Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics AHFE 2014, Kraków, Poland 19-23 July 2014
Edited by T. Ahram, W. Karwowski and T. Marek

Air Traffic Controller Support by Speech Recognition[1]

Oliver Ohneiser, Hartmut Helmke, Heiko Ehr, Hejar Gürlük, Michael Hössl[2], Thorsten Mühlhausen

Institute of Flight Guidance
German Aerospace Center (DLR)
Lilienthalplatz 7, 38108 Braunschweig, Germany

Youssef Oualil, Marc Schulder, Anna Schmidt, Arif Khan, Dietrich Klakow

Department of Computational Linguistics and Phonetics
Saarland University (UdS)
Campus, 66123 Saarbrücken, Germany

ABSTRACT

Air traffic controllers (ATCO) are a core element of the flight guidance process. Decision support systems with accurate output data could increase the controllers' performance or reduce their workload. Nowadays radar data based identification of controllers' intent causes delays of up to 30 seconds. The intents could be predicted better and earlier if their spoken commands would be taken into account. Therefore, the project AcListant® combines an arrival manager (AMAN) with an automatic speech recognizer (ASR). Spoken commands are automatically recognized and forwarded to the AMAN. The AMAN updates its plan (e.g. sequence and aircraft trajectories). The ATCO also receives a direct feedback of ASR recognition performance via an (optional) visual interface. The ATCO gets better support from the assistance system without additional personal effort. This has an impact on the controller's work and behavior like adhering closely to the radio telephony procedures and articulating more clearly. Together with the suggestion of most probable commands depending on the air traffic situation, this would lead to higher speech recognition rates and result in better AMAN assistance. Current results of a pre-study performed by the Düsseldorf Approach Area foreshadow significant positive effects of ATCO support by ASR.

Keywords: Air Traffic Controller, Automatic Speech Recognition, Situation Data Display, Speech Recognition Log, Arrival Manager, Context, AcListant®

INTRODUCTION

Today controllers communicate with pilots via radio telephony (R/T) for flight guidance of aircraft. They are responsible for the safe, punctual, cost efficient, and eco-friendly processing of air traffic movements within their

[1] The work was conducted in the AcListant® project, which is supported by DLR Technology Marketing and Helmholtz Validation Fund. The DLR colleagues were responsible for arrival management system and validation purposes, UdS colleagues provided the automatic speech recognition.

[2] Michael Hössl now works as a first officer on a Boeing 737 aircraft.

assigned airspace. Air traffic controllers obtain information about the status of the air traffic from a radar screen, which displays each aircraft with a configurable label containing information like speed or altitude. Controller commands and clearances via radio telephony for each aircraft are recorded on flight strips, which can either be paper or digital. According to the phase of a flight, there are several controller working positions in order to assure maximum safety. Approach controllers for instance are in charge of the final approach phase of flight. Their flight guidance can be supported by an arrival manager (AMAN), which plans efficient sequences based on radar data input and generates advisories for efficient approaches. The AMAN creates the optimal sequence and trajectories depending on flight status and position of each aircraft. Plan deviations can occur when the controller generates another sequence or "vectors" an aircraft in order to accommodate air traffic management or airline needs. The AMAN then detects these deviations from the radar data resulting from the flown aircraft trajectory. This leads to a new planning cycle. As the clearances given by the controller are unknown to the AMAN (especially the controllers' intent), the plan possibly has to be updated several times until the flight status (e.g. flight level, speed etc.) of the corresponding flight remains stable. This results in a delayed recognition time of the controller intended plan by 30 seconds or even more. For the controller this reduces the benefit of the AMAN.

It is expected that automatic speech recognition (ASR) in the ATC (Air Traffic Control) environment will improve the usage of an AMAN (Helmke et al., 2013), which in turn provides the controller with timely and stable support (see Figure 1). Although the usage of data link in ATC is discussed at least since the 90s, voice communication will definitely remain a pillar of air traffic control. The Strategic Research & Innovation Agenda (SRIA) of ACARE (ACARE, 2012) or Flightpath 2050 (European Commission, 2011) do not expect a fully automated ATM (Air Traffic Management) system in the next decades.

