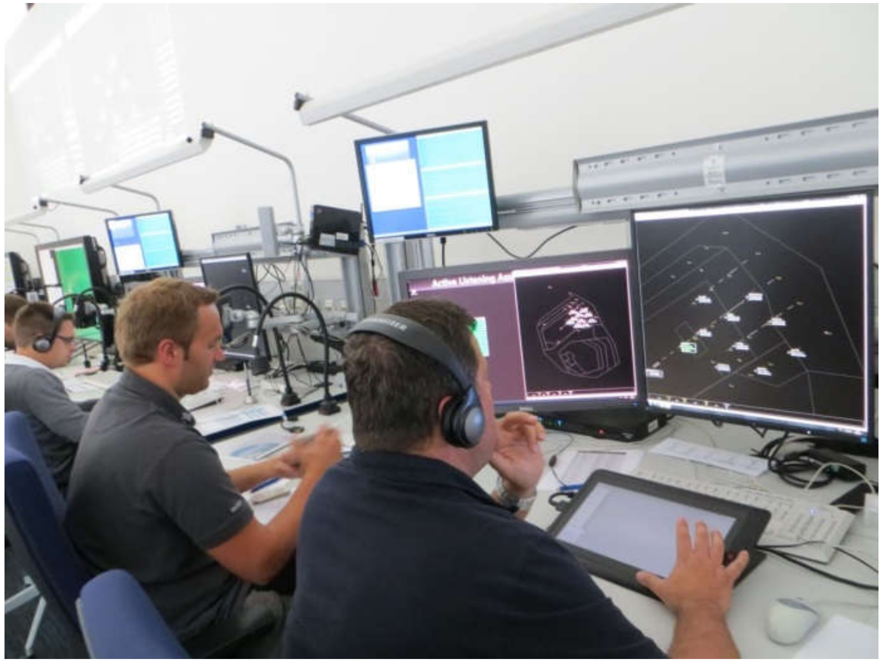
From Speech Signal to HMI



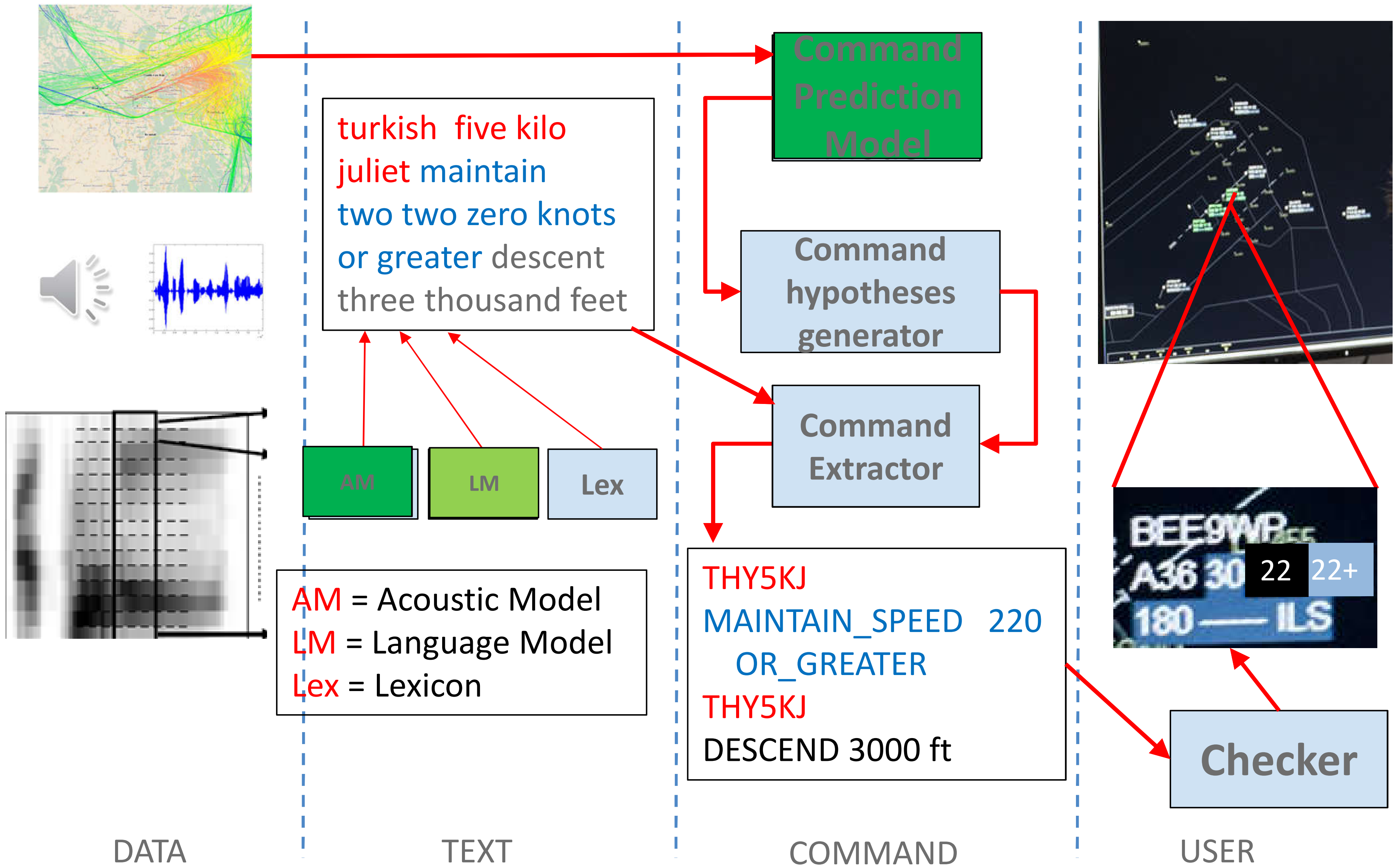
Contents

- Motivation for Speech Recognition in Air Traffic Control
- MALORCA-Project Objectives and Building Blocks of a Basic Speech Recognizer
- Adaptation to Prague and Vienna and Approach Area by Machine Learning
  - Experiment
  - Results
- PJ.16-04 and Speech Recognition
- Next Steps
- Conclusions

