Building Blocks of Assistant Based Speech Recognition for Air Traffic Management Applications

Matthias Kleinert, Hartmut Helmke, Heiko Ehr, Gerald Siol (DLR)
Christian Kern (Austro Control)
Dietrich Klakow, Mittul Singh (UdS)
Petr Motlicek (Idiap)

German Aerospace Center (DLR), Braunschweig
Saarland University (UdS), Saarbrücken
Idiap Research Institute, Martigny
Austro Control, Vienna
Content

• Motivation for Speech Recognition in Air Traffic Control

• Building Blocks of Assistant Based Speech Recognition (ABSR)

• How to adapt ABSR to new environments

• Influence of different components of ABSR on the performance

• Conclusions
All flight information available in digital form in the system (and on the radar screen). This may result in higher controller workload.

Controllers have additional workload. Others have the benefits.
How can Speech Recognition help?

- Automatic Speech Recognition (ASR) can be the solution
  - Analyze controller-pilot communication
  - Extract relevant commands
  - Use commands as label input
  - Manual correction in case ASR fails

- For this DLR and Saarland University developed Assistant Based Speech Recognition (ABSR) to achieve:
  - High recognition and low error rates on command level
  - Reduction of controller workload by automating flight strip management
ABSR-Validation at DLR in 2015

• In 2014/2015 > 20 controllers from DFS, Austro Control and ANS CR validated the ABSR system developed by UdS and DLR for Dusseldorf Approach Area.

• **Goals**  → Quantify the benefits for ABSR with respect to
  • controllers’ workload and
  • ATM efficiency.

• **Baseline:**
  • Commands entered by mouse into radar labels

• **Improved Mode:**
  • Commands entered by ABSR correction if necessary by mouse
ABSR-Results of 2015

• **Performance**
  → Recognition rate: 95%
  → Error rate: 2%

• **ANSPs and controllers**
  → Radar Label maintenance time reduced by a factor of 3

• **Airports**
  → benefit from increase flow of 1 to 2 landings per hour **Performance**

• **Airlines**
  → save 50 to 65 liters of kerosene per flight

• **Society**
  → saves approx. 130 kg of CO₂ per flight*

* A320, 0.8 kg / l, 1 kg kerosene results in 3.15 kg CO₂; 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
What is the Problem?

- In AcListant® 1.3M Euro were spent for ABSR development and validation just for Düsseldorf Approach
- Adaptation is necessary for other airports and working positions because of:
  - Different accents, working procedures, airspace layouts etc.
- Requires lots of time and expert knowledge
- Adaptation will become necessary again even in an already deployed system due to changes in working procedures, airspace layouts etc. (maintenance)
How to solve it?

- Instead of (highly skilled and paid) experts, **machine learning** is used.
- Horizon 2020 funded project MALORCA (Machine Learning of Speech Recognition Models for Controller Assistance)
Content

• Motivation for Speech Recognition in Air Traffic Control

• **Building Blocks of Assistant Based Speech Recognition (ABSR)**

  • How to adapt ABSR to new environments
  
  • Influence of different components of ABSR on the performance
  
  • Conclusions
From Speech Signal to HMI

DATA

TEXT

COMMAND

USER

Feature Extractor
From Speech Signal to HMI

Feature Extractor

ASR Decoder

N-Best Generator

Domain Knowledge

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

DATA

TEXT

COMMAND

USER
From Speech Signal to HMI

Feature Extractor → ASR Decoder → N-Best Generator

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

Domain Knowledge

turkish five kilo juliett maintain two two zero knots or greater
descent three thousand feet
From Speech Signal to HMI

Assistant System

Feature Extractor

ASR Decoder

N-Best Generator

Corrector

Command Hypotheses Generator

Command Prediction Model

Command Filtering

Command Extractor

Domain Knowledge

ATC Grammar

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

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

data text command user

DATA

TEXT

COMMAND

USER
From Speech Signal to HMI

Turkish: Five kilo Juliett maintain two two zero knots or greater descent three thousand feet

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

THY5QJ MAINTAIN_SPEED 220 OR_GREATER
THY5QJ DESCEND 3000 ft
From Speech Signal to HMI

