Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017)

Increasing ATM Efficiency with Assistant Based Speech Recognition

Hartmut Helmke, Oliver Ohneiser
German Aerospace Center (DLR),
Institute of Flight Guidance,
Department Controller Assistance,
Braunschweig, Germany
Hartmut.Helmke@dlr.de
Oliver.Ohneiser@dlr.de

Jörg Buxbaum
DFS Deutsche Flugsicherung GmbH
Joerg.Buxbaum@dfs.de

Christian Kern
Austro Control
Christian.Kern@austrocontrol.at

Abstract—**Initiatives to integrate Automatic Speech Recognition into Air Traffic Management (ATM) exists at least since the late 90s. Some success to replace pseudo pilots have been reported, but its integration into controller assistant tools is missing. German Aerospace Center (DLR) and Saarland University developed Assistant Based Speech Recognition (ABSR) enabling command recognition rates better than 95%. However, good recognition rates are no convincing argument for decision makers. Therefore, we conducted an ABSR validation study with eight air traffic controllers to quantify the benefits with respect to workload and efficiency. The study validates that ABSR does not just reduce controllers' workload, which would already be a lot, but this paper presents that ABSR significantly increases ATM efficiency. Fuel reductions of 60 liters (16 gallons) per flight and a throughput increase by two arrivals per hour are possible.**

Keywords—AcListant®, Automatic Speech Recognition (ASR), Assistant Based Speech Recognition (ABSR), ATM Efficiency, Electronic Flight Strips, Aircraft Label

I. Introduction

The most important task of an air traffic controller (ATCo) is ensuring safety for all involved parties in air traffic. ATCos follow the simple rule "safe, orderly, and expeditious". Hence, efficiency is the next important aspect mainly having economic thoughts of air navigation service providers (ANSP) in mind.

A. Problem

Controllers' tasks and their time distribution are one essential factor. ATCos' core tasks consist of communication and coordination with pilots and other controllers. This can be hampered, however, if ATCos need to spend additional time on subordinate tasks such as documentation. One of those subordinate tasks is maintaining flight information in flight strips and on-screen labels. Strips in electronic or paper form contain static and dynamic flight data. Static data comprises e.g., call sign, weight category, destination, or route information. Dynamic data includes e.g., clearances regarding altitude, speed, direction, rates of climb/descent or procedures, as well as special flight situations like emergencies.

Paper flight strips are still often used, for example, in high density terminal maneuvering areas (TMA). They have the disadvantage of information not being available or transferable in digital form. Modern controller working positions (CWP), therefore, offer digital flight strips. However, they normally have to be managed head down averting one's eyes away from the traffic situation display. For both forms of flight strips, manual documentation of flight data is redundant to what the controller already told or will tell to the aircraft pilot.

B. Solution

The AcListant® project has shown that Assistant Based Speech Recognition (ABSR) support for ATCos is a solution [1]. ABSR helps the controller with partially automated aircraft radar label maintenance, i.e., an automatic speech recognizer (ASR) uses context knowledge about the current situation from a controller assistant system. This enables prediction of the next most probable commands and reduces the search space for the speech recognizer. With this technique, the AcListant® project achieved command error rates below 1.7% [1]. An ASR system with this level of accuracy is feasible for operational use even in the safety-critical air traffic control (ATC) domain. ATCos need to correct the automation only in very few cases. Hence, ATCos' concentration can remain on their main tasks. Furthermore, less time, spent for subordinate tasks, produces free cognitive resources for increasing air traffic demand.

C. Derived Problem

The AcListant® trials of DLR, Saarland University, DFS, Austro Control, and ANS CR (Air Navigation Service of Czech Republic) have shown that Assistant Based Speech Recognition achieves acceptable recognition rates [1] with very positive feedback from involved controllers [2]. However, positive feedback of controllers is a pre-condition, but does not justify a business case. The benefits of speech recognition to the air traffic system also need to be quantified.

D. Solution of Derived Problem

The AcListant®-Strips project [3], the successor of AcListant®, quantifies the benefits of ABSR [4]. Two possible methods to insert given controller commands into the radar labels were compared. The first input method was the baseline. Controllers used the computer mouse for manual input. The second input method automatically worked with ABSR, analyzing the radio communication channel between controller and pilot. The controller may confirm, correct, or reject the output of the speech recognizer. In November and December 2015 the validation trials for benefit quantification were performed in DLR's labs in Braunschweig. The challenge was that the results should not only show a trend, but should be (statistically) significant and of course the project budget was not *unlimited*.

E. Paper Structure

After presenting related work and the concept of ABSR in section II, we explain the performed validation exercise in section III and then briefly summarize the workload improvements in section IV, which were already presented in detail at the 2016 DASC [4]. Section V shows our measurements regarding efficiency and points out results. Section VI investigates the status of speech recognition in ATM applications. We draw conclusions and outline planned future work in the last section.

II. Background

ASR applications can be divided into three different categories: First, dictation software, which is used in the professional market [5]. In consumer products, they are not widely accepted due to their lack of adaptivity. Second, hands-free command and control, which is characterized by short utterances to control technical devices [6], and third, spoken dialog systems, which include Apple's Siri® [7], Google's search by voice [8], and train table dialog systems [9].

A. Speech Recognition Applications in Air Traffic Control

The integrations of ASR in ATC training started in the late 80s [10]. Nowadays enhanced ASR systems are used in ATC training simulators to replace expensive pseudo pilots (e.g., FAA [11], DLR [12], MITRE [13], DFS [14]). ASR applications also go beyond simulation and training. ATC events can automatically be detected in order to assess controller workload. ASR is used to get more objective feedback concerning controllers' workload [15], [16]. Chen and Kopald used speech recognition to build a safety net for airport surface traffic to avoid aircraft using a closed runway [17], [18]. Most recently they presented an approach to detect pilot read back errors [19].

Although ASR systems are widely used in normal life (e.g., Siri®, interface for car navigation systems) and ATC phraseology is standardized, recognizing and understanding controller-pilot communication is still a big challenge and not solved satisfactory. Using common and widely extended ASR tools has not provided acceptable results in terms of word and command detection rates. Reasons for poor performance include the unique ATC vocabulary and syntax, as well as the variety of accents, speakers, and communication channels with different characteristics, and especially controllers' needs to deviate from standard phraseology [20]. Cordero et al. (2012) reported word detection rates not above 20% with different Commercial-off-the-shelf (COTS) recognizers [15].

B. Assistant Based Speech Recognition

One promising approach to improve ASR performance is using context knowledge regarding expected utterances. These attempts go back to the 80s [21], [22]. This information may heavily reduce the search space and lead to fewer missed recognitions [12].

Our ABSR approach uses the output of an Assistant System, i.e., DLR's Arrival Manager (AMAN) 4D-CARMA [23], as context information. Figure 1 describes the concept of assistant based speech recognition. An "Assistant System" analyzes the current situation of the airspace and predicts possible future states used by the "Hypotheses Generator" to predict the set of possible commands. This dramatically reduces the search of the "Lattice Generator" [24], [25]. The search lattice (tree) is dynamically regenerated every 30 seconds and contains a search tree for all possible phoneme sequence determined by the "Hypotheses Generator". The "Speech Recognizer" finds the most probable path in the search tree. We use the public domain speech recognition engine KALDI [26], [27]. The output of the "Command Extractor" is checked again by the "Plausibility Checker", determining whether the recognized commands are reasonable in the current situation, e.g., do not produce conflicts. The "Command Monitor" analyzes the future behavior of the aircraft (via radar data), whether they are in line with the "Command Extractor's" output.

