

Tower Controller Command Prediction for Future Speech Recognition Applications

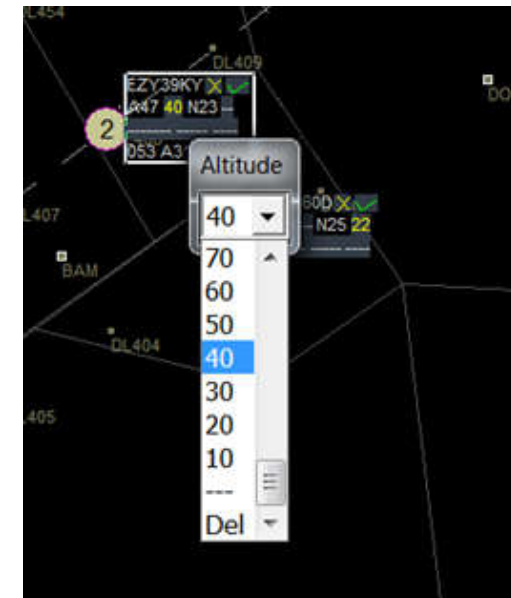
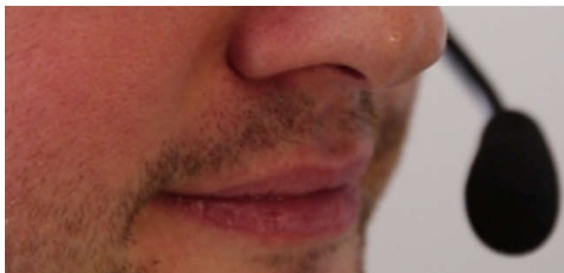
9th SESAR Innovation Days



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

Setup-Project	CPER [%]	CRER [%]
Düsseldorf-AcListant®	0.9	1.7
Vienna-MALORCA	3.2	3.2
Prague-MALORCA	2.3	0.6
Vienna-PJ.16-04	4.8	6.1
Prague-PJ.16-04	0.3	5.4



- 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 **T**ower **C**ommand **H**ypotheses **G**enerator (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](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



founding members



Multiple Remote Tower Setup in DLR Tower Lab

Used airports in two setups:
Budapest / Debrecen / Papa
Vilnius / Kaunas / Palanga



Airport 1

Airport 2

Airport 3



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
→ >20'000 *controller commands*
- 4'400 voice utterances transcribed,
annotated and checked (9 hours net,
respectively 38 hours “with silence”)

<i>Datasets</i>
Hungary-2017-11
Lithuania-2018-03
Hungary-2018-11
Lithuania-2018-12

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

File Edit View Extras ?

Open: D:/PJ16/04-02-ASR/TWR/pj16bwr_20171113_164944_atco1_scn2/wav

Removed: 2017-11-13_16-49-31-44.wav

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

Is Editable: [] Undo: [] Do not check file with TransAnnoOnlineChecker: []

Callsign	Type	2nd Type	Value	Qualifier	Condition
NO_CALLSIGN	CLIMB		2000 ft	ABOVE	
9AJIM	CONFIRM_ACCEPT		3000 ft	BELOW	
AEG550	CONTACT		4000 ft	OR_ABOVE	
BRU8925	CONTACT_FREQUENCY		5000 ft	OR_BELOW	
BUC1182	CONTINUE		6000 ft		
DLH4TN	CORRECTION		7000 ft		
FHY631	CROSS		8000 ft		
HACIZ	DESCEND				
HAWK1	DIRECT				
HBVPA	DIRECT_TO				
IFA1483	DISREGARD				
OKOKV	ENTER_CTR				
SWR225Z	EXPECT				
TRA72Q	FOLLOW				
WZZ1PU	FOLLOW_ROUTE				
WZZ391	GO_AROUND				
	HEADING				
	HOLD				