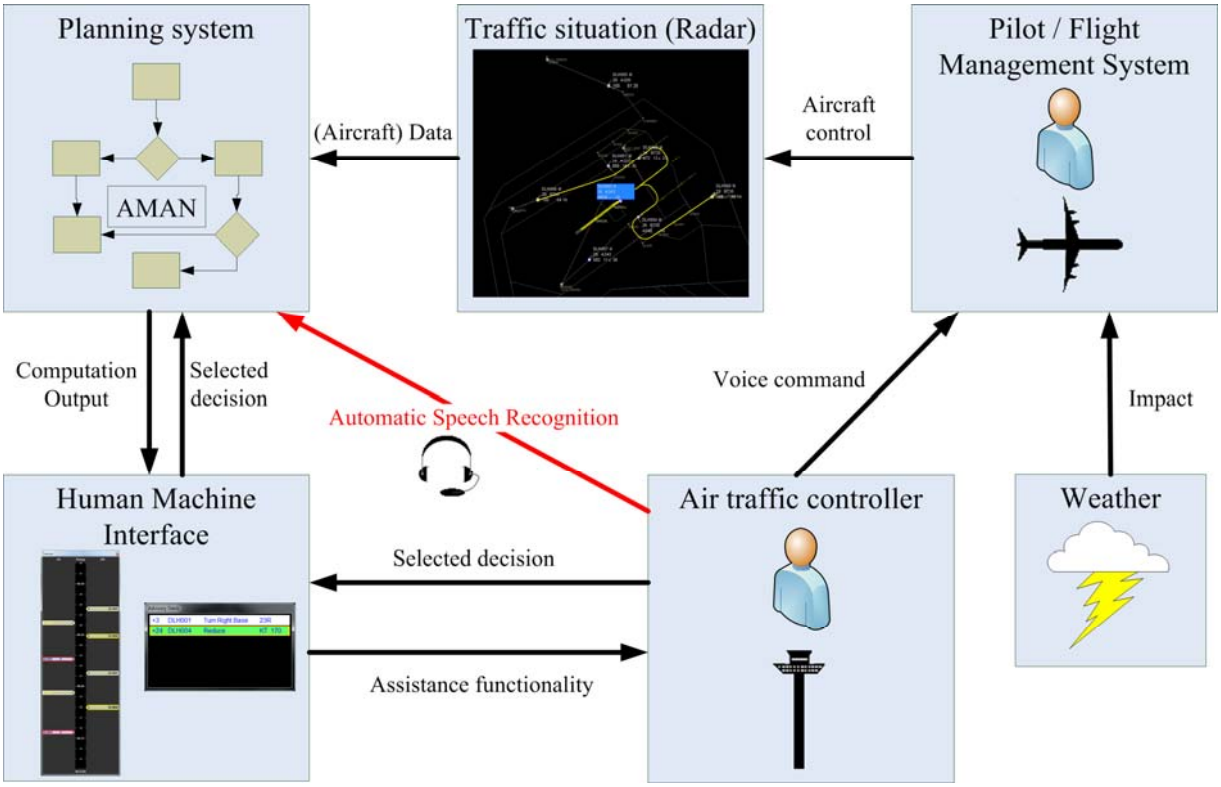

Figure 1. Today's operational procedure with standard AMAN: Interaction of air traffic controller with pilot and machine environment (modified from Elzer, 2001); red arrow: new, with an additional sensor for automatic speech recognition

The following chapters begin with the operational concept of using an ASR-supported AMAN. In addition, the recording of ATCO voice data, the necessary transcription and the extraction of concepts are described. To achieve small error rates, the generation of air traffic context information as well as the inclusion of this context in automatic speech recognition is required. This is explained in two further chapters. Afterwards the visualization of ASR's output in an HMI is detailed. The pre-study evaluating the context-dependent AMAN supported ASR is outlined whereas the last chapter summarizes the work and gives an outlook on future steps of the project.

OPERATIONAL CONCEPT FOR USING AMAN WITH ASR

As described above, the delayed update of the AMAN can be solved by the use of ASR. An assistant system which actively follows the human communication solves this task by analyzing controller-pilot communications and incorporating it as additional sensor information. The project AcListant® follows this new approach, by integrating ASR into the assistant system of an AMAN. The AMAN-ASR-system transforms speech into text and extracts the semantic intent (e.g. descending to a flight level). Listening to controller commands in ATM requires a highly reliable speech recognizer. Therefore, the approach makes explicit use of dynamic context information being created by the AMAN and used by the ASR system. This context information is derived by the AMAN from its other input sources (e.g. radar data, flight plans). It informs the ASR of expected and possible controller advisories, i.e. it continuously creates context information. Consequently, these hypotheses help the speech recognizer to detect spoken commands with improved reliability. Based on the extracted information from the controller-pilot communications the assistant system itself can more quickly adapt its own model, i.e. its knowledge concerning possible future system states.

The described approach solves the problem of delayed system reaction whereby optimal planning is ensured at all times. The better the system is informed about the airspace situation as well as about operator intents, the better the quality of the planning as a combined effort of the AMAN and the human air traffic controller becomes. The improved planning results directly in improved support functionality of the assistant system.

Besides that, the new ASR-based assistance system, AcListant®, operates in a seamless mode, i.e. controllers are not disturbed in their natural way of work. The speech recognizer operates in the background, which means the submission of the detected commands to the AMAN will not be noticed by the controllers. However, it is possible to display the detected commands to the controllers to indicate the performance of the speech recognition and hence to assist the controllers in their understanding of the system. Figure 2 schematically depicts the layout of the AcListant® controller's working position without this additional display. Beneath the situation data display, the flight strip bays are located, depicted as blue rectangles. The blue solid line beneath the flight strip bays merely symbolizes the seamless design of AcListant®, i.e. the method of operation does not change for the air traffic controller. The controllers are actually not aware of the ASR and its coupling to the AMAN, meaning the air traffic controllers operate in their daily routine using R/T and flight strips in order to guide the air traffic.

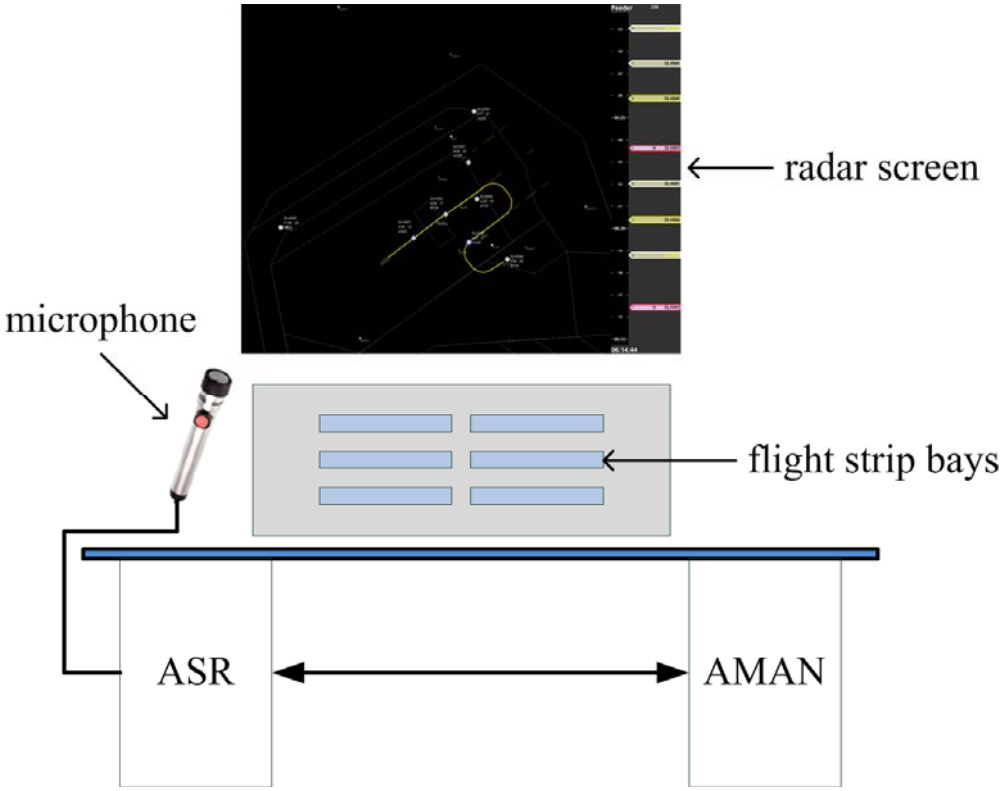

Figure 2. Schematic layout of AcListant® controller working position

The AMAN receives recognized commands from the ASR. After a first plausibility check, missing information which is not available in the speech recognizer process is added by the AMAN, e.g. for the command "SPEED 180"

the AMAN compares the speed value with the aircraft's actual speed and assigns the command type REDUCE or INCREASE. The same procedure happens to the command "FLIGHT LEVEL 80". The AMAN determines by the actual altitude value whether a DESCEND or CLIMB command has been issued. "TURN HEADING 270" can be transformed to TURN_RIGHT_HEADING or TURN_LEFT_HEADING if the present heading of the aircraft is known. If the command is stated as correct, it immediately influences the actual planning cycle, e.g. sequence and trajectory are updated. Derived from the trajectory, the distance-to-go and the distance-to-top-of-descent are computed.

