Adaptation of Assistant Based Speech Recognition to New Domains and its Acceptance by Air Traffic Controllers

Matthias Kleinert[1], Hartmut Helmke[1], Gerald Siol[1], Heiko Ehr[1], Dietrich Klakow[2], Mittul Singh[2], Petr Motlicek[3], Christian Kern[4], Aneta Cerna[5], Petr Hlousek[5]

[1] Institute of Flight Guidance, German Aerospace Center (DLR), Braunschweig, Germany
[2] Spoken Language Systems Group (LSV), Saarland University, Saarbrücken, Germany
[3] Idiap Research Institute, Martigny, Switzerland, [4] Austro Control, Vienna, Austria,
[5] Air Navigation Services of the Czech Republic, Jenec, Czech Republic
firstname.lastname@{dlr.de, lsv.uni-saarland.de, idiap.ch, austrocontrol.at, ans.cz}

Abstract. In air traffic control rooms, paper flight strips are more and more replaced by digital solutions. The digital systems, however, increase the workload for air traffic controllers: For instance, each voice-command must be manually inserted into the system by the controller. Recently the AcListant® project has validated that Assistant Based Speech Recognition (ABSR) can replace the manual inputs by automatically recognized voice commands. Adaptation of ABSR to different environments, however, has shown to be expensive. The Horizon 2020 funded project MALORCA (MAchine Learning Of Speech Recognition Models for Controller Assistance), proposed a more effective adaptation solution integrating a machine learning framework. As a first showcase, ABSR was automatically adapted with radar data and voice recordings for Prague and Vienna. The system reaches command recognition error rates of 0.6% (Prague) resp. 3.2% (Vienna). This paper describes the feedback trials with controllers from Vienna and Prague.

Keywords: Machine Learning ·Assistant Based Speech Recognition · Automatic Speech Recognition · Air Traffic Controller

1 Introduction

Air traffic control (ATC) is a conservative business: Air traffic controller (ATCo) still use paper flight strips in control rooms around the globe. These strips contain vital information about aircraft that are under control of an ATCo in a specified airspace sector (e.g. callsign, type and weight class). For guiding an aircraft through the sector, an ATCo gives instructions (commands) to a pilot via voice communication. These commands are mainly related to speed, direction or flight altitude. Notes about the instructed commands, written on paper flight strips, are only available for the controller herself/himself. If provided in a digital form this information could be valuable for other systems e.g. to provide additional safety functions or for planning purposes. Therefore, many Air Navigation Service Providers (ANSP) already replaced or prepare to replace the traditional paper flight strips with different electronic/digital solutions. These systems, however, require manual digital input and, therefore, increase the controllers' workload.

Recently the AcListant® [1] project has validated that Assistant Based Speech Recognition (ABSR) can offer a solution to reduce manual controller inputs by integrating an assistant system with a speech recognizer [2]. The system analyzes the controller-pilot-communication, automatically extracts the instructed commands and uses them as input for digital flight strip solutions. Since the controller only needs to correct the system in case of a wrongly recognized command, the ATCo has more free cognitive resources for other tasks. For AcListant® ABSR was adapted manually to fit the needs of the Dusseldorf approach area in Germany. In simulation runs with different ATCos the system achieved recognition rates better than 95% and error rates below 2% [3]. However, the whole adaptation process required significant data resources, time and expert knowledge and needs to be repeated every time ABSR will be used at another airport or another ATCo sector.

The Horizon 2020 funded project MALORCA (MAchine Learning Of Speech Recognition Models for Controller Assistance), proposed a cheap and effective solution through a Machine Learning (ML) framework, that takes advantage of the large amounts of radar and voice data being recorded in air traffic control rooms on a daily basis. The developed framework is capable of adapting a generic ABSR system to different airports and controller positions [4]. In order to enable an efficient adaptation process ABSR was divided into several conceptual modules that consist of different building blocks, models and data elements. Only the models need to be adapted to new environments. Prague and Vienna approach area were chosen as first showcases for model adaptation.

In the next section we present related work with respect to speech recognition applications in ATM. Section 3 describes the essential building blocks of an ABSR system, Feedback from the end-users (ATCos) on the completely trained ABSR system was collected during Proof-of-Concept trials in Prague and Vienna [4], [5] and is presented before the conclusions in section 4.

2 Related Work

Large improvements in automatic speech recognition have been reached recently, i.e. also due to the industrial applications such as Google Home or Amazon Echo. A good introduction into the state-of the art of ASR applications in the ATM domain until 2014 is given by Nyuyen and Holone [6]. Recent work has also employed ASR in real applications for pilot read back error detection [7]. Very recently, a speech recognition challenge was released by Airbus to develop ASR for an air traffic control scenario [8], allowing large variety of academy and industry to gain an access to real (i.e. manually transcribed) data and develop new machine learning algorithms in this domain. These proposed algorithms mainly focused on building robust acoustic and language models, combined with some graph (i.e. FST) based algorithms to efficiently deal with command deviations (i.e. especially in case of call-sign detection).

