Semi-supervised Adaptation of Assistant Based Speech Recognition Models for different Approach Areas

Matthias Kleinert, Hartmut Helmke, Gerald Siol, Heiko Ehr (DLR)
Aneta Cerna (ANS CR)
Christian Kern (Austro Control)
Dietrich Klakow, Youssef Oualil, Mittul Singh (Saarland University)
Petr Motlicek, Ajay Srinivasamurthy (Idiap)

Presentation at 37th DASC, London 2018-09-26
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 annotation

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

MALORCA = Machine Learning of Speech Recognition Models for Controller Assistance
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 Approach Area by Machine Learning
  • Experiment Results

• Motivation for ontology in SESAR 2020 PJ.16-04

• Ontology for command annotation

• Next challenges and conclusions
From Paper to Electronic Flight Strips with ASR Support

All flight information is 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. ASR (= Automatic Speech Recognition) is a solution.
**AcListant®-Strips: Validation at DLR in 2015**

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

**Goal:**
Quantify the benefits for ASR with respect to
- 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
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 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
Assistant Based Speech Recognition (ABSR)
Intermediate Results, BUT

- 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, ...)

BUT...

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. This includes the huge effort for validation and benefit analysis.
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 Approach Area by Machine Learning
  • Experiment Results

• Motivation for ontology in SESAR PJ.16-04

• Ontology for command annotation

• Next challenges and conclusions
Motivation – The MALORCA Project

Instead of (highly skilled and paid) experts, machine learning is used.
From Speech Signal to HMI

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

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

> MALORCA> H. Helmke > ATC AIRBUS Challenge Workshop,> Toulouse > 2018-10--04

11
From Speech Signal to HMI

**Command Prediction Model**

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

**AM = Acoustic Model**

**LM = Language Model**

**Lex = Lexicon**

**Checker**

```
THY5KJ
MAINTAIN_SPEED 220 OR_GREATER
THY5KJ
DESCEND 3000 ft
```
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 Approach Area by Machine Learning
  • Experiment Results

• Motivation for ontology in SESAR PJ.16-04

• Ontology for command annotation

• Next challenges and conclusions
Model Training / Learning

DATA

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

Command Prediction Model

Command hypotheses generator

Command Extractor

Checker

THY5KJ
MAINTAIN_SPEED 220 OR_GREATER
THY5KJ
DESCEND 3000 ft

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

AM = Acoustic Model
LM = Language Model
Lex = Lexicon
Semi-supervised learning (AM, LM, CPM)

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

Data → ABSR → Data → Checker → Good Data → Invention of MALORCA

State-of-the-Art with normal Speech Recognizer,
Assistant Based Speech Recognition
Learning Curve for Prague

Command Recognition Rate

Extrapolation:
8 times more data may provide 92.6% (Prague) resp. 90.2% (Vienna)
Effect of Different Components

AM = Acoustic Model
LM = Language Model
Lex = Lexicon

recRate: 87.5%
errRate: 6.7%

Command Prediction Model

With all:
RecRate: 91.8%
ErrRate: 0.57%

Command hypotheses generator

Command Extractor

Checker

USER

DATA

TEXT

COMMAND

AMORCA> H. Helmke > ATC AIRBUS Challenge Workshop, Toulouse > 2018-10-04
Effect of Automatic Learning and Assistant Based Speech Recognition

Baseline Speech Recognizer:
- Recogn. Rate: 85.9%
- Error Rate: 7.9%

Trained by 19 h with/without Cmd. Pred.
- Recogn. Rate: 87.5%
- Error Rate: 6.7%

Prague Approach Area
- With Trained Cmd. Pred.
  - Recogn. Rate: 91.8%
  - Error Rate: 0.6%

Baseline Speech Recognizer:
- Recogn. Rate: 67.2%
- Error Rate: 18.9%

Trained by 17 h with/without Cmd. Pred.
- Recogn. Rate: 71.3%
- Error Rate: 15.7%

Vienna Approach Area
- With Trained Cmd. Pred.
  - Recogn. Rate: 85.2%
  - Error Rate: 3.7%

ABSR = Assistant Based Speech Recognition
CPM = Command Prediction Model
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 Approach Area by Machine Learning
  • Experiment Results

• Motivation for ontology in SESAR PJ.16-04
  • Ontology for command annotation
  • Next challenges and conclusions
Ontology for Transcription of ATC Speech Commands of SESAR 2020 Solution PJ.16-04