DATA

Feature Extractor

Assistant System

TEXT

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

Command Hypotheses Generator

Command Prediction Model

Command Extractor

Command Filter

PLausibility Checker

COMMAND

THY5QJ MAINTAIN_SPEED
220 OR_GREATER
THY5QJ DESCEND 3000 ft

USER

ATC Grammar

Domain Knowledge

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

turkish five kilo juliett
maintain two two zero knots or greater
descent three thousand feet
From Speech Signal to HMI

**Assistant System**

- **Feature Extractor**
- **ASR Decoder**
- **N-Best Generator**
- **Command Hypotheses Generator**
- **Command Extractor**
- **Command Prediction Model**
- **Corrector**
- **Command Filtering**
- **Plausibility Checker**

**Domain Knowledge**

- **ATC Grammar**

**DATA**

**TEXT**

- **AM** = Acoustic Model
- **LM** = Language Model
- **Lex** = Lexicon

**COMMAND**

- **THY5QJ MAINTAIN_SPEED 220 OR_GREATER**
- **THY5QJ DESCEND 3000 ft**

**USER**

**turkish five kilo juliett maintain two two zero knots or greater descent three thousand feet**
Content

- Motivation for Speech Recognition in Air Traffic Control
- Building Blocks of Assistant Based Speech Recognition (ABSR)
- How to adapt ABSR to new environments
- Influence of different components of ABSR on the performance
- Conclusions
Adaptations to DATA-Module

**Domain Knowledge**
- Data that is unique for a given environment (e.g. runways, waypoint names, used frequencies etc.)
- Can be derived from existing data sets

**ATC Grammar**
- Based on ICAO-Phraseology
- No complete recreation, but parts of it have to be adapted to a specific target area.

LM = Language Model  
Lex = Lexicon  
THY5KJ DESCEND 3000 ft
Adaptations to TEXT-Module

- **Acoustic model**
  - Trained from transcribed and/or untranscribed data
  - Speaker-dependent models possible

- **Language model**
  - Learned from labeled ATC text transcriptions
  - Grammar can be used for automatic labeling of training data
  - Data from existing language models can be reused

- **Lexicon**
  - Local Words + pronunciations for a specific environment (e.g. Greetings)

**DATA**

**Assistant System**

**Feature Extractor**

**ASR Decoder**

**N-Best Generator**

**AM** = Acoustic Model

**LM** = Language Model

**Lex** = Lexicon

- **turkish five kilo juliett maintain two two zero knots or greater descent three thousand feet**

**Domain Knowledge**

**ATC Grammar**

**TEXT**

**COMMAND**

**USER**
Adaptations to COMMAND-Module

- **Command prediction model (CPM)**
  - Unique for a given environment
  - Is recreated from scratch
  - Can be learned automatically from controller audio recordings resp. automatic command recognitions

**DATA**

Assistant System

**TEXT**

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

**COMMAND**

THY5QJ MAINTAIN_SPEED 220 OR_GREATER
THY5QJ DESCEND 3000 ft

**USER**

Plausibility Checker

ATC Grammar

Correction

Command Hypotheses Generator

Command Prediction Model

Command Extractor

Command Filtering
Adaptations to USER-Module

- **USER-Module**
  - No core parts of the ABSR system have to be adapted
  - HMI adaptations might be necessary to show output of ABSR
  - A standardization concerning an interface to HMIs and the format of the transmitted commands reduces cost

DATA

Feature Extractor

Domain Knowledge

ATC Grammar

TEXT

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

COMMAND

THY5QJ MAINTAIN_SPEED 220 OR_GREATER
THY5QJ DESCEND 3000 ft

Plausibility Checker

TYK5QJ 30 22+
Content

• Motivation for Speech Recognition in Air Traffic Control

• Building Blocks of Assistant Based Speech Recognition (ABSR)

• How to adapt ABSR to new environments

• Influence of different components of ABSR on the performance

• Conclusions
Speaker-Dependent vs Speaker-Independent Acoustic Model

Assistant System

Feature Extractor

Domain Knowledge

ATC Grammar

ASR Decoder

N-Best Generator

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

Command Prediction Model

Command Hypotheses Generator

Corrector

Command Filtering

Command Extractor

Plausibility Checker

DATA

TEXT

COMMAND

USER

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

THY5QJ MAINTAIN_SPEED 220 OR_GREATER

THY5QJ DESCEND 3000 ft
Speaker-Dependent vs Speaker-Independent Acoustic Model

**Prague Approach Area**

<table>
<thead>
<tr>
<th>Baseline (Full ABSR)</th>
<th>Delta</th>
<th>ABSR with Speaker-Independent AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition Rate: 92.1%</td>
<td>Recognition Rate: -3.5%</td>
<td>Recognition Rate: 88.6%</td>
</tr>
<tr>
<td>Error Rate: 0.60%</td>
<td>Error Rate: +0.48%</td>
<td>Error Rate: 1.07%</td>
</tr>
</tbody>
</table>