MALORCA
Machine Learning of Speech Recognition Models for Controller Assistance

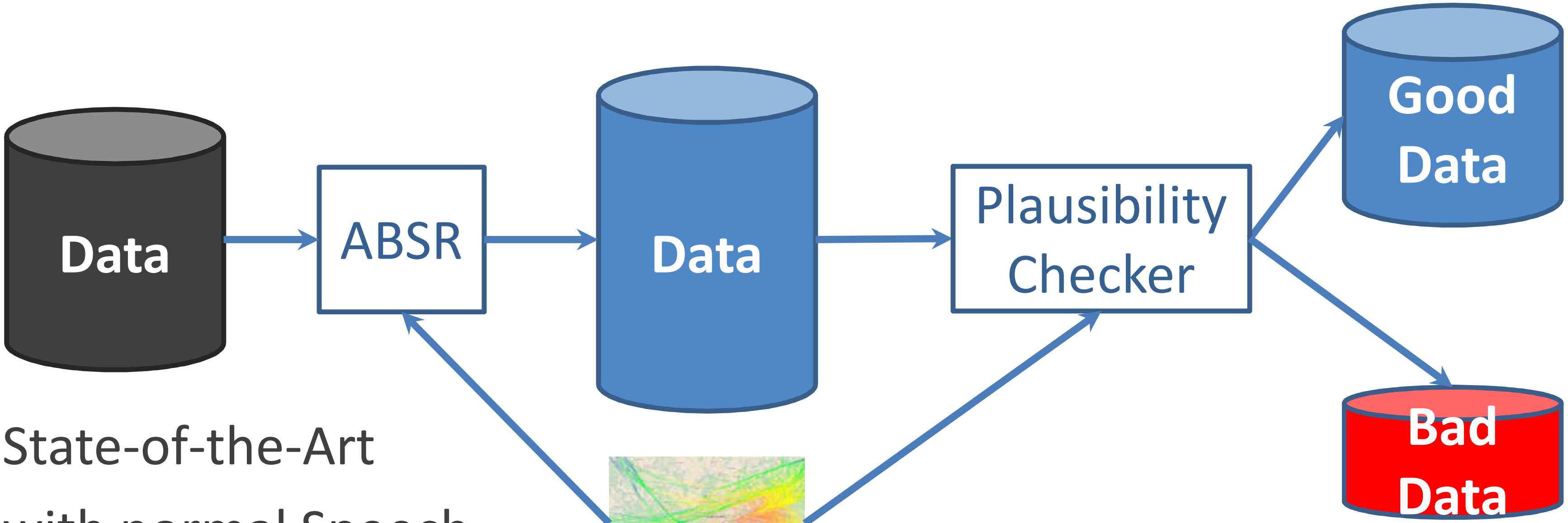


Model Training / Learning



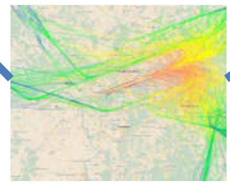
Semi-supervised learning (AM, LM, CPM)

- Exploit untranscribed data
- Generate transcripts (using actual system)
- Data selection: Select “good” or “bad” data ?



State-of-the-Art

with normal Speech Recognizer,

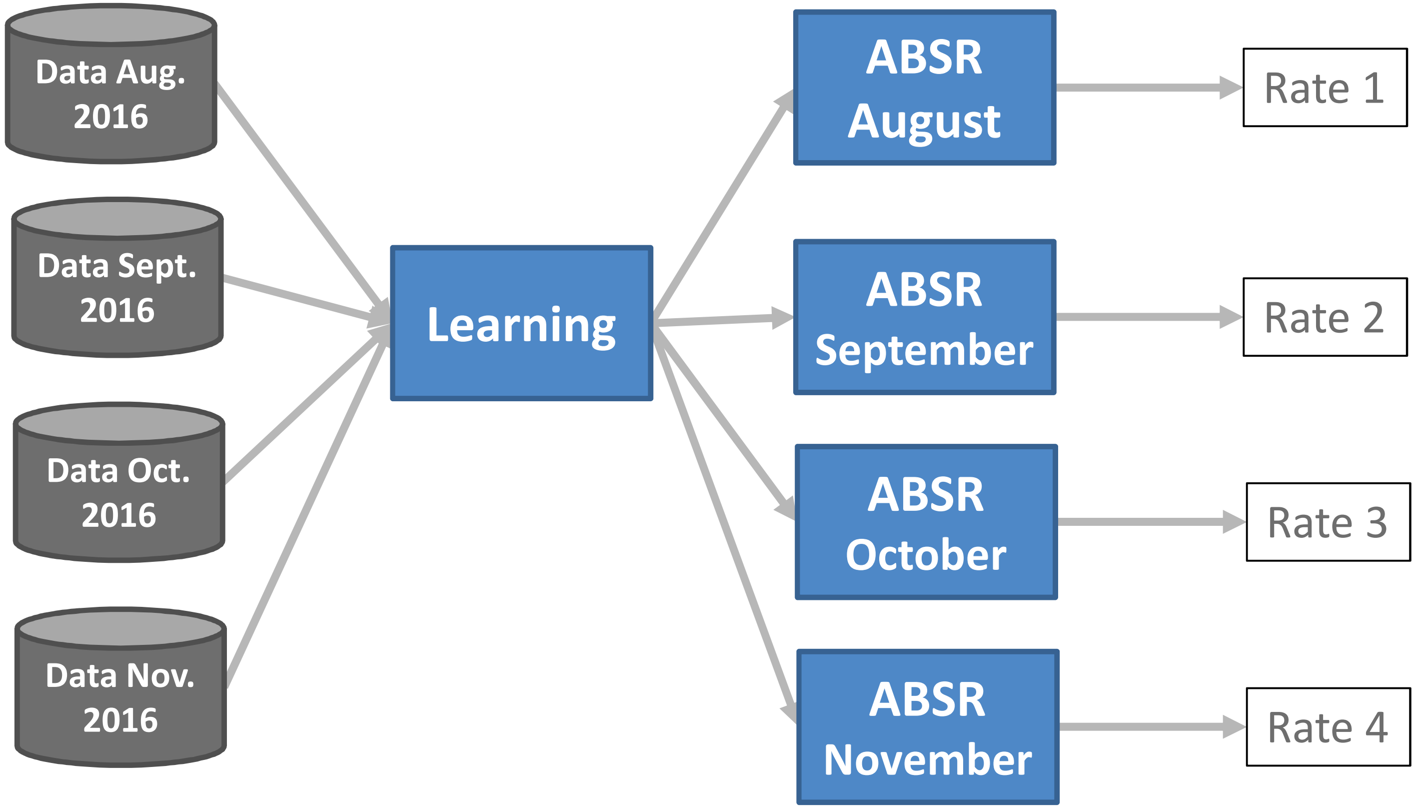


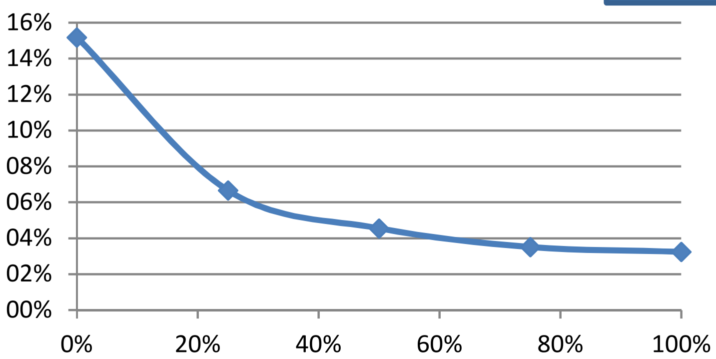
MALORCA: Assistant Based Speech Recognition

