Machine Learning of Controller Command Prediction Models from Recorded Radar Data and Controller Speech Utterances

Matthias Kleinert, Hartmut Helmke, Gerald Siol, Heiko Ehr, Michael Finke, (DLR)
Youssef Oualil, (UdS)
Ajay Srinivasamurthy (Idiap)

German Aerospace Center (DLR), Braunschweig
Saarland University (UdS), Saarbrücken
Idiap Research Institute, Martigny
Motivation – The AcListant® Project

- In 2015 the DLR developed an Active Listening Assistant (AcListant®) in cooperation with Saarland University.
- Automatic Speech Recognition of controller-pilot communication
  → Recognition rate of 95%
- Reduction of controller workload by automating flight strip management
  → Flight strips are integrated into radar labels
  → Automatic insertion of controller commands
  → Radar Label maintenance time reduced by a factor of 3
The AcListant® System uses three different models:

- **Command Prediction Model (CPM)**
  - Predict what an air traffic controller could say
  - Based on radar information
  - Hypotheses Generator derives a set of possible commands e.g.
    - **DLH3ER REDUCE 220 KT**
    - **DLH3ER DESCEND 80 FL**
    - **AUA201 QNH 1022**
    - **AUA201 DESCEND 5000 FT**
  - Reduction of search space for ASR
Motivation – The AcListant® Project

The AcListant® System uses three different models:

- Acoustic Model (AM) and Language Model (LM)
  - **Basic models** for speech recognition
  - Extract the sequence of spoken words e.g.
    
    Hello Austrian two zero one
descend five thousand feet
qnh is one zero two two

- Sequence Labeling Approach to extract relevant commands:
  - AUA201 DESCEND 5000 FT
  - AUA201 QNH 1022
Motivation – The AcListant® Project

**Issue:**
- Large costs of deployment
  - AcListant budget was **1.3M Euro**

→ **Result:** Good working Speech recognizer for one airport
- Models have to be adapted manually for every airport / CWP due to:
  - Different **accents**
  - Different **working procedures**
  - Different **airspace layouts**
- Requires time and expert knowledge
- Adaptation is necessary even in an existing system
Motivation – The MALORCA Project

AcListant

Radar Data
Voice Recordings
Group of experts

Command Prediction Model
Acoustic Model
Language Model

MALORCA

Radar Data
Voice Recordings
Automatic Learning Process

Command Prediction Model
Acoustic Model
Language Model
Motivation – The MALORCA Project

- Focus of the paper lies on the Command Prediction Model (CPM)
  - How can it be learned automatically?
  - How does it influence the outcome of the Speech Recognition System?
- Fixed Acoustic Model (AM) and Language Model (LM) have been used
Command Prediction Model as Decision Tree

- Specific prediction areas for every command type (e.g. DESCEND, REDUCE, CLEARED ILS)
Visual Representation of a CPM Prediction Area

- Every symbol represents an area of 1 nm by 1nm
- The marked areas also contain the corresponding values for a command
Automatic learning of the Command Prediction Model

However: recognized commands could be wrong!

- "sky_travel two five zero nine descend flight level nine zero"
- Could result in: TVS2509 DESCEND 90 FL or TVS2509 REDUCE 190 or…
Expanding learned areas with different windows

Before Expansion

Prediction Area of CPM for Cleared ILS commands of Arrivals
Expanding learned areas with different windows

Before Expansion

After Expansion

Prediction Area of CPM for Cleared ILS commands of Arrivals
Expanding learned areas with different windows

Before Expansion

After Expansion

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

**Actual System**
- Command Prediction Model
- Hypotheses Generator
  - Command Hypotheses
  - Automatic Speech Recognition (ASR)
  - Recognized Commands
- Acoustic Model
- Language Model

**Evaluation System**
- Audio Signal
- Automatic Speech Recognition (ASR)
  - Recognized Commands
  - Command Prediction Model
  - Compare Hypothesis with Recognized Commands
- Hypotheses Generator
  - Command Hypotheses
- Acoustic Model
  - Language Model
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)
Results – Baseline

- Baseline uses full context so almost all possible commands are predicted
- Evaluation of three basic metrics
  - **RecR** = Rate of commands correctly recognized
  - **ErrR** = Rate of recognized commands which were not spoken and not rejected
  - **#NPC** = Average Number of predicted commands

<table>
<thead>
<tr>
<th></th>
<th>RecR [%]</th>
<th>ErrR [%]</th>
<th>#NPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prague</td>
<td>88.12</td>
<td>1.13</td>
<td>2054</td>
</tr>
</tbody>
</table>
Results – Prague different expansion window sizes

- Compared to Baseline loss in RecR is 0.38%, but improvement in ErrR is 0.28%
- Number of predicted commands is reduced by a factor of almost 3.5

<table>
<thead>
<tr>
<th>Window Size</th>
<th>RecR [%]</th>
<th>ErrR [%]</th>
<th>#NPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>88.12</td>
<td>1.13</td>
<td>2054</td>
</tr>
<tr>
<td>3x3</td>
<td>82.68</td>
<td>0.68</td>
<td>464</td>
</tr>
<tr>
<td>5x5</td>
<td>85.31</td>
<td>0.70</td>
<td>511</td>
</tr>
<tr>
<td>7x7</td>
<td>86.05</td>
<td>0.70</td>
<td>541</td>
</tr>
<tr>
<td>9x9</td>
<td>86.58</td>
<td>0.77</td>
<td>562</td>
</tr>
<tr>
<td>11x11</td>
<td>86.93</td>
<td>0.77</td>
<td>576</td>
</tr>
<tr>
<td>13x13</td>
<td>86.47</td>
<td>0.76</td>
<td>552</td>
</tr>
<tr>
<td>15x15</td>
<td>86.78</td>
<td>0.80</td>
<td>564</td>
</tr>
<tr>
<td>17x17</td>
<td>87.14</td>
<td>0.80</td>
<td>574</td>
</tr>
<tr>
<td>19x19</td>
<td>87.37</td>
<td>0.83</td>
<td>583</td>
</tr>
<tr>
<td>21x21</td>
<td>87.56</td>
<td>0.85</td>
<td>591</td>
</tr>
<tr>
<td>25x25</td>
<td>87.74</td>
<td>0.85</td>
<td>604</td>
</tr>
<tr>
<td>29x29</td>
<td>87.78</td>
<td>0.93</td>
<td>615</td>
</tr>
</tbody>
</table>
Results – Effect of the different amount of Training data

- Increase of $\text{RecR}$ between 90% and 100% at window size 29 by 29 nm is still 0.02% for Prague
Conclusion and Future Work

• Recognition Rate of 88% seem low compared to the AcListant® Project (95%) but:
  • AcListant® also used additional information of an AMAN
  • Real world data instead of simulated data
  • Command Prediction Model was only used to filter the output of ASR and not to influence the output itself
  • In the final system all **models will improve each other** over time
    → So recognition rates of the final system will improve
  • Improved filtering of outliers
  • Varying the effect expansion windows have on surrounding areas
  • Using command type dependent expansion windows
Machine Learning of Controller Command Prediction Models from Recorded Radar Data and Controller Speech Utterances

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