**Vienna Approach Area**

<table>
<thead>
<tr>
<th>Baseline (Full ABSR)</th>
<th>Delta</th>
<th>ABSR with Speaker-Independent AM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition Rate: 83.3%</td>
<td>Recognition Rate: -0.7%</td>
<td>Recognition Rate: 82.6%</td>
</tr>
<tr>
<td>Error Rate: 3.21%</td>
<td>Error Rate: +0.21%</td>
<td>Error Rate: 3.42%</td>
</tr>
</tbody>
</table>
Influence of Plausibility Checker

Assistant System

Feature Extractor

ASR Decoder

N-Best Generator

Command Prediction Model

Command Hypotheses Generator

Corrector

Command Filtering

Command Extractor

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

THY5QJ MAINTAIN_SPEED 220 OR_GREATER
THY5QJ DESCEND 3000 ft

Plausibility Checker

DATA

TEXT

COMMAND

USER

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

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

THY5QJ MAINTAIN_SPEED 220 OR_GREATER
THY5QJ DESCEND 3000 ft
Influence of Plausibility Checker

Prague Approach Area

Baseline (Full ABSR)
- Recognition Rate: 92.1%
- Error Rate: 0.60%

Delta
- Recognition Rate: +1.9%
- Error Rate: +1.23%

ABSR without Plausibility Checker
- Recognition Rate: 94.0%
- Error Rate: 1.83%

Vienna Approach Area

Baseline (Full ABSR)
- Recognition Rate: 83.3%
- Error Rate: 3.21%

Delta
- Recognition Rate: +1.6%
- Error Rate: +3.00%

ABSR without Plausibility Checker
- Recognition Rate: 82.6%
- Error Rate: 6.21%
Influence of Command Hypotheses Generator

Assistant System

Feature Extractor

ASR Decoder

N-Best Generator

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

Domain Knowledge

ATC Grammar

Command Prediction Model

Command Hypotheses Generator

Corrector

Command Filtering

Plausibility Checker

DATA

TEXT

COMMAND

USER

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

THY5KJ MAINTAIN_SPEED 220 OR_GREATER
THY5KJ DESCEND 3000 ft
Influence of Command Hypotheses Generator

Prague Approach Area

Baseline (Full ABSR)
- Recognition Rate: 92.1%
- Error Rate: 0.60%

Delta
- Recognition Rate: -4.5%
- Error Rate: +6.07%

ABSR without Hypotheses Generator
- Recognition Rate: 87.5%
- Error Rate: 6.66%

Vienna Approach Area

Baseline (Full ABSR)
- Recognition Rate: 83.3%
- Error Rate: 3.21%

Delta
- Recognition Rate: -11.8%
- Error Rate: +11.78%

ABSR without Hypotheses Generator
- Recognition Rate: 71.5%
- Error Rate: 15.00%
Influence of N-Best Generator