Invention of MALORCA

ABSR = Assistant Based Speech Recognition

T2: Proof-of-Concept for Continuous Learning

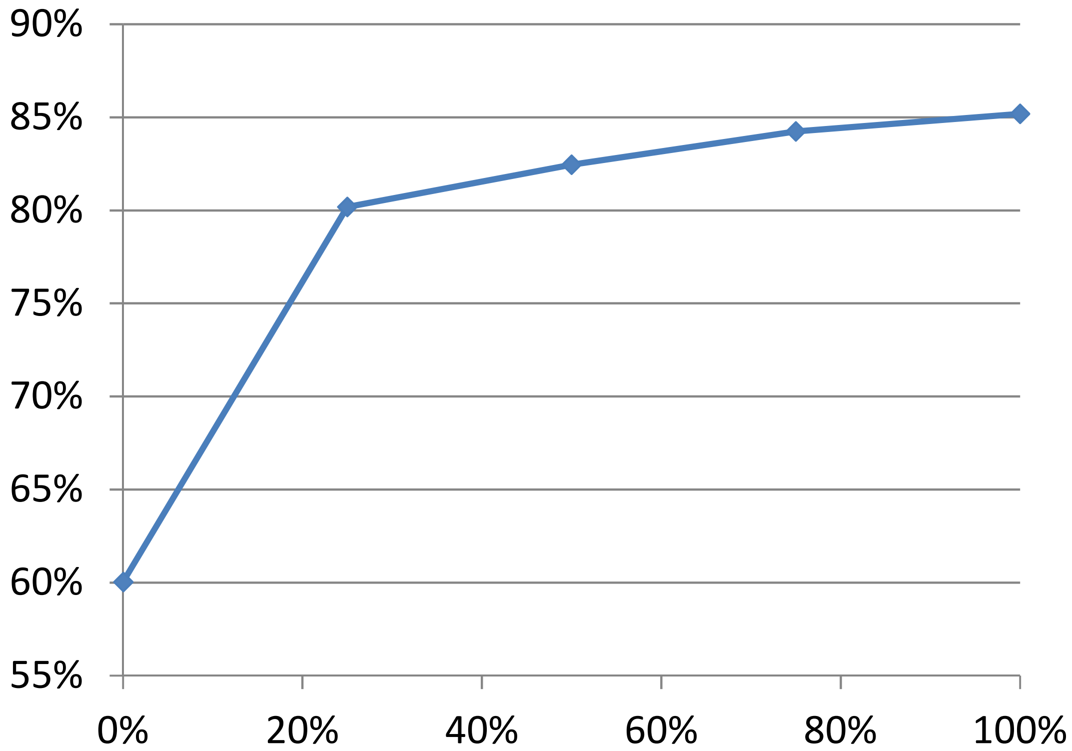




Contents

Learning Curve for Vienna

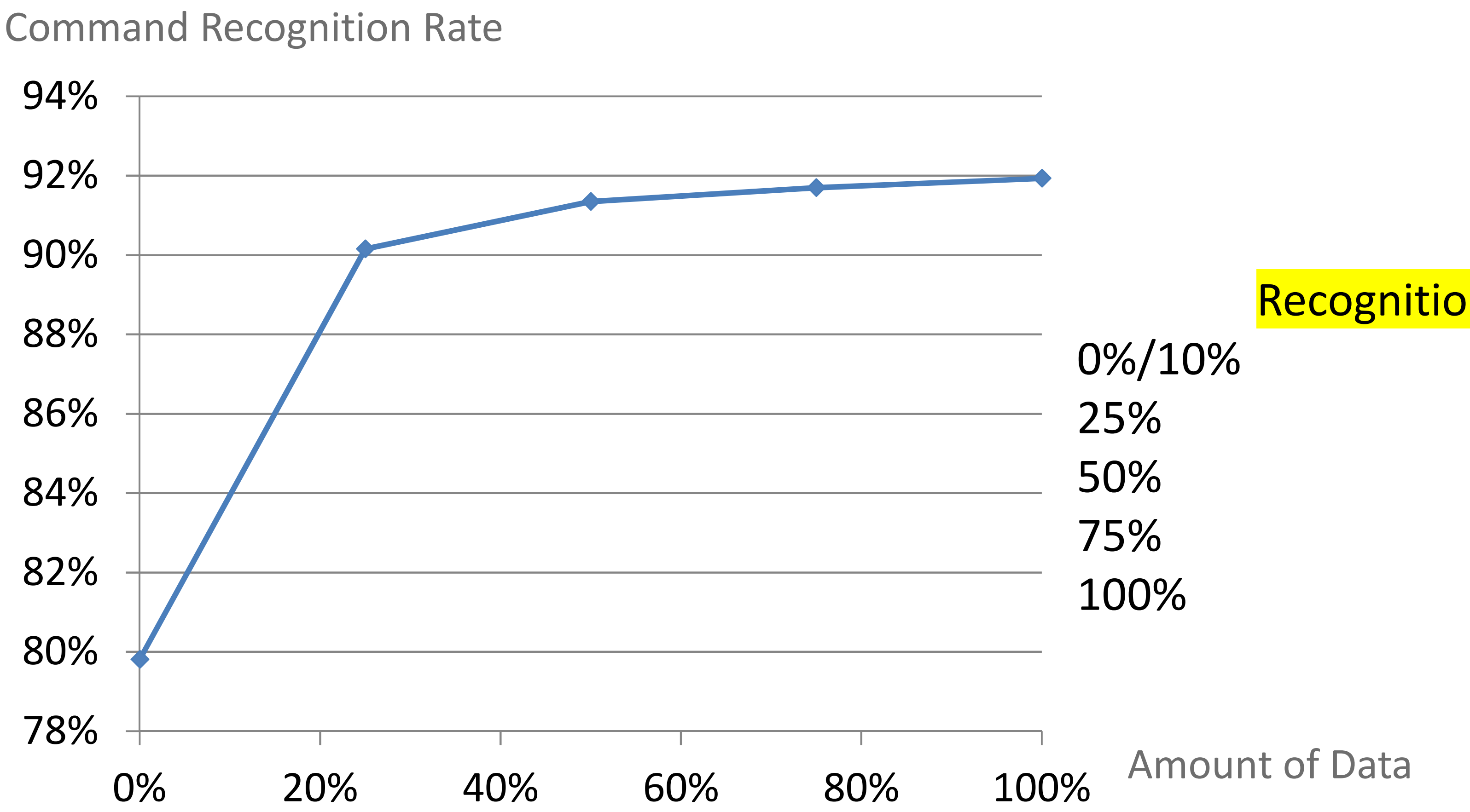
Command Recognition Rate



	Recognition Rate
0%/10%	60.0%
25%	80.2%
50%	82.4%
75%	84.2%
100%	85.2%

Amount of Data

Learning Curve for Prague



	Recognition Rate
0%/10%	79.8%
25%	90.2%
50%	91.3%
75%	91.7%
100%	91.9%

Different Audio Qualities (Signal-to-Noise-Ratio)

SESAR JOINT UNDERTAKING

Prague

