Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013)

Increased Acceptance of Controller Assistance by Automatic Speech Recognition

Hartmut Helmke, Heiko Ehr,
German Aerospace Center (DLR),
Institute of Flight Guidance,
Braunschweig, Germany
Hartmut.Helmke@dlr.de;
Heiko.Ehr@dlr.de

Matthias Kleinert
Ostfalia,
University of Applied Science,
Wolfenbüttel, Germany,
M.Kleinert@ostfalia.de

Friedrich Faubel, Dietrich Klakow
Saarland University (UdS)
Department of Computational Linguistics
and Phonetics, Saarbrücken, Germany
Dietrich.Klakow@lsv.uni-saarland.de
Friedrich.Faubel@lsv.uni-saarland.de

Abstract—**Situation awareness of today's automation relies on sensor information, data bases and the information delivered by the operator using an appropriate HMI. The situation is mostly influenced by voice communications between controller and pilots. Hence, voice communication is an important part for the human operator to implement his plans. Voice communication runs independent and in parallel to the process the automation performs to understand the situation. Therefore, the automation, specifically the support system, is not aware of agreements between human operators. Even worse, the operators have additional effort to inform the support systems about their communication, i.e. their intents. This additional effort can be avoided by using automatic speech recognition systems (ASR). Nowadays, ASR is used in many applications, e.g. Siri® in Apple's iPhone®.**

This paper focuses on the integration of ASR with DLR's arrival manager 4D-CARMA. ASR improves situation awareness of both assistant system and controller. As the controller is responsible for his advisories he sometimes deviates from the recommendations of the automation. The automation often needs at least 40 seconds until it recognizes the deviations from the plan if radar data is available only. Trials performed at DLR's Institute of Flight Guidance have shown that ASR can reduce this deviation time of an Arrival Manager (AMAN) by approx. 90% down to 5 seconds.

As a side-effect, the combination of ASR and AMAN also improves the performance of ASR. The AMAN provides context information about the current and estimated future situations. It creates hypotheses on controller intents and predicts which advisories the controller will probably transmit via voice. First trials have shown that this approach can reduce the word error rate by up to 80%. This can foster the use of ASR in ATM, e.g. as an enabler for the introduction of electronic flight strips.

Keywords- Arrival Management (AMAN), Automatic Speech Recognition (ASR), Situation Awareness (SA)

I. INTRODUCTION

A significant part of human collaboration is coordinated via voice, especially if complex contexts or meta-concepts are involved. By tracing the communication new actors can get an idea of the actual and planned situations so that they can easily integrate themselves. Controlling aircraft in the vicinity of an airport is an example of such a working environment in which two working groups – i.e. pilots and controllers – implement a smooth, efficient and safe traffic flow via radio communication. All pilots in the same sector are controlled by a dedicated controller (team). They use a unique frequency for communication within the sector. This enables the party line effect, i.e. all actors – excluding today's assistant systems – can create a common mental model of the current situation and of future actions.

Various taxonomies have been proposed for automation levels; Parasuraman [23] distinguishes between ten levels of automation where complete departure from human intervention occurs at level 10. Today's ATM systems do not exceed level 3 or 4; hence the operators remain fully in control. This situation accounts for the whole program time frame of SESAR, although both SESAR and NextGen envisage enhanced ground and airborne automation to enhance capacity and maintain safety. According to the Strategic Research & Innovation Agenda (SRIA) of ACARE [1] or Flightpath 2050 [11], a fully automated ATM-System is not considered in the next decades. So voice communication will definitely remain a pillar of air traffic control. SRIA further suggests using Speech Recognition for thread detection of the air traffic system.

One dimension of today's increased system complexity results from the fact that the automation does not follow human communication. This splits the communication into two different worlds: one, in which humans communicate via radio links, and another one, in which machines communicate via computer networks. These worlds are connected by a human machine interface through which humans inform the machines and vice versa. As controllers are responsible for air traffic control they implement their plans even if these deviate from those of the automation. As these deviations especially occur in situations with high workload, the controllers do not have time to inform the assistant system. In this case, the automation may suggest advisories contrary to the intent of the controller. This situation

may persist until the assistant system realizes the deviation through analysis of radar data. Hence, the system needs support from the controllers exactly when the controllers would urgently need the support of the system, due to a high workload.

In this paper, we show how ASR can cut this Gordian knot without additional effort for the operators. This enables common situation awareness without lack of information on the part of the automation and without discrepancies between voice communications and data link information. The following three main functions are required for this purpose:

1. Creation of hypotheses about desired future airspace situations and the corresponding commands
2. Highly reliable speech recognition based on dynamic language models
3. Updating of assistant systems based on the obtained voice communication information

The integration of Automatic Speech Recognition (ASR) into ATM systems has been tried since (at least) the early 90s. We briefly review this prior work in section II. This is followed by section III, which demonstrates how the ASR performance can be improved through use of context information from the assistant system. Section IV is the counterpart which explains the expected benefits for an arrival manager if the assistant system is supplemented by an additional sensor, i.e. ASR.

Section V describes experiments that have been performed with a combination of DLR's arrival manager 4D-CARMA (**4D C**ooperative **ARr**ival **MA**nager), the speech recognizer from Saarland University (UdS) and an approach controller from Austro Control. In these experiments, 4D-CARMA was used for passive shadow mode trials in which 4D-CARMA created sequences and advisories, which were not shown to the controller (shadow mode trials). Therefore, the controller occasionally deviated from the advisories of the automation. If this happened the AMAN often needed at least 40 seconds until it recognized the deviating plan of the controller. Section VI shows that ASR can reduce this deviation time down to 5 seconds. Section VII describes further steps to improve ASR and AMAN performance by exchanging context and intent information before section VIII finally summarizes the results.

II. Related Work

Due to rising demand more sophisticated assistant systems are introduced to support ATM operations, e.g. Arrival Managers (AMAN), Surface Managers (SMAN), and Departure Managers (DMAN). First commercial implementations of an AMAN have been operational at hubs (Frankfurt, Paris) since the early 90s. Today their application is still limited to the coordination of traffic streams between different working positions (e.g. sector and approach controllers) [16]. Implementations in Europe are e.g. OPTAMOS [3], OSYRIS [4], 4D Planner [6], [13], MAESTRO [10]. An extension of their application to support the controllers by advisories in order to implement fuel and noise efficient approaches (e.g. DLR's AMAN 4D-CARMA [17]) currently fails due to insufficient reliability of the advisories. The support quality of such systems highly depends on knowledge of the development of the situation in the airspace. This development is characterized by the intent of the controllers. When controllers deviate from the assistant system the assistant system may need some time to firstly, recognize these deviations and to secondly, determine the controllers' intent. During this deviation time, the support quality is strongly reduced. As controllers tend to deviate from the machine strategy when the situation becomes complicated, the support quality is low exactly when it is really needed. The explicit integration of the assistant system into human communication can compensate this effect. The result of such an approach is an assistant system that actively listens. Facilitating active listening requires highly reliable speech recognition to avoid an increase of controllers' workload due to a high amount of needed corrections.