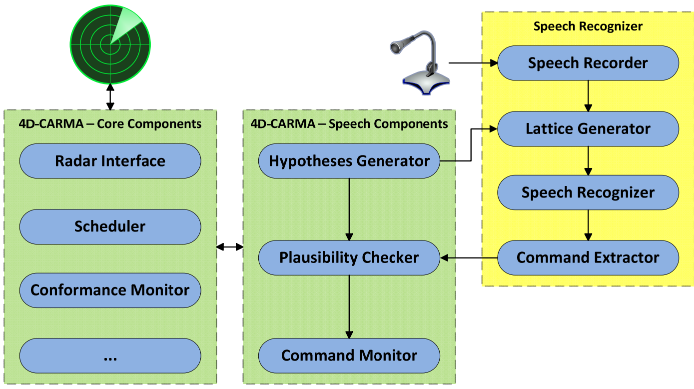

Figure 1. Components of Assistant Based Speech Recognition [1]; in green components of 4D-CARMA; in yellow components of core speech recognizer.

This paper completes our work with respect to ABSR, which started with the study of Shore et al. in 2011 [28]. In a pilot study with a limited set of call signs and commands, Shore [29] reported command (recognition) error rates below 5%. He used an acoustic model derived from the Wall Street Journal recognition corpus. Our ATM Seminar 2013 paper already presented a possible ATC application of ABSR [30]: faster adaptation of an Arrival Manager, if the controller intentionally deviates from the proposal of the assistant system. In 2015 we demonstrated that (1) ABSR is able to generate acceptable speech recognition (>90%) and error rates (<3%), (2) ABSR significantly reduces deviations between the controllers' plan and the plan of the Arrival Manager, and (3) ABSR significantly reduces controllers' workload quantified in [4].

C. Workload Measurement

The often-used ISA score (Instantaneous Self-Assessment) [31] and NASA-TLX score (National Aeronautics and Space Administration Task Load Index [32]) only provide subjective feedback from the controllers themselves. Getting an objective workload measure, we used the secondary task performance measures method. This method identifies the amount of additional work the controller (or operator more generally) can perform in addition to the primary work of air traffic control [33]. Thereby, the secondary task performance serves as an index for the workload of the controller [34]. The advantages of the secondary task method are that it is easy to use and sensitive to variations in workload [35]. The high-end version of workload measurement is physiological measures such as functional near-infrared or electroencephalography which uses special hardware to record different brain activities [36], [37].

D. From Paper Flight Strips to Paperless Systems

Speech recognition in its role as ATCo assistance is in line with the SESAR Concept of Operation (ConOps), sect. "Strategy to reduce Controller task load" [38]. According to the ConOps, automation for routine controller task load will comprise better methods of data input and improved data management. Although manual data input is currently a routine task, the European ATM R&D (research and development) standpoint is that it is undesirable to have ATCo send radio transmissions to cockpit crews and additionally write essential content of these messages on paper or enter it manually into systems, generating extra workload without extra operational benefits.

No plans exist to fully replace radio communication during the course of the next 20 years by CPDLC (Controller-pilot data link communications). And even if ATCos are supported by CPDLC the amount of required system inputs does not decrease. Therefore, speech related assistance would provide substantial operational support. This view is in line with the ATM part of the technology roadmap [39] issued by the German Aerospace Industries Association (BDLI), which derives R&D-needs out of the aerospace strategy issued by the German government [40].

During the first stage of SESAR, a taxonomy for automation levels of ATCos' tasks and ATM system functionalities was developed [38], to allow for proper judgement about current automation levels and possible further advances. According to this taxonomy, ABSR will increase the degree of automation in the "Information Acquisition" task domain due to the automated data entry and the arising additional benefits of instantly putting this data in relation to other available relevant operational information (e.g. conflict detection systems): *"The system supports the human in acquiring info on the process she/he is following. The system integrates data coming from different sources and filters and/or highlights the information items considered relevant for the user."* [38] Due to this and due to its proven relevance, speech recognition will play a decisive role in SESAR 2020 (HMI development, PJ16 [41]). Furthermore, positive results of R&D-projects like AcListant® lead to integration of speech-based assistant systems as a relevant goal of ATM development in the BDLI Technology Roadmap [39].

One example of an ANSP making this transition is Austro Control. Austro Control had been using VAS, the Vienna ATM System, but in 2007 it was decided to replace VAS by a completely new ATM system. The decision was based on financial- and organizational factors. Therefore, Austro Control joined the Cooperation of Air Navigation Service Providers (COOPANS), which is an initiative of the five ANSPs from Ireland, Sweden, Denmark, Croatia, and Austria [42]. These countries mutually agreed on a common Thales Eurocat ATM system called "TopSky". While VAS used paper strips to provide the air traffic controllers with all required flight data, TopSky totally operates paperless. The low failure rates of today's ATM systems allow ANSPs to totally rely on electronic devices and to get rid of paper strips.

In Austria, the area control center made its transition to the paper-less system by end of February 2013, all terminal units followed by end of November 2015. Today only flight information service still uses paper-strips. However, these will also be substituted by electronic subsystems soon. Then all air traffic control units from tower to area control in Austria will operate fully paperless. All important information is now presented on the radar screen. Currently the update process, however, requires manual controller interactions by mouse and keyboard. This is mostly challenging for the approach control units, which need to issue the highest number of instructions in narrow timeframes.

During the first months of operation controllers were occasionally challenged coping with all the required inputs into the system at least in high density traffic situations. This risk, which became an issue, had been identified at a very initial stage before TopSky was operationally used. Each instruction to aircraft (e.g., altitudes, headings) forces the controller moving the cursor precisely to the appropriate locations on the screen and clicking once or several times by mouse. This significantly increases the workload. Previous available workload buffers are reduced resulting in additional sectors in bad weather conditions. Voice recognition was soon identified as a potential solution to this problem. Therefore, Austro Control joined the AcListant® project end of 2014. The very encouraging results with respect to workload reduction were reported at the 2016 DASC [4] and in [43] as well as with respect to efficiency improvements in the following sections.

A challenge for controllers of all units was getting used to a new system that differed significantly from the previous one. Therefore, intensive training was required. Each controller completed a program of 10 different modules lasting for 22 days in total during the year before transition. Early integration of speech recognition into the TopSky system could even reduce this huge effort, because workload requirements are more comparable to the VAS system with paper flight strips.

III. Validation

The main purpose of the AcListant®-Strips project was to quantify the benefits of ABSR (developed by Saarland University (UdS) and DLR) in ATC with respect to efficiency and controllers' workload. The benefit should arise from supporting (approach) controllers with aircraft radar label maintenance. Therefore, we compared the modalities of a "classical manual" *Mouse only* input for controller commands into the human machine interface (HMI) with an *ABSR + Mouse* input. First, we detail the flow of controller commands into the HMI. Second and third, we explain the experiment setup as well as the scenarios and configurations for our experiments. Fourth, we describe the study participants.

A. Controller Command HMI Input

If the controller clicks on one of the five grey aircraft label cells in the HMI, a drop-down menu to enter given clearances opens (see Figure 2).

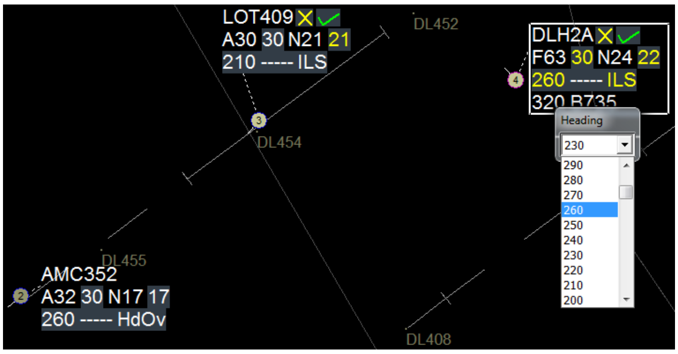

Figure 2. Drop-down menu for heading input in aircraft radar label.