good day easy three one tango hotel ruzyne radar, radar contact roger continue

Vienna

austrian nine four zero nine, there is vfr traffic one o'clock distance five miles opposite aeh round three thousand five hundred feet

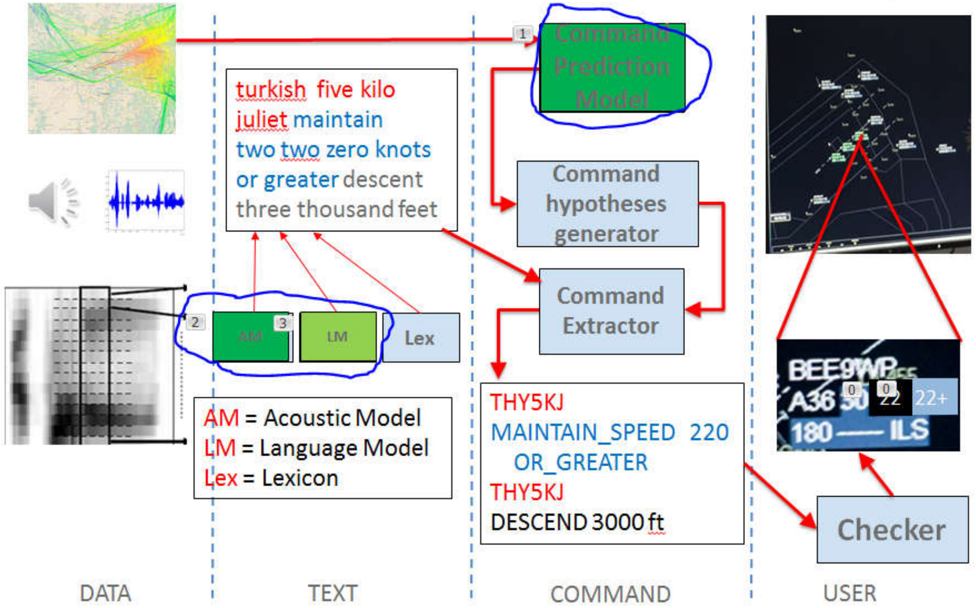
Prague

tango papa, papa turn left heading three, three zero cleared for ils three zero report established and reduce speed one six zero knots

Vienna

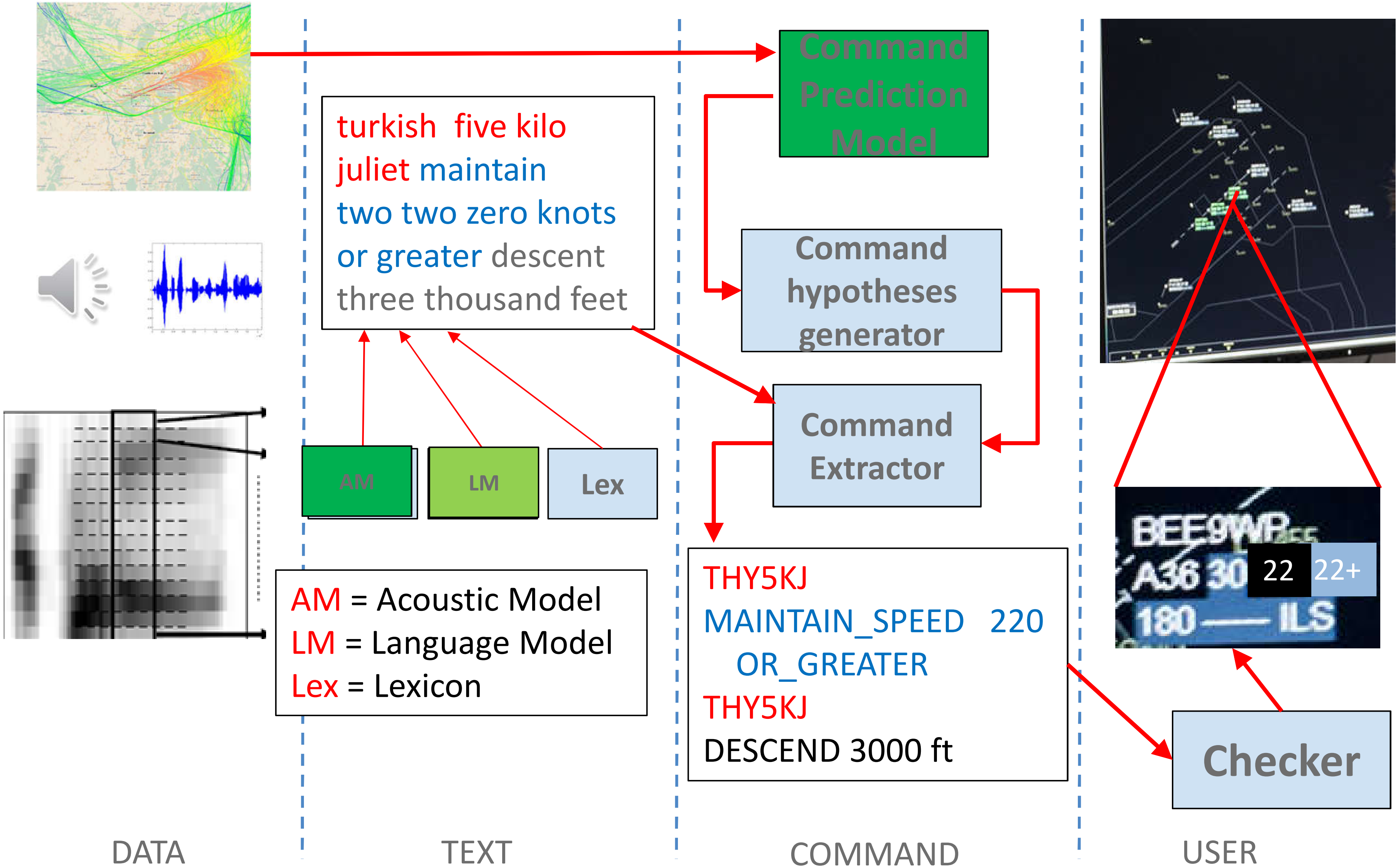
turkish five kilo juliet maintain two two zero knots or greater descent three thousand feet