In the past decade three meanwhile classical application areas of speech recognition have evolved:

- command & control (e.g. for mobile phones, TV sets or navigation systems in cars) [15],
- dictation systems (predominantly for the professional market, as the adaptivity is not yet good enough for widely accepted consumer products) [29],
- spoken dialog systems to access information (e.g. for train time table information) [7]. The recently launched Siri® system is essentially a question answering and spoken dialog system using presumably a Nuance speech recognition engine. Similar systems from Google and Samsung exist. The SmartWeb project completed in 2007 built a first working system for this type of application [28].

At the moment speech recognition is far from being perfect. It still faces a number of problems and challenges. The most commonly-used metric for evaluating ASR is the word error rate (WER). That is a metric of the distance between the word label sequence output of the ASR system and the sequence s of words which were actually said, the gold standard (see pp. 362-364 in [19]). The WER is defined as a derivation of Levenshtein distance [21]:

$$WER(s) \equiv \frac{\mathrm{ins}(s) + \mathrm{del}(s) + \mathrm{sub}(s)}{W(s)}$$

Here, ins(s) is the number of word insertions (words never spoken), del(s) is the number of deletions (words missed by ASR), sub(s) is the number of substitutions needed to align the two sequences, and W(s) is the number of words actually said. Another commonly used metric for evaluating ASR is the sentence error rate (SER), which is the rate of sentences having at least one error (i.e. the rate of not perfectly recognized sentences). Although WER and SER are often related, this is not always the case. Generally, the SER increases with the WER, but one cannot be inferred from the other.

The WER strongly depends on the application context and the recording conditions. Current systems have a WER between 15% and 35% for conversation via telephone lines. Digit string recognition via headset can achieve a word error rate of less than 1% (under optimal conditions). If there is background noise, however, the WER can easily rise to 30% or 50% at a

signal to noise-ratio of 0 dB. Participants of the second Pascal Speech-Separation-Challenge even had to accept a WER of 50%.

In ATM, it is not important that every word is recognized. It is important that the detected concept is correct. It is not important that ASR correctly recognizes "Good morning Lufthansa one two tree descend level one two zero", but that the concept "DLH123 DESCEND FL 120" is extracted. The concept error rate (CER) quantifies this metric. In natural speech processing, the CER is usually smaller than the SER, because natural speech is redundant. ATM advisories are not very redundant and the presented approach will use the context information of the assistant system to already reduce the word error rate. Therefore, CER and SER have comparable values.

ASR is not new in the context of ATM. Hamel [14] and Weinstein [30] have described the application of speech technology in ATC training simulators in the early 90s, however with limited success. Dunkelberger et al. [8] described an intent monitoring system which combines ASR and reasoning techniques to increase recognition performance: In a first step, a speech recognizer analyses the speech signal and transforms it into an N-best list of sentence matches. The second step uses context information to reduce the N-best list. Schäfer [24] applied ASR in the context of an ATC simulation environment to replace simulation pilots (so called *pseudo pilots*) in validation trials with controllers. He used simulation data to predict what a given aircraft's future status will be (e.g. at which altitude and airspeed it will be flying), and which possible conflicts may occur that need to be resolved by the controller.

Currently, different successful applications of ASR in ATC training simulators are available. The FAA reports the successful usage of advanced training technologies in the terminal environment [12]. The Terminal Trainers prototype, developed by CAASD (Center for Advanced Aviation System Development, USA), includes voice synthesis, speech recognition, multimedia lessons, game-based training techniques, simulation, and interactive training tools. The German air navigation service provider DFS uses the system *Voice Recognition and Response* (VRR) of UFA (Burlington, MA) for controller trainings since August 2011 [5]. Less pseudo pilots are needed in its flight service academy. DFS, however, also reports that these systems are currently only usable for training purposes, because they only accept standard phraseology [5].

The presented paper aims at a completely new application area: that of an overhearing speech recognition system which is embedded in a physical context. In contrast to existing applications, there is a human-human communication and the system is listening in to derive knowledge about the plans of the human participants. This aspect is somewhat similar to the goals of the AMIDA project in which meetings were overheard and summaries were generated [2]. The new aspect of the proposed approach is that the speech recognition system is part of an agent (assistant system) that has knowledge about the physical world and at the same time makes plans about actions, which are then proposed to the human participants. Thus, one novelty is that the overheard human-human communication is directly used to improve the assistant system. The other novelty is more of a by-product: the physical context of the planning system (e.g. the aircraft in airspace) is used to improve speech recognition. This is based on the fact that not all interpretations of a speech utterance are equally likely in a given physical context.

III. Benefits of AMAN for ASR

Recent advances in ASR technology continue to be based heavily on data-driven methods, meaning that the full benefits of such research are often not enjoyed in domains for which there is little training data available. In 2011 UdS performed experiments with 16 different speakers to circumvent this problem by using dynamic contextual knowledge to rescore ASR lattice output using a dynamic weighted constraint satisfaction function [25]. We summarize these results in this chapter.

A. Lattice Rescoring

ATC commands are issued using standardized phraseology [9]. In its simplest form a commands consists of a callsign (e.g. DLH496) followed by a goal action (e.g. descent) and a goal value (FL90). Although the communication is standardized, the controller-pilot communication is not easy to follow. The IATA phraseology report [18] shows that controllers often deviate from standard phraseology especially in high workload situations. In our approach, a weighted finite state transducer (WFST) decoder creates a context-dependent phone-to-word transducer lattice, as shown in Figure 1.

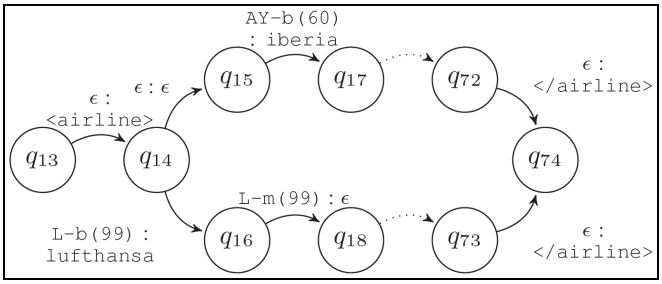


Figure 1. Part of a context-dependent phone-to-word transducer lattice [22] with embedded XML tags. The L-b(99), L-m(99) and AY-b(60) symbols indicate triphone clustered acoustic states. The dotted edges indicate omitted nodes.