Each new but unconfirmed value (altitude / speed / direction / rate of altitude change / miscellaneous) will appear in yellow colour after choosing. These values will turn white as soon as the controller accepts them with a click on the green check mark in the first label line.

When using *ABSR* with *Mouse*, the controller command values are extracted from the radio telephony utterance and automatically appear in the five label cells in yellow. Thus, the controller only needs to check and confirm with a mouse click or correct values in seldom cases of misrecognition.

B. *Setup of Experiments*

The whole AcListant®-Strips project validation cycle included iterative validation trials in accordance with the European Operational Concept Validation Methodology (E-OCVM) [44]. Two pre-validation rounds took place in July and September prior to the final validation in November/December 2015.

Figure 3 shows the basic validation setup for the controller working position in DLR's simulation facility, ATMOS (Air Traffic Management and Operations Simulator). The controller communicated with pseudo pilots in another room via voice-over-IP (VoIP). The headset microphone signal of the controller served as the input signal for the speech recognition engine.

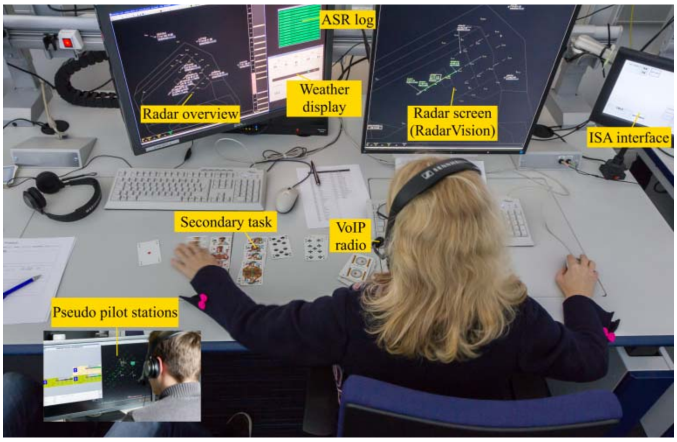

Figure 3. Basic validation setup during final trials.

The main radar screen (right square monitor in Figure 3) displayed the simulated Düsseldorf (EDDL) airspace via the DLR human machine interface RadarVision [45]. The display consisted of TMA structure, routes, GPS waypoints, and radar data of inbound aircraft on runway 23R. The left monitor shows a broader radar overview, automatic speech recognition (ASR) log, and a weather display. The radar overview can be used to estimate the arrival traffic in the next quarter of an hour. The ASR log shows recognized controller utterances with their three elements: call sign, command type, and value. The weather display indicates air pressure (QNH) and wind information.

The ISA (Instantaneous self-assessment) interface (right small monitor in Figure 3) asked for regular study participant feedback regarding individual self-assessed workload [31]. NASA-TLX score [32] was used as a second subjective workload rating measure after each simulation run. A secondary task was evaluated to gather an objective measure for controller workload during the trials. Study participants had to sort a deck of 48 cards into six decks for each playing card type (9-10-Jack-Queen-King-Ace) and name one to four randomly missing cards. Participants were instructed to not let ATC task performance suffer from the secondary task. The time needed to sort the cards and finally identify the missing ones as well as the number of repetitions served as an objective value for user workload.

C. *Scenarios and Configurations of Experiments*

There are two different types of approach controller responsibility areas: initial approach and final approach. *"Final approach"* means responsibility area of a final approach controller in the US, but feeder controller in Europe. *"Initial approach"* means feeder controller in the US, but pickup controller in Europe. "*Complete approach*" is comparable to a combined pickup/feeder controller in Europe as well as a combined feeder/final controller in the US. For generalization we use the terms *Final Ap*, *Initial Ap*, and *Complete Ap*. Study participants had to work as *Final* or *Complete Ap* controller only.

The 60-minute Complete Ap (C) scenario (see Figure 4) comprised medium traffic conditions (roughly 35 inbounds per hour). It included a five minute runway closure phase in the first half and an emergency flight in the beginning of the second half of the scenario.

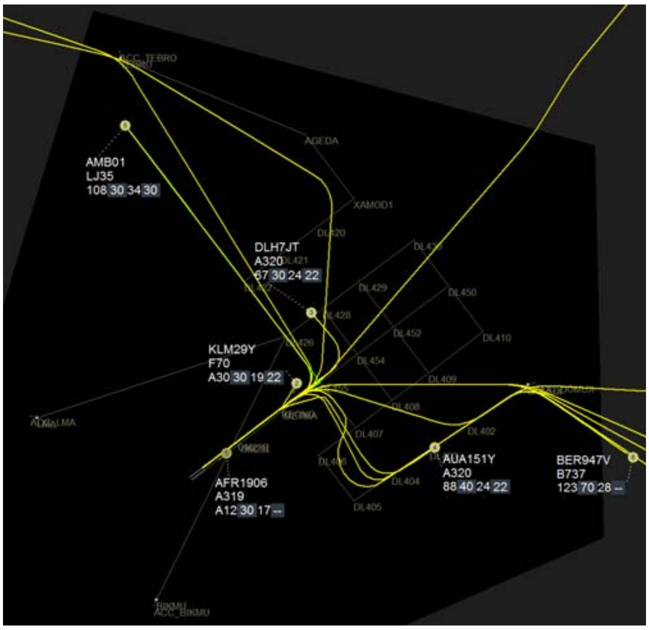

Figure 4. Area of responsibility during Complete Ap scenario.

In the 45-minute Final Ap (F) scenario (see Figure 5) traffic density was very high with 60 inbounds per hour. However, sixty arrivals on one runway are not legal due to ICAO minimum separation requirements. Therefore, controllers had to implement holdings or long path stretching legs to ensure safe traffic operations.

During pre-trials we tried another method to handle excessive traffic flow: Controllers should tell us when there were too many aircraft under their responsibility. The next two aircraft are then deleted from inbound stream as if a virtual Initial Ap controller would have diverted them. Controllers, however, did not ask for intervention at the time it would have been helpful. The extended runway center line often contained large numbers of aircraft up to 40 NM distance to threshold. Virtual Initial Ap coordination of an excessive arrival stream was not initiated until then. Hence, we did not use this "manual" method during the main trials. The deletion of two approaching aircraft was automatically done by a simulated Initial Ap controller if the final got longer than 22 NM.

A second *lesson learned* from the July and September trials concerned the acknowledgement of recognized uttered controller commands. During these pre-trials, controllers had a reaction time of 20 seconds to confirm (ACCEPT) or negate (REJECT) a displayed recognized command via mouse click on a green arrow or yellow cross respectively. Commands were automatically rejected without any manual response after these 20 seconds. However, due to very low ABSR error rates, the majority of controllers told us to change the default behavior. So, we automatically accepted recognized controller commands in the main trials if no controller action was taken within 20 seconds.

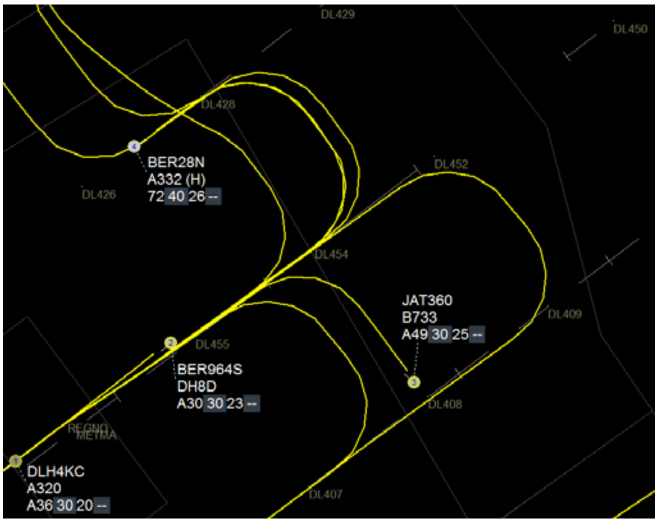

Figure 5. Responsibility area in the Final Ap scenario.

