AcListant with Continuous Learning: Speech Recognition in Air Traffic Control

J. Rataj, H. Helmke, and O. Ohneiser

Abstract Increasing air traffic creates many challenges for air traffic management (ATM). A general answer to these challenges is to increase automation. However, communication between air traffic controllers (ATCos) and pilots is still widely analog and far away from digital ATM components. As communication content is important for the ATM system, commands are still entered manually by ATCos to enable the ATM system to take the content of the communication into account. However, the disadvantage of this procedure is significant additional workload for the ATCos. To avoid this additional effort, automatic speech recognition (ASR) can automatically analyze the communication and extract the content of spoken commands. DLR together with Saarland University invented the AcListant® system, the first assistant based speech recognition (ABSR) with both a high command recognition rate and a low command recognition error rate. Beside the high recognition performance, AcListant® project revealed shortcomings with respect to costly adaptations of the speech recognizer to different air traffic control (ATC) environments. Machine learning algorithms for the automatic adaptation of ABSR to different airports were developed to counteract this disadvantage within the MALORCA project, funded by Single European Sky ATM Research Programme 2020 Exploratory Research (SESAR-ER). To support the standardization of speech recognition in ATM, an ontology for ATC command recognition on semantic level was developed to enable the reuse of expensively manually transcribed ATC communication in the SESAR Industrial Research project PJ.16-04. Finally, results and experiences are used in two further SESAR Wave-2 projects. For the first time, this paper presents the evolution from the idea of ABSR born in an academic environment, starting with the project AcListant®, to industrialization ready research prototype of technology reediness level (TRL) 4. In this course, relevant industrial needs such as costs and necessary

J. Rataj (✉) · H. Helmke · O. Ohneiser
Institute of Flight Guidance, German Aerospace Center (DLR), Braunschweig, Germany
e-mail: juergen.rataj@dlr.de

H. Helmke
e-mail: hartmut.helmke@dlr.de

O. Ohneiser
e-mail: oliver.ohneiser@DLR.de

Air Traffic Management and Systems IV, Lecture Notes in Electrical Engineering 731,
https://doi.org/10.1007/978-981-33-4669-7_6

standardizations supported by tailored European funding scheme are considered. The addressed SESAR projects are MALORCA, PJ.16-04, PJ.10-96 HMI Interaction modes for ATC centre, and PJ.05-97 HMI Interaction modes for Airport Tower.

Keywords Assistant based speech recognition · Machine learning · AcListant® · MALORCA · PJ.16-04 · Ontology

1 Introduction

The increasing air traffic creates many challenges concerning safety, capacity, efficiency, and environmental performance for ATM. Additionally, economic pressure exists to increase productivity in ATC to keep flying affordable. The general answer of the main ATM development programs, such as SESAR (Single European Sky ATM Research) [1] in Europe, NextGen (Next Generation Air Transportation System) in US [2], CARATS (Collaborative Actions for Renovation of Air Traffic System) in Japan [3] or CAAMS (Civil Aviation ATM Modernization Strategy) in China [4] to fulfill these challenges is to increase digitization and automation considerably. In this case, digitization means to transform analog data into digital formats, which, in turn, is the basis for modern automation solutions. Already today, a high degree of digitization exists in ATM. Radar trackers, flight data processing systems (FDPS) as well as other systems represent the real world in digital environment. However, one central element of ATC, the communication between ATCos and pilots, is not digitized yet. The communication still relies on analog radio, which—independent of CPDLC (Controller Pilot Data Link Communications) [5–7]—will exist during the next decade or even longer.

The content of this communication is of utmost importance for the digital ATM systems world. Hence, the spoken commands of the ATCos must be digitized to be available in the digital world. Today, this is manually performed by ATCos via mouse or keyboard in parallel to their voice communication with the pilots. In this way, the digital world understands the impact of human communication on a certain traffic situation. As advantage of digitization, the controller can benefit from decision and negotiation support systems. However, a huge disadvantage is the significant effort for the ATCo concerning additional manual inputs into the digital system. Hence, the question arises whether the advantages by support systems, such as an arrival manager outweigh the disadvantages of additional controller workload.

An approach to avoid the above mentioned disadvantages is to use automatic speech recognition (ASR). ASR enables to automatically extract the content of uttered commands and digitize them for ATC systems without additional ATCo's workload. Therefore, such a technology seems to be very beneficial for further digitization of ATC and will increase automation. Additionally, speech recognition technology gathered a high interest based on popular consumer applications, such as "Siri" or "Alexa". Based on such applications and the large market behind, it can be assumed that the technology will develop rapidly and can be adapted to ATC with moderate effort.

This article describes for the first time the entire development process of assistant based speech recognition (ABSR) in the academic environment and moving towards an industrializable prototype as well as first developments of standards in this context. Furthermore this work presents the references to our original work describing the algorithms, validation trials and the results. The special challenges presented in Sect. 2, which are posed especially to speech recognition in operational air traffic control environment, are the starting point for the novel approach to speech recognition are. In addition, Sect. 2 outlines the overall development process of the ABSR in air traffic control with the associated work in various projects. In Sect. 3, the paper discusses the novel approach utilizing predictions of ATCo behavior to improve speech recognition. An innovative problem solution in the academic environment is not always sufficient for the industrialization. For this reason, further research activities accompanying the industrialization to reduce implementation costs were necessary in order to utilize speech recognition in an operational environment. This is the subject of Sect. 4. The basic approach at this point was machine learning with the intervention that training of acoustic model, language model, and command prediction model iteratively enhance each other. Finally, Sect. 5 describes the standardization efforts required for industrialization, which took place in an industrial environment. This development was enabled by partially coordinated funding instruments mentioned in the article. The resulting projects were AcListant® based on the Helmholtz Validation Fund, MALORCA based on SESAR Exploratory Research and PJ.16-04 based on SESAR Industrial Research. Section 6 closes and gives an outlook.