There are costs associated to each of the edges in Figure 1 (although these are not shown for the sake of readability). The most probable speech recognition hypothesis is defined by the lowest scoring path through the given phone-to-word transducer lattice. Using context information from the AMAN, however, means to rescore the edges, i.e. penalizing hypotheses which are invalid or unlikely in the context in which the utterance was made: For example, the above command "DLH496 descend FL90" would be penalized if there is no aircraft with that callsign in the sector. Likewise, such a command would be unlikely if DLH496 exists, but is already flying at FL80. Details concerning the used rescoring can be found in [25] and [26].

B. Performed Experiments

16 people – all of them no ATC experts – participated in a study in 2011. Self-identified by first language, eight were German speakers, three were North American English speakers, and there were two Greek, one Malayalam, one Romanian

and one Russian speaker. Twelve of the speakers were male and four were female. An approach scenario with 31 inbounds for Frankfurt airport using only one runway was created. 4D-CARMA was used to create sequences and ATC commands which were displayed to the participants (in English). The probands read the commands aloud. These voice commands were recorded using a headset. The commands had no effect on the simulation. They were only used for speech recognition purposes. 1,107 ATC commands were recorded in total, with an average length of 9.5 words per sentence; this corresponds to approximately 100 minutes of speech. Each individual recorded utterance was annotated not only with the true sentence that was read but also with the state of the entire ATC simulation at the time of the recording (e.g. the aircraft on the radar, including theirs speeds, altitudes, heading, position in relation to the radar, etc.). The corresponding aircraft state vectors were retrieved from 4D-CARMA every 5 seconds and then stored to a log file.

As evaluation metrics for speech recognition, both WER and SER were used. In order to measure the improvement through rescoring, we also calculated the mean reciprocal rank (MRR):

$$MRR(Y) = \frac{1}{|Y|} \sum_{y \in Y} \frac{1}{\text{rank}(y)}$$

Here, Y denotes the complete set of utterances (i.e. the set of given commands). The rank of each utterance y is determined as follows: If a command y1 is recognized correctly, i.e. y1 is the highest-scoring hypothesis in the word lattice then rank(y1) is 1. If a command y2 is not recognized correctly and the hypothesis, that this command was given, is only the third best hypothesis in the lattice, then rank(y2) is 3, and so on.

C. Results

TABLE 1 shows the speech recognition results before and after rescoring. The first row indicates the baseline without using context information.

TABLE 1 WER, SER, MRR of experiment

Constraints Used	*WER*	*SER*	*MRR*
none (baseline)	2.81%	22.58%	0.849
constraint "callsign"	0.55%	4.61%	0.966
constraint "callsign, speed, altitude"	0.52%	4.52%	0.967
oracle (best possible)	0.31%	2.07%	0.979

Word resp. sentence error rate, mean reciprocal rank before and after rescoring with constraints

The second row gives results for using the information of available callsigns in the controller's sector. The third row shows the results with additional speed and altitude context information. The last row ("oracle") shows the best possible results that could theoretically be obtained with an optimal rescoring algorithm. This error rate, however, is not 0%, because sometimes the correct word sequence is not contained in the lattice so that rescoring cannot have a positive effect.

Most notable is that the callsign constraint already gives a significant improvement in both word and sentence error rate, with relative reductions of 80% over the baseline. This can be explained by the fact that the callsign constraint effectively reduces the total number of theoretically valid callsigns from more than 200.000 to the number of aircraft which are currently on the controller's frequency.

We will show in the next chapter that a reliable ASR significantly improves the performance of the assistance system. We should, however, keep in mind, that we only considered simple commands. Combined reduce and descend commands, which also contain a heading or frequency change command, were not considered. Without further actions in these cases, the error rate increases significantly. The results, however, show the potential to increase ASR performance if an AMAN provides appropriate context information. Section VII describes our future steps to further improve ASR and AMAN performance.

IV. Expected Benefits of ASR for AMAN

4D-CARMA assigns a sequence number to each inbound. Subsequently, target times are determined at different waypoints (e.g. runway threshold). Trajectories, which meet the target times, are calculated by 4D-CARMA's trajectory predictor and transformed into appropriate controller advisories (see [17] for further details). Input data is the current traffic situation, i.e. radar data and both given and planned advisories. Different input sources for the given advisories are possible; e.g. ASR, mouse or keyboard inputs of a controller. 4D-CARMA monitors the actual situation by comparing it to the planned situation (task of conformance monitoring). If a deviation is detected the challenge is to decide whether the controller intentionally deviates from the plan or whether he tries to implement the actual plan. In the first case a replanning is necessary. If 4D-CARMA takes the wrong decision, due to the needed trade-off between adaptability and stability, the planning of 4D-CARMA differs from the real situation, i.e. from the mental model of the controller. Hence, the exclusive use of radar data can cause a temporary deviation of both worlds, which may last for 40 seconds and more. Even rescheduling will not help in this case, because the reason for the deviation is unknown.

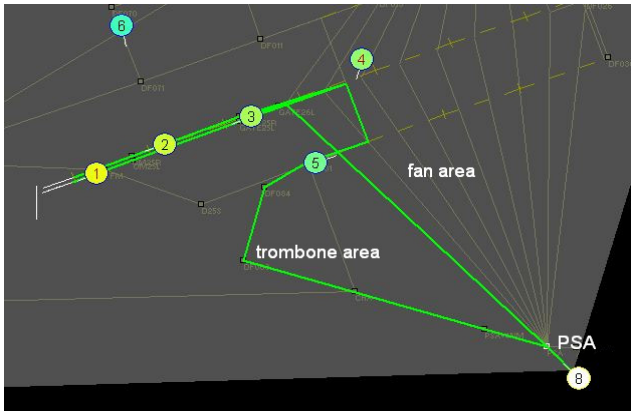


Figure 2. Different path stretching options (fan or trombone routes) exist for the controller

In the following, we show some examples of how the usage of automatic speech recognition can reduce this deviation time down to 5 seconds. For each aircraft, 4D-CARMA and the controller can choose a fan or trombone approach, as shown in

Figure 2. In the first case the controller will give a heading instruction between e.g. 310 and 50 degrees. In the latter case the command will be "Follow Transition". This depends on the traffic situation and also on the individual controller. Let us assume 4D-CARMA plans a trombone approach for the aircraft with sequence no. 8. Without further information 4D-CARMA does not know the controller's intent until the aircraft is one mile behind PSA, the initial approach fix (IAF). By integration of ASR into the AMAN, 4D-CARMA is aware of the heading advisories, before the aircraft is at the IAF. Hence, a rescheduling will happen, which results in a smaller sequence number and an adapted trajectory.

Figure 3 shows a situation which results from a conflict between BMA419 and DLH123. Without further interaction by the controller, 4D-CARMA assumes that BMA419 will be first (sequence no. 11). The controller may however choose to switch this order and therefore reduce the speed of BMA419. Without knowing the advised speed, 4D-CARMA needs approx. 40 seconds to recognize the intended sequence change. If 4D-CARMA gets access to the target speed values the sequence is immediately adapted to the desired sequence. The same improvements occur if descent target values are known.