Results

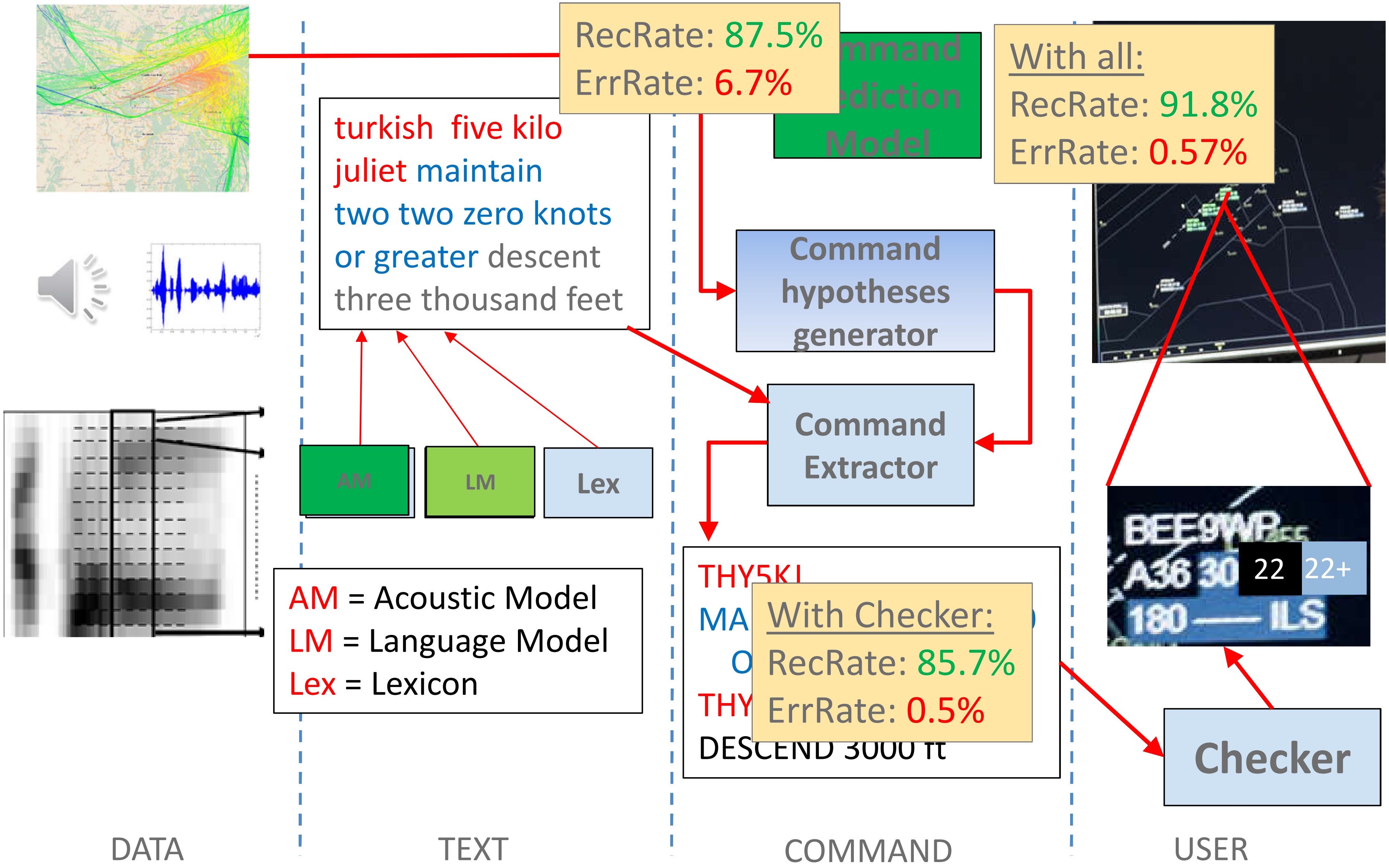


- Machine Learning of Acoustic Model, Language Model, Command Prediction Model is possible
- Command Recognition Rate improves from 80% to 92% (Prague) resp. from 60% to 85% (Vienna)
- By 8 times more data may provide 92.6% (Prague) resp. 90.2% (Vienna)
- Open Question: **Explore combining the data from different airports**