2 Evolution of ASR in ATC

Based on the literature, speech recognition for ATC was used in some places with medium success [8, 9]. First attempts to use standard speech recognition for the controller working position in our labs led to disappointing results concerning the recognition rate. Tests in the DLR research simulator ATMOS (Air Traffic Management and Operation Simulator) with standard ASR systems—adapted to ATC environment—resulted in recognition rates from 65 to 85% per controller command. Such recognition rates will not be accepted by ATCos in an operational environment. It is known from other projects concerning Arrival or Departure Management that ATCos put very high demands on the capabilities of their support systems. If the system could not fulfill the expected abilities, the system will be rejected by the ATCos. Then, it is very difficult to get a second chance to introduce this new technology into ATC.

In order to avoid a rejection by the ATCos, it was assumed that high recognition rates are necessary not knowing exactly what *high* means. Based on this reasoning the first insight to solve the problem was to define a new assessment metric because the metric, word error rate (WER), to evaluate the performance of an ASR system as used in the speech recognition domain [10], is not the deciding value for the ATCo's acceptance of the resulting system. More important is the correctness of the recognition on command level and not of single words. This insight creates a further considerable challenge for ASR, to deliver a high command recognition rate (CRR). The CRR hereby is defined as the percentage of correctly recognized commands divided by all given commands. An ATC command itself consists of several elements (e.g., callsign, command type, and command value) each consisting of several words, hence to achieve a low command recognition error rate (CER) is much more challenging than just a low WER. Details on CRR and CER calculation can be found in [11].

In discussions with ATCos from several European countries within the framework of the SESAR 2020 Industrial Research project PJ.16-04 CWP HMI (Controller Working Position Human Machine Interface), the requirements for ASR applications in ATM context were specified. The most important one is a low CER. Based on statements of ATCos, the CER is especially important, because it causes additional workload to detect an error. Hence, the ATCos prefer to manually input unrecognized commands instead of detecting wrongly recognized commands with additional manual correction effort. The decisive requirement follows from this that the CER of an ASR system should be exceptionally low. On the other hand, an acceptable high CRR is also indispensable.

To achieve both, high CRRs and low CERs, DLR together with Saarland University invented the AcListant® system [11], which will be detailed in the next section. This system bases on a specific context, which is gained using the knowledge of a controller assistant system. Hence, AcListant® (Active Listening Assistant) is denoted as Assistant Based Speech Recognition (ABSR) system, which creates a new class of speech recognition systems. AcListant® validation trials have demonstrated that both, high CRRs (>90%) and low CERs (<3%), are possible. Additionally, it was shown that controller assistant systems, e.g. Arrival Managers, benefit from the knowledge of the content of the communication between controller and pilot [12, 13].

The follow-up project AcListant®-Strips, led by DLR, successfully validates the hypothesis that ABSR reduces ATCo's workload for radar label maintenance. Beyond that, the reduced workload results in an increased controller performance. In Düsseldorf approach scenarios of the validation trials carried out with German and Austrian ATCos, the average flight time in the Terminal Maneuvering Area (TMA) was reduced by 77 s per aircraft and a reduced average flight length of 5 nautical miles was shown [14, 15].

The project also revealed an important shortcoming: The expensive adaptations of ABSR to different environments and user groups with respect to airspace, airports, dialects, local phraseology etc. After achieving the requirements of high CRR and low CER, reducing adaptation costs was the next challenge, which needed to be fulfilled for an industrialization of the research results. Hence, the next development step was driven by the question on how to reduce the costs for deployment and maintenance of an ABSR system. The considerations concerning cost reductions resulted in the idea for the SESAR 2020 Exploratory Research project MALORCA (Machine Learning of Speech Recognition Models for Controller Assistance), which was led by DLR [16] The goal of this project was to substitute the expensive manual adaptation work of AcListant® by automatic procedures. In MALORCA a first set of mechanisms based on machine learning were developed by the project partners (Saarland University, Idiap, Austro Control, Air Navigation Service Provider of Czech Republic (ANS CR) and DLR) to enable an automatic adaption of AcListant® to a certain environment. These mechanisms were exemplarily applied to the approach areas of Vienna and Prague using recorded real controller communications. The resulting CER after learning for Vienna approach was 3.5%. For Prague a CER of 0.6% was achieved [17, 18].

In parallel to the work in SESAR Exploratory Research, activities to foster speech recognition in an industrial environment were performed in SESAR2020 Wave 1 with the ASR Activity in the Industrial Research solution PJ.16-04 CWP HMI led by DLR. The goal of this project was to increase the ATCos' productivity and to support the industrialization of speech recognition in ATC. The process of transforming an audio signal to a sequence of words is called transcription, i.e. the voice to text process. The transformation of the word sequence to the relevant ATC concepts is called annotation. MALORCA has shown that different experts agree on the transcription of a controller utterance, but their annotation results may be different. This creates a problem concerning automatic understanding of controllers' voice. Therefore, a set of rules for annotating a sequence of words to ATC concepts was developed, i.e. an ontology. This ontology was agreed in SESAR project 16-04 by 15 European ATC partners setting the basis for a standard in this field [19]. After having presented the evolution from AcListant® to an agreed ontology for command annotation the projects AcListant®, AcListant®-Strips, MALORCA and PJ.16–04 are presented in more detail.