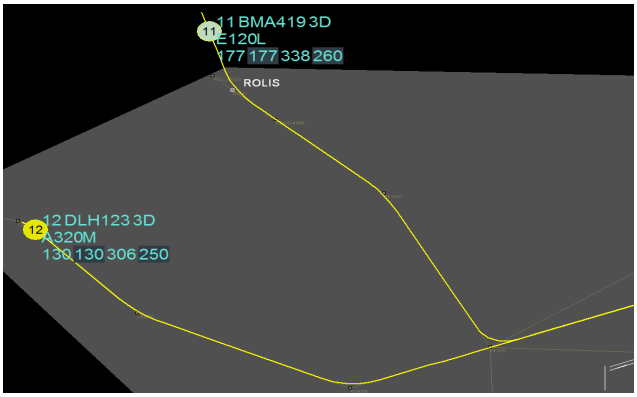


Figure 3. Sequence change caused by reduce advisory

Figure 4 shows an example of a transition approach.

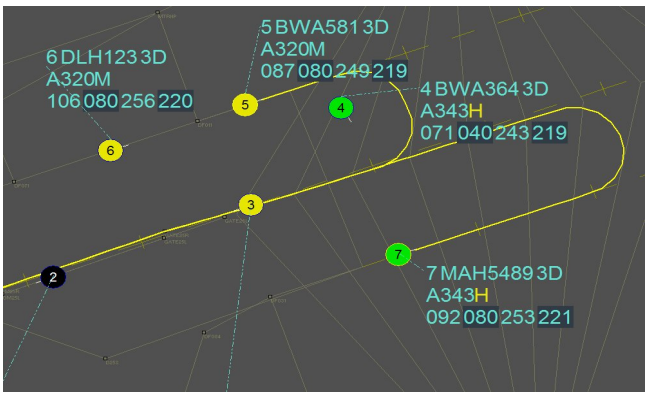


Figure 4. Turn advisories for MAH5489 initiate sequence change

The AMAN assumes that MAH5489 is planned behind DLH123. The controller may decide to change this such that MAH5489 is no. 5 in sequence. In that case the controller will shortly give a turn left command to MAH5489. The AMAN, however, will not recognize the changed intent until the track of MAH5489 significantly changes in the radar data. All in all, considering the read-back time of the pilot, reaction time to start the turn and radar update rate, this may last up to 25 seconds. With ASR we are able to update the sequence immediately after recognizing the voice command.

Conclusion: Extending 4D-CARMA's input by the given controller commands can significantly improve the adaptation speed of the AMAN, which ultimately increases its planning stability.

V. The Experiments

While the previous section has described our expectations, this section describes the performed experiments: (1) evaluation of historical radar data, in which 4D-CARMA was used for passive shadow mode trials, (2) a human-in-the-loop simulation in which 4D-CARMA was supported in shadow mode by ASR. The section closes with a description of the derived measurements which are used to validate the expectations.

A. Passive shadow mode trials with 4D-CARMA

Recorded radar data of approaches to Frankfurt and Cologne/Bonn were used. Two experiments were performed with each data set. Each data set consisted of approx. 30 minutes of radar data along with flight plan information. 4D-CARMA updated its planning every 5 seconds based on the recorded radar data, i.e. it generated new sequences with updated target times, calculated trajectories and generated advisories. As 4D-CARMA ran in passive shadow-mode, its planning output had no effect on the input radar data of future planning cycles. We investigated, how fast and how well the internal picture of 4D-CARMA matched that of the controller. This was used as a baseline scenario without speech recognition.

To evaluate the possible benefits of ASR without available recorded speech data, we simulated the *perfect* speech recognizer by preprocessing the radar data in advance. If the altitude, track or indicated airspeed of the aircraft radar data changed significantly, we generated an appropriate controller advisory and assumed that this advisory was voiced by the controller and subsequently recognized by ASR. Assuming perfect ASR is reasonable because we *only* want to evaluate the possible benefits of ASR concerning the performance of the assistant system.

B. Human-in-the-Loop Simualtions (HITL)

In addition to passive shadow mode trials, HITL simulations were performed with a controller of Austro Control. This was done in June and December 2012, on three different days (Figure 5). The first day was used to (1) check the system, (2) discuss the project and (3) adapt the speech recognizer to the controller's pronunciation. During the adaptation, the controller read approx. 100 predefined commands so that a speaker dependent recognition model could be created. This slightly improved the recognition rate compared to a speaker independent

model. In the following we describe the experiments that were performed in December.

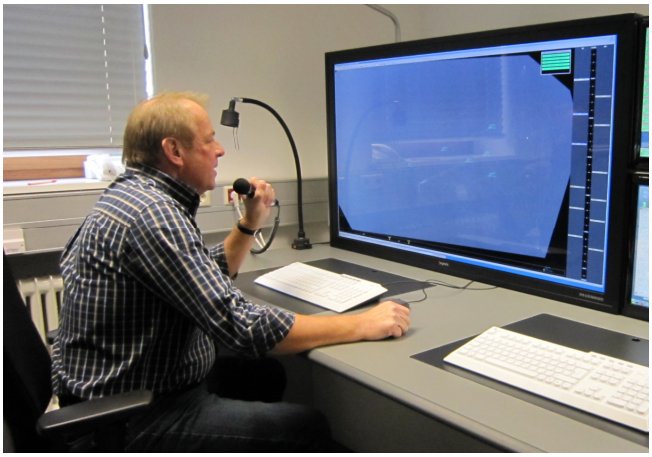


Figure 5. Controller in working environment

For the test runs, the controller was advised to refer to the radio discipline. Multiple instructions and deviations from ICAO phraseology should not be used. His job was to act as a feeder for Frankfurt approach being responsible for all aircraft on the downwind. All aircraft entered the downwind below flight level (FL) 100 and below 250 knots indicated airspeed (IAS). The sector, pick-up and tower controllers were simulated. Only reduce, descend and turn-to-base advisories were allowed. Figure 6 shows the general set-up of the experiment. The controller gave voice commands which were then processed by the speech recognizer. The recognized commands were stored in a data base and used by the simulator to update the radar data. This stands in contrast to the passive shadow mode trials, as 4D-CARMA resp. the controller was able to influence the aircraft's trajectories in each run. Based on the radar data 4D-CARMA updated its sequences and trajectories in the data base and displayed them on the radar screen.