During the main trials we analyzed three different controller input modalities. Those modes were: (1) *Mouse only*, (2) speech recognition plus mouse correction (*ABSR* + *Mouse*), and (3) speech recognition plus correction via multi-touch display (*ABSR* + *MT*).

All scenarios and input modalities were trained in additional training runs (T) in advance. We reduced the number of combinations from 9 (3 scenarios C, F, T times 3 input modalities *Mouse only*, *ABSR* + *Mouse*, *ABSR* + *MT*) to 7 (we call them T-1/2, T-3, F-1, F-2, F-3, C-1, C-2). This enabled us to handle one study participant within two half day sessions. The initial training runs (T-1/2, T-3) and the *ABSR* + *MT* run F-3 were conducted in the first sessions and are not discussed further in this paper, because input modality 3 (*ABSR* + *MT*) was used as an additional training. The study participants did not know the training character of the multi-touch input modality runs in advance. This paper presents results from the following four simulations runs:

- Final Ap with *Mouse only* (F-1)
- Final Ap with *ABSR* + *Mouse* (F-2)
- Complete Ap with *Mouse only* (C-1)
- Complete Ap with *ABSR* + *Mouse* (C-2)

The sequence of those four evaluated runs was randomized for each participant to reduce biased results resulting from training effects.

D. Study Participants

The German and Austrian ANSPs DFS Deutsche Flugsicherung GmbH and Austro Control took part at our main trials. Each ANSP sent four air traffic controllers whereof two in total were female. The average age was 36 (*Standard Deviation SD=Sigma* = 11; age interval between 22 and 53 years). Their professional work experience was 14 years on average (*SD* = 11; experience interval between 1 and 32 years).

IV. Results with Respect to Workload Reduction

As the focus of this paper is on validation results with respect to efficiency, we only briefly summarize the workload reduction results and point to [4] and [43] for more details. Nevertheless, it should be considered that a reduction of workload with the resulting use of free cognitive resources may increase efficiency again.

The subjective workload measure of ISA score showed significance for the C and F scenario as well as combined due to the performed statistical paired t-test (p-values between 10^{-5} and 0.048). Hence, the self-assessed workload of controllers was lower in the *ABSR* + *Mouse* condition than in the *Mouse only* condition. The same is true for the NASA-TLX index. It improved by 20% being supported by ABSR.

Analyzing the secondary task of sorting cards and naming missing ones reveals much and significantly better performance (p-value 0,3%) if controllers were in the *ABSR* + *Mouse* condition. They were roughly by a factor of two faster in their secondary task compared to the *Mouse only* run. This results in an increase of free cognitive resources when ABSR support for radar label maintenance is available. A similar conclusion can be drawn from the time needed to enter controller commands by clicking into radar aircraft labels. In the Mouse only condition controllers needed 30% of their total time (20 of 60 minutes) just to enter and confirm values of their given clearances (more details in [43]). With ABSR support this percentage drastically diminishes to only 10%. The portion was even higher during the Final Ap scenarios due to the greater number of approaching aircraft per hour. The results did not show an influence on the number of given controller clearances or the time gap between consecutive controller clearances.

To put it all in a nutshell, we showed that ABSR support for the radar label maintenance task of air traffic controllers significantly reduces their workload by a factor of three and even improves ATC system data quality. Resulting free cognitive resources may be used to perform additional (safety increasing) controller tasks or to handle a potential higher number of aircraft at a time. Thus, the proven workload reduction could lead to more safety and efficiency. The latter is considered in the following section.

V. Results with Respect to Efficiency

According to the E-OCVM methodology [44] we first derive the validation hypotheses (subsect V.A), then define, take, and calculate the measurements (subsect V.B). Then we test the hypotheses (subsect. V.C) and interpret the results (subsect. V.D).

A. Hypotheses

In the AcListant®-Strips validation plan [46], the basis of our validation trials, the following efficiency related hypotheses were formulated:

ABSR support for radar label maintenance (in contrast to *Mouse only* input) …

1. … increases aircraft throughput (flow),
2. … decreases aircraft flown distance,
3. … decreases aircraft flight time, and
4. … reduces the number of missing inputs in radar label.

B. Measurements

From the hypotheses we derived the following measurements. More details to the measurement values presented in this paper can be found in the final AcListant®-Strips validation report [43].

1) Aircraft Throughput / Flow

Each run starts with lower traffic. Our measurement time starts when the third aircraft has landed and ends with the touchdown time of the last aircraft in the simulation time frame. The scenarios contain heavy and medium aircraft. A separation of 3 NM (medium-medium, medium-heavy), 4 NM (heavy-heavy) and 5 NM (heavy-medium) was required. Therefore, we multiplied the number of landing heavies by a factor of 1.6. Mediums count only 1. Table 1 shows the resulting throughput of both scenarios. Table 2 summarizes the throughput values with respect to mean, standard deviation, and median.

TABLE 1: INBOUND THOUGHPUT IN AIRCRAFT PER HOUR

	Complete Ap		Final Ap	
Controller	ABSR+Mouse	Mouse	ABSR+Mouse	Mouse
A	35.4 (1)	35.3 (4)	46.8 (5)	46.2 (3)
B	34.5 (2)	36.2 (5)	43.7 (3)	42.9 (4)
C	35.7 (4)	34.7 (1)	45.2 (3)	44.3 (5)
D	37.7 (5)	36.2 (2)	44.5 (4)	42.9 (3)
E	31.5 (1)	35.7 (4)	42.4 (3)	45.3 (5)
F	35.2 (5)	31.4 (2)	44.9 (3)	38.4 (4)
G	36.2 (4)	30.0 (1)	46.0 (5)	40.6 (3)
H	35.2 (2)	35.6 (5)	45.4 (4)	41.8 (3)

In brackets we add, when each controller performed the experiment. Controller F e.g., performed the Final Ap scenario with ABSR plus mouse support as his third run, whereas for G it was his last one. Used shading is explained later.

TABLE 2: AVERAGE, MEAN, SIGMA OF RESULTS FOR AIRCRAFT THROUGHPUT

Scenario	Input modality	Mean	Sigma = SD	Median
Complete Ap	Mouse only	34.4	2.2	35.4
Complete Ap	ABSR+Mouse	35.2	1.6	35.3
Final Ap	Mouse only	42.8	2.4	42.9
Final Ap	ABSR+Mouse	44.9	1.3	45.1

2) Flown Distance

TABLE 3: FLOWN DISTANCE IN NM

	Complete Ap		Final Ap	
Controller	ABSR+Mouse	Mouse	ABSR+Mouse	Mouse
A	69.1 (1)	72.5 (4)	58.3 (5)	55.9 (3)
B	67.9 (2)	72.1 (5)	63.9 (3)	65.3 (4)
C	71.3 (4)	73.4 (1)	58.2 (3)	56.9 (5)
D	71.6 (5)	71.7 (2)	62.2 (4)	62.9 (3)
E	81.7 (1)	78.4 (4)	64.3 (3)	60.2 (5)
F	75.4 (5)	97.0 (2)	62.9 (3)	60.4 (4)
G	69.2 (4)	81.7 (1)	53.1 (5)	60.6 (3)
H	72.0 (2)	71.8 (5)	54.9 (4)	63.1 (3)

In Table 3 we show the measured flown distance of each aircraft (again except the first three) from entering into the scenario until touch down and consider only aircraft which landed within the scenario time. Table 4 summarizes mean value, sigma, and median for each scenario and input modality.