As mentioned above, a plausibility check is performed on the recognized commands. Wrong ASR commands may be detected if the aircraft's radar data do not fit the recognized command. If the assumed value of a command type is not reached within a certain time, the command is ignored again by the AMAN. This is the case when the actual time is greater than the command time plus an estimated operation time, e.g. an assigned heading change of 30 degrees should be achieved after at most 40 seconds. All valid commands could be considered for the next computation cycle as far as their values do not exceed certain limits (see chapter "VISUALIZATION OF RECOGNITION OUTPUT IN AMAN HMI").

VOICE DATA RECORDING, TRANSCRIPTION AND CONCEPT EXTRACTION

Achieving good command recognition rates is a basis for delivering correct feedback for an AMAN and, therefore, increasing user acceptance (Hah and Ahlstrom, 2005). The automatic recognition of spoken natural language is difficult for a variety of reasons. Machines process sound very differently than humans and successful recognition of words depends on a variety of factors. Apart from volume, speed and clarity of articulation, the accuracy of machine recognition is also influenced by the speaker's gender, physical and psychic constitution (e.g. stress or fatigue), whether they are native speakers of the language and which dialect they speak. Hence, every single word can be produced in ways, although they all mean the same. In fact, even a single speaker will never say a word exactly the same twice. The automatic speech recognizer thus has to learn from a great amount of training data to successfully extract patterns of pronunciation from individual renditions of sound. The Wall Street Journal WSJ0 corpus (Paul and Baker, 1992) is a commonly used source for training speech recognition systems for American English. It comprises eight hours of recordings by 82 speakers reading articles from the stock exchange journal.

Although this corpus, together with some similar corpora (Hofbauer et al., 2008), is a good basis for speech recognition training, the phraseology and speaking habits of air traffic controllers are different. The scope of content during communication with pilots is limited and regulated. The ICAO (International Civil Aviation Organization) air traffic communication regulation defines a standard (ICAO, 2007). Every radio contact between controller and pilot must follow the defined rules and syntax. However, in practice there are many deviations (Karlsson, 1990). Also, with their individual experience every controller has different flight guiding preferences. The form and semantics of the speech input therefore even deviates at the same controller working position.

Contrary to the specified limited language area, so called out-of-grammar phrases often occur in an ATCO command. Voice communication usually starts with a greeting like "hello" or "bonjour" and ends with "bye" or "tschüs" in different languages. In addition sounds of clearing one's throat and hesitation sounds like "um" or "ähm" may disturb the understandability of commands. As those elements are both completely out-of-grammar (i.e. the word is not part of any allowed sentence) and not part of command-critical phrases, they can be filtered out quite reliably. It is more difficult to interpret deviations (Zokić et al., 2012) in command-critical phrases, e.g. saying "reduce your speed one eighty" instead of the official version "reduce speed one eight zero knots" both introduces illegal words ("eighty") and omits required ones ("knots"). Another challenge is the use of acceptable words in incorrect contexts, e.g. "100" may be spoken as "one hundred" instead of "one zero zero" in flight levels, but not when part of a callsign.

To collect realistic and extensive data (De Cordoba et al., 2006) concerning the concrete test field Düsseldorf approach, we work together with the DFS (Deutsche Flugsicherung GmbH). We recorded several hours of voice communication and corresponding radar data during a simulation of Düsseldorf approach at their place of business in Langen. The speech data was recorded with a headset and a sample rate of 16,000 Hz regarding the "wav"-files. The quality of the recordings varied with the correct and stable "connection" from mouth output to headset input. The complete voice stream was then manually cut into short audio files of individual commands.

During simulation runs in the DLR's ATMOS (Air Traffic Management and Operations Simulator) shown in Figure 3 an input trigger was used for activating ASR recording, capturing start and end times of commands and splitting different commands. The trigger could be set via a foot switch or microphone button, preferably in a way that does not add to the cognitive load of the controller.