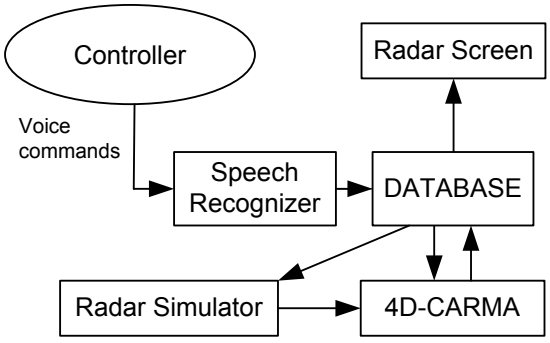


Figure 6. Involved actors of HITL simulation

4D-CARMA did not provide any context information to the speech recognizer in this experiment. Due to the simple grammar allowed in the experiment 96.4 percent of the given commands were recognized correctly. As a fail-safe measure, 4D-CARMA checked the recognized advisories for consistency. Inconsistent advisories were removed from the data base so that they did not influence the radar simulator. Throughout the simulation, only one wrong command was not detected by this additional check (REDUCE Speed 250 instead of 240).

The hardware used in the test runs was installed on the controller working position (CWP) of the DLR's Institute of Flight Guidance. The planning system 4D-CARMA and the radar screen ran on 3 Fujitsu Celsius M470-2 power workstations with Intel Xeon CPU and 6 GB RAM. The processes for speech recognition ran separately, on a similar PC. Windows XP® respectively SUSE Linux 11.2 were the used operating systems.

The controller's human machine interface (HMI) of 4D-CARMA (called RadarVision) is an experimental radar screen that provides all information of the actual traffic situation. 4D-CARMA consists of several modules (e.g. scheduler, trajectory predictor), which are implemented as independent Linux processes. A MySQL data base was used for process interaction. The speech recognizer monitored the speech communication loop between controller and pilot.

C. *Derived measurements*

We defined the following criteria to compare the quality of the simulation runs with and without information about the given controller commands:

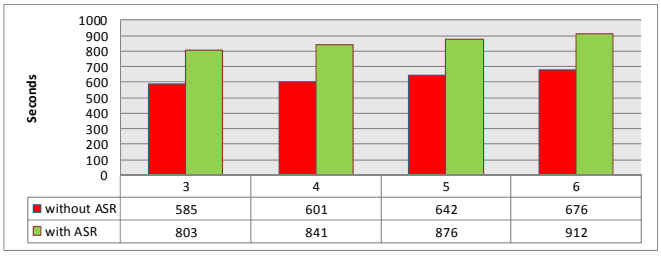
- The time until subsequences were stable (SS-3, SS-4, …) – We define the landing sequence as the order in which aircraft actually touch down and consider subsequences of successive aircraft of the landing sequence of size M. For a landing sequence of size N we can consider N-M+1 subsequences. For each of these subsequences the time is determined until the order of the M elements of the planned subsequence matches with the landing subsequence and is not changed until touchdown of the whole subsequence. We measure the time in seconds until the landing of the last airplane in the subsequence. This derived measurement gives a hint concerning the stability of the AMAN.
- The Non-Conformance Time (NConfT) and Non-Conformance Counter (NConfCnt) – 4D-CARMA determines for each aircraft if the radar data is conform to the actual planned trajectory. The conformance monitoring considers lateral deviations (> 0.5 NM), vertical deviations (> 500 ft.), and temporal deviations (> 10 seconds). Based on these deviations each aircraft gets the status conform, half-conform or non-conform. The status half-conform is assigned if the thresholds are only slightly violated (lateral deviations > 0.25 NM, vertical deviations > 250 ft. or temporal deviations > 5 seconds) and the aircraft is still not in status non-conform. We calculate for each aircraft the total times NConfT (resp. NHConfT) the aircraft is in status non-conform (resp. half-conform) and how often the status changes from conform to non-conform (NConfCnt). These measurements indicate how long resp. how often the internal picture of the controller differs from that of the machine.

VI. RESULTS

A. *Passive shadow mode trials for Frankfurt TMA*

As 4D-CARMA was used in passive shadow mode, it often deviates from the sequence and trajectories actually implemented by the controllers. The fact of many observed deviations is independent whether 4D-CARMA knows earlier in advance the controllers target values, i.e. ASR is available or not. The main point is how often the deviations occur respectively how long they happen.

Figure 7 shows the improved planning stability. Without ASR the AMAN knows 676 seconds (approx. 11 minutes) before touchdown the correct sequence if we consider subsequences with 6 aircraft. With support of ASR this time increases to 912 seconds (approx. 15 minutes).



	3	4	5	6
without ASR	585	601	642	676
with ASR	803	841	876	912

Figure 7. Average values for SS-3, SS-4, SS-5, SS-6 (N=17)

Figure 8 visualizes the differences of these 4 minutes for the controller. The aircraft with sequence no. 8 is 11 minutes before touchdown; no. 11 is 15 minutes before touchdown.

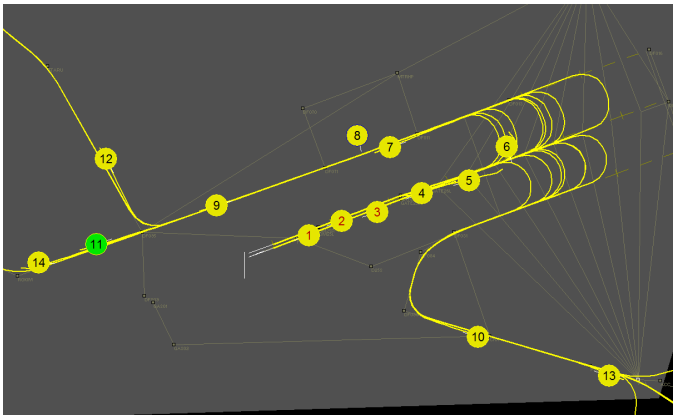


Figure 8. Planned sequence with advisories,

Figure 9 shows the time, when the aircraft are not conform to their planned trajectory (NConfT, NHConfT). We see that in 14% of the time the aircraft are not conform to their trajectory. This high value is based on different effects:

- The unknown wind. As we used historical data from Frankfurt airport the wind speed and directions were unknown on that day. This complicated the calculation of the indicated airspeed, which was advised by the controller, from the ground speed and altitude values.
- The controllers heavily used vectoring to separate the aircraft form each other, resulting in many lateral deviations.
- The controller had no chance to implement the plan of 4D-CARMA (passive shadow mode).

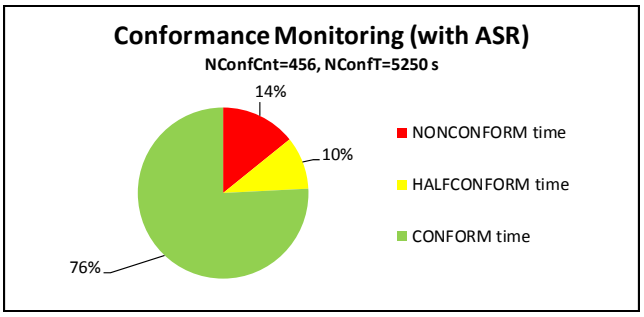


Figure 9. Time of conformance and non-conformance with ASR

Figure 10 demonstrates the conformance monitoring when we have no speech recognizer available, i.e. we do not know the advised target values. In this case one third of the time the aircraft are not conform.