Add Command: 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

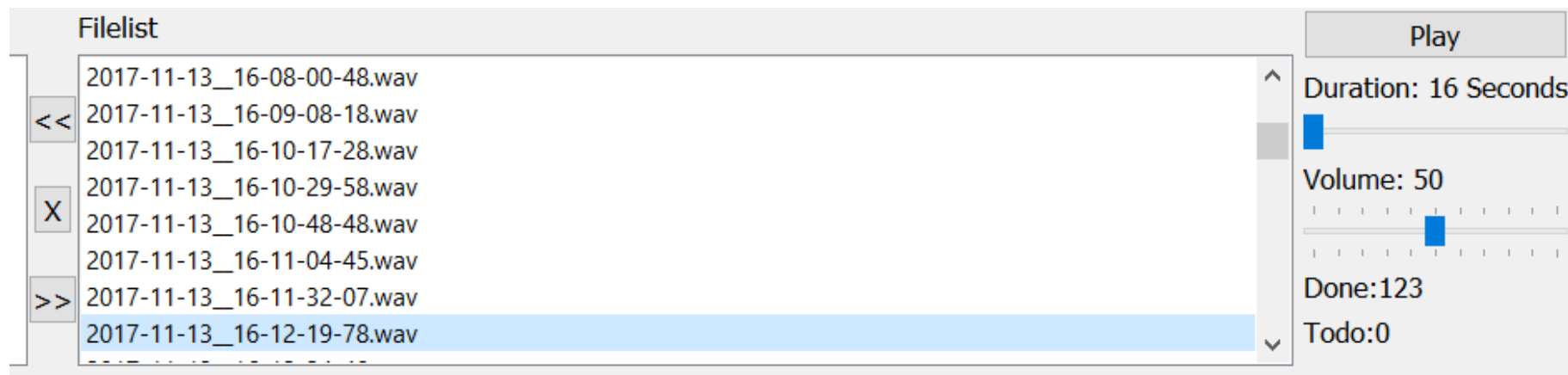
Remove Command: []

Up, Down, Check File, Check Dir, Lock

transavia 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

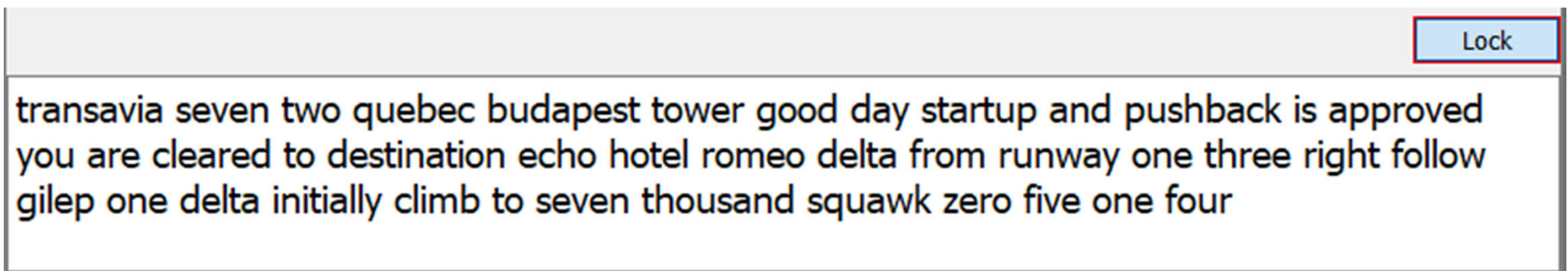
Tool for Transcription and Annotation (2)

- wav-files (audio, “wave”)



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”)

Callsign	Type	2nd Type	Value	Qualifier	Condition
NO_CALLSIGN	CLIMB		2000 ft		
9AJIM	CONFIRM_ACCEPT		3000 ft	ABOVE	
AEG550	CONTACT		4000 ft	BELOW	
BRU8925	CONTACT_FREQUENCY		5000 ft	OR_ABOVE	
BUC1182	CONTINUE		6000 ft	OR_BELOW	
DLH4TN	CORRECTION				
FHY631	CROSS				
HACIZ	DESCEND				
HAWK1	DIRECT				
HBVPA	DIRECT_TO				
IFA1483	DISREGARD				
OKOKV	ENTER_CTR				
SWR225Z	EXPECT				
TRA72Q	FOLLOW				
WZZ1PU	FOLLOW_ROUTE				
WZZ391	GO_AROUND				
	HEADING				
	HOLD				