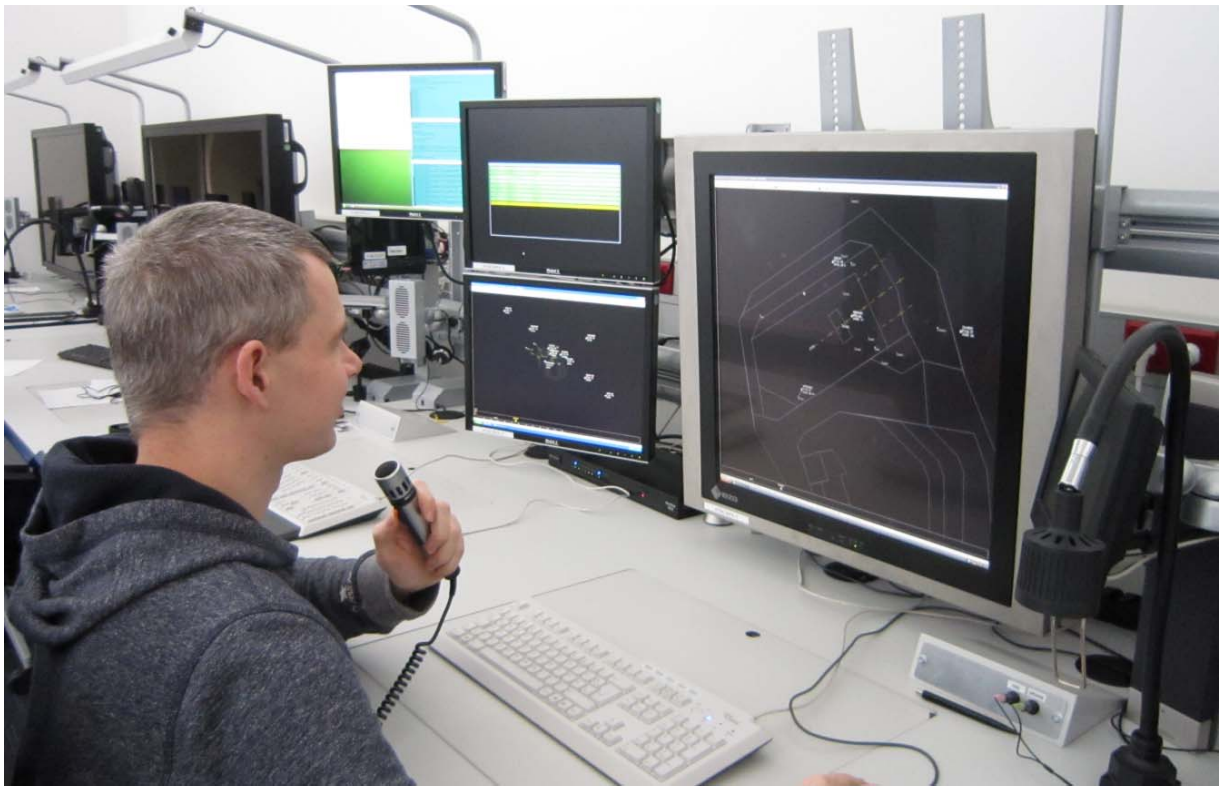

Figure 3. Simulation of Düsseldorf approach with recording of voice and radar data

A large amount of voice data has to be prepared for further electronic processing in different steps. The recorded utterances need to be transcribed as text and be annotated with additional information. The tool "Annotator" (see Figure 4) assists the human auditor with annotating commands.

The upper left window Name shows a list of all command files found in the specified directory. The color of the file names symbolize if the annotation of the utterance was already done (black) or not (blue). Next to it the amplitude and duration of the currently selected recording are shown. The lower half of the window contains the ASR system's prediction on the left and corrected version of the human auditor on the right. Each side consists of fields representing the utterance in plain text, XML (Extensible Markup Language) syntax and as AMAN commands. The XML annotation represents semantic markers that allow easy machine extraction of information from free speech, e.g. which command is meant by a certain utterance or whether a series of digits is part of a flight number, a speed value, a height value, etc. Lastly, the commentary field in the upper right allows auditors to leave remarks about individual recordings, such as unresolved questions or special observations.

The ASR prediction offers a starting point for the auditor, who compares it to the audio recording and can copy it over to the transcription fields if correct. If the prediction contains mistakes, the auditor corrects these in the plain text field (Text transcription). To receive an automatic suggestion on how to annotate the corrected sentence with XML, the auditor presses the GetXML button. The result can again be revised if necessary. In a similar fashion, commands are generated, based on the XML text, by means of the GetCommand button. In cases where a single utterance contains more than one command, they are listed sequentially and each command is headed by the callsign given in the utterance. The resulting auditor-approved utterance representations (text, XML and command) are used either for training the ASR system or to evaluate its word error rate (WER) and command error rate (CmdER).

The annotation effort concentrates on the correct extraction of concepts. Callsigns, command types and the command values constitute concepts and the combination of those three concepts forms a command. Examples for callsigns are DLH437 "Lufthansa four three seven", BAW86V "Speedbird eight seven victor", UAE55 "Emirates five five" or LOT927 "Pollot nine two seven". The command type followed by a possible value is e.g. "Reduce 180 KT ('reduce one eight zero knots')", "Descend FL 80 ('descend flight level eight zero')", "Rate of Descent 1000 ft ('rate of descent one thousand feet')", "Direct DL455 ("direct to delta lima four five five")', "Turn Left heading 260 ('turn left heading two six zero')".

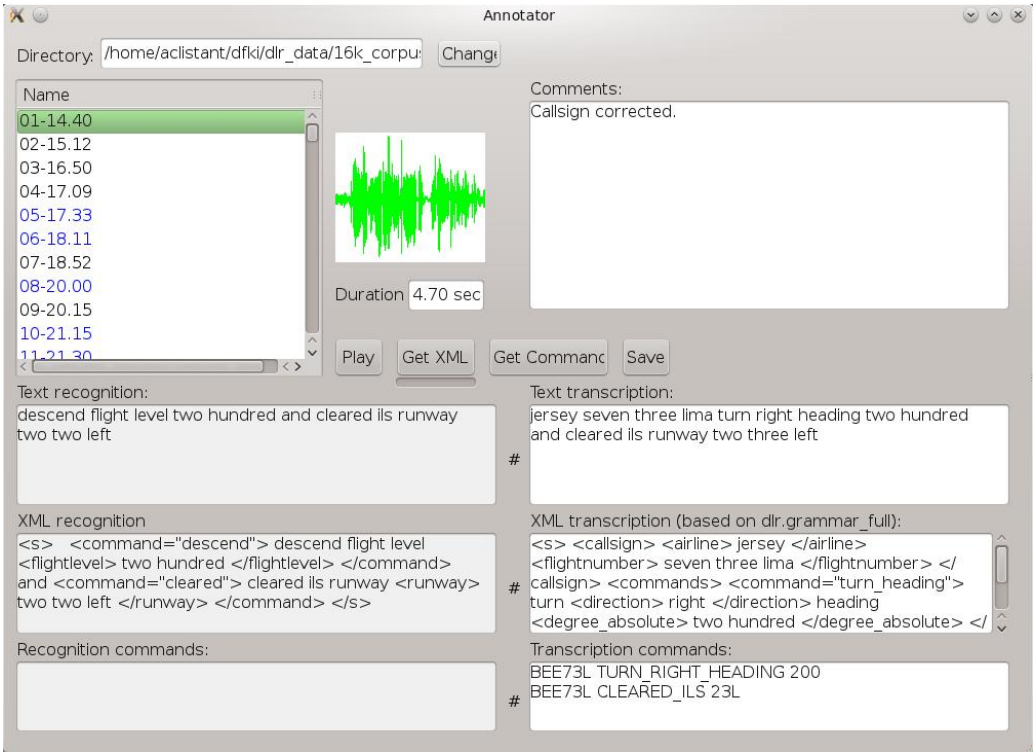

Figure 4. Annotation tool for extraction of concepts out of controller voice commands

GENERATION OF AIR TRAFFIC CONTEXT INFORMATION

Speech recognition for aviation applications benefits from the high level of standardization and procedural structure that is used for all commercial airline operations (ICAO, 2007). Especially the standardized grammar limits the number of possible word combinations. In an effort to further reduce the word error rate in ATC speech recognition, a reduction in the number of possible predefined commands for the speech recognizer to choose from could be beneficial. A smaller number of combinations imply a better chance of recognizing the correct terms. This limitation can be achieved by anticipating air traffic controller commands, i.e. by excluding commands which are not possible due to operational constraints from the context.

