Tower Controller Command Prediction for Future Speech Recognition Applications

Oliver Ohneiser (DLR, Braunschweig), H. Helmke, M. Kleinert, G. Siol, H. Ehr, S. Hobein, A.-V. Predescu, J. Bauer
Motivation for Speech Recognition in Air Traffic Control (ATC)

• Air Traffic Controllers (ATCos) utter verbal clearances to pilots
• Command content needs to be manually entered into ATC system
• Induces (unnecessary) ATCo workload
• Command content is already within utterance
  → Automatic Speech Recognition (ASR)
Background of Command Hypotheses in Approach Environment

• Even better ATCo support with Assistant Based Speech Recognition (ABSR)

• Controller command hypotheses
  • Dramatically decrease command recognition error rate (CRER)
  • Increase command recognition rate (CRR)
  • Quality measurement: command prediction error rate (CPER)
  • Further plausibility checking

• Good results for approach area (Düsseldorf, Prague, Vienna)
Approach Command Hypotheses Generator

<table>
<thead>
<tr>
<th>Setup-Project</th>
<th>CPER [%]</th>
<th>CRER [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Düsseldorf-AcListant®</td>
<td>0.9</td>
<td>1.7</td>
</tr>
<tr>
<td>Vienna-MALORCA</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>Prague-MALORCA</td>
<td>2.3</td>
<td>0.6</td>
</tr>
<tr>
<td>Vienna-PJ.16-04</td>
<td>4.8</td>
<td>6.1</td>
</tr>
<tr>
<td>Prague-PJ.16-04</td>
<td>0.3</td>
<td>5.4</td>
</tr>
</tbody>
</table>

- CPER: command *prediction* error rate
- CRER: command *recognition* error rate
Motivation and Goals for Command Hypotheses in Tower Environment

• No ABSR yet for the tower environment (but reasonable in future, e.g. follow-the-greens)
• Digitization (Commission Implementing Regulation (EU) No 716/2014)

• Goals/Objectives for Tower Command Hypotheses Generator (TCHG)
  • CPER < 10% with standard deviation < 2.5%
  • Context prediction time < 5 s (Context=Set of predicted commands)

• ABSR: Assistant Based Speech Recognition

Source: https://commons.wikimedia.org/wiki/File:Tower_(602732164).jpg
Embedded in SESAR Wave-1 Projects

- PJ.16-04 Controller Working Position / Human Machine Interface
  - Solution Lead: DLR
  - 5 interaction technology related activities (including speech recognition)
  - ASR validation exercise “Command Hypotheses for Multiple Remote Tower”

- PJ.05-02 Remote Tower
  - Project Lead: DLR
  - Simulation setup provided
Multiple Remote Tower Setup in DLR Tower Lab

Used airports in two setups:
Budapest / Debrecen / Papa
Vilnius / Kaunas / Palanga
Human-in-the-Loop Trials with Data Recording

- HungaroControl, Hungary (HC):
  7 ATCos per dataset (pre-trial/trial)
- Oro Navigacija, Lithuania (ON):
  6 ATCos per dataset (pre-trial/trial)
- 111 simulation runs (85 h), 9’300 voice utterances $\rightarrow$ 20’000 controller commands
- 4’400 voice utterances transcribed, annotated and checked (9 hours net, respectively 38 hours “with silence”)

Datasets

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hungary-2017-11</td>
</tr>
<tr>
<td>Lithuania-2018-03</td>
</tr>
<tr>
<td>Hungary-2018-11</td>
</tr>
<tr>
<td>Lithuania-2018-12</td>
</tr>
</tbody>
</table>

Sources of flags:
Hungary: https://commons.wikimedia.org/w/index.php?curid=343602
Lithuania: https://commons.wikimedia.org/w/index.php?curid=343633
### Tool for Transcription and Annotation (1)

#### TransAnno2.0

<table>
<thead>
<tr>
<th>Callsign</th>
<th>Type</th>
<th>2nd Type</th>
<th>Value</th>
<th>Qualifier</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO_CALLSIGN</td>
<td>CLIMB</td>
<td>CONFIRM_ACCEPT</td>
<td>2000 ft</td>
<td>ABOVE</td>
<td></td>
</tr>
<tr>
<td>NARM</td>
<td>CONTACT</td>
<td>3000 ft</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AE0500</td>
<td>CONTACT_FREQUENCY</td>
<td>4000 ft</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRUB925</td>
<td>CORRECTION</td>
<td>5000 ft</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BUC1182</td>
<td>CROSS</td>
<td>6000 ft</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DUL4676</td>
<td>DESCEND</td>
<td>7000 ft</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FHG681</td>
<td>DIRECT</td>
<td>8000 ft</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HACIC</td>
<td>DIRECT_TO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAKW1</td>
<td>DISREGARD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAWK1</td>
<td>ENTER_CTR</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARX68</td>
<td>EXPECT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IFA1483</td>
<td>FOLLOW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORIDY</td>
<td>FOLLOW_ROUTE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SWR2252</td>
<td>GO_AROUND</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRA720Q</td>
<td>HEADING</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W2Z1PU</td>
<td>HOLD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W2Z391</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TransaViva seven two Quebec Budapest Tower good day startup and pushback is approved you are cleared to destination. Echo Hotel Romeo Delta from runway one three right follow Gilep one delta initially climb to seven thousand squawk zero five one four.
Tool for Transcription and Annotation (2)

- wav-files (audio, “wave”)

```
Filelist

2017-11-13_16-08-00-48.wav
2017-11-13_16-09-08-18.wav
2017-11-13_16-10-17-28.wav
2017-11-13_16-10-29-58.wav
2017-11-13_16-10-48-48.wav
2017-11-13_16-11-04-45.wav
2017-11-13_16-11-32-07.wav
2017-11-13_16-12-19-78.wav
```