3 AcListant® and AcListant®-Strips

Currently, ASR in ATC is only used in training, i.e. to replace pseudo-pilots. It is reasonable for training purposes to let an ATCo repeat utterances due to undesired deviations regarding the standard phraseology with resulting incorrect speech recognition. Furthermore, training situations are not as critical as real life situations, hence performance limits of ASR for training are acceptable, but not in an operational environment.

To enable the digital ATC world to understand the communication between ATCos and pilots is very beneficial, even more, if this requires no additional workload for the ATCo, which is possible by using speech recognition. As mentioned above to use ASR in operation, ATC specific requirements have to be taken into account, such as high CRRs and low CERs to be successful. In order to be successful, the gold standard in the ASR community, the WER, for assessment and evaluation has to be extended, because this standard is not descriptive enough as metric for ATC. The specific ATC requirement is to know if the content of a controller command, the concept, is recognized. For example, in the utterance “good morning lufthansa one two three descend flight level one two zero” the meaningful concept from an ATC perspective is “DLH123 DESCEND 120 FL”. Hence, to recognize “good morning” is not necessary, because it contains no relevant information and thus misrecognition is irrelevant. Taking this into account the new metrics CER and CRR [11] were defined for ATC applications.

The work concerning ASR started at DLR and Saarland University with a standard ASR with an acoustic model adapted to real ATCo-pilot communication. Many hours of speech samples were recorded, transcribed word-by-word, and annotated with the included semantic content afterwards. Although already considerable effort was spent it was decided to stop this approach of improving just a standard ASR engine. A radical new approach was necessary.

The new approach—patented and developed by DLR and Saarland University—bases on the intensive use of situational context to improve performance. ASR systems, which use current context are known, but not those that take a prediction of the situation into account. Possible sources for predictive context are controller assistant systems. These systems, such as an arrival manager (AMAN), predict the course of future situations to support the ATCo in planning his next actions. This prediction is considerably dynamic based on changing situation elements.

Using an assistant system, see Fig. 1, results in the new ABSR concept. For ABSR, the DLR AMAN 4D-CARMA (4 Dimensional Cooperative Arrival Manager) was used to provide the current and predicted situation of relevant air traffic. This comprises static and dynamic knowledge of the traffic and airspace situation handled by the 4D-CARMA—Core Components. Static knowledge considers e.g., airspace

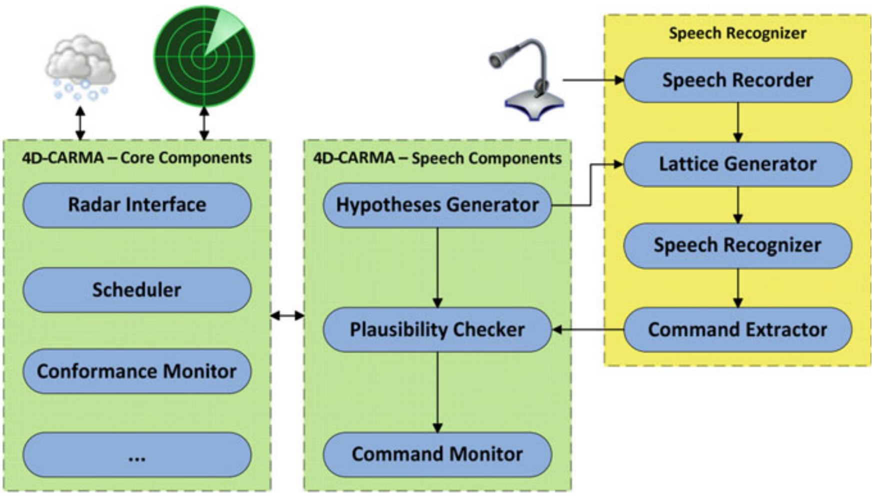

Fig. 1 Components of assistant based speech recognition [11]

structure with routes and waypoints, airspace sector frequencies, minimum separation, etc. Dynamic knowledge bases on the aircraft state vectors and flight phases as well as relevant planning from AMAN modules such as aircraft sequences or distances-to-go. Hence, commands should, e.g., only contain aircraft callsigns that are currently flying in the relevant airspace. Furthermore, knowing an aircraft is in its landing phase, descend and reduce commands are more probable than climb and increase commands. With the knowledge of the airspace structure, also reasonable heading values can be forecasted, because ATCo mostly follow certain routes or direct to certain waypoints.

This above described context knowledge of the assistant system is used by the 4D-CARMA Speech Components, starting with the "Hypotheses Generator" component. The "Hypotheses Generator" does not know exactly which commands the controller will give in the future, but it knows which commands have a higher probability than others in the current and future situation.

These hypotheses are used as input for the "Speech Recognition" block, which consists of the components: "Speech Recorder", "Lattice Generator", "Speech Recognizer", and "Command Extractor". A microphone is connected to "Speech Recorder" to record the signal as wave file. The "Lattice Generator" creates a search space for the "Speech Recognizer" using the output of the "Hypotheses Generator". Hypotheses are of good quality, if they are correct and if just a few commands are forecasted instead of everything that is possible in theory. Hence, the lower the number of hypotheses, the smaller the search space for the speech recognizer. The extracted commands are sent back to the "Plausibility Checker" component, which uses context knowledge and command hypotheses to reject recognized commands. The "Plausibility Checker" divides the recognized commands into three sets:

- Commands immediately accepted, i.e. recognized commands being predicted and also being plausible.
- Commands further monitored with respect to radar data, i.e. recognized commands which are either predicted or have high plausibility values.
- Commands immediately rejected, i.e. recognized commands which are not predicted and with low plausibility values.