Hartmut Helmke et al.
German Aerospace Center (DLR)
Braunschweig, Germany

Presentation at 37th DASC, London 2018-09-27
Controller:

and sky_travel four kilo mike turn right heading three three zero

Pilot 1:

three thirty to the right sky_travel four kilo mike

Pilot 2:

right three three zero to the right four kilo mike
Controller-to-Pilot Voice Annotation

<table>
<thead>
<tr>
<th>Controller:</th>
<th>Transcription</th>
</tr>
</thead>
<tbody>
<tr>
<td>and <strong>sky_travel four kilo mike turn right heading three three zero</strong></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pilot 1:</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>three thirty to the right sky_travel four kilo mike</strong></td>
<td></td>
</tr>
</tbody>
</table>

| Pilot 2: | |
|----------| |
| **right three three zero to the right four kilo mike** | |

What do you prefer as semantic transcription?

1. TVS4KM TURN_RIGHT_HEADING 330
2. **Filler** TVS4KM TURN_RIGHT_HEADING 330
3. TVS4KM HEADING 330
4. TVS4KM HEADING 330 **deg RIGHT**
5. **TVS 4KM TURN_RIGHT_HEADING 330**
Lists of Authors

Hartmut Helmke, Institute of Flight Guidance, DLR, Braunschweig, Germany
Michael Slotty, ATM Simulator Centre, DFS, Langen, Germany,
Michael Poiger, Frequentis, Vienna, Austria
Damián Ferrer Herrer, CRIDA, Madrid, Spain,
Oliver Ohneiser, Institute of Flight Guidance, DLR, Braunschweig, Germany
Nathan Vink, NATS, Whiteley, Fareham, United Kingdom
Aneta Cerna, Air Navigation Services of the Czech Republic, Jeneč, Czech Republic
Petri Hartikainen, Integra, Copenhagen, Denmark
Billy Josefsson, LFV, Stockholm, Sweden
David Langr, Air Navigation Services of the Czech Republic, Jeneč, Czech Republic
Raquel García Lasheras, CRIDA, Madrid, Spain
Gabriela Marin, Romatsa, Bucharest, Romania
Odd Georg Mevatne, Avinor, Oslo, Norway
Sylvain Moos, ATM Innovation Lab, Thales Air Systems, Rungis, France
Mats N. Nilsson, LFV, Stockholm, Sweden
Mario Boyero Pérez, Indra, Madrid, Spain, now with Eurocontrol, Brussels, Belgium
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

Partner/LTP
- only linked via Grant

ATM System provider
ANSPs
Research and Consulting Institutes
What is the concrete PJ.16-04 content?

- Automatic Speech Recognition
- User Profile Management
- Multi-Touch Inputs
- Efficient Process
- Attention Guidance
- CWP Virtualisation
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 Approach Area by Machine Learning
  • Experiment Results

• Motivation for ontology in SESAR PJ.16-04

• Ontology for command annotation

• Next challenges and conclusions
“We annotate an utterance ‘callsign instruction, instruction …’ as

Callsign Instruction Callsign Instruction Callsign Instruction ...

Callsign Instruction

Command

Type Value(s) Unit Qualifier

Condition(s)

Conjunction Requirement
# Vertical Commands (1)

<table>
<thead>
<tr>
<th>Type</th>
<th>Value(s)</th>
<th>Unit</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESCEND</td>
<td></td>
<td>FL</td>
<td>BELOW OR_ABove</td>
</tr>
<tr>
<td>CLIMB</td>
<td>altitude value / FL value</td>
<td>ft</td>
<td>ABOVE OR_BELOW OR_ABOVE</td>
</tr>
<tr>
<td>ALTITUDE</td>
<td></td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>STOP_DESCEND</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STOP_CLIMB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STOP_ALTITUDE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAINTAIN</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PRESENT_ALTITUDE</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Vertical Commands (2)

<table>
<thead>
<tr>
<th>Type</th>
<th>Value(s)</th>
<th>Unit</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAINTAIN</td>
<td>Alt Value / FL Value</td>
<td>FL / ft none</td>
<td>BELOW OR ABOVE</td>
</tr>
<tr>
<td>LEVEL</td>
<td>POS / TIME</td>
<td></td>
<td>ABOVE OR BELOW OR ABOVE</td>
</tr>
</tbody>
</table>

Note: MAINTAIN and LEVEL are the types of commands, ALTITUDE and POS / TIME are the types of value(s), and FL / ft none are the units of measurement.
Command Types Supported