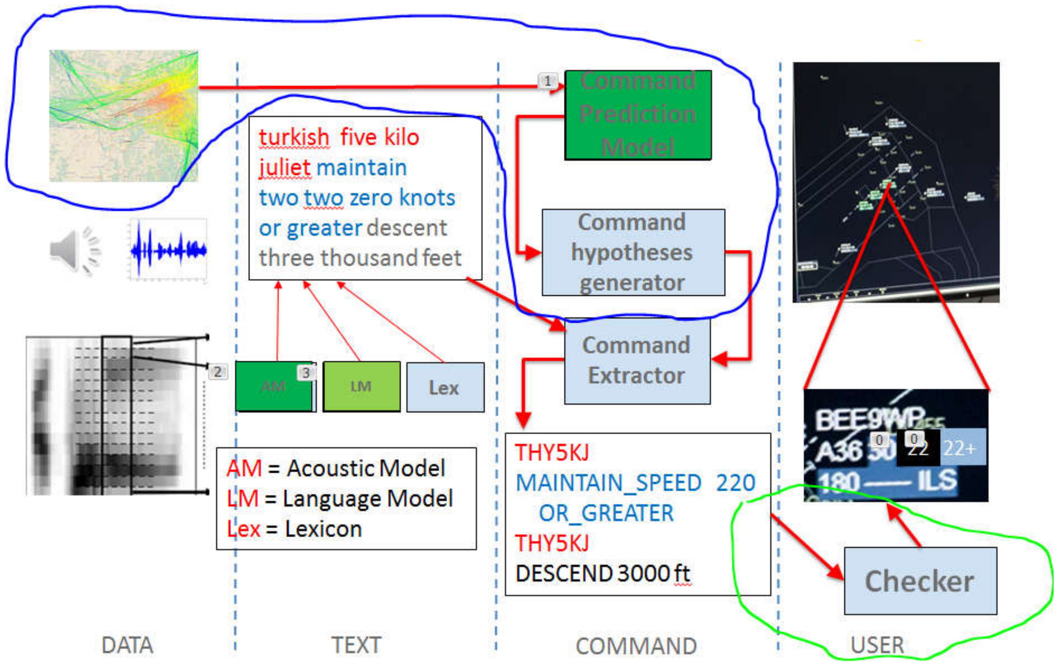
Effect of Different Components



Effect of Different Components



Results (2)

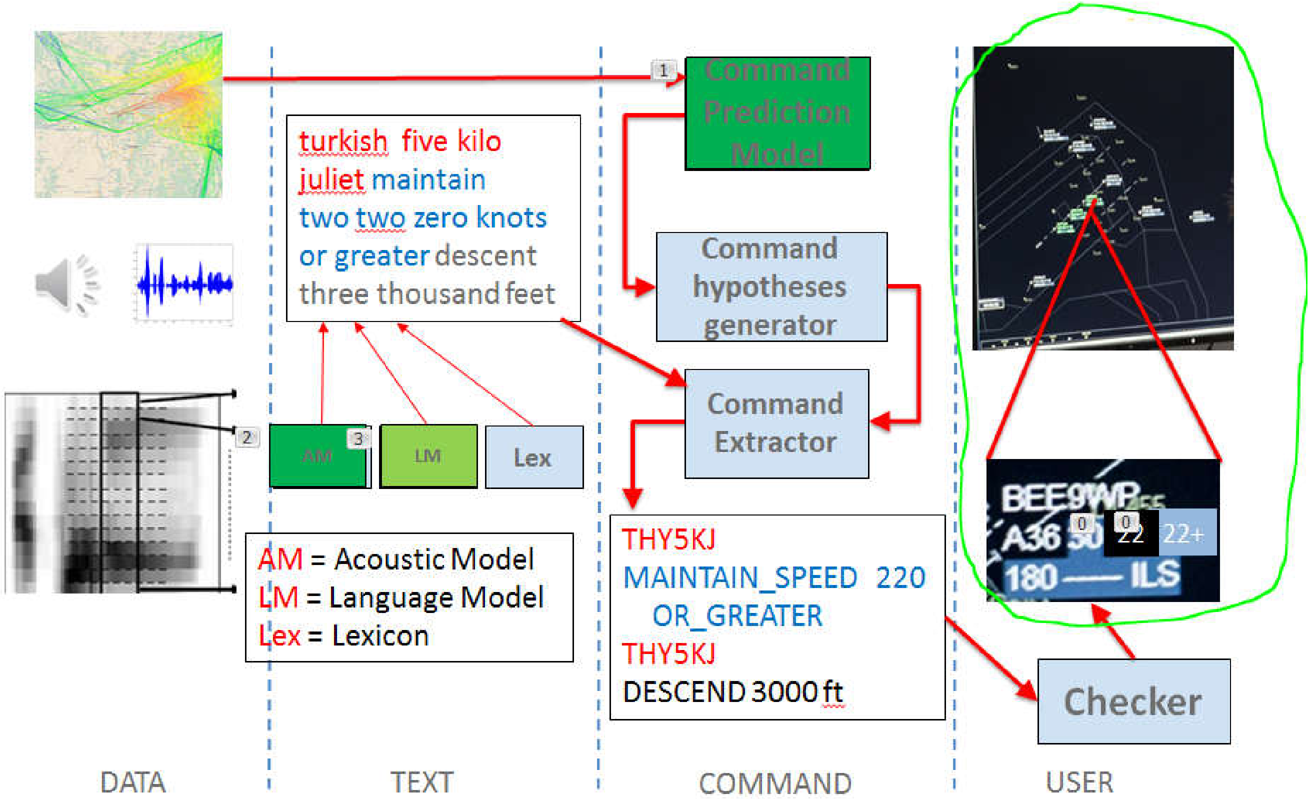


Command hypotheses generator plus Checker improves

- Command Recognition Rate from 85.7% to 91.8% (Prague)
- Command Recogn. Error Rate from 6.7% to 0.6% (Prague)

- Command Recognition Rate from 71.3% to 85.2% (Vienna)
- Command Recogn. Error Rate from 15.7% to 3.2% (Vienna)

User Acceptance (1)



The numbers are clear, BUT

we need to have the end users (controllers) on board from the beginning!!!

User Acceptance (3)

Prague:				
Number of Commands	Number of ABSR Errors/Rejections	Rec Rate	corrected by controller	detected by controller
396	36	90.9%	31	36

Vienna:				
Number of Commands	Number of ABSR Errors/Rejections	Rec Rate	corrected by controller	detected by controller
610	80	86.9%	79	80

No safety issues were observed.

All misrecognitions were detected.

Better Recognition Rate of course would improve even more workload reduction.

ABSR = Assistant Based Speech Recognition (= Speech Recognition with Command Hypotheses Generator)

Contents

SESAR
JOINT UNDERTAKING

- Motivation for Speech Recognition in Air Traffic Control
- MALORCA-Project Objectives and Building Blocks of a Basic Speech Recognizer
- Adaptation to Prague and Vienna and Approach Area by Machine Learning
  - Experiment
  - Results
- **PJ.16-04 and Speech Recognition**
- Next Steps
- Conclusions

MALORCA
Machine Learning of Speech Recognition Models for Controller Assistance

Who contributes to PJ.16-04?