The "Command Monitor" verifies commands monitored by continuous comparison to radar data. If, e.g. a descend command to flight level 90 was recognized and the aircraft did not descend after a predefined time, the command is transferred to the set of "commands rejected".

The validation of the ABSR system was performed in two related projects AcListant® and AcListant®-Strips. In AcListant® the recognized speech was used to support an AMAN as well as the ATCo by avoiding manual inputs to maintain the system. The flight information itself was documented on strips in electronic or paper form or on the radar screen in the aircraft label, depending on the simulation run. The information comprises of, e.g., callsign, destination, or route information, clearances regarding altitude, speed, direction, or procedures, as well as special flight situations like emergencies.

In AcListant® two dimensions of validation questions were addressed (see Fig. 2).

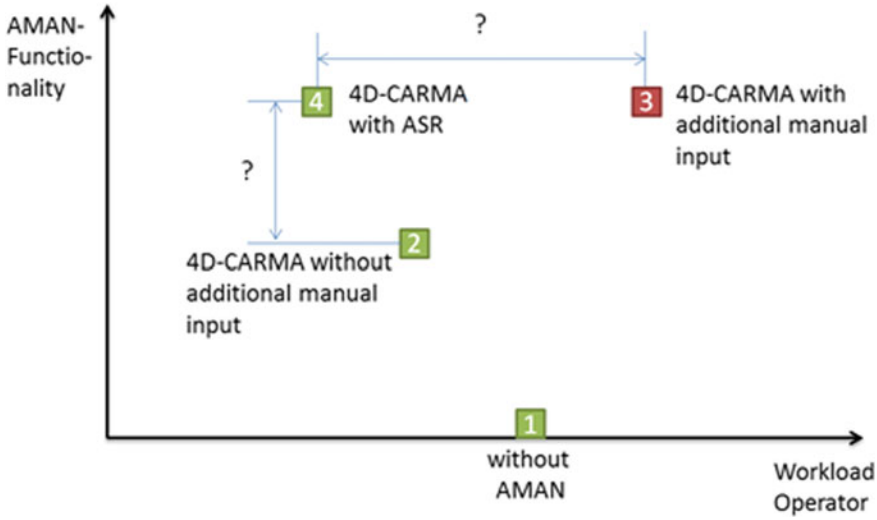

Fig. 2 AMAN functionality versus workload diagram

The first dimension concerns the functionality benefits for the AMAN depending on the input. The second addresses the workload of the ATCo depending on the kind of input. The difference between square 2 and 4 is the additional input for the AMAN based on ATCos' communication. The difference between square 3 and 4 is the kind of input. To validate the increasing functionality, simulation runs with standard AMAN (square 2) and runs with an AMAN supported by ASR creating additional inputs (square 4), were conducted. To quantity the workload reduction, additional runs with an AMAN, either with manual input device (square 3) or with ASR (square 4), were performed. Square 1 illustrates the situation without any controller support.

In the baseline scenario, i.e. square 2 in Fig. 2, for AcListant® trials the flight information were handled with paper flight strips as usual at Düsseldorf approach. In a second scenario, i.e. square 3, the ATCo had to manually input the clearances by mouse and keyboard, which emulates the situation with an electronic flight strip system.

The third scenario, i.e. square 4, based on ABSR usage. In this scenario, the ABSR system listened to the communication between ATCo and pilots. After the speech recognition, the ATCo had the possibility to confirm, correct, or reject the output of the recognizer. Two special test scenarios were chosen to be able to quantify the functionality benefits of a listening AMAN. The first one addressed an emergency situation caused by a sick person on board, the second one a runway closure. In these cases, an early re-planning of the AMAN was necessary to support the ATCo. The re-planning can be triggered by observing the radar data, by manual input of the commands or by speech recognition. Observing the radar data results in delayed system reaction and manual input results in additional ATCo workload. Using ABSR solves both issues by automatic and fast system input.