Assistant Based Speech Recognition integrates the information of an assistant system as context information. DLR and Saarland University first used an arrival manager, which analyses the current situation of the airspace to predict possible future commands of the ATCo [9]. This approach has shown to both significantly increase command recognition rate and reduce command recognition error rate in AcListant® [2][3] and MALORCA [4] project.

3 Conceptual Modules of Assistant Based Speech Recognition

The ABSR systems consists of four main modules, namely, DATA, TEXT, COMMAND and USER module. These modules interact with each other as shown in Fig. 1, in an ATC environment.

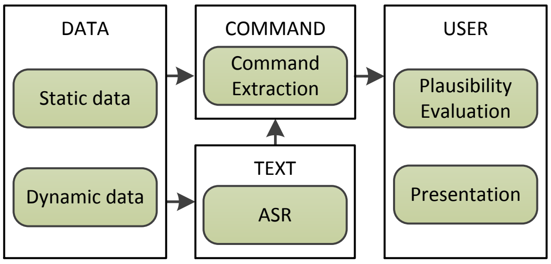

Fig. 1. Conceptual overview of ABSR modules

The DATA module satisfies the airport-specific data requirements of the ABSR system. Static data represents information, which changes rarely like names of way-points and runways, used frequency values etc. Dynamic data represents information, which is time- or space-dependent information like the ATCos' voice signals, radar data, flight plan information, weather information etc. The TEXT module employs this information to perform Automatic Speech Recognition (ASR) on the ATCos' voice signals. The ASR system converts the voice signal to a sequence of words and in process generating many text hypotheses for the same voice signal.

The COMMAND conceptual module uses information from DATA module and extracts the relevant information from the TEXT module's hypotheses to transform them into ATC command hypotheses. Different hypotheses for a voice signal are generated and hence, the USER module selects a unique output, which is adequately presented to the controller. The USER module utilizes the plausibility values and command hypotheses to perform this task. Further details of these modules, especial-ly, its adaptation to different approach areas are described in [5]. The building blocks are setup as generic part of the system. Data elements and models have to be defined resp. automatically adapted by ML to a specific environment.

4 Proof-of-Concept

The ABSR system was presented to ATCos from Prague and Vienna during the Operational Proof-of-Concept trials of the MALORCA project. The ATCos evaluated the system based on different tasks.

4.1 Operational Trial O1

The operational part O1 for the proof of concept trials performed pairwise comparison of the output of two different ABSR systems. Both systems had to execute command recognitions on recorded voice and radar data sets from Prague and Vienna. Only voice recordings that resulted in different command recognitions for both ABSR systems were taken into account for further analysis. N voice recordings (35 for Vienna and 36 for Prague) were randomly selected from those pre-filtered results.

Experimental Setup. Two ABSR systems were trained: (1) Baseline ABSR, trained only with manually transcribed data and (2) ML(Machine Learning)-Improved ABSR, improved with all available both manually and automatically transcribed data. Each ATCo could replay a voice recording as often as necessary and select between four choices.

1. The output of the ML-improved ABSR system is the better one
2. The output of the Baseline ABSR system is the better one
3. Both outputs are equally good
4. Both outputs are equally bad

The controllers did not know for all presented examples which recognition output is generated by which ABSR system (ML-improved or baseline) and the order of the selectable answers was always randomly generated. The experiment was conducted with four ATCos from Prague and five from Vienna.

Results. The ATCos preferred the results of the ML-improved ABSR system (64%) against the baseline ABSR system (21%). In 15% of the cases they could not decide between one of them.

4.2 Operational Trial O2

The operational part O2 for the proof of concept trials is based on monitoring the output of the ABSR system on a radar screen replaying real radar and audio recordings from the air traffic control rooms of Prague and Vienna.

Experimental Setup. The basic idea of the O2 trials was to give the ATCos an impression on how ABSR would work in a real operation room. The ATCos sat in front of a radar screen. Radar data recorded previously was replayed in real time on the radar screen. Additionally audio recordings that correspond to the radar data were used as input for the ML-improved ABSR system. The voice recordings were injected into the ABSR system and real-time speech recognition was performed. The ATCo listened to the voice recordings, saw the corresponding radar data and the ML-improved ABSR system attempted to recognize the voice recordings in real time. The recognitions were displayed to the ATCo in the radar label. The ATCos task was to monitor the recognitions and to correct the recognitions whenever ABSR fails to deliver the right result. The ATCos could provide feedback both with respect to recognition rate and speed. The scenario lasted approx. 30 minutes including a 10 minutes training phase.

Results. The results of the O2 experiments with ATCos from Prague and Vienna are shown in Table 1. Column "corrected by ATCo" indicates that the ATCos did not correct all of the ABSR errors. This happened because the ATCos were not familiar enough with the HMI that was used during the experiments, but all errors were detected by the ATCos and reported to the staff supervising the experiment.