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

Up

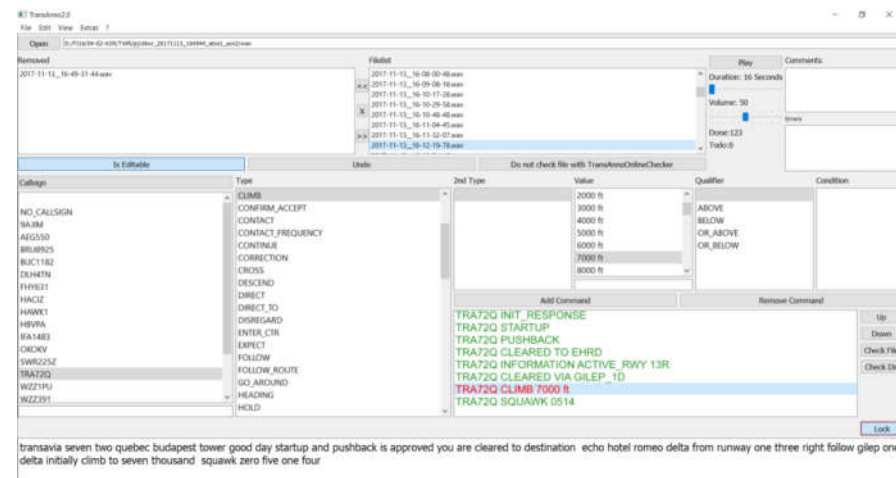
Down

Check File

Check Dir

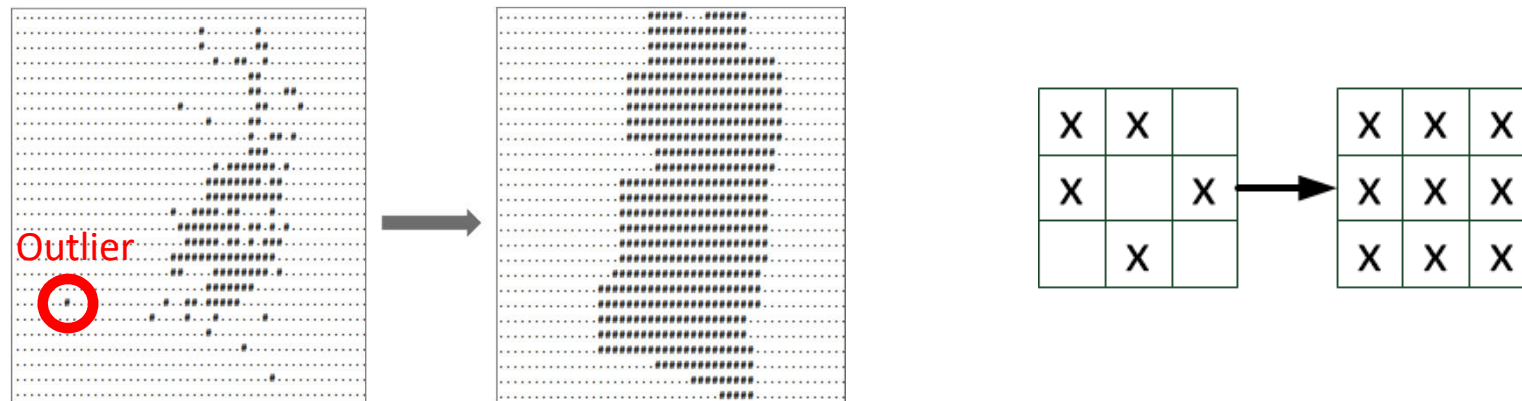
Tool for Transcription and Annotation (5)