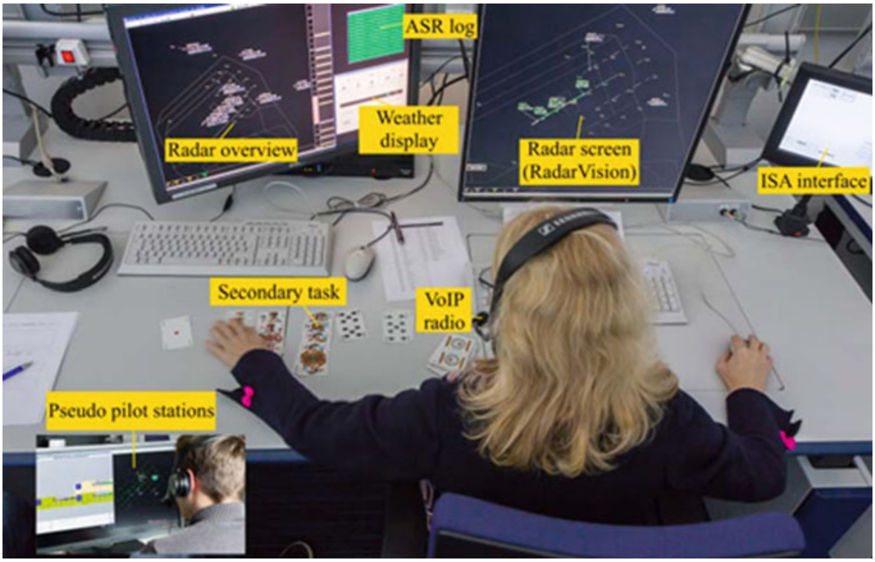

Fig. 3 Basic validation setup during final trials

The set-up for the validation trials consisted of a controller working position (CWP), a traffic simulation and two pseudo-pilot stations. The CWP comprised of radar screen, weather display, radar overview, speech log screen, mouse and keyboard (see Fig. 3). To measure the workload, an instantaneous self-assessment (ISA) test was used. For trials concerning speech recognition, it was necessary to involve different kind of voices. Hence, the participating controllers were selected in a way that there were male and female participants as well as speakers from different countries to take different accents into account.

In AcListant®, it was shown that CRRs of more than 95% are possible using an AMAN to reduce the search space of the speech recognizer. However, the CER was still above 7% and without "assistant based" nearly 20%, which is assumed to be not acceptable. Using also the knowledge from the assistant system to reject commands, i.e. the "Checker" component, the CER was reduced below 2.5%.

The prize for the checker is a decreased recognition rate from 95 to 91%, because correct recognitions were rejected also. The results in Table 1 are based on approx. 4,000 controller commands given in 23 simulation runs. The sum of CER and CRR can be above 100% due to the Levenshtein distance definition [10]. This distance is

Table 1 Command recognition and command recognition error rates

	Recognition rate (%)	Error rate (%)
ASR without AMAN	84.0	19.7
ABSR/AMAN	95.8	7.4
ABSR with checker	91.0	2.5

Table 2 Non-conformance of planned and flown trajectories when comparing different AMAN support levels

Support Condition	Baseline	AMAN	AMAN + ABSR
Average of 3 ATCos based on 69 aircraft (%)	18.7	19.9	8.5

defined here as the minimum number of deletions, substitutions, and insertions to transform one sequence of commands into another one. Hence, if only one command is really said, but three are accidently recognized, we have at least two insertions, which results for this example in a CER of at least 200%.

Furthermore, it was shown that speech recognition improves the adaptation speed of an AMAN on changes in the airspace situation. In the baseline, the AMAN output was not visible to the ATCo. Nevertheless, the AMAN runs in background generating trajectories which are compared with the ones resulting from the ATCo's commands. Table 2 shows the percentages of non-conformance of those trajectories.

Column "AMAN" shows the non-conformance if the AMAN supports the ATCo, but the AMAN gets no input from the speech recognizer. The column "AMAN + ABSR" shows the results, when the AMAN could rely on ABSR. In the case of the visible AMAN, the non-conformance increases from 18.7 to 19.9%. It seems that ATCos tend to slightly deviate if they see AMAN recommendations. When AMAN is supported by ABSR, non-conformance rate is decreased by more than 50%, from 19.9 to 8.5%.

Table 2 clearly shows that the internal plan of the AMAN is more conform to the mental picture of the controller if the AMAN is able to listen to the ATCo. The main results of AcListant® trials [11] are:

- AMAN adapts much faster if the ATCo deliberately deviates from the planning of the assistant system.
- ABSR reduces significantly the deviation between the ATCo's and the assistant system's plan.
- ABSR is able to achieve acceptable CRRs (>90%) and CERs (<3%).
- ABSR significantly reduces ATCo's workload.

In AcListant®-Strips only the difference between the manual input of flight information and ABSR was taken into account. The goal was to quantify benefits of using ABSR as input mechanism to maintain the digital ATC systems. Therefore, the focus was on the workload of the ATCo and the work efficiency. Additionally to known workload measurement tools, we used a secondary task to be performed by the ATCo. The goal of the secondary task was to sort a deck of 48 cards into six decks for each playing card type (9–10-Jack-Queen-King-Ace) and name at the end one to four randomly missing cards. The test subjects were instructed to stay at the ATC task as long as the task requires it. The time needed to sort cards and finally identify the missing ones served as an objective value for user workload. Beside the hypothesis to reduce the workload, it was further assumed that the working efficiency increases based on avoiding head down times and more remaining time to guide the air traffic.

Hence, in one of the two validation scenarios, a very high traffic density was chosen. Eight controllers from Germany and Austria performed different test runs with and without ABSR support [14].

The following results were found in these trials: The ATCos were able to sort twice as many decks of cards as without ABSR support and maintained flight information more precisely. The ATCos invest 30% of their working time to input issued commands by mouse, if no ABSR support is available. This effort is used exclusively to enter known information into an electronic system without any effect on efficiency and quality of the work of the ATCos. Using ABSR changes this situation considerably. The results of the trials have shown that ATCos use only 10% of their working time to maintain flight information when being supported by ABSR. These 10% of working time include the time to check, confirm, and reject outputs of the speech recognizer.

The ATCos were very confident with command recognition rates and command recognition error rates, i.e. they appreciated the automatic aircraft radar label input. They even encouraged having a reaction time of a few seconds to visually check the recognized commands in their HMI instead of actively acknowledging each recognized label input. If ATCos did not intervene during this time, the ATC system should automatically accept the displayed recognition output. It was further found that manual ATC system input by ATCos via mouse and keyboard showed no better quality with respect to accuracy of command values and completeness of inputs.