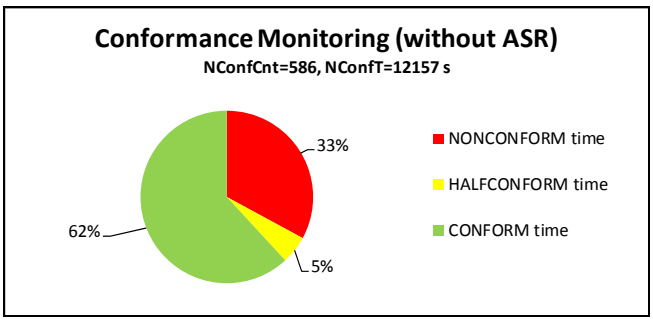


Figure 10. Time of conformance and non-conformance without ASR

B. *Passive shadow mode trials for Cologne/Bonn TMA*

Cologne/Bonn normally uses runway 14L and runway 24 with preference to 14L (approx. 95%). Therefore speech recognition can help to early use the information that the controller has changed from the default runway 14L to 24. Traffic flow in Cologne/Bonn is of course lower than in Frankfurt and no trombone or fan patterns are implemented for path stretching. From time to time, however, the traffic situation requires vectoring to loose time (see Figure 11 for a traffic sample).

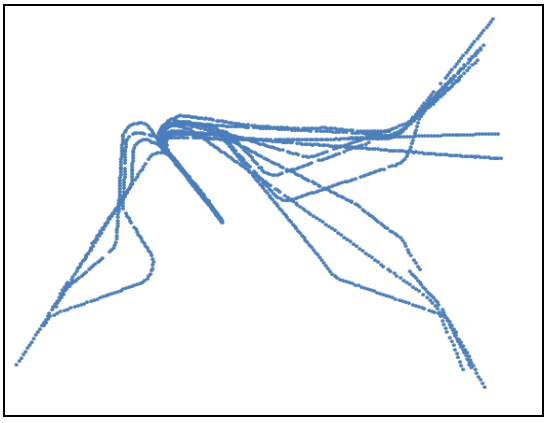


Figure 11. Radar plots of inbound traffic to 14L to Cologne

Figure 12 shows such a situation at the initial approach fix NOR. Currently HLX7JX (no. 3) is approx. 100 knots faster than BER364 (no. 4). Therefore the AMAN suggests a sequence with HLX7JX before BER364. The controller, however, prefers another sequence (BER364 before HLX7JX). Therefore, HLX7JX gets a heading advisory of 70 degrees. As soon as the AMAN knows this heading advisory no further guessing concerning the heading of HLX7JX and the sequence is necessary.

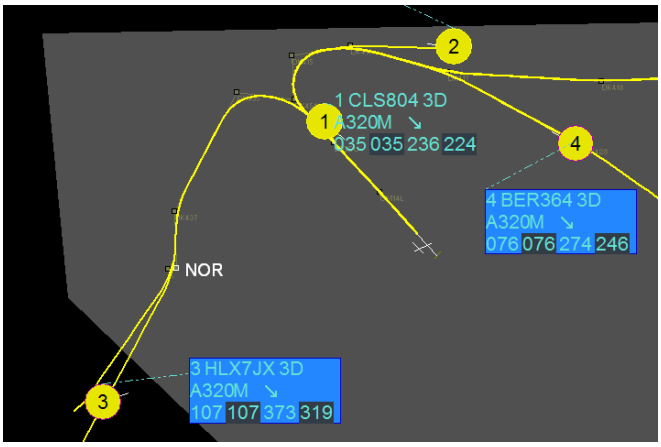


Figure 12. ASR helps to early detect sequence change of no.3 and no.4

Due to low traffic the sequence stability is better than for Frankfurt passive shadow mode trials (even without ASR). Therefore speech recognition only helps in a few cases (apart from the mentioned runway changes and vectoring procedures). Conformance monitoring provides similar results. Details will be published in [20].

C. *Human-in-the-Loop Simualtions (HITL)*

During the human-in-the-loop simulation the controller had no access to the recommended advisories of the AMAN. The output of the speech recognizer was used to control the simulation, i.e. ASR replaced the pseudo-pilots. In a first run 4D-CARMA used the information of the speech recognizer to update its planning (run-A). In a second run the planning was only based on radar signals. The actual controller advisories were no input to 4D-CARMA (run-B). We also used the recorded radar data and controller commands from run-A and run-B to reproduce both simulation runs (with and without ASR) with updated versions of 4D-CARMA (run-A1, A2, B1 and B2). In all reproduced runs 4D-CARMA worked in passive shadow mode. In run-A1 and B1 the recorded radar data and controller commands of run-A resp. B were the input for 4D-CARMA. In run-A2 and B2 4D-CARMA did not receive the advisories as input.

- Run-A: Original run (with ASR output as planning input),
- Run-B: Original run (without usage of ASR output),
- Run-A1/B1 – Reproduced runs with updated version of 4D-CARMA (with ASR output as planning input),
- Run-A2/B2 – Reproduced runs with updated version of 4D-CARMA (without ASR output as planning input).

We therefore evaluated 4 runs. In the following text we combine run-A1 and run-B1 under the headline “with ASR” respectively run-A2 and run-B2 under “without ASR”.

The results of the conformance monitoring show that the controller’s and 4D-CARMA’s intent very often match if and only if the arrival manager has access to the output of ASR (Figure 13 and Figure 14).

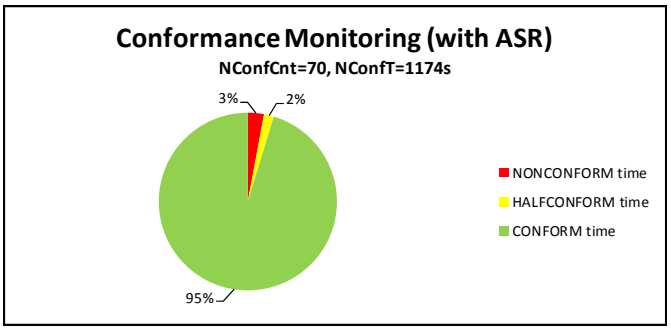


Figure 13. Conformance time for HITL with ASR (run A1 + B1)

Without using ASR the aircraft significantly deviated 170 times from the planned trajectory whereas with ASR only 70 deviations occurred.