Radar flight track analysis and air traffic controller interviews represent the basis of controller command prediction, as real life air traffic control operations include a great number of variables that can cause the air traffic controller to deviate from published procedures. Reasons are e.g. different airplane types and their aerodynamic characteristics, payload, airline standard operating procedures, environmental considerations, weather and above all, the omnipresent necessity to keep approach flow high. Even though these deviations may seem random at first, it was shown (Hössl, 2013) that they do follow certain rules, ultimately allowing a prediction of possible controller commands. Typical deviations from published approach procedures are shown in Figure 5.

Düsseldorf Int. Airport, located in the busy airspace of the Rhine-Ruhr metropolitan area in Western Germany has been selected as example airport for this project. The basic principle of our considerations, however, is universally applicable. This is achieved by strict separation of generic and airport specific data.

Based on the aircraft's parameters, such as position, altitude, airspeed and heading, the amount of possible commands can be significantly reduced. The implementation of this concept is called Mapped Exclusion Method (MEM). The MEM is based on the idea that future ATC commands are primarily influenced by the aircraft's position and configuration. The MEM-function starts with the assumption that all commands are possible all over the approach sector, i.e. the set of possible commands for an aircraft contains all thinkable commands. Subsections of the airspace can be modeled as polygonal areas on a map. Areas have attributes such as preconditions (e.g. "Alt>50" means that only aircraft above 5000 feet are considered to be within the polygon) and restrictions (e.g. "Hdg(20,40)" means that all heading commands have to be within the range of 20° to 40°).

If an aircraft is located within an area and all preconditions are met, the set of possible controller commands is reduced by the restriction-attributes of that area, i.e. the previous set of possible commands is intersected with all restrictions of the area. If it should be required to add commands again in certain sub-areas, expansions can be added as an opposing function to restrictions. Finally the aircraft is checked against all areas and the set of possible controller commands is reduced by the restrictions of respective areas. As a result, only operationally likely commands are added to the list of possible instructions for each aircraft.

To facilitate the MEM, the mapping application JOSM of the OpenStreetMap (OSM) project was used (Ramm et al., 2010). Originally designed to create open source world vector maps, the JOSM interface serves the project's purposes for the allocation of the approach airspace into polygon areas required for the MEM (see Figure 6). These polygons can include properties which are used for preconditions, restrictions and expansions. As the main algorithm remains generic, only airport specific data needs to be created. Using JOSM there is no need of any coding skills of the user.

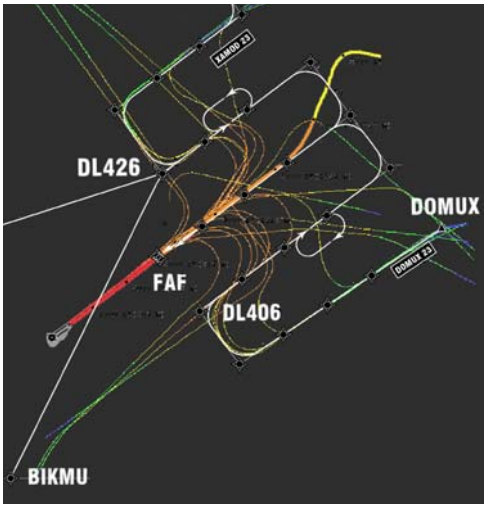

Figure 5. Aircraft flight tracks (colored lines) deviating from published procedures (white lines) at the approach into Düsseldorf Int.

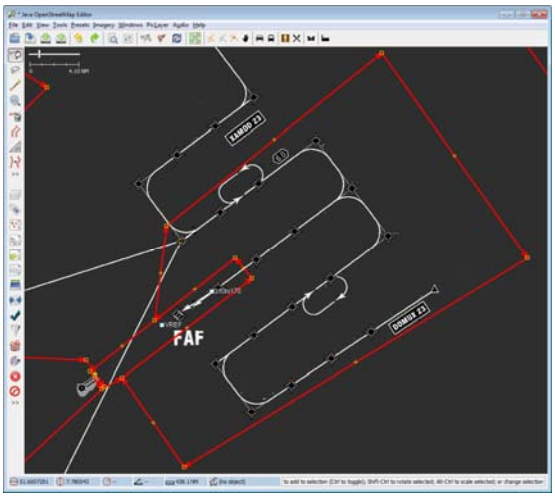

Figure 6. Polygon area (red lines) creation in JOSM

A conducted test series of the MEM algorithm showed an average context error rate of 0.9%, meaning 99.1% of commands issued by the controller in the test cases were predicted (982 out of 991). Additionally, the amount of possible commands could be reduced by an average of 77.3%, meaning only 22.7% of the total context remains within the algorithm's output (204,308 out of 899,577 commands).

AUTOMATIC SPEECH RECOGNITION BASED ON CONTEXT

The proposed Automatic Speech Recognition (ASR) system is based on Kaldi software (Povey et al., 2011), which is an emerging speech recognition toolkit that is freely available under the Apache license. This framework is very flexible and easy to extend to include new ideas and methods. In order to increase the performance of our ASR system, an acoustic model has been trained on five hours of in-domain data, collected in Air Traffic Control (ATC) simulations and annotated using the annotation tool described above.

This model is then combined with an ATC grammar (a language model) that defines the language rules and structure of the ATC commands. The grammar is designed in a scenario-agnostic way that allows easy adaptation for different airports, e.g. callsigns are defined as one airline (out of a list of several dozen), followed by any combination of up to five numbers or letters and digits, irrespective of which airlines and callsigns are actually encountered at the airport in question.

The combination of the acoustic and language model leads to a recognition network (a search space) defining all expected utterances and the different ways of pronouncing them. The recognized sequence of words is returned as the most likely path in this network. The concept extractor is then applied to this sequence to extract the spoken ATC command.