The trials have also shown that a significant reduction of ATCo workload has an effect on throughput and ATCo's efficiency. One to two inbounds per hour for Düsseldorf are possible. Increased throughput and ATCo's efficiency are possible, because released cognitive resources can be used to better guide air traffic. For the Düsseldorf TMA a benefit of 77 s reduced flight time was quantified. This additional flight time is mostly on downwind. If we assume flying in flight level 70 with 250 knots of calibrated air speed, an A320 consumes 2700 L per hour resulting in roughly 50–65 L of reduced fuel consumption per aircraft. One liter of kerosene is 0.8 kg resulting in 3.15 kg of CO_2. Therefore, application of speech recognition can relieve the environment by about 130 kg CO_2 per flight [15].

4 Implementation Costs

Even impressive results concerning ATC performance indicators by automatic maintenance of flight information are not sufficient to avoid critical questions concerning costs. Speech recognition induces costs by procurement, introduction, and maintenance. Procurement costs base on market driven company decisions. Introduction costs occur, because an ABSR system has to be adapted to users and environments. Maintenance costs are driven by adapting the ABSR system if environment changes. According to changes in the user group, an adaption is only necessary if these changes are significant. The main cost driver for the adaption is the manual work performed by experts. Experiences in AcListant® have shown that adaptation and maintenance costs of about one million euros are reasonable adaptation cost for a midsize airport. To reduce such costs the manual work has to be automated.

Cost savings further allow a large number of midsize airports to use ABSR technology because it becomes affordable for them. If they use ABSR for flight information maintenance in the TMA, it is possible to save 130 kg CO_2 per flight to relieve environment, as outlined in Sect. 3. Collecting all additional airports, which are able to afford such a system in case of strongly reduced costs, will have a noticeable impact on the environment. Beside this benefit, additional benefits will occur, using ABSR in ATC, like the increased performance of controller assistant systems as shown above. Furthermore, the availability of transcribed and annotated controller commands can also be used for many off-line analyses.

To enable the work to achieve cost reductions, an extended team around the AcListant® partners gained funding from Horizon 2020 SESAR Exploratory Research for the, DLR coordinated project MALORCA. The goal of the project was to use machine learning algorithms to enable a generic, effective and especially cheap approach to adapt ABSR to a specific environment. A major step to achieve MALORCA goals was to separate environment and user dependent parts of the ABSR software from the independent ones. To achieve this, the ABSR system was disassembled into conceptual modules for the specific tasks INPUT, TEXT, COMMAND, and USER. The INPUT module supplies ABSR with voice signal input, surveillance data (e.g., radar data) and static airport dependent inputs (e.g., waypoints, frequencies). Based on data from INPUT, the TEXT module performs tasks related to the automatic speech recognition, i.e., transcription resulting in different sequences of words for one utterance. The COMMAND module translates sequences of words into controller commands using the output command prediction. Finally, the USER module provides the output of COMMAND to a user with an appropriate human machine interface or to another system. The conceptual modules consist of models, which are application (area) independent and models, which are application dependent. The models are automatically learned by machine learning algorithms.

The acoustic model is based on deep neural networks (DNN) and is automatically trained from transcribed and untranscribed data. If more than two hours of training data were available, speaker dependent acoustic models already outperform speaker independent models provided that the speaker is surely known. The lexicon, i.e. the word list and their pronunciation was manually updated by adding waypoints and some local words for greetings and good-bye. The language model consists of an N-gram statistical langue model and was trained by supervised learning.

For each command type (e.g., DESCEND, HEADING) a prediction area is created and subdivided into subareas of 1 nm by 1 nm. Additionally, a set of predefined rules to each command type is added, e.g. IF flight type is arrival AND controller working position is Feeder AND speed >220 knots. If the "Hypotheses Generator" detects that a lat/long position of an aircraft is inside an area of a specific command type and the rule condition for this area is true, the command values related to that flight and command type are predicted for that aircraft. For each command type the areas are learned by unsupervised learning, i.e. from automatically annotated commands.

The invention of the MALORCA project was that acoustic model, language model, and command prediction model iteratively enhance each other. The basic acoustic model results in automatic annotations of controller utterances. These annotations are used to train the command prediction model, which classifies the automatic annotations into good and bad training data elements. An automatic annotation, which is not predicted, is a bad training example. This classification is used in the next iteration to improve the acoustic model, which results in better annotations to improve the command prediction model etc. More details of model adaptation by machine learning are provided in [18] and [20].

After adapting all models the basic ABSR system could iteratively be improved with machine learning increasing the CRR from 80% to 92% for Prague, and from 60% to 83% for Vienna respectively. The 80% for Prague correspond to the case that no automatically transcribed data was available and the 92% include the usage of 18 h, i.e. 100%, of the automatically transcribed data set. The starting point of 60% CRR for Vienna data was on the one hand caused by worse audio quality and on the other hand by the higher variability of deviations from standard phraseology by Vienna ATCos. The CER could be reduced from 4.1 to 0.6% for Prague and from 10.9 to 3.2% for Vienna. For Vienna also 18 h of untranscribed and four hours of transcribed and annotated data were available.

5 ASR Towards Industrialization

In the SESAR2020 Industrial Research project PJ.16-04 the ASR activity is fostered on a broad basis by many project partners. Nineteen European affiliations from fifteen different countries contributed to maturing the technology readiness level (TRL) to TRL4. The overall aim of the project was to increase ATCo's productivity. Supporting companies consisted of European air navigation service providers, three ATM system providers, and research/consultancy organizations [21].