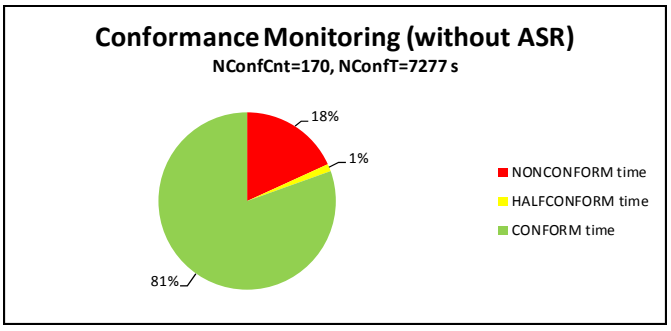


Figure 14. Conformance time for HITL without ASR (run A2+ B2)

Concerning sequence stability ASR offers only slight benefits, because the controller only deviated in one case from the planned sequence of the AMAN, see Figure 15.

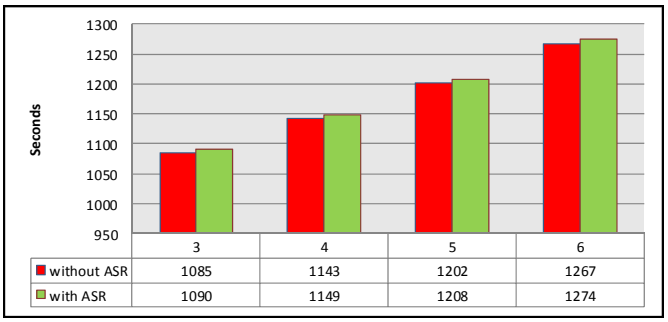


Figure 15. Average values for SS-3, SS-4, SS-5, SS-6 (N=13, both times)

In the situation in Figure 16 there was a gap between QFA764 (no. 3) and DLH645 (no. 4). Therefore the controller decided to change the sequence, i.e. DLH645 before QFA764. With ASR the sequence update happened 15 seconds earlier than without ASR. The higher value of SS-6 compared to SS-3 does not mean that a subsequence with 6 elements is more stable than a subsequence with 3 elements. The value of SS-i strongly depends on the number of aircraft in the subsequence, because the sequences are very stable and therefore, SS-i is mostly determined by the remaining flight time of the i^{th} aircraft, when the last deviation of the controller from the planned subsequence occurred.

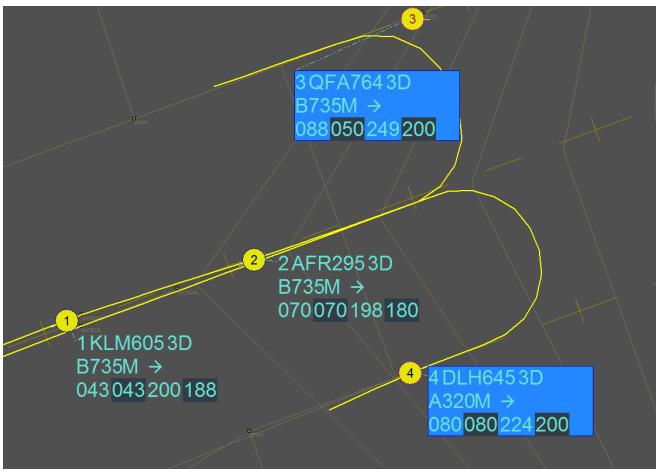


Figure 16. Controller deviates from plan by turn command to DLH645

VII. Next Steps

On the one hand the previous chapters have shown that the context information of an assistant system can significantly improve the performance of ASR. On the other hand an assistant system like an AMAN can benefit from the additional sensor of the speech recognizer. The next steps will be to bring the whole system into a real operational environment. A prototype application was demonstrated at the ATC Global fair in Amsterdam in 2011. This prototype, which did not consider non-functional requirements and used a limited grammar, will be extended and validated in the AcListant™ project (Active Listening Assistant), which will start in February 2013 with duration of two years. Current partners are DLR, University of Saarland, German Flight Services DFS and a Venture Capital Investor. The objectives are to improve the demonstrator, to validate the demonstrator in a real application environment at a German airport, and to prepare a know-how-transfer.

We showed that WER was reduced from 2.8% to 0.5% if context information is available. In a real application with long advisories sequences like "*Good morning DLH456, identified, Descend FL 100, Reduce to 250 knots after passing Gedern and turn left heading 250*" we expect higher word error rates. More import than recognizing each single word is of course the concept error rate, i.e. we need the information "*DLH456 Descend FL100*", "*DLH456 Reduce 250 after GED*" and "*DLH456 Turn Left Heading 250 after GED*". Currently the AMAN has no additional sensor, i.e. it is guessing the controllers advisories. We can assume that the support system only guess correct in 50% or even less of the cases. Therefore, concept error rates of 10% or less will already be a significant improvement. Validation will show which error rates will result in a workload reduction respectively, which will increase controller workload and therefore, will be rejected by the controllers.

VIII. Conclusions

One reason for the huge difficulties of introducing higher levels of automation in the ATM world refers to the intensive use of spoken language. Mainly two aspects have to be taken into account. The first aspect addresses the parallelism of the world of the situational knowledge between the operators and the one between the operators and the systems. This aspect is based on the fact that in the operator's world a common picture of the situation, including the intentions, is created by direct communication and listening to the communication of the others. In the machine world the picture of the situation is based on sensor information without any knowledge about the intentions of the operators. This difference in the end creates misunderstandings between operators and systems which lead to failures and further on to a lack of acceptance for automation. The second aspect regards the transition from today's procedures to highly automated future ones which may be executed in huge steps with considerable reduced voice communication. These steps may, however, be too big to be acceptable for the operators, may create a problem according to the integration of General Aviation and may create huge investments to retrofit old aircraft. One solution for these problems is the introduction of speech recognition as an integral part of automation following the "motto": Only those who can listen in can be in the loop!

Speech recognition is not a new technology. It is used with a considerable success in aviation e.g. in ATM training simulators. If the recognition rate is good enough speech recognition could be even an enabler for the introduction of higher levels of automation. This results in the need to use the situational picture and the knowledge of the assistant system to create dynamic speech hypotheses, which provide the needed high recognition rates. We showed that the dynamic context information provided by an AMAN can reduce error rates by a factor of 5.

Speech recognition provides an additional sensor for a controller assistant tool which reduces down-time by 35 seconds, i.e. the time the pictures of the situation in the machine's world do not match to the picture in the controllers' world. As the speech supported assistant system can be seen as an upgrade for existing assistant systems, like AMAN or DMAN, the hurdles for an introduction should be low. Furthermore, the used basic technologies are well known hence the time frame for a transition into practice can also be low, i.e. quick gains can be achieved.