TABLE 4: RESULTS FOR FLOWN DISTANCE IN NM

Scenario	Input modality	Mean	Sigma = SD	Median
Complete Ap	Mouse only	77.3	8.2	72.9
Complete Ap	ABSR+Mouse	72.3	4.2	71.4
Final Ap	Mouse only	60.7	2.9	60.5
Final Ap	ABSR+Mouse	59.7	4.0	60.2

3) Flight Time

In the same way, we compare the flight times. We calculate for each aircraft (except the first three landings) the difference between its flight time (from entering into scenario until touch down) and its earliest possible time predicted by AMAN 4D-CARMA (see Table 5 and Table 6).

TABLE 5: ADDITIONAL FLIGHT TIME IN SECONDS

	Complete Ap		Final Ap	
Controller	ABSR+Mouse	Mouse	ABSR+Mouse	Mouse
A	121.6 (1)	190.1 (4)	285.1 (5)	266.6 (3)
B	111.1 (2)	163.6 (5)	375.0 (3)	401.6 (4)
C	187.8 (4)	215.4 (1)	286.6 (3)	291.3 (5)
D	147.9 (5)	150.7 (2)	316.8 (4)	344.8 (3)
E	305.3 (1)	271.9 (4)	437.6 (3)	347.4 (5)
F	208.6 (5)	529.1 (2)	376.2 (3)	340.3 (4)
G	141.8 (4)	345.3 (1)	250.7 (5)	360.2 (3)
H	178.0 (2)	156.4 (5)	250.4 (4)	364.7 (3)

TABLE 6: AVERAGE, MEAN AND SIGMA OFADDITIONAL FLIGHT TIME [S]

Scenario	Input modality	Mean	Sigma = SD	Median
Complete Ap	Mouse only	253	122	203
Complete Ap	ABSR+Mouse	175	58	163
Final Ap	Mouse only	340	40	346
Final Ap	ABSR+Mouse	322	63	302

In the Final Ap scenario the controller only influences a small part of the flight (see Figure 5). In the Complete Ap scenario (Figure 4) we save, however, 77 seconds per flight. These unplanned flight time extensions take place in level flight. An A320 needs 2,700 liters of kerosene per hour in a level flight in FL70 with CAS (calibrated airspeed) of 250 knots, i.e., we save 57 liters per flight. DLR's test aircraft (A320-232) needs 3,100 liters per hour. We can expect fuel savings of 10.5 million liters per year for a medium airport with 500 arrivals per day!

4) Missing Radar Label Inputs

Paper flight strips only provide limited access to clearance information. Electronic flight strips promise that more stakeholders can benefit from clearances available in digital form. This, however, requires that the controller inputs all given clearances.

TABLE 7:MISSING RADAR LABEL INPUTS

	Complete Ap		Final Ap	
Controller	ABSR+Mouse	Mouse	ABSR+Mouse	Mouse
A	6.1% (1)	7.2% (4)	7.9%	3.2% (3)
B	3.4% (2)	8.4% (5)	1.5%	1.0% (4)
C	4.2% (4)	6.3% (1)	1.3%	3.0% (5)
D	1.4% (5)	10.4% (2)	1.2%	2.9% (3)
E	10.0% (1)	9.0% (4)	6.9%	5.9% (5)
F	5.6% (5)	25.9% (2)	8.1%	14.1% (4)
G	3.3% (4)	23.4% (1)	4.4%	14.9% (3)
H	5.2% (2)	6.6% (5)	3.5%	8.3% (3)

The number of given clearances that are not inserted into the radar label (either by mouse or by ABSR), is, therefore, also an efficiency measure. We manually transcribed all given clearances. Table 7 shows the percentage of commands which were neither manually by mouse nor automatically by ABSR entered into the radar labels. Table 8 summarizes the average, mean and standard deviation.

TABLE 8: AVERAGE, MEAN, SIGMA OF MISSING RADAR LABEL INPUTS

Scenario	Input modality	Mean	Sigma = SD	Median
Complete Ap	Mouse only	12.1%	7.4%	8.7%
Complete Ap	ABSR+Mouse	4.9%	2.4%	4.7%
Final Ap	Mouse only	6.7%	5.0%	4.5%
Final Ap	ABSR+Mouse	4.4%	2.8%	4.0%

5) *Overall Efficiency*

Throughput, flown distance, flight time, and missing radar label information are competing efficiency parameters. Reducing flight distances by diverting aircraft to other airports (in our case deleting them from scenario) normally decreases runway throughput and vice versa. Therefore, we define an overall efficiency measurement *OE* combining the four measurements. For each of the four measurements we calculate the best b_j and the worst value w_j across all trials of a particular scenario. From Table 1, for the Final Ap scenario we get b_{Flow} = 46.8 and w_{Flow}=38.4. For each of the four measurements (flow, distance, time, label info) we have for the Final Ap scenario 16 values m_{jk} with j=1...4 and k=1...16. For each measure the best value is set to 100%, the worst value is 0% and the values in between are linearly transformed to OE_{jk}. The overall efficiency OE_k for the 16 Final Ap experiments is the weighted average of all four values OE_{jk} (The sum of all c_j coefficients is 1.0):

$$OE_k = \sum_{j=1}^{4} \frac{m_j - w_j}{b_j - w_j} * c_j \quad (1)$$

For controller E with the mouse input modality in Table 1 we calculate ((45.3-38.4)/(46.8-38.4)=) 81.5%. In the same way we also transform the values for the Complete Ap scenario, see Table 9.

TABLE 9: OVERALL EFFICIENCY VALUES

	Complete Ap		Final Ap	
Controller	ABSR+Mouse	Mouse	ABSR+Mouse	Mouse
A	82.9% (1)	75.8% (4)	73.2% (5)	87.0% (3)
B	83.5% (2)	79.6% (5)	60.4% (3)	54.3% (4)
C	82.7% (4)	73.2% (1)	82.8% (3)	76.5% (5)
D	96.4% (5)	77.6% (2)	72.0% (4)	58.0% (3)
E	46.0% (1)	68.9% (4)	36.2% (3)	63.9% (5)
F	75.3% (5)	6.1% (2)	50.6% (3)	17.1% (4)
G	88.9% (4)	19.5% (1)	88.7% (5)	21.7% (3)
H	79.1% (2)	80.0% (5)	85.9% (4)	38.8% (3)

Table 10 shows the median, mean, and sigma overall efficiency values, grouped by scenario type and input modality. As flown distance and flight time highly depend on each other, we choose c_{Flow}=$c_{labelInfo}$ = 0.33 and c_{Dist}=c_{Time}=0.67.

TABLE 10: AVERAGE, MEAN, SIGMA OF OVERALL EFFICIENCY VALUES

Scenario	Input modality	Mean	Sigma = SD	Median
Complete Ap	Mouse only	60.1%	27.7%	74.5%
Complete Ap	ABSR+Mouse	79.4%	13.9%	82.8%
Final Ap	Mouse only	52.2%	23.2%	56.2%
Final Ap	ABSR+Mouse	68.7%	17.2%	72.6%

C. *Results*

We perform paired t-tests. Each hypothesis was tested three times: (1) for the Complete Ap scenario, (2) for the Final Ap scenario, and (3) for both scenarios together. As an example we take the hypothesis that ABSR support increases throughput. Our null hypothesis is "ABSR support does not increase aircraft throughput in the Final Ap scenario compared to the mouse input modality". Our test value is defined by

$$T = (D - \mu_0)\,\frac{\sqrt{n}}{SD} \quad (2)$$

We calculate the differences of the flows (ABSR-supported minus mouse-only run) of Table 1 for the Final Ap scenario, e.g., 44.9 minus 38.4 for controller F. The number of differences (controllers) is n (8 in our case). D is the mean value of the flow differences (45.1 minus 42.9 = 2.2). SD is the standard deviation of the difference (2.8 is calculated in [43]). We are just interested in checking whether ABSR input enables a higher flow than mouse input. Therefore, we set μ_0 to 0 arrivals per hour. We calculate a value T of 2.1.