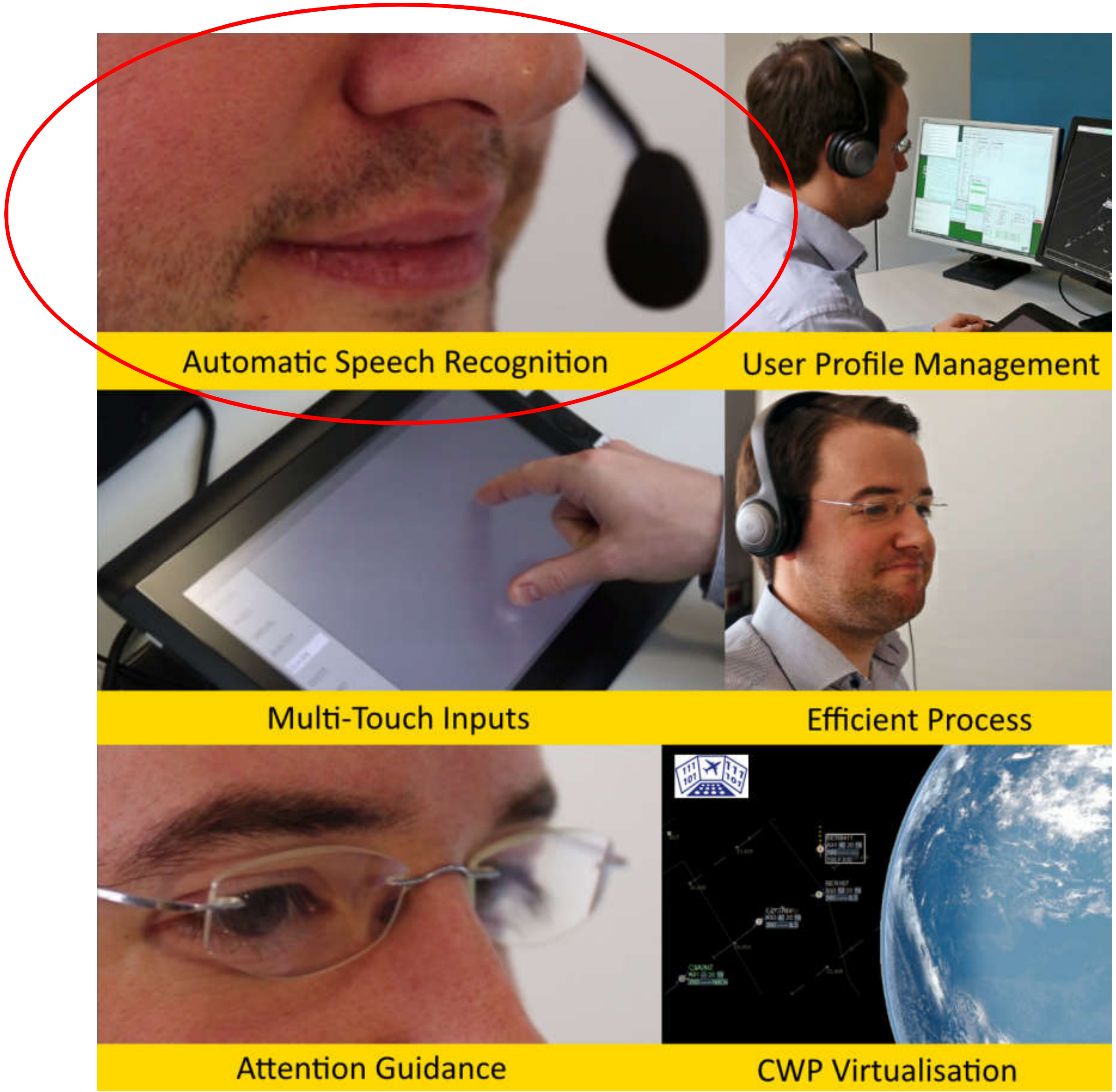
23 partners from 16 European countries

PJ.16-04 Team
THALES AIR SYS
ANS CR (B4)
Integra
LPS SR (B4)
ACG/COOPANS
CCL/COOPANS
LFV/COOPANS
Naviair/COOPANS
DFS
ENAIRE
CRIDA
NATS
Avinor ANS
SKYGUIDE
SKYSOFTATM
EUROCONTROL
DLR (AT-One)
FRQ (FSP)
HC (FSP)
SINTEF (NATMIG)
INDRA
ROMATSA
LEONARDO

Logos of PJ.16 Members

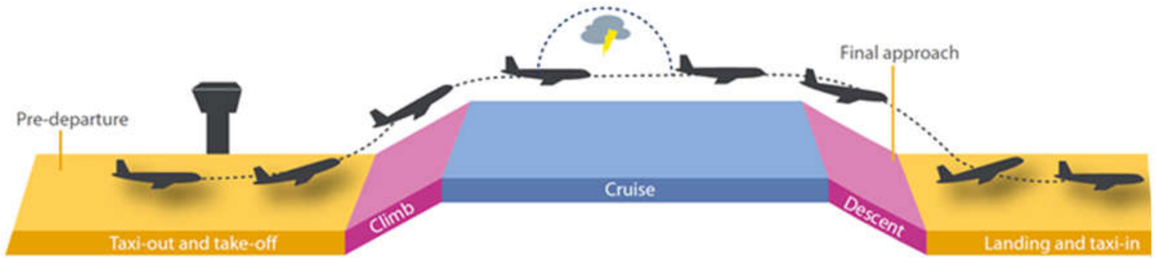
ATM System provider
ANSPs
Research and Consulting Institutes

Activities in PJ.16-04



ASR Exercises of PJ.16-04

No Learning planned
Everything manually adapted



Pre-departure | Taxi-out and take-off | Climb | Cruise | Descent | Landing and taxi-in

Tower	APP	En-Route	APP	Tower

From UFA (US)	Munich Approach (DFS / PJ.10-02b / DFS Simulator)
From Nuance (US)	Prague Approach – (Vienna Approach) (THALES / DLR / ANS-CR / INTEGRA / Coopans / PJ.16-04 internal / SHAPE Sim
From EML/IBM (US)	Madrid FIR (Route/TMA) (CRIDA / ENAIRE / PJ.10-01b / CRIDA Lab)
No ASR	Hungary Remote Tower Environment (DLR / HUNGAROCONTROL / PJ.05-02 / DLR TowerLab)
COTS	Germany Remote Tower Environment (FREQUENTIS / DFS / PJ.05-03 / DFS Premises)

Contents

- Motivation for Speech Recognition in Air Traffic Control
- MALORCA-Project Objectives and Building Blocks of a Basic Speech Recognizer
- Adaptation to Prague and Vienna and Approach Area by Machine Learning
    - Experiment
    - Results
- PJ.16-04 and Speech Recognition
- **Next Steps**
- Conclusions

SESAR JOINT UNDERTAKING

MALORCA

Machine Learning of Speech Recognition Models for Controller Assistance

Challenges for Speech Recognition (ASR) From MALORCA to IBICA

- Confidence scores for ASR output

How sure ASR is? Plausibility values for recognized commands and for predicted commands ?