The scheme described above is rather naïve as it does not include any ATC knowledge, e.g. how frequently commands are spoken, how likely commands are, radar information, etc. The presence of such information can

strongly improve the recognition performance (Shore et al., 2012). This idea is integrated into our system using the context information (described in chapter "GENERATION OF AIR TRAFFIC CONTEXT INFORMATION") that is updated every 40-60 seconds. The received context is used to build a lattice (see example in Figure 7) that replaces the recognition network, strongly reducing both the number of possible callsigns and commands.

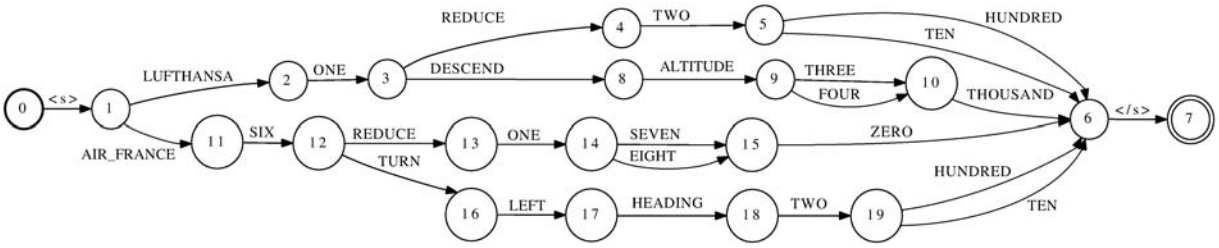

Figure 7. A simplified lattice resulting from the application of context to the recognition network. Each path represents a sentence that can be recognized by the ASR. Note that for the Lufthansa aircraft a reduce or descend command is allowed, while the Air France can receive reduce or turn. In the original network, any callsign could have received any command.

The search space for callsigns (airline and flight number) is limited to the aircrafts that appear on the radar, reducing it from millions of possible combinations to between ten and twenty. Similarly, the number of possible command-value combinations can be reduced to 30 to 40 per aircraft by discarding those that are not anticipated in the context information. Furthermore, the possible probability values that could be received with the context information can be used to distinguish between more likely callsigns and commands in the reduced search space itself, e.g. aircrafts that are close to the terminal are more likely to be addressed than ones that are still far away.

The proposed approach reduces the Word Error Rate (WER) by 47% relative, dropping it from 12.0% to 6.3%. More importantly, it reduces the Command Error Rate (CmdER), i.e. the number of commands with at least one mistake in callsign, command or value, by 75% relative, decreasing it from 61% to 15%. The gap between the improvement of the CmdER compared to the WER is due to the fact that the recognizer mostly commits errors while recognizing keywords such as digits, which define flight numbers and command values. As a result, spoken sentences that were correctly recognized except for a single digit will still be counted as completely miss-recognized in the CmdER, leading to this discrepancy in the two metrics.

VISUALIZATION OF RECOGNITION OUTPUT IN AMAN HMI

The human machine interface RadarVision visualizes output data of the database used by the arrival management system. RadarVision mainly serves as a prototypic situation data display for air traffic controllers. The largest part of the screen is filled with a conventional radar screen. The radar display includes all aircraft positions shown in a two-dimensional way completed by alphanumeric details on callsign, aircraft type, altitude and speed.

In addition to the current runways in use, significant waypoints, route segments and borders of important airspace areas can be shown. Beyond simply displaying the current situation, the calculated support functionalities of the AMAN can be visualized to enhance the radar screen as shown in Figure 8. This includes for instance the planned touchdown sequence, the computed trajectory for every aircraft or a distance-to-go prediction. The timescale on the right consists of aircraft labels assigned to a certain landing time on a specific runway. The whole scale is moving on downwards to the actual touchdown time as time goes by.

Regarding the automatic speech recognition process running in the background, the controllers have no additional work. They neither have to click or type anything nor need to change habits. Nevertheless, the ATCOs can choose to receive direct feedback about their voice input. The recognized command is therefore prepared for a so called speech recognition log (see Figure 9). This display stack visualizes the system's understanding of the controller's utterance. The speech recognition log also distinguishes between different categories of recognized commands. Every command is checked for plausibility and compared to the given context.

A command is only plausible if its type-dependent value lies within predefined limits within the air traffic approach environment. Speed commands span a valid range between 150 and 300 knots. Altitude commands must lie within the interval from 2000 feet to 400 flight levels. Rates of descent are valid in steps of 500 feet from 1000 up to 3500 feet per minute.

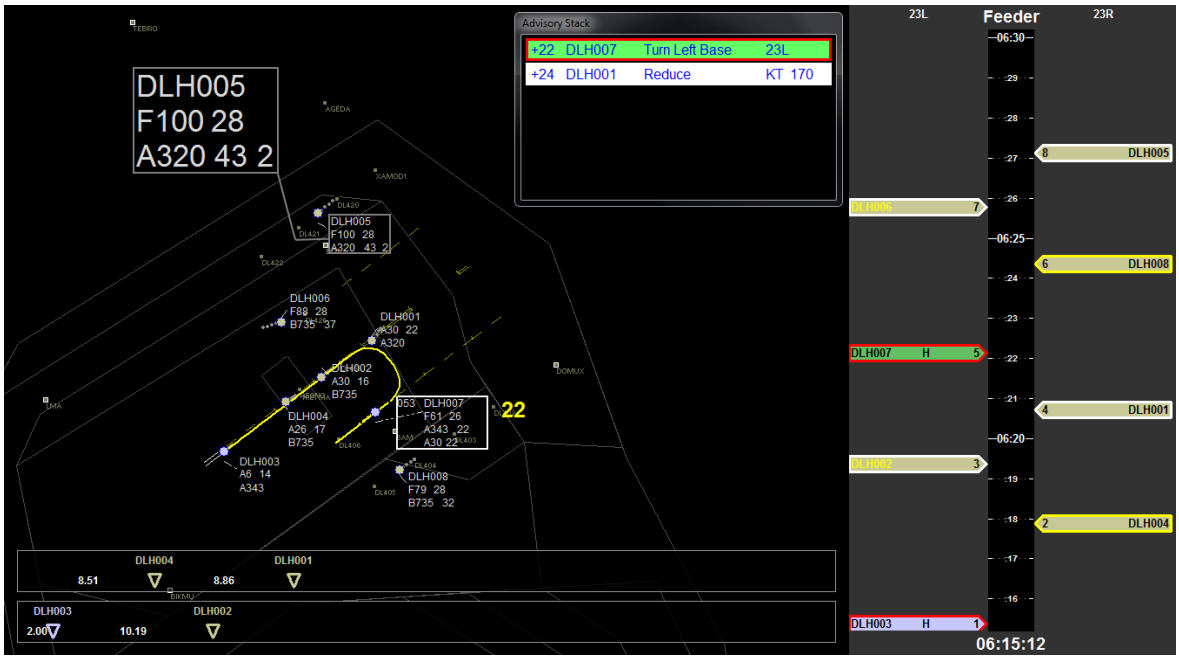

Figure 8. *RadarVision* shows a radar screen with AMAN data in the form of an advisory stack, a timeline with sequences and planned touchdown times, values for distance-to top-of-descent and threshold and a centerline separation range