As T obeys a t-distribution with n-1 degrees of freedom we can reject our null hypothesis H_0 with probability of α (p-value), if the calculated value for T is bigger than the value of the inverse t-distribution with n-1 degrees of freedom at position $t_{n-1,\, 1-\alpha}$ (in our case 1.9 for α=0.05). Therefore, the counter hypothesis H_0 is rejected, because T=2.1 > 1.9 holds. We could even calculate the minimal α so that T > $t_{n-1,\, 1-\alpha}$ is still valid. This is in our case α=0.038. Our results support the hypothesis.

We also calculated the probability to reject the null hypotheses for the Complete Ap scenario and for both scenarios together (row "Increased flow" in Table 11). In the same way we calculated α for reduced flight distance, reduced flight time, and less missing inputs in the radar label. Table 11 shows the minimal α values, which we marked in green if α is less than 5%, in light green for α < 10% and in yellow otherwise.

TABLE 11: MIN α FOR THE DIFFERENT HYPOTHESES

Hypotheses	Complete Ap	Final Ap	Both
Increased flow	23.8%	3.8%	3.7%
Flight distance	5.1%	27.7%	4.3%
Flight time	4.9%	23.8%	3.6%
Label information	1.9%	9.3%	0.7%
Overall efficiency	6.1%	7.7%	1.6%

D. *Interpretation*

The results presented in Table 11 are statistically significant. Nevertheless, we observe that in some cases the performance with only mouse support seems to be better than with ABSR support. In Table 1 we shaded these experiment pairs darker. An explanation might be the order of experimental conditions. In those cases the controllers started with speech recognition support and the mouse-only supported scenarios followed. We, therefore, calculated the mean throughput of all experiments which were performed as the first trial, as the second trial and so on (row "Observed Averages" in Table 12).

TABLE 12: CORRECTION VALUE TO COMPENSATE ORDER EFFECTS

Experiment Number	1	2	3	4	5
Observed Averages	32.9	34.3	43.5	39.3	40.9
Expected Averages	34.8	34.8	43.8	39.3	39.3
Correction Values	1.9	0.5	0.4	0.0	-1.6

As we only performed Complete Ap scenarios as trial number 1 and 2, we had to subtract the mean of all Complete Ap experiments. All trials with number 3 were with Final Ap, so we had to subtract the mean of all Final Ap experiments. Trials number 4 and 5 were equally mixed between Complete Ap and Final Ap. Thus, we subtracted the mean of all number 4 and 5 experiments (row "Expected Averages" in Table 12). The resulting "Correction Values" are added to each throughput value in Table 1. Table 13 summarizes all corrected weighted flow values.

TABLE 13: CORRECTED FLOW VALUES IN AIRCRAFT PER HOUR

	Complete Ap		Final Ap	
Controller	**ABSR+Mouse**	**Mouse**	**ABSR+Mouse**	**Mouse**
A	37.3 (1)	35.3 (4)	45.2 (5)	46.6 (3)
B	35.0 (2)	34.6 (5)	44.0 (3)	43.0 (4)
C	35.8 (4)	36.6 (1)	45.6 (3)	42.8 (5)
D	36.1 (5)	36.7 (2)	44.5 (4)	43.3 (3)
E	33.4 (1)	35.8 (4)	42.8 (3)	43.7 (5)
F	33.6 (5)	31.9 (2)	45.3 (3)	38.5 (4)
G	36.2 (4)	31.8 (1)	44.4 (5)	40.9 (3)
H	35.7 (2)	34.0 (5)	45.4 (4)	42.2 (3)

In the same way we correct flown distance, flight time and missing radar label information. The corrected overall efficiency is presented in Table 14.

TABLE 14: CORRECTED OVERALL EFFICIENCY VALUES

	Complete Ap		Final Ap	
Controller	**ABSR+Mouse**	**Mouse**	**ABSR+Mouse**	**Mouse**
A	97.2% (1)	72.7% (4)	59.1% (5)	93.0% (3)
B	91.6% (2)	65.4% (5)	66.4% (3)	51.2% (4)
C	79.6% (4)	87.5% (1)	88.8% (3)	62.4% (5)
D	82.3% (5)	85.7% (2)	68.9% (4)	64.1% (3)
E	60.4% (1)	65.7% (4)	42.2% (3)	49.8% (5)
F	61.2% (5)	14.3% (2)	56.6% (3)	14.0% (4)
G	85.8% (4)	33.8% (1)	74.5% (5)	27.7% (3)
H	87.2% (2)	65.9% (5)	82.8% (4)	44.8% (3)

By considering order effects we change the individual values of each experiment, but not the average values. Therefore, the time saving in the Complete Ap scenario is still 77 seconds per flight. Finally Table 15 shows the resulting α values, when the measurements are corrected due to order effects.

TABLE 15: MIN α FOR THE DIFFERENT HYPOTHESES AFTER ORDER EFFECT COMPENSATION

Hypotheses	Complete Ap	Final Ap	Both	Old Both
Increased flow	14.3%	2.5%	1.3%	3.7%
Flight distance	3.2%	24.1%	2.6%	4.3%
Flight time	2.9%	19.0%	1.9%	3.6%
Label info	0.5%	9.7%	0.2%	0.7%
Overall effic.	2.0%	5.7%	0.5%	1.6%

Adjusting for the order effect in this way shows an even stronger effect of ABSR over mouse-only (comparison with Table 11 (or the last two columns of Table 15)). We should emphasize that order effects are a reasonable explanation for the variation between different controllers. We checked the hypotheses, whether order effects exists for throughput, flight distance, addition flight time, radar label deviation and overall efficiency by an analysis of variances (ANOVA-test, i.e., F-Test) and could not falsify the null hypotheses. Only for the overall efficiency we got a tendency that the measurement depend on the sequence number (α=17%). We have two overlapping effects: The measurements depend both (slightly) on the sequence number and (significantly) on the input modality. The α-values should be expected between the values in Table 11 and Table 15. Additionally it must be emphasized that it is of course always better to control for order effects, instead of addressing them by calculations. One way of controlling would be just increasing the number of participants, which, however, would have violated our budget constraint.

The principle benefit of ABSR lies in the reduction of manual data input by ATCo, resulting in further availability of cognitive resources and reduction of head-down times. Benefits are concentrated on situations with high workload, high share of radio transmissions and a high rate of short term decisions concerning the control of air traffic. Therefore, speech recognition will – at first hand - mainly be of substantial use in approach units, serving high traffic airports. Benefits are also expected for ATC tower controllers that pose high-workload conditions.

Furthermore, ABSR bears the potential to enhance HMI processes in ATM systems, which integrate the air situation display together with relevant flight information on a single main window. ABSR in this context could also be used to form a migration path between radio transmission-based control and datalink-based control, enabling the system to send radio transmissions additionally as a datalink message to the addressed aircraft.

VI. AUTOMATIC SPEECH RECOGNITION: WHAT IS ACHIEVED AND WHERE RESEARCH STILL IS NEEDED?

While speech recognition in its ATM-tailored functionality itself shows a high degree of maturity and vicinity to first implementation steps, further applications of ABSR can be identified that need further R&D effort or prerequisite technological changes in the basic ATM system and the way communication between cockpit and ground is organized. Table 16 shows basic ideas of application of ASR at the controller working position in either upper or lower airspace control centers and explains the variety of maturity levels among different possible applications of ABSR.