On the one hand dynamic context information provided by an AMAN improves the performance of ASR; on the other hand ASR improves the performance of an AMAN. Both effects can increase acceptance of advanced controller assistant tools.

ACKNOWLEDGMENT

We would like to thank Todd Shore who creates the basis for the results, presented in section III, in his master thesis [26].

REFERENCES

[1] ACARE, "Realizing Europe's vision for aviation – Strategic Research & Innovation Agenda (SRIA)," Vol. 2, Sep. 2012.

[2] AMIDA, "Augmented Multi-party Interaction with Distance Access," http://www.ercim.eu/activity/projects/amida.html, 2009.

[3] Avibit "AMAN OPTATMOS," http://www.avibit.com/Solutions/OPTAMOS.htm.

[4] Barco ,"AMAN OSYRIS," http://www.barco.com/en/product/1229.

[5] S. Ciupka, "Siris große Schwester erobert die DFS," in German, Engl. title „Siris big sister captures DFS," transmission, Vol. 1, 2012.

[6] DFS, "AMAN 4D Planner," http://www.dfs.de/dfs/internet2008/module/worldwide_solutions/deutsch/worldwide_solutions/download/4d_planner.pdf.

[7] DialRC, "The Dialog Research Center,", http://dialrc.org

[8] K. Dunkelberger, and R. Eckert, "Magnavox Intent Monitoring System for ATC Applications," Magnavox, 1995.

[9] Eurocontrol, "All Clear? The path to clear communication. ICAO Standard Phraseology A Quick Reference Guide for Commercial Air Transport Pilots," http://www.skybrary.aero/bookshelf/books/115.pdf, 2011.

[10] Egis-Avia, "*AMAN MAESTRO*," http://www.egis-avia.com/products/ATC-Systems

[11] European Commision, "Flightpath 2050, Europe's Vision for Aviation Maintaining Global Leadership & Serving Society's Needs -- Report of the High Level Group on Aviation Research," 2011.

[12] FAA, "2012 National Aviation Research Plan (NARP),", March 2012.

[13] W. Gerling and D. Seidel, "Project 4D-Planner," Scient. Seminar, Braunschweig, 2002.

[14] 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.

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

[16] N. Hasevoets, and P. Conroy, "AMAN Status Review 2010," Eurocontrol, Edition number 0.1, 17 December, 2010.

[17] H. Helmke, R. Hann, M. Uebbing-Rumke, D. Müller, and D. Wittkowski, "Time-based arrival management for dual threshold operation and continous descent approaches," 8th USA/Europe ATM R&D Seminar, 29. Jun. - 2. Jul. 2009, Napa, California (USA), 2009.

[18] IATA, "Phraseology -- Pilots/Air Traffic Controllers Phraseology Study,", 2011.

[19] D. Jurafsky and J. H. Martin, "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition," 2nd edition. Englewood Cliffs, NJ, USA: Prentice-Hall, 9th Feb. 2008.

[20] M. Kleinert: "Integration of speech recogntion and controller assistance – quantification of the benefits of controller intent recognition", german title "Integration von Spracherkennung und Lotsenassistenz - Quantifizierung der Vorteile einer Lotsenabsichtserkennung", bachelor thesis, University of Applied Science Ostfalia Wolfenbüttel, March 2013.

[21] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," in: Soviet Physics -- Doklady 10.8, Feb. 1966.

[22] A. Ljolje, F. Pereira, and M. Riley, "Efficient general lattice generation and rescoring," in Eurospeech 1999, Sep. 1999, pp. 1251–1254.

[23] R. Parasuraman, T.B. Sheridan, and C .D. Wickens, "A model for types and levels of human interaction with automation," IEEE Transactions on Systems, Man and Cybernetics, Volume 30, Issue 3, 2000, pp. 286-297.

[24] 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.

[25] T. Shore, Fr. Faubel, H. Helmke, D. Klakow, "Knowledge-Based Word Lattice Rescoring in a Dynamic Context," Interspeech 2012, Sep. 2012, Portland, Oregon.

[26] T. Shore, "Knowledge-based word lattice re-scoring in a dynamic context," master thesis, Saarland University (UdS), 2011.

[27] SESAR Consortium, "The ATM Target Concept, D3," Ref: MGT-0707-010-01-00, Sept. 2007.

[28] SmartWeb: http://www.smartwebprojekt.de/start_de.html

[29] 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.

[30] C. Weinstein, "Opportunities for Advanced Speech Processing in Military Computer-Based Systems," Proceedings of the IEEE, Vol. 79, No. 11, 1991.

AUTHOR BIOGRAPHY

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, and started to work on different Expert System applications. Since 1999 he concentrates on Controller Assistant Systems, especially Arrival Management being responsible of the Arrival Manager 4D-CARMA. Prof. Helmke is an assistant professor for Computer Science since 2001. He is author of several text books on software engineering and programming languages.

Heiko Ehr is Mathematic-technical Assistant. He joined DLR's Institute of Flight Guidance in 1997. Mr. Ehr worked on several projects like 4D-Planer and its successor 4D-CARMA, Flow Monitor and Traffic Monitor. He is an expert in system integration and interface development. Additionally he is the IT-Manager of the department of Controller Assistant.

Matthias Kleinert is a student of Computer Science at the Ostfalia University of Applied Science, Wolfenbüttel, Germany since 2009. After he finished his semester abroad at the University of Wisconsin Parkside, in Kenosha, USA he joined DLR's Institute of Flight Guidance in 2012 to finish his bachelor thesis on controller intent recognition. Currently he is supporting the development of DLR's advanced research prototype 4D-CARMA.

Friedrich Faubel received his Diploma degree in Computer Science from the University of Karlsruhe, Germany, in 2006. After a six month research stay at Carnegie Mellon University in Pittsburgh, USA, he worked at Deutsche Telekom Laboratories in Berlin, Germany, until October 2007. He finally joined the Spoken Language Systems group at Saarland University, Germany, in late 2007 where he is currently pursuing his PhD in engineering. From April to July 2009 he had a research stay at the Centre for Speech Technology Research in Edinburgh, UK.

Dietrich Klakow studied Physics from 1987 until 1991 at the Universities of Erlangen (Germany) and York (UK). He completed his PhD at the University of Erlangen in 1994. Until 1996 he did a post doc at the Weizmann-Institute in Israel. In 1996 he changed to the area of speech and language research, joined the Philips research lab and he was holding a lecturer position at Aachen University since 1999. Since May 2003 Prof. Klakow is professor at Saarland University in Saarbrücken (Germany) where he builds up a new research team which deals with algorithms for the human machine interaction. His team scored first in the speech separation challenge. The other research focus is statistical natural language processing in particular question answering. The system built by the group was one of the top ranking systems in past TREC Question Answer competitions. In 2011 he received a Google Research Award.