Table 1. Results of operational Proof-of-Concept O2

Airspace	ATCo	Number of Commands	Number of ABSR Errors	corrected by ATCo	detected by ATCo
Prague	C1	99	9	9	9
	C2	99	9	6	9
	C3	99	9	7	9
	C4	99	9	9	9
Vienna	C1	122	16	16	16
	C2	122	16	16	16
	C3	122	16	16	16
	C4	122	16	15	16
	C5	122	16	16	16
All	9	1006	116	110	116

4.3 Debriefing sessions with ATCos

After the operational Proof of Concept parts O1 (pairwise comparison baseline/ML-improved ABSR) and O2 (monitoring ABSR) a questionnaire was presented to the ATCos. They were asked to answer different questions by assigning a digital value between one and six to each of the questions. “1” means “totally disagree”, whereas”6” means “totally agree”.

Table 2. Mean values and standard deviations

	Prague		**Vienna**		**Both**	
	µ	σ	µ	σ	µ	σ
I could imagine to work with Speech Recognition support for radar label maintenance	5.8	0.5	5.0	1.0	5.3	0.9
I understood the application of Speech Recognition Support for Radar Label Maintenance	6.0	0.0	5.4	0.5	5.7	0.5
Today’s support of Speech Recognition was adequate for the presented scenario.	4.8	1.3	5.2	0.4	5.0	0.9
Speech Recognition Support of MALORCA system would be (already) helpful for my workplace.	4.0	1.4	4.4	1.1	4.2	1.2
The application of the MALORCA system would provide an improvement for my work.	4.8	1.0	5.2	0.8	5.0	0.9
The number of command corrections was proper with respect to the scenario (traffic density…)	4.3	1.0	5.0	0.7	4.7	0.9
It was easy to do a corrective action and I was able to maintain situational awareness.	5.5	1.0	5.8	0.4	5.7	0.7
ASR will cause safety problems	2.0	0.8	2.5	0.6	2.3	0.7

5 Conclusions

The MALORCA project validated for different controller positions of Prague and Vienna approach area that the automatic adaptation of ABSR is possible by using machine learning technologies. This paper presented the results of the operational Proof of Concept trials from the MALORCA project with ATCos from Prague and Vienna. It showed that ATCos prefer the output of a machine learning adapted ABSR system compared to a basic generic ABSR system. Furthermore the trials with nine different ATCos monitoring recorded real life air traffic from Prague and Vienna showed that recognition rates of machine learning trained system are high enough to reduce controller workload. Even though the ABSR system mostly recognized (88.5%) the given commands, the situation awareness of the ATCos were not negatively affected. They were able to detect all misrecognitions.

The feedback from ATCos after trials showed that they were satisfied with the performance of ABSR and that the system would help in their daily work if available in their operational environment. Currently no system supplier for ANSPs can provide such a technology. The next step must be the integration of ABSR into operational environment.

References

[1] The project AcListant® (Active Listening Assistant) http://www.aclistant.de

[2] Helmke, H., Ohneiser, O., Mühlhausen, Th., Wies, M.: Reducing controller workload with automatic speech recognition. In: IEEE/AIAA 35th Digital Avionics Systems Conference (DASC). Sacramento, California, 2016

[3] Helmke, H., Ohneiser, O., Buxbaum, J., Kern, C.: Increasing ATM efficiency with assistant-based speech recognition. In: 12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017). Seattle, Washington, 2017

[4] Kleinert, M., Helmke, H., Siol, G., Ehr, H., Cerna, A., Kern, C., Klakow, D., Motlicek, P. et al. Semi-supervised Adaptation of Assistant Based Speech Recognition Models for different Approach Areas. In: IEEE/AIAA 37th Digital Avionics Systems Conference (DASC). London, England, 2018

[5] Kleinert, M., Helmke, H., Ehr. H., Kern, C., Klakow, D., Motlicek, P., Singh, M., Siol, G.: Building Blocks of Assistant Based Speech Recognition for Air Traffic Management Applications. 8th SESAR Innovation Days, Salzburg, 2018, to be published

[6] Nguyen, V. N., Holone, H.: Possibilities, challenges and the state of the art of automatic speech recognition in Air Traffic Control. In: International Journal of Computer and Information Engineering, Vol 9, No. 8, 2015, pp.1940-1949

[7] Chen, S., Kopald, H.D. , Chong, R., Wei, Y., Levonian, Z: Read back error detection using automatic speech recognition. In:12th USA/ Europe Air Traffic Management Research and Development Seminar (ATM2017), Seattle, WA, USA, 2017

[8] 2018 AIRBUS Air Traffic Control Challenge Workshop: https://www.irit.fr/recherches/SAMOVA/pagechallenge-airbus-atc-workshop.html

[9] Shore, T., Faubel, F., Helmke, H., Klakow, D.: Knowledge-based word lattice rescoring in a dynamic context. In: Interspeech 2012, Sep. 2012, Portland, Oregon