- Hesitation/Uncertainty/Aeh detection

Controller certainty, professionality, performance analysis, good phraseology

- Pilot Recognition

Different accents, cockpit noise, more phraseology deviations

- Learning of Statistical-Language models and phraseology (deviations)

Not changing the controller

- Standardization
  Exchanging of transcriptions

Ontology of PJ.16-04 for ASR output? Data exchange between different partners? Agreed interfaces?

- Real-time Speech Recognition

Online versus offline decoding? First output to controller before releasing push-to-talk-button

- Needed Performance :

80% enough? 1% of Error Rate? Which rates are really needed or which prize?

- Safety Issues

Speech Recognizer is shut off? Over trusting in ASR? Bad ASR performance can increase workload and cause safety issue?

Challenges for Machine Learning From MALORCA to IBICA

- Introspective Learning Algorithm, trust, self-diagnosis
  - Problem out of problem area, result out of envelope, early integration of regulators
- Daily Learning and Update
  - When installing new version, Using 1000s of hours
- Learning directly in ops rooms also for data privacy
  - Voice and radar data stays in ops room, only learned models (abstraction) leave ops room
- Other Input Modalities as second sensor
  - ADS-B, Mode-S, Electronic Flight Strips,
- Automatic Radar Data processing resp. other second sensors
  - Parallel processing of offline data, more than real time needed for pre-processing
- Artificial Data
  - There is no data, than more data, artificial is even cheaper, image recognition disturbs images or avoids labels by synthetic data
- End-to-end-Learning
  - Direct learning Commands from features, without indirection of phoneme and word learning

What else: End-to-end modeling Application of speech+other data from daily recordings

End-to-end modeling

DLH8EK REDUCE 220

Benefiting Applications from MALORCA's Speech Recognition and Learning Competence

Individual Controller Education

- Immediate feedback after shift/simulation run

Holistic Workload Assessment and Prediction

- Monitoring and Planning tool for supervisors

Permanent **Online Learning** (in more Complex Areas / Environments)

- Training on daily basis directly in the ops room

Machine Learning for Controller Assistant Tool Configuration and Maintenance

- Reducing xMAN deployment and maintenance costs, they are used and not only bought

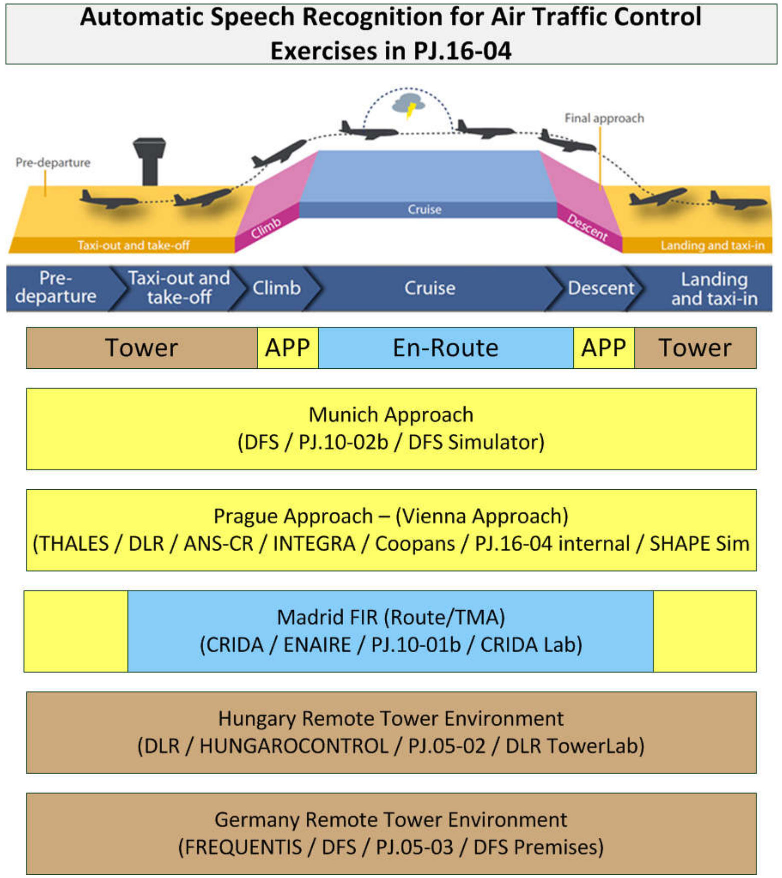
Contents

- Motivation for Speech Recognition in Air Traffic Control
- MALORCA-Project Objectives and Building Blocks of a Basic Speech Recognizer
- Adaptation to Prague and Vienna and Approach Area by Machine Learning
    - Experiment
    - Results
- PJ.16-04 and Speech Recognition
- Next Steps
- Conclusions

European competence

SESAR JOINT UNDERTAKING

- 16.04 DFS task: ASR system is from UFA (US)
- 16.04 CRIDA/ENAIRE task: ASR system is from EML/IBM (US)
- 16.04 Frequentis task: ASR system not chosen yet (COTS)
- 16.04 task in Rungis: ASR system is from Nuance (US)



- MALORCA's ASR is based on KALDI which is public domain, plus extensions done in MALORCA and before
- Competence stays in Europe

Conclusions

- **Machine Learning** of Acoustic Model, Language Model, Command Prediction Model is possible
- Command Recognition Rate improves from 80% to 92% (Prague) resp. from 60% to 85% (Vienna)

Command Hypotheses Generator plus Checker improves (Context integration)

- Command Recognition Rate from 85.7% to 91.8% (Prague)
- Command Recogn. Error Rate from 6.7% to 0.6% (Prague)

No safety issues were observed.

Machine Learning can ease Adaptation and Maintenance of ATC tools (e.g. adaptation of an AMAN).

MALORCA's competence

- Machine Learning (Big Data)
- Automatic Speech Recognition
- Air Traffic Management
- Integration of new technologies into the ATM world
- Project Management

- Besides "algorithmic" competence MALORCA has data competence on board, i.e. MALORCA team has access to real data.
- Different to PJ-16-04 MALORCA works on data from the ops room

Promising further applications for machine learning and Automatic Speech Recognition exists.

MALORCA team is ready for IBICA.

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