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



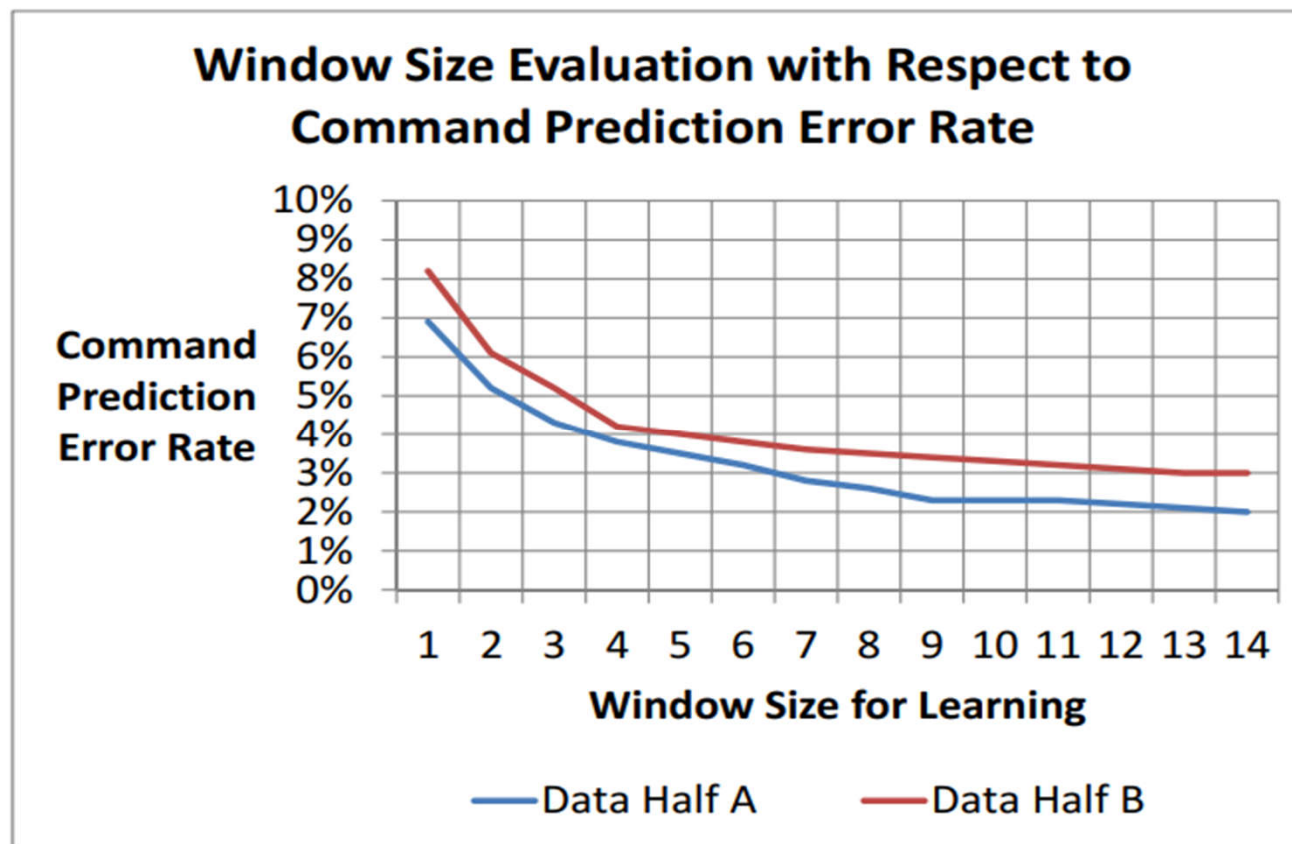
Applying Machine Learning

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



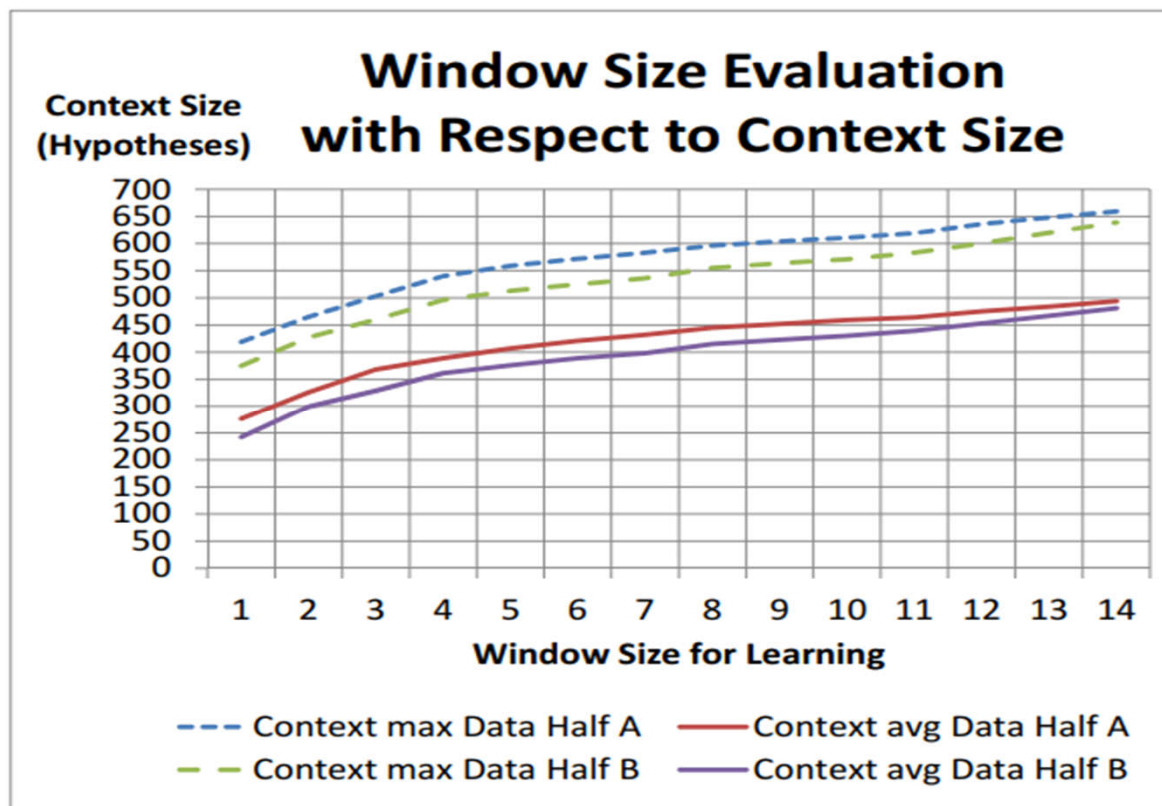
- Need for HUGE amount of data (first training, then testing)
- But only small amount in ATC ASR research