Assistant System

ASR Decoder

N-Best Generator

Command Hypotheses Generator

Command Prediction Model

Corrector

Command Filtering

Command Extractor

Assistant System

Feature Extractor

Domain Knowledge

ATC Grammar

DATA

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

TEXT

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

COMMAND

THY5QJ MAINTAIN_SPEED 220 OR_GREATER
THY5QJ DESCEND 3000 ft

USER

Plausibility Checker

TYK5QJ

Assistant System

Feature Extractor

Domain Knowledge

ATC Grammar

DATA

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

TEXT

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

COMMAND

THY5QJ MAINTAIN_SPEED 220 OR_GREATER
THY5QJ DESCEND 3000 ft

USER

Plausibility Checker

TYK5QJ
Influence of N-Best Generator

**Prague Approach Area**

- **N-Best = 1**
  - Recognition Rate: 91.4%
  - Error Rate: 0.56%

- **N-Best = 5** (Baseline)
  - Recognition Rate: 92.1%
  - Error Rate: 0.60%

- **N-Best = 10**
  - Recognition Rate: 92.2%
  - Error Rate: 0.57%

- **N-Best = 20**
  - Recognition Rate: 92.1%
  - Error Rate: 0.73%

- **N-Best = 50**
  - Recognition Rate: 92.0%
  - Error Rate: 0.62%

**Vienna Approach Area**

- **N-Best = 1**
  - Recognition Rate: 83.7%
  - Error Rate: 3.21%

- **N-Best = 5** (Baseline)
  - Recognition Rate: 83.3%
  - Error Rate: 3.21%

- **N-Best = 10**
  - Recognition Rate: 83.7%
  - Error Rate: 3.21%

- **N-Best = 20**
  - Recognition Rate: 83.7%
  - Error Rate: 3.22%

- **N-Best = 50**
  - Recognition Rate: 83.9%
  - Error Rate: 3.16%
Conclusion

• The use of **Machine Learning** for an ABSR-System is possible

• **ABSR** divided into different conceptual modules with
  
  • Reusable Building Blocks (No adaptation necessary)
  
  • Automatically learned Models (Acoustic-, Language-, Command Prediction Model)
  
  • Partially manual updates of Data elements

• Not all parts of the system have to be implemented (if decrease in performance is acceptable)

• Command Hypotheses Generator has the biggest influence on the overall performance
Building Blocks of Assistant Based Speech Recognition for Air Traffic Management Applications

Thank you for your attention!

Matthias Kleinert, Hartmut Helmke
DLR, Institute of Flight Guidance
Controller Assistance

Matthias.Kleinert@dlr.de
Phone: +49 531 295-2567

Hartmut.Helmke@dlr.de
Phone: +49 531 295-2599
Command Prediction Model as Decision Tree

- Specific prediction areas for every command type (e.g. DESCEND, REDUCE, CLEARED ILS)
Expanding learned areas with different windows

Prediction Area of CPM for Cleared ILS commands of Arrivals
Experimental Set-Up for Proof-of-Concept

- Real world data provided ANS CR and Austro Control
- Four different controller positions
- 18.7h training data for Vienna and 18.1h trainings data for Prague

<table>
<thead>
<tr>
<th>Configuration</th>
<th># Total_cmds</th>
<th># Descend_cmds</th>
<th># ILS clearances</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEC (Prague)</td>
<td>11103</td>
<td>2184</td>
<td>351</td>
</tr>
<tr>
<td>PEC (Prague)</td>
<td>5365</td>
<td>920</td>
<td>458</td>
</tr>
<tr>
<td>BALAD (Vienna)</td>
<td>5929</td>
<td>1062</td>
<td>13</td>
</tr>
<tr>
<td>Feeder (Vienna)</td>
<td>6959</td>
<td>1100</td>
<td>245</td>
</tr>
</tbody>
</table>

Automatically transcribed (used for learning)

<table>
<thead>
<tr>
<th>Approach Area</th>
<th># Utterances</th>
<th># given commands</th>
<th># sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prague</td>
<td>2582</td>
<td>4563</td>
<td>27</td>
</tr>
<tr>
<td>Vienna</td>
<td>2427</td>
<td>3556</td>
<td>19</td>
</tr>
</tbody>
</table>

Transcribed manually (test data)
### User Acceptance (2)

<table>
<thead>
<tr>
<th></th>
<th>Prague:</th>
<th>Vienna:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Commands</td>
<td>Number of ABSR Errors/Rejections</td>
<td>Number of ABSR Errors/Rejections</td>
</tr>
<tr>
<td>396</td>
<td>36</td>
<td>610</td>
</tr>
<tr>
<td>Rec Rate</td>
<td>90.9%</td>
<td>86.9%</td>
</tr>
<tr>
<td>Corrected by controller</td>
<td>31</td>
<td>79</td>
</tr>
<tr>
<td>Detected by controller</td>
<td>36</td>
<td>80</td>
</tr>
</tbody>
</table>

No safety issues were observed.
All misrecognitions were detected.
Better Recognition Rate of course would improve even more workload reduction.
Building Blocks of Assistant Based Speech Recognition for Air Traffic Management Applications

Matthias Kleinert, Hartmut Helmke, Heiko Ehr, Gerald Siol (DLR)
Christian Kern (Austro Control)
Dietrich Klakow, Mittul Singh (UdS)
Petr Motlicek (Idiap)

German Aerospace Center (DLR), Braunschweig
Saarland University (UdS), Saarbrücken
Idiap Research Institute, Martigny
Austro Control, Vienna