A heading command consists of valid values between 0 and 360 degrees. In case of a *transition*, a *direct to*, a *handover* or a *cleared ILS* command, the respective route, waypoint, controller position or runway has to exist in the database of the modelled approach control area. During the further process, the combination of type, value and actual aircraft state is checked. An aircraft already flying a speed of 180 knots cannot reduce its speed to 220 knots. This speed value would only be possible when accelerating with an increase command. The same way a command with a descent to FL 90 is negated if the aircraft is already at flight level 80. Such commands will be assigned as invalid.

Speech Recognition Log

DLH2NC	Descend	Alt 3000
AUA151Y	Reduce	KT 150
UAE55	Descend	Alt 2000
BER630	Descend	FL 100
AUA151Y	Reduce	KT 150
DLH2NC	Reduce	KT 210
SXS1CF	Reduce	KT 320
AUA151Y	Reduce	KT 170
BER947V	Descend	Alt 2000
AFR1906	Descend	Alt 2000
BER2356	Turn Left Base	78L
DLH437	Descend	FL 270
BAW938	Descend	FL 110
DLH61A	Reduce	KT 150
UAE55	Descend	FL 70
AUA151Y	Descend	Alt 2000

Figure 9. Speech Recognition Log shows valid, out-of-context and invalid commands as a direct controller feedback

In the second phase of investigating plausibility, the AMAN's context is used. The command is compared with the context information regarding the current timeslot. If the command is not included in this context, the command is marked as "not in context". This could only happen if ASR is used without context information for demonstration and comparing purposes. All other commands that do not fit the range of invalidity or out-of-context are supposed to be valid.

All spoken and recognized commands of the last five minutes then are displayed in a vertical list on the speech recognition log (see Figure 9). Valid commands have a green, out-of-context commands a turquoise and invalid commands a yellow background. The color red is not used here, because a missed recognition should not provoke an alarming effect. During the first five seconds the font of the command text is bold and the background color brighter to highlight the last commands. By using the mouse-over functionality, the background of the aircraft is highlighted in purple (see example "AFR1906 Descend Alt 2000" in Figure 9). All corresponding information of this aircraft is also highlighted on the radar screen, the timescale and in other supporting windows. The recognition log allows the controllers to get direct feedback on how well their commands are understood by the system. Nevertheless, there is no need to repetitively check the evaluation of the recognition output. Therefore, current work attitudes are not intruded into. The visualization of the voice recognition outcome, however, may lead to a cognitive response on good and bad recognition. The personal awareness of a controller may be that strictly following the ICAO aircraft radio regulations causes good results in being automatically understood. However, the recognition rate should still be acceptable when making intentional deviations from the allowed syntax. Perhaps this presents an incentive to achieve better recognition rates by articulating more clearly and closer to the standard.

Another effect on the controller could be a better acceptance of the decision support system AMAN due to more accurate and up-to-date information particularly in high-workload (Cordero et al., 2012) air traffic situations. Low word error rates of the automatic speech recognizer also mean good command recognition quality. The AMAN analyses these mostly correct commands and includes them in the current calculation cycle. Discovering the controller's intent through this data should then lead to a better controller support. Relevant values would be more up-to-date regarding the true plan of the controller. The shown values, as for instance the distance-to-threshold for each aircraft, would become more reliable to the controller. The distance-to-threshold is often useful for controllers and pilots in situations of sequence rescheduling due to emergencies or unexpected behaviors. So controllers could experience some kind of benefits with the described better support by an assistance system using automatic speech recognition. Following this, their acceptance and active encouragement of support systems might rise. As a last consequence even their behavior could adapt to profit from an air traffic controller support system with speech recognition.

PRE-STUDY FOR VALIDATION OF CONTEXT-DEPENDENT SPEECH RECOGNITION SUPPORTED AMAN

Depending on the operational concept (described in chapter "OPERATIONAL CONCEPT FOR USING AMAN WITH ASR") and the correlated validation goals, six scenarios of 45 minutes each, which are clustered in two groups, are envisaged. These two groups differ in traffic and the exceptional event occurring during the run. However, both groups contain a base scenario with no controller assistance and an ASR-scenario with AMAN support. In addition, the first scenario cluster also contains an AMAN scenario with no additional input to validate the benefit of ASR. The second scenario cluster includes an AMAN scenario, where the AMAN can get additional input via mouse and keyboard. By including this scenario, the workload difference between manual input and the automatic input via speech recognition can be validated. Table 1 gives an overview of the six scenarios and their main characteristics.

Table 1: Scenario overview

Base scenario without AMAN	Traffic with 23 arrivals with one priority flight
Standard AMAN scenario	
ASR AMAN scenario	
Base scenario without AMAN	Traffic with 20 arrivals and a short runway closing
Manual input AMAN scenario	
ASR AMAN scenario	

To access the benefits described in the operational concept, several performance parameters describing the air traffic control and the speech recognition processes are analyzed (as listed in Table 2). These parameters include air traffic information, characteristics of spoken commands or different error rates which describe the quality of the speech recognition.