The degree of maturity of the functionality is very different in relation to various possible applications. "ASR for pilots" in Table 16 includes detection of cockpit crew transmissions, adaption to different pilots and English skills. For automatic voice recognition of controllers AcListant®-Strips has shown that sufficient recognition rates are possible. Recognition of pilot utterances, however, will require addition research effort to allow for sufficient recognition rates. It is noteworthy that some of the applications expected to be most useful to support operations (e.g., automated input of releases to the ATM system) have a very high degree of maturity. Here, further efforts will focus on ensuring the coherence of the automated support with the working processes of the controllers and describing a migration path into the operational environment.

TABLE 16: POTENTIAL APPLICATION OF ABSR IN ATM IN RELATION TO REMAINING R&D EFFORTS AND TECHNOLOGICAL PREREQUISITES

Automatic Speech Recognition based functionality	Maturity concerning operational usability	Remaining R&D needs / Further technological prerequisites
Transfer of clearances to ATM System	Proven by project AcListant®	none
Warning, if voice clearances differ from system inputs	Basically proven by project AcListant	none

Marking of R/T addressed flights in the air situation window	Proven by project AcListant®	Usability aspects
Warning, if clearances will cause possible conflicts	Low level concerning alarm mechanisms and dialogue between different systems, ABSR-functionality proven by AcListant®	Interaction with conflict detection tools not yet described
ABSR controlled display functions (Weather, Sectors)	Low level + operational benefits not yet determined	Usability aspects
Bridge technology between radio transmissions and datalink (CPDLC)	Low level, status: idea	Relevant effort for concept, roadmap, system design, and interaction, etc. needed
Matching of clearances and read back	Low level	Comparison of clearances & read backs, ASR for pilots
Keyword by pilots triggers attention guidance (“Wake”, “Wind shear”, “Go Around”, “TCAS”)	Low level	Relevant R&D and development effort needed, ASR for pilots
Matching of weight category at initial call	Low level	None, ASR for pilots

TCAS = Traffic Collision Avoidance System; R/T = Radio Telephony

VII. CONCLUSIONS AND OUTLOOK

This paper concludes our work in the context of Assistant Based Speech Recognition (ABSR). In 2015, we demonstrated that command recognition rates better than 95% are possible with command error rates below 1.7% [1]. In the same paper we showed that speech recognition improves the adaptation speed of an Arrival Manager. In 2016, we published that ABSR significantly reduces controller workload [4]. Finally, this work shows that the workload reduction directly results in an increase of ATM efficiency with respect to number of movements per hour and aircraft flight time. For the Düsseldorf approach area (after Frankfurt and Munich the biggest airport in Germany) we quantified the benefits to 77 seconds reduced flight time and a throughput increase of one to two inbounds per hour. 77 seconds less flight time directly results in a fuel reduction of 50 to 65 liters per aircraft. This paper impressively demonstrates how a technology from research delivers benefits – measurable benefits – in air traffic control.

We present the maintenance of radar labels as the show case to quantify benefits of ABSR and examine the status of further applications of speech recognition to support ATM. In these areas further research is needed, but industry and ANSPs can already benefit now, given that ABSR could be integrated into the ops rooms. An interface between the TopSky ATM system and ABSR is still missing. ABSR implemented as research prototype also needs to meet prescribed levels related to software quality as it is a standard for all operational systems in aviation industry [47].

In order to reduce adaptation costs to approach area, different from Düsseldorf, DLR, Saarland University, Idiap Research Institute together with the air navigation service providers from Austria and Czech Republic have started the SESAR funded project MALORCA (Machine Learning of Recognition Models for Controller Assistance) in April 2016. This project aims at automatically learning models for recognition and for the Arrival Manager from recorded radar data and untranscribed controller pilot voice communication [48]. Prague and Vienna approach area are selected as demonstration airports to show that automatic learning from recorded radar and speech data reduces start-up costs.

ACKNOWLEDGMENT

The work was conducted in the AcListant® and AcListant-Strips project, which is supported by DLR Technology Marketing and Helmholtz Validation Fund. We also like to thank our reviewers Jon Rein (FAA William J. Hughes Technical Center, Atlantic City, NJ, USA) and Lucas Klamert (Austro Control) for their valuable inputs.

REFERENCES

[1] H. Helmke, J. Rataj, T. Mühlhausen, O. Ohneiser, H. Ehr, M. Kleinert, Y. Oualil, and M. Schulder, “Assistant-Based Speech Recognition for ATM Applications,” in 11th USA/ Europe Air Traffic Management Research and Development Seminar (ATM2015), Lisbon, Portugal, 2015.

[2] H. Gürlük, H. Helmke, M. Wies, H. Ehr, M. Kleinert, T. Mühlhausen, K. Muth, and 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.

[3] AcListant homepage: www.AcListant.de, AcListant = Active Listening Assistant, n.d.

[4] H. Helmke, O. Ohneiser, T. Mühlhausen, and M. Wies, “Reducing Controller Workload with Automatic Speech Recognition,” in IEEE/AIAA 35th Digital Avionics Systems Conference (DASC), Sacramento, California, USA, 2016.

[5] Speech Technology, “Nuance healthcare expands dragon medical portfolio,”http://www.speechtechmag.com/Articles/News/News-Feature/Nuance-Healthcare-Expands-Dragon-Medical-Portfolio-77155.aspx, 2011.

[6] S.W. Hamerich, “Towards advanced speech driven navigation systems for cars,” in Intelligent Environments, IE 07, 3rd IET International Conference, Sept. 2007, pp. 247-250.

[7] SRI International “Siri-based virtual personal assisstant technology,” http://www.sri.com/engage/ventures/siri, n.d.

[8] J. Schalkwyk, D. Beeferman, F. Beaufays, B. Byrne, C. Chelba, M. Cohen, M. Kamvar, and B. Strope, “Google search by voice: A case study,” in Advances in Speech Recognition: Mobile Environments, Call Centers and Clinics, Springer, 2010, pp. 61-90.

[9] DialRC, “The Dialog Research Center,” http://dialrc.org, n.d.

[10] C. Hamel, D. Kotick, and M. Layton, “Microcomputer System Integration for Air Control Training,” Special Report SR89-01, Naval Training Systems Center, Orlando, FL, USA, 1989.

[11] FAA, “2012 National Aviation Research Plan (NARP),” March 2012.

[12] D. Schäfer, “Context-sensitive speech recognition in the air traffic control simulation,” Eurocontrol EEC Note No. 02/2001 and PhD Thesis of the University of Armed Forces, Munich, 2001.

[13] R. Tarakan, K. Baldwin, and R. Rozen, “An automated simulation pilot capability to support advanced air traffic controller training,” in 26th Congress of the International Council of the Aeronautical Sciences, Anchorage, AK, 2008.

[14] S. Ciupka, “Siris big sister captures DFS“ original German title: “Siris große Schwester erobert die DFS,” transmission, Vol. 1, 2012.

[15] J.M. Cordero, M. Dorado, and 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, 2012, pp. 46-53.

[16] J.M. Cordero, N. Rodríguez, J.M. de Pablo, and M. Dorado, “Automated Speech Recognition in Controller Communications applied to Workload Measurement,” 3rd SESAR Innovation Days, Stockholm, Sweden, 2013.

[17] S. Chen and H. Kopald, “The Closed Runway Operation Prevention Device: Applying Automatic Speech Recognition Technology for Aviation Safety,” in 11th USA/ Europe Air Traffic Management Research and Development Seminar (ATM2015), Lisbon, Portugal, 2015.

[18] S. Chen, H.D. Kopald, A. Elessawy, Z. Levonian, and R.M. Tarakan, “Speech Inputs to Surface Safety Logic Systems,” in IEEE/AIAA 34th Digital Avionics Systems Conference (DASC), Prague, Czech Republic, 2015.