“Window” Size (CPER dependent)



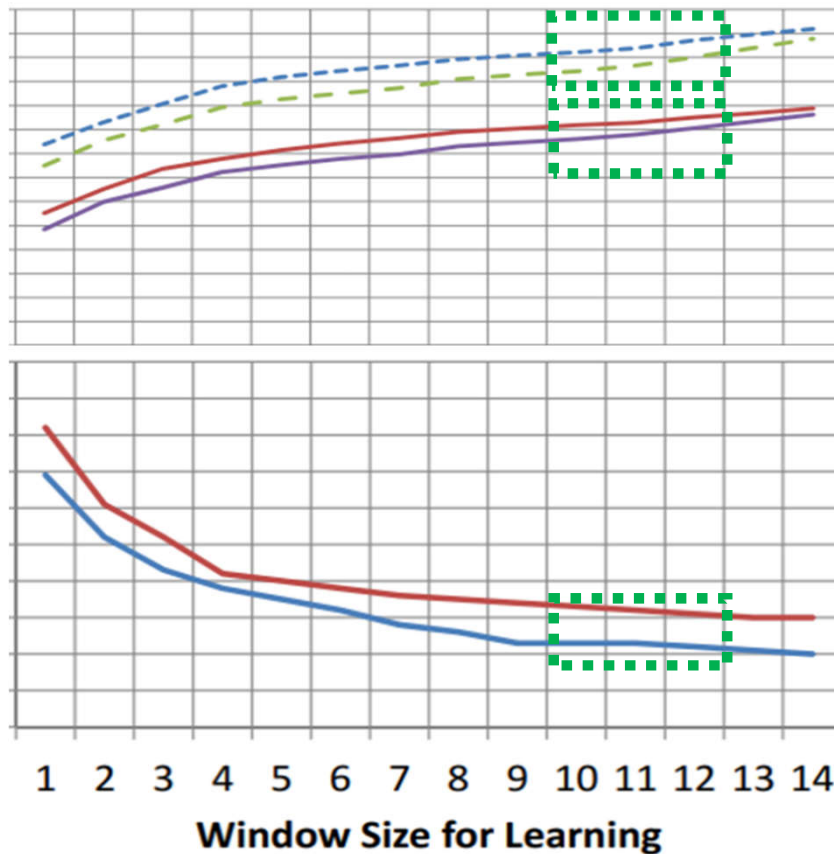
- Hungary-2017-11 data was split into two halves with 50% data amount each (A/B) for training and testing the other half

“Window” Size (Context Size dependent)



- max=biggest
- avg=average
- ...number of hypotheses in context
- Always lower than maximum possible number of hypotheses

“Window” Size 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

<i>Name of Evaluation</i>	<i>Training with Dataset</i>	<i>Test with Dataset</i>
HUNGARY	Hungary-2017-11	Hungary-2018-11
LITHUANIA	Lithuania-2018-03	Lithuania-2018-12

- What you learn, is what you expect!
- Runway change in test data not equal to training data