1. **Vertical Commands (Approach, Enroute)**
2. Changing Direction Commands (Approach, Enroute)
3. Vertical Speed Commands (Approach, seldom in Enroute)
4. Speed-Commands (Approach, Enroute)
5. CLEARED and CANCEL-Type-Family
6. Frequency-Changing Commands
7. Information Commands
8. Airborne-Holdings (Approach)
9. VFR Clearances (Tower)
10. REPORT
11. Request Commands (Pilot)
12. Miscellaneous Commands
13. Taxi Commands (Ground, Tower)
14. Ground-Command-Type-Family (Ground)
# Speed Commands (1)

<table>
<thead>
<tr>
<th>Transcription</th>
<th>tunair five one four <em>reduce</em> one eighty knots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>TAR514 REDUCE 180 kt</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transcription</th>
<th>tunair five one four <em>reduce</em> one eighty</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>TAR514 REDUCE 180 none</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transcription</th>
<th>tunair five one four <em>speed</em> two three zero knots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>TAR514 SPEED 230 kt</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Transcription</th>
<th>tunair five one four two three zero knots</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>TAR514 SPEED 230 kt</td>
</tr>
</tbody>
</table>
## Speed Commands (2)

<table>
<thead>
<tr>
<th>Transcription</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>tunair five one four <strong>reduce</strong> one eighty knots or less</td>
<td>TAR514 REDUCE 180 kt OR_LESS</td>
</tr>
<tr>
<td>wizzair seven zero one we have departing traffic please <strong>reduce</strong> speed</td>
<td>WZZ701 INFORMATION TRAFFIC</td>
</tr>
<tr>
<td></td>
<td>WZZ701 REDUCE none none</td>
</tr>
<tr>
<td>austrian seven six eight zulu <strong>reduce</strong> to final approach speed</td>
<td>AUA768Z REDUCE_FINAL_APPROACH_SPEED</td>
</tr>
</tbody>
</table>
Speed Commands Overview

<table>
<thead>
<tr>
<th>Type</th>
<th>Type 2</th>
<th>Value(s)</th>
<th>Unit</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAINTAIN</td>
<td>SPEED</td>
<td>Speed-Value</td>
<td>kt</td>
<td>LESS, OR_LESS, GREATER, OR_GREATER</td>
</tr>
<tr>
<td>INCREASE</td>
<td>REDUCE</td>
<td>Speed-Value</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>SPEED</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Value(s)</th>
<th>Unit</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCREASE _BY</td>
<td>Speed-Value-2</td>
<td>kt</td>
<td>OR_GREATER</td>
</tr>
<tr>
<td>REDUCE _BY</td>
<td>none</td>
<td>MA</td>
<td>none</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type</th>
<th>Value(s)</th>
<th>Unit</th>
<th>Qualifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAINTAIN</td>
<td>PRESENT_SPEED</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

CWP HMI SESAR
JOINT UNDERTAKING
## Conditional Clearances (1)

<table>
<thead>
<tr>
<th>Transcription</th>
<th>CSA seven zero seven hotel <strong>reduce speed one six zero knots</strong> to maintain <strong>until four miles final</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Annotation</strong></td>
<td>CSA707H <strong>REDUCE 160 kt UNTIL TOUCHDOWN DISTANCE 4 NM</strong></td>
</tr>
</tbody>
</table>
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 Approach Area by Machine Learning
  • Experiment Results

• Motivation for ontology in SESAR PJ.16-04

• Ontology for command annotation

• Next challenges and conclusions
Conclusions

• There is no data than more data!!!
• We need (automatic) annotations > 1000 hours (at least for ABSR)
• We need to join our efforts
• Suggested ontology of PJ.16-04 partners is only a first step, BUT a significant one
• This is a European answer. Contributions from US, Asia etc. are encouraged.
• We focus on ground side (the controller). Which challenge can we expect from pilot side?
• We focus on approach (and tower)? Surprises on en-route side?
• Only English. What about Chinese or Dutch or French?
• Common standard enables harmonized implementations (e.g. replace one ASR implementation by another)
Conclusions

• **Machine Learning** of Acoustic Language Model & Command Prediction Model is possible

• Command Recognition Rate improves from 80% to 92% (Prague) resp. from 60% to 85% (Vienna)

Assistant Based Speech Recognition improves ASR performance (Context integration)
  • Command Recognition Rate **improves** from 86% to 92% (Prague)
  • Command Recognition Error Rate **improves** from 7% to 0.6% (Prague)

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