[19] S. Chen, H.D. Kopald, R. Chong, Y. Wei, and Z. Levonian, “Read Back Error Detection using Automatic Speech Recognition,” in 12th USA/ Europe Air Traffic Management Research and Development Seminar (ATM2017), Seattle, WA, USA, 2017.

[20] H. Said, M. Guillemette, J. Gillespie, C. Couchman, and R. Stilwell, “Pilots & Air Traffic Control Phraseology Study”, International Air Transport Association, 2011.

[21] S.R. Young, W.H. Ward, and A.G. Hauptmann, “Layering predictions: Flexible use of dialog expectation in speech recognition,” in Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI89), Morgan Kaufmann, 1989, pp. 1543-1549.

[22] S.R. Young, A.G. Hauptmann, W.H. Ward, E.T. Smith, and P. Werner, “High level knowledge sources in usable speech recognition systems,” in Commun. ACM, vol. 32, no. 2, Feb. 1989, pp. 183-194.

[23] H. Helmke, R. Hann, M. Uebbing-Rumke, D. Müller, and D. Wittkowski, “Time-based arrival management for dual threshold operation and continuous descent approaches,” in 8th USA/Europe ATM Research and Development Seminar (ATM2009), , Napa, CA, USA, 2009.

[24] A. Schmidt, “Integrating Situational Context Information into an Online ASR System for Air Traffic Control,” Master Thesis, Saarland University (UdS), 2014.

[25] Y. Oualil, M. Schulder, H. Helmke, A. Schmidt, and D. Klakow, “Real-Time Integration of Dynamic Context Information for Improving Automatic Speech Recognition,” Interspeech, Dresden, Germany, 2015.

[26] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, et al., “The Kaldi Speech Recognition Toolkit,” in IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, Hilton Waikoloa Village, Big Island, Hawaii, US, Dec. 2011.

[27] KALDI homepage: www.kaldi-asr.org, n.d.

[28] T. Shore, F. Faubel, H. Helmke, and D. Klakow, “Knowledge-Based Word Lattice Rescoring in a Dynamic Context,“ Interspeech 2012, Sep. 2012, Portland, Oregon.

[29] T. Shore, “Knowledge-based word lattice re-scoring in a dynamic context,” Master Thesis, Saarland University (UdS), 2011.

[30] H. Helmke, H. Ehr, M. Kleinert, F. Faubel, and D. Klakow, “Increased Acceptance of Controller Assistance by Automatic Speech Recognition,” in 10th USA/ Europe Air Traffic Management Research and Development Seminar (ATM2013), Chicago, IL, USA, 2013.

[31] C.S. Jordan and S.D. Brennen, “Instantaneous self-assessment of workload technique (ISA),” Defence Research Agency, Portsmouth, 1992.

[32] S.G. Hart and L.E. Staveland, “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” in Human mental workload, P.A. Hancock and N. Meshkati, Eds., Amsterdam, North-Holland, pp. 139-183.

[33] W.B. Knowles, “Operator Loading Tasks,” in Human Factors, 1963.

[34] G. D. Ogden, J. M. Levine, and E. J. Eisner, “Measurement of Workload by Secondary Tasks,” in Human Factors, 1979.

[35] N. A. Stanton, P. M. Salmon, G. H. Walker, C. Baber, and D. P. Jenkins, “Human Factors Methods - A Practical Guide for Engineering and Design,” Ashgate, 2005.

[36] H. Ayaz, B. Willems, S. Bunce, P. A. Shewokis, K. Izzetoglu, S. Hah, A.R. Deshmukh, B. Onaral, “Cognitive Workload Assessment of Air Traffic Controllers Using Optical Brain Imaging Sensors,” in Advances in understanding human performance: Neuroergonomics, human factors design, and special populations, CRC Press Taylor & Francis Group, 2010, pp. 21-31.

[37] P. Aricò, G. Borghini, G. Di Flumeri, A. Colosimo, I. Graziani, J.-P. Imbert, et al., “Reliability over time of EEG-based mental workload evaluation during Air Traffic Management (ATM) tasks,” in Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE, 2015.

[38] SESAR Concept of Operation Step 1, Edition 01.00.00, 2012

[39] BDLI: "BDLI technology roadmap for civil aviation technology in Germany", in German "BDLI-Technologie-Roadmap für die zivile Luftfahrttechnologie in Deutschland“, September 2015.

[40] German Government: "Aviation Strategy of Federal Government“, in German ”Die Luftfahrtstrategie der Bundesregierung“, Federal Ministry for Economics and Energy, March 2014.

[41] SESAR Joint Undertaking, “Second draft of the SESAR 2020 R&I Programme”, Edition 0.21, July 2014.

[42] COOPANS homepage: www.coopans.com, n.d.

[43] H. Helmke, O. Ohneiser, M. Wies, and M. Kleinert, “AcListant-Strips: Validation Results of Main Trials,” internal report, DLR-IB-FL-BS-2016-19, Braunschweig, 2016.

[44] Euroncontrol, “E-OCVM Version 3.0 Volume I – European Operational Concept Validation Methodology”, Februar 2010.

[45] O. Ohneiser, H. Helmke, H. Ehr, H. Gürlük, M. Hössl, T. Mühlhausen, Y. Oualil, M. Schulder, A. Schmidt, A. Khan, and D. Klakow, “Air Traffic Controller Support by Speech Recognition,” in Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014, Advances in Human Aspects of Transportation: Part II, N. Stanton, S. Landry, G. Di Bucchianico, and A. Vallicelli, Eds. Krakow, Poland: CRC Press, 2014, ISBN: 978-1-4951-2098-5, pp. 492-503.

[46] M. Wies, “AcListant-Strips: Validation Plan,” in German, DLR; IB-112-2015/50, Braunschweig, 2015.

[47] EUROCAE “ Guidelines for ANS software safety assurance,” ED 153, 2009 Edition, August 2009.

[48] MALORCA homepage: www.malorca-project.de, MALORCA = Machine Learning of Recognition Models for Controller Assistance, n.d.

AUTHOR BIOGRAPHY

Jörg Buxbaum leads the “R&D Air Traffic Management Team” at DFS Deutsche Flugsicherung GmbH in Langen (Germany). He received his Diploma degree in Aerospace Engineering in 1997 after studying at Technische Universität Berlin and Arizona State University (USA). Subsequently, he joined Frankfurt airport operator FAG, where he accounted for optimization of runway utilization. In 1999, he began at DFS as assistant of the Managing Director Technical Operations, followed by different executive functions in technical and operational responsibilities, e.g., head of Frankfurt approach control unit.

Hartmut Helmke received his Diploma degree in Computer Science from the University Karlsruhe (Germany) in 1989 and his doctor degree (PhD) from the chemical engineering faculty of the Technical University of Stuttgart in 1999. In 1989 he joined DLR’s Institute of Flight Guidance in Braunschweig. Since 1999 he concentrates on Controller Assistant Systems, especially Arrival Management being responsible of the Arrival Manager 4D-CARMA and lead the AcListant® and AcListant®-Strips project. Now he leads the MALORCA project [48]. Prof. Helmke is an assistant professor for Computer Science since 2001.

Christian Kern joined Austro Control in 1997 and is the Manager of the Vienna approach control unit. In 2013 he received his master degree in Business Administration with a specialization on Air Traffic Management from the Danube University of Krems (Austria).

Oliver Ohneiser received his master degree in Computer Science (2011) and his doctor degree (PhD) in Aerospace Engineering (2017) both from the Technical University of Braunschweig (Germany). He joined DLR in 2006 and is a scientific assistant in the department “Controller Assistance” of the Institute of Flight Guidance in Braunschweig since 2010. Mr. Ohneiser is lead of SESAR’s “PJ16.04 CWP HMI” Automatic Speech Recognition activity.