One achievement of the ASR activity was the definition of an ontology for annotation of ATCo commands [19]. The ontology is a set of rules on how to formally understand the content of an ATCo utterance which can consist of multiple concepts. Before extracting concepts, transcription of utterances is required. An example of a transcription and the agreed annotation of concepts from this example are shown in Fig. 4. Each utterance is annotated as a series of callsign-instruction pairs. The instruction can consist of a mandatory command part and optional conditions. The command itself is composed of a type (see example in Fig. 5) and in most cases of a value, a unit, a qualifier and a condition as shown in the given example above.

The developed ontology currently consists of 120 different command types for the en-route, approach and tower phase. It takes the ICAO phraseology and CPLDC protocol into account. However, the ontology sometimes goes beyond or is more general to satisfy the needs to harmonize integration of ASR into controller working positions.

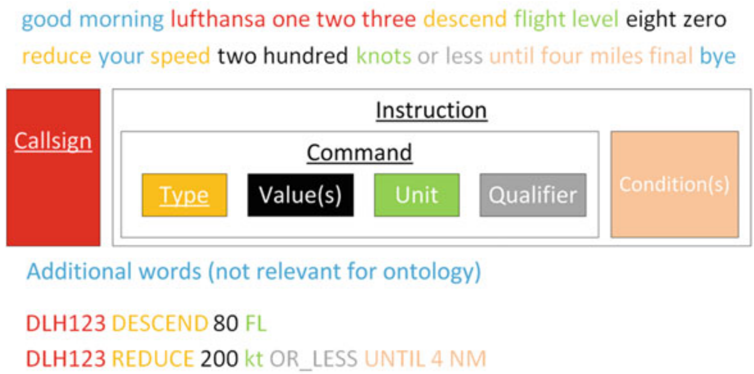

Fig. 4 Basic scheme of the ontology for controller command annotation with sub-parts of an instruction and example commands

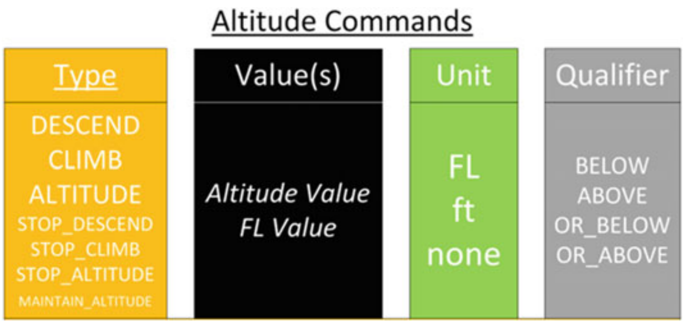

Fig. 5 Sub-parts of command annotation; example: altitude commands

The PJ.16-04 partners conducted several different validation exercises in 2018 and 2019 concerning ASR. One exercise from THALES, DLR, ANS-CR, Integra, and Austro Control (ACG) integrated different components to an ABSR system for Prague and Vienna approach [22]. DLR provided the hypotheses generator to predict controller commands and the checker component. They were used to improve the commercial ASR engine used by THALES. Validation trials with Czech and Austrian ATCos in the THALES SkyCentre proved that the hypotheses generator and the command checker significantly reduced the CER and thus in an environment similar to real ATC operations rooms.

Another exercise of PJ.16-04 compared issued clearances from Hungarian and Lithuanian ATCos in multiple remote tower environments with controller command predictions, developed by DLR [23]. To the best of our knowledge this was the first time that controller command prediction has been developed for a tower CWP.

Fig. 6 Multiple remote tower trials at DLR Braunschweig

Furthermore, it was the first to deal with a multiple remote tower environment forecasting controller commands for different airports in parallel. The command prediction was tested in a set-up for the PJ.05-02 multiple remote tower trials at DLR Braunschweig, see Fig. 6. The complete trials generated 107 recorded simulation runs. The command prediction error rate for annotated trials was 7.3%, i.e. 93% of the commands given by the ATCo were predicted [23].

6 Conclusions and Outlook

The paper presents the evolution of Assistant Based Speech Recognition (ABSR) introduced by DLR and Saarland University. AcListant® project has shown that both acceptable CRRs (>90%) and CERs (<3%) are possible.

AcListant®-Strips even improves ASR performance (above 95% and below 1.7%) and quantifies the benefits of ABSR: Controllers' "clicking time" is reduced by a factor of three resulting in two landings more per hour and 60 L of kerosene saving per inbound flight based on released cognitive ATCo resources. The command recognition and error rates were classified as totally sufficient by ATCos that participated in the ABSR trials.

MALORCA developed generic reusable modules and models. The latter ones can automatically be trained by machine learning algorithms. This result in reduced adaptation costs. SESAR2020s Wave 1 funded project 16-04 enables exchange of training data and reduced transcription and annotation effort, because the main European ATM players agreed on an ontology for command annotation.

SESAR2020s Wave 2 further promotes activities on ABSR with solutions PJ.10-96 and PJ.05-97 that were started end of 2019. Solution 97 foresees validation trials with an ABSR system integrated into a tower environment. This comprises trials at DLR in Braunschweig and EUROCONTROL in Brétigny. ACG controllers will also perform ABSR trials in the Vienna approach operation's room in solution 96, the first time that an ABSR system will be directly integrated into the ops room of an air navigation service provider.