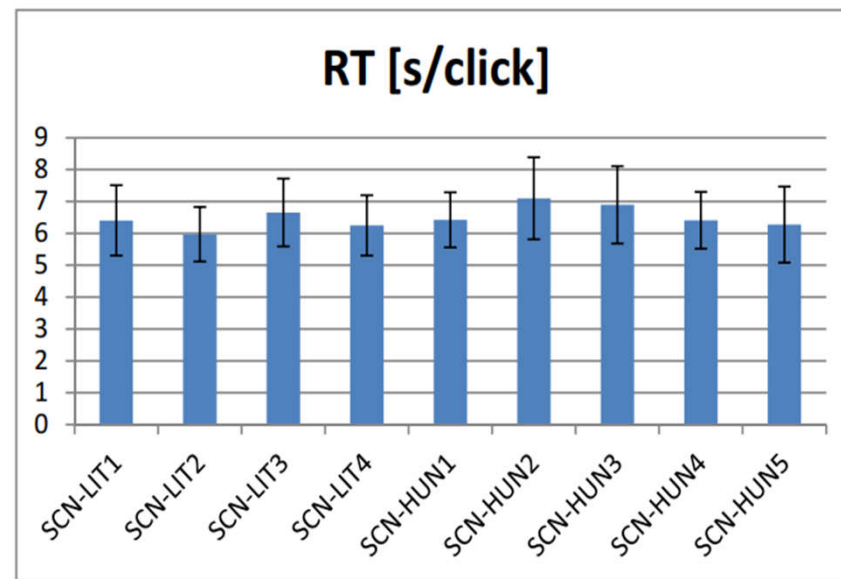
Results Overview: CPER and Context Size

<i>Name of Evaluation</i>	<i>CPER</i>	<i>SD</i>	<i>ctx_avg</i>	<i>ctx_max</i>
HUNGARY	<u>7.8%</u>	2.57%	450	593
LITHUANIA	12.5%	3.6%	340	498

- CPER: command *prediction* error rate
- SD: standard deviation

Results: Context Generation

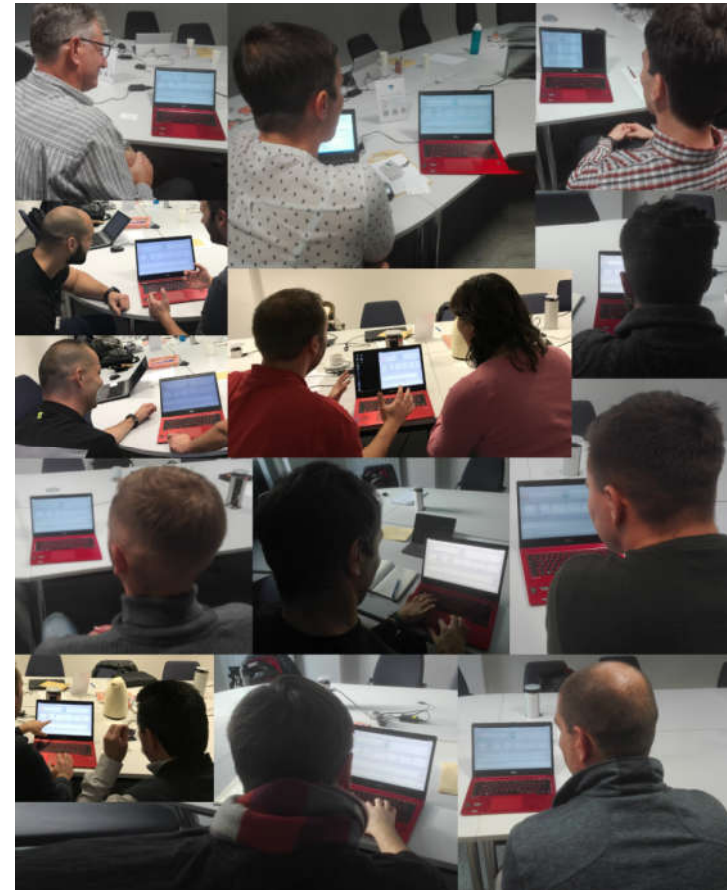
- 21'000 times in 59 simulation runs of final trials
- Each 119 ms (in average)
- Requirement: Faster than 5 s (due to radar update rate, duration of utterances, and expected situation change)