Table 2: Overview of validation goals and accessed parameter

Validation goals	**Accessed parameter**
Improved prediction quality of ETA (Estimated Time of Arrival)	Comparison of ATA (Actual Time of Arrival) and ETA for each aircraft
Improved vertical flight path	Comparison of flight time and trajectory length for each aircraft
Improved flight duration	
Reduced number of clearances	Number of clearances given
Reduced frequency utilization time	Accumulation of controller radio time
Earlier detection of sequence change	Time, when landing sequence is established in AMAN
Reduced error rate (at least 5%) already improves AMAN performance	
Higher degree of conformance of planned and actual flown trajectory	Deviation between radar position and AMAN position of each aircraft
ASR reduces workload due the absence of manual data entry	Number of mouse and/or keyboard inputs
Number of active controller corrections is reduced	Number of corrections
Time of active controller corrections is reduced	Accumulated time of corrections
Context information keeps response time of speech recognition process in acceptable boundaries	Computation time of speech recognition process with and without context information
Use of context information for robust Digits Error Rate (DER)	Calculation of DER
Context information improves Word Error Rate (WER) as well as Concept and Command Error Rate (ConER/CmdER)	Calculation of Error Rates
Even imperfect context information improves WER and ConER/CmdER	
Acoustic models for non-native speakers improve WER and ConER/CmdER	
Different acoustic models for male and female speakers improve WER and ConER/CmdER	
Increasing robustness against out-of-grammar utterances improves WER and ConER/CmdER	

During the pre-study for validation, which was conducted with a controller from the Düsseldorf approach sector, several aspects of improvement for the validation cycle were detected and implemented. These changes include scenario-related topics like the adaptation of the flight plan, the aircraft behavior or the look and feel of the controller working position, but also validation related topics like the implementation of a training scenario or the alignment of the validation storyboard. With these results the next validation cycle will be approached. The intention is to have three more validation cycles in total, with a steady increase in test persons. The final validation trials will be conducted in early spring 2015 with at least ten controllers.

CONCLUSIONS AND OUTLOOK

Air traffic controllers are responsible for the flight guidance process. They could be supported by decision support systems. The assistance quality of these systems depends on the accuracy of input data, e.g. the update rate of the radar information. Taking into account the controller's intent would improve the support quality. This intent could be derived by analyzing the controller-pilot voice channel via speech recognition. Therefore, the project AcListant® combines an arrival manager (AMAN) with an automatic speech recognizer (ASR). For training of the acoustic model of ASR, speech data is recorded and transcribed to extract flight guidance concepts. The ASR uses AMAN generated air traffic context information to achieve high command recognition rates. The recognition performance is visualized in an HMI as an optional direct feedback for the controller. Due to high speech recognition quality, the assistance system delivers better support. As voice communication is the usual task of an ATCO, no additional workload is caused, which highly increased the acceptance probability. The performed pre-study using the Düsseldorf Approach Area foreshadows that the controller's behavior could change positively with regard to articulation and radio telephony procedures. Higher speech recognition performance improves the AMAN assistance and vice versa.

As a next step it is planned to take the contextual likelihood of each possible air traffic command into account. This would further enhance the ASR performance. Final validation trials at the end of the project should reveal the effect of the AMAN-ASR combination on ATCOs.

REFERENCES

ACARE (2012), "*Realizing Europe's vision for aviation - Strategic Research & Innovation Agenda (SRIA)*", Vol. 2, Sep. 2012.

Cordero, J.M., Dorado, M., de Pablo, J.M. (2012), "*Automated Speech Recognition in ATC Environment*", International Conference on Application and Theory of Automation in Command and Control Systems (ATACCS'2012), ISBN 978-2-917490-20-4.

De Cordoba, R., Ferreiros, J., San-Segundo, R., Macias-Guarasa, J., Montero, J.M., Fernandez, F., D'Haro, L.F., Pardo, J.M. (2006), "*Air Traffic Control Speech Recognition System Cross-Task & Speaker Adaptation*", in: IEEE A&E Systems Magazine, Universidad Politécnica de Madrid, IEEE.

Elzer, P. (2001), "*Recent Developments in the Field of Human-Machine-Interfaces*", in: at - Automatisierungstechnik 49, No. 1/2001, pp. 15–29. – ISSN 0178–2312. – Oldenbourg Verlag.

European Commission (2011), "*Flightpath 2050, Europe's Vision for Aviation Maintaining Global Leadership & Serving Society's Needs -- Report of the High Level Group on Aviation Research*".

Hah, S., Ahlstrom, V. (2005), "*Comparison of Speech with Keyboard and Mouse as the Text Entry Method*", in: Proceedings of the Human Factors and Ergonomics Society, 49th Annual Meeting.

Helmke, H., Ehr, H., Kleinert, M., Faubel, F., Klakow, D. (2013), "*Increased Acceptance of Controller Assistance by Automatic Speech Recognition*", in: Tenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2013).

Hofbauer, K., Petrik, S., Hering, H. (2008), "*The ATCOSIM Corpus of Non-Prompted Clean Air Traffic Control Speech*", in: Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08), Eds.: Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Tapias, D., European Language Resources Association (ELRA).

Hössl, M. (2013), "*ATCpredict: Predicting Air Traffic Control Anticipation of Approach Controller Intentions at the Example of Düsseldorf Airport*", DLR-IB 112-2013/40, Bachelor Thesis, University of Applied Sciences Bremen.

ICAO (International Civil Aviation Organization) (2007), "*Doc 4444, ATM/501: Procedures for Air Navigation Services, Air Traffic Management, Fifteenth Edition*", Chapter 12: Phraseologies.

Karlsson, J. (1990), "*The Integration of Automatic Speech Recognition into the Air Traffic Control System (FTL Report R90-1)*", Flight Transportation Laboratory, Department of Aeronautics and Astronautics, M.I.T. Cambridge, Massachusetts.

Paul, D.B., Baker, J.M. (1992), "*The design for the wall street journal-based CSR corpus*", in: Proceedings of the workshop on Speech and Natural Language, ACL HLT '91.

Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K. (2011), "*The Kaldi Speech Recognition Toolkit*", in: IEEE 2011 Workshop on Automatic Speech Recognition and Understanding.

Ramm, F., Topf, J., Chilton, S. (2010), "*OpenStreetMap: Using and Enhancing the Free Map of the World*", UIT Cambridge, ISBN 978-1906860110.

Shore, T., Faubel, F., Helmke, H., Klakow, D. (2012), "*Knowledge-Based Word Lattice Rescoring in a Dynamic Context"*, Interspeech 2012, ISCA.

Zokić, M., Boras, D., Lazić, N. (2012), "*Say Again*", in: International Journal of Education and Information Technologies, Issue 4, Volume 6, 2012, International Association of Computer Science and Information Technology Press (IACSIT Press).