Acknowledgments The projects AcListant® and AcListant®-Strips were supported by DLR Technology Marketing and Helmholtz Validation Fund. The SESAR Exploratory Research project MALORCA and Industrial Research project PJ.16-04 CWP HMI also comprising the Automatic Speech Recognition activity (PJ.16-04-02) have received funding from the SESAR Joint Undertaking under the European Union's grant agreement No. 698824 respectively 734141.

References

1. Single European Sky ATM Research Joint Undertaking, https://www.sesarju.eu. SESAR Joint Undertaking, n.d.
2. Federal Aviation Administration, https://www.faa.gov/nextgen/, Modernization of U.S. Airspace, n.d.
3. Ministry of Land, Infrastructure, Transport and Tourism, https://www.mlit.go.jp/en/koku/koku_fr13_000000.html, Collaborative Actions for Renovation of Air Traffic Systems (CARATS), n.d.
4. International Civil Aviation Organization, https://www.icao.int/Meetings/a39/Documents/WP/wp_304_en.pdf, China's Strategy for Modernizing Air Traffic Management, n.d.
5. Eurocontrol, "LINK2000+: ATC data link operational guidance in support of DLS regulation," No 29/2009, vol. 6, 17 December 2012, online available at https://www.skybrary.aero/bookshelf/books/2383.pdf
6. O. Veronika Prinzo, Data-linked pilot reply time on controller workload and communication in a simulated terminal option, Civil Aeromedical Institute, Federal Aviation Administration, Oklahoma City, Oklahoma, USA, May 2001
7. ICAO, Global operational data link document (GOLD), 2nd edn. (2013)
8. D. Schäfer, Context-sensitive speech recognition in the air traffic control simulation, in *Eurocontrol EEC Note No. 02/2001 and PhD Thesis of the University of Armed Forces* (Munich, Germany, 2001)
9. J.M. Cordero, M. Dorado, J.M. de Pablo, Automated speech recognition in ATC environment, in *Proceedings of the 2nd International Conference on Application and Theory of Automation in Command and Control Systems (ATACCS '12)* (IRIT Press, Toulouse, France), pp. 46–53
10. V.I. Levenshtein, Binary codes capable of correcting deletions, insertions, and reversals. Soviet Phys. Doklady **10**(8) (1966)
11. H. Helmke, J. Rataj, T. Mühlhausen, O. Ohneiser, H. Ehr, M. Kleinert, Y. Oualil, M. Schulder, Assistant-based speech recognition for ATM applications, in *11th USA/Europe Air Traffic Management Research and Development Seminar (ATM2015)* (Lisbon, Portugal, 2015)
12. H. Gürlük, H. Helmke, M. Wies, H. Ehr, M. Kleinert, T. Mühlhausen, K. Muth, O. Ohneiser, Assistant based speech recognition—another pair of eyes for the Arrival Manager, in *IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)* (Prague, Czech Republic, 2015)
13. AcListant homepage, www.AcListant.de, AcListant = Active Listening Assistant, n.d.
14. H. Helmke, O. Ohneiser, T. Mühlhausen, M. Wies, Reducing controller workload with automatic speech recognition, in *IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)* (Sacramento, CA, USA, 2016)

15. H. Helmke, O. Ohneiser, J. Buxbaum, C. Kern, Increasing ATM efficiency with assistant based speech recognition, in *12th USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)* (Seattle, WA, USA, 2017)
16. The project MALORCA, https://www.malorca-project.de, n.d.
17. M. Kleinert, H. Helmke, G. Siol, H. Ehr, M. Finke, Y. Oualil, A. Srinivasamurthy, Machine learning of controller command prediction models from recorded radar data and controller speech utterances, in *8th SESAR Innovation Days* (Belgrade, Serbia, 2017)
18. M. Kleinert, H. Helmke, G. Siol, H. Ehr, A. Cerna, C. Kern, D. Klakow, P. Motlicek et al., Semi-supervised adaptation of assistant based speech recognition models for different approach areas, in *37th AIAA/IEEE Digital Avionics Systems Conference (DASC)* (London, UK, 2018)
19. H. Helmke, M. Slotty, M. Poiger, D. Ferrer Herrer, O. Ohneiser, N. Vink, A. Cerna, P. Hartikainen, B. Josefsson, D. Langr, R. García Lasheras, G. Marin et al., Ontology for transcription of ATC speech commands of SESAR 2020 solution PJ.16-04, in *37th AIAA/IEEE Digital Avionics Systems Conference (DASC)* (London, UK, 2018)
20. M. Kleinert, H. Helmke, H. Ehr, C. Kern, D. Klakow, P. Motlicek, M. Singh, G. Siol, Building blocks of assistant based speech recognition for air traffic management applications, in *8th SESAR Innovation Days*, Salzburg, Austria, 2018.
21. The SESAR Project PJ.16-04, https://www.sesarju.eu/projects/cwphmi, n.d.
22. M. Kleinert, H. Helmke, S. Moos, P. Hlousek, C. Windisch, O. Ohneiser, H. Ehr, A. Labreuil, Reducing controller workload by automatic speech recognition assisted radar label maintenance, in *9th SESAR Innovation Days* (Athens, Greece, 2019)
23. O. Ohneiser, H. Helmke, M. Kleinert, G. Siol, H. Ehr, S. Hobein, A.-V. Predescu, J. Bauer, Tower controller command prediction for future speech recognition applications, in *9th SESAR Innovation Days* (Athens, Greece, 2019)