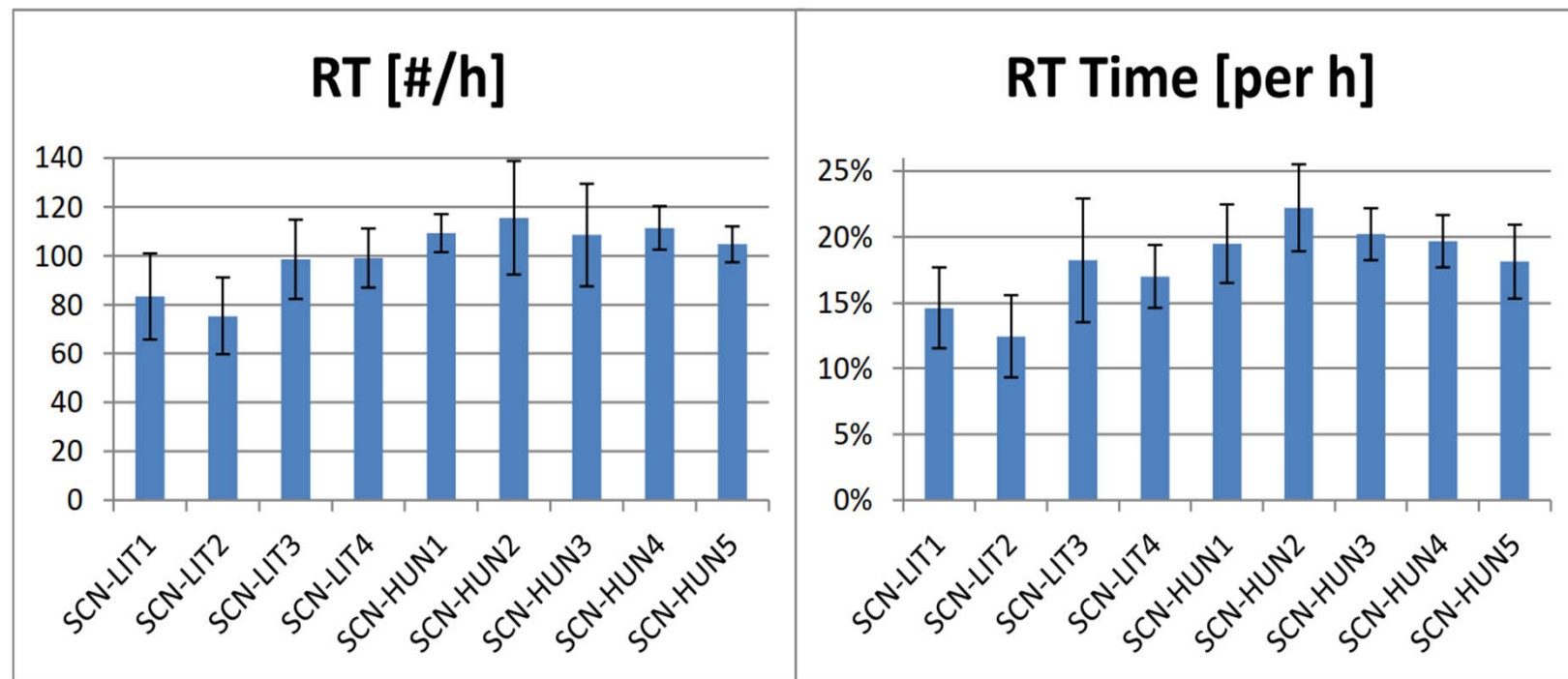
- RT: Radio telephony

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

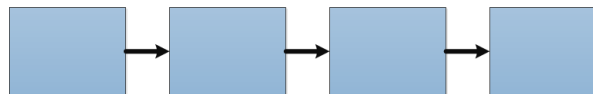
<i>Controller Command Type</i>	<i>Best Window Size</i>
INFORMATION	9
CLEARED	8
INIT_RESPONSE	11
REPORT	12
TAXI	11
CLIMB	2
CONTACT_FREQUENCY	7
VACATE	9
PUSHBACK	1

- Integration of enhanced TCHG with speech recognizer to generate first ABSR system for tower environment



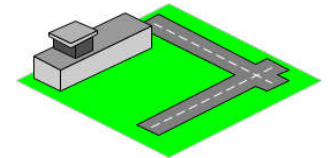
Outlook on Future Work (2)

- 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



founding members



Thank you



CWP HMI PJ.16-04-02 ASR-EXE-240

“?”



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.