Play

Duration: 16 Seconds
Volume: 50
Done: 123
Todo: 0
Tool for Transcription and Annotation (3)

- cor-files (transcription, “correct”)

- Word sequences to be transferred into meanings
Tool for Transcription and Annotation (4)

- cmd-files (annotation, “command”)

```plaintext
TRA72Q INIT_RESPONSE
TRA72Q STARTUP
TRA72Q PUSHBACK
TRA72Q CLEARED TO EHRD
TRA72Q INFORMATION ACTIVE_RWY 13R
TRA72Q CLEARED VIA GILEP_1D
TRA72Q CLIMB 7000 ft
TRA72Q SQUAWK 0514
```
Tool for Transcription and Annotation (5)

- Quality assurance and follow defined processes
  - Ontology (annotation rules) support
- Acceleration for efficient usage
- Automatic context check
- Automatic plausibility check
Applying Machine Learning

- Determination of parameter “window” size
- Best window (raster size as certain rectangle in terms of latitude and longitude) for command predictions

- Need for HUGE amount of data (first training, then testing)
- But only small amount in ATC ASR research
“Window” Size (CPER dependent)

- Hungary-2017-11 data was split into two halves with 50% data amount each (A/B) for training and testing the other half.
“Window” Size (Context Size dependent)

- max = biggest
- avg = average

...number of hypotheses in context

- Always lower than maximum possible number of hypotheses

Window Size Evaluation with Respect to Context Size

- Context max Data Half A
- Context avg Data Half A
- Context max Data Half B
- Context avg Data Half B
“Window” Size Determination

- Trade-off between
  - Low CPER
  - Low Context Size

- “Window” size 11 was chosen (changes in percent after decimal compared to predecessor)
Training and Test Data Sets

<table>
<thead>
<tr>
<th>Name of Evaluation</th>
<th>Training with Dataset</th>
<th>Test with Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>LITHUANIA</td>
<td>Lithuania-2018-03</td>
<td>Lithuania-2018-12</td>
</tr>
</tbody>
</table>

- What you learn, is what you expect!
- Runway change in test data not equal to training data
Results Overview: CPER and Context Size

<table>
<thead>
<tr>
<th>Name of Evaluation</th>
<th>CPER</th>
<th>SD</th>
<th>ctx_avg</th>
<th>ctx_max</th>
</tr>
</thead>
<tbody>
<tr>
<td>HUNGARY</td>
<td>7.8%</td>
<td>2.57%</td>
<td>450</td>
<td>593</td>
</tr>
<tr>
<td>LITHUANIA</td>
<td>12.5%</td>
<td>3.6%</td>
<td>340</td>
<td>498</td>
</tr>
</tbody>
</table>

- CPER: command *prediction* error rate
- SD: standard deviation
Results: Context Generation

- 21,000 times in 59 simulation runs of final trials
- Each 119 ms (in average)
- Requirement: Faster than 5 s (due to radar update rate, duration of utterances, and expected situation change)

- RT: Radio telephony
13 ATCos in Debriefing

• Discussion about transcription, annotation, and checking with hands-on experience on tool

• Potential use of ABSR in tower environment
  • Comparisons (Safety)
  • List of clearances (Safety)
  • ATC system maintenance (Workload)
  • Etc.
Potential for Workload Reduction
Radio Telephony (Number/Portion per Hour)
Summary of Results

- Command Prediction Error Rate (CPER) was sufficiently low → down to < 8% (first time for tower environment)
- Command Prediction Time was sufficiently fast → 120ms, so faster than required

⚠ Positive effects on tower controller’s workload in the future can be assumed, but – conform to planning – not proven with this exercise
Outlook on Future Work (1)

• Implementation of enhanced TCHG
  • Set of hypotheses should be minimized with individual window sizes per command type

<table>
<thead>
<tr>
<th>Controller Command Type</th>
<th>Best Window Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>INFORMATION</td>
<td>9</td>
</tr>
<tr>
<td>CLEARED</td>
<td>8</td>
</tr>
<tr>
<td>INIT_RESPONSE</td>
<td>11</td>
</tr>
<tr>
<td>REPORT</td>
<td>12</td>
</tr>
<tr>
<td>TAXI</td>
<td>11</td>
</tr>
<tr>
<td>CLIMB</td>
<td>2</td>
</tr>
<tr>
<td>CONTACT_FREQUENCY</td>
<td>7</td>
</tr>
<tr>
<td>VACATE</td>
<td>9</td>
</tr>
<tr>
<td>PUSHBACK</td>
<td>1</td>
</tr>
</tbody>
</table>

• Integration of enhanced TCHG with speech recognizer to generate first ABSR system for tower environment
Outlook on Future Work (2)

• Bigger amount of data
  • Available radar and speech data for training of all scenarios and environments tested
  • Annotated speech data to then optimize CPER

• Aerodrome ATC includes more types of clearances
  • Comparable results than in approach area?
Outlook on Succeeding Project

• SESAR Wave 2 project PJ.05 “Digital Technologies for Tower”, solution 97 “HMI Interaction modes for Airport Tower”; TRL2 → TRL4

• Validation exercises with partners B4 (ANS-CR, ON), COOPANS (ACG, CCL) at Remote TowerLab DLR, Braunschweig and Tower platform EUROCONTROL, Brétigny
Thank you

The opinions expressed herein reflect the author’s view only. Under no circumstances shall the SESAR Joint Undertaking be responsible for any use that may be made of the information contained herein. The project has received funding from the SESAR Joint Undertaking under the European Union's Horizon 2020 research and innovation programme under grant agreement No 734141.